



การตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม



วิทยานิพนธ์เสนอปัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร

เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาบริหารคุณภาพเพื่อธุรกิจ

ปีการศึกษา 2562

ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต^๑
สาขาวิชาวิทยาการคอมพิวเตอร์
ปีการศึกษา 2562
ติดต่อที่เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง "การตรวจจับและรักษาความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม"

ของ สรวีญ์ พธ์พยอม

ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต สาขาวิชาการคอมพิวเตอร์

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการสอบวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.กันต์พงษ์ วรรัตน์ปัญญา)

ประธานที่ปรึกษาวิทยานิพนธ์

(ดร.วนัชรี มากรัม)

กรรมการผู้ทรงคุณวุฒิภายใน

(ผู้ช่วยศาสตราจารย์ ดร.เกรียงศักดิ์ เตเมีย)

กรรมการผู้ทรงคุณวุฒิภายใน

(ผู้ช่วยศาสตราจารย์ ดร.ธนาธร พอค้า)

กรรมการผู้ทรงคุณวุฒิภายใน

(รองศาสตราจารย์ ดร.ไกรศักดิ์ เกษร)

อนุมัติ

(ศาสตราจารย์ ดร.ไพบูล มุณีสว่าง)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม

ผู้วิจัย สรวิญ พิเชฐพยอม

ประธานที่ปรึกษา ดร. วันสุรีย์ มาศกรรัม

ประเภทสารนิพนธ์ วิทยานิพนธ์ วท.ม. สาขาวิชาวิทยาการคอมพิวเตอร์, มหาวิทยาลัยนเรศวร, 2562

คำสำคัญ การแปลงเรดอนหลายขนาด, การตรวจจับจุดภาพที่แสดงผลในรูปแบบเจ็ดส่วน, การรูปแบบเจ็ดส่วน, การรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน, การเพิ่มข้อมูล, การตัดแบบกำหนดช่วงบนแบบจำลองสี HSV

บทคัดย่อ

การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจุดภาพแหล่งอีดีนั้นมีความท้าทายโดยเฉพาะอย่างยิ่งในการค้นหาตัวเลขท่ามกลางวัตถุที่หลากหลายบนภาพซึ่งอาจมีส่วนประกอบหรือรูปทรงที่คล้ายคลึงกัน ดังนั้น ในงานวิจัยนี้จะนำเสนอวิธีการนำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนออกมายกจากภาพ โดยการตัดแบบกำหนดช่วงบนแบบจำลองสี HSV (Predefined HSV color slicing technique) เพื่อการตรวจจับ และประยุกต์ใช้โครงข่ายประสาทเทียมแบบconvoluzion (Convolutional neural network) ในการรู้จำ ทางผู้วิจัยนำเสนอวิธีการเพิ่มข้อมูลโดยการแปลงเรดอนหลายขนาด (Multiscale Radon transform) เพื่อลดปัญหาด้านข้อจำกัดของข้อมูลภาพสอนสำหรับโครงข่ายประสาทเทียม รวมถึงการลดปัญหาพฤติกรรมการถ่ายภาพซึ่งส่งผลให้ภาพเอียง มีขนาดภาพที่หลากหลาย มุ่งมองของภาพ และผลจากการตัดกระบทองแสงที่แตกต่างกัน ซึ่งอาจส่งผลต่อการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน และผลการทดสอบแสดงให้เห็นว่าวิธีการตรวจจับที่ผู้วิจัยนำเสนอ มีค่าความแม่นยำ และค่าการเรียกคืนตัวอักษรที่สูงที่สุดอยู่ที่ 94.46% และ 92.24% ตามลำดับ รวมไปถึงวิธีการรู้จำที่ผู้วิจัยนำเสนอ มีค่าเฉลี่ยของอัตราความถูกต้องอยู่ที่ 87.05% อีกด้วย

Title	TEXT DETECTION AND RECOGNITION FOR INDUSTRIAL AUTOMATION
Author	SORAWEE POPAYORM
Advisor	Wansuree Masagram, Ph.D.
Academic Paper	Thesis M.S. in Computer Science, Naresuan University, 2019
Keywords	Predefined HSV color slicing, multiscale Radon transform, seven segment display detection, seven segment number recognition, data augmentation

ABSTRACT

Detection of LED seven segment panels presents a particular challenge of locating characters among background clutter. This study proposed a framework based on a predefined HSV color slicing technique for detection and recognition by using the convolutional neural network. The problem of limited data in creating rotation, scale, perspective, and illumination invariant of the neural network training sets are alleviated by the data augmentation technique. The multiscale Radon transform is proposed in this study to enhance the data augmentation for seven segment display recognition. The results demonstrate the framework's superiority over other color slicing methods with 94.46% precision and 92.24% recall for the part of detection. And the accuracy rates for the recognition part demonstrate the overall accuracy performance of 87.05%.

ประกาศคุณปการ

ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงในความกรุณาของ ดร.วันสรีร์ มาศกรัม อาจารย์ที่ปรึกษาวิทยานิพนธ์ และกรรมการผู้ทรงคุณวุฒิทุกท่าน ที่ให้ความรู้ พร้อมทั้งให้คำแนะนำต่าง ๆ ตลอดระยะเวลาในการทำวิทยานิพนธ์ ตลอดจนในการตรวจสอบแก่ไขวิทยานิพนธ์ฉบับนี้ และกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์อันประกอบไปด้วย รองศาสตราจารย์ ดร.ไกรศักดิ์ เกษร ผู้ช่วยศาสตราจารย์ ดร.เกรียงศักดิ์ เมมีย ผู้ช่วยศาสตราจารย์ ดร.ธนาธร พอค้า และ ผู้ช่วยศาสตราจารย์ ดร.กันต์พงษ์ วรรัตน์ปัญญา กรรมการผู้ทรงคุณวุฒิภายนอก ที่ได้ให้คำแนะนำและให้แนวคิดในวิทยานิพนธ์ฉบับนี้ ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์

กราบขอบพระคุณ ดร.ราวิเชษฐ์ อิทธิจรูญโรจน์ ที่ได้กรุณาเสียเวลาอันมีค่าให้คำปรึกษา และแนะนำในการทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงสมบูรณ์

ขอขอบคุณ Dr.Noah Hafner สำหรับคำแนะนำ และแนวทางในการวิเคราะห์ปัญหาในงานวิจัย

ขอขอบคุณ คณะวิทยาศาสตร์ และภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ในคำแนะนำ และการสนับสนุนทั้งเรื่องเอกสาร อุปกรณ์ และทุนส่งเสริมงานวิจัย

ขอขอบคุณ นายธีรเมธร์ อินซัยศรี และ นางสาววิภาวดี ดำมินเศษ ที่สละเวลาให้คำปรึกษา เกี่ยวกับทฤษฎีและหลักการณ์ทางคณิตศาสตร์ที่เกี่ยวข้องในงานวิจัยฉบับนี้

ขอขอบคุณ เพื่อน ๆ และพี่ ๆ ในสาขาวิชาเทคโนโลยีสารสนเทศและวิทยาการคอมพิวเตอร์ ทุกคนที่เคยช่วยเหลือ ให้คำปรึกษา และเป็นกำลังใจในการทำงานวิจัยตลอดมา

ขอกราบขอบคุณพระคุณ บิดา มารดา และบุคคลในครอบครัวอันเป็นที่รัก ซึ่งให้กำลังใจผู้วิจัย มาโดยตลอด รวมไปถึงอาจารย์และผู้มีพระคุณทุกท่านที่ให้การสนับสนุน ผู้วิจัยขอกราบขอบพระคุณ ด้วยความรักและความเคารพเป็นอย่างสูง

สรวีร์ โพธิ์พยอม

สารบัญ

หน้า	
บทคัดย่อภาษาไทย	ค
บทคัดย่อภาษาอังกฤษ	ง
ประกาศคุณูปการ	จ
สารบัญ	ฉ
สารบัญตาราง	ญ
สารบัญภาพ	ภ
บทที่ 1	1
บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 จุดมุ่งหมายของการวิจัย	4
1.3 ขอบเขตของงานวิจัย	4
1.4 นิยามศัพท์เฉพาะ	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	5
1.6 แผนการดำเนินงาน	6
บทที่ 2	7
เอกสารและงานวิจัยที่เกี่ยวข้อง	7
2.1 ทฤษฎีที่เกี่ยวข้องกับการตรวจจับตัวเลขที่แสดงผลแบบเจ็ดส่วน	7
2.1.1 การแสดงผลแบบเจ็ดส่วน (Seven segment)	7
2.1.2 การตรวจจับและรู้จำวัตถุ (Object detection and recognition)	7

2.1.3 แบบจำลองสี (Color model)	8
2.1.3.1 RGB	8
2.1.3.2 HSV	8
2.2 ทฤษฎีที่เกี่ยวข้องกับการรู้จำตัวเลขที่แสดงผลแบบเจ็ดส่วน	10
2.2.1 การแปลงภาพเชิงเรขาคณิตบนระนาบสองมิติ (2-D geometric image transform).....	10
ขั้นตอนที่ 2.2.1.1 การแปลงภาพเชิงมิติ (Isometry transformation).....	11
ขั้นตอนที่ 2.2.1.2 การประมาณค่าในช่วง (Interpolation)	13
2.2.2 การแปลงเรดอน (Radon transform).....	15
2.2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional neural network) .	18
2.3 งานวิจัยที่เกี่ยวข้องกับการตรวจจับ	19
2.4 งานวิจัยที่เกี่ยวกับการรู้จำ	22
บทที่ 3	26
วิธีดำเนินงานวิจัย	26
3.1 การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี	27
3.1.1 ขอบเขตของการดำเนินการ	27
3.1.1.1 ขอบเขตของข้อมูล	27
3.1.1.2 ขอบเขตของอุปกรณ์.....	28
3.1.2 การออกแบบและพัฒนาวิธีการ	28
3.1.3 วิธีการวัดประสิทธิภาพ.....	34
3.2 การรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี.....	35
3.2.1 ขอบเขตของการดำเนินการ.....	35

3.2.1.1 ขอบเขตของข้อมูล	35
3.2.1.2 ขอบเขตของอุปกรณ์.....	38
3.2.2 การออกแบบและพัฒนาวิธีการ	38
บพที่ 4.....	49
ผลการทดลอง.....	49
การทดลองที่ 4.1 การทดลองการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน	49
4.1.1 ผลการวัดประสิทธิภาพโดยไม่มีการปรับค่าของตัวแปร	49
4.1.2 ผลการวัดประสิทธิภาพโดยการปรับค่าของตัวแปร	51
การทดลองที่ 4.2 การทดลองการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน.....	53
4.2.1 ผลการวัดประสิทธิภาพรูปแบบที่ 1	54
4.2.2 ผลการวัดประสิทธิภาพรูปแบบที่ 2	59
บพที่ 5	65
บพสรุป	65
5.1 สรุปผลการวิจัย	65
5.2 ข้อเสนอแนะ	66
บรรณานุกรม	67
ภาคผนวก.....	70
ภาคผนวก ก.....	71
ภาคผนวก ก.....	71
ภาคผนวก ข.....	73
บทความที่นำเสนอในงานประชุมวิชาการ	73
ประวัติผู้วิจัย	85



สารบัญตาราง

หน้า

ตาราง 1 แผนการดำเนินงาน	6
ตาราง 2 ค่าขอบเขต หรือช่วงของสีที่ผู้วิจัยนำเสนอ	31
ตาราง 3 ลักษณะข้อมูลที่ใช้ในการสร้างแบบจำลองการรู้จำรูปแบบที่ 1	46
ตาราง 4 ลักษณะข้อมูลที่ใช้ในการสร้างแบบจำลองการรู้จำรูปแบบที่ 2	47
ตาราง 5 การเปรียบเทียบรูปแบบการวัดประสิทธิภาพการรู้จำ	47
ตาราง 6 ผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจจับโดยไม่มีการปรับค่าของตัวแปร	49
ตาราง 7 ผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจจับโดยการปรับค่าของตัวแปร	51
ตาราง 8 ผลการเปรียบเทียบเวลาในการประมวลผลการตรวจจับ	52
ตาราง 9 ผลการเปรียบเทียบประสิทธิภาพของวิธีการรู้จำรูปแบบที่ 1	54
ตาราง 10 ผลการเปรียบเทียบประสิทธิภาพของวิธีการรู้จำรูปแบบที่ 2	59
ตาราง 11 ผลการเปรียบเทียบเวลาในการประมวลผลการรู้จำ	62
ตาราง 12 สรุปข้อจำกัดการตรวจจับภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน	65
ตาราง 13 สรุปข้อจำกัดการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน	65

สารบัญภาพ

	หน้า
ภาพ 1 ตัวอย่างหน้าจอแสดงผลแบบเจ็ดส่วน	1
ภาพ 2 ตัวอย่างผลลัพธ์การตรวจจับเส้นขอบด้วย Canny edge detection	2
ภาพ 3 ตัวอย่างลักษณะของตัวแสดงผลแบบเจ็ดส่วน.....	7
ภาพ 4 แบบจำลองสี RGB.....	8
ภาพ 5 แบบจำลองสี HSV และลักษณะของสีจาก Hue	9
ภาพ 6 การเลื่อนภาพ.....	11
ภาพ 7 การหมุนภาพ.....	12
ภาพ 8 การย่อและขยายภาพ	12
ภาพ 9 การบิดภาพ	13
ภาพ 10 ตัวอย่างภาพนำเข้าและภาพผลลัพธ์	13
ภาพ 11 ผลลัพธ์ตำแหน่งพิกเซลของภาพนำเข้า.....	15
ภาพ 12 ผลลัพธ์การประมาณค่าในช่วงแบบเพื่อนบ้านใกล้สุด	15
ภาพ 13 การหาพื้นที่ตลอดแนวเส้นตรงของการแปลงเรเดอน	16
ภาพ 14 ตัวอย่างการหมุนภาพในการแปลงเรเดอน	16
ภาพ 15 การหาผลรวมของข้อมูลในการแปลงเรเดอน	17
ภาพ 16 การนำเวกเตอร์หลักมาเข้มต่อ กันในการแปลงเรเดอน	17
ภาพ 17 ใช้โน้ตบุ๊กจาก การแปลงเรเดอนสำหรับการประมวลผลภาพ	18
ภาพ 18 ผลลัพธ์การแปลงองศาของสีในแบบจำลองสี HSV โดย Mu และคณะ	21
ภาพ 19 ตัวอย่างผลลัพธ์การใช้งานวิธีการตรวจจับของ Mu	21

ภาพ 20 ตัวอย่างบริเวณที่สนใจ	22
ภาพ 21 กรอบแนวคิดของ Rakhi P. Ghugardare	23
ภาพ 22 กรอบแนวคิดของ Kulkarni	23
ภาพ 23 ความหนาแน่นของพิกเซลเพื่อระบบการรู้จำตัวเลข	24
ภาพ 24 กรอบแนวคิดของวิธีการที่นำเสนอด้วย	27
ภาพ 25 ตัวอย่างข้อมูลสำหรับการตรวจจับตัวเลขที่แสดงผลแบบเจ็ดส่วน	28
ภาพ 26 ผลลัพธ์การตัดภาพด้วยสีจากแบบจำลองสี HSV	29
ภาพ 27 ตัวอย่างผลลัพธ์แบบจำลองสี HSV จากแบบจำลองสี RGB	29
ภาพ 28 ผลลัพธ์แบบจำลองสี HSV จากแบบจำลองสี RGB จากการถ่ายวัตถุเดิม	30
ภาพ 29 ผลลัพธ์การตัดแบบกำหนดช่วงของสี และความสว่างของแบบจำลองสี HSV	32
ภาพ 30 ผลลัพธ์การตรวจจับวัตถุโดยการเชื่อมต่อสองมิติ	32
ภาพ 31 ตัวอย่างผลลัพธ์จากการระบุการกรองขั้นต้นโดยเงื่อนไขขนาดของวัตถุ	33
ภาพ 32 ตัวอย่างผลลัพธ์จากการระบุการกรองขั้นต้นโดยเงื่อนไขรูปทรงของวัตถุ	34
ภาพ 33 ตัวอย่างข้อมูลภาพถ่ายตัวเลขปกติ	35
ภาพ 34 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการหมุน	36
ภาพ 35 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการย่อ-ขยาย	36
ภาพ 36 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการเอียงเชิงมุมมอง	36
ภาพ 37 ตัวอย่างข้อมูลภาพใช้โน้ตเคนมาร์กจากกระบวนการที่ผู้วิจัยนำเสนอ	36
ภาพ 38 ตัวอย่างข้อมูลภาพถ่ายตัวเลขปกติ	37
ภาพ 39 ตัวอย่างข้อมูลภาพถ่ายตัวเลขที่เอียง	37
ภาพ 40 ตัวอย่างข้อมูลภาพถ่ายหลายขนาด	38
ภาพ 41 ตัวอย่างข้อมูลภาพถ่ายหลายมุมมอง	38

ภาพ 42 ตัวอย่างข้อมูลภาพถ่ายที่ได้รับผลกระทบจากแสง	38
ภาพ 43 การปรับขนาดของภาพโดยไม่รักษารูปทรงของตัวเลข	39
ภาพ 44 ผลลัพธ์การขยายภาพด้วยศูนย์	39
ภาพ 45 การขยายภาพโดยข้อมูลเวกเตอร์ແກ່ທີ່ມີຄ່າເປັນ 0	39
ภาพ 46 การขยายภาพโดยข้อมูลเวกเตอร์ຫລັກທີ່ມີຄ່າເປັນ 0	40
ภาพ 47 ผลลัพธ์การแปลงขนาดภาพ	40
ภาพ 48 ตัวกรองເກາສ්ເຊື່ອນຂອງວິທີການທີ່ນຳເສັນອນນຳມາໃຊ້ງານ	41
ภาพ 49 ตัวอย่างผลลัพธ์จากการกรองภาพจากตัวขนาดต่างกัน	41
ภาพ 50 (ก) ผลลัพธ์ໃຈໂນແກຣມຈາກການແປ່ງເຮດອນຂອງກລຸ່ມຕົວເລີຂ ແລະ (ข) ຕັ້ງເລີກກລຸ່ມ ເດືອກກັນທີ່ມີອັນດາຄວາມເອີ່ມທີ່ແຕກຕ່າງກັນ	42
ภาพ 51 การເລືອກເວັກເຕູຮ່ອງຫລັກທີ່ມີຄ່າສູງສຸດ	43
ภาพ 52 ການຈັດເຮືອງສັນປະສົງທີ່ໃໝ່ກຣົນທີ່ 1	44
ກາພ 53 ການຈັດເຮືອງສັນປະສົງທີ່ໃໝ່ກຣົນທີ່ 2	44
ກາພ 54 ການຈັດເຮືອງສັນປະສົງທີ່ໃໝ່ຂອງເລີຂ 8 ທີ່ມີອັນດາຄວາມເອີ່ມທ່າງກັນ	44
ກາພ 55 ຜົບຜົນການຈັດເຮືອງສັນປະສົງທີ່ໃໝ່ເຮດອນທີ່ຂໍຍາຍພາບດ້ວຍศູນຍົງ	45
ກາພ 56 ຜົບຜົນການຕຽບຕໍ່ການທີ່ໄດ້ຮັບຜົນການຈັດເຮືອງສັນປະສົງທີ່	50
ກາພ 57 ການປັບປຸງທີ່ໄດ້ຮັບຜົນການຈັດເຮືອງສັນປະສົງທີ່	56
ກາພ 58 ການປັບປຸງທີ່ໄດ້ຮັບຜົນການຈັດເຮືອງສັນປະສົງທີ່	57
ກາພ 59 ການປັບປຸງທີ່ໄດ້ຮັບຜົນການຈັດເຮືອງສັນປະສົງທີ່	58
ກາພ 60 ຜົບຜົນການທີ່ໄດ້ຮັບຜົນການຈັດເຮືອງສັນປະສົງທີ່	63

ภาพ 61 ปั้นหาในการจัดเรียนสัมประสิทธิ์ใหม่เรดอน 72



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

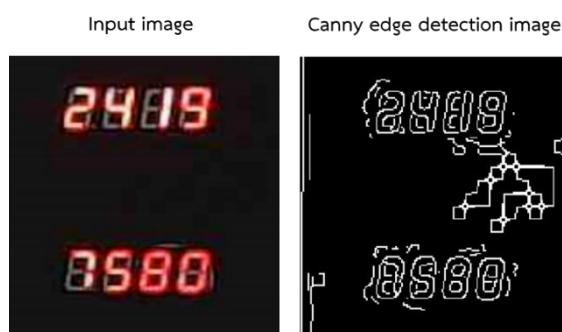
ในภาคอุตสาหกรรมมีการใช้งานเครื่องจักรหรืออุปกรณ์ที่มีหน้าจอแสดงผลแบบเจ็ดส่วนเป็นจำนวนมาก เช่น เครื่องผสมผลิตภัณฑ์ โดยทั่วไปหากโรงงานต้องการใช้งานข้อมูลตัวเลขเหล่านี้ และไม่มีการส่งตัวเลขผ่านไปยังระบบโดยตรงจำเป็นต้องใช้มนุษย์ในการเก็บข้อมูลโดยการจดบันทึกตัวเลขที่แสดงผลผ่านทางหน้าจอ ซึ่งในกระบวนการนี้อาจเกิดความผิดพลาดอันเนื่องมาจากพฤติกรรมการจดบันทึกหรือเกิดความเสียหายต่อเอกสารการจดบันทึกได้ และหากข้อมูลตัวเลขมีการเปลี่ยนแปลงอย่างต่อเนื่อง Roge งานอุตสาหกรรมส่วนมากจะทำการกำหนดช่วงเวลาในการจดบันทึกตัวเลขเหล่านี้จึงส่งผลให้ข้อมูลที่ได้รับเกิดความคลาดเคลื่อนจากข้อมูลจริง ดังนั้น การระบบคอมพิวเตอร์มาใช้งานในการตรวจจับและรู้จักตัวเลขที่แสดงผลแบบเจ็ดส่วนจึงเป็นทางเลือกในการหลีกเลี่ยงปัญหาดังกล่าวที่อาจเกิดขึ้นได้ อย่างไรก็ตาม การเปลี่ยนเครื่องจักรที่มีอยู่เป็นเครื่องจักรที่สามารถส่งตัวเลขผ่านระบบไปยังคอมพิวเตอร์หรือเซิร์ฟเวอร์ได้โดยตรงอาจมิใช่ตัวเลือกที่ดีที่สุดเนื่องจากเครื่องจักรที่มีเทคโนโลยีทันสมัยนั้นอาจมีราคาแพงมากเกินความจำเป็นของ Roge งาน อุตสาหกรรม ในขณะที่เครื่องอ่านอักษรด้วยแสง (Optical character recognition) เป็นอุปกรณ์ที่มีความสามารถในการรู้จักตัวอักษรแล้วนำไปเก็บไว้ในหน่วยความจำมีค่าใช้จ่ายที่น้อยกว่าเครื่องจักรขนาดใหญ่ แต่ด้วยลักษณะที่เป็นเจ็ดส่วนนั้นทำให้ตัวเลขมีโครงสร้างที่ไม่ต่อเนื่อง และหากจอภาพแสดงผลเป็นแอลอดีที่มีส่วนประกอบคือแสงและสีดังภาพ 1 ส่งผลให้เครื่องอ่านอักษรด้วยแสงมีประสิทธิภาพด้อยลง ยิ่งไปกว่านั้นเครื่องอ่านอักษรด้วยแสงมีข้อจำกัดเมื่อนำไปใช้งานกับภาพที่ได้รับผลจากสัญญาณรบกวน



ภาพ 1 ตัวอย่างหน้าจอแสดงผลแบบเจ็ดส่วน

จากการวิจัยที่ผ่านมาทำให้ทราบปัญหาการตรวจจับและรู้จำตัวเลขที่แสดงผลแบบเบ็ดส่วนบนของภาพแล็ซีดีและจากภาพแล็อดีในภาพถ่ายธรรมชาติอันเนื่องมาจากสภาพแวดล้อมพฤษิตกรรมการถ่ายภาพ และความบกพร่องของอุปกรณ์ในการถ่ายภาพ จึงมีการพัฒนาเครื่องมือและวิธีการต่าง ๆ เพื่อนำไปใช้กับภาพที่พบปัญหาดังกล่าว เช่น งานวิจัยดังต่อไปนี้ ในปี 2009 ได้มีการนำเสนอวิธีการที่พัฒนามาจากเครื่องอ่านอักษรระดับด้วยแสง โดยโทโพโลยีของโครงข่ายประสาทเทียมเพื่อให้สามารถทำการรู้จำตัวเลขที่แสดงผลในรูปแบบเบ็ดส่วนแบบบัตโนมัติ (Bonacic et al., 2009) ในขณะเดียวกันได้มีการพัฒนาวิธีการแบ่งส่วนภาพที่เรียกว่า Scissoring algorithm ซึ่งใช้ในการตัดส่วนภาพที่เป็นตัวเลขที่แสดงในรูปแบบเบ็ดส่วนทั้งบนของภาพแล็อดีและแล็ซีดีจากอุปกรณ์อิเล็กทรอนิกส์ เช่น ดิจิตอลมัลติมิเตอร์ แล้วนำการสกัดคุณลักษณะจากโครงสร้างของตัวเลขมาใช้เพื่อทำการรู้จำโดยการเปรียบเทียบเชิงสถิติหรือเชิงโครงสร้างแทนการใช้งานโครงข่ายประสาทเทียม (Ghugardare, Narote, Mukherji, & Kulkarni, 2009) เช่นเดียวกันกับ นำเสนอวิธีการรู้จำตัวเลขที่แสดงผลในรูปแบบเบ็ดส่วนบนของภาพแล็ซีดีโดยใช้เทคนิค Y Profile และ X Profile ในการแบ่งส่วนภาพหลังจากนั้นจึงสกัดคุณลักษณะจากความหนาแน่นของพิกเซลเชิงโครงสร้างของตัวเลขในการรู้จำซึ่งได้ผลลัพธ์เป็นที่น่าพอใจ โดยที่โครงข่ายประสาทเทียมใช้เวลาประมวลผลมากเมื่อมีข้อจำกัดด้านชาร์ดแวร์ (Kulkarni & Kute, 2016)

จากการวิจัยข้างต้นมีเพียงงานวิจัยเดียวที่ทำการทดลองกับหน้าจอแสดงผลแล็อดี (Ghugardare et al., 2009) ซึ่งแตกต่างจากหน้าจอแสดงผลแล็ซีดีในส่วนของแสงและสี แต่ยังพบข้อจำกัดในด้านระยะห่างจากการถ่ายภาพที่จำเป็นต้องให้ได้ส่วนของตัวเลขในลักษณะที่สมบูรณ์โดยไม่มีวัตถุอื่นเกี่ยวข้องในภาพ นั่นคือ หากในภาพประกอบด้วยวัตถุที่หลากหลายแล้ววิธีการนี้อาจไม่เหมาะสม ทางผู้วิจัยจึงได้ทดลองใช้การตรวจจับเส้นขอบ (Edge detection) โดย Canny edge detection กับภาพถ่ายหน้าจอแสดงผลแบบเบ็ดส่วนบนของภาพแล็อดีแล้วพบปัญหาการได้มาซึ่งเส้นขอบที่ไม่ใช่ตัวเลขที่เปล่งแสงของหน้าจอแสดงผลแบบเบ็ดส่วนดังภาพ 2



ภาพ 2 ตัวอย่างผลลัพธ์การตรวจจับเส้นขอบด้วย Canny edge detection

ดังนั้น การวิจัยครั้งนี้ทางผู้วิจัยจึงนำเสนอเทคนิคการแบ่งส่วนภาพโดยการกำหนดขอบเขตของสีและความสว่างจากภาพถ่าย บนจำลองสี HSV เพื่อให้ได้ส่วนที่เป็นตัวเลขที่สมบูรณ์มากที่สุด ซึ่งสีจากแหล่งอื่นนั้นมีความหลากหลาย เช่น สีแดง สีเขียว หรือสีฟ้า ทางผู้วิจัยจึงได้กำหนดช่วงของสีที่เหมาะสมกับสีต่าง ๆ เพื่อรับกับความหลากหลายของสีจากแหล่งอื่น โดยสามารถใช้งานกับภาพถ่ายที่ประกอบด้วยวัตถุอื่นนอกจากหน้าจอแสดงผลแบบเจ็ดส่วนบนจอภาพแหล่งอื่น

จากการศึกษางานวิจัยที่ได้กล่าวไว้ข้างต้นทางผู้วิจัยพบว่างานวิจัยโดยส่วนใหญ่มีข้อจำกัดด้านฮาร์ดแวร์มาเกี่ยวข้อง (Kulkarni & Kute, 2016) เนื่องจากวิธีการเหล่านี้เหมาะสมในการนำไปใช้งานกับอุปกรณ์ที่มีขนาดเล็ก หรืออุปกรณ์จำพวกประมวลผลข้อมูลที่อยู่ใกล้กับแหล่งข้อมูล (Edge computing) เช่น Raspberry pi หรือ Arduino board โดยไม่กีดขวางการทำงานของพนักงานหรือเครื่องจักร และอาจมีความสามารถในการประมวลผลที่ไม่มากจึงเป็นข้อจำกัดที่สำคัญในการพัฒนา วิธีการเหล่านี้ แต่เมื่อพิจารณาจากความสามารถของโครงข่ายประสาทเทียมแล้วพบว่า ส่วนที่ใช้ระยะเวลามากที่สุดคือการสร้างแบบจำลอง ซึ่งการนำแบบจำลองไปใช้นั้นสามารถทำงานได้รวดเร็วไม่ต่างจากวิธีการรู้จำอื่น ดังนั้น ทางผู้วิจัยจึงต้องการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแหล่งอื่นโดยไม่มีข้อจำกัดด้านการประมวลผลของฮาร์ดแวร์ และเพื่อให้วิธีการนี้มีประสิทธิภาพมากขึ้น ทางผู้วิจัยตระหนักว่าการถ่ายภาพหรือจับภาพในภาคอุตสาหกรรมอาจมีการจัดรูปแบบเพื่อที่จะได้ภาพถ่ายที่มีตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนที่สมบูรณ์ ตั้งตรง รวมถึงมีระยะห่างในการถ่ายภาพคงที่ แต่ในการถ่ายภาพโดยทั่วไปนั้นเป็นไปไม่ได้ที่จะวัดระยะห่างในการถ่ายภาพ และได้ภาพตัวเลขที่ตั้งตรงในทุกครั้งที่ถ่ายภาพ จึงส่งผลให้ตัวเลขที่ได้รับจากภาพถ่ายมีมุมมอง องศาในการเอียง รวมถึงขนาดที่หลากหลาย ซึ่งส่งผลให้การรู้จำภาพนั้นมีประสิทธิภาพด้อยลงอย่างมาก ดังนั้น ผู้วิจัยจึงนำปัญหาเหล่านี้มายิเคราะห์และจัดการเพื่อการนำไปใช้งานในสภาพแวดล้อมที่หลากหลายมากขึ้น

จากการค้นคว้าพบว่าในปี 2018 ได้มีงานวิจัยนำเสนอการเพิ่มประสิทธิภาพในการรู้จำของโครงข่ายประสาทเทียม โดยใช้การเพิ่มปริมาณข้อมูลภาพ (Data Augmentation) เป็นการเพิ่มความหลากหลายของข้อมูลสอนและข้อมูลทดสอบเพื่อส่งให้โครงข่ายประสาทเทียม แล้วใช้การลงคะแนน (voting) ในการทำนายผลลัพธ์ เพื่อใช้ในการรู้จำภาพสกู๊นเน็ต แล้ววน (Dong, Wu, Dong, Feng, & Wu, 2018) และในขณะเดียวกันการเพิ่มปริมาณข้อมูลภาพยังสามารถนำมาใช้ในการแก้ปัญหาข้อมูลภาพสอนที่มีจำนวนจำกัดอีกด้วย ดังที่นำเสนอใน การรู้จำข้อมูลภาพ SAR (T. Jiang, Cui, Zhou, & Cao, 2018) แต่โครงข่ายประสาทเทียมจะมีประสิทธิภาพน้อยลงหากข้อมูลที่ต้องการทำนายเป็นข้อมูลที่มีสภาพแวดล้อมแตกต่างจากข้อมูลที่ทำการเรียนรู้ ดังนั้น ผู้วิจัยจึงนำเสนอรูปแบบมาตรฐานให้กับข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนที่มีลักษณะที่แตกต่างกันในแต่ละกลุ่ม

ตัวเลข แต่มีความใกล้เคียงกันในกลุ่มตัวเลขเดียวกันถึงแม้จะได้รับผลกระทบจากสภาพแวดล้อมที่ต่างกันเพื่อนำไปใช้กับโครงข่ายประสาทเทียมเพื่อการรู้จำตัวเลขในรูปแบบเจ็ดส่วน

1.2 จุดมุ่งหมายของการวิจัย

- 1.2.1 เพื่อศึกษาวิธีการการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแล็อกอีดี โดยการศึกษาทฤษฎีที่เกี่ยวข้องกับการตรวจจับและการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน การตรวจจับวัตถุด้วยสี และการทดสอบการใช้งานจริง
- 1.2.2 เพื่อพัฒนาวิธีการการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแล็อกอีดี โดยใช้โปรแกรมภาษา Python ในการพัฒนาระบวนการทั้งหมด

