

การศึกษารูปแบบมือจับและลักษณะรูปทรง
ของวัตถุในระบบอีโกเซ็นทริกวิชั่น

THE INVESTIGATION OF HAND GRASP TYPE AND OBJECT
ATTRIBUTES IN EGOCENTRIC VISION SYSTEM

ชนะ จันทร์ศรี

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ปีการศึกษา 2564
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

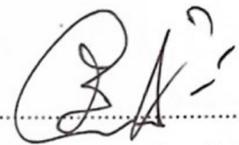
การศึกษารูปแบบมือจับและลักษณะรูปทรง
ของวัตถุในระบบอีโก้เซ็นทริกวิชั่น

ชนะ จันทร์ศรี

ดุขฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ปีการศึกษา 2564
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

วิทยานิพนธ์ฉบับนี้เป็นงานวิจัยที่เกิดจากการค้นคว้าและวิจัย ขณะที่ข้าพเจ้าศึกษาอยู่ในคณะ
วิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ดังนั้นงานวิจัยในวิทยานิพนธ์ฉบับนี้ถือเป็น
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี และข้อความต่าง ๆ ในวิทยานิพนธ์ฉบับนี้ ข้าพเจ้า
ขอรับรองว่าไม่มีการคัดลอกหรือนำงานวิจัยของผู้อื่นมานำเสนอในชื่อของข้าพเจ้า

This thesis consists of research materials conducted at the Faculty of
Engineering, Rajamangala University of Technology Thanyaburi and hence the copyright
owner. I hereby certify that the thesis does not contain any forms of plagiarism.



(นายชนะ จันท์ศรี)



หัวข้อวิทยานิพนธ์	การศึกษารูปแบบมือจับและลักษณะรูปทรงของวัตถุในระบบอีโก้เซ็นทริกวิชั่น
ชื่อ - นามสกุล	นายชนะ จันทร์ศรี
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์จักรี ศรีนนท์ฉัตร, Ph.D.
ปีการศึกษา	2564

บทคัดย่อ

การตรวจจับในระบบอีโก้เซ็นทริกวิชั่น (Egocentric Vision) เป็นหนึ่งในสาขาของคอมพิวเตอร์วิทัศน์ (Computer Vision) ที่มีการพัฒนาวิจัยกันอย่างแพร่หลาย เนื่องจากการพัฒนาเทคโนโลยีระบบกล้องถ่ายรูปและวิดีโอให้มีขนาดเล็ก ง่ายและสะดวกต่อการใช้งาน จึงมีการนำอุปกรณ์ดังกล่าวมาติดบริเวณศีรษะหรือหน้าอกเสมือนการบันทึกผ่านการมองเห็นของตัวเอง (First-Person Vision) เพื่อศึกษาการใช้มือจับวัตถุต่าง ๆ แต่เนื่องด้วยมือของมนุษย์มีการเปลี่ยนรูปร่างและภาพที่บันทึกผ่านกล้องมีมุมมองที่แตกต่างกัน ทำให้ยากต่อการตรวจจับท่ามือ อีกทั้งปริมาณความเข้มแสงสว่างในสภาพแวดล้อม ทำให้ประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุลดลง

งานวิจัยนี้จึงนำเสนอเทคนิคการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นทริกวิชั่นในสภาวะความเข้มแสงสว่างต่ำ โดยการทำงานร่วมกับเทคนิค Adaptive Gamma Correction with Weighting Distribution และความต้านทานแบบโค้งงอ (Flex Sensor) ในขั้นตอนการปรับปรุงภาพในระบบอีโก้เซ็นทริกวิชั่น ได้นำค่าเฉลี่ยแรงดันไฟฟ้าที่เปลี่ยนแปลงของความต้านทานแบบโค้งงอที่ติดตั้งไว้ที่นิ้วโป้งขณะจับวัตถุมากำหนดเป็นน้ำหนักพารามิเตอร์ในการปรับค่าแกมมาของภาพ เนื่องด้วยการจัดวางนิ้วโป้งที่มีลักษณะเฉพาะในแต่ละท่ามือจับวัตถุ ทำให้ได้ภาพที่มีคุณลักษณะเฉพาะหลังจากผ่านขั้นตอนการปรับปรุงภาพที่มีความเข้มแสงสว่างต่ำ ส่งต่อให้ในส่วนการตรวจจับและจดจำท่ามือจับวัตถุ ซึ่งใช้เทคนิคการเรียนรู้เชิงลึกแบบ YOLOv2

ผลการทดลองพบว่า เทคนิคที่นำเสนอสามารถเพิ่มประสิทธิภาพในการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นทริกวิชั่นได้เป็นอย่างดี โดยสามารถตรวจจับและจดจำท่ามือจับวัตถุตามลักษณะรูปทรงต่าง ๆ จำนวน 18 ท่ามือ ได้ความถูกต้องสูงสุดร้อยละ 86.28

คำสำคัญ: ระบบอีโก้เซ็นทริกวิชั่น ท่ามือจับวัตถุ ความต้านทานแบบโค้งงอ การเรียนรู้แบบเชิงลึก การเพิ่มประสิทธิภาพในสภาวะความเข้มแสงสว่างต่ำ

Dissertation Title	The Investigation of Hand Grasp Type and Object Attributes in Egocentric Vision System
Name – Surname	Mr.Chana Chansri
Program	Electrical Engineering
Dissertation Advisor	Assistant Professor Jakkree Srinonchat, Ph.D.
Academic Year	2021

ABSTRACT

Egocentric recognition is one of the computer vision fields which has been studied pervasively. Due to its development in photography system and a small size video that made it simple to utilize, the egocentric camera is attached to the head or the chest to record the first-person point of view. The gadget is used to study hand grasp type. However, as human hands can be transformed into shapes through different camera angles, detecting the hand grasp type is not easy. Moreover, an environmental light under low illumination conditions makes it difficult to detect the hand grasp type.

This study addressed the techniques which increased the efficiency of egocentric hand grasp type recognition in low light conditions by utilizing the flex sensor and image enhancement algorithm with adaptive gamma correction with weighting distribution. In the enhancement stage, the flex sensor was attached to the thumb for object manipulation. The average voltages were used to configure the weighting parameter to improve the gamma function in the image enhancement stage. As the thumb placement differed in positions of each grasp on the objects, distinct images were made after the enhancement of low light conditions. Moreover, the YOLOv2 of the deep learning technique was used for the grasp detection and recognition system.

The research results showed that the proposed method could significantly improve grasp detection and recognition in egocentric vision. It can detect 18 gestures according to different shapes of objects with the highest accuracy of about 86.28%.

Keywords: egocentric vision, hand grasp, flex sensor, image enhancement, deep learning

กิตติกรรมประกาศ

การจัดทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยการสนับสนุนเป็นอย่างดีจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.จักรี ศรีนนท์ฉัตร ที่ได้ช่วยให้คำแนะนำและติดตามการทำวิทยานิพนธ์ครั้งนี้อย่างใกล้ชิดตลอดมา นับตั้งแต่เริ่มต้นจนกระทั่งสำเร็จเรียบร้อยสมบูรณ์ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของท่านเป็นอย่างยิ่ง

ขอขอบพระคุณและมอบความดีทั้งหมดนี้ให้แก่ พ่อแม่ญาติพี่น้องและคณาจารย์ ที่ให้การสนับสนุนและประสิทธิ์ประสาทวิชาความรู้ รวมทั้งสมาชิกร่วมห้องปฏิบัติการ Signal Processing Research Laboratory (SPRL) คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีทุกคน ที่ช่วยให้คำแนะนำและเป็นกำลังใจที่ดีตลอดมา

สุดท้ายนี้ ผู้วิจัยหวังเป็นอย่างยิ่งว่างานชิ้นนี้จะเป็นประโยชน์สำหรับผู้สนใจ หากวิทยานิพนธ์ในครั้งนี้อาจตกบกพร่องหรือไม่สมบูรณ์ประการใด ผู้วิจัยขอกราบขออภัยมา ณ โอกาสนี้ด้วย

ชนะ จันทร์ศรี



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(3)
บทคัดย่อภาษาอังกฤษ	(4)
กิตติกรรมประกาศ	(5)
สารบัญ	(6)
สารบัญตาราง	(8)
สารบัญรูป	(9)
คำอธิบายสัญลักษณ์และคำย่อ	(12)
บทที่ 1 บทนำ	13
1.1 ความเป็นมาและความสำคัญของปัญหา	13
1.2 วัตถุประสงค์การวิจัย	14
1.3 ขอบเขตของการวิจัย	14
1.4 ขั้นตอนการวิจัย	14
1.5 ประโยชน์ที่คาดว่าจะได้รับ	15
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	16
2.1 ทำมือจับวัตถุ	16
2.2 ลักษณะรูปทรงของวัตถุ.....	19
2.3 ระบบกล้องไอ้โก้เซ็นทริควิชัน	21
2.4 ความต้านทานแบบโค้งงอ	23
2.5 เครื่องมือทางการประมวลผลภาพ.....	24
2.6 เทคนิคการเรียนรู้แบบเชิงลึก	27
2.7 งานวิจัยที่เกี่ยวข้อง	36
บทที่ 3 วิธีดำเนินการวิจัย	41
3.1 การออกแบบระบบประมวลผลภาพเบื้องต้น	41
3.2 วจรความต้านทานแบบโค้งงอ	47
3.3 ระบบเพิ่มประสิทธิภาพการตรวจจับและจดจำทำมือจับวัตถุ	57
3.4 การเก็บรวบรวมฐานข้อมูลภาพและสัญญาณวจรความต้านทานแบบโค้งงอ	58

สารบัญ (ต่อ)

	หน้า
3.5 การออกแบบอัลกอริทึม	64
3.6 วิธีการทดสอบงานวิจัย	66
บทที่ 4 ผลการทดสอบและการวิเคราะห์	70
4.1 ผลการทดสอบการประมวลผลภาพเบื้องต้น	70
4.2 ผลการทดสอบถุงมือเซนเซอร์แบบ 16 DoF.....	74
4.3 ผลการทดสอบตรวจจับและจดจำท่ามือจับวัตถุ	78
4.4 การวิเคราะห์ผลการทดสอบของอัลกอริทึมที่นำเสนอ	91
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	93
5.1 สรุปผลการวิจัย	93
5.2 ข้อเสนอแนะ	94
บรรณานุกรม	95
ภาคผนวก	103
ภาคผนวก ก วารสารทางวิชาการและบทความการประชุมวิชาการ	104
ประวัติผู้เขียน	110

สารบัญตาราง

	หน้า
ตารางที่ 2.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยถุงมือเซนเซอร์	20
ตารางที่ 3.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยนี้	54
ตารางที่ 3.2 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยเพิ่มประสิทธิภาพตรวจจับท่ามือ	59
ตารางที่ 3.3 พารามิเตอร์ของโครงข่ายประสาทคอนโวลูชันแบบเชิงลึกสำหรับฝึกการเรียนรู้	69
ตารางที่ 4.1 ผล Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN โมเดล VGG16	82
ตารางที่ 4.2 ผล Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN โมเดล ResNet50	83
ตารางที่ 4.3 ผล Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN โมเดล GoogLeNet	84
ตารางที่ 4.4 ผล Mean Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN ในทุกท่ามือจับวัตถุ	85
ตารางที่ 4.5 เปรียบเทียบอัตราความถูกต้องของผลการตรวจจับและจดจำท่ามือจับระหว่าง เทคนิคที่นำเสนอกับเทคนิคเดิม	90

สารบัญรูป

	หน้า
รูปที่ 2.1 ตัวอย่างท่ามือจับแบบกำลังแม่นยำ	17
รูปที่ 2.2 ท่ามือจับวัตถุจำนวน 18 ท่ามือที่ใช้ทดลองในระบบอิกโก้เซ็นทริกวิชั่น	18
รูปที่ 2.3 การจับวัตถุในตำแหน่ง a, b และ c ตามลักษณะรูปทรงของวัตถุ	19
รูปที่ 2.4 ตัวอย่างรูปทรงลักษณะตามขนาดมิติของวัตถุ	20
รูปที่ 2.5 ตัวอย่างกล่องอิกโก้เซ็นทริกวิชั่นที่มีขายในเชิงพาณิชย์	21
รูปที่ 2.6 ตัวอย่างการสวมใส่กล่องในระบบอิกโก้เซ็นทริกวิชั่น	22
รูปที่ 2.7 โครงสร้างส่วนประกอบของความต้านทานแบบโค้งงอ	23
รูปที่ 2.8 การโค้งตัวและมุมการโค้งตัวของความต้านทานแบบโค้งงอ	24
รูปที่ 2.9 ที่ 50% ของแสงที่เซนเซอร์ของกล่องดิจิตอลจับได้จากสีจริง	25
รูปที่ 2.10 กราฟพลอตความสัมพันธ์การแปลงความเข้มแสงสว่างที่ค่า γ ต่าง ๆ	27
รูปที่ 2.11 เปรียบเทียบการทำงานของกระบวนการ Machine Learning กับ Deep learning	28
รูปที่ 2.12 ลำดับการพัฒนา Artificial Intelligence, Machine Learning และ Deep learning	29
รูปที่ 2.13 การดำเนินการคอนโวลูชันระหว่างรูปภาพ (Sub-region) กับเคอร์เนล (Kernel)	30
รูปที่ 2.14 การดำเนินการคำนวณในชั้นพูลลิ่ง (Pooling Layer)	31
รูปที่ 2.15 โครงสร้างสถาปัตยกรรมภายในอัลกอริทึม YOLOv2	34
รูปที่ 2.16 พื้นที่การคำนวณหา Intersection over Union ที่ตรวจจับได้	35
รูปที่ 2.17 การทำงานของอัลกอริทึม YOLOv2 ในการตรวจจับวัตถุภายในภาพ	35
รูปที่ 3.1 ขั้นตอนการตรวจจับและจดจำท่าภาษามือด้วยกล่อง Microsoft Kinect	42
รูปที่ 3.2 ลำดับภาพที่ใช้ในการตรวจจับท่าทางมือด้วยกล่อง Microsoft Kinect	43
รูปที่ 3.3 ท่ามือสะกดนิ้วไทยที่เป็นท่ามือเดียวใช้ในการทดลอง	43
รูปที่ 3.4 ระยะเวลาการทำงานของกล่องถ่ายภาพความลึกทั้งสองโหมด	44
รูปที่ 3.5 ขั้นตอนการทำงานการระบบควบคุมอุปกรณ์ด้วยท่าทางมือ	45
รูปที่ 3.6 จุดตำแหน่งบนฝ่ามือที่ใช้ในการคำนวณท่ามือ	47
รูปที่ 3.7 วงจรมัลติเพล็กซ์สำหรับอ่านค่าสัญญาณจากความต้านทานแบบโค้งงอ	48
รูปที่ 3.8 ถู่มือที่ติดตั้งความต้านทานแบบโค้งงอแบบ 16 DoF	49
รูปที่ 3.9 รูปจากการจัดเรียงเมทริกซ์ข้อมูลขนาด 16x1200	50

สารบัญรูป (ต่อ)

	หน้า
รูปที่ 3.10 สัญญาณและรูปสี่ขนาด 227x227 พิกเซล ที่ได้แปลงภาพด้วยวิธี Color mapping	51
รูปที่ 3.11 สัญญาณและรูปสี่ขนาด 227x227 พิกเซล ที่ได้จากการแปลงภาพด้วยเทคนิค GAF.....	53
รูปที่ 3.12 ขั้นตอนการทำงานระบบตรวจจับและจดจำท่ามือจับวัตถุในสภาวะความเข้มแสงสว่างต่ำ	58
รูปที่ 3.13 ภาพตัวอย่างบันทึกขณะใช้มือจับวัตถุพร้อมความต้านทานแบบโค้งงอ	59
รูปที่ 3.14 ตัวอย่างสัญญาณวงจรความต้านทานแบบโค้งงอขณะใช้มือจับวัตถุ	63
รูปที่ 3.15 สัญญาณเอาต์พุตที่ได้จากเซนเซอร์ที่ติดตั้งไว้ขณะจับวัตถุในท่ามือที่แตกต่างกัน	64
รูปที่ 3.16 ไดอะแกรมการทำงานของอัลกอริทึมที่นำเสนอปรับปรุงภาพ	65
รูปที่ 3.17 ตัวอย่างภาพที่ได้ปรับความเข้มสว่างให้ลดต่ำลงในแต่ละรูปแบบ	67
รูปที่ 3.18 ตัวอย่างภาพแล Histogram ที่ผ่านการปรับความสว่างด้วยเทคนิค AGCWD	68
รูปที่ 4.1 ผลการจดจำท่าทางมือของแต่ละตัวอักษรจำนวนทั้งหมด 42 ตัว	71
รูปที่ 4.2 ผลการทดลองตรวจจับท่าภาษามือของการสะกดนิ้วไทยในแต่ละระยะทางตั้งแต่ 0.6 ถึง 2.8 เมตร	72
รูปที่ 4.3 ผลการทดลอง Confuse Matrix ของจดจำท่าสะกดนิ้วมืออเมริกาจำนวน 12 ท่ามือ	73
รูปที่ 4.4 ผลการทดสอบถุงมือเซนเซอร์แบบ 16 DoF ด้วยวิธีการ Nearest interpolation	74
รูปที่ 4.5 กราฟแสดงความถูกต้องในการจดจำท่ามือจับวัตถุ 20 ท่ามือ	75
รูปที่ 4.6 กราฟแสดงความถูกต้องในการใช้ถุงมือเซนเซอร์จดจำท่ามือจับวัตถุ 20 ท่ามือ	76
รูปที่ 4.7 กราฟแสดงความถูกต้องในการใช้ถุงมือเซนเซอร์จดจำท่าสัญลักษณ์ภาษามือจำนวน 12 ท่ามือ	77
รูปที่ 4.8 กราฟแสดงความถูกต้องในการใช้ถุงมือเซนเซอร์จดจำท่ามือตามลักษณะรูปทรง	78
รูปที่ 4.9 ตัวอย่างภาพที่ตรวจจับและผลการทำนายท่ามือจับวัตถุที่ค่า IoU เท่ากับ 0.5	81
รูปที่ 4.10 Precision-recall curve of VGG 16 ที่สภาวะระดับความสว่างต่าง ๆ	85
รูปที่ 4.11 เปรียบเทียบผลการจดจำท่ามือจับขณะปรับค่าแกมมา	86
รูปที่ 4.12 เปรียบเทียบผลการจดจำท่ามือจับขณะปรับค่าความคมชัด	87
รูปที่ 4.13 ค่า False Positive ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับค่า IoU	88

สารบัญญรูป (ต่อ)

	หน้า
รูปที่ 4.14 ผล Overlap Matching ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับค่า IoU	88
รูปที่ 4.15 ค่า Miss Rate ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับเปลี่ยนค่า IoU	89



คำอธิบายสัญลักษณ์และคำย่อ

AGCWD	Adaptive Gamma Correction with Weighting Distribution
AP	Average Precision
AUC	Area Under the Curve
CH	Center of Hand
CNN	Convolutional Neural Network
CW	Center of Wrist
DCNN	Deep Convolutional Neural Network
DoF	Degree of Freedom
GAF	Gramian Angular Fields
GF	Grasp Factor
GPIO	General Purpose Input Output
GPU	Graphics Processing Unit
HOG	Histograms of Oriented Gradients
IoU	Intersection over Union
mAP	mean Average Precision
PM	Proposed Method
RFA	Radian Finger Analysis
RGB	Red Green Blue
ROI	Region of Interest
TM	Traditional Method
WD	Weighting Distribution
YCbCr	Luminance Chroma blue Chroma red
YOLO	You Only Look Once
YOLOv2	You Only Look Once Version2

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันงานวิจัยด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) ได้รับความสนใจและมีการพัฒนากันอย่างต่อเนื่อง โดยเฉพาะการตรวจจับในระบบอีโก้เซ็นทริกวิชั่น (Egocentric Vision) ซึ่งเป็นอีกหนึ่งสาขาที่มีการพัฒนาวิจัยกันอย่างแพร่หลาย เนื่องจากการพัฒนาเทคโนโลยีระบบกล้องถ่ายรูปและวิดีโอให้มีขนาดเล็ก ง่ายและสะดวกต่อการใช้งาน จึงมีการนำอุปกรณ์ดังกล่าวมาติดตามตัว บริเวณศีรษะหรือหน้าอกเสมือนการบันทึกผ่านการมองเห็นของตนเอง โดยมีหลายหลากรุ่น เช่น GoPro, Microsoft Hololens, Google Glass เป็นต้น เพื่อศึกษาการจับวัตถุต่าง ๆ ในการดำเนินชีวิตประจำวันเพื่อนำไปประยุกต์ใช้กับเทคโนโลยีหุ่นยนต์และเทคโนโลยีโลกเสมือนในอนาคต โดยข้อดีของระบบอีโก้เซ็นทริกวิชั่นเมื่อเทียบกับการที่ถ่ายภาพทั่วไป มีหลากหลายประการ เช่น การบันทึกภาพเป็นสิ่งที่อยู่ตรงหน้าเป็นตำแหน่งคงที่เสมอ กล้องเคลื่อนที่ที่เกิดขึ้นจากกิจกรรมและความสนใจของผู้สวมใส่ มือและวัตถุที่ถูกจัดการมักจะปรากฏที่ศูนย์กลางของภาพ การบิดเบี้ยวของภาพเมื่อลดลง เป็นต้น การตรวจจับและจดจำท่ามือจับวัตถุของมนุษย์เป็นงานที่ค่อนข้างยากมีข้อจำกัด เนื่องจากมือมีการเปลี่ยนแปลงรูปร่างภายใต้มุมมองของปริภูมิและช่วงเวลาที่แตกต่างกันไป ตัวแปรสำคัญอีกสิ่งหนึ่งคือ ปริมาณความเข้มแสงสว่างในสิ่งแวดล้อมที่ต่ำส่งผลทำให้ประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุที่ลดลง ทำให้ระบบการตรวจจับและจดจำไม่สามารถจำแนกท่ามือจับที่มีการจัดเรียงนิ้วมือลักษณะใกล้เคียงกันได้

ในงานวิจัยนี้จึงนำเสนอเทคนิคที่สามารถเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นทริกวิชั่นในสภาวะความเข้มแสงสว่างต่ำ โดยการทำงานร่วมกันระหว่างเทคนิคการปรับปรุงภาพกับความต้านทานแบบโค้งงอ (Flex Sensor) โดยใช้การตรวจจับและจดจำด้วยเทคนิคการเรียนรู้แบบเชิงลึก (Deep Learning) เนื่องจากการใช้มือจับวัตถุของมนุษย์มีการจัดวางนิ้วโป้งที่มีลักษณะเฉพาะในแต่ละท่ามือจับวัตถุ จึงนำคุณสมบัติดังกล่าวใช้เป็นพารามิเตอร์ในการกำหนดน้ำหนักการปรับปรุงภาพ โดยการคำนวณค่าเฉลี่ยแรงดันไฟฟ้าที่เปลี่ยนแปลงของความต้านทานแบบโค้งงอที่ติดบนนิ้วโป้งขณะจับวัตถุ ทำให้มีคุณลักษณะเฉพาะในแต่ละภาพของท่ามือจับวัตถุและช่วยเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุของเทคนิคการเรียนรู้แบบเชิงลึก

1.2 วัตถุประสงค์การวิจัย

ศึกษาการตรวจจับและจดจำท่ามือจับวัตถุแบบต่าง ๆ ตามลักษณะรูปร่างของวัตถุในระบบอีโก้เซ็นทริกวิชั่นที่ใช้ในชีวิตประจำวัน โดยใช้เทคนิคการประมวลผลภาพร่วมกับการเรียนรู้เชิงลึก เพื่อพัฒนาเทคนิคใหม่ในการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุที่อยู่ในสภาพแวดล้อมที่มีความเข้มแสงสว่างต่ำ

1.3 ขอบเขตของการวิจัย

1.3.1 ใช้เทคนิคการประมวลผลภาพควบคู่กับเทคนิคการเรียนรู้เชิงลึก เพื่อใช้ในการตรวจจับและจดจำท่ามือจับวัตถุ

1.3.2 สามารถตรวจจับท่ามือจับวัตถุในรูปแบบต่าง ๆ ตามลักษณะรูปร่างของวัตถุได้อย่างน้อย 15 ท่ามือจับวัตถุ

1.3.3 สามารถจดจำท่ามือจับวัตถุตามลักษณะรูปร่างต่าง ๆ ได้อย่างน้อยร้อยละ 60

1.4 ขั้นตอนการวิจัย

1.4.1 ศึกษาการใช้มือจับวัตถุรูปแบบต่าง ๆ ในระบบอีโก้เซ็นทริกวิชั่น เครื่องมือทางการประมวลผลภาพ หลักการทำงานของความต้านทานแบบโค้งงอ เทคนิคการเรียนรู้แบบเชิงลึกและรวบรวมวรรณกรรมที่เกี่ยวข้อง

1.4.2 ออกแบบวงจรความต้านทานแบบโค้งงอและการเก็บรวบรวมสัญญาณไฟฟ้าจากการใช้มือจับวัตถุรูปแบบต่าง ๆ

1.4.3 ออกแบบระบบการเก็บรวบรวมฐานข้อมูลภาพมือจับวัตถุ การประมวลผลภาพและโปรแกรมการประมวลผลการเรียนรู้แบบเชิงลึก

1.4.4 วิเคราะห์ผลการทดลองเบื้องต้น

1.4.5 สร้างอัลกอริทึมในการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุ

1.4.6 ทดสอบและปรับปรุงอัลกอริทึมในการเพิ่มประสิทธิภาพการทำงาน

1.4.7 วิเคราะห์และสรุปผลการทดสอบ

1.4.7 จัดทำรูปเล่มดุษฎีนิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 เพื่อเป็นข้อมูลในการศึกษาทำมือจับวัตถุในระบบอีโก้เซ็นทริกวิชั่น
- 1.5.2 เพื่อเป็นแนวทางในการออกแบบและการใช้งานความต้านทานแบบโค้งงอ
- 1.5.3 เพื่อเป็นแนวทางในการออกแบบและการใช้งานเทคนิคการเรียนรู้แบบเชิงลึก
- 1.5.4 เพื่อเป็นแนวทางในการพัฒนาอัลกอริทึมสำหรับเพิ่มประสิทธิภาพการตรวจจับและจดจำทำมือจับวัตถุในสภาวะความเข้มแสงสว่างต่ำ
- 1.5.5 เพื่อพัฒนาเทคโนโลยีไปสู่การใช้งานในระบบการมองเห็นแบบเสมือนจริงและระบบการจับของมือหุ่นยนต์ในอนาคต



บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงหลักการและทฤษฎีที่ใช้ในการทำดัชนีนิพนธ์ โดยเนื้อหาจะประกอบด้วย ทำมือจับวัตถุที่ใช้ในชีวิตประจำวัน ลักษณะรูปทรงของวัตถุ ระบบกลองอีโก้เซ็นทริกวิชั่น ความต้านทานแบบโค้งงอ เครื่องมือทางการประมวลผลภาพ เทคนิคการเรียนรู้แบบเชิงลึกและในหัวข้อสุดท้ายจะเป็นงานวิจัยที่เกี่ยวข้อง

2.1 ทำมือจับวัตถุ

มือเป็นอวัยวะของมนุษย์ที่มีความสำคัญอย่างมากในการดำรงชีวิต ซึ่งทำให้เราสามารถโต้ตอบกับสิ่งของและสิ่งแวดล้อม สื่อสารกับผู้อื่น และดำเนินกิจกรรมในชีวิตประจำวัน เช่น การกิน การอาบน้ำ และการแต่งตัว เป็นต้น การศึกษารูปร่างและข้อจำกัดที่เกี่ยวข้องในการจับวัตถุแต่ละครั้ง และการรู้รูปแบบการใช้งานทั่วไปเป็นสิ่งสำคัญ ซึ่งสามารถมาใช้ในการพัฒนาหลายด้านตั้งแต่การแพทย์ ระบบกลไกหุ่นยนต์ ระบบโลกเสมือนจริง การออกแบบผลิตภัณฑ์และอื่นอีกมากมาย มนุษย์ปฏิสัมพันธ์กับคอมพิวเตอร์ในส่วนติดต่อประสานกับผู้ใช้ทั่วไป โดยเฉพาะการใช้มือในการโต้ตอบกับเทคโนโลยีที่ใช้ในชีวิตประจำวัน ในเรื่องการออกแบบระบบปฏิสัมพันธ์ ไม่ใช่แค่สิ่งสำคัญที่ต้องเข้าใจท่าทางจับของมือระหว่างการใช้งานปกติ แต่ต้องรู้ว่าท่าทางมีอนั้นเป็นอย่างไรเพื่อปรับตามความต้องการของงานด้วย เนื่องจากความซับซ้อนและความหลากหลายของการใช้มือของมนุษย์ทำให้การจัดหมวดหมู่และการจำแนกประเภทของท่ามือที่มีความท้าทาย ซึ่งมีมนุษย์มือมีทั้งหมด 15 ข้อมนิ้ว ส่งผลให้มีจุดหมุนอิสระมากกว่า 20 ระดับ เพราะเหตุนี้การสร้างแบบจำลองรูปร่างมือโดยตรงนั้นยากและต้องมีการระบุพารามิเตอร์จำนวนมาก อย่างไรก็ตาม การรวมกันของวิธีที่มีมือโต้ตอบกับวัตถุที่จับนั้นมักจะถูกจำกัดและอาจแบ่งออกเป็นคลาสย่อย ที่เรียกว่า ประเภทของการจับซึ่งมักใช้เพื่ออธิบายการใช้มือและการเสนอการจำแนกประเภทการจับที่แตกต่างกันมากมายในหลายสาขาวิชาการ ตั้งแต่วิทยาการหุ่นยนต์ การแพทย์ และชีวกลศาสตร์ เป็นต้น โดยการเคลื่อนไหวและการทำงานของมือไม่ได้เป็นเพียงของระดับจุดหมุนอิสระภายในของมือเท่านั้น แต่ยังรวมถึงการเคลื่อนไหวของร่างกายและแขนตลอดจนการสัมผัสกับสิ่งแวดล้อมรอบๆ การเคลื่อนไหวของมือมนุษย์ที่พยายามจับวัตถุนั้น มีการจำแนกประเภทฟังก์ชันของมือในการจัดการยึดเหนี่ยวกับวัตถุอยู่ในความสัมพันธ์คงที่กับมือ โดยไม่มีการเคลื่อนไหวของมือ การจับแต่ละครั้งสามารถจำแนกได้ตามความต้องการของความแม่นยำหรือกำลังของผู้จับวัตถุ [1] [2] ในการจับแบบกำลัง (Power) มีความสัมพันธ์ที่แนบแน่นระหว่างวัตถุกับมือ ซึ่งหมายความว่า การเคลื่อนที่

ทั้งหมดของการจัดการวัตถุด้วยมือต้องถูกกระตุ้นส่งเสริมด้วยแขนของผู้จับ ส่วนสำหรับการจัดการแบบแม่นยำ (Precision) มือที่ใช้จับวัตถุสามารถจัดการวัตถุโดยไม่ต้องขยับแขนเพื่อส่งแรงออกไป สิ่งเหล่านี้เป็นความเข้าใจที่วัตถุอยู่ในความสัมพันธ์คงที่กับมือที่ใช้จับวัตถุ โดยวัตถุแต่ละแบบจะมีการจัดการวัตถุด้วยการจับแบบความแม่นยำหรือกำลังก็ขึ้นอยู่กับรูปทรงและการนำไปใช้งานของวัตถุนั้น ดังตัวอย่างการจับวัตถุในรูปที่ 2.1

	Prismatic			Round	Flat
Power	 Large Wrap	 Large Wrap	 Index Finger Extension	 Power Sphere	 Extension Type
Precision	 Writing Tripod	 Thumb-n Finger	 Precision Sphere	 Lateral Pinch	

รูปที่ 2.1 ตัวอย่างท่ามือจับแบบกำลังและแม่นยำ [2]

การศึกษาจับวัตถุด้วยมือเป็นสิ่งจำเป็นสำหรับเทคโนโลยีการสร้างมือจับหุ่นยนต์และการสร้างเทคโนโลยีโลกเสมือนจริงที่สามารถโต้ตอบและปรับเปลี่ยนความเหมาะสมได้ตามสภาพแวดล้อมของมันได้ การศึกษาเกี่ยวกับการจับวัตถุของมนุษย์มีเป็นจำนวนมากและรวมทั้งการวิจัยเกี่ยวกับหุ่นยนต์โดยทำงานภายใต้การตรวจจับความเคลื่อนไหวของมือ ในการศึกษาเหล่านี้มีจุดสนใจ คือรูปร่างมือและหน้าสัมผัสระหว่างมือกับวัตถุ ซึ่งศึกษาเกี่ยวข้องกับพฤติกรรมของมือและวัตถุสามารถโต้ตอบกันได้อย่างไร การเคลื่อนไหวและฟังก์ชันของมือไม่เพียงแต่เป็นผลมาจากองศาอิสระของมือ แต่ยังรวมถึงการเคลื่อนไหวของร่างกายและแขนเช่นเดียว การสัมผัสนั้นสร้างความเข้าใจเกี่ยวกับการเคลื่อนไหวของมือมนุษย์ที่พยายามจับองค์ประกอบของวัตถุเหล่านั้น ภายในหมวดหมู่ที่กว้างของการจับวัตถุ การใช้มือจับวัตถุของมนุษย์เป็นการใช้นิ้วมือและฝ่ามือสัมผัสกับตำแหน่งต่าง

ๆ ของวัตถุเพื่อให้สามารถจัดการกับวัตถุนั้นได้ ซึ่งการจัดวางตำแหน่งนิ้วมือต้องขึ้นอยู่กับลักษณะรูปทรงของวัตถุนั้นด้วย การแบ่งประเภทของการใช้มือจับเป็นสิ่งสำคัญสำหรับการทำความเข้าใจการจัดการวัตถุด้วยมือ เนื่องจากมีลักษณะวิธีการใช้นิ้วมือและฝ่ามือจับวัตถุแตกต่างกันไปในระหว่างการจัดการกับวัตถุ โดยการใช้มือจับแต่ละแบบอยู่ที่จัดการวัตถุนั้นด้วยท่ามือแบบความแม่นยำหรือกำลังก็ขึ้นอยู่กับ

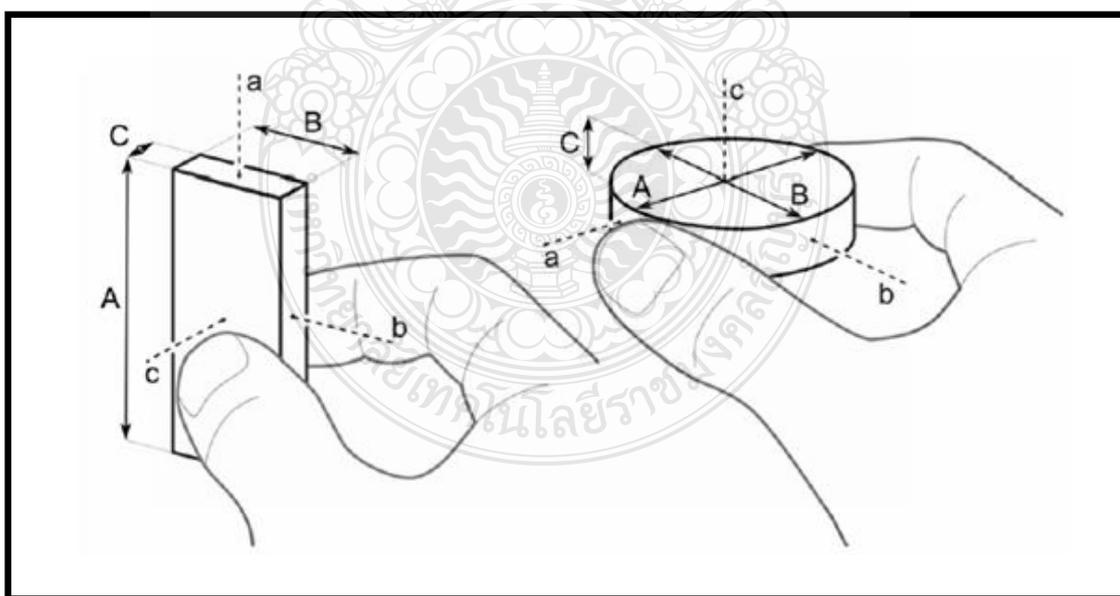
รูปทรงและการนำไปใช้งานของวัตถุนั้น งานวิจัยนี้ได้นำรูปแบบท่ามือจับวัตถุที่ใช้ในชีวิตประจำวัน [3] [4] มาใช้ตรวจสอบและจดจำลักษณะท่ามือจับในระบบอีโก้เซ็นทริกวิชั่นโดยมีรูปแบบท่ามือจับจำนวน 18 ท่ามือ ดังในรูปที่ 2.2



รูปที่ 2.2 ท่ามือจับวัตถุจำนวน 18 ท่ามือ ที่ใช้ทดลองในระบบอีโก้เซ็นทริกวิชั่น

2.2 ลักษณะรูปทรงของวัตถุ

รูปทรงของวัตถุที่ถูกจับหรือจัดการด้วยมือนั้น การทำความเข้าใจกระบวนการดังกล่าวมีความสำคัญมากในการศึกษา เนื่องจากมันบ่งบอกถึงการเคลื่อนไหวของมือที่เป็นไปได้ในการโต้ตอบระหว่างมือกับวัตถุที่จะจับ โดยวัตถุมักจะถูกจัดการได้หลายวิธี วิธีจัดการแต่ละวิธีก็อาจแตกต่างกันตามขนาด น้ำหนักและวัตถุประสงค์การจับ วัตถุสามารถมีตำแหน่งที่จับได้หลายตำแหน่งและมนุษย์จะเลือกตำแหน่งที่ยึดตามการใช้งานและตัวแปรอื่นๆ ตัวอย่างเช่น กรรไกรแก้วมีหูจับ มีตำแหน่งจับที่แตกต่างกันอย่างน้อย 3 ตำแหน่ง คือ ด้านข้าง (รูปทรงกระบอก), ด้านบน (รูปร่างดิสก์) และด้ามจับ (รูปทรงโค้งมน) ถึงแม้ว่ามันอาจจะดูแปลกที่จะแยกวัตถุโดยการจับจากตำแหน่ง โดยวัตถุส่วนใหญ่มีชุดเรขาคณิตที่มักเกี่ยวข้องกับการจับของมือ โดยเฉพาะอย่างยิ่งสำหรับวัตถุที่ออกแบบมาเพื่อปฏิสัมพันธ์ของมนุษย์ ดังนั้นจะมองเฉพาะที่เรขาคณิตของวัตถุโดยรวมและส่วนเรขาคณิตของวัตถุสำหรับพฤติกรรมกรจับของมนุษย์ [5] นอกจากนี้การจับวัตถุของมนุษย์จะคำนึงถึงน้ำหนัก รูปร่างและขนาดที่จะนำมาพิจารณาในการจับ จากที่กล่าวมาแล้วนั้น วัตถุหนึ่งชิ้นสามารถมีตำแหน่งที่จับได้ตามธรรมชาติหลายตำแหน่งขึ้นอยู่กับวัตถุประสงค์ของงานและตัวแปรอื่นๆ มนุษย์เลือกหนึ่งตำแหน่งจับเพื่อจัดการกับวัตถุตัวแปรเดียวในระดับวัตถุคือ น้ำหนักของวัตถุ ส่วนตัวแปรอื่นๆ ขึ้นอยู่กับแต่ละตำแหน่งการจับ เมื่อพิจารณาการจับวัตถุตามลักษณะรูปทรงดังรูปที่ 2.3

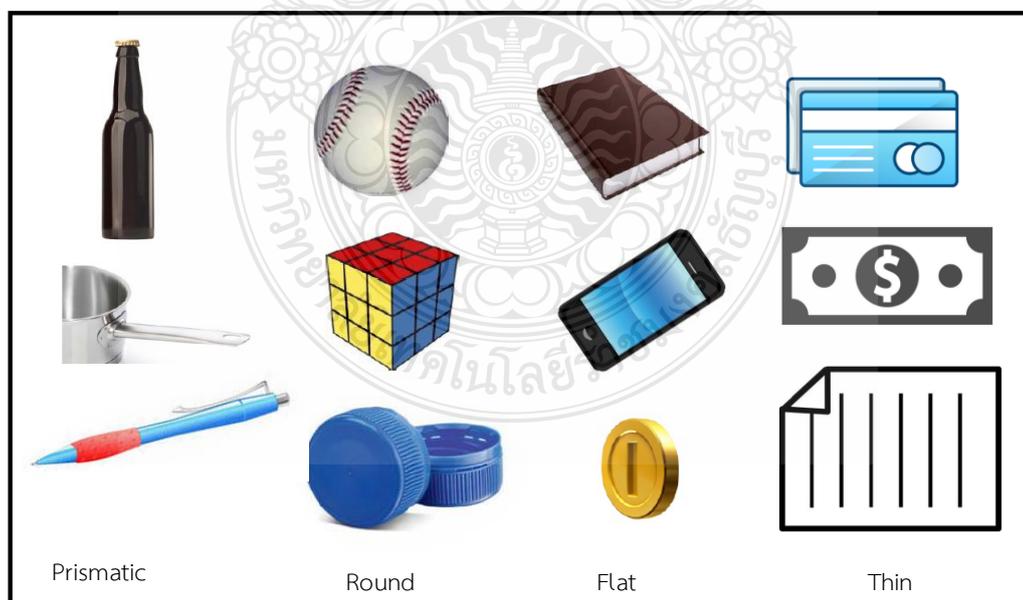


รูปที่ 2.3 การจับวัตถุในตำแหน่ง a, b และ c ตามลักษณะรูปทรงของวัตถุ [6]

ตามรูปที่ 2.3 แกนของวัตถุแทนด้วย a , b และ c โดยที่ a เป็นขนาดของมิติวัตถุที่ยาวที่สุด และ c จะอยู่ตามแนวที่สั้นที่สุด ($a \geq b \geq c$) ซึ่งเป็นความยาวของวัตถุในสามมิติ โดยสามารถแบ่งลักษณะรูปทรงของวัตถุที่แตกต่างกันเป็นจำนวน 4 รูปทรง คือ Prismatic, Round, Flat และ Thin ดังตารางที่ 2.1 และรูปที่ 2.4 ซึ่งวัตถุหนึ่งชิ้นอาจมีได้หลายแบบของรูปทรงเรขาคณิตและหลายวิธีในการจัดการขึ้นกับวัตถุประสงค์ในการใช้งานวัตถุนั้น ตัวอย่างเช่น การจับวัตถุรูปร่างของขวดที่มีความยาวและรูปร่างหน้าบ่งบอกถึงการจับเคลื่อนที่โดยการถือ แต่ในขณะที่ฝาขวดที่มีขนาดเล็กและกลมการจับเคลื่อนที่โดยการบิดหมุน ในขณะที่วัตถุสามารถประเมินได้โดยคุณสมบัติที่หลากหลาย เช่น รูปร่าง, น้ำหนัก, ความเรียบของพื้นผิว เป็นต้น ซึ่งงานวิจัยนี้เน้นเฉพาะลักษณะที่เกี่ยวข้องกับการจับเท่านั้น

ตารางที่ 2.1 การจำแนกลักษณะรูปทรงของวัตถุที่ใช้มือจับ [2]

ลำดับ	ลักษณะรูปทรง	ขนาดมิติของวัตถุ
1	Prismatic	$a > 2b$
2	Round	$b \leq a < 2b, c \leq a < 2c$
3	Flat	$b > 2c$
4	Thin	$b \gg c$



รูปที่ 2.4 ตัวอย่างลักษณะรูปทรงตามขนาดมิติของวัตถุ

กำหนดให้ด้านที่จับได้เป็นส่วนหนึ่งของวัตถุที่อยู่ระหว่างนิ้วมือเมื่อจับ โดยอ้างอิงแกนวัตถุที่กำหนดไว้ก่อนหน้านี้ในตำแหน่ง a, b และ c เพื่อพิจารณาว่าแกนใดสามารถกำหนดได้ดีที่สุดของการเปิดมือสำหรับจับดังรูปที่ 2.3 แสดงให้เห็นตัวอย่างการจับวัตถุโดยรูปทางซ้ายวัตถุถูกจับในตำแหน่งที่สั้นที่สุด ดังนั้นด้านที่จับได้สำหรับวัตถุจำนวนโดยมากจะมีมากกว่าหนึ่งด้านที่กำหนดการเปิดมือในการจับวัตถุ อย่างไรก็ตาม งานวิจัยนี้ใช้วัตถุที่สามารถจัดการได้ง่ายในการจับด้วยมือเพื่อใช้ทดสอบการตรวจจับและรู้จำประเภทท่ามือจับที่ปรากฏในระบบอีโก้เซ็นทริกวิชั่น

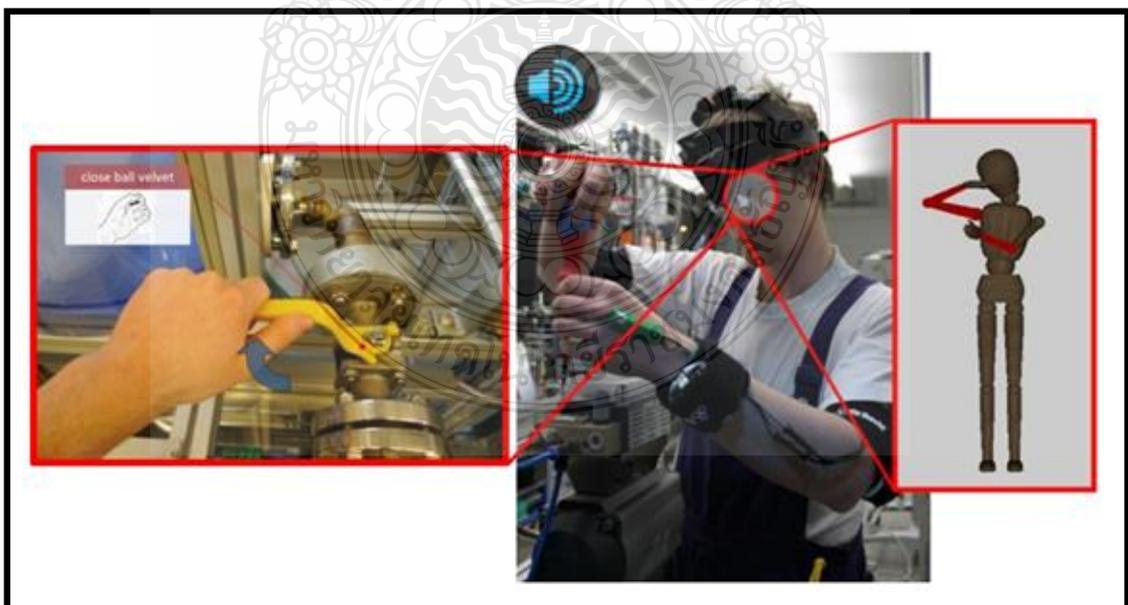
2.3 ระบบกล้องอีโก้เซ็นทริกวิชั่น

ระบบอีโก้เซ็นทริกวิชั่นหรือสามารถเรียกอีกอย่างว่า First-Person Vision (FPV) [7] [8] เป็นสาขาหนึ่งของคอมพิวเตอร์วิชั่น ซึ่งเกี่ยวข้องกับการวิเคราะห์ภาพและวิดีโอที่ถ่ายโดยกล้องที่สวมใส่ได้ ซึ่งโดยทั่วไปแล้วจะสวมใส่ที่ศีรษะหรือบนหน้าอก โดยให้ภาพที่ใกล้เคียงกับระยะการมองเห็นของผู้สวมใส่กล้อง ดังนั้นข้อมูลภาพจะบันทึกส่วนของฉากที่ผู้สวมใส่เน้นเพื่อทำงานด้วยมือและเสนอมุมมองเพื่อทำความเข้าใจกิจกรรมของผู้ใช้และบริบทของพวกเขาในสภาพแวดล้อมทั่วไป กล้องที่สวมใส่จะมองไปข้างหน้า ซึ่งมักจะสัมพันธ์กับภาพที่มองด้วยดวงตาของผู้ใช้ ซึ่งเป็นประโยชน์ให้เข้าใจถึงกิจกรรมและความตั้งใจของผู้ใช้ได้ดีขึ้น โดยเริ่มมีงานวิจัยเกี่ยวกับระบบอีโก้เซ็นทริกวิชั่นที่ตีพิมพ์ตั้งแต่ปี ค.ศ 1997 [24] แต่ได้รับความนิยมมากขึ้นหลังจากปี ค.ศ. 2012 ด้วยการเกิดขึ้นของแว่นตาอัจฉริยะและกล้องแอคชั่นแคม เช่น Google Glass and GoPro cameras เป็นต้น ดังตัวอย่างในรูปที่ 2.5