1.3 ขอบเขตของงานวิจัย

การตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม ผู้วิจัยมุ่งเน้นการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแล็อกอีดี และข้อมูลภาพตัวเลขที่ผู้วิจัยนำมาใช้พัฒนาจะมีลักษณะดังนี้

- เป็นภาพสี
- มีตัวเลขจากหน้าจอแสดงผลแบบเจ็ดส่วนบนจอภาพแล็อกอีดีสีแดง
- มีตัวเลขที่แสดงในหน้าจอแสดงผลทั้งหมด 8 ตัวเลข
- ตัวเลขที่ปรากฏบนจอภาพแล็อกอีดีแต่ละตำแหน่งคือ 0 ถึง 9
- ภาพหน้าจอแสดงผลในรูปแบบเจ็ดส่วนมีขนาดภาพ 1280×720 พิกเซล
- ระยะการถ่ายภาพอยู่ในช่วง 0.25 เมตร ถึง 1.5 เมตร จากหน้าจอแสดงผล
- ภาพที่ตัดมาเพียงตัวเลขมีขนาดภาพอยู่ระหว่าง 2×6 พิกเซล จนถึง 36×52 พิกเซล
- ภาพที่ตัดมาเพียงตัวเลขมีความเอียงในช่วง -80° จนถึง 80°
- ภาพที่ตัดมาเพียงตัวเลขมีการเอียงเชิงมุมมองในช่วง -80° จนถึง 80°
- ภาพถ่ายทั้งหมดจะใช้อุปกรณ์ Logitech HD pro webcam c920 ในการถ่ายภาพ

ผู้วิจัยพัฒนาวิธีการทั้งหมดโดยโปรแกรมภาษา Python และ MATLAB และผู้วิจัยนำเสนอรายละเอียดทั้งหมดของข้อมูลในการทดลองไว้ในบทที่ 3 ของงานวิจัยฉบับนี้

1.4 นิยามศัพท์เฉพาะ

- 1.4.1 การตรวจจับวัตถุ (Object detection) คือ การใช้คุณลักษณะของวัตถุในภาพ เช่น โครงสร้าง หรือ สี ในการแยกและประเภทของวัตถุเพื่อทำการระบุตำแหน่งของวัตถุประเภทที่ต้องการ

1.4.2 การรู้จำวัตถุ (Object recognition) คือ กระบวนการในการเรียนรู้วัตถุของเครื่องจักรโดยการสกัดคุณลักษณะเฉพาะจากวัตถุ เพื่อใช้ในการตัดสินแล้วจัดหมวดหมู่แก่วัตถุนั้น

1.4.3 การตรวจจับเส้นขอบ (Edge detection) คือ การแปลงภาพนำเข้าให้เหลือเพียงส่วนที่เป็นเส้นขอบภายในภาพ หรือการตัดส่วนที่เป็นเนื้อสีออกจากภาพโดยการหาความแตกต่างของข้อมูลภาพ นิยมนำมาใช้ในการหานบริเวณของวัตถุภายในภาพ

1.4.4 จอภาพแอลซีดี (Liquid crystal display หรือ LCD) คือ หน้าจอแสดงผลแบบบิดจิตอลชี้งภาพที่ปราศจากแสงจากหลอดไฟด้านหลังจอภาพผ่านชั้นกรองแสงไปยังคริสตัลเหลวที่เรียงตัวกัน 3 เซลล์ได้แก่ แสงสีแดง แสงสีเขียว และแสงสีน้ำเงินกล้ายเป็นพิกเซลที่สร้างขึ้นมา (เสริฐ จำเริญ, 2012)

1.4.5 จอภาพแอลอีดี (Light-emitting diode หรือ LED) คือ หน้าจอแสดงผลที่เกิดจากแสงจากไดโอดที่สามารถเปลี่ยนพลังงานไฟฟ้าเป็นพลังงานแสงได้โดยตรง ชี้งสีที่ปราศจากขั้นบนของภาพสัมพันธ์กับวัสดุกึ่งตัวนำที่ใช้ (Production, 2017)

1.4.6 การเพิ่มปริมาณข้อมูลภาพ (Data augmentation) คือ เทคนิคการเพิ่มปริมาณข้อมูลภาพสอน หรือข้อมูลภาพทดสอบ โดยใช้การแปลงภาพเชิงเรขาคณิต (Geometric transformations) หรือการแปลงภาพในรูปแบบอื่น เข้ามาช่วยในการแปลงข้อมูลภาพที่มีอยู่ให้มีความหลากหลายในรูปแบบที่ทำการแปลง และเพิ่มจำนวนข้อมูลภาพให้มากขึ้น ชี้งเทคนิคนี้สามารถเพิ่มประสิทธิภาพในการท่านายหรือการรู้จำให้กับเครื่องจักรได้ (Dong et al., 2018; T. Jiang et al., 2018; Sánchez, Ruiz, & Meza, 2005)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้รับความรู้ที่ศึกษามาประยุกต์ใช้ในงานด้านการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแอลอีดีได้

1.5.2 ได้รับวิธีการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแอลอีดีอย่างมีประสิทธิภาพ

1.6 แผนการดำเนินงาน

ตาราง 1 แผนการดำเนินงาน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

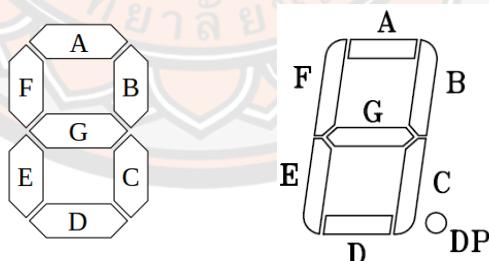
ในงานวิจัยนี้ผู้วิจัยได้ศึกษาแนวคิดและทฤษฎีรวมไปถึงงานวิจัยที่เกี่ยวข้อง เพื่อนำมาประยุกต์ในการวิจัย ซึ่งประกอบด้วยเนื้อหาดังต่อไปนี้

- 2.1 ทฤษฎีที่เกี่ยวข้องกับการตรวจจับตัวเลขที่แสดงผลแบบเจ็ดส่วน
- 2.2 ทฤษฎีที่เกี่ยวข้องกับการรู้จำตัวเลขที่แสดงผลแบบเจ็ดส่วน
- 2.3 งานวิจัยที่เกี่ยวข้องกับการตรวจจับตัวเลขที่แสดงผลแบบเจ็ดส่วน
- 2.4 งานวิจัยที่เกี่ยวข้องกับการรู้จำตัวเลขที่แสดงผลแบบเจ็ดส่วน

2.1 ทฤษฎีที่เกี่ยวข้องกับการตรวจจับตัวเลขที่แสดงผลแบบเจ็ดส่วน

2.1.1 การแสดงผลแบบเจ็ดส่วน (Seven segment display)

หน้าจอแสดงผลแบบเจ็ดส่วน (Seven segment display) เป็นหน้าจอแสดงผลที่นำมาใช้ในงานกับการแสดงผลตัวเลขดิจิตอลจากอุปกรณ์อิเล็กทรอนิกส์ซึ่งอาจแสดงตัวอักษรได้บางชนิด อุปกรณ์ชนิดนี้เกิดจากการเรียงตัวของส่วนประกอบ 7 ส่วนติดต่อกันในลักษณะเลขแปดแสดงผลบนจอภาพแอลอีดี ส่วนประกอบแต่ละส่วนอาจมีลักษณะเป็นหกเหลี่ยมหรือคล้ายสี่เหลี่ยมคงทุม ซึ่งอาจเรียงตัวกันในลักษณะตรงหรือลักษณะเอียง



ภาพ 3 ตัวอย่างลักษณะของตัวแสดงผลแบบเจ็ดส่วน

ที่มา : (KMUTNB, 2013; Uln2003, 2018)

2.1.2 การตรวจจับและรู้จำวัตถุ (Object detection and recognition)

การตรวจจับวัตถุ เป็นเทคโนโลยีคอมพิวเตอร์ที่สัมพันธ์กับคอมพิวเตอร์วิทัค์และ การประมวลผลภาพ นิยมนำมาใช้ในการบอกตำแหน่งของวัตถุที่ต้องการซึ่งปรากฏในภาพดิจิตอลหรือวิดีโอด้วยใช้ลักษณะ รูปทรง หรือองค์ประกอบของวัตถุในการแยกแยะ เช่น การตรวจจับสี่เหลี่ยมโดย

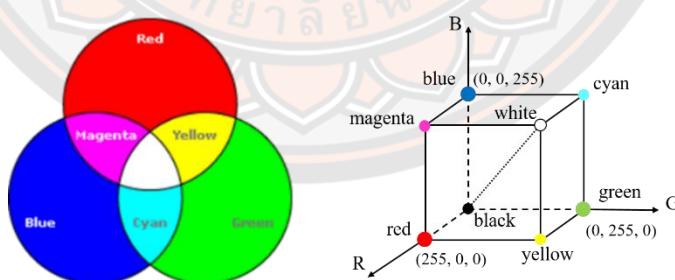
รูปทรงที่ประกอบด้วยมุมภายใน 4 มุม ด้วยรูปทรงดังกล่าวทำให้วงกลม หรือรูปทรงเรขาคณิตอื่นไม่ถูกตรวจจับ

การรู้จำวัตถุ เป็นการหาลักษณะเฉพาะของแต่ละวัตถุที่ถูกตรวจจับซึ่งอาจอยู่ในประเภทเดียวกันแต่มีรายละเอียดลักษณะเฉพาะที่แตกต่างกันเพื่อคัดแยกประเภทของวัตถุ แล้วคุณลักษณะเฉพาะที่ใช้ในการคัดแยกเหล่านั้นจะถูกจัดเก็บไว้เพื่อสร้างแบบจำลองในการรู้จำวัตถุ เช่น สีเหลี่ยมจากการตรวจจับอาจได้สีเหลี่ยมด้านบน สีเหลี่ยมคงที่ หรือสีเหลี่ยมจัตุรัสปะปนอยู่ด้วยกัน แต่ลักษณะของสีเหลี่ยมด้านบน คือ ด้านของสีเหลี่ยมที่อยู่ตรงกันข้ามกันต้องขนาดกันทั้ง 2 คู่ สีเหลี่ยมคงที่ คือ สีเหลี่ยมที่มีด้านขนาดกันอย่างน้อย 1 คู่ และสีเหลี่ยมจัตุรัสด้านทั้ง 4 จะต้องมีขนาดเท่ากันและมุมภายในทุกมุมต้องมีขนาดเท่ากันทำให้แต่ละมุมเป็นมุมฉาก ด้วยคุณลักษณะเฉพาะเหล่านี้ทำให้คอมพิวเตอร์สามารถเรียนรู้จากข้อมูลและแยกแยะวัตถุเป็นประเภทต่าง ๆ ได้

2.1.3 แบบจำลองสี (Color model)

2.1.3.1 RGB

เป็นแบบจำลองสีที่อ้างอิงมาจากแสงสี มีแม่สี 3 สี คือ แดง (Lefèbvре, Macq, & Legat) เขียว (Green) และน้ำเงิน (Blue) ระบบแสงสีนี้ได้รับความนิยมนำมาใช้ในการแสดงผลผ่านหน้าจอ เช่น คอมพิวเตอร์ และโทรศัพท์ สีอื่นนอกเหนือจากแม่สีจะเกิดจากการผสมสีจากแม่สีในระบบพิกัดจากในรูปแบบ (r , g , b) โดยที่ r g และ b คือ ค่าของสีที่อยู่ระหว่าง 0 ถึง 255 ของสีแดง เขียว และน้ำเงินตามลำดับ ซึ่งมีสีดำเป็นค่าที่ต่ำที่สุด คือ $(0, 0, 0)$ และสีขาวเป็นค่าที่สูงที่สุด คือ $(255, 255, 255)$



ภาพ 4 แบบจำลองสี RGB

ที่มา : (Corporation, 2011; Doughty, 2016)

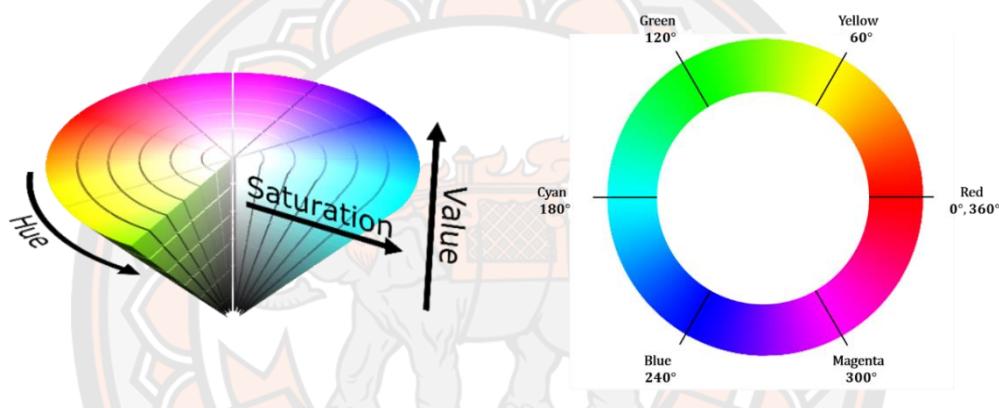
2.1.3.2 HSV

เป็นแบบจำลองสีที่มีแนวคิดมาจากการความสามารถในการมองเห็นของมนุษย์ซึ่งแบ่งออกเป็น 3 ส่วนประกอบ ดังนี้

ส่วนประกอบที่ 1 สี (Hue) เป็นส่วนที่มีเนื้อสีของทุกสีมีลักษณะแบบจำลองเป็นวงกลมเริ่มต้นที่สีแดง 0 องศา ไปจนถึงสีม่วงแดง 300 องศา แล้ววนกลับมาที่สีแดง 360 องศา ซึ่งเมื่อพิจารณาจากกํากลางของสี จะมีระยะห่างจากสีที่ติดกันอยู่ 60 องศา ดังภาพ 5

ส่วนประกอบที่ 2 ความอิมสี (Saturation) แสดงผลสีที่สมสีขาว และการเป็นเนื้อสีแท้ โดยที่ความอิมสีมีค่าจาก 0 ถึง 255 เมื่อความอิมสีเป็น 0 จะแสดงผลสีขาว และความอิมสีเป็น 255 จะแสดงเนื้อสีแท้ เมื่อค่าของความสว่าง (Value) มีค่าเป็น 255 ดังภาพ 5

ส่วนประกอบที่ 3 ความสว่าง (Value หรือ Brightness) แสดงผลความสว่างของสี ซึ่งมีค่าระหว่าง 0 ถึง 255 โดยขึ้นกับค่าความอิมสี (Saturation) หากค่าความอิมสีเป็น 0 ค่าความสว่างที่ 0 จะแสดงผลสีดำ และค่าความสว่างที่ 255 จะแสดงผลสีขาว ดังภาพ 5



ภาพ 5 แบบจำลองสี HSV และลักษณะของสีจาก Hue

ข้อมูลภาพที่พับได้ทั่วไปส่วนใหญ่เป็นระบบแบบจำลองสี RGB ซึ่งค่าของสีนี้จะบรรจุค่าการผสมสี แดง เขียว และน้ำเงินโดยที่ความอิมสี และความสว่างปะปนอยู่ภายในค่าสีด้วย เช่นกัน สมการ 1 – 5 ต่อไปนี้ เป็นสมการการคำนวณแบบจำลองสี HSV จากแบบจำลองสี RGB

$$I_{min}(x, y) = \min\{R(x, y), G(x, y), B(x, y)\} \quad (1)$$

$$I_{max}(x, y) = \max\{R(x, y), G(x, y), B(x, y)\} \quad (2)$$

สมการคำนวณสี

$$H(x, y) = \begin{cases} 0, & I_{max} = I_{min} \\ 60^\circ \left(\frac{G(x, y) - B(x, y)}{I_{max} - I_{min}} \right) + 0^\circ, & I_{max} = R(x, y) \\ 60^\circ \left(\frac{B(x, y) - R(x, y)}{I_{max} - I_{min}} \right) + 120^\circ, & I_{max} = G(x, y) \\ 60^\circ \left(\frac{R(x, y) - G(x, y)}{I_{max} - I_{min}} \right) + 240^\circ, & I_{max} = B(x, y) \end{cases} \quad (3)$$

สมการคำนวณความอิมสี

$$S(x, y) = \begin{cases} 0, & I_{max} = 0 \\ \frac{I_{max}}{I_{max} - I_{min}}, & otherwise \end{cases} \quad (4)$$

สมการคำนวณความสว่าง

$$V(x, y) = I_{min} \quad (5)$$

เมื่อ	H, S, V	คือ ค่าในส่วนของสี ความอิมสี และความสว่างที่ได้ตามลำดับ
	(x, y)	คือ ตำแหน่งของพิกเซลที่พิจารณา
	R, G, B	คือ ค่าของสีแดง เขียว และน้ำเงินของพิกเซลที่พิจารณาตามลำดับ
	I_{min}	คือ ค่าของสีที่น้อยที่สุดเมื่อเปรียบเทียบค่าสีแดง สีเขียว และสีน้ำเงิน ณ ตำแหน่งพิกเซล (x, y) ซึ่งสามารถคำนวณดังสมการที่ 1
	I_{max}	คือ ค่าของสีที่สูงที่สุดเมื่อเปรียบเทียบค่าสีแดง สีเขียว และสีน้ำเงิน ณ ตำแหน่งพิกเซล (x, y) ซึ่งสามารถคำนวณดังสมการที่ 2

2.2 ทฤษฎีเกี่ยวกับการรู้จำตัวเลขที่แสดงผลแบบเจ็ดส่วน

2.2.1 การแปลงภาพเชิงเรขาคณิตบนระนาบสองมิติ (2-D geometric image transform)

เทคนิคการแปลงภาพเชิงเรขาคณิตบนระนาบสองมิติได้พัฒนาขึ้นเพื่อตอบสนองกับความหลากหลายของภาพที่สามารถพบได้โดยทั่วไป เช่น การแปลงภาพให้มีความเอียง หรือมีขนาดที่ใหญ่

ขึ้นหรือเล็กลง โดยอาศัยภาพต้นแบบมาผ่านกระบวนการแปลงทางคณิตศาสตร์ ซึ่งการแปลงเชิงเรขาคณิตประกอบด้วย 2 ขั้นตอน ดังต่อไปนี้

ขั้นตอนที่ 2.2.1.1 การแปลงภาพเชิงมิติ (Isometry transformation)

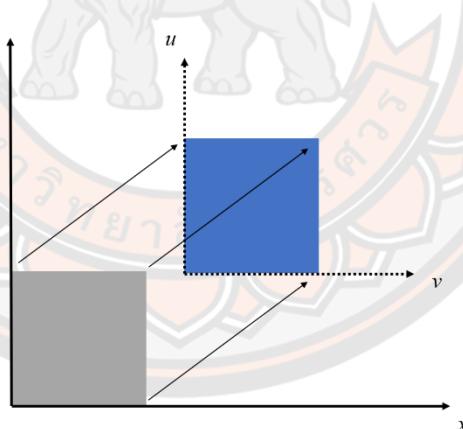
กระบวนการแปลงภาพเชิงมิติอยู่บนพื้นฐานของการแปลงเชิงเส้น (Linear transformation) และการประมาณค่าในช่วง (Interpolation) ซึ่งเมื่อพิจารณาบนพิกัดเอกพันธุ์ (Homogeneous coordinates) บนระบบสองมิติ ณ จุด (x, y) ได้ ๆ (Sánchez et al., 2005) แต่ละรูปแบบการแปลงจะสามารถคำนวณได้ดังนี้ เมื่อ (u, v) เป็นผลลัพธ์การแปลงภาพเชิงมิติในรูปแบบที่กำหนด

การแปลงที่ 1 การเลื่อนภาพ (Translation)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

เมื่อ t_x คือ ค่าคงที่ที่ส่งผลต่อการเลื่อนในแนวแกน x

เมื่อ t_y คือ ค่าคงที่ที่ส่งผลต่อการเลื่อนในแนวแกน y

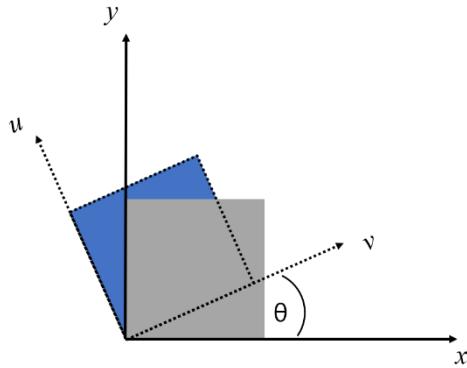


ภาพ 6 การเลื่อนภาพ

การแปลงที่ 2 การหมุนภาพ (Rotation)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

เมื่อ θ คือ ค่าองศาของการหมุนจากแนวแกน x

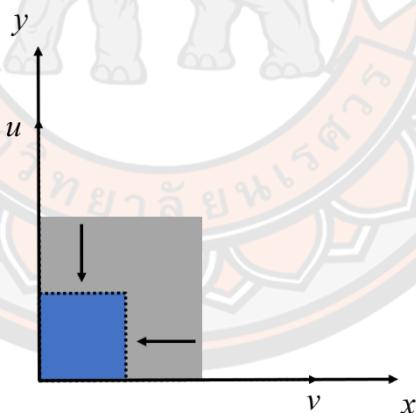


ภาพ 7 การหมุนภาพ

การแปลงที่ 3 การย่อและขยายภาพ (Scaling)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (8)$$

เมื่อ S_x คือ ตัวประกอบที่ส่งผลต่อขนาดภาพตามแนวแกน x
 S_y คือ ตัวประกอบที่ส่งผลต่อขนาดภาพตามแนวแกน y

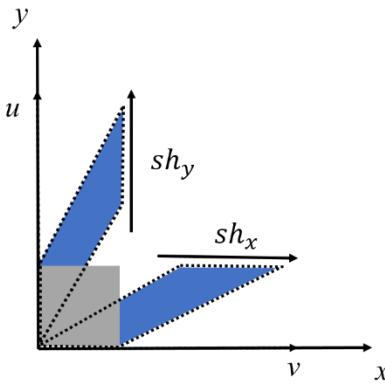


ภาพ 8 การย่อและขยายภาพ

การแปลงที่ 4 การบิดภาพ (Shearing)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & sh_x & 0 \\ sh_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (9)$$

เมื่อ sh_x คือ ค่าสัมประสิทธิ์ที่ส่งผลต่อการบิดรูปร่างตามแนวแกน x
 sh_y คือ ค่าสัมประสิทธิ์ที่ส่งผลต่อการบิดรูปร่างตามแนวแกน y



ภาพ 9 การบิดภาพ

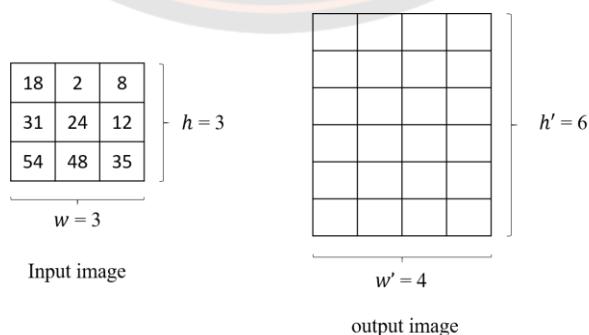
โดยที่สมการที่ 6 – 9 สามารถเขียนอยู่ในรูปแบบมาตรฐานได้ดังสมการที่ 10 และสามารถประยุกต์ใช้ในการแปลงความเอียงเชิงมุมมอง (Perspective) ได้

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} S_x \cdot \cos\theta & -sh_x \cdot \sin\theta & t_x \\ sh_y \cdot \sin\theta & S_y \cdot \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (10)$$

ขั้นตอนที่ 2.2.1.2 การประมาณค่าในช่วง (Interpolation)

จากการบวนการแปลงเชิงมิติจะได้ผลลัพธ์ตำแหน่งของภาพที่ผ่านการแปลงหลังจากนั้นกระบวนการนี้จะทำการประมาณค่า ณ จุดต่าง ๆ บนภาพที่ผ่านการแปลงเชิงมิติโดยใช้การประมาณค่าในช่วงด้วยรูปแบบเพื่อนบ้านใกล้สุด (Nearest-neighbor) เพื่อเป็นการปรับค่าของแต่ละพิกเซลบนภาพโดยไม่มีการคำนวณค่าใหม่ ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 2.2.1.2.1 หากค่าสัดส่วน (Proportion) ระหว่างความกว้างและความสูงของภาพนำเข้ากับความกว้างและความสูงของภาพผลลัพธ์ดังตัวอย่างต่อไปนี้



ภาพ 10 ตัวอย่างภาพนำเข้าและภาพผลลัพธ์

จากภาพ 10 ภาพนำเข้ามีความกว้าง 3 พิกเซล และความสูง 3 พิกเซล ภาพผลลัพธ์มีความกว้าง 4 พิกเซล และความสูง 6 พิกเซล สัดส่วนของความสูง ($Ratio_{Row}$) คำนวณดังสมการที่ 11 และสัดส่วนของความกว้าง ($Ratio_{Column}$) ดังสมการที่ 12

$$Ratio_{Row} = \frac{h}{h'} \quad (11)$$

$$Ratio_{Column} = \frac{w}{w'} \quad (12)$$

เมื่อ	h	คือ ความสูงของภาพนำเข้า
	h'	คือ ความสูงของภาพผลลัพธ์
	w	คือ ความกว้างของภาพนำเข้า
	w'	คือ ความกว้างของภาพผลลัพธ์

จากภาพ 10 ค่าสัดส่วนของความสูง คือ $\frac{6}{3}$ มีค่าเป็น 2 และค่าสัดส่วนของความกว้าง $\frac{4}{3}$ มีค่าประมาณ 1.33

ขั้นตอนที่ 2.2.1.2.2 คำนวณหาตำแหน่งอ้างอิงจากภาพนำเข้าบนพิกเซลของภาพผลลัพธ์ทำให้ค่าของแต่ละพิกเซลอยู่ในช่วงค่าพิกเซลของภาพนำเข้าโดยสมการที่ 13 และ 14

$$Row_{position} = \frac{Rv}{Ratio_{Column}} \quad (13)$$

$$Column_{position} = \frac{Cv}{Ratio_{Row}} \quad (14)$$

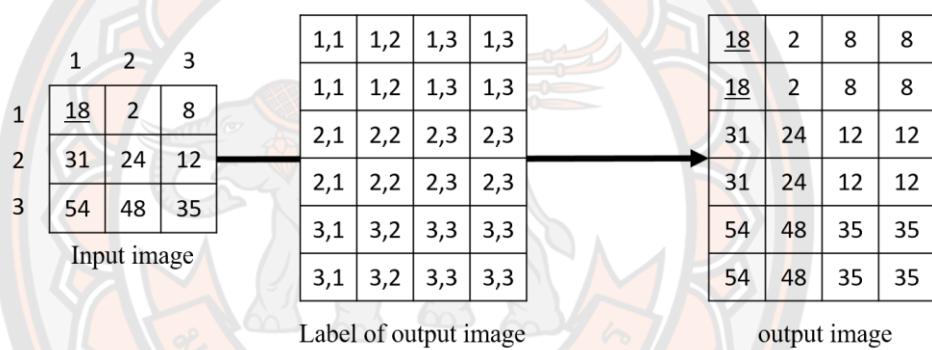
เมื่อ	$Row_{position}$	คือ ผลลัพธ์ตำแหน่งแรกของภาพนำเข้า
	$Column_{position}$	คือ ผลลัพธ์ตำแหน่งหลักของภาพนำเข้า
	Rv	คือ ตำแหน่งพิกเซลของเวกเตอร์แครบันภาพผลลัพธ์
	Cv	คือ ตำแหน่งพิกเซลของเวกเตอร์หลักบนภาพผลลัพธ์

จากสมการที่ 13 และ 14 เมื่อนำมาใช้กับตัวอย่างจากภาพ 10 ได้ผลลัพธ์ตำแหน่งหลักของภาพนำเข้า เป็น $\frac{[1 \ 2 \ 3 \ 4]}{1.33} = [0.75 \ 1.5 \ 2.25 \ 3]$ ทำให้เป็นจำนวนเต็มโดยการปัดเศษนิยมให้ค่าผลลัพธ์มีค่ามากกว่าหรือเท่ากับค่าก่อนการปัดเศษนิยม จึงได้ผลลัพธ์ของตัวอย่างเป็น $[1 \ 2 \ 3 \ 3]$ และ $\frac{[1 \ 2 \ 3 \ 4 \ 5 \ 6]^T}{2} = [0.5 \ 1 \ 1.5 \ 2 \ 2.5 \ 3]^T$ เป็นผลลัพธ์ตำแหน่งแรกของภาพนำเข้าเมื่อทำการปัดเศษนิยมจึงได้ผลลัพธ์เป็น $[1 \ 1 \ 2 \ 2 \ 3 \ 3]^T$ และนำตำแหน่งทั้งหมดและหลักมาจัดเรียงบนภาพผลลัพธ์จะได้ผลลัพธ์ดังภาพ 11

1,1	1,2	1,3	1,3
1,1	1,2	1,3	1,3
2,1	2,2	2,3	2,3
2,1	2,2	2,3	2,3
3,1	3,2	3,3	3,3
3,1	3,2	3,3	3,3

ภาพ 11 ผลลัพธ์ตำแหน่งพิกเซลของภาพนำเข้า

ขั้นตอนที่ 2.2.1.2.3 ค่าของแต่ละพิกเซลจากภาพ 11 เป็นตำแหน่งอ้างอิงพิกเซลของภาพนำเข้า ขั้นตอนนี้จะทำการนำค่าจากตำแหน่งที่อ้างอิงจากภาพนำเข้ามาแทนลงในพิกเซลบนภาพผลลัพธ์ ดังภาพ 12



ภาพ 12 ผลลัพธ์การประมาณค่าในช่วงแบบเพื่อบ้านใกล้สุด

2.2.2 การแปลงเรดอน (Radon transform)

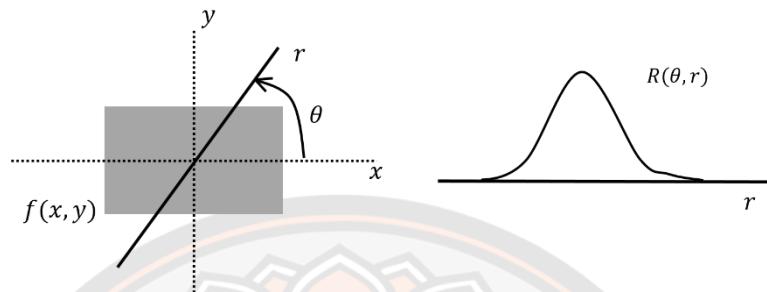
การแปลงเรดอนเป็นการแปลงบนระนาบสองมิติได้รับการคิดค้นในปี 1917 ซึ่งอยู่บนพื้นฐานของการหาพื้นที่ตลอดแนวเส้นตรงในทิศทางหรือองศาที่สามารถกำหนดได้แล้วแปลงฟังก์ชันบนโดเมนปกติให้อยู่ในโดเมนความถี่ในรูปแบบกราฟไซน์ (Sine) และเรียกผลลัพธ์จากการแปลงนี้ว่า ไซโนแกรม (Sinogram) (Radon, 1986) โดยสมการการหาพื้นที่ของฟังก์ชัน $f(x, y)$ ในทิศทางองศา (θ) จากแนวแกน x ดังต่อไปนี้

$$R(\theta, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(r - x \cos \theta - y \sin \theta) dx dy \quad (15)$$

- เมื่อ $R(\theta, r)$ คือ ผลลัพธ์พื้นที่ตลอดแนวเส้นตรง r
 δ คือ เดลต้าฟังก์ชัน (Dirac delta function) ดังสมการที่ 16
 r คือ เส้นตรงที่ตัดผ่านจุดกำเนิดซึ่งทำมุมองศา θ จากแนวแกน x

$$\delta(c) = \begin{cases} +\infty & , c = 0 \\ 0 & , c \neq 0 \end{cases} \quad (16)$$

เมื่อ c คือ สมาชิกของจำนวนจริง (Real number)



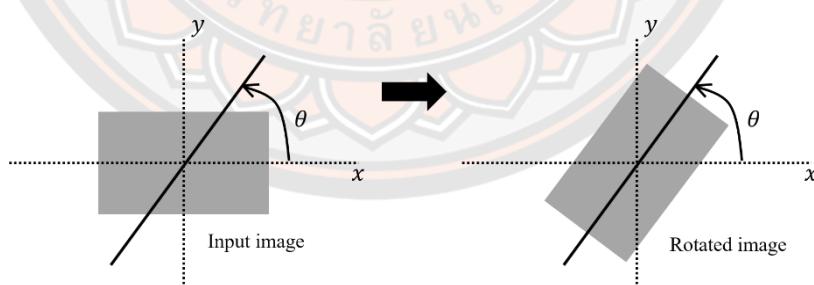
ภาพ 13 การหาพื้นที่ตลอดแนวเส้นตรงของการแปลงเรเดอน

พื้นที่ $R(\theta, r)$ จะถูกนำไปแปลงให้อยู่ในโดเมนความถี่โดยการแปลงฟูเรียร์ (Fourier transform) และทฤษฎีฟูเรียร์สไลซ์ (Fourier slice theorem) และเมื่อนำการแปลงเรเดอนมาประยุกต์ใช้ในการประมวลผลภาพ (Image processing) ทำให้กระบวนการแปลงเรเดอนมีขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดคลุ่มของทิศทางหรือองศา (Degree) จากแนวแกน x

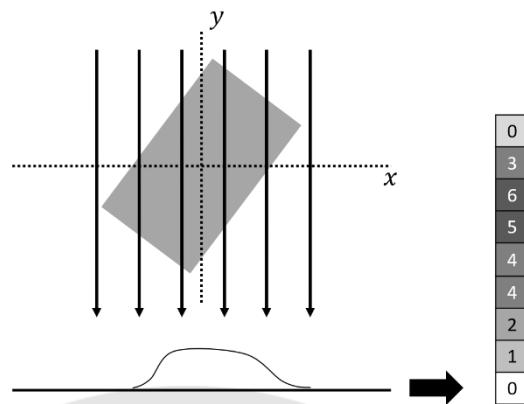
ขั้นตอนที่ 2 ใช้การหมุนภาพจากการแปลงภาพเชิงมิติในองศาที่กำหนด โดยที่ขนาดของภาพ

ผลลัพธ์จากการหมุนจะไม่เปลี่ยนแปลงจากภาพนำเข้าดังภาพ 14



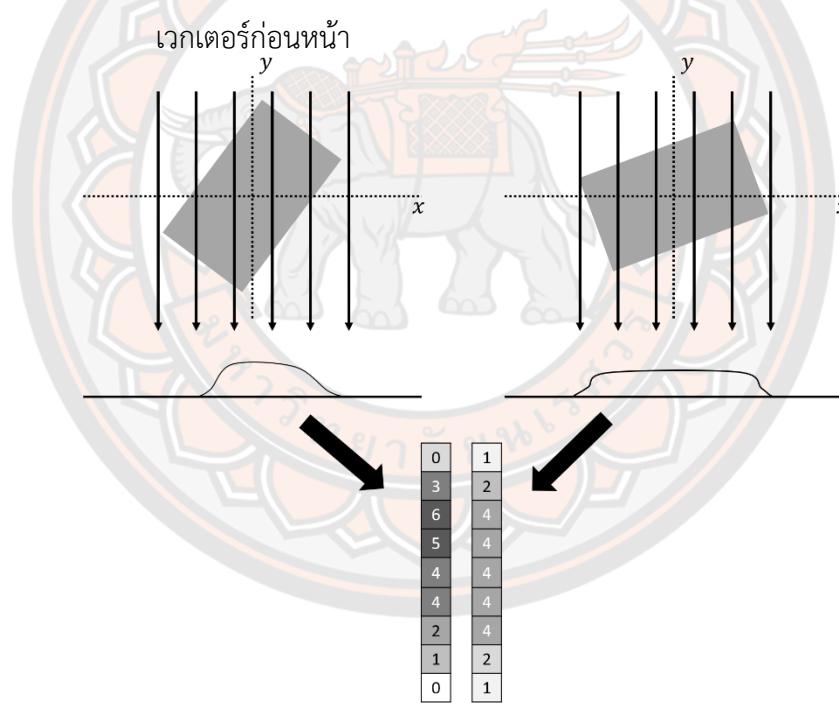
ภาพ 14 ตัวอย่างการหมุนภาพในการแปลงเรเดอน

ขั้นตอนที่ 3 หาผลรวมของค่าบนภาพทั้งหมดในทิศทางตั้งฉากกับแกน x เก็บไว้ในรูปเวกเตอร์หลัก (Column vector) ดังภาพ 15 โดยขนาดของเวกเตอร์เท่ากับค่าสูงสุดจากการเปรียบเทียบความสูงและความกว้างของภาพนำเข้า



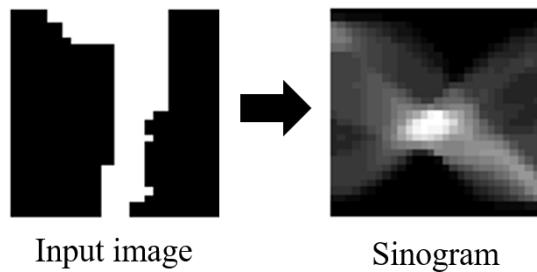
ภาพ 15 การหาผลรวมของข้อมูลในการแปลงเรดอน

ขั้นตอนที่ 4 ทำซ้ำขั้นตอนที่ 2 และ 3 แล้วนำเวกเตอร์หลักจากขั้นตอนที่ 3 มาเข้ามต่อ กับ



ภาพ 16 การนำเวกเตอร์หลักมาเข้ามต่อ กันในการแปลงเรดอน

ขั้นตอนที่ 5 ทำซ้ำขั้นตอนที่ 4 จนครบทิศทางหรือองศาที่กำหนดไว้ในขั้นตอนที่ 1 ผลลัพธ์จากการวนการแปลงเรดอนจะมีลักษณะเป็นคลื่นความถี่ใช้ ซึ่งเรียกว่าไซ-โนแกรมดังภาพ 17



ภาพ 17 ใช้โน้ตเ gramm จากการแปลงเรดอนสำหรับการประมวลผลภาพ

การแปลงเรดอนได้นำการนำมาใช้งานอย่างหลากหลายด้วยความสามารถในการสร้างภาพ (Image reconstruction) จากรังสีเอกซ์ (X-ray) ในด้านการแพทย์ ซึ่งสามารถพัฒนาได้ในกระบวนการทำงานของเครื่องเอกซ์เรย์คอมพิวเตอร์ (Computer tomography scanner) งานวิจัยในปี 2003 นำเสนอการรักษาภาพอากาศยานที่มีความหลากหลายในด้านขนาด การเอียง และตำแหน่งของอากาศยานในภาพ โดยได้นำวิธีการแปลงเรดอนมาใช้เพื่อลดผลกระทบจากความหลากหลายตั้งกล่าวในโดเมนความถี่ (Junhong, Quan, Hongcai, & Peiling, 2003)

2.2.3 โครงข่ายประสาทเทียมแบบคอนโวโลชัน (Convolutional neural network)

โครงข่ายประสาทเทียมแบบคอนโวโลชันเป็นโครงข่ายประสาทเทียมที่เพิ่มส่วนของการสกัดคุณลักษณะโดยการคอนโวโลชันด้วยตัวกรองแบบต่าง ๆ ก่อนนำไปเชื่อมต่อกับโครงข่ายประสาทเทียมแบบปกติจึงทำให้โครงข่ายประสาทเทียมแบบคอนโวโลชันมีส่วนประกอบดังนี้

ส่วนประกอบที่ 1 ชั้นคอนโวโลชัน (Convolutional layer) ใช้ตัวกรองที่หลากหลายในการสกัดคุณลักษณะจากภาพ ซึ่งตัวกรองภายในชั้นเดียวกันมีขนาดเท่ากัน แต่อาจมีขนาดต่างกันในชั้นที่แตกต่างกัน

ส่วนประกอบที่ 2 ชั้นการลดขนาดและรวมข้อมูล (Pooling layer) เป็นชั้นการตรวจสอบผลลัพธ์จากชั้นคอนโวโลชันแล้วทำการเลือกในรูปแบบต่าง ๆ จากผลลัพธ์ทั้งหมดแล้วทำการลดขนาดของข้อมูลภาพลงเพื่อความรวดเร็วของกระบวนการ และเก็บรักษาคุณลักษณะของภาพ เช่น การเลือกจากค่าสูงสุด (Max pooling) การเลือกจากค่าเฉลี่ย (Average pooling)

ส่วนประกอบที่ 3 พังค์ชันกระตุ้น (Activation function) เป็นส่วนช่วยลดระยะเวลาให้กับกระบวนการเรียนรู้ โดยการปรับค่าน้ำหนัก (Weight) ให้อยู่ในช่วงที่เหมาะสม

สถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวโลชันถูกสร้างขึ้นเพื่อนำไปใช้งานกับการรักษาภาพตัวเลขจากลายมือแล้วได้ผลลัพธ์ที่มีประสิทธิภาพ (LeCun, Bottou, Bengio, & Haffner, 1998) และเป็นต้นแบบให้กับการพัฒนาโครงข่ายประสาทเทียมแบบคอนโวโลชันเพื่อ

ประสิทธิภาพในการใช้งานที่หลากหลายมากขึ้น ซึ่ง LeNet-5 ประกอบด้วย 7 เลเยอร์ที่ไม่รวมกระบวนการนำเข้าภาพมีขนาด 28×28 พิกเซล ดังนี้

2.2.3.1 ชั้นคอนโวลูชัน (Convolutional layer) : จะนำข้อมูลภาพนำเข้าไปผ่านกระบวนการกรองภาพเพื่อให้ได้คุณลักษณะของภาพ (feature maps) รวม 6 รูปแบบ ซึ่งภาพผลลัพธ์มีขนาด 28×28 จำนวน 6 ภาพ

2.2.3.2 ชั้นการลดขนาดและรวมข้อมูล (Pooling layer) : กรองภาพด้วยตัวกรองค่าเฉลี่ยหรือการสุ่มตัวอย่างแบบบ่อย (Sub-sampling) ขนาด 2×2 พิกเซล นำมาใช้กับทุกภาพจากชั้นก่อนหน้าจะได้ภาพผลลัพธ์ขนาด 14×14 พิกเซล จำนวน 6 ภาพ

2.2.3.3 ชั้นคอนโวลูชัน : ใช้การจับคุณลักษณะขนาด 5×5 พิกเซลจำนวน 16 ตัวกรอง แต่เลือกเพียง 10 ภาพเท่านั้นจาก 16 ผลลัพธ์

2.2.3.4 ชั้นการลดขนาดและรวมข้อมูล : กรองภาพด้วยตัวกรองค่าเฉลี่ยขนาด 2×2 พิกเซล จำนวน 16 ตัวกรอง จึงได้ผลลัพธ์ 16 ภาพ

2.2.3.5 ชั้นการคอนโวลูชัน และเชื่อมต่ออย่างสมบูรณ์ (Fully connected convolution layer) : ใช้ตัวกรองคุณลักษณะ 120 ตัวกรอง ซึ่งมีขนาด 1×1 พิกเซล ทำให้ได้ผลลัพธ์ 120 ภาพ แล้วทำการเชื่อมต่อกับโครงข่ายประสาทเทียมทั้งหมด 400 โนนด

2.2.3.6 ชั้นการเชื่อมต่ออย่างสมบูรณ์ (Fully connection layer) : นำผลลัพธ์คุณลักษณะทั้งหมดจากชั้นก่อนหน้ามาเชื่อมต่อกับโครงข่ายประสาทเทียม 84 โนนด

2.2.3.7 ชั้นผลลัพธ์ (Output layer) : จะนำผลลัพธ์จากชั้นตอนก่อนหน้าทั้งหมดมาคำนวณหาความน่าจะเป็น แล้วกำหนดให้อยู่ในช่วง 0 – 1 โดยอาศัยฟังก์ชันซอฟท์แมกซ์ (Softmax function)

2.3 งานวิจัยที่เกี่ยวข้องกับการตรวจจับ

ในการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแล็อดีทางผู้วิจัยได้ทำการวิเคราะห์ข้อมูลภาพแล้วได้สังเกตถึงลักษณะของจุดภาพแล็อดีที่มีส่วนประกอบหลักคือแสงและสีโดยข้อมูลภาพจากจุดภาพแล็อดีที่ผู้วิจัยเก็บรวบรวมเพื่อพัฒนาคือสีแดง ดังนั้น การตรวจจับด้วยเงื่อนไขของสีจึงเป็นส่วนที่ผู้วิจัยพิจารณาศึกษางานวิจัยแล้วพบว่าในปี 2009 ได้มีการนำเสนอวิธีการตรวจจับและรู้จำสัญญาณไฟจราจรแบบแล็อดีที่ประกอบด้วยสีแดง เหลือง เขียว โดยการทำให้แบบจำลองสี RGB อยู่ในรูปแบบมาตรฐาน (Normalized RGB) เพื่อกำจัดเงื่อนไขทางแสง แล้วใช้เงื่อนไขเชิงขอบเขตของสีแดง (R) สีเขียว (G) และสีน้ำเงิน (B) ในการแบ่งส่วนภาพด้วย 3 ช่วงสี ดังนี้

Red : $R > 200$ and $G < 150$ and $B < 150$

Yellow : $R > 200$ and $G > 150$ and $B < 150$

Green : $R < 150$ and $G > 240$ and $B > 220$

ภาพที่ได้รับในขั้นตอนข้างต้นจะเป็นภาพขาวดำที่มีส่วนที่ผ่านเงื่อนไขเป็นสีขาว และส่วนที่เหลือเป็นสีดำ หลังจากนั้นตรวจจับเส้นขอบโดยใช้ตัวกรองโซเบล (Sobel mask) เพื่อตรวจจับบริเวณของสัญญาณไฟจราจร (Omachi & Omachi, 2009) ต่อมาในปี 2012 การตรวจจับและรู้จำแอปเปิลจากภาพถ่ายได้รับการพัฒนาเพื่อการใช้งานกับหุ่นยนต์เก็บเกี่ยว โดยใช้วิธีการที่เรียกว่า 2R-G-B แปลงภาพถ่ายสีให้เป็นภาพระดับสีเทาดังสมการต่อไปนี้

$$(x, y) = \begin{cases} 0, & 2R \leq G + B \\ 2R - G - B, & \text{other} \\ 255, & 2R \geq G + B + 255 \end{cases} \quad (17)$$

เมื่อ R, G, B คือ ค่าของสีแดง สีเขียว และสีเหลือง ตามลำดับบนแบบจำลองสี RGB

แล้วใช้วิธีการของ Otsu (Nobuyuki Otsu, 1979) ในการแปลงภาพให้เป็นขาวดำเพื่อแบ่งส่วนภาพ และทำการกำจัดสัญญาณรบกวน โดยกระบวนการการเปิดและการปิด (Opening and closing operation) หลังจากนั้นตรวจจับบริเวณของแอปเปิลโดยการตรวจจับเส้นขอบ (Edge detection) และการตรวจจับวงกลม (Circle detection) โดยใช้การแปลงชัฟแบบวงกลม (The circle Hough Transform) ตามลำดับ (G.-Q. Jiang & Zhao, 2012) ซึ่งทั้งสองงานวิจัยข้างต้นนี้ใช้แบบจำลองสี RGB ในการนำส่วนที่ต้องการออกมากจากภาพแล้วให้ผลลัพธ์การตรวจจับบริเวณที่สนใจได้เป็นอย่างดี แต่การนำวิธีการนี้มาใช้กับการตรวจจับตัวเลขซึ่งมีรายละเอียดของโครงสร้างที่มากกว่าการตรวจจับบริเวณวงกลม รวมไปถึงการนำแบบจำลองสี RGB มาใช้งานกับตัวเลขที่มีส่วนประกอบเป็นแสงจากแหล่งอิฐน้ำทำได้ยาก และทางผู้วิจัยได้ทดลองใช้การตรวจจับเส้นขอบกับข้อมูลตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของการแปลงอิฐแล้วพบว่าส่วนที่ไม่ใช่ตัวเลขปะปนอยู่อีกด้วยดังภาพ 2 ทางผู้วิจัยจึงศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจจับโดยใช้แบบจำลองสีรูปแบบอื่น จึงพบว่าได้มีงานวิจัยที่นำแบบจำลองสี HSV มาพัฒนาในการตรวจสอบสัญญาณไฟจราจรเพื่อการใช้งานบนยานพาหนะไร้คนขับในปี 2015 โดยแบ่งส่วนภาพโดยการกำหนดเกณฑ์สีที่เหมาะสม (Transcendental color threshold method) เพื่อตรวจจับภาพ โดยมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 แปลงช่วงของสี (H) โดยการเพิ่มองศาให้กับทุกค่าสี 60 องศาดังสมการต่อไปนี้

$$H_m = \begin{cases} H + 60, & H \leq 300 \\ H - 300, & H > 300 \end{cases} \quad (18)$$

เมื่อ H_m คือ ผลลัพธ์จากการแปลงช่วงของสีซึ่งให้ผลลัพธ์ดังภาพ 18



ภาพ 18 ผลลัพธ์การแปลงของสีในแบบจำลองสี HSV โดย Mu และคณะ

ขั้นตอนที่ 2 นำส่วนที่คาดว่าจะเป็นสัญญาณไฟจราจรออกจากภาพโดยการพิจารณาค่าของสี ความอิมสี และความสว่างของแต่ละพิกเซล (x, y) ดังนี้

$$(x, y) = \begin{cases} red, & H_m \in [41.8, 66.6], S \geq 100, V \geq 50 \\ yellow, & H_m \in [89.1, 136.1], S \geq 100, V \geq 50 \\ green, & H_m \in [173.1, 216.3], S \geq 100, V \geq 50 \\ other, & otherwise \end{cases} \quad (19)$$

หลังจากนั้นใช้ Support vector machine (SVM) ในการรู้จำคุณลักษณะจากการไอลร์ดับสีไฮสโต-แกรม (Histogram oriented gradients หรือ HOG) ซึ่งผู้วิจัยได้นำวิธีการกำหนดเกณฑ์สีที่เหมาะสมนี้มาใช้งานกับข้อมูลของผู้วิจัยเองแล้วได้ผลลัพธ์การแบ่งส่วนภาพดังภาพ 19 (Mu, Xinyu, Deyi, Tianlei, & Lifeng, 2015)



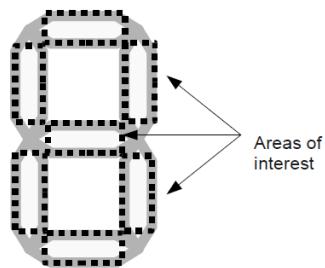
ภาพ 19 ตัวอย่างผลลัพธ์การใช้งานวิธีการตรวจจับของ Mu

ผู้วิจัยได้ทำการวิเคราะห์สาเหตุของผลลัพธ์ข้างต้นแล้วพบว่าการนำส่วนของความอิมสีมาใช้งานนั้น ส่งผลกระทบต่อบริเวณโดยรอบของตัวเลข รวมไปถึงการกำหนดเกณฑ์สีที่กำหนดไว้บ้านไม่สอดคล้อง กับข้อมูลภาพมากพอเนื่องจากรายละเอียดของตัวเลขที่มากกว่าและมีลักษณะไม่อยู่ร่วมกันเป็นกลุ่ม เมื่อเปรียบเทียบกับสัญญาณไฟจราจร ทำให้ผลลัพธ์ที่ได้เป็นตัวเลขที่ไม่สมบูรณ์

2.4 งานวิจัยที่เกี่ยวกับการรู้จำ

ผู้วิจัยได้ศึกษาวิธีการในการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนพบว่าในปี 2009 มีการนำวิธีการอ่านอักษรระดับด้วยแสง (OCR) มาพัฒนาให้สามารถรู้จำตัวเลขรูปแบบเจ็ดส่วนได้ (Bonacic et al., 2009) โดยนำ 3 เทคนิคที่สามารถสกัดคุณลักษณะที่แตกต่างกันมาใช้งาน ได้แก่

เทคนิคที่ 1 พิจารณาบริเวณที่สนใจ (Interest area method) จะพิจารณาคุณลักษณะจากผลรวมของพิกเซลในบริเวณที่สนใจดังภาพ 20



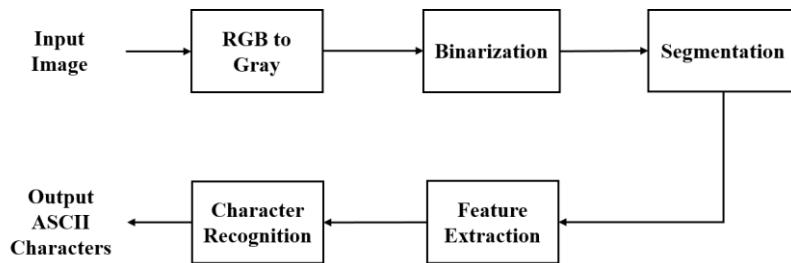
ภาพ 20 ตัวอย่างบริเวณที่สนใจ

ที่มา : (Bonacic et al., 2009)

เทคนิคที่ 2 การลดความละเอียด (The resolution reduction technique) จะทำการลดขนาดของภาพ (Down-sampled image) ที่มีตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนหนึ่งหลัก ซึ่งค่าในภาพที่ลดขนาดมาแล้วนั้น คือ คุณลักษณะของตัวเลข

เทคนิคที่ 3 การวิเคราะห์การกระจายบนแกน (The axis distribution analysis method) คือ การสกัดคุณลักษณะจากผลรวมของพิกเซลเชิงโครงสร้างของตัวเลขในแนวตั้งและแนวนอนเก็บค่าไว้ในເຕັກເຫຼັກແລະເວັກເຫຼັກແລ້ວມາດູວ່າມານີ້ມີພົບພັນທີ່ສຳເນົາ

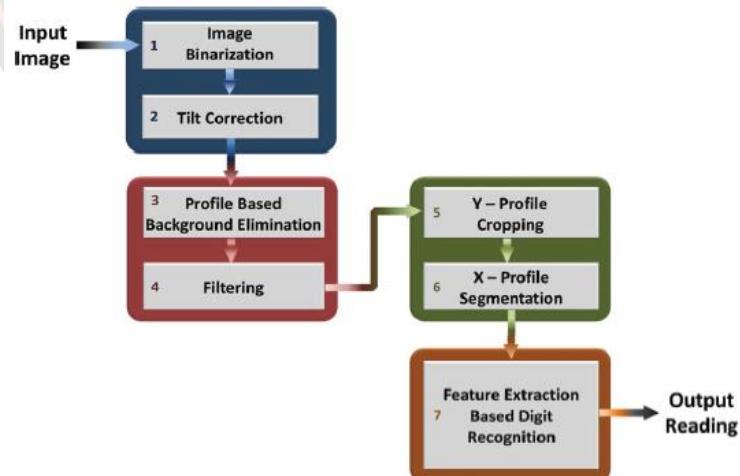
หลังจากการสกัดคุณลักษณะแล้วจะนำคุณลักษณะเหล่านั้นไปที่โครงข่ายประสาทเทียม (Feed-forward neural network) เพื่อทำการรู้จำตัวเลขจากคุณลักษณะที่หลากหลาย และใช้โครงข่ายประสาทเทียมแบบต่าง ๆ ในการกำหนดค่าน้ำหนัก (Weight) ที่เหมาะสมที่สุดให้กับคุณลักษณะเหล่านั้น (Committee neural networks) และใช้ Topology of neural networks ใน การเปรียบเทียบคุณภาพจากผลการใช้งานโครงข่ายประสาทเทียม และในระยะเวลาที่ใกล้เคียงกัน ได้มีการพัฒนาการทำงานของการอ่านอักษรระดับด้วยแสงให้สามารถทำงานแบบอัตโนมัติโดยการสแกนหน้าจอแสดงผลแบบเจ็ดส่วนผ่านภาพจากกล้อง (Ghugardare et al., 2009) โดยมีแนวคิดดังต่อไปนี้



ภาพ 21 กรอบแนวคิดของ Rakhi P. Ghugardare

ที่มา : (Ghugardare et al., 2009)

เริ่มจากการแปลงภาพนำเข้าเป็นภาพระดับสีเทาแล้วแปลงเป็นภาพขาวดำด้วยวิธีการของ Otsu (Otsu's method) เปรียบเทียบการแบ่งส่วนภาพ (blob method และ Peak to valley) แล้วเลือกผลลัพธ์ที่ดีที่สุดหลังจากนั้นทำการจำจัดสัญญาณรบกวนภาพก่อนนำไปสกัดคุณลักษณะโดย การแบ่งภาพตัวเลขแต่ละหลักออกเป็น 12 บริเวณที่เท่ากัน ซึ่งลักษณะพิกเซลสีดำหรือส่วนที่เป็น ตัวเลขภายในแต่ละบริเวณ เช่น ขอบ และมุม จะเป็นส่วนสำคัญในการแยกแยะหมวดหมู่ของตัวเลข เพื่อทำการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน หลังจากนั้นในปี 2016 ได้มีการนำเสนองานวิจัย เกี่ยวกับการรู้จำตัวเลขจากหน้าจอแสดงผลแบบเจ็ดส่วนบนซอฟต์แวร์ Kulkarni & Kute, 2016) โดยมี 7 กระบวนการดังภาพ 22



ภาพ 22 กรอบแนวคิดของ Kulkarni

ที่มา : (Kulkarni & Kute, 2016)

กระบวนการที่ 1 Image binarization เป็นการดึงส่วนที่เป็นตัวเลขออกจากภาพโดยใช้ขอบเขตของสีจากโมเดลสี RGB ซึ่งจะเปลี่ยนให้ส่วนที่เป็นตัวเลขเป็นสีดำ และส่วนอื่นที่อยู่นอกขอบเขตของสีมีสีขาว

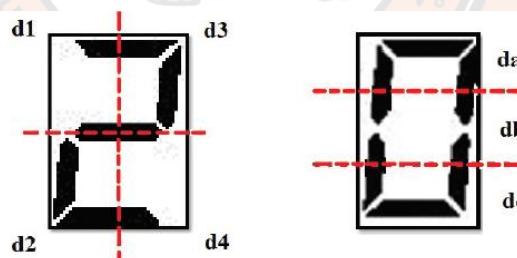
กระบวนการที่ 2 Tilt correction ลดปัญหาการเอียงของภาพถ่ายโดยใช้การแปลงเรเดอน (Radon transform) ซึ่งกำหนดช่วงมุมจากแกนแนวตั้งไว้ที่ 90° ในทิศทางทวนเข็มนาฬิกา ถึง 90° ในทิศทางตามเข็มนาฬิกา

กระบวนการที่ 3 Profile based background elimination เป็นการกำหนดขอบเขตของส่วนที่พิจารณาในแนวตั้งและแนวนอนเพื่อกำจัดส่วนรอบนอกสีดำที่ไม่ใช่ตัวเลข

กระบวนการที่ 4 Filtering เป็นขั้นตอนการกรองภาพด้วยตัวกรองค่าเฉลี่ยความถี่ต่ำ (low pass filter) ที่มีขนาด 3×3 เพื่อกำจัดสัญญาณรบกวนที่ปะปนอยู่กับตัวเลขออกจากภาพ

กระบวนการที่ 5 และ 6 Cropping and segmentation เป็นการตัดส่วนที่เป็นเพียงตัวเลขออกจากภาพ โดยการกำหนดขอบเขตของพิกเซลสีดำตามความสูง และความกว้างตามลำดับ แล้วค่าความกว้างของแต่ละตัวเลขที่แบ่งออกจากจะนำมาใช้เป็นเกณฑ์กำหนดว่าวัตถุนั้นเป็นตัวเลขหรือไม่

กระบวนการที่ 7 Feature extraction based digit recognition สกัดคุณลักษณะจากความหนาแน่นของพิกเซลสีดำเพื่อระบบการรู้จำตัวเลข โดยจะแปลงภาพตัวเลขเพื่อนำไปพิจารณาทีละส่วนด้วยต้นไม้ตัดสินใจ (decision tree) ในการรู้จำตัวเลขดังภาพ 23



ภาพ 23 ความหนาแน่นของพิกเซลเพื่อระบบการรู้จำตัวเลข

ที่มา : (Kulkarni & Kute, 2016)

โดยที่ d_i เป็นความหนาแน่นของพิกเซลสีดำ เมื่อ $i \in \{1, 2, 3, 4, a, b, c\}$ ซึ่งทางผู้วิจัยได้มุ่งเน้นไปที่การแก้ปัญหาของภาพถ่ายที่อาจได้รับผลกระทบจากการเอียงของการถ่ายภาพโดยใช้การแปลงเรเดอน ผู้วิจัยจึงศึกษาทฤษฎีรวมถึงการใช้งานการแปลงเรเดอนเพื่อนำมาประยุกต์ใช้ในงานวิจัย

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนทำให้เห็นว่า รูปแบบของตัวเลขเป็นส่วนสำคัญในการนำมาสกัดคุณลักษณะเพื่อทำการรู้จำ และผู้วิจัยได้ทดลองนำคุณลักษณะเหล่านี้มาทำการรู้จำโดยการจับคู่คุณลักษณะ (Features matching) ซึ่งให้ผลลัพธ์ที่น่าพึงพอใจกับข้อมูลที่เป็นภาพตัวเลขในมุมมองตรงและอาจมีสัญญาณรบกวนเล็กน้อย แต่เมื่อพิจารณา กับภาพถ่ายโดยธรรมชาติที่ไม่สามารถกำหนดระยะเวลาถ่ายภาพได้อย่างคงที่ในทุกการถ่ายภาพ รวมถึง มุมมองที่หลากหลายในการถ่ายภาพจะส่งผลให้การรู้จำจากคุณลักษณะเหล่านี้มีประสิทธิภาพน้อยลง อย่างมาก ดังนั้น ทางผู้วิจัยจึงมุ่งเน้นการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการรู้จำ



บทที่ 3

วิธีดำเนินงานวิจัย

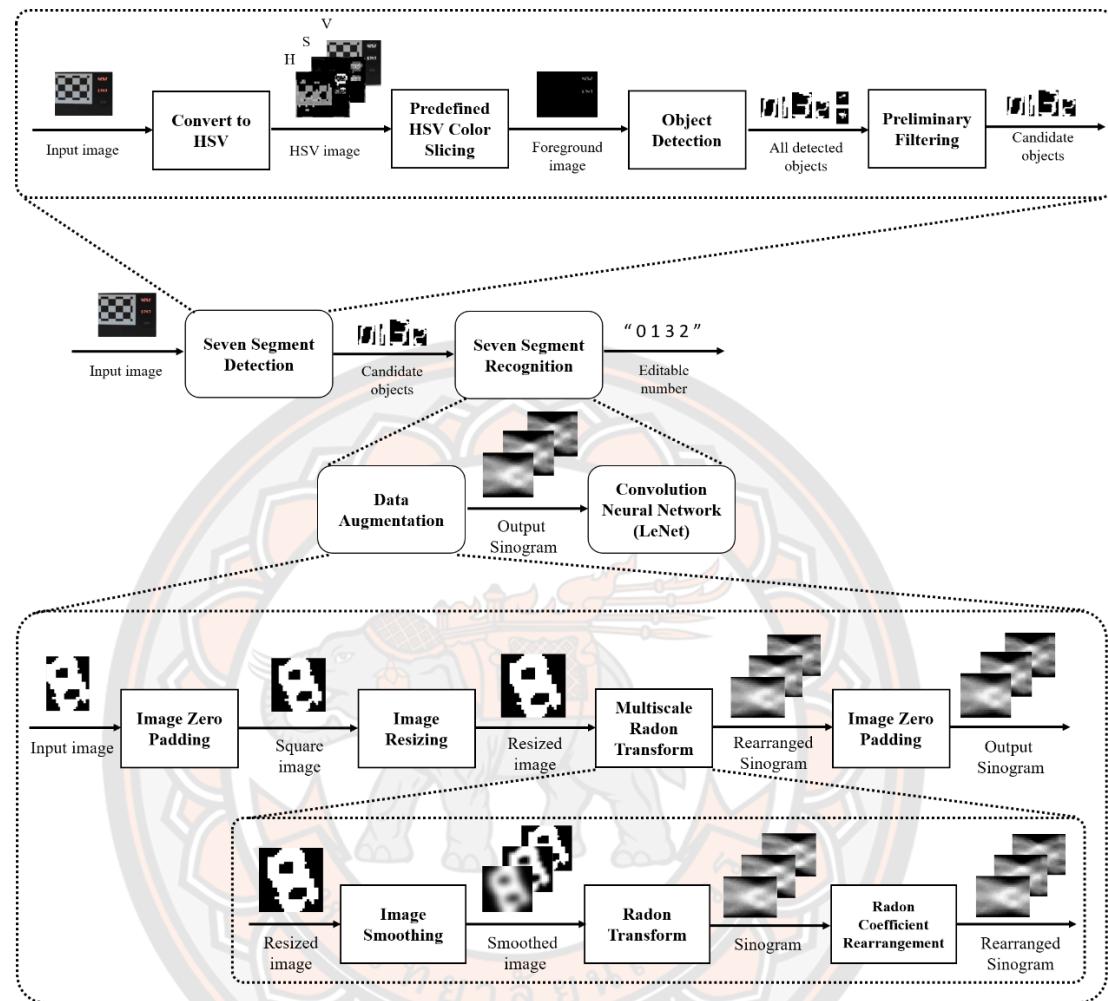
งานวิจัยเรื่องการตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม
ผู้วิจัยได้แบ่งการดำเนินงานวิจัยออกเป็น 2 ส่วน ดังนี้

**ส่วนที่ 1 การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี ซึ่งผู้วิจัยได้
นำเสนอบริการตัดแบบกำหนดช่วงบนแบบจำลองสี HSV (Predefined HSV Color Slicing) ในงาน
ประชุมวิชาการ 2019 - 16th International Joint Conference on Computer Science and
Software Engineering (JCSSE) ในหัวข้อ Seven Segment Display Detection and Recognition
using Predefined HSV Color Slicing Technique (Popayorm, Titijaroonroj, Phoka, &
Massagram, 2019)**

**ส่วนที่ 2 การรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี ซึ่งผู้วิจัยได้นำเสนอ
การแปลงเรเดอนหลายขนาด (Multiscale Radon transform หรือ MRT) ในงานประชุมวิชาการ
2020 - 12th International Conference on Knowledge and Smart Technology (KST) ใน
หัวข้อ Data Augmentation Based on Multiscale Radon Transform for Seven Segment
Display Recognition (Popayorm, Titijaroonroj, Phoka, & Massagram, 2020)**

วิธีการดำเนินงานของแต่ละส่วนจะประกอบไปด้วย ขอบเขตของข้อมูล การออกแบบและ
พัฒนาวิธีการ และวิธีการวัดประสิทธิภาพตามลำดับ โดยกระบวนการที่ผู้วิจัยนำเสนอ มีกรอบแนวคิด
แสดงดังภาพ 24