รูปที่ 2.5 ตัวอย่างกล้องอีโก้เซ็นทริกวิชั่นที่มีขายในเชิงพาณิชย์ [9]

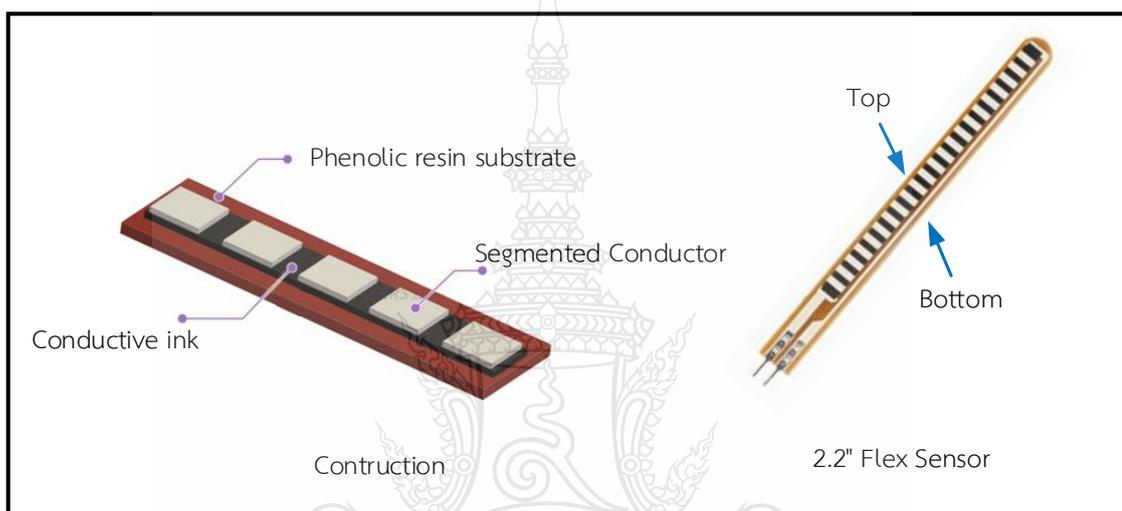
ภาพที่บันทึกด้วยกล้องในระบบอีโก้เซ็นทริกวิชั่นมีข้อดีหลายประการเมื่อเปรียบเทียบกับภาพที่บันทึกด้วยกล้องในการถ่ายทั่วไป โดยปกติแล้วตำแหน่งกล้องจะอยู่คงที่และไม่ขึ้นต่อผู้สวมใส่ ตัวอย่างเช่น สิ่งบันทึกภาพคือสิ่งที่อยู่ต่อหน้าผู้สวมใส่ กล้องการเคลื่อนไหวเกิดขึ้นจากกิจกรรมและความสนใจของผู้สวมใส่กล้องและวัตถุที่ถูกมือจัดการมักจะปรากฏที่ศูนย์กลางของภาพ ซึ่งลดการบิดเบี้ยวของภาพมือ อย่างไรก็ตาม เมื่อทำงานใน FPV นักวิจัยต้องยังเจอปัญหาสำคัญคือกล้องไม่นิ่งเนื่องจากเคลื่อนที่ไปกับตัวผู้สวมใส่ ทำให้เคลื่อนไหวเร็วและกะทันหันการเปลี่ยนแปลงความเข้มแสงสว่าง ซึ่งจะลดคุณภาพของภาพได้อย่างมาก ทำให้ยากต่อการแยกแยะมือและวัตถุที่สนใจจากพื้นหลัง (Background) เนื่องด้วยความก้าวหน้าของเทคโนโลยีกล้องที่สวมใส่ได้ในช่วงไม่กี่ปีที่ผ่านมา ทำให้กล้องชนิดนี้ได้รับความนิยมมากขึ้นโดยตัวอย่างกล้องที่มีขายในเชิงพาณิชย์ดังในรูปที่ 2.5 อุปกรณ์กล้องสวมใส่ที่มีอยู่ในปัจจุบันมีขนาดเล็กและน้ำหนักเบา เหล่านี้ช่วยให้สามารถบันทึกภาพที่มีคุณภาพสูงในแบบแฮนด์ฟรีได้ กล้องสวมใส่ได้แบบ GoPro และ Looxcie ซึ่งมีอัตราเฟรมค่อนข้างสูงตั้งแต่ 25 ถึง 60 fps ส่วนใหญ่จะใช้สำหรับบันทึกกิจกรรมผู้ใช้ได้ไม่กี่ชั่วโมง แต่ถ้าเป็นกล้องสวมใส่ได้แบบ Narrative Clip และ SenseCam ที่จับภาพเพียง 2 หรือ 3 fpm ซึ่งส่วนใหญ่จะใช้สำหรับการบันทึกภาพในเวลาที่น่าสนใจหรือทั้งวัน โดยปกติในแต่ละวันแล้วจะมีการบันทึกภาพถึง 2,000 ภาพ นั่นคือมากถึง 60,000 ต่อเดือน ซึ่งการเปลี่ยนแปลงความเข้มแสงสว่างในแต่ละช่วงเวลาจะลดคุณภาพของภาพได้ ทำให้ยากต่อการจำแนกมือและวัตถุที่สนใจจากพื้นหลัง ตัวอย่างการสวมใส่กล้องในระบบอีโก้เซ็นทริกวิชั่นเพื่อนำภาพจากกล้องที่ติดกับตัวไปใช้งานวิเคราะห์ปัญหาเหมือนการมองเห็นเช่นเดียวกับผู้สวมใส่ดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างการสวมใส่กล้องในระบบอีโก้เซ็นทริกวิชั่น [10]

2.4 ความต้านทานแบบโค้งงอ

อุปกรณ์ความต้านทานแบบโค้งงอเป็นเซนเซอร์ที่วัดปริมาณการโก่งตัวหรือโค้งงอ โดยปกติเซนเซอร์จะติดอยู่กับพื้นผิวและความต้านทานขององค์ประกอบเซนเซอร์จะแปรผันตามการดัดของพื้นผิว ซึ่งความต้านทานเป็นสัดส่วนโดยตรงกับปริมาณการโค้งงอมีลักษณะดังรูปที่ 2.7 โดยโครงสร้างเป็นหมึกพิมพ์ที่นำไฟฟ้าถูกพิมพ์อยู่บนแผ่นฉนวนรองรับ (Substrate) แบบยืดหยุ่น ผงหมึกพิมพ์มีคุณสมบัตินำไฟฟ้า กระบวนการพิมพ์ทำให้ผงหมึกมีการรวมตัวกันอย่างหลวมๆ ในขณะที่แผ่นเซนเซอร์อยู่ในแนวราบปกติจึงมีค่าความต้านทานที่ปลายทั้ง 2 ข้าง ประมาณ $25k\Omega$ [11]



รูปที่ 2.7 โครงสร้างส่วนประกอบของความต้านทานแบบโค้งงอ [12]

เมื่อมีการโก่งตัวของเซนเซอร์ความยาวของเซนเซอร์จะเพิ่มขึ้นผงหมึกที่รวมตัวอยู่จะกระจายตัวออกซึ่งจะทำให้ระยะห่างของผงหมึกมากขึ้น คุณสมบัติการนำไฟฟ้าจึงลดลง ทำให้ค่าความต้านทานเพิ่มขึ้น โดยค่าความต้านทานจะแปรเปลี่ยนไปตามมุมที่เกิดการโก่งตัวดังรูปที่ 2.8 ซึ่งมีค่าความต้านทานเป็นดังสมการที่ 2.1-2.3

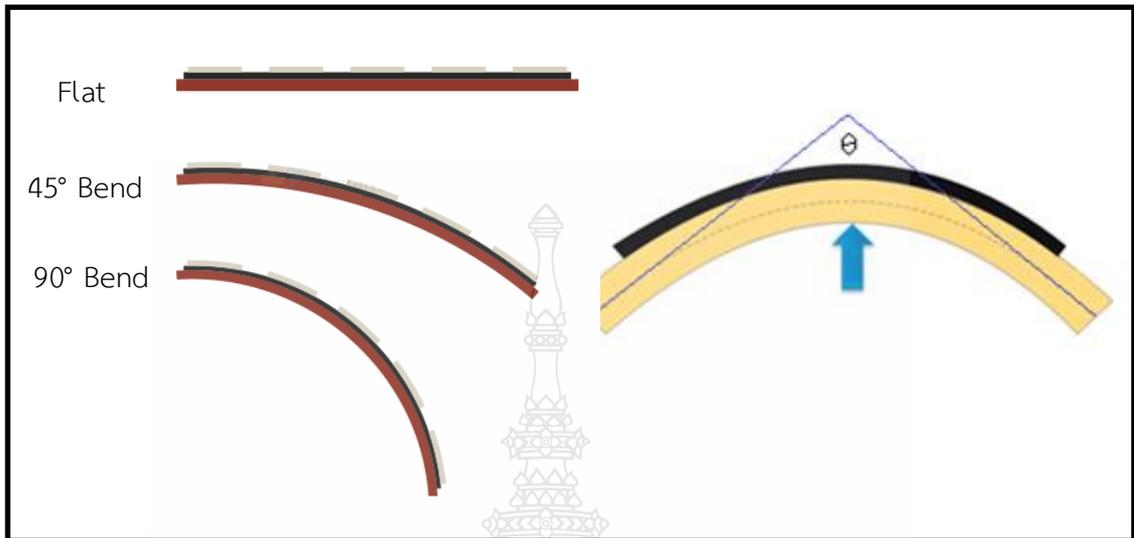
$$\theta = 0^\circ \rightarrow R_s \approx 25k\Omega \quad (2.1)$$

$$\theta = 45^\circ \rightarrow 25k\Omega < R_s < 125k\Omega \quad (2.2)$$

$$\theta = 90^\circ \rightarrow 25k\Omega \ll R_s < 125k\Omega \quad (2.3)$$

ที่มุม 45° ความต้านทานนี้จะอยู่ที่ประมาณ $75 k\Omega$ ที่มุม 90° ความต้านทานนี้จะอยู่ที่ประมาณ $125 k\Omega$ เมื่อเซนเซอร์ถูกปรับให้อยู่ในแนวราบอีกครั้ง ความต้านทานจะกลับสู่ค่าเดิม ดังนั้นเราจึงสามารถวัดมุมงอได้โดยดูจากค่าความต้านทานที่เกิดขึ้น ในทางทฤษฎี ค่าของ R_s ($k\Omega$) สัมพันธ์กับมุมดังนี้

$$R_s = 100 \cdot \frac{\theta}{180} + 25. \quad (2.4)$$



รูปที่ 2.8 การโก่งตัวและมุมการโก่งตัวของความต้านทานแบบโค้งงอ [12]

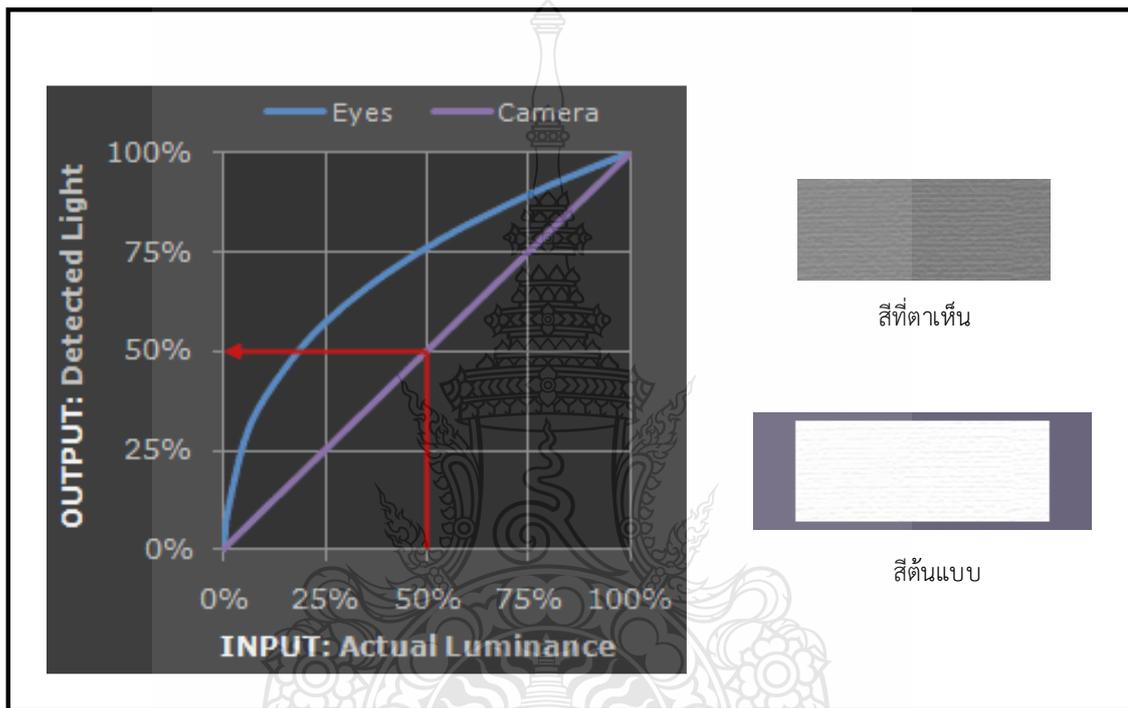
2.5 เครื่องมือทางการประมวลผลภาพ

การประมวลผลภาพ (Image processing) เป็นวิธีการที่นิยมใช้ในการตรวจจับและจดจำวัตถุต่าง ๆ โดยเฉพาะในงานคอมพิวเตอร์วิทัศน์ที่ใช้ในระบบไอทีเช่นทรูวิชั่น การประมวลผลภาพเป็นการนำภาพมาประมวลผลหรือคิดคำนวณด้วยคอมพิวเตอร์เพื่อให้ได้ข้อมูลที่ต้องการ โดยมีวิธีการที่ใช้กันอยู่มีหลายรูปแบบมาก อย่างไรก็ตามในหัวข้อนี้จะนำเสนอเทคโนโลยีที่ใช้หรือเกี่ยวข้องกับงานวิจัยนี้เท่านั้น โดยเน้นในเรื่องการปรับปรุงคุณภาพของรูปภาพ (Image Enhancement) ซึ่งเป็นกระบวนการแปลงข้อมูลภาพที่เป็นตัวเลข เพื่อที่จะสร้างภาพที่เน้นรายละเอียดหรือปรับโทนแสงของภาพที่ต้องการ โดยอาจจะต้องพิจารณาใช้เทคนิคปรับปรุงภาพหลายเทคนิคและนำมาประมวลผลเพื่อให้ได้ภาพที่ต้องการ ซึ่งแต่ละเทคนิคจะมีการปรับปรุงภาพในแง่มุมที่แตกต่างกัน ในงานวิจัยนี้จะใช้การปรับปรุงภาพจากการปรับค่าแกมมาเพื่อชดเชยแสงสว่าง ซึ่งมีงานวิจัยและมีการนำไปประยุกต์ใช้อย่างกว้างขวาง [13] [14]

2.5.1 ค่าแกมมาของภาพ

ค่าแกมมา (Gamma) คือ ค่าความสัมพันธ์ระหว่างค่าตัวเลขในพิกเซล (Pixel) ของภาพและค่าความแสงสว่าง (luminance) ตามความเป็นจริงของภาพที่ได้จากเซนเซอร์รับภาพของกล้อง ถ้าไม่มีค่าแกมมา เฉดสีที่ถูกเซนเซอร์กล้องดิจิทัลบันทึกไว้จะเห็นไม่ตรงกับที่ตามนุษย์หรือในจอมอนิเตอร์แสดงผล เราจึงต้องมีระบบที่เรียกว่า แกมมาคอลเรคชัน (Gamma correction) โดยการทำงานของเซนเซอร์กล้องดิจิทัล เมื่อมีจำนวนแสงโปรตอนกระทบ image sensor ที่ 2 เท่า กล้องได้รับข้อมูลเป็น

2 เท่าเช่นกัน (ความสัมพันธ์แบบ Linear) แต่ตาของมนุษย์ไม่ได้ทำงานแบบเดียวกัน ซึ่งสายตามนุษย์รับรู้แสงเพียงส่วนหนึ่งเท่านั้น และปรับตามความสว่างเพิ่มตามความเข้มแสงของแสงนั้น (ความสัมพันธ์แบบ Nonlinear) ดังกราฟในรูปที่ 2.9 ยกตัวอย่างเช่น ที่ 50% หรือสีเทา สายตาของมนุษย์จะเห็นแสงได้มืดกว่ากล้องถ่ายภาพอย่างเห็นได้ชัด เมื่อเทียบกันแล้วตามนุษย์อ่อนไหวต่อแสงในส่วนโทนที่มีได้มากกว่า จากเหตุผลในข้อนี้ทำให้ทัศนวิสัยในการมองของเราจะมีช่วงรับแสงที่กว้างกว่า ไม่อย่างนั้น ความสว่างในที่โล่งแจ้งจะมากเกินกว่าสายตาจะรับได้



รูปที่ 2.9 ตัวอย่างความเข้มแสงที่ 50% ที่เซนเซอร์ของกล้องดิจิทัลจับได้จากสีจริง [15]

ดังที่กล่าวมาแล้วค่าแกมมาเป็นการแปรค่าระหว่างการรับแสงที่ตามนุษย์เห็นกับภาพจากเซนเซอร์ของกล้องดิจิทัล เมื่อภาพจากกล้องดิจิทัลถูกบันทึกข้อมูลเก็บไว้ หรือเรียกว่า Gamma encode ซึ่งนั่นจะทำให้ค่าที่ได้ใกล้เคียงกับสิ่งที่เห็นจากจอแสดงผล เนื่องจากระบบการมองเห็นหรือการตอบสนองต่อความเข้มแสงของมนุษย์นั้นไม่เป็นเชิงเส้น กล่าวคือ มนุษย์มองเห็นส่วนมืดสว่างกว่าความเป็นจริง แต่ Image sensor ตอบสนองต่อความเข้มแสงเป็นเชิงเส้น ยกตัวอย่างเช่น สีขาว มีความเข้มแสง 100 หน่วย กับสีเทามีความเข้มแสง 50 หน่วย สีเทาที่มนุษย์เห็นจะไม่ใช้สีเทากลางหรือจะมองสีเทานั้นมีความเข้มแสงหรือสว่างมากกว่าครึ่งหนึ่งของสีขาว สมมติในสถานะการณ์เดียวกัน หากมีสีเทาที่มีความเข้มแสง 25 หน่วย มนุษย์จะรับรู้หรือมองเห็นสีเทานี้ว่าสว่างเป็นครึ่งหนึ่งของสีขาวหรือเห็นเป็นสีเทากลางเช่นเดียวกับฟิล์มกรองแสงที่ยอมให้แสงผ่าน 50% เมื่อมนุษย์มองผ่านฟิล์มดังกล่าว

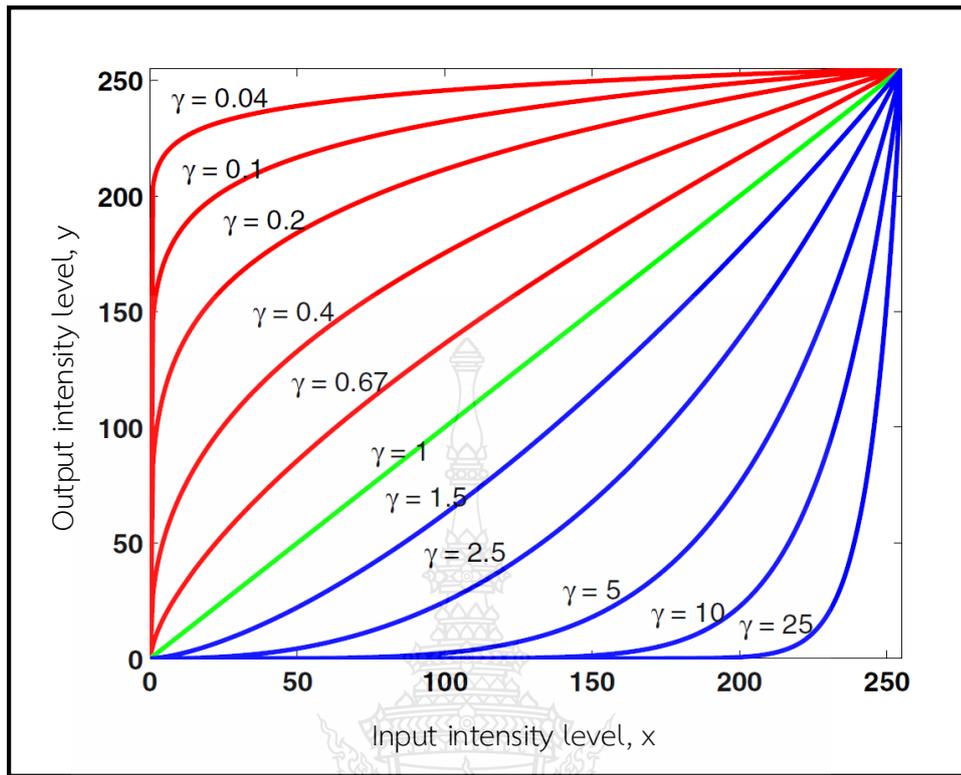
จะรับรู้ว่ามันไม่ได้ทำให้มืดลงครึ่งนึง แต่จะรู้สึว่ามันมืดลงเพียง 30% ในขณะที่อุปกรณ์ต่าง ๆ เช่น จอแสดงผล รวมทั้ง image sensor นั้นมีพฤติกรรมเป็นเชิงเส้นจึงจำเป็นต้องทำการ Gamma correction เพื่อเลียนแบบการรับรู้ของมนุษย์ซึ่งจะใช้การแปลงภาพแบบสมการยกกำลัง (Power law) ในการปรับปรุงภาพ ซึ่งจะเป็นการปรับความเข้มแสงให้สว่างขึ้น โดยมีฟังก์ชันการแปลงภาพในแบบยกกำลังแบ่งออกได้เป็นยกกำลังเพิ่มขึ้น (Power law transformations) และยกกำลังเศษส่วนหรือการหาราก (nth root transformation)

2.5.2 เทคนิค Adaptive Gamma Correction

เทคนิค Gamma correction เป็นวิธีที่ใช้กันทั่วไปในการแปลงภาพระดับสีเทา (Gray Scale) ที่มีช่วงค่าพิกเซล 0-255 โดยใช้รูปแบบพื้นฐานในสมการที่ 2.5 โดยที่ x คือค่าระดับสีเทาก่อนการแก้ไขแกมมา ส่วน y คือ ค่าพิกเซลสีเทาหลังทำ Gamma correction โดยมีความสัมพันธ์ระหว่างค่า x และ y ดังในรูปที่ 2.10 ซึ่งมีความสัมพันธ์กับค่า γ ที่ 3 กรณี คือ กรณีแรกเมื่อค่า $\gamma = 1$ จะได้ค่าภาพระดับสีเทาเหมือนก่อนการปรับค่าของพิกเซลภาพ ส่วนกรณีที่ $2 \gamma < 1$ จะได้ค่าระดับสีเทามีค่าพิกเซลของภาพมากกว่าก่อนปรับและจะเพิ่มขึ้นอย่างเห็นได้ชัดในช่วงค่าสีเทาระดับต่ำ และกรณีสุดท้าย $\gamma > 1$ ค่าระดับสีเทาของภาพมีค่าพิกเซลของภาพน้อยกว่าก่อนปรับและจะลดลงอย่างเห็นได้ชัดช่วงค่าสีเทาระดับสูง [16]

$$y = 255 \left(\frac{x}{255} \right)^\gamma \quad (2.5)$$

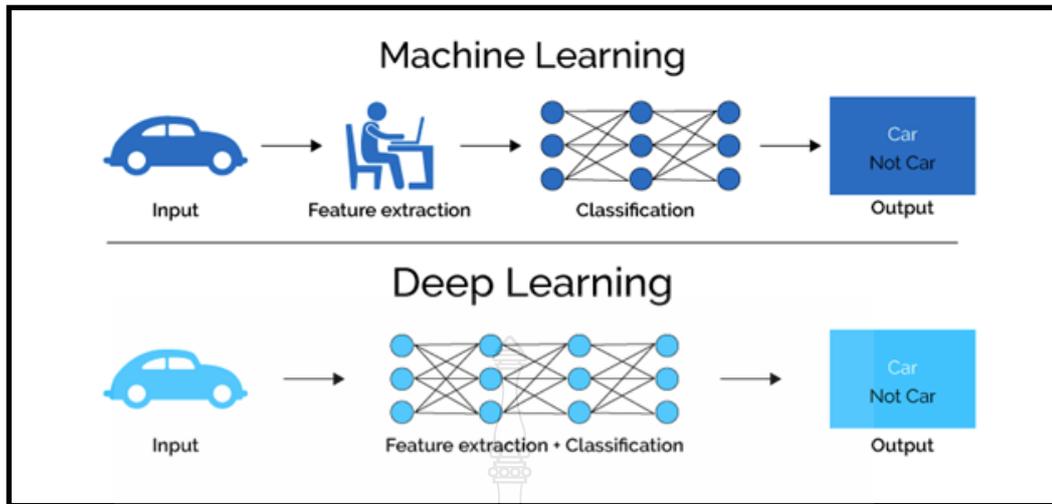
ถ้าทำดำเนินการปรับค่าแกมมาด้วยค่าคงที่กับพิกเซลของภาพทั้งหมด สำหรับผลที่ได้จะทำให้ภาพทั้งหมดจะมีมืดลงหรือสว่างขึ้น อย่างไรก็ตามพิกเซลของภาพส่วนใหญ่มีแสงสว่างที่ไม่สม่ำเสมอโดยถ้ามีบางพื้นที่สว่างเกินไปและบางส่วนมืดไป ค่าแกมมาคงที่จะไม่สามารถใช้ชดเชยได้อย่างมีประสิทธิภาพ ในกรณีที่มามีบริเวณแบบดังกล่าวทั้งสองพร้อมกัน ดังนั้นในกรณีนี้จะสร้างฟังก์ชันที่เหมาะสมเพื่อสร้างความสัมพันธ์ระหว่างค่าแกมมากับค่าระดับพิกเซลของภาพสีเทา ค่าแกมมานั้นน้อยกว่า 1, $\gamma < 1$ ใช้กับส่วนของภาพมีค่าระดับสีเทาต่ำและค่าแกมมาสูงกว่า 1, $\gamma > 1$ ใช้ในส่วนของภาพมีค่าระดับสีเทาสูง ซึ่งจะสามารถเพิ่มค่าพิกเซลระดับภาพสีเทาในส่วนพื้นที่มืดของภาพและลดค่าพิกเซลระดับภาพสีเทาของบริเวณที่สว่างเกินไป ซึ่งจะทำให้สามารถชดเชยแสงในพื้นที่มืดและสว่างเกินไปได้อย่างมีประสิทธิภาพในเวลาเดียวกัน



รูปที่ 2.10 กราฟพลอตความสัมพันธ์การแปลงความเข้มแสงสว่างที่ค่า γ ต่าง ๆ [54]

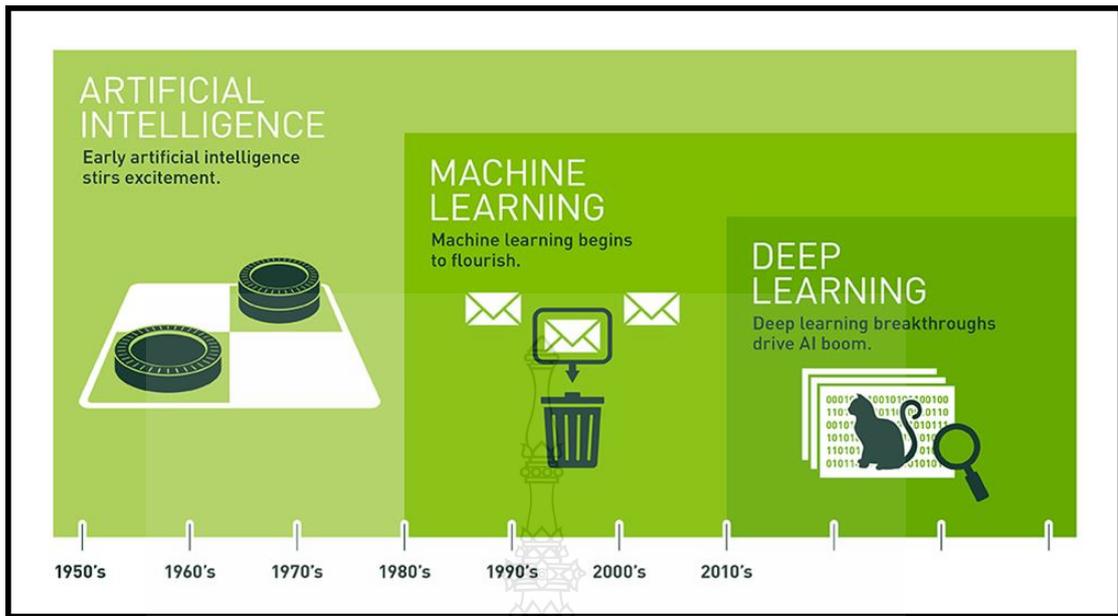
2.6 เทคนิคการเรียนรู้แบบเชิงลึก

เทคนิคการเรียนรู้เชิงลึก คือ ซอฟต์แวร์คอมพิวเตอร์ที่เลียนแบบให้เหมือนกับเครือข่ายเซลล์ประสาท (Network of neuron) ในสมองมนุษย์ ซึ่งเป็นซับเซ็ตย่อยของ Machine Learning และถูกเรียกว่า Deep Learning เพราะสร้างโครงข่ายประสาทที่ซับซ้อนแบบลึก (Deep Neural Network) ขึ้นมา โดยใช้จำนวนชั้น (Layers) ที่แตกต่างกันเพื่อเรียนรู้ข้อมูล ความลึกล้ำของโมเดล (Model) แสดงออกมาตามจำนวนชั้นที่มากขึ้น ส่วน Machine Learning (ML) หมายถึง ความสามารถของเครื่องมือที่สามารถเรียนรู้สิ่งต่าง ๆ ได้จากข้อมูลที่มีอยู่ โดยที่ไม่ต้องเขียนโค้ดอธิบายกฎเกณฑ์ต่าง ๆ ของสิ่งนั้น ML ทำให้เครื่องคอมพิวเตอร์สามารถเรียนรู้สิ่งต่าง ๆ ได้ด้วยตนเอง โดยใช้ประโยชน์จากพลังในการประมวลผลของคอมพิวเตอร์ในยุคปัจจุบัน ทำให้สามารถประมวลผลข้อมูลจำนวนมากได้นั่นเอง โดยข้อแตกต่างระหว่าง Machine Learning และ Deep Learning จุดที่เห็นได้ชัดเลยก็คือการทำ Feature extraction ที่ตัว Deep learning จะทำด้วยตนเอง แต่ถ้าเป็น Machine Learning จะต้องทำมาก่อนที่จะนำไปเข้าโมเดลของระบบ แต่ Deep Learning จะพยายามหาจุดเด่นของอินพุตโดยอัตโนมัติ ดังรูปที่ 2.11



รูปที่ 2.11 เปรียบเทียบการทำงานของ Machine Learning กับ Deep learning [17]

Deep Learning ได้รับความสนใจและพัฒนาขึ้นเป็นอย่างมากในช่วงไม่กี่ปีมานี้ แต่ข้อเสียของ Deep Learning คือ การใช้ทรัพยากรในการทำงานมาก ไม่ว่าจะเป็นทรัพยากรของคอมพิวเตอร์ หรือจะเป็นจำนวนข้อมูลที่ต้องการประมวลผล โดยหาก Neural Network มีโครงสร้างที่ใหญ่มากก็จำเป็นจะต้องมีข้อมูลที่เยอะมากเช่นกัน โดยรูปแบบย่อยของ Machine Learning ที่มีการคิดตั้งแต่ยุคบุกเบิก คือ Artificial Neural Networks (ANN) เป็นการจำลองวิธีการทำงานของสมอง ให้คอมพิวเตอร์รู้จักคิดและจดจำในแนวเดียวกับโครงข่ายประสาทของมนุษย์ ซึ่งทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งที่รับรู้ภายนอกทำให้กระแสประสาทที่วิ่งไปถึงตัวนิวเคลียสจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอนิวเคลียสก็จะกระตุ้นเซลล์อื่น การทำงานของ ANN จะไม่ซับซ้อนเท่าระบบประสาทของมนุษย์ แต่ใช้การจำลองว่าอินพุตแต่ละตัวมีน้ำหนัก (weight) และเซลล์แต่ละหน่วยจะมีตัวกำหนด (Threshold) ว่าน้ำหนักรวมของอินพุตต้องมากขนาดไหน ถึงจะส่งเอาต์พุตต่อไปได้ อย่างไรก็ตามกว่าแนวคิดนี้ ซึ่งมีมานานแล้วจะใช้งานได้จริง ก็ต้องรอจนกระทั่งเทคโนโลยีด้าน GPU ถูกพัฒนาจนสามารถตอบสนองได้เมื่อไม่กี่ปีที่ผ่านมาเอง หนึ่งในการค้นพบครั้งสำคัญของ Deep Learning เกิดขึ้นเมื่อ Andrew Ng ศาสตราจารย์จากมหาวิทยาลัยสแตนฟอร์ด ก่อตั้งโครงการ Google Brain ขึ้นในปี 2012 และฝึกปัญญาประดิษฐ์ ในระดับที่ลึกกว่าผ่านภาพต่าง ๆ จากวิดีโอใน YouTube กว่า 10 ล้านคลิป ซึ่ง Deep Learning ก็คือ การนำแนวคิดของ ANN มาใช้ในระดับที่ลึกและมี Layer ที่ซับซ้อนกว่านั่นเองและด้วยวิธีการเรียนรู้ครั้งแล้วครั้งเล่าแบบนี้เอง ที่ทำให้ AI สามารถแยกแยะสิ่งต่าง ๆ ได้อย่างแม่นยำยิ่งขึ้นจนเหนือกว่ามนุษย์ และนำไปใช้งานในหลายๆ สาขา เช่น การวินิจฉัยโรค การควบคุมรถไร้คนขับ การเรียนรู้กติกาของเกมแบบใน AlphaGo เป็นต้น ลำดับการพัฒนาเทคนิคการเรียนรู้เชิงลึกเป็นดังรูปที่ 2.12



รูปที่ 2.12 ลำดับการพัฒนา Artificial Intelligence, Machine Learning และ Deep learning [17]

อัลกอริทึมที่น่าสนใจที่มี Neural Network เป็นส่วนประกอบ ซึ่งในยุคปัจจุบันก็ได้มีอัลกอริทึมย่อยๆ ของ Neural Network อีกจำนวนมาก เช่น Convolution Neural Network (CNN), Recurrent Neural Network (RNN) เป็นต้น ซึ่งงานวิจัยนี้จะใช้โครงข่ายประสาทแบบ CNN ซึ่งได้รับความนิยมเป็นอย่างมาก เนื่องจากมีประสิทธิภาพการทำงานที่สูงกว่าโครงข่ายแบบอื่นๆ

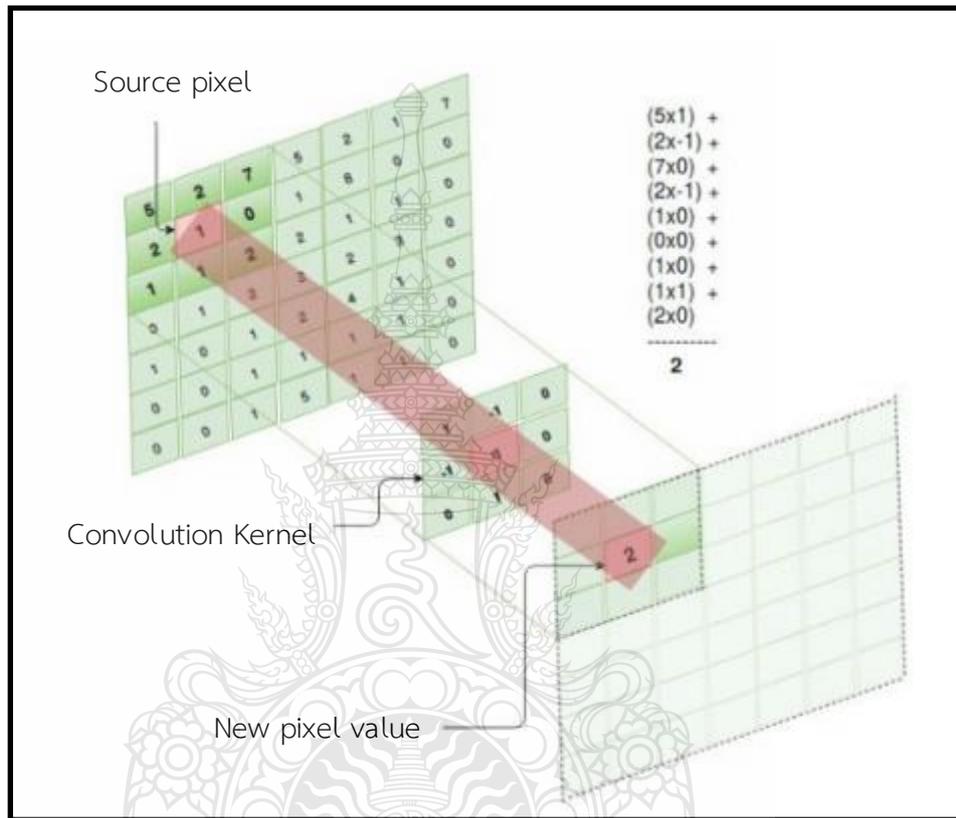
2.6.1 โครงข่ายประสาทแบบ CNN

โครงข่ายประสาทแบบ CNN เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม Bio-inspired ซึ่งได้รับความนิยมเป็นอย่างมาก เนื่องจากมีประสิทธิภาพในการทำงานที่ดี โดยที่ CNN จะจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อยๆ และนำกลุ่มของพื้นที่ย่อยๆ มาผสมกัน เพื่อดูว่าสิ่งที่เห็นอยู่เป็นอะไรกันแน่ การมองพื้นที่ย่อยของมนุษย์จะมีการแยกคุณลักษณะ (Feature) ของพื้นที่ย่อยนั้น เช่น ลายเส้นและการตัดกันของสี ซึ่งการที่มนุษย์รู้ว่าพื้นที่ตรงนี้เป็นเส้นตรงหรือสี่เหลี่ยม เพราะมนุษย์ดูทั้งจุดที่สนใจและบริเวณรอบๆ โดยวิธีการจะใช้เทคนิคการทำคอนโวลูชันข้อมูล คือการคำนวณแบบ dot product ระหว่างพื้นที่ส่วนย่อยของรูปภาพ (Sub-region) กับเคอร์เนล (Kernel) เพื่อดึงลักษณะเด่น (Feature Extraction) ออกมาจากรูปภาพ โดยเคอร์เนลที่นำมาคำนวณจะมีขนาดเล็กกว่าพื้นที่ส่วนย่อยของรูปภาพ โดยกำหนดเป็น เมทริกซ์จัตุรัส (Square Matrix) การคำนวณ Convolutional สามารถทำได้ดังสมการต่อไปนี้ [18]

$$G = h * F \quad (2.6)$$

$$G[i, j] = \sum_{u=-k}^k \sum_{v=-k}^k h[u, v] F[i - u, j - v]$$

โดยที่ h คือ รูปภาพที่ดำเนินการและ F คือ เคอร์เนล (Kernel) ขนาด $k * k$



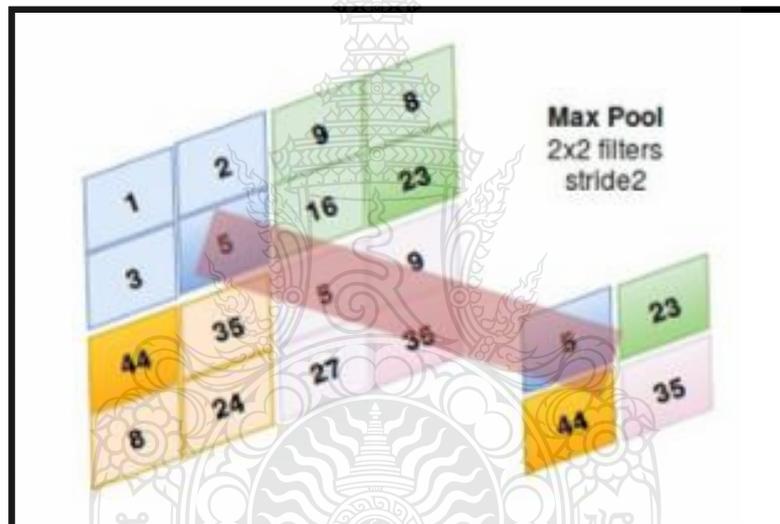
รูปที่ 2.13 การดำเนินการคอนโวลูชันระหว่างรูปภาพ (Sub-region) กับเคอร์เนล (Kernel) [18]

จุดเริ่มต้นของโครงข่ายประสาทเทียมแบบคอนโวลูชัน นำเสนอโดย LeCun et al. (1990) [55] ได้นำวิธีการคำนวณแบบคอนโวลูชันมาผนวกเข้ากับโครงข่ายประสาทเทียมเพื่อใช้คำนวณหา ลักษณะเด่นของรูปภาพ โดยขั้นที่ได้จากการคำนวณด้วยวิธีคอนโวลูชันเรียกว่า ชั้นคุณลักษณะเด่น (Feature Map) ในงานวิจัยได้ทดสอบกับข้อมูลตัวเลข โดยเก็บรวบรวมรหัสไปรษณีย์ จากซองจดหมาย ในประเทศสหรัฐอเมริกา โครงข่ายประสาทเทียมแบบคอนโวลูชัน สามารถทำได้ทั้งสกัดคุณลักษณะ (Feature Extraction) และจำแนกประเภท (Classification) ซึ่งถือได้ว่าเป็นจุดเด่นของโครงข่ายนี้ โดยทั่วไปโครงข่ายประสาทเทียมแบบคอนโวลูชันจะประกอบไปด้วย 3 ชั้นที่สำคัญดังนี้

2.6.1.1 ชั้นคอนโวลูชัน (Convolutional Layer) เป็นชั้นการคำนวณเพื่อหาผลลัพธ์ของ Neuron ที่เชื่อมต่อ (Connected) มาจากพื้นที่ส่วนย่อย (Local Region) ของรูปภาพด้วยวิธีการคอน

โวลูชัน ซึ่งเป็นการคำนวณแบบ dot product กับเคอร์เนลแบบจัตุรัสขนาดเล็ก เช่น 3x3, 5x5, หรือ 7x7 เป็นต้น ผลลัพธ์ที่ได้จากการคำนวณเรียกว่า Feature map

2.6.1.2 ชั้นพูลลิ่ง (Pooling Layer) เป็นชั้นที่อยู่คั่นกลางระหว่างชั้นคอนโวลูชัน โดยมีวัตถุประสงค์เพื่อลดขนาด (Down sampling) ของ Feature Map ให้มีขนาดเล็กลง ซึ่งสามารถใช้ฟังก์ชันค่าเฉลี่ย ฟังก์ชันค่าต่ำสุดและฟังก์ชันค่าสูงสุดในการคำนวณ ดังนั้น หากเลือกใช้ฟังก์ชันค่าสูงสุดในการคำนวณ จะเรียกว่า Max pooling ดังในรูปที่ 2.14 ดังนั้นการทำ Max Pooling สามารถทำได้โดยนำ Feature Map มาแบ่งออกเป็นพื้นที่ย่อย โดยกำหนดให้มีขนาดเป็น $p \times p$ ส่วน จากตัวอย่างในรูป กำหนดให้มีขนาด 2x2 ส่วน จากนั้น หาค่าที่มากที่สุดในแต่ละพื้นที่ย่อย หรือ Pool จากพื้นที่แรกที่ประกอบด้วยตัวเลข 1, 2, 3, 5 สังเกตได้ว่าตัวเลขที่มากที่สุดคือ 5 ดังนั้นใน Pool นี้จึงเลือกค่า 5 มาพิจารณา



รูปที่ 2.14 การดำเนินการคำนวณในชั้นพูลลิ่ง (Pooling Layer) [18]

2.6.1.3 ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) เป็นชั้นที่เชื่อมโยงระหว่าง Feature Map และ Output แบบสมบูรณ์ นั่นหมายถึงทุกๆ Neuron ที่อยู่ในชั้นสุดท้ายของ Feature Map จะถูกนำไปเปลี่ยนรูป (Reshape) หรือเรียกว่า Flatten เพื่อส่งไปคำนวณต่อในชั้นถัดไป ซึ่งก็คือชั้นของโครงข่ายประสาทเทียมแบบ Multilayer perceptron ดังนั้นในชั้นนี้ Feature Map ที่ได้ในชั้นสุดท้ายจึงเปรียบเสมือนข้อมูลนำเข้า (Input) ไปยังโครงข่ายประสาทเทียม

2.6.2 การถ่ายโอนการเรียนรู้

การถ่ายโอนการเรียนรู้ (Transfer learning) สาขาหนึ่งของการเรียนรู้ของ Machine Learning และปัญญาประดิษฐ์ ซึ่งมีจุดมุ่งหมายเพื่อนำข้อมูลที่ได้รับการเรียนรู้ในงานหนึ่งไปใช้กับงานอื่นที่คล้ายคลึงกันแค่ปรับปรุงส่วนที่เรียนรู้แล้วใช้เป็นส่วนหนึ่งของงานใหม่ที่ผ่านการถ่ายทอดความรู้จากงานที่เกี่ยวข้องที่ได้เรียนรู้ไปแล้ว ในงานวิจัยนี้จะใช้การพัฒนาส่วน Pretrain CNN ซึ่งใช้โมเดลโครงข่ายประสาท CNN จำนวน 5 โมเดล จากโปรแกรม MATLAB 2021a โดยมีรายละเอียดแต่ละโมเดลดังนี้