กรอบแนวคิด



ภาพ 24 กรอบแนวคิดของวิธีการที่นำเสนอ

3.1 การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี

3.1.1 ขอบเขตของการดำเนินการ

3.1.1.1 ขอบเขตของข้อมูล

ในการเก็บรวบรวมข้อมูลเพื่อใช้ในการดำเนินงานวิจัย ผู้วิจัยใช้อุปกรณ์ในการจำลองข้อมูลที่มีลักษณะดังนี้

- เป็นภาพสี
- มีตัวเลขจากหน้าจอแสดงผลแบบเจ็ดส่วนบนจากการแอลอีดีสีแดง
- หมุนมองตรงกับหน้าจอแสดงผล
- มีตัวเลขที่แสดงในภาพแต่ละภาพทั้งหมด 8 ตัวเลข

- ขนาดภาพ 1280 × 720 พิกเซล
- จำนวน 240 ภาพ
- ระยะห่างระหว่างจ่อแสดงผลกับกล้องถ่ายภาพประมาณ 1 เมตร
- ไม่มีการรับ光จากสภาพแวดล้อม สัญญาณรบกวน และพฤติกรรมการถ่ายภาพ

3.1.1.2 ขอบเขตของอุปกรณ์

อุปกรณ์ในการเก็บภาพ คือ Logitech HD pro webcam c920



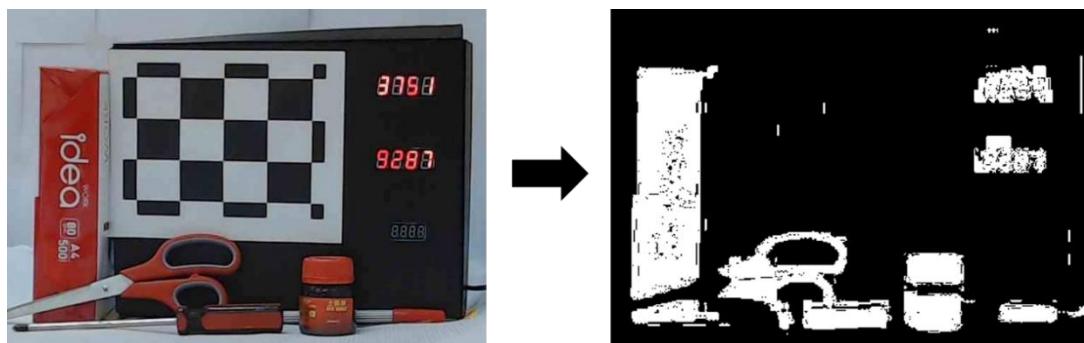
ภาพ 25 ตัวอย่างข้อมูลสำหรับการตรวจสอบจับตัวเลขที่แสดงผลแบบเจ็ดส่วน

3.1.2 การออกแบบและพัฒนาวิธีการ

ผู้วิจัยมุ่งเน้นไปที่การพิจารณาคุณลักษณะโดยทั่วไปของตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดีเป็นสำคัญเพื่อให้ได้มาเพียงตัวเลขที่ปรากฏบนภาพเท่านั้นแล้วส่งไปยังกระบวนการรุ้งจำในส่วนถัดไป ซึ่งขั้นตอนในการตรวจสอบจับหน้าจอแสดงผลแบบเจ็ดส่วนแบ่งเป็น 4 ขั้นตอนดังนี้

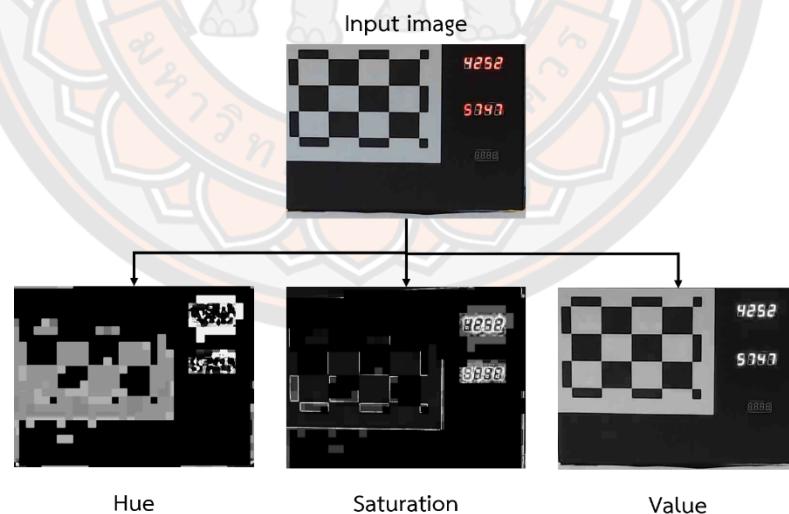
ขั้นตอนที่ 3.1.2.1 การแปลงแบบจำลองสี RGB เป็น HSV (Convert to HSV)

ผู้วิจัยได้ทำการวิเคราะห์โครงสร้างของตัวเลขที่แสดงผลแบบเจ็ดส่วนบนจากการแอลอีดีแล้วได้ทำการแบ่งส่วนประกอบของตัวเลขนี้ออกเป็น 2 ส่วนคือ สี และแสงที่มีแหล่งกำเนิดจากแอลอีดี ด้วยส่วนประกอบเหล่านี้ทำให้การนำส่วนที่เป็นตัวเลขออกจากภาพด้วยเงื่อนไขทางสีเพียงอย่างเดียว นั้นทำได้ยาก เช่น ถ้าหากภาพถ่ายมีวัตถุอื่นที่มีสีเดียวกันกับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนดังภาพ 26



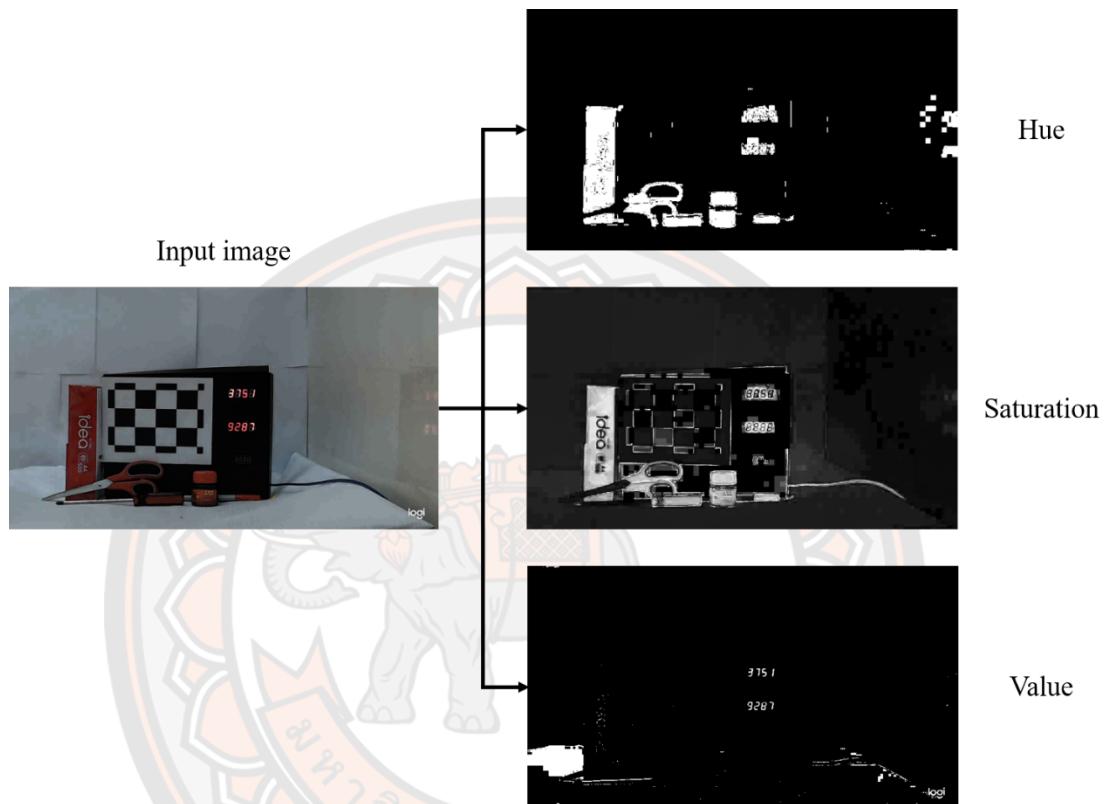
ภาพ 26 ผลลัพธ์การตัดภาพด้วยสีจากแบบจำลองสี HSV

ผู้วิจัยจึงนำแบบจำลองสี HSV ที่สามารถควบคุมทั้งส่วนของสีและความสว่างได้โดยตรงมาใช้ในการวิจัยครั้งนี้ โดยนำเข้ารูปภาพจากนั้นคำนวณส่วนประกอบของแบบจำลองสี HSV จากแบบจำลองสี RGB ซึ่งจะแบ่งส่วนประกอบออกเป็น 3 ส่วน คือ Hue ส่วนของสี Saturation ส่วนของความอิมสีส่งผลต่อความสดของเนื้อสี และ Value หรือ Brightness ส่วนของค่าความสว่างและความเข้มของสีดังภาพ 27 (ผู้วิจัยได้อธิบายการคำนวณแบบจำลองสี HSV จากแบบจำลองสี RGB ไว้ในหัวข้อที่ 2.1.3.2.2 บทที่ 2)



ภาพ 27 ตัวอย่างผลลัพธ์แบบจำลองสี HSV จากแบบจำลองสี RGB

ขั้นตอนที่ 3.1.2.2 การตัดแบบกำหนดช่วงบนแบบจำลองสี HSV (Predefined HSV Color Slicing) เนื่องจากภาพนำเข้าประกอบไปด้วยตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน และพื้นหลังรวมไปถึงวัตถุอื่นที่ปรากฏอยู่บนภาพ และปัญหาที่เกิดขึ้นเมื่อใช้เพียงส่วนใดส่วนหนึ่งของแบบจำลองสี HSV ดังภาพ 28



ภาพ 28 ผลลัพธ์แบบจำลองสี HSV จากแบบจำลองสี RGB จากภาพถ่ายวัตถุสีแดง

จากการจะเห็นว่าส่วนของสีจะนำเสนอบริเวณของช่วงสีที่กำหนดเท่านั้น ส่วนของความอิมสีนั้นมีค่าสูงเฉพาะวัตถุอื่นและบริเวณโดยรอบของตัวเลขเนื่องจากแหล่งกำเนิดของตัวเลขเป็นแหล่งอีดี ทำให้ส่วนกลางของตัวเลขมีสีแดงที่ความอิมสีต่ำ และส่วนของความสว่างนั้นให้ผลลัพธ์ตัวเลขที่สมบูรณ์แต่ยังมีบริเวณอื่นที่ไม่ใช่ตัวเลขด้วยเช่นกัน ดังนั้น ขั้นตอนนี้จึงเป็นการนำส่วนที่เป็นตัวเลขออกจากภาพด้วยวิธีการกำหนดค่าเชิงขอบเขตของสี H และค่าความสว่าง V จาก HSV โดยเลือกสีที่ต้องการพิจารณาตามสีของแหล่งอีดี ให้ x และ y แทนตำแหน่งพิกเซลของภาพแล้วและหลักตามลำดับ แล้วใช้สมการต่อไปนี้ในการนำส่วนที่เป็นตัวเลขออกจากภาพ

$$I_h(x, y) = \begin{cases} 1, & t_{left} \leq H(x, y) \leq t_{right} \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

$$I_v(x, y) = \begin{cases} 1, & 204 \leq V(x, y) \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

โดยที่ $I_h(x, y)$	คือ ค่าสี ณ พิกเซล (x, y)
$I_v(x, y)$	คือ ค่าความสว่าง ณ พิกเซล (x, y)
t_{left} และ t_{right}	คือ ค่ากึ่งกลางระหว่างสีที่เลือกับสีที่อยู่ใกล้เคียงกันทางซ้าย และทางขวาตามลำดับ

การหาค่ากึ่งกลางระหว่างสีที่เลือกับสีใกล้เคียงทางซ้ายและทางขวา เช่น เมื่อพิจารณาสีเหลืองมีค่าของสีเป็น 60° และสีที่อยู่ใกล้เคียงกับสีเหลืองทางซ้ายและทางขวา คือ สีแดงมีค่าสีเป็น 0° และสีเขียวมีค่าสีเป็น 120° ตามลำดับ จึงสามารถคำนวณค่ากึ่งกลางทางซ้ายของสีเหลืองเป็น $\frac{60^\circ+0^\circ}{2}$ นั่นคือ 30° และค่ากึ่งกลางทางขวาของสีเหลืองเป็น $\frac{60^\circ+120^\circ}{2}$ มีค่าเป็น 90° ดังนั้น สีเหลืองที่ทางผู้วิจัยนำเสนอจึงมีค่าอยู่ในช่วง 31° จนถึง 90° ซึ่งช่วงของสีต่าง ๆ ได้นำเสนอในตาราง 2

ตาราง 2 ค่าขอบเขต หรือช่วงของสีที่ผู้วิจัยนำเสนอ

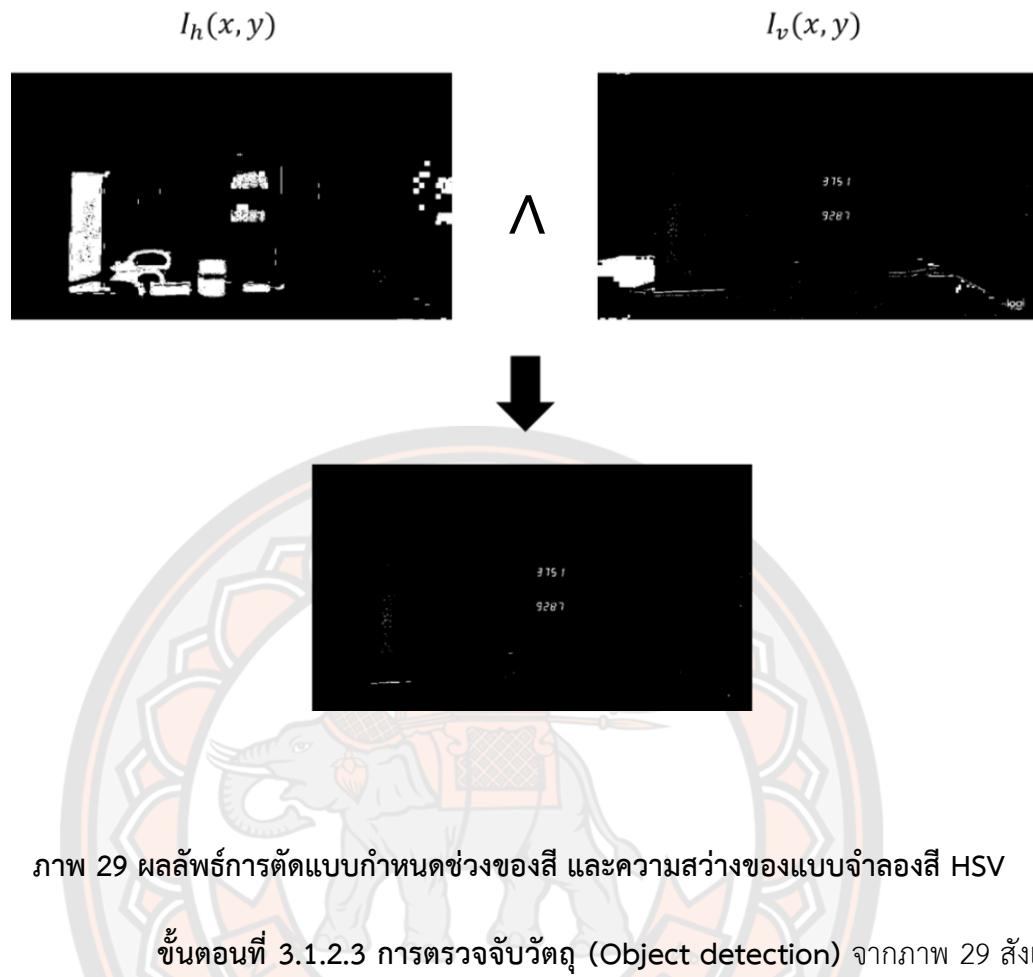
Color type	Red	Yellow	Green	Cyan	Blue	Magenta
t_{left}	331°	31°	91°	151°	221°	271°
t_{right}	30°	90°	150°	210°	270°	330°

การตัดแบบกำหนดช่วงสีทำได้ดังสมการต่อไปนี้

$$I_{out}(x, y) = I_h(x, y) \wedge I_v(x, y) \quad (22)$$

เมื่อ $I_{out}(x, y)$ คือ ผลลัพธ์การแบ่งส่วนภาพ
 \wedge คือ ความสัมพันธ์ “และ” ซึ่งจะให้ค่าเป็นจริงเมื่อ p เป็นจริง และ q
 เป็นจริงในความสัมพันธ์ $p \wedge q$ เท่านั้น

ผลลัพธ์จากการที่ 22 มีลักษณะดังภาพ 29



ภาพ 29 ผลลัพธ์การตัดแบบกำหนดช่วงของสี และความสว่างของแบบจำลองสี HSV

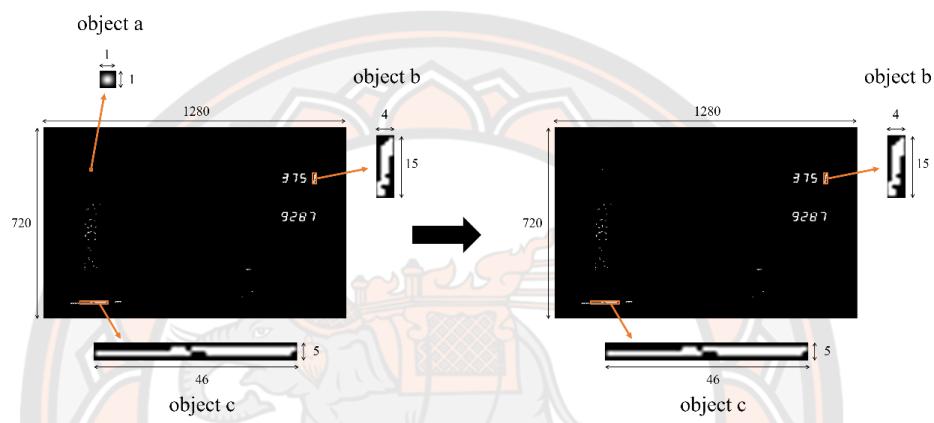
ขั้นตอนที่ 3.1.2.3 การตรวจจับวัตถุ (Object detection) จากภาพ 29 สังเกตได้ว่าส่วนที่ได้มาจากการตรวจจับวัตถุในรูปแบบเจ็ดส่วน เพื่อการจัดการกับตัวเลขและส่วนประกอบต่าง ๆ บนภาพ ขั้นตอนนี้จึงจัดกลุ่มพิกเซลจากเส้นทางที่อยู่ติดกันเป็นวัตถุเดียวกันโดยการเชื่อมต่อสองมิติ (Two-dimensional connectivity) ซึ่งจะทำการเชื่อมต่อพิกเซลที่มีค่าเดียวกันและอยู่ติดกันใน 4 หรือ 8 ทิศทางรอบพิกเซลที่พิจารณาให้เป็นวัตถุเดียวกัน (Haralick & Shapiro, 1992) โดยจะได้ผลลัพธ์เป็นวัตถุในภาพทั้งหมดซึ่งอาจมีวัตถุที่ไม่ใช่ตัวเลขปะปนอยู่ด้วย ซึ่งในงานวิจัยฉบับนี้ผู้วิจัยจัดกลุ่มโดยใช้ 8 ทิศทางรอบพิกเซลที่พิจารณาได้รับผลลัพธ์ดังภาพ 30



ภาพ 30 ผลลัพธ์การตรวจจับวัตถุโดยการเชื่อมต่อสองมิติ

ขั้นตอนที่ 3.1.2.4 การกรองขั้นต้น (Preliminary Filtering) จากขั้นตอนก่อนหน้าทางผู้วิจัยจะทำการตัดออกวัตถุเหล่านั้นโดยพิจารณาจากโครงสร้างของวัตถุที่อาจมีลักษณะไม่เหมือนตัวเลข ดังนี้

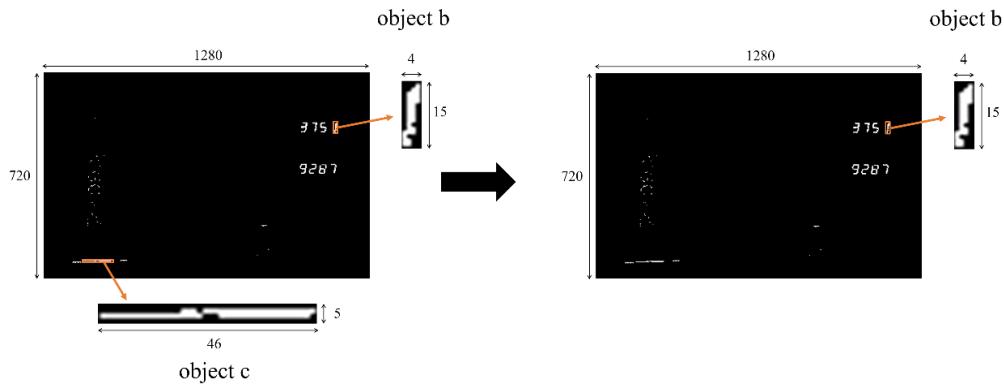
เงื่อนไขที่ 3.1.2.4.1 ขนาดของวัตถุ วัตถุที่มีขนาดโดยรวมของพิกเซลน้อยกว่า 0.005% จากขนาดภาพ (Rong-Chi, 2011) ยกเว้นการทดลองในกรณีที่ข้อมูลทดสอบมีความหลากหลายของขนาดหรือระเบยการถ่ายภาพ ซึ่งขนาดโดยรวมของพิกเซลน้อยกว่า 0.001 % ตัวอย่างเช่น ภาพ 31



ภาพ 31 ตัวอย่างผลลัพธ์จากการกรองขั้นต้นโดยเงื่อนไขขนาดของวัตถุ

จากภาพ 31 มีวัตถุที่ตรวจจับได้ทั้งหมด 3 วัตถุ คือ a b และ c ภาพถ่ายมีขนาด $1,280 \times 720$ พิกเซล ทำให้มีพิกเซลทั้งหมด $921,600$ พิกเซล ดังนั้น 0.005% ของขนาดภาพมีค่า $921,600 \times 0.00005$ นั่นคือ 46.08 พิกเซล เมื่อพิจารณาขนาดของแต่ละวัตถุจะได้ว่า วัตถุ a มีขนาด 1 พิกเซล วัตถุ b มีขนาด 60 พิกเซล และวัตถุ c มีขนาด 230 พิกเซล ดังนั้น วัตถุ a จะถูกตัดออกเหลือเพียง วัตถุ b และ c เท่านั้น

เงื่อนไขที่ 3.1.2.4.2 รูปทรงของวัตถุ วัตถุที่มีความยาวในแนวอนมากกว่าความยาวในแนวตั้ง ยกเว้นในกรณีของข้อมูลทดสอบที่มีความหลากหลายของการเอียงเนื่องจากตัวเลขที่เอียงอาจมีความก้างมากกว่าความสูง



ภาพ 32 ตัวอย่างผลลัพธ์จากการกรองขึ้นต้นโดยเงื่อนไขรูปทรงของวัตถุ

จากภาพ 32 วัตถุที่มีความกว้างมากกว่าความสูงจะถูกตัดออก ดังนั้น วัตถุ c ที่มีความกว้าง 46 และความสูง 5 จะถูกตัดออกเหลือเพียงวัตถุ b เท่านั้นที่ถูกส่งไปยังขั้นตอนการรู้จำ

3.1.3 วิธีการวัดประสิทธิภาพ

ผู้วิจัยได้ทำการวัดประสิทธิภาพ และเปรียบเทียบผลลัพธ์ของวิธีที่ผู้วิจัยได้นำเสนอ กับผลลัพธ์จากการนำวิธีการตัดภาพของงานวิจัยที่เกี่ยวข้อง 3 งานวิจัยมาประยุกต์ใช้ ได้แก่

งานวิจัยที่ 1 วิธีการตรวจจับและรู้จำไฟจราจรโดยใช้แบบจำลองสี RGB (Omachi & Omachi, 2009)

งานวิจัยที่ 2 วิธีการตรวจจับและรู้จำแอปเปิลสีแดงโดยใช้แบบจำลองสี RGB (G.-Q. Jiang & Zhao, 2012)

งานวิจัยที่ 3 วิธีการตรวจจับไฟจราจรโดยใช้แบบจำลองสี HSV (Mu et al., 2015)

การวัดประสิทธิภาพของวิธีการดังกล่าวรวมถึงวิธีการที่ได้นำเสนอผู้วิจัยได้ทำการคำนวณค่าประสิทธิภาพในแต่ละภาพแล้วนำหาค่าเฉลี่ยเทียบกับจำนวนภาพทั้งหมดที่ใช้ในการทดลอง

การประเมินผลประสิทธิภาพของการตรวจจับที่ผู้วิจัยนำมาใช้มีการวัดประสิทธิภาพใน 2 ด้านได้แก่ 1. ความแม่นยำในการตรวจจับ (Precision) ซึ่งเป็นการวัดประสิทธิภาพว่าวิธีการนั้นสามารถตรวจจับวัตถุได้ถูกต้องมากน้อยเพียงใดเมื่อเทียบกับวัตถุที่ตรวจจับได้ทั้งหมด และ 2. ค่าการเรียกคืนอักขระ (Recall) ซึ่งใช้วัดประสิทธิภาพว่าวิธีการนั้นสามารถตรวจจับวัตถุได้ถูกต้องมากน้อยเพียงใดเมื่อเทียบกับอักขระทั้งหมดในภาพทั้งหมดที่ถูกตรวจจับและไม่ถูกตรวจจับ โดยมีการคำนวณจากสมการต่อไปนี้

$$\text{Precision} = \frac{N_{CD}}{N_{AD}} \times 100 \quad (23)$$

$$Recall = \frac{N_{CD}}{N_{CD} + N_{OD}} \times 100 \quad (24)$$

- เมื่อ N_{CD} คือ จำนวนวัตถุที่เป็นอักษรที่ต้องการและตรวจจับได้ถูกต้อง
 N_{AD} คือ จำนวนวัตถุทั้งหมดที่ถูกตรวจจับทั้งเป็นและไม่เป็นอักษรที่ต้องการ
 N_{OD} คือ จำนวนวัตถุที่เป็นอักษรที่ต้องการแต่ไม่ถูกตรวจจับ

3.2 การรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี

3.2.1 ขอบเขตของการดำเนินการ

3.2.1.1 ขอบเขตของข้อมูล

ข้อมูลที่ใช้ในการทดลองแบ่งออกเป็น 2 ส่วนได้แก่ ข้อมูลสอนและข้อมูลทดสอบ

3.2.1.1.1 ข้อมูลสอน

ภาพปกติ (None) ข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดีซึ่งเป็นข้อมูลเดียวกับการทดลองการตรวจจับ แต่ทำการแบ่งตัวเลขออกจากพื้นหลังเรียบร้อยแล้วจึงมีลักษณะดังนี้

- เป็นภาพขาวดำ
- มีตัวเลขเป็นสีขาวและพื้นหลังสีดำ
- มุ่มมองตรง
- ขนาดภาพระหว่าง 3×11 พิกเซล จนถึง 10×14 พิกเซล
- ระยะห่างระหว่างจ่อแสดงผลกับกล้องถ่ายภาพประมาณ 1 เมตร
- ไม่มีการรบกวนจากสภาพแวดล้อม สัญญาณรบกวน และพฤติกรรมการถ่ายภาพ



ภาพ 33 ตัวอย่างข้อมูลภาพถ่ายตัวเลขปกติ

ภาพเหล่านี้จะนำไปผ่านกระบวนการเรียนรู้เพื่อสร้างแบบจำลองในการรู้จำ และนำมาสร้างข้อมูลชุดใหม่ โดยกระบวนการเพิ่มปริมาณข้อมูลภาพโดยการหมุน (Rotation) การย่อ-ขยาย (Scaling) และ การเอียงเชิงมุ่มมอง (Perspective) รวมถึงข้อมูลภาพไซโนแกรม จึงได้ข้อมูลสอนชุดใหม่ดังนี้

ภาพที่เพิ่มข้อมูลโดยการหมุน (Rotation)

- หมุนภาพในช่วง -60° จนถึง 60°
- ขนาดภาพระหว่าง 4×7 พิกเซล จนถึง 13×15 พิกเซล



ภาพ 34 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการหมุน

ภาพที่เพิ่มข้อมูลโดยการย่อ-ขยาย (Scaling)

- ปรับขนาดภาพในช่วง 30% จนถึง 300% ของภาพปกติ
- ขนาดภาพระหว่าง 3×7 พิกเซล จนถึง 36×52 พิกเซล



ภาพ 35 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการย่อ-ขยาย

ภาพที่เพิ่มข้อมูลโดยการเอียงเชิงมุมมอง (Perspective)

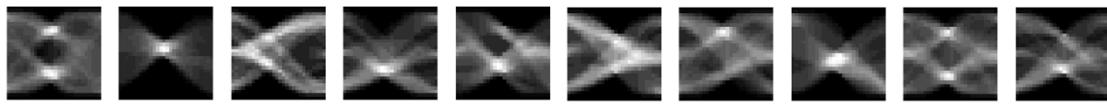
- เอียงมุมมองของภาพประมาณ -80° จนถึง 80°
- ขนาดภาพระหว่าง 2×6 พิกเซล จนถึง 10×14 พิกเซล



ภาพ 36 ตัวอย่างข้อมูลภาพที่เพิ่มปริมาณข้อมูลโดยการเอียงเชิงมุมมอง

ภาพไซโนแกรมจากวิธีการที่ผู้วิจัยนำเสนอ (MRT)

- ภาพไซโนแกรมที่ผ่านกระบวนการแปลงเรเดอน Hariyadi
- ขนาดภาพ 28×28 พิกเซล



ภาพ 37 ตัวอย่างข้อมูลภาพไซโนแกรมจากกระบวนการที่ผู้วิจัยนำเสนอ

3.2.1.1.2 ข้อมูลทดสอบ

ข้อมูลชุดนี้ทางผู้วิจัยเก็บรวบรวมโดยการถ่ายภาพตัวเลขที่แสดงผลแบบเจ็ตส่วนบนของภาพแล็ปอีดีที่อยู่ในสภาพแวดล้อมที่ผู้วิจัยจำลองขึ้น 5 ประเภท ประกอบด้วย

ภาพถ่ายปกติ (Normal) ข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ตส่วนบนของภาพแล็ปอีดีซึ่งเป็นข้อมูลที่แบ่งจากข้อมูลสอน (ภาพปกติ) มีลักษณะของข้อมูลดังนี้

- เป็นภาพขาวดำ
- มีตัวเลขเป็นสีขาวและพื้นหลังสีดำ
- มุ่งมองตรง
- ขนาดภาพระหว่าง 3×11 พิกเซล จนถึง 10×14 พิกเซล
- จำนวน 400 ภาพ
- ระยะห่างระหว่างจ่อแสดงผลกับกล้องถ่ายภาพประมาณ 1 เมตร
- ไม่มีการรบกวนจากสภาพแวดล้อม สัญญาณรบกวน และพฤติกรรมการถ่ายภาพ



ภาพ 38 ตัวอย่างข้อมูลภาพถ่ายตัวเลขปกติ

ภาพถ่ายเอียง (Rotate)

- หมุนกล้องในการถ่ายภาพในช่วงประมาณ -80° จนถึง 80° ช่วงระยะห่าง 20°
- ขนาดภาพระหว่าง 4×12 พิกเซล จนถึง 15×15 พิกเซล
- จำนวน 1,000 ภาพ



ภาพ 39 ตัวอย่างข้อมูลภาพถ่ายตัวเลขที่เอียง

ภาพถ่ายขยายขนาด (Scale)

- ความหลากหลายของระยะห่างในการถ่ายภาพในช่วง 0.25 เมตร จนถึง 1.5 เมตร ช่วงระยะห่าง 0.25 เมตร
- ขนาดภาพระหว่าง 2×6 พิกเซล จนถึง 36×52 พิกเซล
- จำนวน 1,000 ภาพ



ภาพ 40 ตัวอย่างข้อมูลภาพถ่ายหลายขนาด

ภาพถ่ายหลายมุมมอง (Perspective)

- เอียงมุมมองของภาพ -60° จนถึง 60° ช่วงระยะห่าง 20°
- ขนาดภาพระหว่าง 3×9 พิกเซล จนถึง 10×15 พิกเซล
- จำนวน 1,000 ภาพ



ภาพ 41 ตัวอย่างข้อมูลภาพถ่ายหลายมุมมอง

ภาพถ่ายที่ได้รับผลกระทบของแสง (Illumination)

- ภาพถ่ายจօภาพแอลอีดีในความมืด และจօภาพแอลอีดีที่มีความหลากหลายของการส่องสว่างของแสง
- ขนาดภาพระหว่าง 5×12 พิกเซล จนถึง 10×15 พิกเซล
- จำนวน 500 ภาพ



ภาพ 42 ตัวอย่างข้อมูลภาพถ่ายที่ได้รับผลกระทบจากแสง

3.2.1.2 ขอบเขตของอุปกรณ์

อุปกรณ์ในการเก็บภาพ คือ Logitech HD pro webcam c920

3.2.2 การออกแบบและพัฒนาวิธีการ

ผู้วิจัยมุ่งเน้นไปที่การรู้จำภาพตัวเลขที่อาจได้รับผลกระทบจากสภาพแวดล้อม หรือ พฤติกรรมการถ่ายภาพ เช่น ภาพถ่ายเอียง มีมุมมองภาพถ่ายที่หลากหลาย และมีระยะห่างจากการถ่ายภาพที่ไม่คงที่ เป็นต้น ดังนั้น ผู้วิจัยจึงใช้ข้อมูลที่ไม่ได้รับผลกระทบ และได้รับผลกระทบดังกล่าวในการทดสอบ และผู้วิจัยพัฒนาการเพิ่มปริมาณข้อมูลภาพ ที่สามารถเพิ่มความหลากหลายของข้อมูล และแปลงข้อมูลให้อยู่ในรูปแบบมาตรฐานเดียวกันซึ่งมีความทนทานต่อสภาพแวดล้อมต่าง ๆ

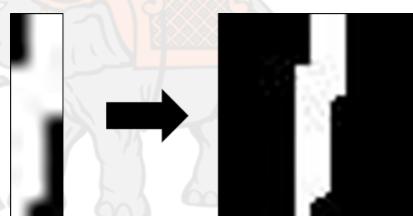
หลังจากนั้นจึงนำการเรียนรู้ของเครื่องจักรมาประยุกต์ใช้ในการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแล้วอีดี ซึ่งกระบวนการรู้จำแบ่งออกเป็น 5 ขั้นตอนดังนี้

ขั้นตอนที่ 3.2.2.1 การขยายภาพด้วยศูนย์ (Image zero padding) เนื่องจาก การหาคุณลักษณะเฉพาะของภาพตัวเลขนั้นมีความสำคัญอย่างมาก การนำข้อมูลผ่านกระบวนการต่าง ๆ อาจทำให้รูปทรง (Shape) ของตัวเลขเสียหายดังภาพ 43



ภาพ 43 การปรับขนาดของภาพโดยไม่รักษารูปทรงของตัวเลข

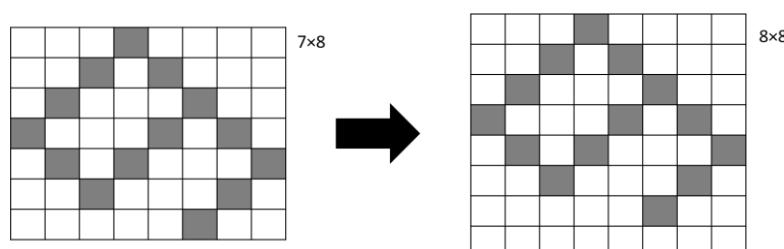
ดังนั้น จึงมีความจำเป็นต้องรักษารูปทรงของตัวเลขไว้เพื่อใช้ในการแยกแยะกลุ่มของตัวเลข ซึ่งผู้วิจัย มุ่งเน้นรักษาให้ข้อมูลอยู่ในส่วนกลางของภาพมากที่สุดให้ผลลัพธ้มีลักษณะดังภาพ 44



ภาพ 44 ผลลัพธ์การขยายภาพด้วยศูนย์

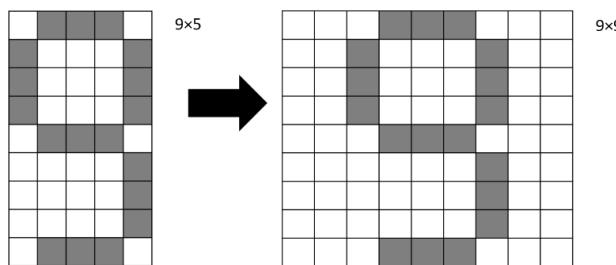
โดยการเพิ่มข้อมูลเวกเตอร์แคลว์หรือหลักที่มีค่าเป็น 0 ทั้งหมดจนข้อมูลเป็นจำนวนเต็มโดยเงื่อนไขดังนี้

เงื่อนไขที่ 3.2.2.1.1 กรณีที่ข้อมูลมีความกว้างมากกว่าความสูงจะทำการเพิ่มข้อมูลเวกเตอร์แคลว์ทั้งบนและล่างจนกระทำต่อไปที่ข้อมูลเป็นจำนวนเต็ม



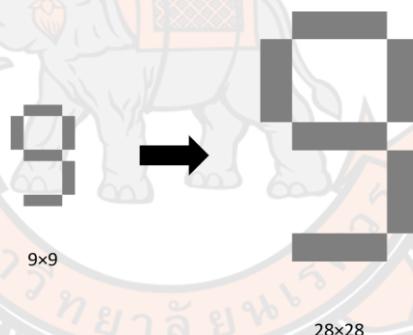
ภาพ 45 การขยายภาพโดยเพิ่มข้อมูลเวกเตอร์แคลว์ที่มีค่าเป็น 0

เงื่อนไขที่ 3.2.2.1.2 กรณีที่ข้อมูลมีความสูงมากกว่าความกว้างจะทำการเพิ่มข้อมูลเวกเตอร์หลักทั้งซ้ายและขวาจนกระทำต่อไปที่ข้อมูลเป็นจำนวนเต็ม



ภาพ 46 การขยายภาพโดยข้อมูลเวกเตอร์หลักที่มีค่าเป็น 0

ขั้นตอนที่ 3.2.2.2 การแปลงขนาดภาพ (Image resizing) ข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนที่นำเข้ามาจากการตรวจสอบหน้ามีขนาดที่หลากหลาย และสถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบconvolutoinal ที่องการข้อมูลภาพนำเข้าที่มีขนาด 28×28 พิกเซล ดังนั้น ผู้วิจัยจึงทำการปรับขนาดข้อมูลภาพตัวเลขเป็น 28×28 พิกเซล ด้วยเทคนิคการย่อ-ขยายภาพจากการแปลงภาพเชิงเรขาคณิต



ภาพ 47 ผลลัพธ์การแปลงขนาดภาพ

ขั้นตอนที่ 3.2.2.3 การแปลงเรเดอนหลายขนาด (Multiscale Radon transform) ในขั้นตอนนี้จะเพิ่มความหลากหลายของข้อมูล และแปลงข้อมูลให้อยู่ในรูปของไซโน-แกรม (sinogram) แล้วทำการแก้ไขให้อยู่ในรูปแบบมาตรฐานบนไซโนแกรม ซึ่งผู้วิจัยได้แบ่งขั้นตอนนี้ออกเป็น 3 ส่วน ได้แก่

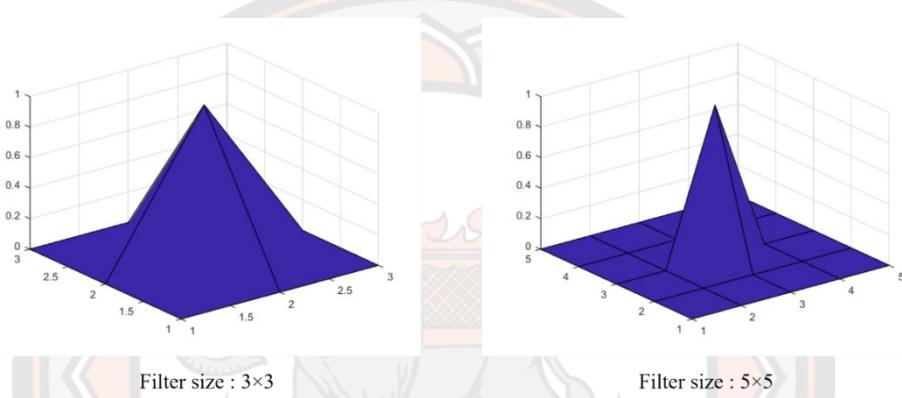
3.2.2.3.1 การทำภาพให้เรียบ (Image smoothing) เพิ่มความหลากหลายของข้อมูล โดยการกรองภาพด้วยตัวกรอง gauss เซียน (Gaussian filter) ซึ่งภาพผลลัพธ์ที่ได้จะมีลักษณะของรูปทรงกลม เคียงกับภาพถ่ายหลายขนาด (Scaling image) ดังนั้น ผู้วิจัยจึงมีแนวคิดว่าขั้นตอนนี้สามารถเพิ่มความทนทานด้านความหลากหลายของขนาดภาพได้ ซึ่งมีกระบวนการดังนี้

ขั้นตอนที่ 1 สร้างตัวกรองเกาส์เชียนโดยสมการที่ 25

$$G(u, v, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \quad (25)$$

- เมื่อ $G(u, v, \sigma)$ คือ ตัวกรองเกาส์เชียน
- (u, v) คือ ตำแหน่งที่พิจารณากรองบนภาพ
- σ คือ ส่วนเบี่ยงเบนมาตรฐาน

โดยที่ผู้วิจัยได้กำหนดค่าของขนาดของตัวกรองเป็น 3×3 และ 5×5 ซึ่งแต่ละตัวกรองจะมีค่า σ เป็น 0.2



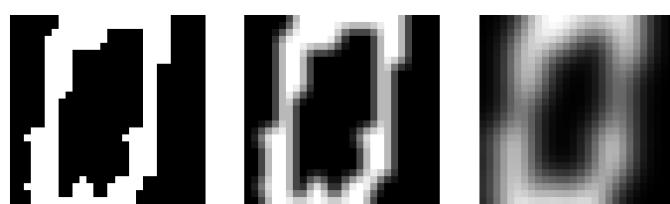
ภาพ 48 ตัวกรองเกาส์เชียนของวิธีการที่นำเสนอสำหรับใช้งาน

ขั้นตอนที่ 2 กรองภาพโดยใช้ตัวกรองเกาส์เชียน G จากสมการที่ 25 ได้ดังสมการที่ 26

$$I' = I * G \quad (26)$$

- เมื่อ I' คือ ภาพผลลัพธ์ที่ถูกกรองภาพให้เรียบ
- I คือ ภาพนำเข้า
- $*$ คือ ตัวดำเนินการ convolution

ทางผู้วิจัยกรองภาพด้วยตัวกรองเกาส์เชียนที่มีขนาดแตกต่างกัน 2 ขนาดเพื่อเพิ่มความทนทานด้านความหลากหลายของระยะห่างการถ่ายภาพ



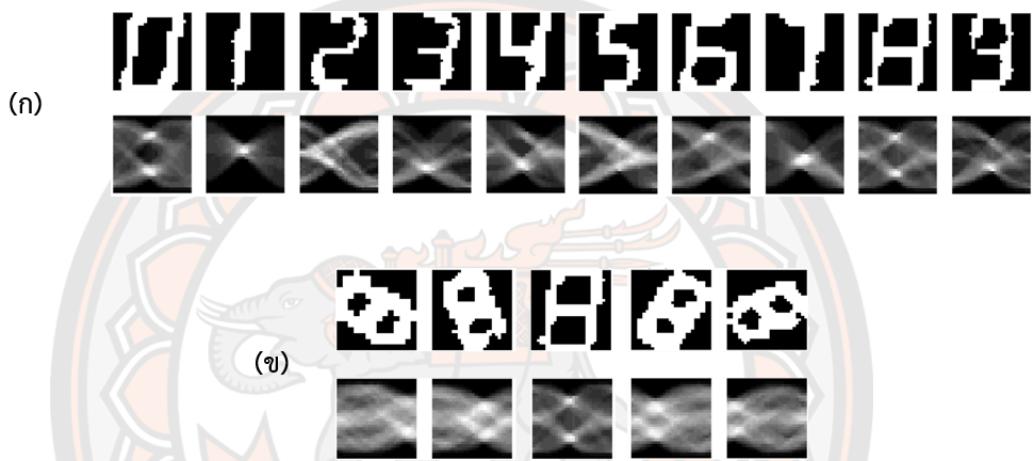
ภาพ 49 ตัวอย่างผลลัพธ์จากการกรองภาพจากตัวขนาดต่างกัน

3.2.2.3.2 การแปลงเรดอน (Radon transform) ขั้นตอนนี้เป็นการแปลงภาพให้อยู่ในรูปคลื่นความถี่ซึ่งที่เรียกว่าไซโนแกรม ซึ่งมีรูปแบบที่สามารถจัดการกับความเอียงของภาพได้สะดวกกว่าการจัดการบนโดเมนเวลา (Spatial domain) ดังที่ผู้วิจัยได้อธิบายไว้ในบทที่ 2 หัวข้อที่ 2.2.2 การแปลงเรดอนคำนวณดังสมการที่ 27 ต่อไปนี้

$$I_r = R(I')$$
 (27)

เมื่อ I_r คือ ภาพไซโนแกรมจากการแปลงเรดอน

$R(\cdot)$ คือ พัฟฟ์ชันการแปลงเรดอน



ภาพ 50 (ก) ผลลัพธ์ไซโนแกรมจากการแปลงเรดอนของกลุ่มตัวเลข และ (ข) ตัวเลขกลุ่มเดียวกันที่มีองศาความเอียงที่แตกต่างกัน

3.2.2.3.3 การจัดเรียงสัมประสิทธิ์ใหม่เรดอน (Radon coefficient rearrangement) จากภาพ 50 สร้างเกตตี้ได้ว่าภาพไซโนแกรมจากตัวเลขกลุ่มเดียวกันที่มีองศาความเอียงที่แตกต่างกันแต่มีรูปทรงที่คล้ายคลึงกันกัน และเมื่อพิจารณาจากจุดที่มีค่าสูงที่สุดบนไซโนแกรม (ส่วนที่สว่างที่สุด) จะพบในตำแหน่งที่แตกต่างกัน ดังนั้น ผู้วิจัยจึงพัฒนาขั้นตอนนี้เพื่อสร้างรูปแบบมาตราฐานบนไซโนแกรมซึ่งได้รับผลกระทบจากการเอียงน้อยที่สุด โดยการลีนตำแหน่งที่มีค่าสูงที่สุดไปยังบริเวณส่วนกลางของภาพโดยการเลื่อนวน (Circular shift) ด้วยกระบวนการดังนี้

3.2.2.3.3.1 หาตำแหน่งของค่าสูงสุดบนไซโนแกรมแต่ค่าที่สูงที่สุดอาจปรากฏมากกว่า 1 ตำแหน่ง และอาจอยู่บนหลักที่แตกต่างกันดังภาพ 51

กรณีที่ 1							กรณีที่ 2						
0	1	2	3	4	5	6	0	1	2	3	4	5	6
2	4	0	1	4	2	3	2	4	0	1	4	2	3
3	5	1	0	4	3	4	3	5	1	0	4	3	4
2	6	0	0	4	4	5	2	6	0	0	5	4	5
1	6	0	1	4	3	5	1	6	0	1	6	3	5
1	5	1	1	4	3	5	1	5	1	1	4	3	5
0	5	0	2	4	2	4	0	5	0	2	4	2	4
1	4	0	1	4	3	5	1	4	0	1	4	3	5

ภาพ 51 การเลือกเวกเตอร์หลักที่มีค่าสูงสุด

ดังนั้น ผู้วิจัยจึงแบ่งขั้นตอนนี้ออกเป็น 2 กรณี

กรณีที่ 1 กรณีที่ค่าสูงสุดอยู่บนหลักเดียว จะนำตำแหน่งของค่าสูงสุดมาใช้ในการคำนวณ

กรณีที่ 2 กรณีที่ค่าสูงสุดอยู่บนหลักที่แตกต่างกัน จะคำนวณหาค่าเฉลี่ยของหลักที่มีค่าสูงสุด และนำมาเปรียบเทียบเพื่อเลือกหลักที่มีค่าเฉลี่ยที่มากที่สุดมาคำนวณต่อไป

ต่อไปนี้

3.2.2.3.3.2 คำนวณค่าระยะห่างจากจุดศูนย์กลางโดยสมการ

$$c_{shift} = c^{center} - c^{max} \quad (28)$$

เมื่อ c_{shift} คือ ระยะห่าง (จำนวนหลัก) จากจุดศูนย์กลาง c^{max}

c^{center} คือ ตำแหน่งของเวกเตอร์หลักที่อยู่กึ่งกลางไซโนแกรม

c^{max} คือ ตำแหน่งของเวกเตอร์หลักที่มีค่าสูงสุดจากหัวข้อที่ 3.2.2.3.1

จากนั้นทำการเลื่อนวนเวกเตอร์หลักที่มีค่าสูงสุดไปที่ตำแหน่งกึ่งกลางของไซโนแกรม โดยสามารถคำนวณตำแหน่งของหลักใหม่ c'_i หลังจากการเลื่อนวนได้ดังนี้

$$c'_i = (c_i + c_{shift}) \% n_c \quad (29)$$

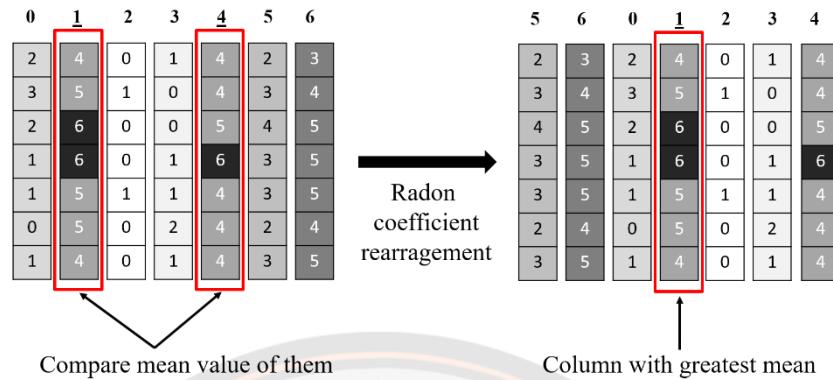
เมื่อ c_i คือ ตำแหน่งของเวกเตอร์หลัก i ก่อนการเลื่อนวน

i คือ จำนวนนับที่มีค่าไม่สูงกว่า n_c

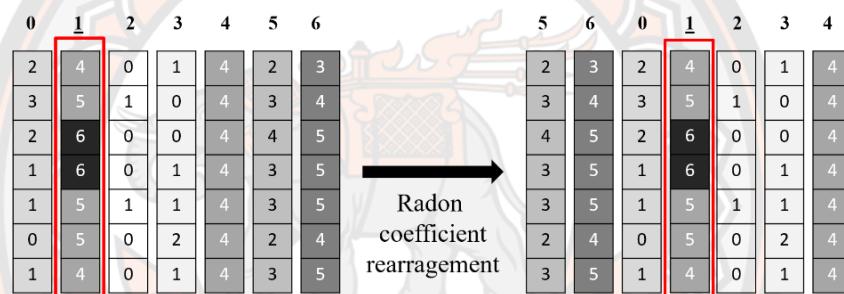
n_c คือ จำนวนของหลักทั้งหมดของไซโนแกรม

% การดำเนินการหารเอาเศษ (modulo)

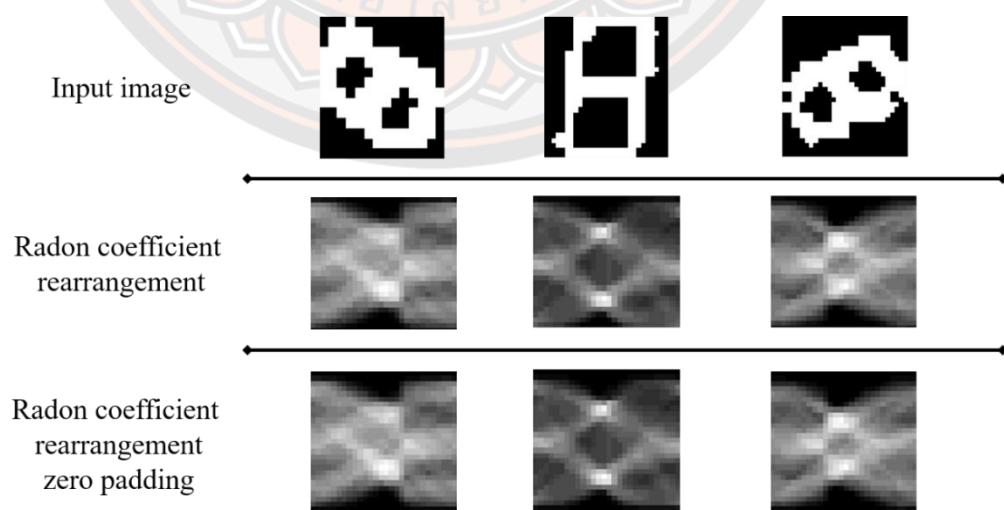
ผลลัพธ์จากขั้นตอนนี้แสดงดังภาพ 52 และ 53



ภาพ 52 การจัดเรียงสัมประสิทธิ์ใหม่กรณีที่ 1

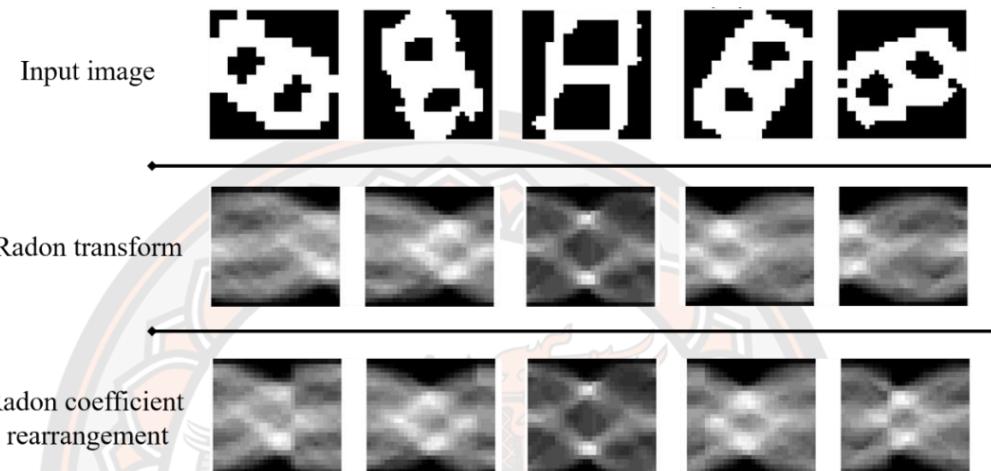


ภาพ 53 การจัดเรียงสัมประสิทธิ์ใหม่กรณีที่ 2



ภาพ 54 การจัดเรียงสัมประสิทธิ์ใหม่ของเลข 8 ที่มีองค์ความเอียงต่างกัน

ขั้นตอนที่ 3.2.2.4 การขยายภาพด้วยศูนย์ (Image zero padding) ผลลัพธ์ไซโอนแกรมที่ผ่านการจัดเรียงสัมประสิทธิ์ใหม่จะมีขนาดของภาพแปรผันตามองศาที่กำหนด และจำเป็นต้องเปลี่ยนแปลงขนาดเพื่อนำเข้าไปสู่กระบวนการเรียนรู้ ซึ่งอาจเกิดความเสียหายกับรูปทรงของไซโอนแกรม ดังนั้น ขั้นตอนนี้จะเป็นการรักษารูปทรงของไซโอนแกรม เพื่อให้การเปลี่ยนแปลงขนาดส่งผลกระทบกับข้อมูลน้อยที่สุดในทำนองเดียวกันกับขั้นตอนที่ 3.2.2.1



ภาพ 55 ผลลัพธ์การจัดเรียงสัมประสิทธิ์ใหม่เพื่อรองรับการขยายภาพด้วยศูนย์

ขั้นตอนที่ 3.2.2.5 สถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวลูชัน (LeNet-5 CNN) ทางผู้วิจัยได้ศึกษาวิธีการรู้จำที่สามารถนำมาประยุกต์ใช้กับข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแล็ปท็อป ที่ได้รับผลกระทบต่าง ๆ จากพฤติกรรมการถ่ายภาพ แล้วได้พิจารณาว่าโครงข่ายประสาทเทียมมีความสามารถในการแยกแยะ และจัดการกับผลกระทบเหล่านี้ได้ในระดับหนึ่ง ดังนั้น ผู้วิจัยจึงนำสถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวลูชันซึ่งต้องการข้อมูลภาพนำเข้าที่มีขนาด 28×28 พิกเซล และไม่ใช้ทรัพยากรในการคำนวณมากมาใช้งานกับข้อมูลภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน โดยผู้วิจัยได้กำหนดค่าของตัวแปรดังต่อไปนี้

ตัวแปรที่ 1 คือ Learning rate มีค่าเท่ากับ 0.02

ตัวแปรที่ 2 คือ Batch size มีค่าเท่ากับ 50

ตัวแปรที่ 3 คือ Epochs มีค่าเท่ากับ 10

3.2.3 วิธีการวัดประสิทธิภาพ

การวัดประสิทธิภาพการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนผู้วิจัยประยุกต์ใช้อัตราความถูกต้อง (Accuracy rate) ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการวัดประสิทธิภาพ

ความสามารถในการรู้จำให้กับแบบจำลองในการรู้จำ ซึ่งทางผู้วิจัยได้นำข้อมูลตัวเลขปกติ วิธีการเพิ่มปริมาณข้อมูลภาพโดยการแปลงภาพเชิงเรขาคณิตบนระบบสองมิติ 3 วิธีการ และวิธีการเพิ่มความหลากหลายของข้อมูลที่ผู้วิจัยได้นำเสนอ เพื่อนำไปสร้างแบบจำลองข้อมูลการสอนโดยสถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวลูชัน จึงมีชนิดของข้อมูลที่นำไปสร้างแบบจำลองในการรู้จำ ดังนี้

- ข้อมูลชนิดที่ 1 ข้อมูลตัวเลขปกติ (N)
- ข้อมูลชนิดที่ 2 ข้อมูลตัวเลขจากการหมุนภาพ (R)
- ข้อมูลชนิดที่ 3 ข้อมูลตัวเลขจากการย่อ-ขยายภาพ (S)
- ข้อมูลชนิดที่ 4 ข้อมูลตัวเลขจากการเอียงเชิงมุมมอง (P)
- ข้อมูลชนิดที่ 5 ข้อมูลตัวเลขจากการแปลงเรเดอนหลายขนาด (MRT)

ผู้วิจัยได้นำวิธีการที่ 1 – 4 มาสร้างแบบจำลองในการรู้จำรูปแบบต่าง ๆ เพื่อเปรียบเทียบกับแบบจำลองที่ผู้วิจัยนำเสนอ ซึ่งสร้างจากข้อมูลชนิดที่ 1 และ 5 ซึ่งแต่ละแบบจำลองจะนำทดสอบกับข้อมูลทดสอบ 5 ประเภทได้แก่ 1.ภาพถ่ายปกติ 2.ภาพถ่ายเอียง 3.ภาพถ่ายหลายขนาด 4.ภาพถ่ายหลายมุมมอง 5.ภาพถ่ายที่ได้รับผลกระทบของแสง โดยที่ผู้วิจัยได้วัดประสิทธิภาพการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนใน 2 รูปแบบ ซึ่งมีลักษณะดังนี้

รูปแบบที่ 1 : ทำการเพิ่มปริมาณข้อมูลภาพโดยวิธีการแปลงเชิงเรขาคณิตผ่านโปรแกรม MATLAB และสร้างแบบจำลองการรู้จำด้วยโปรแกรมภาษา PYTHON ซึ่งการสร้างแบบจำลองในการรู้จำจะมีข้อมูลตัวเลขปกติ 1,600 ภาพร่วมกับจำนวนข้อมูลชนิดที่ 2 – 4 ชนิดละ 4,000 ภาพ และชนิดที่ 5 จำนวน 3,200 ภาพ ซึ่งผู้วิจัยได้วัดประสิทธิภาพกับแบบจำลองที่มีลักษณะดังต่อไปนี้

ตาราง 3 ลักษณะข้อมูลที่ใช้ในการสร้างแบบจำลองการรู้จำรูปแบบที่ 1

การสร้างแบบจำลอง	จำนวนข้อมูลที่ใช้	การสร้างแบบจำลอง	จำนวนข้อมูลที่ใช้
N	1,600	N + R	5,600
N + S	5,600	N + P	5,600
R + S	8,000	R + P	8,000
S + P	8,000	N + R + S	9,600
N + R + P	9,600	N + S + P	9,600
N + R + S + P	13,600	N + MRT	4,800

รูปแบบที่ 2 : ทุกรอบวนการของรูปแบบที่ 2 นี้จะดำเนินการโดยโปรแกรมภาษา PYTHON ทั้งหมด ทั้งการเพิ่มปริมาณข้อมูลภาพและการสร้างแบบจำลองในการรู้จำ เนื่องจากการวัดประสิทธิภาพด้านเวลาในการประมวลผล และมีจำนวนข้อมูลที่ใช้ในการสร้างแบบจำลองในการรู้จำ เท่ากันในทุกแบบจำลองคือ 4,800 ภาพ โดยผู้วิจัยจะเลือกให้ข้อมูลแต่ละชนิดมีจำนวนเท่ากันยกเว้นแบบจำลองที่สร้างโดยข้อมูล 1 ชนิด และ 2 ชนิดที่มีข้อมูลภาพปกติในการสร้างแบบจำลองในการรู้จำ จึงมีลักษณะแบบจำลองในการรู้จำดังตาราง 4

ตาราง 4 ลักษณะข้อมูลที่ใช้ในการสร้างแบบจำลองการรู้จำรูปแบบที่ 2

การสร้างแบบจำลอง	จำนวนข้อมูลที่ใช้	การสร้างแบบจำลอง	จำนวนข้อมูลที่ใช้
N	1,600	N + R	1,600 + 3,200
N + S	1,600 + 3,200	N + P	1,600 + 3,200
R + S	2,400 × 2	R + P	2,400 × 2
S + P	2,400 × 2	N + R + S	1,600 × 3
N + R + P	1,600 × 3	N + S + P	1,600 × 3
N + R + S + P	1,200 × 4	N + MRT	1,600 + 3,200

ตาราง 5 การเปรียบเทียบรูปแบบการวัดประสิทธิภาพการรู้จำ

ลักษณะวิธีการวัด ประสิทธิภาพ	โปรแกรมที่ใช้ในการพัฒนา	จำนวนข้อมูลการสอนจาก การแปลงข้อมูลเชิงเรขาคณิต
รูปแบบที่ 1	MATLAB และ Python	ไม่เท่ากันในแต่ละแบบจำลอง
รูปแบบที่ 2	Python	4,800 ภาพ

วิธีการคำนวณค่าอัตราความถูกต้อง

$$A_c = \frac{N_c}{N_c + N_i} \times 100 \quad (30)$$

- เมื่อ A_c คือ ค่าอัตราความถูกต้อง (%)
 N_c คือ จำนวนภาพที่ทำนายได้ถูกต้อง
 N_i คือ จำนวนภาพที่ทำนายผิด

วิธีการคำนวณค่าเฉลี่ย

$$M = \frac{\sum_i^n A_{c_i}}{n} \quad (31)$$

- เมื่อ M คือ ค่าเฉลี่ยของอัตราความถูกต้อง^{*}
 A_{c_i} คือ ค่าอัตราความถูกต้องที่คำนวณได้จากข้อมูลทดสอบที่ i
 n คือ จำนวนครั้งที่ทำการทดสอบ

วิธีการคำนวณค่าส่วนเบี่ยงเบนมาตรฐาน

$$\sigma = \sqrt{\frac{\sum_i^n (A_{c_i} - M)^2}{n - 1}} \quad (32)$$

- เมื่อ σ คือ ส่วนเบี่ยงเบนมาตรฐานของอัตราความถูกต้อง^{*}
 A_{c_i} คือ ค่าอัตราความถูกต้องที่คำนวณได้จากข้อมูลทดสอบที่ i

บทที่ 4

ผลการทดลอง

ผลการทดลองของงานวิจัยเรื่องการตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม ผู้วิจัยได้ทำการวัดประสิทธิภาพและเวลาการประมวลผลของวิธีการทั้ง 2 การทดลองประกอบด้วย

4.1 การทดลองการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

4.2 การทดลองการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

ในแต่ละการทดลองผู้วิจัยพัฒนาด้วยโปรแกรมภาษา MATLAB และ PYTHON

การทดลองที่ 4.1 การทดลองการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

ในส่วนของการวัดประสิทธิภาพการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนนี้ได้ทำการทดสอบกับข้อมูลภาพหน้าจอแสดงผลในรูปแบบเจ็ดส่วนทั้งหมด 240 ภาพ และแต่ละภาพประกอบด้วยตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน 8 ตัวเลข โดยผู้วิจัยได้ทำการวัดประสิทธิภาพ 2 รูปแบบ ดังต่อไปนี้

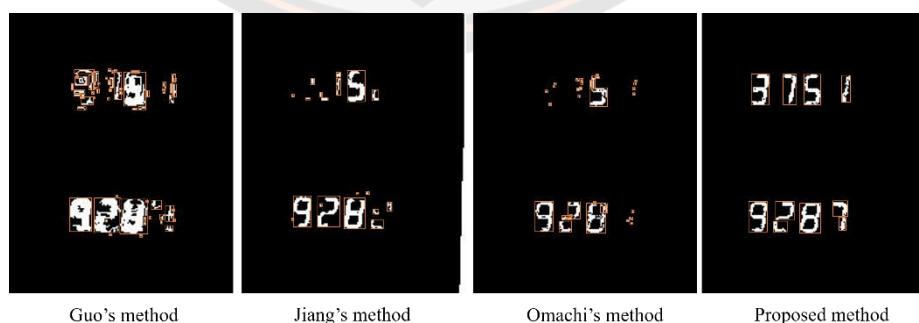
4.1.1 ผลการวัดประสิทธิภาพโดยไม่มีการปรับค่าของตัวแปร