2.6.2.1 โมเดลแบบ Alexnet ในปี ค.ศ. 2012 นักวิจัย Krizhevsky และคณะ [48] นำเสนอโครงข่ายที่มีจำนวนทั้งสิ้น 8 ชั้น โดยโครงสร้างนี้เรียกว่า AlexNet ประกอบไปด้วยชั้นคอนโวลูชัน และชั้นเชื่อมโยงแบบสมบรูณ์ จำนวน 5 และ 3 ชั้น ตามลำดับ โดยผลลัพธ์ของชั้นเชื่อมโยงแบบสมบรูณ์ ถูกกำหนดให้มีจำนวน 100 โหนด ภาพที่ใช้จะมีขนาด $224 \times 224 \times 3$ พิกเซล ดังนั้น ชั้นคอนโวลูชันแรกจะถูกกำหนดให้มี 96 kernel โดยที่แต่ละ kernel มีขนาด $11 \times 11 \times 3$ ในการคำนวณจะเลื่อน kernel ครึ่งละ 4 พิกเซล ผลลัพธ์จากการคอนโวลูชันจะมีขนาด $55 \times 55 \times 96$ ต่อมาในชั้นคอนโวลูชันที่สองของ ถูกกำหนดให้มี 256 kernel โดยที่แต่ละ kernel มีขนาด $5 \times 5 \times 48$ และจากผลลัพธ์ของการคอนโวลูชันจะทำให้มีขนาด $27 \times 27 \times 256$ หลังจากการคำนวณชั้นคอนโวลูชันที่ 1 และ 2 ข้อมูลจะถูกส่งไปชั้นพูลลิ่งเพื่อทำ Normalized และคำนวณค่าสูงสุดในแต่ละ Pool (Max Pooling) โดยที่แต่ละ Pool จะถูกแบ่งเป็น 2×2 ส่วนในคอนโวลูชันชั้นที่สามจะมีจำนวน 384 kernel โดยที่แต่ละ kernel มีขนาด $3 \times 3 \times 256$ ทำให้ได้ผลลัพธ์จากการคอนโวลูชันจะมีขนาด $13 \times 13 \times 256$ โดยชั้นที่สี่ ถูกกำหนดให้มี 384 kernel โดยมีขนาด $3 \times 3 \times 192$ ผลลัพธ์จากการคอนโวลูชันทำให้ได้ขนาด $13 \times 13 \times 384$ ชั้นที่ห้า มีจำนวน 256 kernel โดยมีขนาด $3 \times 3 \times 192$ สุดท้ายจะได้ผลลัพธ์จากการคอนโวลูชัน มีขนาด $13 \times 13 \times 256$ โดยที่ชั้นคอนโวลูชันในชั้นที่ 1 ถึง 5 ใช้ฟังก์ชัน Rectified Linear Units (ReLU) เป็นฟังก์ชันกระตุ้น หรือ Activation Function ชั้นเชื่อมโยงแบบสมบรูณ์ถูกแบ่งออกเป็น 2 ชั้น และกำหนดให้มีจำนวนชั้นละ 4,096 Node

2.6.2.2 โมเดลแบบ VGG ในปี ค.ศ. 2015 Simonyan และ Zisserman [49] ได้นำเสนอโครงข่ายประสาทแบบ VGG เป็นโครงสร้างแบบลึก เนื่องจากมีชั้นคอนโวลูชันจำนวน 16 ชั้น มีการใช้ kernel ในการทำคอนโวลูชันขนาด 3×3 โดยมีโครงสร้างประกอบด้วยชั้นคอนโวลูชัน จำนวน 5 กลุ่ม โดยในการคอนโวลูชันกลุ่มที่ 1 เป็นการทำคอนโวลูชันที่ 1 และ 2 มีจำนวน kernel ชั้นละ 64 ในการคอนโวลูชันกลุ่มที่ 2 เป็นการทำคอนโวลูชันที่ 1 และ 2 มีจำนวน kernel ชั้นละ 128 ในการคอนโวลูชันกลุ่มที่ 3 เป็นการทำคอนโวลูชันที่ 1 ถึง 3 มีจำนวน kernel ชั้นละ 256 เคอร์เนล ในการคอนโวลูชันกลุ่มที่ 4 เป็นการทำคอนโวลูชันที่ 1 ถึง 3 มีจำนวน kernel ชั้นละ 512 และ ในการคอนโวลูชัน

กลุ่มที่ 5 เป็นการทำคอนโวลูชันที่ 1 ถึง 3 มีจำนวน kernel ชั้นละ 512 โดยการทำคอนโวลูชันในกลุ่มที่ 1 ถึง 5 จะตามด้วยชั้นแมกซ์พูลลิ่ง จากนั้นข้อมูลจะส่งไปยัง ชั้นเชื่อมโยงแบบสมบูรณ์ ที่มีจำนวน 4,096 Node และในชั้นสุดท้ายของชั้นเชื่อมโยงแบบสมบูรณ์ ถูกกำหนดให้มีจำนวน 1,000 โหนด โดยคำนวณหาผลลัพธ์ด้วยวิธีซอฟต์แวร์ โดยเรียกโครงสร้างที่กล่าวมานี้ว่า VGG 16 ซึ่งมีการทำคอนโวลูชัน 16 ชั้นและต่อมีการเพิ่มการทำคอนโวลูชันอีกสามชั้น เรียกว่า VGG 19 โดยทั้งสองโมเดลใช้ขนาดภาพอินพุต 224x224x3

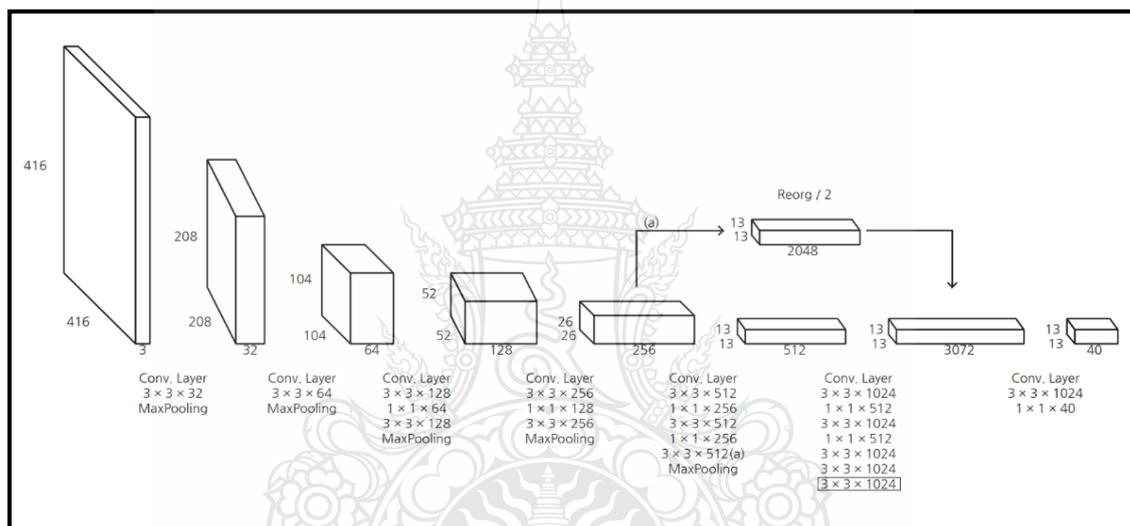
2.6.2.3 GoogLeNet ในปี ค.ศ. 2014 ได้ถูกนำเสนอโดยใช้โครงสร้างของโมเดล Inception ทำให้มีโครงข่ายประสาท CNN ที่มีความลึกและกว้าง [51] แต่ไม่ทำให้การประมวลผลช้าลง โดยในแต่ละ Inception มีการคำนวณคอนโวลูชันแบบขนานที่ใช้ kernel ขนาด 1x1, 3x3 และ 5x5 และยังทำให้จำนวนมิติข้อมูลลดลง โดยออกแบบให้มีโครงสร้างของ Inception จำนวน 9 ชั้น จากทั้งหมด 22 ชั้น และใช้ฟังก์ชันการคำนวณชนิด ReLU เพื่อใช้เป็นฟังก์ชันกระตุ้นและชั้นเชื่อมโยงแบบสมบูรณ์กำหนดให้มีจำนวน 1024 Node ใช้ขนาดภาพอินพุต 224x224x3

2.6.2.4 Resnet 50 นำเสนอโดย Kaiming He [50] ในปี ค.ศ. 2015 โดยโครงสร้างต่างจากโครงข่ายแบบ Sequential ดั้งเดิม เช่น AlexNet และ VGG ตรงที่ ResNet เป็นรูปแบบสถาปัตยกรรมที่ต้องอาศัยโมดูลสถาปัตยกรรมที่เรียกว่า Microarchitecture คือ ใช้ชุด building blocks ในการสร้างเครือข่ายใหม่ทั้งหมด โดยใช้โมดูลที่มีความแม่นยำมากขึ้นด้วยการอัปเดตโมดูลที่เหลือเพื่อใช้ Identity mapping โดยโมเดล ResNet นั้นมีโครงข่ายประสาท CNN ที่ลึกกว่าตระกูล VGG มาก แต่ขนาด Weights size เล็กกว่าอันเนื่องมาจากการใช้ชั้นพูลลิ่งแบบ Global average มากกว่าชั้นเชื่อมต่อแบบสมบูรณ์ ทำให้ขนาดโมเดล ResNet50 ซึ่งมีการคำนวณชั้นคอนโวลูชันจำนวน 50 ชั้น ได้รับความนิยมอย่างมาก

2.6.3 อัลกอริทึม YOLO

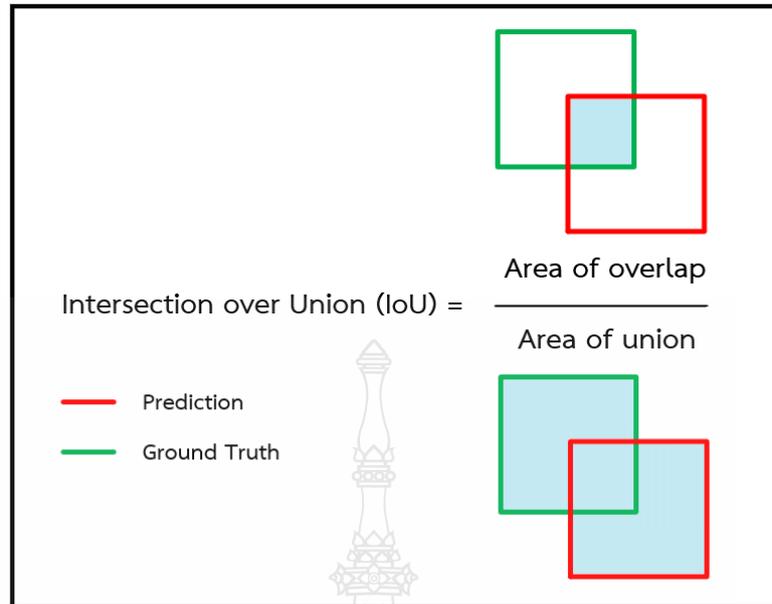
YOLO หรือ You Only Look Once เป็นเทคนิคการรู้จำวัตถุที่ใช้วิธีการที่เรียกว่า Fast Single-Shot Detection ซึ่งจะเป็นวิธีการที่สามารถตรวจจับวัตถุได้ จากการส่งผ่านรูปภาพเข้าไปในระบบเพียงครั้งเดียว โดยใช้หลักการโครงข่ายประสาทแบบ CNN โดยอัลกอริทึม YOLO มีการพัฒนาอย่างต่อเนื่อง ซึ่ง YOLO มีหลายรุ่น (Version) ในแต่ละรุ่นจะมีโครงสร้างสถาปัตยกรรมที่แตกต่างกัน มีประสิทธิภาพทั้งด้านความแม่นยำและความเร็วในการประมวลผล โดย ในปี 2016 Redmon และ Farhadi [19] ได้เสนอ YOLO ตามโครงสร้างของ SSD (Single Shot Multibox Detector) โดยต่อมา YOLOv2 มีอัตราการรับรู้ที่เร็วขึ้นในขณะเมื่อเทียบกับวิธีการ RCNN, Fast RCNN และ Faster RCNN และยังคงความแม่นยำเทียบเท่ากับ SSD500 ความแม่นยำเฉลี่ยสูงถึง 76.8% และความเร็วในการรับรู้

67 เฟรมต่อวินาที บนชุดข้อมูล VOC2007 จากโมเดลเครือข่าย Darknet-19 ซึ่ง YOLOv2 ประกอบด้วย ชั้นของ Convolutional 19 ชั้นและมี Pool layers สูงสุด 5 ชั้นและให้ความแม่นยำในการตรวจจับที่สูงขึ้น ในขณะที่รักษาความเร็วในการตรวจจับเมื่อใช้ภาพที่มีความละเอียด 544x544 พิกเซล รุ่น YOLOv2 [20] [21] สามารถทำความเร็วได้ 40 เฟรมต่อวินาที และมีความแม่นยำในการตรวจจับสูงกว่าเร็วกว่า RCNN ในอัลกอริทึม YOLO การดำเนินการจำแนกวัตถุว่าเป็นชนิดอะไร (Classification) และดำเนินการหาตำแหน่งของวัตถุ (Localization) โดยใช้กรอบล้อมวัตถุ (Bounding Box) จะทำไปพร้อมๆ กัน โดยวิธีของ YOLO ไม่ได้พิจารณาดำเนินการจากภาพทั้งภาพ แต่จะแบ่งภาพออกเป็น ส่วน ๆ ซึ่งวิธีการแบบนี้ส่งผลดีในด้านความเร็วของการประมวลผล โดยมีโครงสร้างสถาปัตยกรรม YOLOv2 ดังรูปที่ 2.15



รูปที่ 2.15 โครงสร้างสถาปัตยกรรมภายในอัลกอริทึม YOLOv2 [22]

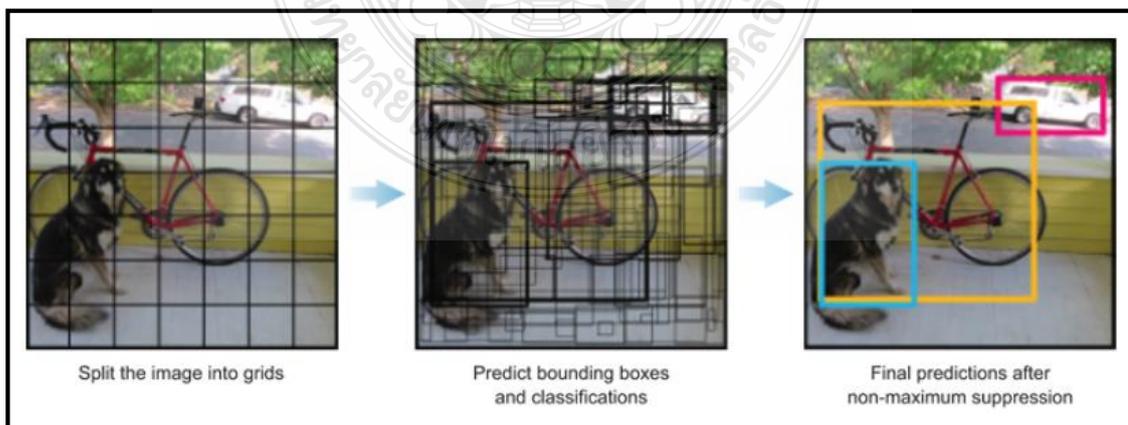
ขั้นตอนการทำงานของ YOLO จะเป็นในลักษณะการแบ่ง รูปภาพเป็นส่วนๆ หรือกริด (grid) จากนั้นทำการเลื่อนการคำนวณ ไปทีละจุดตามที่แบ่งกริดไว้ (Sliding Windows) พร้อมกับ คำนวณหา ว่าวัตถุจะมีอยู่จริงหรือไม่ พื้นที่ที่มากน้อยเพียงไรจากความน่าจะเป็นของวัตถุที่ปรากฏในพื้นที่นั้นๆ ด้วย กระบวนการ Intersection over Union (IoU) ซึ่งคำนวณจากอัตราส่วนของพื้นที่ซ้อนทับของพื้นที่ที่ทำนายกับพื้นที่จริงกับพื้นที่รวมของพื้นที่ที่ทำนายกับพื้นที่จริงดังรูปที่ 2.16 ในการทำนายกรอบล้อม วัตถุจะได้ข้อมูลเป็นชุดข้อมูลประเภท Array ซึ่งประกอบด้วยข้อมูลการมีอยู่จริงของวัตถุตำแหน่งและ ขนาดของกรอบล้อมวัตถุและชนิดของวัตถุ



รูปที่ 2.16 พื้นที่การคำนวณหา Intersection over Union ที่ตรวจจับได้

กรณีถ้าในแต่ละกริดมีวัตถุมากกว่าหนึ่งอย่าง YOLO มีกระบวนการ Anchor box เพื่อแก้ปัญหาดังกล่าว แนวคิดคือการสร้าง Anchor box ในรูปทรงต่าง ๆ และคำนวณใหม่ เพื่อให้ได้ผลลัพธ์ที่ครอบคลุมและแม่นยำมากขึ้น เมื่อระบบสร้างกรอบล้อมวัตถุจนหมดแล้วดังรูปที่ 2.17 จะเข้าสู่กระบวนการ Non-max suppression ซึ่งขั้นตอนนี้จะเป็นการลดจำนวนกรอบล้อมวัตถุที่เป็นวัตถุเดียวกันและอยู่ในพื้นที่เดียวกันที่มีความน่าจะเป็นน้อยกว่าค่าที่กำหนดไว้ (Threshold) [23] ดังสมการที่ 2.7

$$d(box, centroid) = 1 - IOU(box, centroid) \quad (2.7)$$



รูปที่ 2.17 การทำงานของอัลกอริทึม YOLOv2 ในการตรวจจับวัตถุภายในภาพ [19]

2.7 งานวิจัยที่เกี่ยวข้อง

งานวิจัยด้านวิเคราะห์การใช้มือจับวัตถุในระบบอีโก้เซ็นเซอร์ที่ขึ้นในในงานคอมพิวเตอร์วิทัศน์ที่มีนักค้นคว้าวิจัยกันหลากหลายในปัจจุบันเพื่อทำความเข้าใจการจัดวางของนิ้วมือต่อวัตถุที่จับผ่านมุมมองของผู้สวมใส่กล้องติดตัว โดยมีการศึกษาเรื่องของมือและนิ้วมือกันมาอย่างยาวนานเพื่อเข้าใจกระบวนการความสัมพันธ์ลักษณะการใช้มือจับ ชนิดวัตถุและการกระทำต่อวัตถุ ปัจจุบันเทคโนโลยีกล้องพัฒนาให้มีขนาดเล็กจึงนำมาใช้ในระบบอีโก้เซ็นเซอร์ที่ขึ้นในการตรวจจับวัตถุและระบุกิจกรรมของผู้ที่สวมใส่ โดยลักษณะเด่นของการบันทึกภาพแบบนี้ คือ การได้เห็นมุมมองของบุคคลที่สวมใส่โดยการวางกล้องหันไปทางข้างหน้าติดบนหน้าอกหรือศีรษะ ซึ่งให้มุมมองที่ไม่เหมือนการบันทึกภาพทั่วไป โดยเน้นที่ตัวบุคคลเพื่อบันทึกข้อมูลที่เกี่ยวข้องกับผู้สวมใส่กล้องมากขึ้น ระบบระบบอีโก้เซ็นเซอร์ที่ขึ้นถูกนำไปใช้กับแอปพลิเคชันต่าง ๆ มากมาย แม้กระทั่งการดูแลสุขภาพทางการแพทย์ ดังนั้นผู้วิจัยจึงได้ศึกษางานวิจัยที่เกี่ยวข้องเพื่อใช้เป็นแนวทางในการทำวิจัย โดยมีรายละเอียดดังนี้

Steve Mann (1997) นำเสนอการใช้กล้องถ่ายภาพติดที่บนศีรษะสำหรับถ่ายทอดกิจกรรมในการใช้ชีวิตประจำวันโดยมีวัตถุประสงค์เพื่อที่จะสามารถประมวลผลภาพวิดีโอได้จากระยะไกลด้วยการสร้างวิดีโอพลาตูปอสต์ โดยลิงก์การสื่อสารเพื่อส่งวิดีโอจากกล้องไปยังคอมพิวเตอร์จากระยะไกลและส่งสัญญาณที่ประมวลผลจากคอมพิวเตอร์กลับไปเพื่อแสดงผลแบบสวมศีรษะ [24]

Thad Starner และคณะ (1998) จากห้องปฏิบัติการมีเดีย มหาวิทยาลัยแมสซาชูเซตส์ (MIT) นำกล้องถ่ายภาพขนาดเล็กติดศีรษะเพื่อเพิ่มศักยภาพการตรวจการโดยทำงานร่วมกับระบบคอมพิวเตอร์เป็นระบบคอมพิวเตอร์วิทัศน์ในยุคแรก การประมวลผลภาพทำงานร่วมกับเทคนิค Hidden Markov Model (HMM) [25]

W.W.Mayol และคณะ (2000) ได้นำเสนออุปกรณ์ต้นแบบที่ใช้สวมใส่งานอีโก้เซ็นเซอร์ที่ขึ้นซึ่งแยกการเคลื่อนไหวของกล้องออกจากท่าทางและการเคลื่อนไหวของผู้สวมใส่ โดยการรวมเคลื่อนที่และการตอบสนองของเซนเซอร์ภาพภายใต้ตำแหน่งของเซนเซอร์และความสัมพันธ์ขณะทำการสวมใส่โดยประสิทธิภาพของหุ่นยนต์ต้นแบบจะประเมินจากสถานการณ์การใช้งานหุ่นยนต์จากการควบคุมระยะไกลที่มองเห็นด้วยกล้องที่สวมใส่ ซึ่งมีความสำคัญมากสำหรับการช่วยเหลือหุ่นยนต์ที่ทำงานจากการควบคุมจากระยะทางไกล [26]

Michael F. Land และ Mary Hayhoe (2001) นำเสนอการศึกษาความสัมพันธ์ของการเคลื่อนไหวของสายตาในการจ้องมองและการจัดการของมือในขณะที่จับวัตถุ โดยทดลองในงานเตรียมอาหารด้วยกล้องอีโก้เซ็นเซอร์ที่ขึ้น โดยสามารถแบ่งออกเป็นชุดของการกระทำต่อวัตถุ โดยมีข้อสรุปว่าสายตาจะมองไปที่วัตถุตามลำดับก่อนหลัง ซึ่งบ่งชี้ว่าการเคลื่อนไหวของดวงตาได้รับการวางแผนใน

รูปแบบมอเตอร์และดวงตามักจะจับจ้องไปที่วัตถุเดียวกันตลอดการกระทำแม้ว่าวัตถุจะเคลื่อนที่ไป โดยการควบคุมจะถูกมองจากบนลงล่างเป็นหลัก [27]

Ryuhei Tenmoku และคณะ (2005) สร้างระบบต้นแบบวิเคราะห์วิธีค้นหาตำแหน่งและอธิบายวิธีการจัดการมุมมองที่เน้นวัตถุจริงในระบบ Augmented Reality (AR) โดยการรวมสภาพแวดล้อมจริงกับวัตถุเสมือนเข้าด้วยกันในเวลาเดียวกัน ซึ่งผู้ใช้สามารถดูและพร้อมคำอธิบายประกอบโดยใช้โมเดล 3 มิติของฉากจริง ด้วยการถ่ายภาพที่สร้างจากฉาก 3 มิติ ด้วยกล้องที่สวมใส่ในระบบ AR [28]

Steve Hodges และคณะ (2006) นำเสนอการใช้กล้อง Sense Cam เพื่อจดจำกิจกรรมที่มนุษย์ใช้ในการดำเนินชีวิตโดยเพิ่มชุดเซนเซอร์กล้องถ่ายภาพนิ่งแบบสวมใส่ได้ ช่วยให้ผู้ใช้สวมใส่จดจำแง่มุมต่างๆ และสร้างเครื่องช่วยความจำย้อนหลังที่มีประสิทธิภาพ โดยการศึกษาผู้ใช้เชิงลึกกับอุปกรณ์นี้ครั้งแรกในผู้ป่วยที่เป็นโรคความจำเสื่อมและผู้ที่มีภาวะเสื่อมของระบบประสาท เช่น โรคอัลไซเมอร์ เป็นต้น [29]

Natasha Alves และ Tom Chau (2008) นำเสนอใช้กล้องดิจิทัลถ่ายในส่วนที่เป็นเฉพาะท่อนแขนลงไปถึงมือร่วมกับตัวตรวจวัดสัญญาณของกล้ามเนื้อเพื่อหาความสัมพันธ์ของสัญญาณที่เกิดขึ้นขณะการใช้มือจับวัตถุเปรียบเทียบกับภาพที่ได้จากกล้องดิจิทัลโดยการติดตั้งกล้องบนโต๊ะในห้องปฏิบัติการเพื่อติดตามการจับวัตถุของมือโดยการตรวจจับสีผิวมนุษย์ในการตรวจหามือจับภายในภาพเพื่อเก็บรวบรวมข้อมูลและแบ่งแยกลักษณะของการจับวัตถุจากสัญญาณ Mechanomyogram (MMG) และภาพจากกล้องดิจิทัล [30]

Yuichi Motai และ Akio Kosaka (2008) นำเสนอเทคนิคการสอบเทียบมือหุ่นยนต์สำหรับจับวัตถุในพารามิเตอร์ต่างๆ ทั้งแกน XZY โดยใช้การทำงานควบคู่กันของกล้องดิจิทัลและมือจับของหุ่นยนต์ อีกทั้งยังใช้ตัวกรองคาลมาน (Kalman filter) ช่วยในการคำนวณความถูกต้องของตำแหน่งในการเคลื่อนที่ของแขนจับทั้ง 3 แกน ซึ่งได้ทำการติดตั้งกล้องเหนือแขนของมือจับเพื่อใช้บันทึกภาพตำแหน่งบนแผ่นสอบเทียบเสมือนการมองผ่านทางสายตาของมนุษย์ [31]

Sudeep Sundaram และ Walterio W. Mayol Cuevas (2009) นำเสนอกรอบการรับรู้กิจกรรมโดยสังเกตการกระทำและการจัดการของมือที่กระทำ โดยผู้ทดสอบใช้มือเปล่าควบคุมผ่านกล้องที่สวมใส่จากภาพความละเอียดต่ำ (160x120 พิกเซล) ทำงานร่วมกับ Bayesian Network เพื่ออนุมานตำแหน่ง วัตถุ ปฏิสัมพันธ์ เหตุการณ์ และกิจกรรม โดยการรวบรวมชุดลำดับกิจกรรมประจำวันที่มีขนาดใหญ่ของความสัมพันธ์ระหว่างกิจกรรมในแต่ละวัน เพื่อสำรวจวิธีการมองเห็นที่มีประสิทธิภาพ แต่ราคาไม่แพงเพื่อทำการอนุมานกิจกรรมที่ได้ทำ โดยไม่ต้องพึ่งการตรวจจับวัตถุหรือมือที่ซับซ้อน [32]

Alireza Fathi และคณะ (2011) ศึกษาการแก้ปัญหาการเรียนรู้การจับวัตถุแบบจำลองจากวิดีโอของกิจกรรมในครัวเรือนด้วยกล้องอีโก้เซ็นเซอร์ โดยใช้การเรียนรู้หลายอินสแตนซ์เพื่อจับคู่แต่ละอินสแตนซ์ของแต่ละวัตถุในลำดับต่าง ๆ เพื่อตรวจจับวัตถุในภาพโดยลดขั้นตอนการทำ Labeling ในการจดจำชนิดวัตถุ [33]

Hamed Pirsiavash และ Deva Ramanan (2012) นำเสนอฐานข้อมูลชุดใหม่และอัลกอริทึมใหม่สำหรับการตรวจจับและจดจำกิจกรรมในชีวิตประจำวัน (Activities of Daily Living) ด้วยกล้องระบบอีโก้เซ็นเซอร์ โดยรวบรวมชุดข้อมูล 1 ล้านภาพของอาสาสมัครจำนวนมากที่ทำกิจกรรมต่างๆ โดยไม่มีสคริปต์ ฐานข้อมูลมีคำอธิบายเพิ่มเติมกำกับประกอบพร้อมกิจกรรมติดตามวัตถุ ตำแหน่งมือ และเหตุการณ์การโต้ตอบ นำเสนอเทคนิค Temporal pyramids ทำงานร่วมกับแบบจำลองวัตถุแบบ Composite object models ทำให้เกิดการปรับปรุงประสิทธิภาพมากกว่าวิธีการเดิมถึงสองเท่า [34]

Giuseppe Serra และคณะ (2013) นำเสนออัลกอริทึมการแบ่งส่วนมือตามการจำแนกประเภทแบบ Random Forest Superpixel ซึ่งมีการประมวลผลทางแสง เวลาและพื้นที่อย่างต่อเนื่อง ในภาพวิดีโอระบบอีโก้เซ็นเซอร์ นอกจากนี้ยังเสนอวิธีการจดจำท่าทางมือด้วย Exemplar SVMs จากการแบ่งตัวอย่างเป็นชุดเล็ก ๆ ซึ่งให้ประสิทธิภาพดีที่สุดในขณะนั้นและปรับปรุงความแม่นยำในการจดจำท่าทางมือให้เหมาะนำไปใช้งานต่าง ๆ ของระบบอีโก้เซ็นเซอร์ [35]

Minjie Cai และคณะ (2015) นำเสนอครั้งแรกในการพัฒนาเทคนิคจำแนกการจับและควบคุมมือในการจัดการวัตถุในชีวิตประจำวันโดยบันทึกด้วยกล้อง RGB ที่สวมใส่บนศีรษะ โดยใช้การตรวจจับสีผิวมนุษย์ในการแบ่งแยกภาพเฉพาะในส่วนมือจับและใช้เทคนิค Histogram of Oriented Gradient (HOG) และ Scale Invariant Feature Transform (SIFT) ในการสกัดคุณลักษณะของภาพ ซึ่งได้ผลการทำงานที่น่าพึงพอใจกับคุณลักษณะตามแต่ละวัตถุและใช้การวิเคราะห์ข้อมูลเพื่อแบ่งแยกชนิดมือจับที่ได้จากกล้องดิจิทัลด้วยเทคนิค Principle Component Analysis (PCA) และ Support Vector Machine (SVM) [36]

Suriya Singh และคณะ (2016) นำเสนอระบบโครงข่ายประสาทแบบ CNN สำหรับการเรียนรู้และการจำแนกแบบ end to end ภาพกิจกรรมของผู้สวมใส่กล้องอีโก้เซ็นเซอร์ โดยตรวจจับท่ามือ การเคลื่อนไหวของศีรษะและจุดเด่นของภาพ โดยใช้ฐานข้อมูลสำหรับฝึกการเรียนรู้จากวิดีโอที่ทำ Labeling ในจำนวนน้อยได้ [37]

Minjie Cai และคณะ (2016) นำเสนอแบบจำลองประเภทของมือและคุณลักษณะของการจัดการวัตถุเพื่อให้รับรู้ การกระทำเมื่อเปลี่ยนแปลงรูปร่างของมือที่มีต่อวัตถุได้อย่างถูกต้อง โดยมุ่งเน้นการเข้าใจความสัมพันธ์ตามบริบทระหว่างประเภทและคุณลักษณะของวัตถุเพื่อกำหนดลักษณะการ

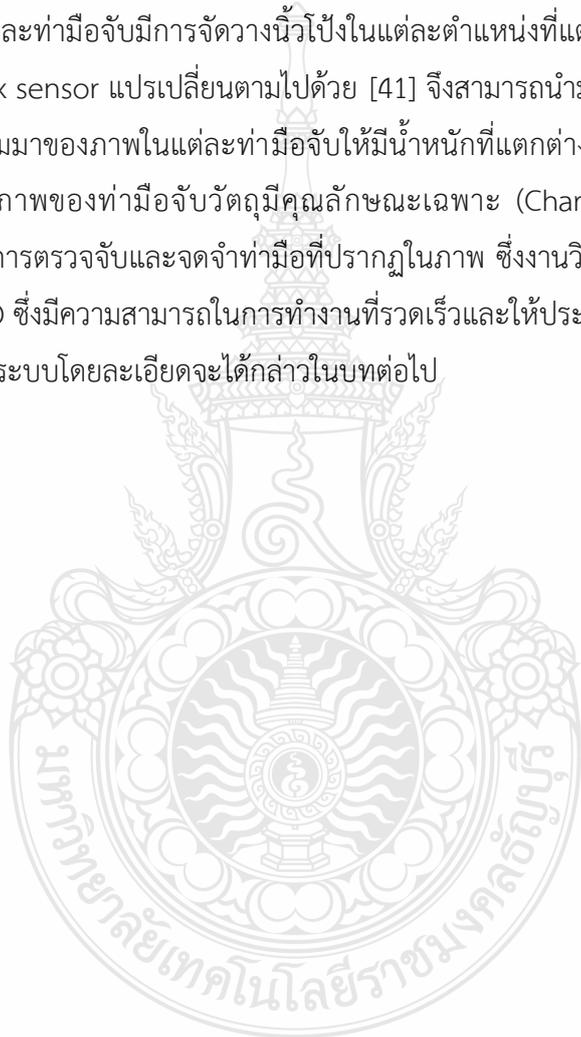
กระทำต่างๆ ของมือต่อวัตถุในท่าทางมือต่างๆ ในการทดลองใช้ท่ามือจับวัตถุจำนวน 9 ท่ามือ โดยทำงานร่วมกับโครงข่ายประสาทแบบ CNN [2]

Minjie Cai และคณะ (2017) ได้ปรับปรุงเทคนิคเพิ่มประสิทธิภาพการจดจำท่ามือจับแบบต่างๆ ในการจัดการต่อวัตถุ โดยบันทึกด้วยกล้อง RGB ที่สวมใส่บนศีรษะที่ได้เคยนำเสนอไว้ โดยการเพิ่มประสิทธิภาพตรวจจับภาพท่ามือจับที่ปรากฏในแต่ละเฟรมด้วยเทคนิค Dense Hand Trajectory (DHT) ในส่วนการสกัดคุณลักษณะของภาพท่ามือจับจากที่เคยใช้แต่เทคนิค HOG และ SIFT และจะมีการใช้โครงข่ายประสาทแบบ CNN ในการจดจำท่ามือ โดยการสอนการเรียนรู้จากภาพท่ามือจับของฐานข้อมูล จึงทำให้สามารถนำการทำงานของระบบไปใช้ในสิ่งแวดล้อมจริงได้ [38]

Meng Wang และคณะ (2018) นำเสนอการตรวจจับและจดจำกิจกรรมของมนุษย์ในการจัดการวัตถุผ่านการมองเห็นด้วยกล้องอีโก้เซ็นเซอร์วีชัน โดยการผสมผสานการสกัดคุณลักษณะของคุณสมบัติวัตถุและการเคลื่อนที่ ด้วยเทคนิคใหม่ Manipulated Object Histogram (MOH) ซึ่งใช้เทคนิคการไหลของแสง (Optical Flow) ร่วมกับโครงข่ายประสาทแบบ CNN ตรวจจับการเคลื่อนที่ของมือและวัตถุขณะทำกิจกรรม โดยสามารถแก้ไขปัญหาสัญญาณรบกวนที่เกิดจากการเคลื่อนไหวของกล้องและเพิ่มประสิทธิภาพในการจดจำกิจกรรมของมนุษย์ [39]

จากงานวิจัยที่ผ่านมาของระบบกล้องอีโก้เซ็นเซอร์วีชันจะเห็นได้ว่ามีการศึกษาและพัฒนาการนำไปใช้งานอย่างหลากหลาย เนื่องจากระบบกล้องอีโก้เซ็นเซอร์วีชันมีความน่าสนใจและข้อดีหลายประการดังที่ได้กล่าวไว้ในหัวข้อ 2.3 โดยงานวิจัยมีจำนวนเพิ่มขึ้นอย่างรวดเร็ว หลังจากที่มีการผลิตกล้องหรืออุปกรณ์ FPV ออกมาสู่ขายในเชิงพาณิชย์จากโรงงานผลิตต่างๆ เช่น Google GoPro Microsoft เป็นต้น โดยในงานวิจัยที่มีการศึกษาจะแบ่งเป็นได้ 2 ประเภท คือ การจำกิจกรรมต่าง ๆ ในกิจวัตรของผู้สวมใส่กล้องกับการตรวจจับและจดจำการจัดการวัตถุด้วยมือในขณะทำกิจกรรม โดยในงานวิจัยนี้จะดำเนินการวิจัยเพื่อศึกษารูปแบบมือจับวัตถุและลักษณะรูปร่างของวัตถุในระบบอีโก้เซ็นเซอร์วีชัน ซึ่งสามารถนำไปสู่การพัฒนาได้หลายด้านตั้งแต่การแพทย์ การออกแบบผลิตภัณฑ์ ระบบกลไกหุ่นยนต์ ระบบโลกเสมือนจริง เป็นต้น เนื่องจากมือของมนุษย์มีการเปลี่ยนแปลงรูปร่างและเคลื่อนไหวตลอดเวลาทำให้เป็นสิ่งที่ยากต่อการตรวจจับและเมื่อต้องการตรวจจับและจดจำท่ามือด้วยกล้อง FPV ซึ่งส่วนใหญ่กล้องจะติดที่บริเวณศีรษะของผู้สวมใส่ ทำให้ภาพค่อนข้างมีการเคลื่อนไหวที่รวดเร็วส่งผลให้คุณภาพของภาพที่บันทึกได้อาจจะไม่ดีเท่าที่ควรเป็นผลทำให้ประสิทธิภาพในการตรวจจับและจดจำได้ไม่ได้สูงมาก [2][38] และอีกหนึ่งปัจจัยสำคัญที่ทำให้การตรวจจับและจดจำท่ามือจับวัตถุมีประสิทธิภาพลดลง คือ การเปลี่ยนแปลงความเข้มของแสงสว่างของสภาพแวดล้อม ซึ่งมีผลต่อการตรวจจับมือที่ปรากฏในภาพ การจำแนกนิ้วมือที่จัดเรียงในขณะจับวัตถุรูปแบบต่างๆ มีความยากลำบากมากขึ้น งานวิจัยนี้จึงได้นำเสนอเทคนิคใหม่ในการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือในสภาวะที่มี

ความสว่างน้อยหรือในเวลากลางคืน ซึ่งจากการศึกษางานวิจัยในระบบอีโก้เซ็นเซอร์ที่ผ่านมายังไม่มีการแก้ปัญหาในประเด็นนี้มาก่อนโดยระบบที่นำเสนอเป็นการทำงานร่วมกันของการปรับค่าแกมมาของภาพด้วยเทคนิค Adaptive Gamma Correction with Weighting Distribution (AGCWD) [13] และการใช้คุณสมบัติอุปกรณ์เซนเซอร์ที่มีการเปลี่ยนแปลงความต้านทานตามการโค้งงอของอุปกรณ์ โดยการนำความต้านทานแบบโค้งงอไปติดบริเวณนิ้วมือหัวแม่มือหรือนิ้วโป้ง (Thump) ขณะจับวัตถุ ซึ่งในการจับวัตถุพบว่านิ้วโป้งทำหน้าที่หลักในการใช้จัดการวัตถุ [40] และมีการใช้นิ้วโป้งในการจับในทุก ๆ ท่ามือจับวัตถุ ซึ่งในแต่ละท่ามือจับมีการจัดวางนิ้วโป้งในแต่ละตำแหน่งที่แตกต่างกัน ทำให้ได้สัญญาณเอาต์พุตของวงจร Flex sensor แปรเปลี่ยนตามไปด้วย [41] จึงสามารถนำมาใช้เป็นพารามิเตอร์ในการกำหนดการปรับค่าแกมมาของภาพในแต่ละท่ามือจับให้มีน้ำหนักที่แตกต่างกัน ด้วยวิธีการของเทคนิค AGCWD ทำให้แต่ละภาพของท่ามือจับวัตถุมีคุณลักษณะเฉพาะ (Characteristic) แล้วส่งต่อภาพดังกล่าวให้กับขั้นตอนการตรวจจับและจดจำท่ามือที่ปรากฏในภาพ ซึ่งงานวิจัยนี้ได้ใช้เทคนิคการเรียนรู้เชิงลึกที่เรียกว่า YOLO ซึ่งมีความสามารถในการทำงานที่รวดเร็วและให้ประสิทธิภาพในการทำงานที่สูง ซึ่งหลักการทำงานของระบบโดยละเอียดจะได้กล่าวในบทต่อไป



บทที่ 3

วิธีดำเนินการวิจัย

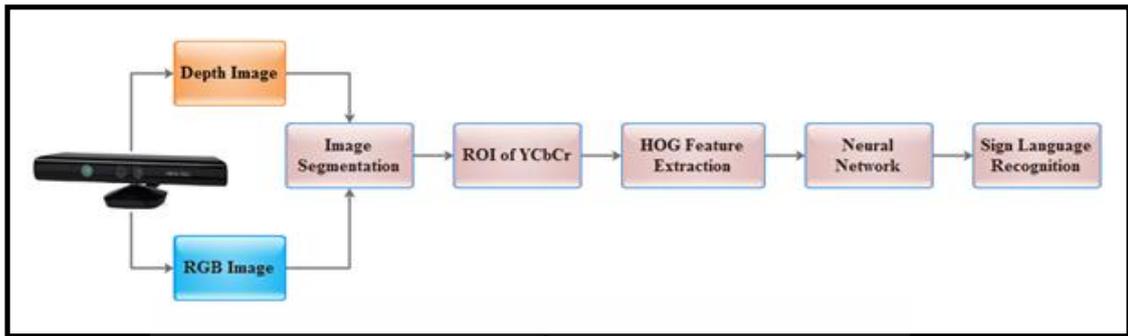
ในบทนี้จะกล่าวถึงขั้นตอนและวิธีดำเนินการวิจัยทั้งหมด ซึ่งประกอบด้วย การออกแบบระบบประมวลผลภาพเบื้องต้น การออกแบบระบบการทำงานในภาพรวม การเก็บรวบรวมข้อมูล การออกแบบวงจรความต้านทานแบบโค้งงอ การเก็บรวบรวมฐานข้อมูลภาพและสัญญาณวงจรความต้านทานแบบโค้งงอ การออกแบบอัลกอริทึมสำหรับการตรวจจับและจดจำท่ามือในระบบอ็อบเจกต์เซ็นเซอร์วิชั่นในสภาวะความเข้มของแสงสว่างต่ำและหัวข้อสุดท้ายจะเป็นวิธีการทดสอบงานวิจัย

3.1 การออกแบบระบบประมวลผลภาพเบื้องต้น

การประมวลผลภาพดิจิทัลเป็นการแปลงข้อมูลภาพให้อยู่ในรูปแบบข้อมูลดิจิทัลหรือข้อมูลเชิงตัวเลข โดยวัตถุประสงค์ของการประมวลผลภาพ คือการปรับปรุงคุณภาพของภาพเพื่อให้มนุษย์ ซึ่งงานวิจัยนี้ได้เริ่มศึกษาและทดลองการประมวลผลภาพจากข้อมูลกล้องดิจิทัลเพื่อเป็นพื้นฐานในงานวิจัยต่อไป จำนวน 3 หัวข้อ ดังต่อไปนี้

3.1.1 การตรวจจับและจดจำท่าภาษามือ

งานนี้นำเสนอการพัฒนาการจดจำภาษามือสำหรับการสะกดภาษามือไทย [42] โดยการทำงานของระบบเป็นดังรูปที่ 3.1 การทำงานใช้กล้องถ่ายภาพความลึกของ Microsoft Kinect ที่อินพุตวิดีโอภาพ 30 เฟรมต่อวินาที แต่ละเฟรมประกอบด้วยภาพความลึก (Depth) และสี RGB ซึ่งมีขนาด 640x480 พิกเซล แล้วได้ปรับปรุงคุณภาพก่อนที่จะประมวลผลทั้งภาพความลึกและภาพสี ดังนั้นตำแหน่งมือในภาพ RGB จะถูกใช้เป็นข้อมูลตำแหน่งในภาพความลึก โดยใช้ความแตกต่างของความสว่างในภาพความลึก ซึ่งภาพสี RGB จะถูกเลือกเอาเฉพาะในพื้นที่ Region of Interest (ROI) ที่ขนาด 160x160 พิกเซล ซึ่งเป็นเพียงส่วนพื้นที่มือเท่านั้น จากนั้นพื้นที่มือนั้นจะถูกแปลงเป็น YCbCr ซึ่งมีคุณสมบัติทนทานในสภาวะแสงสว่างที่เปลี่ยนแปลงไป [43] จากนั้นจึงใช้เทคนิค Histograms of Oriented Gradients (HOG) [44] สำหรับการดึงคุณลักษณะของภาพเพื่อให้มีความทนทานในการเปลี่ยนแปลงความสว่างเช่นกัน โดยเทคนิค HOG คำนวณสกัดคุณลักษณะของการไล่ระดับสีของภาพและพื้นสีภาพ ความคมชัดของขอบภาพและและรูปร่างของภาพ ซึ่งให้ความต้านทานต่อความแปรผันของแสงสว่าง ผลที่ได้จาก HOG นี้จะได้รับการดึงคุณลักษณะสำหรับฝึกการเรียนรู้การจำแนกประเภทด้วยโครงข่ายประสาทเทียม (Artificial Neural Networks)



รูปที่ 3.1 ขั้นตอนการตรวจจับและจดจำท่าภาษามือด้วยกล้อง Microsoft Kinect

กล้อง Kinect มีกล้องสองตัว คือ ตัวแรกสำหรับถ่ายภาพสี RGB และอีกตัวถ่ายภาพอินฟราเรดสำหรับภาพความลึก ซึ่งภาพสีและความลึกที่ได้จากกล้องเป็นดังตัวอย่างรูปที่ 3.2 ภาพสีและภาพความลึกไม่ได้ซ้อนทับกันอย่างสมบูรณ์ตามพิกเซล เนื่องด้วยมุมกล้องที่วางห่างกันเล็กน้อย ภาพถ่ายจะมีพิกัดที่เริ่มต้นที่สอดคล้องกันในภาพความลึก (0,0) และภาพสี (640,480) ทั้งสองภาพที่มีขนาดภาพ 1280x480 พิกเซล ตำแหน่งวัตถุใช้อ้างอิงเพื่อค้นหาความต่างระหว่างคอลัมน์และแถวในภาพความลึกและภาพสี โดยพิกัดตำแหน่งจุดอ้างอิงของภาพความลึกจะน้อยกว่าภาพสีเพียงเล็กน้อย ซึ่งค่านี้จะเป็นดัชนีการชดเชยเพื่อใช้ในการแก้ไขตำแหน่งในภาพสีและภาพความลึกให้มีตำแหน่งที่ตรงกันจากนั้นทำการตรวจจับตำแหน่งส่วนมือที่ปรากฏในภาพ โดยใช้ภาพความลึกที่สัมพันธ์กับระยะห่างจากมือถึงตัวกล้อง ซึ่งวัตถุที่อยู่ใกล้กล้องจะมีความสว่างมากกว่าวัตถุที่อยู่ไกลและความสว่างจะน้อยลงเมื่อวัตถุอยู่ไกลออกไป คุณสมบัตินี้ทำให้สามารถระบุตำแหน่งมือได้ เนื่องจากมือจะอยู่ข้างหน้าฉากเสมอ ทำให้เกิดความสว่างสูงสุดในภาพความลึก จากนั้นใช้ในเทคนิคการตรวจจับ Blob Analysis สำหรับการกำหนดขอบเขต ROI เพื่อกำหนดตำแหน่งของมือ โดยศูนย์กลางของมือจะถูกคำนวณเพื่อจะได้ค่าตำแหน่งมือในภาพสี โดยจะถูกรวมเข้ากับดัชนีการชดเชยเพื่อใช้คำนวณตำแหน่งของมือในภาพสีที่ต้องการ หลังจากการทดลอง พบว่า เทคนิคนี้สามารถตรวจจับภาพสัญลักษณ์มือที่ปรากฏในแต่ละเฟรมได้เป็นอย่างดี แม้ว่าจะมีพื้นหลังที่ซับซ้อนหลากหลายสี แต่ก็สามารถให้ผลการตรวจจับที่ต้องการ