การวัดประสิทธิภาพครั้งนี้จะนำค่าของตัวแปรค่าสี ความอิมสี หรือความสว่างจากงานวิจัยที่ทางผู้วิจัยนำมาเปรียบเทียบมาใช้งานโดยไม่มีการปรับเปลี่ยนค่าตัวแปรเหล่านี้ จึงมีผลลัพธ์ การวัดประสิทธิภาพแสดงดังตาราง 3

ตาราง 6 ผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจจับโดยไม่มีการปรับค่าของตัวแปร

วิธีการ	ความแม่นยำ	การเรียกคืน
วิธีการของ Jiang	70.94%	69.48%
วิธีการของ Omachi	91.01%	28.91%
วิธีการของ Mu	92.20%	55.31%
วิธีการของที่นำเสนอ	94.46%	92.24%

จากการเปรียบเทียบประสิทธิภาพของวิธีการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแล้วดีสีแดงแสดงให้เห็นว่าวิธีการกำหนดช่วงของแบบจำลองสี HSV ให้ผลลัพธ์ที่น่าพอใจ โดยมีความแม่นยำที่ 94.46% ซึ่งสูงกว่าคู่เปรียบที่ใช้บประมาณ 2% – 20% และมีค่าการเรียกคืนอยู่ที่ 92.24% ซึ่งสูงกว่าคู่เปรียบที่ใช้บประมาณ 20% - 60% และแสดงให้เห็นว่าวิธีการที่นำมาเปรียบเทียบนั้นยังไม่สามารถนำตัวเลขทั้งหมดออกจากภาพได้ในการวัดประสิทธิภาพครั้งนี้เนื่องจากวิธีการของ Jiang เป็นการกำหนดช่วงของสีบนแบบจำลองสี RGB เพื่อการตรวจจับบริเวณที่เป็นแอปเปิลสีแดง ซึ่งไม่จำเป็นต้องคำนึงถึงค่าความสว่างของบริเวณนั้น แต่ข้อมูลที่ใช้ในการทดสอบเป็นตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแล้วดีที่มีส่วนประกอบที่สำคัญคือแสง และวิธีการของ Omachi จะเป็นการกำหนดช่วงของสีบนแบบจำลองสี RGB เพื่อทำการตรวจจับบริเวณที่เป็นสัญญาณไฟจราจร ถึงแม้ว่าจะมีลักษณะเป็นแอปเปิลดีเข่นเดียวกับหน้าจอแสดงผลมีความสว่างที่มากกว่าสัญญาณไฟจราจร รวมไปถึงการเลือกบริเวณที่มีค่าความสว่างจากแบบจำลองสี RGB นั้นทำได้ยากกว่าแบบจำลองสี HSV หาก และวิธีการของ Mu เป็นการกำหนดช่วงของสี ความอิมสี และความสว่างบนแบบจำลองสี HSV พบร่วมกับการกำหนดค่าความอิมสีที่สูงจนเกินไปในการตรวจจับซึ่งทางผู้วิจัยได้อธิบายไปแล้วในบทที่ 3 ว่าความอิมนั้นนำมาใช้งานได้ยากและอาจได้ส่วนประกอบอื่นปะปนอยู่กับตัวเลขอีกด้วยเนื่องจากบริเวณที่เป็นตัวเลขมีความแปรปรวนของค่าความอิมสี ทั้งนี้วิธีการที่ผู้วิจัยนำเสนอสามารถกำหนดช่วงของสี และความสว่างโดยตรงจากแบบจำลองสี HSV โดยไม่คำนึงถึงค่าความอิมสี ซึ่งเมื่อพิจารณาจากภาพ 46 และแสดงให้เห็นว่าวิธีการที่ผู้วิจัยนำเสนอได้รับส่วนที่เป็นตัวเลขที่สมบูรณ์ และมีความผิดพลาดเพียงเล็กน้อยเท่านั้น จึงส่งผลให้กระบวนการที่ผู้วิจัยนำเสนอ มีประสิทธิภาพสูงที่สุด



ภาพ 56 ผลลัพธ์การตรวจจับของวิธีการต่าง ๆ

4.1.2 ผลการวัดประสิทธิภาพโดยการปรับค่าของตัวแปร

การวัดประสิทธิภาพครั้งนี้จะนำค่าของตัวแปรค่าสี ความอิมสี หรือความสว่างจากงานวิจัยที่ผู้วิจัยนำมาเปรียบเทียบมาทำการปรับเปลี่ยนค่าตัวแปรให้มีความเหมาะสมกับข้อมูลที่ผู้วิจัยได้นำมาทดสอบ โดยที่งานวิจัยที่มีการปรับเปลี่ยนค่าของตัวแปรประกอบด้วย

งานวิจัยที่ 1 วิธีการตรวจจับและรู้จำไฟจราจรโดยใช้แบบจำลองสี RGB (Omachi & Omachi, 2009)

งานวิจัยที่ 2 วิธีการตรวจจับไฟจราจรโดยใช้แบบจำลองสี HSV (Mu et al., 2015)

จึงมีผลลัพธ์การวัดประสิทธิภาพแสดงดังตาราง 4

ตาราง 7 ผลการเปรียบเทียบประสิทธิภาพวิธีการตรวจจับโดยการปรับค่าของตัวแปร

วิธีการ	ความแม่นยำ	การเรียกคืน
วิธีการของ Jiang	70.94%	69.48%
วิธีการของ Omachi	96.72%	94.00%
วิธีการของ Mu	95.92%	94.53%
วิธีการของที่นำเสนอ	94.46%	92.24%

ในการวัดประสิทธิภาพครั้งนี้จะไม่ทำการปรับเปลี่ยนค่าตัวแปรของวิธีการของ Jiang เนื่องจาก Jiang นำเสนอเพื่อการแปลงภาพสีให้เป็นภาพระดับสีเทา จึงมีความแตกต่างจากวิธีการจากงานวิจัยอื่น (สามารถตรวจสอบรายละเอียดวิธีการของ Jiang ได้ในหัวข้อที่ 2.3)

จากการเปรียบเทียบประสิทธิภาพของวิธีการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแล้วอีสีแดงแสดงให้เห็นว่า หลังจากที่ผู้วิจัยได้ทำการปรับเปลี่ยนค่าตัวแปรต่าง ๆ ของแบบจำลองสี RGB และ HSV ให้มีความเหมาะสมกับข้อมูลที่นำมาทดสอบนั้นส่งผลให้วิธีการของ Omachi และ วิธีการของ Mu มีประสิทธิภาพที่สูงขึ้นมากกว่าวิธีการที่ผู้วิจัยนำเสนอ โดยที่วิธีการของ Omachi มีความแม่นยำเท่ากับ 96.72% และค่าการเรียกคืนเท่ากับ 94% และวิธีการของ Mu มีค่าความแม่นยำและค่าการเรียกคืนเท่ากับ 95.92% และ 94.53% ตามลำดับ ซึ่งสูงกว่าวิธีการที่ผู้วิจัยนำเสนอประมาณ 1% – 3% อย่างไรก็ตามการปรับช่วงของตัวแปรนี้จะเหมาะสมกับข้อมูลที่นำมาทดสอบเพียงทดสอบเท่านั้น หากนำวิธีการนี้ไปทดลองใช้กับข้อมูลที่แตกต่างจากข้อมูลที่นำมาทดสอบเพียง

เล็กน้อยอาจทำให้ผลลัพธ์วิธีการจากงานวิจัยที่ทำการปรับเปลี่ยนค่าตัวแปรนี้มีค่าความแม่นยำและค่าการเรียกคืนที่น้อยลงได้

ผู้วิจัยได้ทำการวัดประสิทธิภาพในด้านเวลาในการประมวลผลของการตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแออลอีดีจำนวน 240 ภาพ ซึ่งมีผลลัพธ์ดังตาราง 8

ตาราง 8 ผลการเปรียบเทียบเวลาในการประมวลผลการตรวจจับ

วิธีการ	ระยะเวลารวม (วินาที)	ระยะเวลาต่อภาพ (วินาที)
วิธีการของที่นำเสนอ	30.0993	0.1254
วิธีการของ Mu	40.1470	0.1673
วิธีการของ Omachi	34.0906	0.1420
วิธีการของ Jiang	54.9032	0.2288

จากตาราง 8 แสดงให้เห็นว่าวิธีการที่ผู้วิจัยนำเสนอันมีเวลาในการประมวลผลที่น้อยที่สุด เนื่องจากวิธีการของ Mu มีการแปลงช่วงของสีไป 60 องศา และการเลือกช่วงของความอิ่มสีทำให้เวลาในการประมวลผลมากกว่าวิธีการที่ผู้วิจัยนำเสนอ วิธีการของ Omachi มีการเลือกช่วงของส่วนประกอบทั้งหมด 3 ส่วน นั่นคือ สีแดง สีเขียว และสีน้ำเงิน จึงทำให้ใช้เวลาในการประมวลผลมากกว่าวิธีการที่ผู้วิจัยนำเสนอซึ่งใช้ส่วนประกอบเพียงสี และความสว่างเท่านั้น และวิธีการที่ Jiang นำเสนอเป็นกระบวนการการปรับภาพสีให้เป็นภาพระดับสีเทาซึ่งต้องผ่านกระบวนการแปลงภาพให้เป็นขาวดำโดยวิธีการของ Otsu จึงทำให้ใช้ระยะเวลาในการประมวลผลมากกว่าวิธีการอื่นที่กล่าวมา

การทดลองที่ 4.2 การทดลองการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

ในส่วนของการวัดประสิทธิภาพการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนนี้ผู้วิจัยได้ทำการวัดประสิทธิภาพใน 2 รูปแบบดังที่อธิบายในหัวข้อที่ 3.2.3 โดยใช้สถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวลูชันในการสร้างแบบจำลองการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน และมีข้อมูลที่ใช้ในการสร้างแบบจำลอง ดังนี้

ข้อมูลชนิดที่ 1 ข้อมูลตัวเลขปกติ (N)

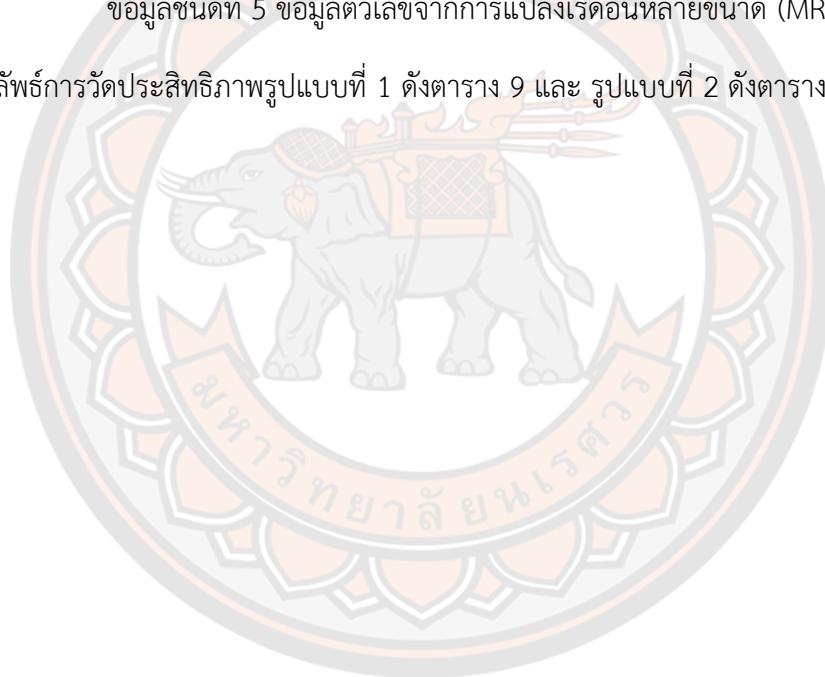
ข้อมูลชนิดที่ 2 ข้อมูลตัวเลขจากการหมุนภาพ (R)

ข้อมูลชนิดที่ 3 ข้อมูลตัวเลขจากการย่อ-ขยายภาพ (S)

ข้อมูลชนิดที่ 4 ข้อมูลตัวเลขจากการเอียงเชิงมุมมอง (P)

ข้อมูลชนิดที่ 5 ข้อมูลตัวเลขจากการแปลงเรเดอนหลายขนาด (MRT)

ซึ่งมีผลลัพธ์การวัดประสิทธิภาพรูปแบบที่ 1 ดังตาราง 9 และ รูปแบบที่ 2 ดังตาราง 10



4.2.1 ผลการวัดประสิทธิภาพรูปแบบที่ 1

ตาราง 9 ผลการเปรียบเทียบประสิทธิภาพของวิธีการรักษาแบบที่ 1

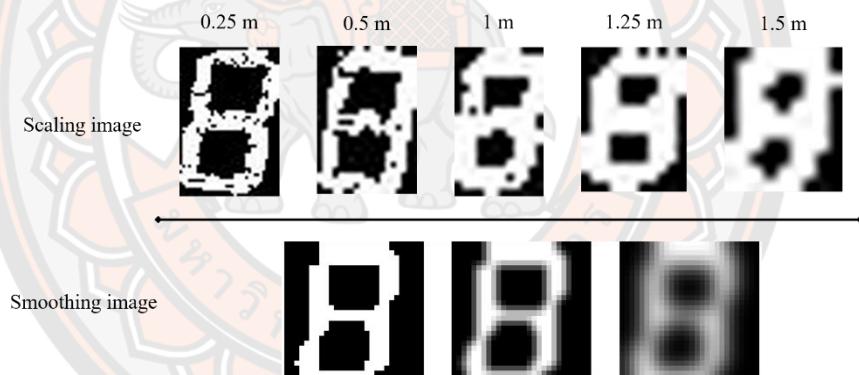
ชื่อ muestra	ค่าอัตราความถูกต้อง (%)	ค่าเฉลี่ย				ส่วนเบี่ยงเบนมาตรฐาน
		ภาพถ่ายปกติ	ภาพถ่ายอี้ยง	หลาຍญามมอง	ภาพถ่ายที่ได้รับผลกระทบของแสง	
N	99.50	29.00	89.00	61.50	87.76	73.35
N + R	99.75	82.10	88.20	66.30	88.78	85.03
N + S	99.75	27.20	92.90	63.30	91.63	74.96
N + P	99.75	32.50	92.90	86.30	94.49	81.19
R + S	99.75	84.30	89.00	66.80	95.71	87.11
R + P	99.75	85.70	92.90	89.59	89.90	91.57
S + P	99.70	31.20	95.10	83.70	97.35	81.41
N + R + S	99.50	86.60	92.20	71.80	94.49	88.92
N + R + P	99.75	80.70	91.60	87.10	95.60	90.95
						7.41

ตาราง 9 (ต่อ)

ชื่อนวัตกรรมสืบ	ค่าอัตราความถูกต้อง (%)		ค่าเฉลี่ย		ส่วนเบี่ยงเบน	
	ภาพถ่ายปกติ	ภาพถ่ายเอียง	ภาพถ่าย	ภาพถ่าย	ผลกรองทบทวน	มาตรฐาน
N + S + P	99.75	31.70	94.80	83.70	96.53	81.30
R + S + P	99.75	87.30	93.50	85.90	93.88	92.07
N + R + S + P	99.75	88.20	93.20	84.80	94.69	92.13
N + MRT	99.75	74.3	90.8	72.7	98	87.11
						12.88

จากตาราง 9 เมื่อเปรียบเทียบแบบจำลองที่ผู้วิจัยได้นำเสนอ กับแบบจำลองที่สร้างโดยข้อมูล 2 ชนิด ซึ่งประกอบด้วยข้อมูลตัวเลขปกติ และข้อมูลตัวเลขจากการแปลงเชิงเรขาคณิต แสดงให้เห็น ว่าการเพิ่มข้อมูลภาพเพื่อแก้ไขปัญหาพฤติกรรมการถ่ายภาพโดยตรงให้ผลลัพธ์อัตราความถูกต้องที่สูง ที่สุด ยกเว้นข้อมูลทดสอบภาพถ่ายที่ได้รับผลกระทบจากแสงที่วิธีการที่ผู้วิจัยนำเสนอ มีอัตราความถูกต้องสูงที่สุดที่ 98% และแม้ว่าในกรณีนี้ วิธีการที่ผู้วิจัยนำเสนอประสิทธิภาพอาจไม่สูงที่สุด แต่ มีอัตราความถูกต้องสูงเป็นอันดับสองรองจากการเพิ่มข้อมูลเพื่อแก้ไขปัญหาพฤติกรรมการถ่ายภาพ โดยตรง และมีค่าเฉลี่ยสูงที่สุดอีกด้วย โดยที่ผลกระทบจากการถ่ายภาพต่างๆ ได้ลดลงจาก กระบวนการดึงต่อไปนี้

4.2.1.1 การถ่ายภาพในระยะห่างที่ใกล้ขึ้น ส่งผลให้ภาพตัวเลขมีเส้นขอบที่หนาขึ้น
เนื่องจากความสว่างของแสงจากแหล่งอีดี รวมไปถึงการมีรายละเอียดบริเวณตัวเลขที่น้อยลง ซึ่ง สอดคล้องกับกระบวนการกรองภาพให้เรียบในการประมวลผลภาพ ดังนั้น กระบวนการกรองภาพ ด้วยตัวกรองหลายขนาดจึงส่งผลให้ผลลัพธ์การรักษาภาพถ่ายในระยะห่างที่แตกต่างกัน มีผลลัพธ์ที่ดีขึ้น ดังภาพ 57

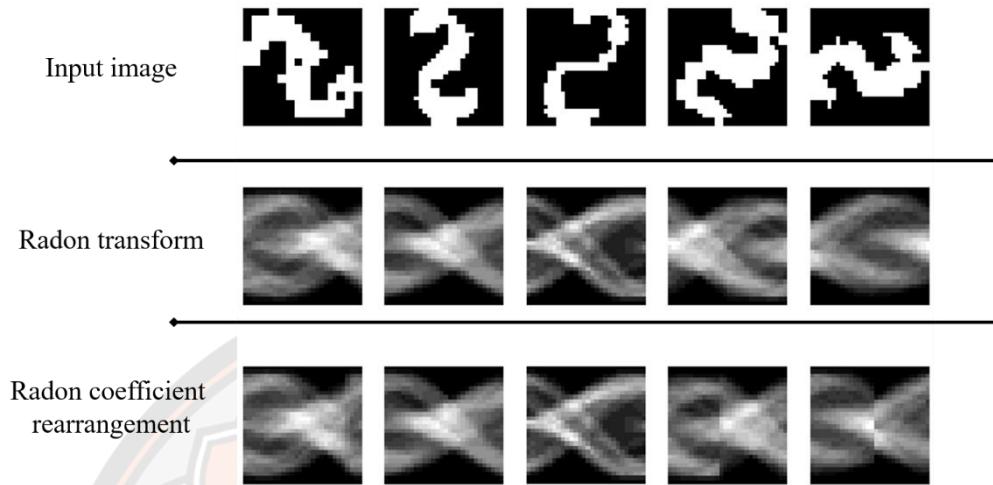


ภาพ 57 การเปรียบเทียบภาพหลายขนาดและภาพที่ผ่านกระบวนการกรองภาพให้เรียบ

อย่างไรก็ตามกระบวนการนี้ดำเนินการกับข้อมูลภาพทั่วไป เช่น ข้อมูลภาพ JPEG ซึ่งได้รับการบีบอัด ข้อมูลจากข้อมูลภาพดิบ จึงมีความละเอียดน้อยกว่าข้อมูลภาพดิบ และอาจเป็นเหตุให้รูปทรงเสียหาย ซึ่งกระบวนการกรองภาพให้เรียบสามารถลดผลกระทบดังกล่าวเพื่อรักษาภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนได้อีกด้วย

4.2.1.2 ภาพถ่ายเอียงส่งผลให้รูปทรงของตัวเลขเปลี่ยนแปลงไป จึงทำให้การสกัดคุณลักษณะจากความหนาแน่นเชิงโครงสร้างในแนวตั้ง หรือแนวนอนไม่เหมาะสมในการนำมาใช้งาน แต่กระบวนการการแปลงเรเดอนเป็นการเก็บคุณลักษณะความหนาแน่นของพิกเซลเชิงโครงสร้างของตัวเลขในทิศทางที่ครอบคลุมลักษณะของการเอียงของภาพ และการจัดเรียงค่าสัมประสิทธิ์ใหม่เรเดอน

ทำให้คุณลักษณะที่สกัดมาในลักษณะไซโนแกรมมีลักษณะที่คล้ายคลึงกันกับไซโนแกรมของภาพตัวเลขปกติมากขึ้นจากการเลื่อนวน จึงทำให้ประสิทธิภาพของการรู้จำภาพถ่ายเอียงสูงขึ้น ดังภาพ 58



ภาพ 58 การเปรียบเทียบภาพเอียงและภาพไซโนแกรมที่จัดเรียงสัมประสิทธิ์ใหม่

4.2.1.3 ภาพถ่ายหลายมุมมองเกิดจากการรวมกันของการย่อ-ขยายภาพและการหมุนภาพในเชิงทฤษฎี ดังนั้น การแก้ปัญหา ก่อนหน้า จึงทำให้รู้จำภาพถ่ายหลายมุมมองอย่างมีประสิทธิภาพที่สูงขึ้นได้ แต่วิธีการที่ผู้วิจัยนำเสนอ นั้นยังไม่มีการลดปัญหาจากผลกระทบนี้โดยตรง จึงทำให้ผลลัพธ์การทดสอบกับข้อมูลภาพตัวเลขหลายมุมมอง มีค่าอัตราความถูกต้องเพียง 72.40%

4.2.1.4 ตัวเลขจากภาพถ่ายที่ได้รับผลกระทบจากแสง มีความหนาของเส้นขอบที่มาก แต่รูปทรงของตัวเลขยังคงไม่เปลี่ยนแปลงสำหรับข้อมูลที่ผู้วิจัยเก็บรวบรวมมาใช้ในการทดสอบ ซึ่งทางผู้วิจัยคาดว่ากระบวนการกรองภาพให้เรียบด้วยตัวกรองหลายขนาด มีส่วนช่วยในการรู้จำภาพถ่ายที่ได้รับผลกระทบจากแสง เนื่องจากการกรองภาพให้เรียบ ส่งผลต่อเส้นขอบของรูปทรงของตัวเลขให้มีลักษณะที่หนาขึ้น ดังภาพ 59



ภาพ 59 การเปรียบเทียบภาพที่ได้รับผลกระทบจากแสงและภาพที่ผ่านกระบวนการกรอง

ภาพให้เรียบ

อย่างไรก็ตาม เมื่อพิจารณาแบบจำลองที่สร้างด้วยข้อมูลที่มีความหลากหลายมากขึ้น พบว่ามีประสิทธิภาพที่สูงกว่าวิธีการที่ผู้วิจัยนำเสนออย่างชัดเจน เช่น แบบจำลองที่สร้างจากข้อมูลภาพเอียงและภาพหลายมุมมอง ($R + P$) มีค่าอัตราความถูกต้องสูงถึง 91.57% เนื่องจากกระบวนการแปลงขนาดภาพให้สอดคล้องกับโครงข่ายประสาทเทียมแบบคอนโวลูชันมีส่วนช่วยในการลดผลกระทบจากระยะห่างในการถ่ายภาพที่แตกต่างกัน และจากการใช้ข้อมูลภาพเอียงเพียงชนิดเดียวในการสร้างแบบจำลองให้ผลลัพธ์การรู้จำภาพถ่ายหลายมุมมองที่อัตราความถูกต้องเพียง 66.30% ดังนั้น การเพิ่มปริมาณข้อมูลภาพตัวเลขจากการเอียงเชิงมุมมองจึงส่งผลให้แบบจำลองนี้มีประสิทธิภาพที่ดีขึ้นอย่างมาก แต่ใช้จำนวนข้อมูลในการสร้างแบบจำลองที่มากขึ้นด้วยเช่นเดียวกัน และแบบจำลองที่มีอัตราความถูกต้องสูงที่สุด คือ แบบจำลองที่สร้างโดยข้อมูลทุกชนิด ได้แก่ ข้อมูลตัวเลขปกติ ข้อมูลตัวเลขจากการหมุนภาพ ข้อมูลตัวเลขจากการย่อ-ขยายภาพ ข้อมูลตัวเลขจากการหมุนภาพเชิงมุมมอง ($N + R + S + P$) ซึ่งมีอัตราความถูกต้องอยู่ที่ 92.13%

4.2.2 ผลการวัดประสิทธิภาพรูปแบบที่ 2
ตาราง 10 ผลการเปรียบเทียบประสิทธิภาพของวิธีการรักษาแบบที่ 2

ชื่อสูตรทดสอบ	ค่าอัตราความถูกต้อง (%)			ส่วนเบี่ยงเบนมาตรฐาน		
	ภาพถ่ายปกติ	ภาพถ่ายอึด	หลาຍชູນມອງ	ภาพถ่ายที่ได้รับ	ผลการทดสอบ	ค่าเฉลี่ย
N	99.50	29.00	89.00	61.50	87.76	73.35
N + R	99.75	69.70	87.50	84.60	87.37	85.78
N + S	99.75	26.10	84.70	72.70	92.80	75.21
N + P	99.75	24.50	88.00	81.40	96.80	78.09
R + S	99.75	69.40	91.20	87.80	96.20	88.87
R + P	99.00	68.30	<u>92.30</u>	87.10	91.20	87.58
S + P	99.75	30.10	85.00	80.50	94.21	77.91
N + R + S	99.75	68.90	90.20	86.40	92.20	87.49
N + R + P	99.75	<u>80.70</u>	91.60	87.10	40.00	79.83
						23.32

ตาราง 10 (ต่อ)

ชื่อมาตรฐานทดสอบ	ค่าอัตราความถูกต้อง (%)					ส่วน เบี่ยงเบน
	ภาพถ่ายปกติ	ภาพถ่ายเสียง	หลาຍນາດ	ภาพถ่าย	ภาพถ่ายที่ได้รับ ^{ผลการแบบของ}	
N + S + P	99.75	29.20	87.40	81.30	96.84	78.90
R + S + P	99.75	70.20	91.30	89.70	98.95	89.98
N + R + S + P	99.50	69.00	91.10	90.10	99.47	89.83
N + MRT	99.75	74.30	90.80	72.40	98.00	87.05

จากตาราง 10 จะเป็นการทดสอบที่แต่ละแบบจำลองสร้างจากข้อมูลจำนวนเท่ากัน ในทุกแบบจำลอง ดังนั้น ระยะเวลาในการสร้างแต่ละแบบจำลองจะใกล้เคียงกัน ซึ่งแตกต่างจากการวัดประสิทธิภาพรูปแบบที่ 1 จึงทำให้ผลการวัดประสิทธิภาพอาจมีความแตกต่างกัน ดังเช่น แบบจำลองที่สร้างจากข้อมูลภาพอ่อน弱 มีประสิทธิภาพในการรู้จำข้อมูลภาพตัวเลขหลายมุมมองสูงขึ้น จากการวัดประสิทธิภาพรูปแบบที่ 1 กว่า 18% แต่ประสิทธิภาพโดยรวมยังคงใกล้เคียงกัน และผลการเปรียบเทียบประสิทธิภาพแสดงให้เห็นว่า เมื่อเปรียบเทียบประสิทธิภาพแบบจำลองที่ผู้วิจัยได้นำเสนอ กับแบบจำลองที่สร้างโดยข้อมูล 2 ชนิด ซึ่งประกอบด้วยข้อมูลตัวเลขปกติ และข้อมูลตัวเลขจากการแปลงเชิงเรขาคณิต วิธีการที่ผู้วิจัยนำเสนอ นั้นมีประสิทธิภาพสูงที่สุดในทุกข้อมูลทดสอบ ยกเว้นข้อมูลภาพตัวเลขหลากหลายมุมมอง เนื่องจากวิธีการที่ผู้วิจัยนำเสนอ นั้นยังไม่มีการลดผลกระทบจากส่วนนี้โดยตรงดังที่กล่าวไว้ในผลการวัดประสิทธิภาพรูปแบบที่ 1 จึงมีค่าอัตราความถูกต้องต่ำกว่าแบบจำลองอื่น ๆ ประมาณ 1% – 18% อย่างไรก็ตามแม้ว่าจำนวนข้อมูลในการสร้างแบบจำลองจะน้อยกว่าแบบจำลองจากการวัดประสิทธิภาพรูปแบบที่ 1 แต่การเพิ่มความหลากหลายของข้อมูลยังคงเพิ่มประสิทธิภาพในการรู้จำให้กับแบบจำลองที่สร้างโดยสถาปัตยกรรม LeNet-5 บนโครงข่ายประสาทเทียมแบบคอนโวลูชันให้สูงขึ้น และแบบจำลองที่มีค่าอัตราความถูกต้องสูงที่สุด คือแบบจำลองที่สร้างจากข้อมูลทุกชนิด ยกเว้นข้อมูลภาพตัวเลขปกติ ($R + S + P$) มีอัตราความถูกต้องเท่ากับ 89.98% ซึ่งมีค่าสูงกว่าแบบจำลองที่สร้างจากข้อมูลทุกชนิด ($N + R + S + P$) ประมาณ 0.2% เท่านั้น ผู้วิจัยจึงสรุปได้ว่าผลลัพธ์การวัดประสิทธิภาพรูปแบบที่ 1 และ รูปแบบที่ 2 มีประสิทธิภาพในทิศทางเดียวกัน

นอกจากนี้ผู้วิจัยยังได้ทำการวัดประสิทธิภาพด้านเวลาในการประมวลผลการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน ซึ่งแสดงดังตาราง 11

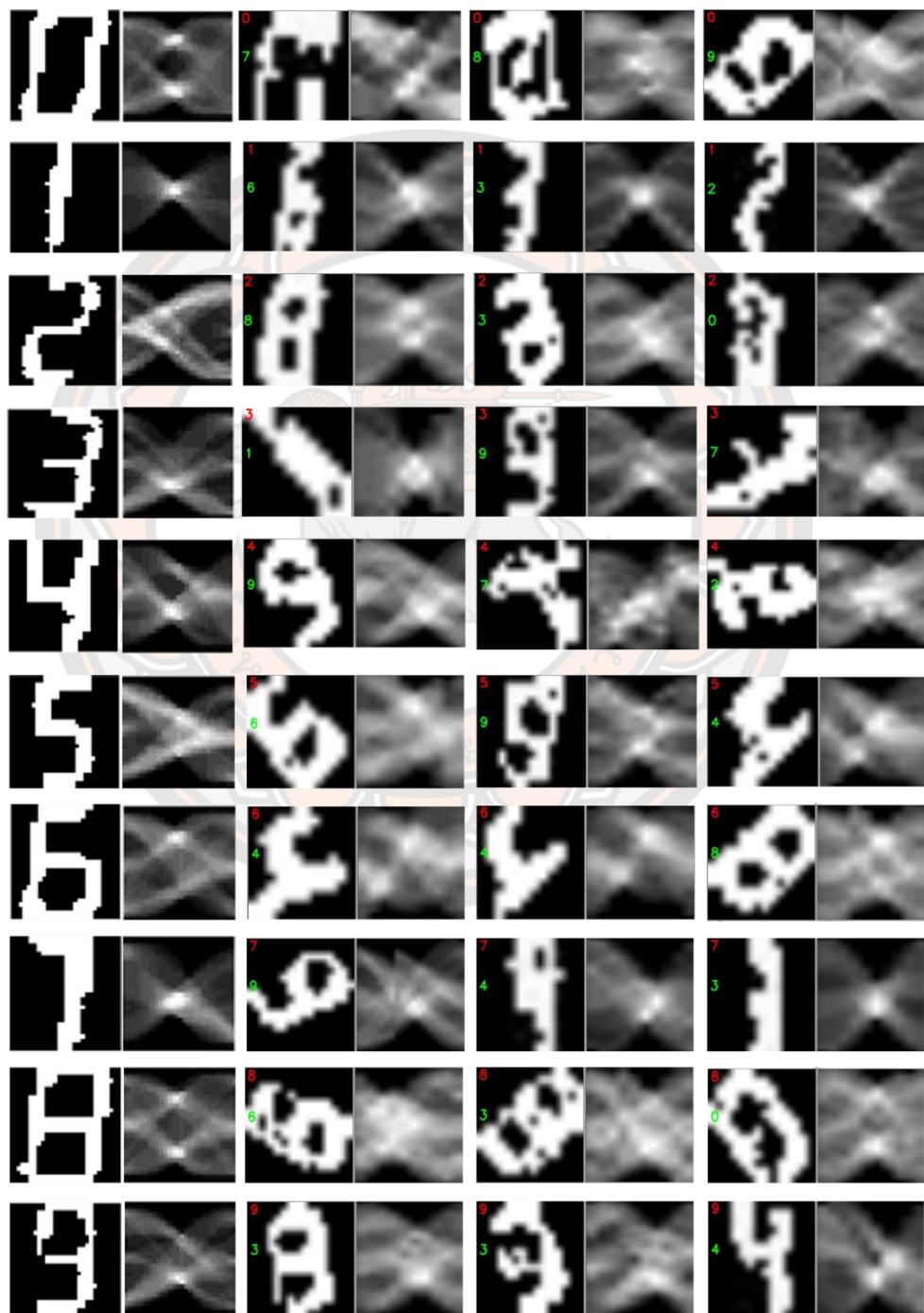
ตาราง 11 ผลการเปรียบเทียบเวลาในการประมวลผลการรู้จำ