ในการทดลองใช้ฐานข้อมูลประกอบด้วยภาพท่าทางมือ 720 ภาพ สำหรับชุดฝึกการเรียนรู้ซึ่งมี 24 รูปแบบท่ามือ ซึ่งใช้เป็นท่าภาษามือสะกดคำภาษาไทย ดังตัวอย่างในภาพที่ 3.2 ภาพมือที่ตรวจจับได้จะนำไปแยกคุณลักษณะด้วยเทคนิค HOG ซึ่งเทคนิคนี้จะให้ค่าคุณลักษณะเวกเตอร์ของภาพ ในท่าทางมือแต่ละท่าแตกต่างกันไปตามการจัดเรียงของนิ้วและทิศทางของเวกเตอร์ในส่วน ROI ของภาพ และค่าการกระจายตัวจะถูกสร้างขึ้นเป็นแผนภาพโพลาร์ ซึ่งมีรูปแบบที่แตกต่างกันในแต่ละท่าทางมือ ค่าเวกเตอร์จากการแยกคุณลักษณะด้วยเทคนิค HOG จะถูกใช้เป็นอินพุตในการจดจำด้วยเทคนิคโครงข่ายประสาทเทียมสำหรับการประมวลผลท่าทางมือในแต่ละเฟรม



รูปที่ 3.2 ลำดับภาพที่ใช้ในการตรวจจับท่าทางมือด้วยกล้อง Microsoft Kinect

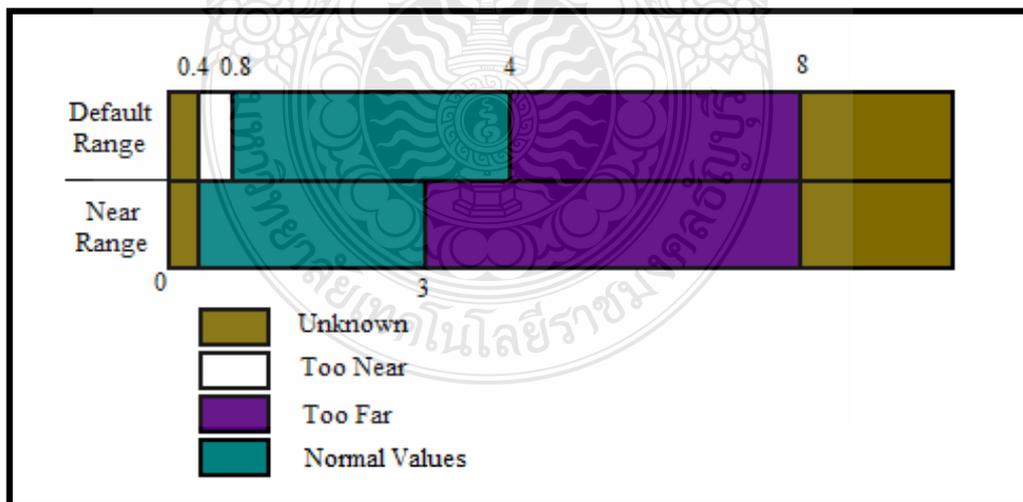
3.1.2 ความถูกต้องและน่าเชื่อถือของกล้อง Microsoft Kinect

การทำงานของกล้อง Microsoft Kinect ภาพที่บันทึกจากกล้องถ่ายภาพความลึกประกอบด้วยข้อมูลที่เกี่ยวข้องกับระยะห่างระหว่างกล้องกับวัตถุในฉาก วัตถุที่อยู่ใกล้กล้องจะมีความสว่างมากกว่าวัตถุที่อยู่ไกลกล้องออกไป กล้องถ่ายภาพความลึกประกอบด้วยเซนเซอร์ CMOS แบบขาวดำและเครื่องฉายแสงอินฟราเรดที่มีโครงสร้างเพื่อใช้คำนวณความลึกของฉาก ทำให้ได้ภาพความลึกซึ่งช่วยให้สามารถค้นหาตำแหน่งมือในฉากได้ดังการทดลองในหัวข้อ 3.1.1 คุณสมบัตินี้ทำให้สามารถระบุตำแหน่งของมือในภาพสี RGB ได้ เนื่องจากมือจะอยู่ข้างหน้าฉากเสมอ แต่อย่างไรก็ตาม ยังจำเป็นต้องหาความน่าเชื่อถือและความแม่นยำของระยะทางที่เหมาะสมในการใช้งานกล้อง Kinect ในการตรวจจับและจดจำท่าทางมือสะกดนิ้วไทยที่เป็นเฉพาะท่ามือเดียวจำนวน 16 ท่ามือ ดังในรูปที่ 3.3



รูปที่ 3.3 ท่ามือสะกดนิ้วไทยที่เป็นท่ามือเดียวใช้ในการทดลอง

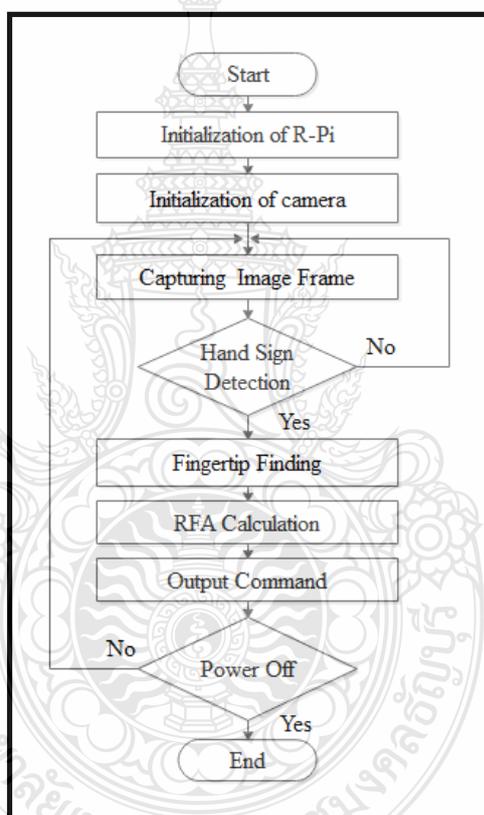
งานวิจัยนี้จึงออกแบบเพื่อใช้ทดสอบการทำงานของกล้องถ่ายภาพความลึกในการตรวจจับท่าทางมือ เทคนิคที่เป็นจุดเด่นของงานที่นำเสนอนี้ คือ ไม่มีขั้นตอนการสอนเรียนรู้ (Supervise learning) เพื่อลดเวลาในการคำนวณ การตรวจจับมือใช้การทำงานร่วมกันของภาพสี RGB และภาพความลึก เพื่อตรวจจับในส่วน ROI ของมือซึ่งจะอยู่เบื้องหน้าของภาพเสมอ ดังนั้นเมื่อข้อมูลภาพความลึกที่ได้รับการประมวลผลร่วมกับข้อมูลภาพสี RGB ก็จะสามารถจัดพื้นที่หลังของภาพสีและส่วนต่าง ๆ ของร่างกายได้ ซึ่งระบบที่นำเสนอนี้ทนทานต่อสภาพแวดล้อมที่ซับซ้อน หลังจากตรวจพบตำแหน่งมือในภาพสี RGB จะใช้เทคนิค HOG เพื่อแยกคุณลักษณะของภาพมือ เนื่องจากเป็นอัลกอริทึมให้ผลลัพธ์ที่ดีกับการตรวจจับสีผิวมนุษย์ [44] และจะส่งต่อคุณลักษณะเหล่านี้ให้กับอินพุตของโครงข่ายประสาทเทียมเพื่อดำเนินการประมวลผลการจดจำท่ามือที่ปรากฏในภาพ หลักการทำงานของกล้อง Microsoft Kinect มีอุปกรณ์ออปติคัลแบบ Diffractive ร่วมกับเลเซอร์อินฟราเรดในการสร้างรูปแบบจุดสำหรับฉายขนาด 2 เมกะพิกเซล แล้วกำหนดความแตกต่างระหว่างจุดแสงอินฟราเรดที่ปล่อยออกมาจากตำแหน่งอ้างอิง โดยแสงอินฟราเรดนี้มี 640 จุดในแนวนอนและ 480 จุดในแนวตั้ง และระยะห่างกัน 3 มิลลิเมตรที่จากแหล่งกำเนิดแสง 2 เมตร โดยมีโหมดการทำงานวัดความลึกของกล้องจำนวนสองช่วง คือ ช่วงปกติทั่วไปและระยะใกล้ ดังรูปที่ 3.4 การทดลองเพื่อประเมินความถูกต้องของการรู้จำภาษามือไทย อาสาสมัครจะสวมเสื้อยืดแขนสั้นและยืนอยู่หน้ากล้อง Kinect ทำการสะกดท่าภาษามือไทยในฉากพื้นหลังที่ซับซ้อนมีสีหลากหลายและมีการทำท่าทาง 16 ท่ามือ โดยทำการทดสอบระยะห่างระหว่างกล้อง Kinect ตั้งแต่ 0.6 ถึง 2.8 เมตร โดยมีการเพิ่มขึ้นละ 0.2 เมตร รวมระยะทั้งหมด 12 ช่วง



รูปที่ 3.4 ระยะการทำงานของกล้องถ่ายภาพความลึกทั้งสองโหมด

3.1.3 การควบคุมอุปกรณ์ด้วยท่าทางมือ

ระบบควบคุมอุปกรณ์ด้วยท่าทางมือทำงานโดยใช้กล้องถ่ายภาพสี RGB และ Raspberry pi เพื่อตรวจจับท่ามือของผู้สั่งการควบคุมซึ่งอยู่ที่หน้ากล้อง ระบบจะแปลความหมายคำสั่งที่ตรวจพบแล้วส่งออกไปยังพอร์ต GPIO ของ Raspberry Pi ซึ่งไดอะแกรมการทำงานและอัลกอริทึมของโปรแกรมเป็นดังรูปที่ 3.5 หลังจากระบบเริ่มทำงานโดยการเปิด Raspberry Pi 3 รุ่น B ซึ่งเป็นคอมพิวเตอร์ขนาดเล็กจะส่งงานผ่านโปรโตคอลของเครื่องและส่งข้อมูลคำสั่งการไปยังกล้องที่ติดตั้งไว้ โดยเป็นกล้องรุ่น Raspberry Pi Camera V2 ผ่านพอร์ตเชื่อมต่อ CSI กล้องจะส่งข้อความเตรียมพร้อมไปที่ Raspberry pi แล้วระบบจึงเริ่มต้นทำงาน ภาพแต่ละเฟรมมีขนาดภาพ 640x480 พิกเซล



รูปที่ 3.5 ขั้นตอนการทำงานของระบบควบคุมอุปกรณ์ด้วยท่าทางมือ

โดยโปรแกรมจะทำการประมวลผลภาพตรวจจับท่ามือที่ปรากฏ ซึ่งใช้เทคนิคการตรวจจับสีผิวมนุษย์ควบคู่ไปกับเทคนิคการไหลของแสง (Optical Flow) เพื่อค้นหาตำแหน่งและเครื่องหมายมือที่ปรากฏในแต่ละเฟรม การตรวจจับสีผิวต้องแปลงภาพจากระบบสี RGB ที่ได้รับจากกล้องไปสู่ระบบสี YCbCr ซึ่งเหมาะสมสำหรับแบ่งส่วนสีผิวมนุษย์ในภาพ เนื่องจากทนทานต่อการเปลี่ยนแปลงแสงสว่างในภาพแวดล้อมได้ดีกว่าระบบสี RGB ส่วนเทคนิคการไหลของแสงเป็นการคำนวณความเข้มของแสงที่

เปลี่ยนไประหว่างสองเฟรมติดต่อกัน โดยคำนวณการเคลื่อนที่ของพิกเซลที่มีค่าความเข้มของแสงคงที่ เมื่อเทียบกับพิกเซลรอบข้างด้วยเทคนิค Dense Optical Flow [45] หลังจากได้ตำแหน่งและภาพของมือจากทั้งสองเทคนิคที่ทำงานร่วมกันแล้วจะคำนวณหาเฉพาะเครื่องหมายของมือและลบภาพส่วนอื่นๆ ออกไป โดยใช้เทคนิคการคำนวณรูปทรงของท่ามือที่ได้จากภาพไบนารีเพื่อหาตำแหน่งปลายนิ้ว แล้วใช้เทคนิค Radian Fingertip Analysis (RFA) ที่นำเสนอในงานวิจัยนี้ในการจดจำท่าทางของมือ

หลังจากได้ภาพไบนารีแล้วจะต้องทำการคำนวณหาจำนวนนิ้วมือ, ตำแหน่งศูนย์กลางฝ่ามือ, ระยะทางระหว่างปลายนิ้วกับศูนย์กลางฝ่ามือและตำแหน่งตรงกลางข้อมือในภาพที่ตรวจจับได้เพื่อใช้สำหรับจดจำท่ามืองดรูปที่ 3.6 ซึ่งมีขั้นตอนกระบวนการทำงานดังนี้

1. คำนวณหาตำแหน่งจุดศูนย์กลางฝ่ามือ CH และปลายนิ้วมือทั้งหมดที่มีในภาพ โดยใช้การหา Convexity Defects ของปลายนิ้วมือ พร้อมด้วยการคำนวณหาพื้นที่และความยาวของรัศมีฝ่ามือ R

2. คำนวณหาระยะทางของแต่ละปลายนิ้วมือที่ตรวจจับได้จากตำแหน่งศูนย์กลางฝ่ามือ CH โดยใช้ตำแหน่งพิกเซลของจุดเริ่มต้นในแต่ละ Convexity (p_s) ของปลายนิ้วกับตำแหน่งของศูนย์กลางฝ่ามือ CH แล้วทำการหารด้วยรัศมีของฝ่ามือ R เรียกว่า ระยะทาง Normalize แทนด้วย D_{nor} ดังสมการที่ 3.1 ซึ่งแต่ละนิ้วจะต้องมีค่ามากกว่า 1.9, $D_{nor} > 1.9$

$$D_{nor} = \frac{\sqrt{(x_{ps}-x_{CH})^2+(y_{ps}-y_{CH})^2}}{R} \quad (3.1)$$

3. คำนวณหาตำแหน่งกลางข้อมือ CW เนื่องจากข้อมือเป็นส่วนที่จะอยู่ในทิศทางตรงกันข้ามกับนิ้วมือและจะอยู่ด้านล่างของภาพเสมอ โดยจะคำนวณจาก 1.3 เท่าของระยะรัศมีฝ่ามือ R และหาตำแหน่งสี่ทางด้านซ้ายสุดและขวาสุดของข้อมือ ซึ่งตรงกลางข้อมือก็คือจุดตรงกลางของตำแหน่งทั้งสอง

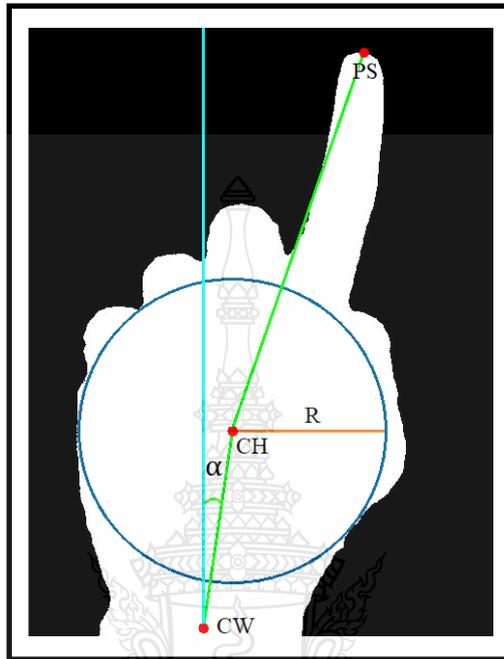
4. คำนวณหามุมเอียงของฝ่ามือ โดยใช้เส้นตรงระหว่างจุดตรงกลางข้อมือ CW กับจุดศูนย์กลางฝ่ามือ CH เป็น y_t และ x_t แทนด้วยค่า α_{rad}

$$\alpha_{rad} = \tan^{-1} \left(\frac{y_t}{x_t} \right) \quad (3.2)$$

5. คำนวณหามุมเรเดียสของปลายนิ้วมือแต่ละนิ้ว โดยใช้เส้นตรงระหว่างจุดเริ่มต้นของ Convexity (p_s) กับจุดศูนย์กลางฝ่ามือ CH เป็น y_p และ x_p และรวมเข้ากับมุมเอียงของฝ่ามือ แทนด้วยค่า θ_{rad}

$$\theta_{rad} = \tan^{-1} \left(\frac{y_p}{x_p} \right) + \alpha_{rad} \quad (3.3)$$

6. คำนวณหาชื่อนิ้วมือในภาพจากระยะทาง D_{nor} และมุม θ_{rad} เพื่อประมวลผลท่ามือที่ปรากฏในภาพที่ตรวจจับได้



รูปที่ 3.6 จุดตำแหน่งบนฝ่ามือที่ใช้ในการคำนวณตรวจจับท่ามือ

3.2 วงจรความต้านทานแบบโค้งงอ

อุปกรณ์ความต้านทานแบบโค้งงอมีคุณสมบัติความต้านทานเปลี่ยนแปลงตามการงอตัวของอุปกรณ์ดังได้กล่าวไว้ในหัวข้อที่ 2.4 ซึ่งในส่วนหัวข้อนี้จะกล่าวถึงการออกแบบนำมาใช้ในงานวิจัยนี้ โดยมีรายละเอียดดังนี้

3.2.1 การออกแบบวงจร

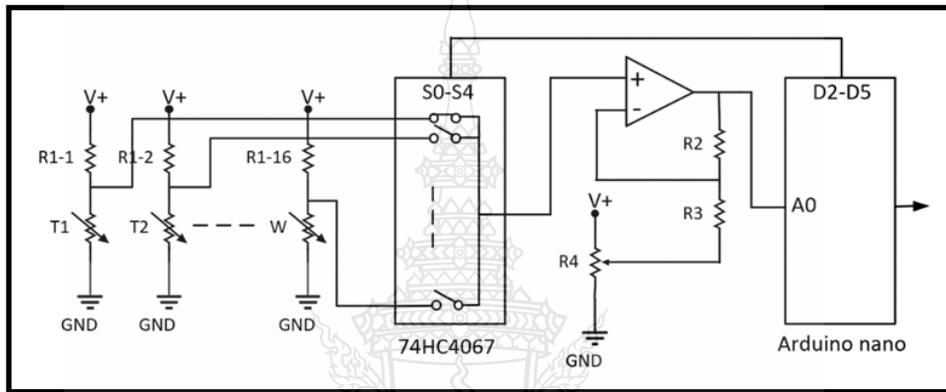
โครงสร้างความต้านทานแบบโค้งงอเป็นหมึกพิมพ์นำไฟฟ้าถูกพิมพ์อยู่บนแผ่นฉนวนรองรับแบบยืดหยุ่น ผงหมึกพิมพ์มีคุณสมบัตินำไฟฟ้า โดยถ้ามีการงอตัวที่มุม 45° ความต้านทานจะอยู่ที่ประมาณ $75 \text{ k}\Omega$ ที่มุม 90° ความต้านทานนี้จะอยู่ที่ประมาณ $125 \text{ k}\Omega$ เมื่อเซนเซอร์ถูกปรับให้อยู่ในแนวราบอีกครั้ง ความต้านทานจะกลับสู่ค่าเดิม ดังนั้นเราจึงสามารถวัดมุมงอได้โดยดูจากค่าความต้านทานที่เกิดขึ้น ในทางทฤษฎี ค่าของ R_s ($\text{k}\Omega$) ในวงจร Operational Amplifier ที่ใช้งานสัมพันธ์กับมุมที่เกิดขึ้นดังสมการที่ 3.4

$$R_s = 100 \cdot \frac{\theta}{180} + 25. \quad (3.4)$$

ส่วนวงจรบัฟเฟอร์แบบปรับได้ (Adjustable Buffer) ที่ใช้งานสามารถเพิ่มโพเทนชิโอมิเตอร์ (Potentiometer) ลงในวงจรเพื่อปรับช่วงแรงดันขาออก V_{out} โดยสามารถหาค่าได้จากสมการที่ 3.5

$$V_{out} = \left(\frac{R_5}{R_5 + R_1} V \right) \cdot \left(1 + \frac{R_2}{R_3 + R_4} \right) - V_{R_4} \quad (3.5)$$

โดยที่ R_1 คือความต้านทานไบอัสให้กับความต้านทานแบบโค้งงอมีค่าคงที่และ R_5 คือความต้านทานไฟฟ้าผันแปร ซึ่งเปลี่ยนค่าตามการโค้งงอที่ทำให้เกิดความเครียดแรงดันตกคร่อม R_5 เป็นแรงดันขาเข้า V_{in} ส่งต่อไปที่วงจรขยาย



รูปที่ 3.7 วงจรมัลติเพล็กซ์สำหรับอ่านค่าสัญญาณจากความต้านทานแบบโค้งงอ

ดังนั้นจากสมการที่ 3.4 และ 3.5 แสดงถึงความสัมพันธ์เชิงเส้นตรงระหว่างมุม θ และแรงดันแรงดันขาออก V_{out} โดยการติดตั้งตำแหน่งการอ่านค่าการเปลี่ยนแปลงแรงดันขาออก V_{out} จากความต้านทานแบบโค้งงอ แต่ละตัวใช้ IC Multiplex เบอร์ 74HC4067 ดังรูปที่ 3.7 เพื่อเลือกตำแหน่งของสัญญาณที่ต้องการอ่านค่าเอาต์พุตจาก Multiplexer เข้าสู่วงจรบัฟเฟอร์แบบปรับค่าได้ ซึ่งได้เพิ่ม Potentiometer (R_4) ลงไปเพื่อปรับช่วงความไว (Sensitivity) ได้ จากนั้นสัญญาณที่ออกจากวงจรบัฟเฟอร์ถูกวัดค่าแรงดันไฟฟ้าโดยใช้ตัวแปลง Analog to Digital (A/D) ที่มีอยู่ในอุปกรณ์ Arduino microcontroller สัญญาณที่ได้ถูกบันทึกเพื่อนำไปทดสอบในลำดับถัดไป เนื่องจากมีคุณสมบัติที่โค้งงอได้ของความต้านทานแบบโค้งงอ จึงสามารถนำมาติดตั้งบริเวณจุดหมุนที่ข้อมือได้ ซึ่งขณะการจับวัตถุหรือจัดการของมือ เช่น เซอร์จะมีการโค้งงอรองรับกับการจัดเรียงข้อมือไปตามรูปทรงของวัตถุทำให้เกิดการเปลี่ยนแปลงความต้านทาน ซึ่งจะมีความเฉพาะตัวแตกต่างกันไปตามแต่ละท่ามือ ในงานวิจัยนี้ได้ใช้คุณสมบัติดังกล่าวใน 2 หัวข้อการทดลอง คือ การสร้างถุงมือเซนเซอร์เพื่อจดจำท่ามือจากถุงมือแบบ 16 DoF (Degree of Freedom) และการใช้ความต้านทานแบบโค้งงอติดตั้งที่นิ้วโป้งเพื่อใช้การเปลี่ยนแปลงของสัญญาณเอาต์พุตไปเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุในท่ามือต่าง ๆ