เวลาที่ใช้ แบบจำลองการรู้จำ (วินาที)	เวลาที่ใช้สร้าง สร้างข้อมูล (วินาที)	เวลาที่ใช้สร้าง แบบจำลอง (วินาที)	เวลาประมวลผล รวม (วินาที)
N + R	0.5269	52.4524	52.9793
N + S	0.5043	53.3836	53.8879
N + P	0.5949	51.1052	<u>51.7001</u>
R + S	0.4186	53.1569	53.5755
R + P	0.4849	53.2769	53.7618
S + P	0.4612	52.3934	52.8546
N + R + S	0.2834	53.1621	53.4455
N + R + P	0.3169	55.8724	56.1893
N + S + P	0.3123	54.0313	54.3436
R + S + P	0.3541	53.5046	53.8587
N + R + S + P	0.2513	55.0412	55.2925
N + MRT	4.0192	52.6039	56.6231

จากการวัดประสิทธิภาพด้านเวลาแสดงให้เห็นว่าวิธีการที่ผู้วิจัยได้นำเสนอใช้เวลาในการประมวลผลมากที่สุด เนื่องจากกระบวนการที่ทางผู้วิจัยพัฒนาขึ้นมาประกอบด้วยกระบวนการเพิ่มข้อมูลโดยการแปลงเชิงเรขาคณิต นั่นคือ การหมุนภาพจากกระบวนการแปลงเรเดอน การย่อ - ขยายภาพจากกระบวนการแปลงขนาดภาพ เมื่อพิจารณารวมกับกระบวนการขยายภาพด้วยศูนย์ การกรองภาพให้เรียบ และการจัดเรียงสัมประสิทธิ์ใหม่เรดอนทำให้เวลาในการประมวลผลของวิธีการที่ผู้วิจัยนำเสนอสูงกว่าวิธีการอื่น

อย่างไรก็ตามเวลาในการสร้างแบบจำลองนี้จะไม่ส่งผลต่อการใช้งาน เนื่องจากผู้ใช้งานสามารถนำแบบจำลองการรู้จำที่สร้างเรียบร้อยแล้วมาใช้งานได้โดยตรง

จากการรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน ผู้วิจัยได้พิจารณาถึงความสามารถในการรู้จำของแบบจำลองที่สร้างจากวิธีการที่ผู้วิจัยนำเสนอโดยการวิเคราะห์ความผิดพลาดในการทำนายผลลัพธ์ของแบบจำลองซึ่งให้ผลลัพธ์ดังภาพ 60



ภาพ 60 ผลลัพธ์การทำนายที่ผิดพลาดของวิธีการที่นำเสนอ

จากภาพ 60 เป็นภาพตัวเลขการทำนายที่ผิดพลาดของแบบจำลองที่สร้างจากข้อมูลภาพไซโนแกรมที่ผู้วิจัยนำเสนอ ซึ่งตัวเลขสีเขียวแสดงถึงคลาสของตัวเลข และตัวเลขสีแดงแสดงถึงผลลัพธ์การทำนาย

ผู้วิจัยได้ทำการวิเคราะห์เบื้องต้นและสันนิษฐานถึงสาเหตุของผลการทำนายที่ผิดพลาด ซึ่งปัจจัยหลักที่ส่งผลต่อการทำนายที่ผิดพลาดนั้นเกิดจากความคล้ายคลึงกันของลักษณะไซโนแกรมของภาพตัวเลข โดยมีสาเหตุดังต่อไปนี้

- การหมุนภาพ
- การเอียงเชิงมุมมอง
- การกระเจิงของแสง ซึ่งเกิดจากการหมุน หรือการเอียงเชิงมุมมอง

เนื่องจากสาเหตุข้างต้นส่งผลกระทบต่อกระบวนการตรวจสอบภาพตัวเลขที่แสดงผลในรูปแบบเบ็ดส่วนบนของภาพแลอล็อตที่ผู้วิจัยนำเสนอทำให้มีผลลัพธ์ภาพตัวเลขมีรูปทรงที่เสียหาย ซึ่งอาจมีโครงสร้างที่เพิ่มขึ้นจากการกระเจิงของแสง หรือลดลงจากมุมมองของภาพที่เปลี่ยนไป ด้วยเหตุนี้จึงส่งผลให้ลักษณะไซโนแกรมของภาพตัวเลขที่นำมาทดสอบมีความคลาดเคลื่อนและคล้ายคลึงกับไซโนแกรมของภาพตัวเลขปกติในคลาสอื่น ๆ ที่ได้นำไปสร้างแบบจำลองในการรู้จำ

อย่างไรก็ตามสาเหตุที่ผู้วิจัยได้กล่าวไว้ในงานวิจัยฉบับนี้ยังเป็นเพียงข้อสันนิษฐานของผู้วิจัยเท่านั้น ซึ่งต้องการการพิสูจน์ทางทฤษฎี และหลักการทางคณิตศาสตร์ต่อไป

บทที่ 5

บทสรุป

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการตรวจจับและรู้จำข้อความที่แสดงผลแบบเจ็ดส่วนบนจากการแอลอีดี ซึ่งทางผู้วิจัยเริ่มพัฒนาจากการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดี

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจากการแอลอีดีโดยผู้วิจัยได้สรุปข้อจำกัดในการใช้งานวิธีการในการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนไว้ดังตาราง 12 และ ตาราง 13

ตาราง 12 สรุปข้อจำกัดการตรวจจับภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

หัวข้อ	ประสิทธิภาพวิธีการที่นำเสนอ
สีของแอลอีดี	ทำการทดลองเฉพาะแอลอีดีสีแดงเท่านั้น
ระยะห่างในการถ่ายภาพ	ประมาณ 1 เมตรจากหน้าจอแสดงผล
ลักษณะของภาพถ่าย	มุ่งมองตรง ไม่มีผลกระทบจากการเอียง
ขนาดภาพที่ใช้ทดลอง	1280 × 720 พิกเซล

ตาราง 13 สรุปข้อจำกัดการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน

หัวข้อ	ประสิทธิภาพวิธีการที่นำเสนอ
สีของภาพตัวเลข	ภาพตัวเลขระดับสีเทา หรือขาวดำ
ระยะห่างในการถ่ายภาพ	อยู่ในช่วง 0.25 จนถึง 1.5 เมตรจากหน้าจอแสดงผล
ขนาดของภาพตัวเลข	อยู่ในช่วง 2×6 พิกเซล จนถึง 36×52 พิกเซล
ความเอียงจากการหมุน	อยู่ในช่วง -80° จนถึง 80°
ความเอียงเชิงมุมมอง	อยู่ในช่วง -60° จนถึง 60°
สภาพของแสง	แสงสว่างจากไฟภายในห้อง และกลางคืนสภาพแสงน้อยหรือไม่มีแสง

วิธีการตรวจจับที่ผู้วิจัยนำเสนอสามารถนำตัวเลขของมาจากการภาพได้โดยใช้ค่าของสี และความสว่างในช่วงที่กำหนดบนแบบจำลองสี HSV ซึ่งตัวเลขที่ได้รับจะถูกส่งไปยังกระบวนการสกัดคุณลักษณะที่เรียกว่า การแปลงเรดอน helynx น้ำดึงสามารถลดผลกระทบจากภาพถ่ายเอียง ความหลากรายของระยะห่างในการถ่ายภาพ ความหลากหลายของมุมมองการถ่ายภาพ รวมถึงภาพที่ได้รับผลกระทบจากแสงได้ แล้วคุณลักษณะที่สกัดได้จะถูกส่งไปยังกระบวนการเรียนรู้ด้วยโครงข่ายประสาทเทียม ซึ่งกระบวนการทั้งหมดที่ทางผู้วิจัยนำเสนอ มีผลลัพธ์ที่น่าพึงพอใจ โดยมีค่าความแม่นยำและค่าการเรียกคืนตัวอักษรที่ 94.46% และ 92.24% ตามลำดับสำหรับการตรวจจับ และมีค่าเฉลี่ยของอัตราความถูกต้องที่ 87.05% สำหรับการรู้จำ

5.2 ข้อเสนอแนะ

1. ใน การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน เมื่อนำไปใช้งานกับภาพตัวเลขที่มีผลกระทบจากการเอียง ขนาดภาพ และมุมมองภาพ ทำให้กระบวนการกรองขั้นต้นจากขั้นตอนที่ 3.1.4 ในบทที่ 3 มีความผิดพลาดในการแยกแยะและตัดออกวัตถุที่ไม่ใช่ตัวเลข
2. สำหรับกระบวนการจัดเรียงสัมประสิทธิ์ใหม่เรดอน ผู้วิจัยมุ่งพิจารณาเวกเตอร์หลักจากค่าที่สูงที่สุดบนไซโโนแกรม และส่วนที่ส่งผลต่อค่าบนไซโโนแกรมคือความหนาแน่นของพิกเซลของวัตถุ ในทิศทางที่กำหนด ซึ่งในบางกรณีถึงแม้วัตถุบนภาพนำเข้ามีรูปทรงคล้ายคลึงกันแต่ผลลัพธ์จากกระบวนการจัดเรียงสัมประสิทธิ์ใหม่นั้นแตกต่างกัน ซึ่งนำเสนอในภาคผนวก ก ดังภาพ 61
3. ในงานวิจัยฉบับนี้ได้นำเสนอเรื่องการตรวจจับและรู้จำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนของภาพแออลอีดี โดยมีวิธีการหลักในการตรวจจับคือการตัดแบบกำหนดช่วงบนแบบจำลองสี HSV และวิธีการหลักในการรู้จำคือการแปลงเรดอน helynx (MRT) ซึ่งในกระบวนการรู้จำนี้มีความแตกต่างของข้อมูลภาพตัวเลขปกติเพียงเล็กน้อยผ่านวิธีการที่ผู้วิจัยนำเสนอเพื่อสร้างแบบจำลองการรู้จำ และมีเพียงกระบวนการเพิ่มความหลากหลายของข้อมูลด้วยตัวกรองเกาส์เซียนหลายขนาด เท่านั้น ในขณะที่ความหลากหลายของข้อมูลสามารถส่งเสริมให้โครงข่ายประสาทเทียมแบบconvolutional ซึ่งมีประสิทธิภาพที่สูงขึ้นได้ ดังนั้น การเพิ่มความหลากหลายของข้อมูลจึงยังเป็นแนวทางในการพัฒนาวิธีการรู้จำให้มีประสิทธิภาพสูงขึ้น

บรรณานุกรม

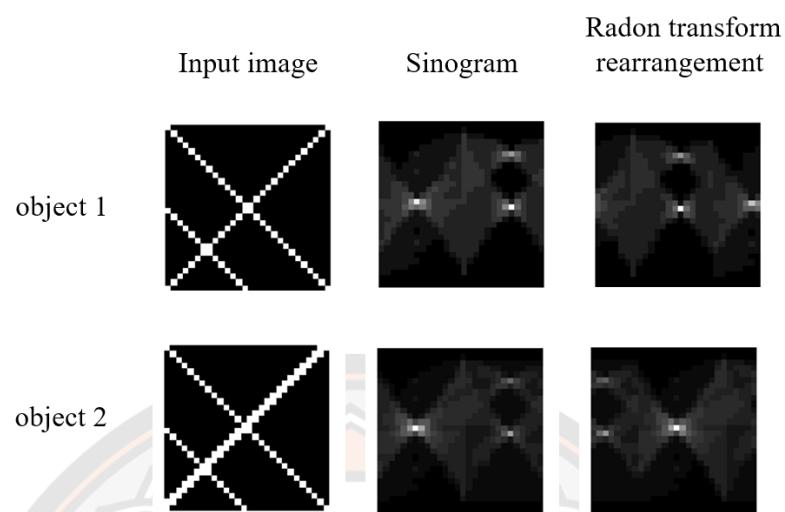
- Bonacic, I., Herman, T., Krznar, T., Mangić, E., Molnar, G., & Cupic, M. (2009). **Optical Character Recognition of Seven – segment Display Digits Using Neural Networks.** Paper presented at 2009 - 32st International Convention on Information and Communication (MIPRO), Technology, Electronics and Microelectronics, 323–328.
- Corporation, I. (2011). **Color Models.** Retrieved from https://scc.ustc.edu.cn/zlsc/sugon/intel/ipp/ipp_manual/index.htm#IPPI/ippi_ch6/ch6_color_models.htm
- Dong, Q., Wu, A., Dong, N., Feng, W., & Wu, S. (2018). **A Convolution Neural Network for Parts Recognition Using Data Augmentation.** Paper presented at the 2018 13th World Congress on Intelligent Control and Automation (WCICA).
- Doughty, M. (2016). Two Kinds of Color. Retrieved from <https://www.sketchpad.net/basics4.htm>
- Ghugardare, R. P., Narote, S. P., Mukherji, P., & Kulkarni, P. M. (2009). **Optical character recognition system for seven segment display images of measuring instruments.** Paper presented at the TENCON 2009 - 2009 IEEE Region 10 Conference.
- Haralick, R. M., & Shapiro, L. G. (1992). **Computer and robot vision** (Vol. 1): Addison-wesley Reading.
- Jafari ,K. K., & Soltanian ,Z. H. (2005). **Radon transform orientation estimation for rotation invariant texture analysis.** IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6), 1004-1008.
- Jiang, G. Q., & Zhao, C. J. (2012). **Apple recognition based on machine vision.** Paper presented at the 2012 International Conference on Machine Learning and Cybernetics.
- Jiang, T., Cui, Z., Zhou, Z., & Cao, Z. (2018). **Data Augmentation with Gabor Filter in Deep Convolutional Neural Networks for Sar Target Recognition.** Paper

- presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.
- Junhong, L., Quan, P., Hongcai, Z., & Peiling, C. (2003). **Image recognition using Radon transform.** Paper presented at the Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems.
- KMUTNB, E. S. L. E. (2013). **การแสดงตัวเลขสองหลักบน 7-Segment Display ด้วยวิธี Time-Multiplexing.** Retrieved from <http://cpre.kmutnb.ac.th/esl/learning/index.php?article=74hc595-7seg-multiplexing>
- Kulkarni, P. H., & Kute, P. D. (2016). **Optical numeral recognition algorithm for seven segment display.** Paper presented at the 2016 Conference on Advances in Signal Processing (CASP).
- LeCun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). **Gradient-Based Learning Applied to Document Recognition.** 46.
- Lefèuvre, F., Macq, B., & Legat, J. D. (2002). **RASH: RAdon soft hash algorithm.** Paper presented at the 2002 11th European Signal Processing Conference.
- Mu, G., Xinyu, Z., Deyi, L., Tianlei, Z., & Lifeng, A. (2015). **Traffic light detection and recognition for autonomous vehicles.** The Journal of China Universities of Posts and Telecommunications, 22(1), 50-56.
- Omachi, M., & Omachi, S. (2009). **Traffic light detection with color and edge information.** Paper presented at the 2009 2nd IEEE International Conference on Computer Science and Information Technology.
- Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv:1712.04621 [cs].
- Popayorm, S., Titijaroonroj, T., Phoka, T., & Massagram, W. (2019). **Seven Segment Display Detection and Recognition using Predefined HSV Color Slicing Technique.** Paper presented at the 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE).
- Popayorm, S., Titijaroonroj, T., Phoka, T., & Massagram, W. (2020). **Data Augmentation Based on Multiscale Radon Transform for Seven Segment Display**

- Recognition.** Paper presented at the 2020 12th International Conference on Knowledge and Smart Technology (KST).
- Production, N. (2017). **LED Display គីអេអ្រ**. Retrieved from <https://www.nmproduction9.com/led-display-%E1%EB%ED%EA%EC/>
- Radon, J. (1986). **On the Determination of Functions from Their Integral Values along Certain Manifolds.** IEEE transactions on medical imaging, 5(4), 170-176.
- Rong-Chi, C. (2011). **Intelligent text detection and extraction from natural scene images.** Paper presented at the 2011 16th North-East Asia Symposium on Nano, Information Technology and Reliability (NASNIT).
- Sánchez, S. A. G., Ruiz, C. A., & Meza, C. E. A. (2005). **GEOMETRIC TRANSFORMATIONS AND ITS APPLICATION IN DIGITAL IMAGES**, 60, 4.
- Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017). **Research on data augmentation for image classification based on convolution neural networks.** Paper presented at the 2017 Chinese Automation Congress (CAC).
- Taylor, L., & Nitschke, G. (2018). **Improving Deep Learning with Generic Data Augmentation.** Paper presented at the 2018 IEEE Symposium Series on Computational Intelligence (SSCI).
- Uln2003. (2018). **Seven-segment display.** Retrieved from https://en.wikipedia.org/wiki/Seven-segment_display
- Zhang, Y., & Wu, L. (2011). **A Rotation Invariant Image Descriptor based on Radon Transform.** International Journal of Digital Content Technology and its Applications. 209-217.
- សេរីវិទ្យា ជាអើយ, ន. (2012). **ចែ LCD គីអេអ្រ ?** Retrieved from <https://www.gotoknow.org/posts/51805>







ภาพ 61 ปัญหาในการจัดเรียงสัมประสิทธิ์ใหม่เรดอน

ภาคผนวก ข
บทความที่นำเสนอในงานประชุมวิชาการ

1. บทความเรื่อง “Seven Segment Display Detection and Recognition using Predefined HSV Color Slicing Technique” นำเสนอในงานประชุมวิชาการ The 16th International Joint Conference on Computer Science and Software Engineering (JCSSE2019) จัดที่โรงแรม Amari Pattaya จังหวัดชลบุรี ประเทศไทย ระหว่างวันที่ 10-12 กรกฎาคม พ.ศ.2562 บทความนี้ตีพิมพ์ไว้ใน Proceedings of 2019 - 16th International Joint Conference on Computer Science and Software Engineering หน้า 224-229
2. บทความเรื่อง “Data Augmentation Based on Multiscale Radon Transform for Seven Segment Display Recognition” นำเสนอในงานประชุมวิชาการ The 12th International Conference on Knowledge and Smart Technology (KST-2020) จัดที่โรงแรม Amari Pattaya จังหวัดชลบุรี ประเทศไทย ระหว่างวันที่ 29 มกราคม – 1 กุมภาพันธ์ พ.ศ.2563 บทความนี้ตีพิมพ์ไว้ใน Proceedings of 2020 - 12th International Conference on Knowledge and Smart Technology หน้า 47-51

Seven Segment Display Detection and Recognition using Predefined HSV Color Slicing Technique

Sorawee Popayorm ^{*}, Taravichet Titijaroonroj[†], Thanathorn Phoka[‡] and Wansuree Massagram[§]

^{*†§}Faculty of Science, Department of Computer Science and Information Technology,

Naresuan University, Phitsanulok, 65000, THAILAND

Email: sorawee61@email.nu.ac.th^{*}, thanathornp@nu.ac.th[†], and wansureem@nu.ac.th[§]

[†]Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, THAILAND. Email: taravichet@it.kmitl.ac.th[†]

Abstract—Detection and recognition of LED seven segment panels present a particular challenge of locating characters among background clutter. This study proposed a framework based on a predefined HSV color slicing technique. The results demonstrate the framework's superiority over other color slicing methods with 94.46% precision, 92.24% recall, and 87.17% accuracy rates. The digit detection and recognition with the proposed predefined HSV color slicing is simple and straightforward — making it an attractive solution for the future deployment for edge-computing.

Keywords—Predefined HSV color slicing; seven segment display detection; seven segment number recognition; edge-computing

I. INTRODUCTION

Industrial machinery and household appliances use seven-segment displays predominantly as indicators of their statuses. Machine operators and appliance users are required to read the liquid-crystal displays (LCD) or the light-emitting diode (LED) based displays to correctly operate them. Optical character readers (OCR) have emerged as a solution to detection and recognition of printed text documents; however, they are not yet suitable to tackle the problems light-source displays present. The challenges of detecting and recognizing characters from the LCD and LED include, but not limited to, (1) locating the characters from background clutter and (2) handling brightness and contrast of the image.

This particular study proposes a solution for detecting and recognizing characters from a red LED seven segment display. While, initially, this problem may seem intuitive and unchallenging, the characters are difficult to detect and read under typical conditions. The common seven-segment characters consist of digit 0 through 9. They are fixed font which should not present any difficulty in reading. Nonetheless, in each digit, the image is filled with background clutter, not the text of interest, which makes it hard to achieve text recognition through normal OCR techniques.

Additionally, most LED displays have very high contrast, i.e. black background with bright lights. Most captured images suffer from saturation. The edges around the digit segments are fuzzy. This problem is known as “bleeding”.

These two main challenges are addressed straightforwardly in this paper using color slicing technique. The past approaches for the specific problem of reading LED/LCD displays in natural images include the works from [1], [2], and [3]. Bonai

et al. [1] introduced an optimal topology of neural networks to recognize seven-segment digits. The results from [1] was impressive with above 90% success rate. Nonetheless, neural networks are time-consuming and require high-performance hardware as explained in [2], [3]. Ghugardare et al. [2] presented a scissoring algorithm for image segmentation to recognize the seven segment display from digital multimeters. This alternative approach is less taxing than neural networks because the local features could be easily extracted from the precise edges and corners. Similarly, Kulkarni et al. [3] proposed a simplistic cropping technique for seven segment recognition of LCD-based meters. The cropping for segmentation was based on X-and-Y profiles. The number recognition was based on the pixel density feature and a decision table.

Most of these mentioned approaches were performed on personal computers or workstations. However, the ultimate goal for this research is to deploy an algorithm to detect and recognize characters on seven segment display in real-time using a less powerful edge computing, e.g. a NodeMCU or a Raspberry Pi, where the processing power is strictly limited. Thus, a simple predefined HSV (hue, saturation, value) color slicing approach was considered for this study to evaluate the algorithm's feasibility before applying to real-world problems.

The framework for the proposed system along with the background theory of each step are described in Section II. Section III shows the experimental results with the discussion of their significances. The conclusion and the direction of the future work for this study are discussed in Section IV.

II. PROPOSED METHOD

The work in this study is divided into two major elements: detection and recognition. Fig. 1 illustrates the proposed framework. The seven segment detection consists of four steps: HSV conversion, predefined HSV color slicing, object detection, and preliminary filtering. After the detection is performed correctly, the candidate objects will then go through the recognition element of the framework. The recognition part includes three steps: resizing, feature extraction with projection technique, and image classification using the Euclidean distance.

This section explains the background theory and technical details for both detection and recognition parts.

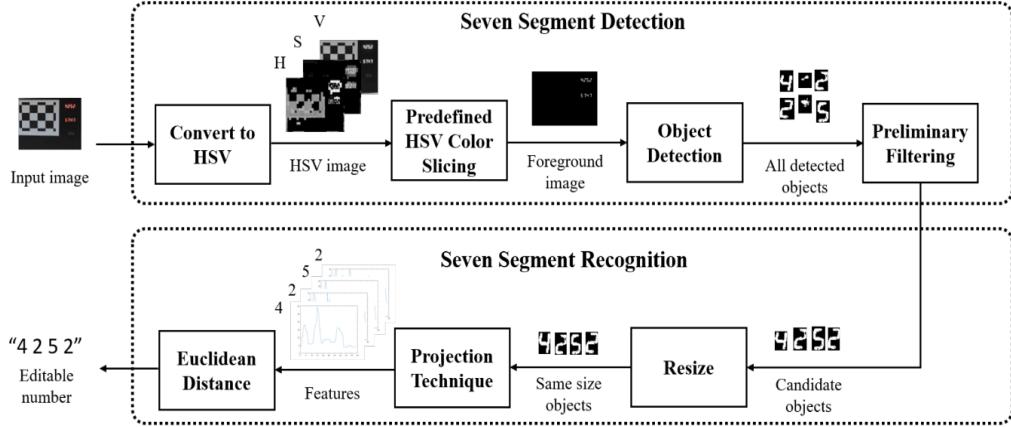


Fig. 1: The proposed framework for the seven segment display detection and recognition based on a predefined HSV color slicing technique

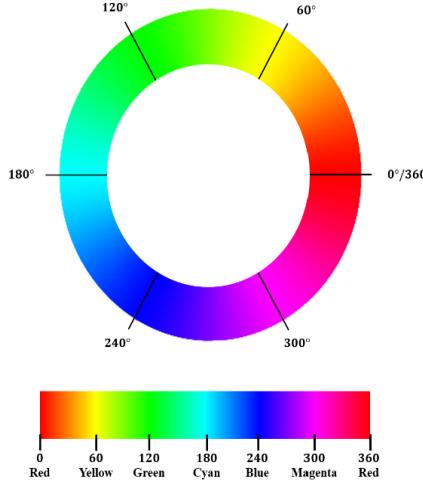


Fig. 2: The color ranges in hue components between 0° to 360° degrees.

A. Seven Segment Detection

To detect the seven segment display in any given image, this study proposed a simple predefined HSV color slicing technique, which could be achieved as follows:

1) HSV Conversion: the HSV color model consists of hue, saturation, and value. These three components correspond to the color type, the purity of color, and the brightness. The HSV model is widely used in image segmentation and image detection [4].

The hue components are in radial slices ranging from 0° to 360°. For this particular color model, each color has its own centroid as shown in Fig. 2. The centroids of yellow, green, cyan, blue, and magenta are at 60°, 120°, 180°, 240°

and 300° respectively. Due to the cyclical nature of the hue-to-angle mapping, red is centered at 0° and 360°.

The first step of the seven segment character detection is to convert the input image $I(x, y)$ with size $M \times N$ pixels in the RGB color model into the HSV color model using Eqs. (1) - (3).

$$H(x, y) = \begin{cases} 0, & I_{max} = I_{min} \\ 60^\circ \frac{G(x, y) - B(x, y)}{I_{max} - I_{min}} + 0^\circ, & I_{max} = R(x, y) \\ 60^\circ \frac{B(x, y) - R(x, y)}{I_{max} - I_{min}} + 120^\circ, & I_{max} = G(x, y) \\ 60^\circ \frac{R(x, y) - G(x, y)}{I_{max} - I_{min}} + 240^\circ, & I_{max} = B(x, y) \end{cases} \quad (1)$$

$$S(x, y) = \begin{cases} 0, & I_{max} = 0 \\ \frac{I_{max}}{I_{max} - I_{min}}, & \text{otherwise} \end{cases} \quad (2)$$

$$V(x, y) = I_{min} \quad (3)$$

where H , S , and V are the hue, saturation, and value components. R , G , and B are the red, green, and blue channels. I_{min} and I_{max} are defined by Eqs. (4) - (5).

$$I_{min}(x, y) = \min\{R(x, y), G(x, y), B(x, y)\} \quad (4)$$

$$I_{max}(x, y) = \max\{R(x, y), G(x, y), B(x, y)\} \quad (5)$$

The advantage of this color model is that the separation of colors can be performed easily via the hue component. Moreover, since the hue components are not sensitive to illumination, the separation of colors is robustness to lighting conditions [5]. Therefore, the HSV color model is selected in the proposed method. The concept of predefined HSV color slicing is explained below.

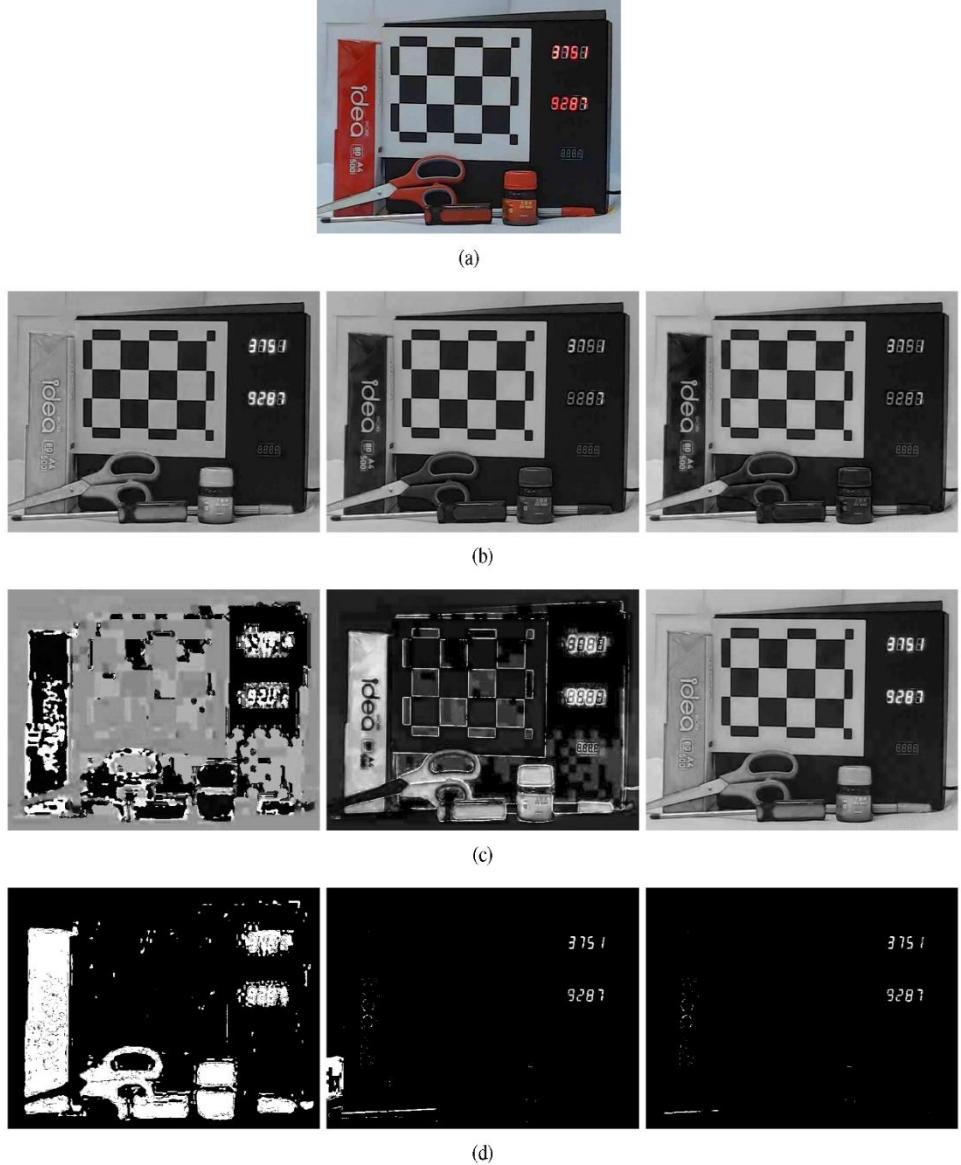


Fig. 3: Images of (a) a seven segment display with cluttering objects, (b) R-G-B components, (c) H-S-V components and (d) outputs from predefined HSV color slicing.

2) *Predefined HSV Color Slicing:* to divide the image pixels into two groups—foreground (seven segment characters) and background (clutter residues), color slicing technique is proposed. For this particular study, the seven segment LEDs are in red, hence, a simple predefined HSV color slicing is suitable for this situation.

The predefined HSV color slicing process begins with boundary setting. First, the left and right boundaries of the color range are set according to the color type as shown in Table I. The values of the boundaries are defined from the middle degree between two nearest neighbor color types. For

example, in case of blue color, the left boundary, t_{left} , is defined from the average between the centroid of cyan and blue colors whereas the right boundary, t_{right} , is defined from the average between the centroid of magenta and blue colors.

In this work, the color of interest is red. Therefore, its t_{left} is set to 331° and its t_{right} is set to 30° .

Once the color boundaries are set, the foreground image I_{out} containing the candidate pixel at the coordinate (x, y) can be defined by Eq. (6).

$$I_{out}(x, y) = I_h(x, y) \cap I_v(x, y) \quad (6)$$

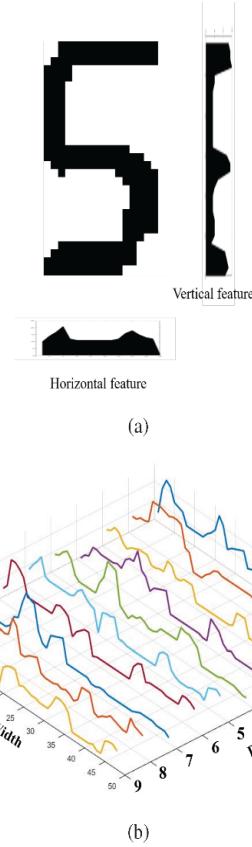


Fig. 4: An illustration of (a) seven segment display feature and (b) comparison of features from digit zero to nine.

where \cap is an intersection operation. $I_h(x, y)$ and $I_v(x, y)$ correspond to hue and value conditions which are defined by Eqs. (7) and (8), respectively.

$$I_h(x, y) = \begin{cases} 1, & t_{left} \leq H(x, y) \leq t_{right} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$I_v(x, y) = \begin{cases} 1, & 0.8 \leq V(x, y) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $H(x, y)$ and $V(x, y)$ are the hue and the value components. Fig. 3 shows an input image which contains many red objects such as a pen, a screwdriver, a bottle, a pair of scissors, and two seven segment panels. Distinguishing red objects from one another could be difficult. However, the redness from the seven segment display is noticeably different from other red objects. The color type of the panels occurs from the LED light source, providing a much higher brightness than the residual red objects. This is the main reason why the hue component must be considered along with the value component. Interestingly, the saturation component is noticeably missing from this step. $S(x, y)$ represents the color purity and, hence, unnecessary in for a single-color LED display.

TABLE I: The predefined left and right boundaries of color range in hue component.

Color Types	red	yellow	green	cyan	blue	magenta
t_{left}	331°	31°	91°	151°	221°	271°
t_{right}	30°	90°	150°	210°	270°	330°

TABLE II: The conditions of the noisy elimination.

Condition Rules	Meaning
$cH > cW$	incorrect proportion object
$cN < (0.005 \times (M \times N))$	small object

After transforming the input image $I(x, y)$ into the HSV color model during the previous step, the predefined HSV color slicing extracts the foreground image with the hue and value components calculated from Eqs. (6) - (8). The result of the predefined HSV color slicing is displayed in Fig. 3(d) with I_h , I_v , and I_{out} arranging from left to right.

3) *Object Detection*: all candidate pixels in the foreground image are grouped into a set of candidate objects based on the analysis of connected pixels, as described in [6]. The two-dimensional binary image acts as an input for this analysis. Eight neighboring pixels are considered whether they are a part of the same object if its binary value is equal to that of the center.

4) *Preliminary Filtering*: the preliminary filtering for noisy object elimination is defined as shown in Table II where cH , cW , and cN are the candidate objects height, width, and pixel number. Generally, a height of seven segment object is higher than its width. Additionally, the small objects are assumed to be noisy objects and are recommended to be removed if the size is approximately 0.5% of the image size as explained in [7].

B. Seven Segment Recognition

The image recognition research domain can be categorized into two groups: (1) learning and (2) handcrafting approaches. The learning approaches, e.g. artificial neural network or convolution neural network, automatically extract the abstract features and feed them into a classification for prediction. The advantage of the learning approach is its effective performance. The main disadvantage of this approach is its demanding processing requirement for both time and power [8], [2]. Contrarily, the handcrafting approaches are simpler and do not require vast training data or computing time. Image recognition through handcrafting is flexible and can be tailored specifically for hardware implementation, thus suitable for this study. There are three proposed steps in recognizing the seven segment characters with the handcrafting approach as explained below.

1) *Resizing*: since the candidate objects from the seven segment detection are various in sizes, the extracted feature size would subsequently be different. This problem could be avoided by resizing all candidate objects to be 30×18 pixels by using bicubic interpolation to standardize the image sizes.

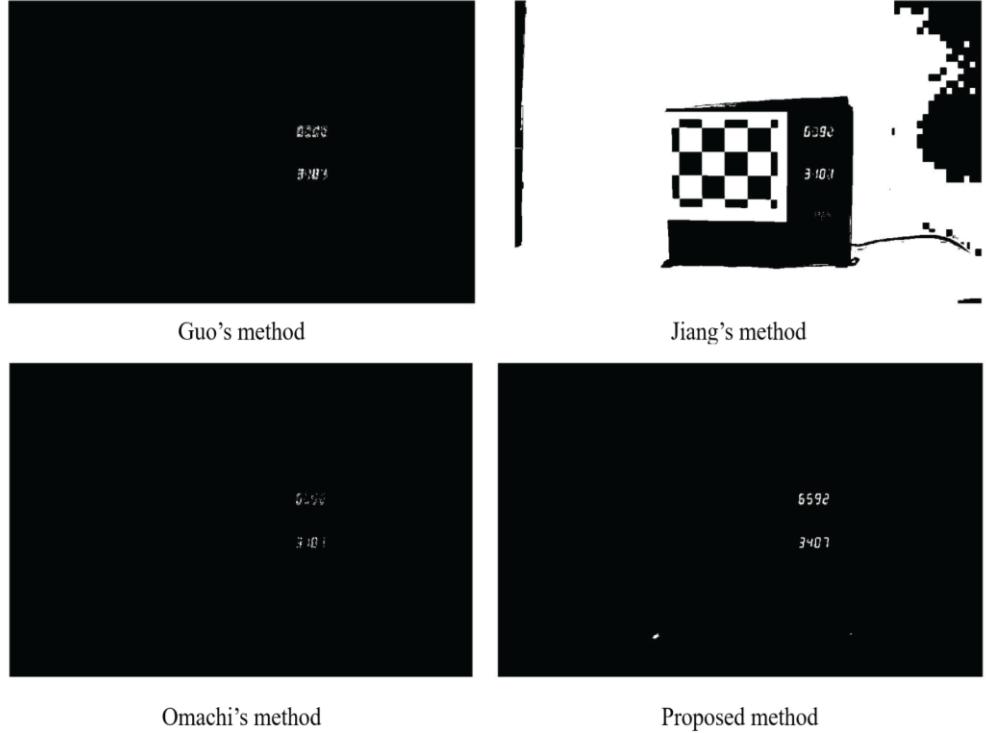


Fig. 5: The foreground images computed by the baseline and the proposed methods.

2) *Projection Technique*: the feature f is extracted using a projection technique in vertical and horizontal directions. This step could also be considered as the feature extraction phase. Previously during the resizing phase, the height and width of the candidate object are 30 and 18, therefore 30 and 18 features in vertical and horizontal directions are obtained — making the total number of features to be 48 dimensions as defined in Eq. (9).

$$f = [f_{vertical} \ f_{horizontal}] \quad (9)$$

where $f_{vertical}$ and $f_{horizontal}$ are the vertical and the horizontal features computed in Eqs. (10) and (11), respectively.

$$f_{vertical}(l) = \sum_{k=1}^K B(k, l) \quad (10)$$

$$f_{horizontal}(k) = \sum_{l=1}^L B(k, l) \quad (11)$$

where B is the resized image of the candidate object with K and L being the height and the width of that particular object. The feature f , extracted from Eq. (9), is the summation of the intensity in vertical and horizontal directions as shown in Fig. 4(a). This feature represents the candidate objects shape and structure and can be used to differentiate the distinct features in each group as shown in Fig. 4(b).

TABLE III: A comparison of the performance of the seven segment detection and recognition with proposed and baseline methods.

Methods	Detection		Recognition
	Precision	Recall	Accuracy rate
Jiang's method	70.94%	69.48%	18.56%
Omachi's method	91.01%	28.91%	40.08%
Guo's method	92.20%	55.31%	10.58%
Proposed method	94.46%	92.24%	87.17%

3) *Euclidean Distance*: to classify the candidate object into the correct class, the Euclidean distance is required. This method is widely used in many recognition applications such as face, character, or fingerprint. Moreover, the Euclidean distance computation is less time consuming and less taxing on hardware performance. The distance d between the features from testing and training objects can be computed in Eq. (12).

$$d(f_{tr}, f_{ts}) = \sqrt{\sum_{p=1}^{K+L} (f_{tr}(p) - f_{ts}(p))^2} \quad (12)$$

where f_{ts} and f_{tr} are the testing and the training features obtained from the projection technique. The feature f_{ts} is classified into the same group as f_{tr} when their Euclidean distance is at a minimum. Hence, this step is also known as the image classification phase.

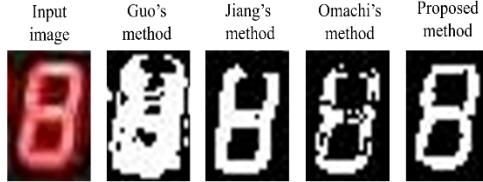


Fig. 6: The seven segment display object extracted from the baseline and the proposed methods.

III. EXPERIMENT AND DISCUSSION

The proposed framework was tested on 240 images with a size of 1280×720 pixels. The images were taken from the Logitech HD Pro C920 webcam without any rotation or special lighting conditions. The performance of the proposed method is evaluated and compared against existing red color slicing methods including Guo's [9], Jiang's [10], and Omchi's [11] in terms of precision, P , recall, R , and accuracy, A_c , as calculated by Eqs. (13) to (15).

$$P = \frac{N_{DC}}{N_{Actual}} \times 100 \quad (13)$$

$$R = \frac{N_{DC}}{N_{DC} + N_{DI}} \times 100 \quad (14)$$

$$A_c = \frac{N_c}{N_{DC} + N_{DI}} \times 100 \quad (15)$$

where N_{DC} is the number of correctly detected objects and N_{DI} is the number of incorrectly detected objects. N_{Actual} and N_c are the number of existing object and correctly recognized object, respectively.

As shown in Table III, the experimental results are divided into two parts: detection and recognition. The first part of the table shows the precision and recall of the detection algorithm. The second part shows the recognition accuracy rate. All performance metrics of the proposed predefined HSV color slicing were better than all three existing methods with the recall and accuracy rate higher than the second-best methods by more than 20 and 45 percent.

The proposed predefined HSV color slicing, while fundamentally simple, could achieve greater performance due to its ability to better separate the foreground from the background clutter. As shown in Fig. 5, the foreground images derived from Jiang's and Omachi's methods, which were based on the RGB color model, were deficient. Since the RGB color models represent the combination of pure color, saturation, and brightness, many unwanted objects could be obtained — leading to an incomplete character as shown in Fig. 6. Guo's, despite being based on the HSV color model, was not nearly as effective as the proposed method due to its use of all H, S, and V components.

Nevertheless, the proposed predefined HSV color slicing involves an undeniable limitation. This particular work is limited to a red-color LED seven-segment display. For a broader range and more variety of colors, a better and more universal algorithm must be adopted. However, with respect to the red LED display, the proposed framework was justifiably satisfactory.

IV. CONCLUSION

The proposed predefined HSV color slicing was proven to successfully achieve accurate detection and recognition of the LED seven segment display. The separation of the foreground and background clutter was the key element that helps the proposed framework achieve greater accuracy than other baseline methods. Since this work was aimed to be deployed for edge-computing, the future for this study includes the porting of the algorithm to be run on a NodeMCU and/or a Raspberry Pi.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Computer Science and Information Technology, Naresuan University, and the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, for funding, and Dr. Noah Hafner for his insightful vision and valuable discussion.

REFERENCES

- [1] I. Bonai, T. Herman, T. Krzna, E. Mangi, G. Molnar, and M. Cupic, "Optical Character Recognition of Seven segment Display Digits Using Neural Networks," 2009.
- [2] R. P. Ghugardare, S. P. Narote, P. Mukherji, and P. M. Kulkarni, "Optical character recognition system for seven segment display images of measuring instruments," in *TENCON 2009 - 2009 IEEE Region 10 Conference*, Jan. 2009, pp. 1–6.
- [3] P. H. Kulkarni and P. D. Kute, "Optical numeral recognition algorithm for seven segment display," in *2016 Conference on Advances in Signal Processing (CASP)*, Jun. 2016, pp. 397–401.
- [4] P. Ganeshan and V. Rajini, "Assessment of satellite image segmentation in RGB and HSV color space using image quality measures," in *2014 International Conference on Advances in Electrical Engineering (ICAEE)*, Jan. 2014, pp. 1–5.
- [5] P. Wonghabut, J. Kumphong, R. Ung-arunyawee, W. Leelapatra, and T. Satiennam, "Traffic Light Color Identification for Automatic Traffic Light Violation Detection System," in *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, Jul. 2018, pp. 1–4.
- [6] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992, pp. 28–48.
- [7] R. C. Chang, "Intelligent text detection and extraction from natural scene images," in *The 16th North-East Asia Symposium on Nano, Information Technology and Reliability*, Oct 2011, pp. 23–28.
- [8] P. Kulkarni, A. Khatri, P. Banga, and K. Shah, "Automatic Number Plate Recognition (ANPR) system for Indian conditions," in *2009 19th International Conference Radioelektronika*, Apr. 2009, pp. 111–114.
- [9] G. Mu, Z. Xinyu, L. Deyi, Z. Tianlei, and A. Lifeng, "Traffic light detection and recognition for autonomous vehicles," *The Journal of China Universities of Posts and Telecommunications*, vol. 22, no. 1, pp. 50–56, Feb. 2015.
- [10] G. Jiang and C. Zhao, "Apple recognition based on machine vision," in *2012 International Conference on Machine Learning and Cybernetics*, vol. 3, Jul. 2012, pp. 1148–1151.
- [11] M. Omachi and S. Omachi, "Traffic light detection with color and edge information," in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, Aug. 2009, pp. 284–287.

Data Augmentation Based on Multiscale Radon Transform for Seven Segment Display Recognition

Sorawee Popayorm ^{*}, Taravichet Titijaroonoj [†], Thanathorn Phoka [‡] and Wansuree Massagram [§]

^{*†§}Faculty of Science, Department of Computer Science and Information Technology,

Naresuan University, Phitsanulok, 65000, THAILAND

Email: sorawee61@email.nu.ac.th^{*}, thanathornp@nu.ac.th[†], and wansureem@nu.ac.th[§]

[†]Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, THAILAND. Email: taravichet@it.kmitl.ac.th[†]

Abstract—To alleviate the problem of limited data in creating rotation, scale, perspective, and illumination invariant of the neural network training sets, the multiscale Radon transform is proposed in this study to enhance the data augmentation for seven segment display recognition. Resizing, smoothing, and coefficient shifting generate the desired invariant effects for the training model. The accuracy rates from the experiment demonstrate the superiority of the proposed method over other data augmentation techniques with the best overall accuracy performance of 87.05% — outperforming other data augmentation techniques by 6-13%. The convolutional neural network model generated from the proposed multiscale Radon transform data augmentation is suitable for seven segment display recognition and could become beneficial to other type of self-luminous type of images.

Keywords—multiscale Radon transform; data augmentation; convolutional neural network; seven segment display recognition

I. INTRODUCTION

One of the main challenges in any imaging processing and pattern recognition study is the availability of the data set. Especially when employing machine learning techniques, the algorithms would require a large amount of annotated sample data [1]. Collecting image data is a time-consuming procedure. Experts involved in the field of interest are often asked to label the images during this tedious process in order to gain a reliable training data set. To overcome this issue, researchers could use a technique called “data augmentation,” which involves simple modifications of data set images such as scaling, cropping, rotation, and translation. Using data augmentation to improve the neural networks’ training process is becoming a standard procedure in computer vision domain [2].

This paper presents data augmentation technique using Gaussian filter in conjunction with Radon transform to increase the training images for seven segment recognition with a convolutional neural network (CNN). The proposed multiscale Radon transform (MRT) provides variety of scaled and rotated images in sinograms. The Radon transform is traditionally used in computed axial tomography (CAT) scan. The tomographic images from a CAT scan are similar to images of seven segment displays with dark background and high-contrast self-luminous objects.

To detect and recognize characters from seven segment display, which are ubiquitous in daily life, the previous study

in [3] addresses the challenges straightforwardly using a predefined color slicing technique. Ghugardare et al. [4] developed the optical character recognition system for seven segment recognition in 2009 based on the statistical feature extraction. The aforementioned research reported that an artificial neural network (ANN) was time-consuming on less powerful hardware. However, the ANN would be an attractive option for recognition implementation on non-restrictive hardware as presented in the study by Bonacic et al. [5] in which the seven-segment digit recognition was achieved by feeding characteristic of digit from three techniques: the areas of interest, the resolution reduction, and the axis distribution analysis, to the neural network. Various multilayer feed-forward ANNs optimized by the genetic algorithm were employed. The prediction of each network, which is a vector size 10 weight of 10 digit classes, is used to predict the final result by using the average technique from those vectors.

In 2016, an optical numeral recognition algorithm for seven-segment displays was proposed by Kulkarni et al. [6]. The study focused on robustness to variations of digit positions, brightness, contrast, tilt, and noise. Their algorithms consisted of four main parts: preprocessing, noise removal, segmentation, and recognition. Firstly, the Radon transform was used to solve the inherent tilt problem of seven-segment images. Secondly, X and Y profiles of white pixels were used to background removal before denoising it by low pass filter. Thirdly, the vertical and horizontal cropping was used based on the X and Y profiles of black pixels for the display segmentation. Lastly, the black pixel density is feed to a decision tree for digit recognition.

Most recognition studies face one common problem the limitation of data. To significantly increase the diversity of data available for training models, Jiang et al. [7] used data augmentation to generate the limited synthetic aperture radar automatic target data and improve the deep convolution neural network in 2018. The Gabor filter is used to generate the diversity of data with multi-scale and multi-direction Gabor filters.

Increasing the data set size with augmentation could strengthen the performance and increase the accuracy of CNNs [8]. The main contribution of this paper is the proposed MRT-based data augmentation technique to create a training model which encourages less dependency on sizes, rotations, and lighting conditions, i.e., scale, rotation, and illumination invariant.

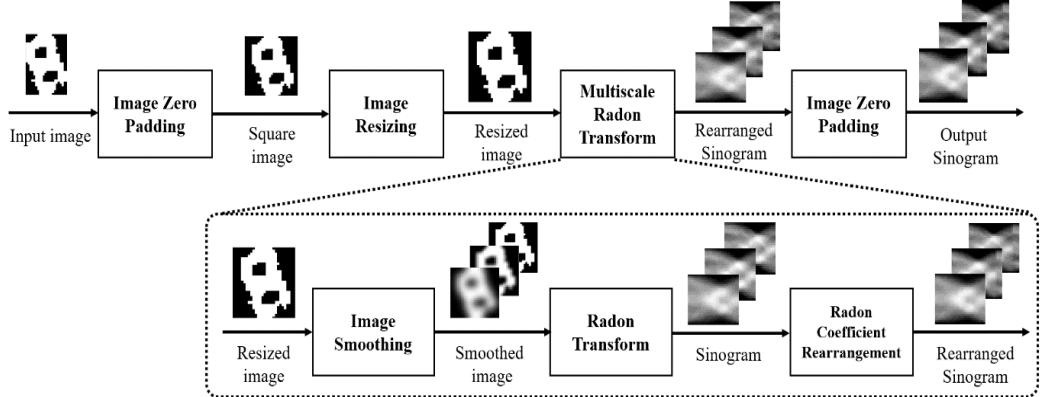


Fig. 1: The proposed data augmentation based on multiscale Radon transform for seven segment display recognition

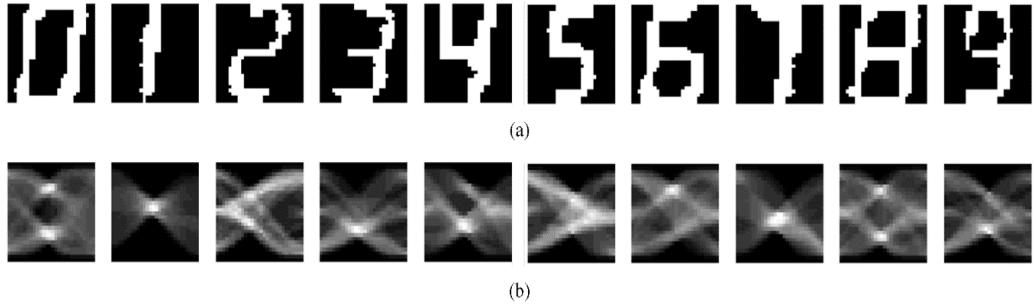


Fig. 2: Images of (a) digit 0-9 from seven segment display detection and (b) sinograms of digits from (a) displaying distinct features of each digit

In this study, the proposed data augmentation based on MRT along with the background theory are described in Section II. The experimental setup and results and their significance are shown in Section III. The conclusion and the direction of the future work for this study are discussed in Section IV.

II. PROPOSED METHOD

This section is divided into two main parts: (1) an overview of the proposed data augmentation for seven segment display recognition and (2) the background details of the multiscale Radon transform. Both are described as follows:

A. Data Augmentation for Seven Segment Display Recognition

The workflows of the proposed data augmentation is illustrated in Fig. 1. The binary input image, with the best-fit bounding-box to ensure translation invariant, is processed through a zero padding step to create a square image. This image is then resized to suit the neural networks. The next box represents where the transform (e.g. scaling, rotation, perspective, etc.) is performed. The image after the transformation will need to be padded again before being processed with a neural network.

1) Image zero padding: to preserve the original shape and the aspect ratio of the seven segment digit, an image zero padding technique is required. The input image is padded by zero on both left and right sides to achieve a square image if its height is greater than its width. Similarly, the input image is padded on top and bottom if its width is greater than its height.

2) Image resizing: the input images are, naturally, various in sizes. Image resizing is necessary to increase or decrease the total number of pixels, essentially providing the same effects as augmentation with scaling. In this particular work, linear transformation is used to modify the spatial relationship between pixels. Sánchez, Ruiz, and Meza [9] suggested that the geometric transform should include two operations: (1) a spatial transformation to relocate the pixels in the image pane and (2) an interpolation to estimate the intensity of the pixel in the transformed image. To achieve the best approximation of a pixel's intensity based on the values at surrounding pixels, the Nearest-neighbor interpolation was selected for this work.

3) Multiscale Radon transform: the resized image is transformed into a sinogram by the MRT method, described in the following section in details. The output obtained from this step will be zero padded again to preserve the sinogram characteristics. The rearranged and padded sinogram becomes a standard form in which the content of data shape is stored and

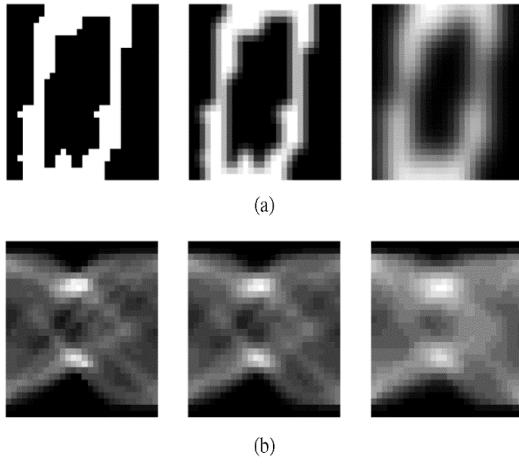


Fig. 3: Images of (a) smoothed images from three different Gaussian filter scales and (b) sinogram outputs of Radon transform from the smoothed images in (a)

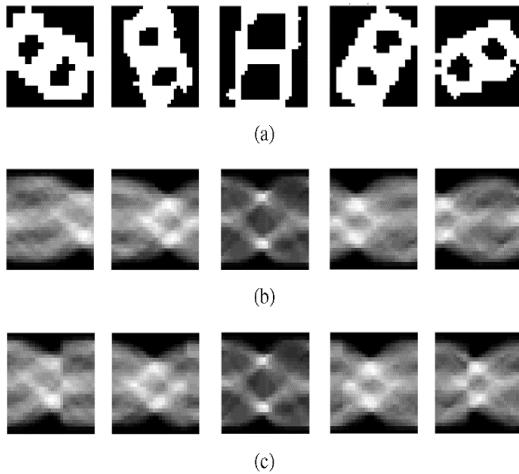


Fig. 4: Images of digit 8 with (a) various rotations, (b) sinogram features from Radon transform of those rotated images, and (c) Radon transform features with rearranged coefficients

is distinctive for different seven segment digit, as illustrated in Fig. 2. Furthermore, this standard form can retain the original content and withstand any effects taken on the data.

B. The Proposed Multiscale Radon Transform

In this paper, we propose the multiscale Radon transform for data augmentation on seven segment display recognition. This work is an extension of the Radon transform, a powerful method to match an object with various rotations by projecting its intensity of pixels on to a plane. The Radon transform has been applied to many applications such as image reconstruction of a CAT scan, object recognition in geophysics and astronomy, etc. However, the Radon transform itself is not

adequately robust to the variety of scales, rotations, perspectives, or lighting condition. To improve the performance of data augmentation for the training set, the multiscale Radon transform (MRT) is proposed. The main objective of the proposed method is transforming the input data from spatial domain to the standard form which is robust to the various effects. The proposed method consists of three main processes: image smoothing, Radon transform, and Radon coefficient rearrangement.

1) Image smoothing with Gaussian filter: aims to increase the diversity of the training images and enable the proposed solution to be more robust to the variety of image sizes. It consists of two steps as follows.

Step 1: Generate a Gaussian filter by Eq. (1).

$$G(u, v, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(u^2+v^2)/(2\sigma^2)} \quad (1)$$

where (u, v) is a location of an $S \times S$ filter and σ is a standard deviation of such filter.

Step 2: Smooth the input image (\tilde{I}) by Eq. (2).

$$\tilde{I} = I * G \quad (2)$$

where I is an input image and $*$ is a convolution operation. The smoothed images in three different scales, exhibited in Fig. 3, show that the pattern of each sinogram, which is the output of Radon transforming, are highly similar. In other words, the transformation is robust to variety of scales, i.e. scale invariant.

2) Radon transform: introduced by Johann Radon in 1917 [10] to process two-dimensional data onto one-dimensional projection. This technique is widely used in image reconstruction for medical image processing and recognition.

The Radon transform, R , maps an image in spatial domain to sinusoidal waves with different amplitudes and phases by computing the integral intensity of the image at a specific degree, and is defined in Eq. (3) by

$$R(\theta, \rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \quad (3)$$

where $i(x, y)$ is an input image data, δ is a Dirac delta function with θ as a specific degree, and ρ is the distance of the one-dimensional output.

The output of a Radon transform is called a “sinogram”, I_r , that be calculated from the smoothed image by Eq. (4).

$$I_r = R(\tilde{I}) \quad (4)$$

where R is the radon transform from Eq.(3).

3) Radon coefficient rearrangement: the sinograms of the same digit in different angles are distinct, yet their similarities are noticeable, as seen in Fig. 4a. The shapes are similar but the locations of the most intense spots are different. To solve this problem, the Radon transform coefficient could be rearranged to shift the position similar to that of lexicographical order problems. The coefficient rearrangement consists of two steps:

Step 1: search the position (c_x^{max}, c_y^{max}) of the maximum coefficient in the sinogram. The maximum coefficient appears in two cases: (1) a single maximum position (c_x) on the x -axis and (2) multiple maximum positions. In the first case, the

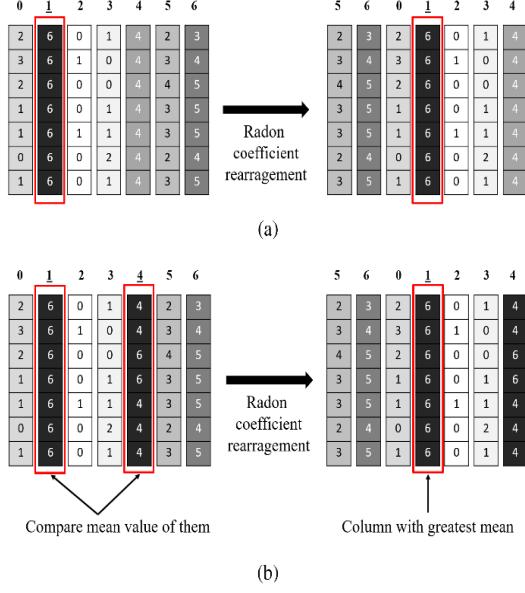


Fig. 5: The Radon coefficient rearrangement with (a) single maximum coefficient and (b) multiple maximum coefficients

parameter c_x^{max} is set to c_x corresponding to the position of the maximum coefficient, as shown in Fig. 5a. In the second case, the parameter c_x^{max} is determined from the c_x that the column vector returns the highest mean value as shown in Figs. 5b.

Step 2: shift the coefficient based on column vector by c_{shift} to rearrange the sinogram. The parameter c_{shift} is defined by (6) as follows.

$$c_{shift} = c_x^{center} - c_x^{max} \quad (5)$$

where c_x^{center} is the sinogram center on the x -axis. However, because a sinogram is essentially a sinusoid, a circular shift is demanded in this step. Thus, each new column vector, c_i' , becomes

$$c_i' = (c_i + c_{shift}) \% n_c \quad (6)$$

where c_i is the original individual column vector and n_c is the total number of columns.

The results of these rearranged sinograms, shown in Fig. 4c, are closely similar to the original sinograms in Fig. 4b. Hence, the sinograms are now rotation invariant.

III. EXPERIMENT AND DISCUSSION

This section shows the performance evaluation of the proposed MRT method for data augmentation in seven segment display recognition. The experimental setup and the results are discussed as follows.

A. Experimental Setup

The proposed data augmentation with MRT was tested against other types of augmentation on five different data

sets: normal, rotate, scale, perspective, and illumination variations. In this paper, LeNet architecture, one of well-known convolutional neural networks, is used. LeNet is the basic convolutional neural network architecture introduced by Lecun et al. in 1998 for character recognition [11]. Since LeNet does not need GPU support for computing, requires less computing time, and requires a small image to be an input, and thus suitable for the particular digit data in this study.

All digit image in this experiment was taken by Logitech HD Pro C920 webcam. The 2,000 normal digit images without inherent tilt and lighting conditions were divided into 1,600 images and 400 images for training and testing, respectively. There are four testing image categories- (i) rotate, (ii) scales, (iii) perspective, and (iv) illumination effect. In the rotate category, the image was taken by tilt the camera about -80° to 80° with 20° steps for 1,000 digit images. The scaled image was taken with a variety of distance between the camera and the seven-segment display from 25 to 150 meters with 25 steps for 1,000 digit images the scales category. The perspective category contained 1,000 digit image data which was skew tilt the seven segment display to left and right with 20° to 60° by 20° steps. For the illumination category, 500 digit images were taken in a dark environment with a diversity of light incidents. The image sizes ranged from 3×7 to 36×52 pixels.

The learning rate, batch size, and epoch are 0.02, 50, and 10, respectively. For MRT, the parameters of the Gaussian function including σ and ρ that were set to 0.2 and [2,7]. The performance of the proposed method was compared with the baseline augmentations including rotated images with θ range of -80 to 80 degrees, scaled images with the sizes of 0.5 to 4 pixels, perspective with θ of -80 to 80 degrees [12]. The accuracy rate, A_c , was calculated by Eq. 7 as follows

$$A_c = \frac{N_c}{N_c + N_i} \times 100 \quad (7)$$

where N_c and N_i is a correctly and an incorrectly predicted objects respectively.

B. Result and Discussion

As shown in Table I, the experimental setups are divided into two parts: training and testing. The first part of the table shows five categories of training sets. The all training sets, with and without any data augmentation, achieved the best accuracy of above 99% with normal images excepting rotation augmentation case, which is 92.5%.

Upon a closer look at the experimental results, the training set with data augmentation from the proposed MRT appeared to yield the best overall performance with best performance in two categories: normal and illumination. Additionally, the MRT did not suffer much in other categories unlike other training sets. Its performance was comparable to the best one in each test. Gaussian filter enable the training model to be able to tolerate the blooming effects whereas the Radon transform with rearranged coefficient make the model more robust to sizes and rotation. Since perspective variation is a combination of rotation and scaling, the proposed MRT also performed well in that category. MRT also appears to create a CNN that is strongly robust to illumination variation.

TABLE I: Performance comparison of different data augmentation on various test data

Training Set	Accuracy Rate (%) of Test Data					Mean	S.D.
	Normal	Rotate	Scale	Perspective	Illumination		
None	99.50	29.00	89.00	61.50	87.76	73.35	28.48
Rotation	92.50	83.30	84.50	67.20	78.98	81.30	9.27
Scaling	99.20	27.40	87.50	63.60	91.22	73.78	29.13
Perspective	99.50	33.90	93.90	85.50	93.88	81.34	26.98
MRT	99.75	74.30	90.80	72.40	98.00	87.05	12.97

IV. CONCLUSION

The data augmentation with multiscale Radon transform was proven to successfully achieve accurate recognition of the seven segment digits on the convolution neural network. The training set created from data augmentation of the proposed MRT method was able to generate a CNN model that is scale and rotation invariant. The proposed method is also robust to lighting conditions, which is a key factor in recognizing a self-luminous object as an LED seven segment digit. The study has demonstrated that data augmentation with the proposed multiscale Radon transform could improve the neural network training process for seven segment display. Researchers should be able to applied the proposed technique to other self-luminous type of images such as CAT scans or x-ray pictures. Future work should include the investigation on such experiments to provide more insights and explanations as well as the assumption of illumination.

REFERENCES

- [1] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *2017 Chinese Automation Congress (CAC)*, Oct. 2017, pp. 4165–4170.
- [2] L. Taylor and G. Nitschke, "Improving deep learning with generic data augmentation," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Nov 2018, pp. 1542–1547.
- [3] S. Popayorm, T. Titijaroomroj, T. Phoka, and W. Massagram, "Seven segment display detection and recognition using predefined hsv color slicing technique," in *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, July 2019, pp. 224–229.
- [4] R. P. Ghugardare, S. P. Narote, P. Mukherji, and P. M. Kulkarni, "Optical character recognition system for seven segment display images of measuring instruments," in *TENCON 2009 - 2009 IEEE Region 10 Conference*, Jan. 2009, pp. 1–6.
- [5] I. Bonačić, T. Herman, T. Krznar, E. Mangić, G. Molnar, and M. Cupic, "Optical Character Recognition of Seven – segment Display Digits Using Neural Networks," in *Proceedings of MIPRO 2009 - 32st International Convention on Information and Communication, Technology, Electronics and Microelectronics*, pp. 323–328, 2009.
- [6] P. H. Kulkarni and P. D. Kute, "Optical numeral recognition algorithm for seven segment display," in *2016 Conference on Advances in Signal Processing (CASP)*, Jun. 2016, pp. 397–401.
- [7] T. Jiang, Z. Cui, Z. Zhou, and Z. Cao, "Data augmentation with gabor filter in deep convolutional neural networks for sar target recognition," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 689–692, 2018.
- [8] Q. Dong, A. Wu, N. Dong, W. Feng, and S. Wu, "A Convolution Neural Network for Parts Recognition Using Data Augmentation," in *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, Jul. 2018, pp. 773–777.
- [9] S. A. G. Sánchez, C. A. Ruiz, and C. E. A. Meza, "Geometric transformations and its application in digital images," *Journal of Theoretical and Applied Information Technology*, vol. 60, pp. 150–153, Feb 2014.
- [10] J. Radon, "Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten," *Akad. Wiss.*, vol. 69, pp. 262–277, 1917.
- [11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
- [12] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," Dec. 2017.