3.2.2 การพัฒนาถุงมือเซนเซอร์แบบ 16 DoF

การพัฒนาถุงมือเซนเซอร์ในงานวิจัยนี้ได้นำความต้านทานแบบโค้งงอติดตั้งบริเวณจุดหมุนที่ข้อนิ้วมือและข้อมือรวมจำนวน 16 DoF ซึ่งขณะการจับวัตถุหรือจัดการของมือ เซนเซอร์จะมีการโค้งตัวรองรับกับการจัดเรียงนิ้วมือไปตามรูปทรงของวัตถุ ซึ่งจะมีความเฉพาะตัวแตกต่างกันไปตามแต่ละท่ามือ ถุงมือเซนเซอร์ที่พัฒนาเพื่อใช้ในการทดลองในงานวิจัยนี้เป็นดังรูปที่ 3.8



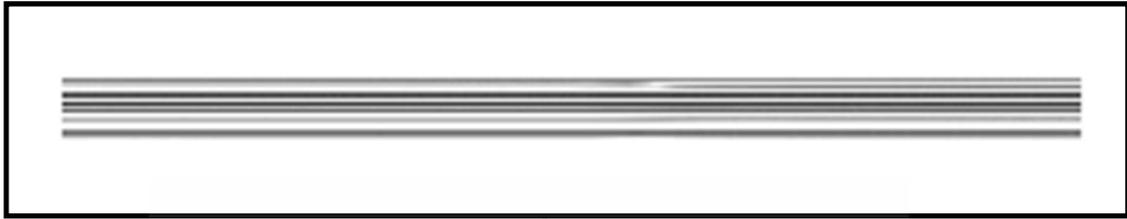
รูปที่ 3.8 ถุงมือที่ติดตั้งความต้านทานแบบโค้งงอแบบ 16 DoF

หลังจากการจับวัตถุในท่ามือต่าง ๆ ก็จะได้สัญญาณเอาต์พุตซึ่งจะแปลง A/D บันทึกข้อมูลของเซนเซอร์ในแต่ละตัวจำนวน 16 แถว ทำการจัดเรียงเป็นเมทริกซ์ ดังสมการที่ 3.6 ซึ่งในงานวิจัยนี้จะใช้ 2 เทคนิคในการแปลงข้อมูลสัญญาณดังกล่าวเป็นรูปภาพ

$$I = \begin{bmatrix} T_1(1) & T_1(2) & T_1(3) & \dots & T_1(1200) \\ T_2(1) & T_2(2) & T_2(3) & \dots & T_2(1200) \\ T_3(1) & T_3(2) & T_3(3) & \dots & T_3(1200) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W(1) & W(2) & W(3) & \dots & W(1200) \end{bmatrix} \quad (3.6)$$

เมื่อจัดเรียงเมทริกซ์ข้อมูลที่ได้จากเซนเซอร์เป็นข้อมูลภาพจะได้ดังรูปที่ 3.9 โดยแต่ละแถวจะเป็นตำแหน่งของเซนเซอร์ ซึ่งมีจำนวน 16 แถวและในแนวคอลัมน์จะเป็นข้อมูลที่เกิดขึ้นตามเวลาที่บันทึก 0-10 วินาที จำนวน 1200 คอลัมน์ หลังจากนั้นใช้เทคนิค Nearest interpolation และ

Gramian Angular Fields (GAF) [56] ในการแปลงสัญญาณที่ได้เป็นภาพเพื่อใช้สำหรับทดสอบด้วยโครงข่ายประสาทแบบ CNN ต่อไป



รูปที่ 3.9 รูปจากการจัดเรียงเมทริกซ์ข้อมูลขนาด 16x1200

3.2.2.1 การแปลงด้วยวิธี Nearest interpolation เนื่องจากภาพที่ได้จากการแปลงด้วยวิธี Nearest interpolation จะไม่ทำให้เกิดการรวมกันของข้อมูลในแต่ละช่องสัญญาณ ภาพที่ขยายจึงยังคงมี 16 ช่อง แต่มีขนาด $227 \times 227 \times 3$ พิกเซล เป็นไปตามข้อกำหนดของ AlexNet โดยวิธีการแปลงขนาดภาพด้วยวิธี Nearest interpolation เป็นดังสมการที่ 3.7

$$I^{(x,y)} = I(\text{round}\{f_x^{-1}(x,y)\}\text{round}\{f_y^{-1}(x,y)\}) \quad (3.7)$$

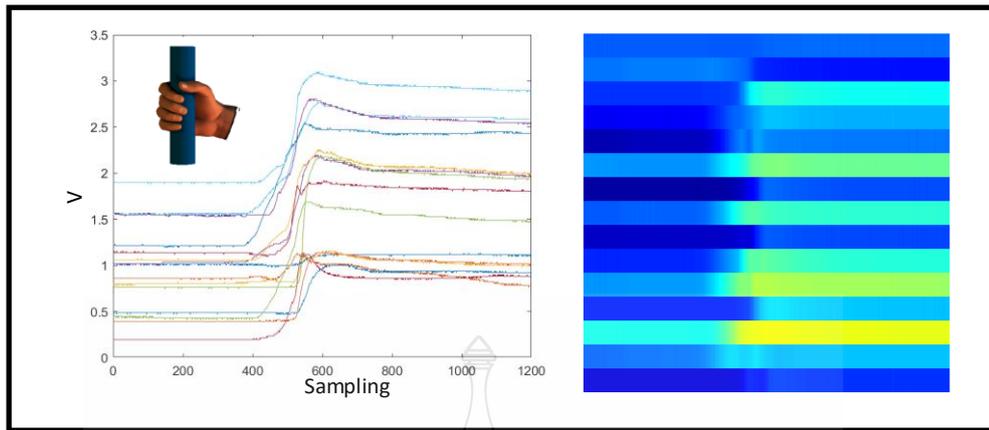
โดย $I^{(x,y)}$ คือ ภาพที่ได้หลังจากแปลงขนาด

I คือ ภาพที่แปลงขนาด

x' คือ ค่าในตำแหน่งพิกัดแกน X ของภาพต้นฉบับ

y' คือ ค่าในตำแหน่งพิกัดแกน Y ของภาพต้นฉบับ

ผลของการเปลี่ยนแปลงความต้านทานถูกบันทึกในรูปแบบของแรงดันเอาต์พุต ซึ่งเปลี่ยนแปลงตามการงอนิ้วมือ ซึ่งในรูปที่ 3.10 เป็นภาพตัวอย่างหลังแปลงสัญญาณที่ได้ในบางท่ามือจับวัตถุ พบว่าเมื่อมีการงอนิ้วมือที่แตกต่างกันทั้งลักษณะและความสูงของเส้นกราฟย่อมแตกต่างกัน ดังนั้นเมื่อมีการจับวัตถุซึ่งท่ามือที่ใช้ในการจับวัตถุจะสามารถวิเคราะห์ได้จากลักษณะกราฟที่เกิดขึ้น โดยลักษณะท่ามือจับแบบ Large Diameter, Medium Wrap และ Small Diameter จะมีลักษณะในการจับที่เหมือนกัน แตกต่างกันแค่ขนาดของวัตถุ ซึ่งสังเกตได้ว่าความสูงของกราฟที่เกิดจากท่า Small Diameter จะสูงที่สุด รองลงมาคือ Medium Wrap และต่ำสุดคือ Large Diameter เนื่องจากท่าจับแบบ Small Diameter จะมีมุมงอที่เกิดขึ้นมากที่สุด เป็นผลให้ความสูงของสัญญาณมากที่สุดตามไปด้วย ในขณะที่เดียวกันสำหรับ ขั้นตอนสุดท้ายจะแปลงเป็นข้อมูลภาพสีด้วยวิธี Color mapping ในระบบสี RGB ดังรูปที่ 3.10 ซึ่งภาพที่ได้จะถูกกำหนดขนาดที่ 227×227 pixels และเป็นภาพ 3 เลเยอร์นำไปเป็นชุดข้อมูลสำหรับทดสอบด้วยโครงข่าย CNN ต่อไป



รูปที่ 3.10 สัญญาณและรูปสี่ขนาด 227x227 พิกเซล ที่ได้แปลงภาพด้วยวิธี Color mapping

3.2.2.2 การแปลงด้วยวิธี Gramian Angular Fields (GAF) เทคนิคนี้เป็นการแปลงข้อมูลอนุกรมเวลาหนึ่งมิติเป็นภาพสองมิติ GAF เป็นการเข้ารหัสอนุกรมเวลาลงในภาพโดยใช้เมทริกซ์ที่อ้างอิงตามพิกัดเชิงชั่ว ซึ่งอัลกอริทึมนี้นิยมใช้สำหรับการแปลงข้อมูลอนุกรมเวลาเป็นข้อมูลภาพ โดยยังสามารถรักษา Temporal correlation แบบสมบูรณ์ไว้ได้ จากอนุกรมเวลา $X = \{x_1, x_2, \dots, x_n\}$ ของข้อมูลค่าจริง n จำนวน อนุกรมเวลาดั้งเดิม x จะถูกทำให้เป็นมาตรฐานในช่วง $[-1,1]$ หรือ $[0,1]$ ซึ่งสามารถแสดงได้ ดังต่อไปนี้

$$\tilde{x}_{-1}^i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (3.8)$$

$$\tilde{x}_0^i = \frac{(x_i - \min(X))}{\max(X) - \min(X)} \quad (3.9)$$

จากนั้นก็จะต้องแปลงอนุกรมเวลาหนึ่งมิติจากระบบพิกัดคาร์ทีเซียนเป็นระบบพิกัดเชิงชั่ว ซึ่งสามารถแสดงเป็น

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in X' \\ r_i = \frac{i}{N} & i \in N \end{cases} \quad (3.10)$$

โดยที่ฟังก์ชัน Cosine ของข้อมูลที่ทำ Normalize แล้ว โดย \tilde{x}_i ถูกนำมาเป็นมุม ϕ_i ในระบบพิกัดเชิงชั่วและทำการ Label เวลาที่ i/N จะใช้เป็นรัศมี การประมวลผลข้อมูลโดยดำเนินการทำให้เป็น Normalize ทั้งสองรูปแบบข้อมูลที่มีช่วงมุมต่างกัน เมื่อแปลงเป็นระบบพิกัดเชิงชั่ว ช่วงมุมของฟังก์ชัน

Cosine ที่มีการสอดคล้องกับข้อมูลภายในช่วง $[0, 1]$ คือ $[0, \pi/2]$ และมุมที่สอดคล้องกับข้อมูลในช่วง $[-1, 1]$ คือ $[0, \pi]$ วิธีการนี้ขึ้นอยู่กับระบบพิกัดเชิงขั้ว ซึ่งให้มุมมองใหม่สำหรับการทำความเข้าใจอนุกรมเวลา นั่นคือ เมื่อเวลาผ่านไป ค่าของลำดับจะแปรผันจากการเปลี่ยนแปลงของแอมพลิจูดเดิมไปจนถึงการเปลี่ยนแปลงเชิงมุมในระบบพิกัดเชิงขั้ว โดยการคำนวณผลรวมต่อส่วนต่างของฟังก์ชันตรีโกณมิติระหว่างจุดสุ่มตัวอย่าง ความสัมพันธ์ของเวลาระหว่างจุดสุ่มตัวอย่างจะถูกกระบุจากมุมมองของมุม โดยจะได้ Gramian Angular Summation Field (GASF) ดังนี้

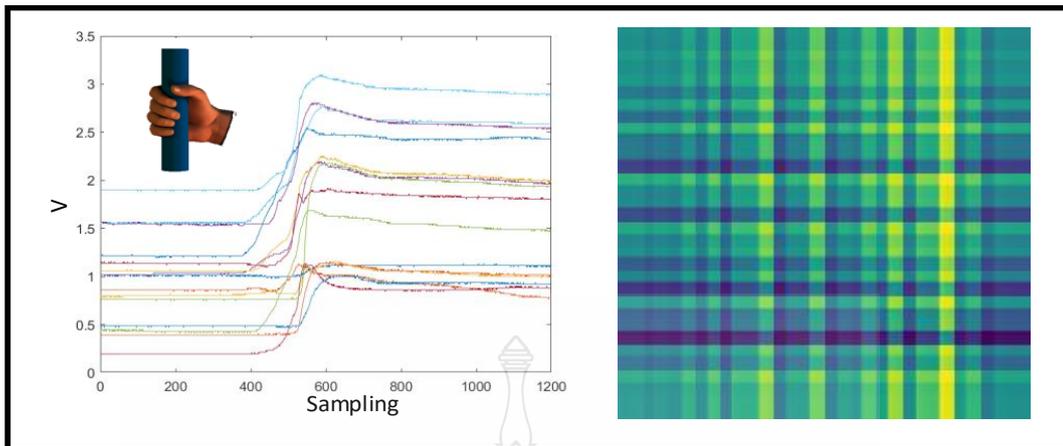
$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \dots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \dots & \cos(\phi_2 + \phi_n) \\ \dots & \dots & \dots \\ \cos(\phi_n + \phi_1) & \dots & \cos(\phi_n + \phi_n) \end{bmatrix} \quad (3.11)$$

$$GASF = \bar{X}' \cdot \bar{X} - \sqrt{I - \bar{X}^2} \cdot \sqrt{I - \bar{X}^2} \quad (3.12)$$

I คือเวกเตอร์แถวหน่วย $[1, 1, \dots, 1]$ หลังจากแปลงเป็นระบบพิกัดเชิงขั้วแล้ว เรานำอนุกรมเวลาที่แต่ละขั้นตอนเป็นพื้นที่เมตริกซ์ 1 มิติ โดยการกำหนดให้ $\langle x, y \rangle = x \cdot y - \sqrt{1 - x^2} \cdot \sqrt{1 - y^2}$ และ G คือเมตริกซ์ Gramian

$$G = \begin{bmatrix} \langle \tilde{x}_1, \tilde{x}_1 \rangle & \dots & \langle \tilde{x}_1, \tilde{x}_n \rangle \\ \langle \tilde{x}_2, \tilde{x}_1 \rangle & \dots & \langle \tilde{x}_2, \tilde{x}_n \rangle \\ \dots & \dots & \dots \\ \langle \tilde{x}_n, \tilde{x}_1 \rangle & \dots & \langle \tilde{x}_n, \tilde{x}_n \rangle \end{bmatrix} \quad (3.13)$$

เทคนิค GAF ถูกนำมาใช้เพื่อแปลงข้อมูลอนุกรมเวลาหนึ่งมิติเป็นข้อมูลสองมิติ โดยมีสามขั้นตอนดังนี้ การปรับขนาด การแปลงแกนพิกัดและฟังก์ชันตรีโกณมิติ เพื่อนำมาใช้กับงานประมวลผลภาพของคอมพิวเตอร์ ซึ่งจะได้ภาพตัวอย่างดังรูปที่ 3.11



รูปที่ 3.11 สัญญาณและรูปสี่ขนาด 227x227 พิกเซล ที่ได้จากการแปลงภาพด้วยเทคนิค GAF

3.2.3 ท่ามือจับวัตถุสำหรับถูงมือเซนเซอร์

งานวิจัยถูงมือเซนเซอร์สำหรับจดจำท่ามือขณะจับวัตถุเพื่อนำไปใช้ในการควบคุมการจับวัตถุของระบบหุ่นยนต์ได้ใช้ท่ามือในการทดสอบ 20 ท่ามือ ได้แก่ Large Diameter , Medium Wrap, Small Diameter, Fixed Hook, Index Finger Extension, Thump Index Finger, Thump 2Finger, Thump 3Finger, Thump 4Finger, Writing Tripod, Tip Pinch, Power Sphere, Power Disk, Precision Disk, Tripod, Lateral Tripod, Parallel Extension, Extension Type, Literal Pinch และ Ring รวมทั้งหมดเป็น 20 ท่ามือ ในแต่ละท่าจับมีลักษณะการงอนิ้วที่แตกต่างกัน ซึ่งจุดหมุนของข้อนิ้วในแต่ละจุดที่ติดตั้งเซนเซอร์ไว้จะทำให้ค่าความต้านทานเปลี่ยนแปลงไปตามด้วย โดยจะค่อยๆ เพิ่มขึ้นเมื่อมีการงอนิ้วเพื่อจัดการวัตถุตามลักษณะรูปทรงของวัตถุและการนำไปใช้งานในแต่ละท่ามือจับ โดยในแต่ละท่ามือจับใช้วัตถุ 5 ชิ้น แล้วจับวัตถุละ 40 ครั้ง ซึ่งจะได้จำนวน 200 ชุดข้อมูลต่อท่ามือจับ แต่บางท่ามืออาจใช้วัตถุชนิดเดียวกัน ซึ่งสัญญาณเอาต์พุตแต่ละท่ามือจับจะมีขนาดแรงดันไฟฟ้าที่แตกต่างกันอย่างเฉพาะเจาะจง โดยฐานข้อมูลสัญญาณการเปลี่ยนแปลงแรงดันไฟฟ้าขณะจับวัตถุของถูงมือเซนเซอร์ที่ใช้แปลงเป็นภาพด้วยเทคนิค Nearest interpolation และ Gramian Angular Fields เพื่อทดสอบระบบถูงมือเซนเซอร์ รวมทั้งหมดจำนวน 4,000 ชุดข้อมูล ซึ่งท่ามือจับวัตถุและชนิดวัตถุที่ใช้ในงานวิจัยนี้เป็นดังตารางที่ 3.1

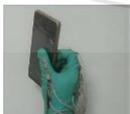
ตารางที่ 3.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยถุงมือเซนเซอร์

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Large Diameter	Can 	Cup 	Box 	Bottle 1 	Bottle 2 
Medium Wrap	Can 	Remote TV 	Knife 	Phone 	Bottle 
Small Diameter	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Fixed Hook	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Index Finger Extension	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Thumb Index Finger	Key 	Spoon 	Credit Card 	Bill 	Paper 

ตารางที่ 3.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยนี้ถุงมือเซนเซอร์(ต่อ)

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Thumb 2 Finger	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Thumb 3 Finger	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Thumb 4 Finger	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Writing Tripod	Pen 1 	Pen 2 	Wrench 	Toothbrush 	Spoon 
Tip Pinch	Key 	Spoon 	Credit Card 	Bill 	Paper 
Power Sphere	Ball 1 	Ball 2 	Ball 3 	Ball 4 	ball 5 

ตารางที่ 3.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยถุงมือเซนเซอร์ (ต่อ)

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Power Disk	Can 	Cup 	Disk 	Tap 1 	Tap 2 
Precision Disk	Can 1 	Cup 	Disk 	Tap 1 	Tap 2 
Tripod	Watch 	Lock 	Bottle Cap 1 	Bottle Cap 2 	Bottle Cap 3 
Lateral Tripod	Watch 	Lock 	Bottle Cap 1 	Bottle Cap 2 	Bottle Cap 3 
Parallel Extension	Bowl 	Smart Phone 	Book 	Wallet 	Disk 
Extension Type	Bowl 	Smart Phone 	Book 	Wallet 	Disk 

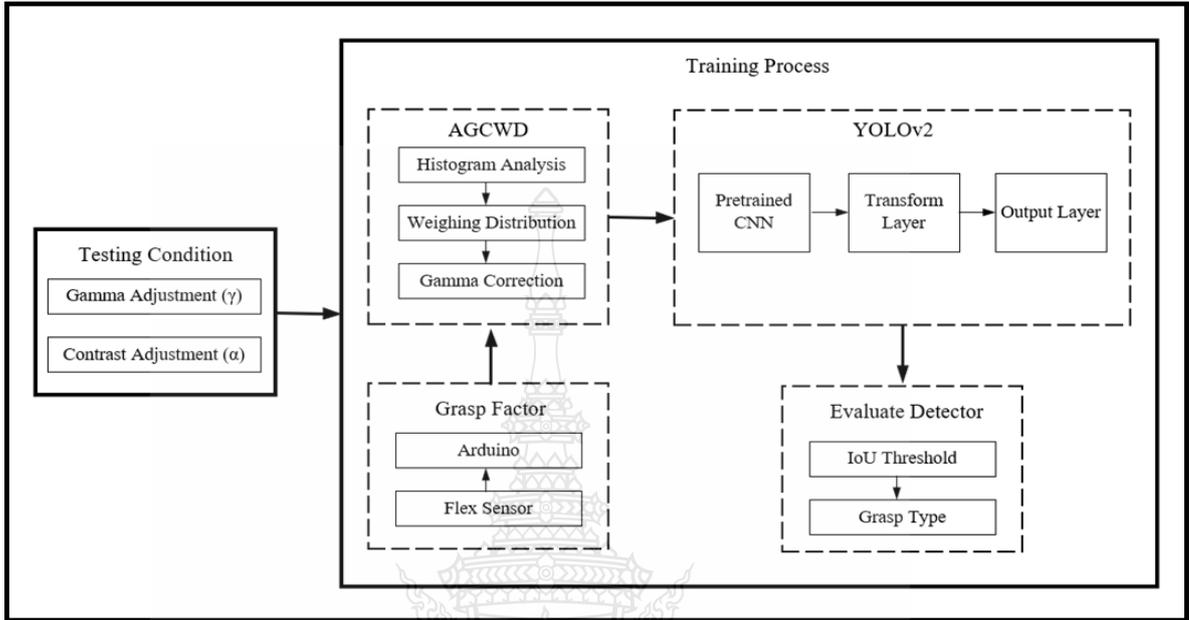
ตารางที่ 3.1 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยนี้ถึงมือเซนเซอร์(ต่อ)

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Literal	Key	Spoon	Credit Card	Bill	Paper
Pinch					
Ring	Can 1	Can 2	Can 3	Bottle 1	Bottle 2
					

3.3 ระบบเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุ

งานวิจัยนี้นำเสนอระบบการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่าทางมือจับวัตถุในระบบอีโก้เซ็นเซอร์ควิซันในสภาพแวดล้อมที่มีความเข้มแสงสว่างต่ำ โดยทำงานร่วมกันระหว่างการปรับค่าแกมมาด้วยเทคนิค AGCWD กับความต้านทานแบบโค้งงอที่ติดอยู่กับนิ้วโป้ง ซึ่งข้อมูลภาพจะบันทึกโดยใช้กล้องติดศีรษะและใช้ท่ามือจับวัตถุ 18 ท่ามือ ที่แตกต่างกันในแต่ละชนิดตามรูปร่างของวัตถุ โดยแต่ละท่ามือจับใช้วัตถุ 5 ชิ้น แต่บางท่ามืออาจใช้วัตถุชนิดเดียวกัน ซึ่งสัญญาณเอาต์พุตแต่ละท่ามือจับจะมีขนาดแรงดันไฟฟ้าที่แตกต่างกันอย่างเฉพาะเจาะจง โดยฐานข้อมูลภาพที่ใช้ในการทดสอบระบบมี 200 ภาพต่อท่ามือจับ รวมทั้งหมดจำนวน 3600 ภาพ เทคนิคที่นำเสนอประกอบด้วย 4 ขั้นตอนหลัก คือ การปรับปรุงค่าแกมมาของภาพ การคำนวณพารามิเตอร์การจับ โครงข่ายประสาทคอนโวลูชัน YOLOv2 และการประเมินและจดจำท่ามือจับ ซึ่งไดอะแกรมการทำงานเป็นดังรูปที่ 3.12 โดยระบบจะรับภาพที่มีความเข้มแสงสว่างในระดับต่าง ๆ ซึ่งถูกจำลองจากการปรับค่า Contrast (α) และค่าแกมมา (γ) หลังจากนำเข้าภาพอินพุตระบบจะปรับปรุงภาพให้มีความสว่างมากขึ้นโดยใช้เทคนิค AGCWD ซึ่งใช้น้ำหนักของพารามิเตอร์ gf ที่ได้จากค่าเฉลี่ยสัญญาณการเปลี่ยนแปลงแรงดันไฟฟ้าเอาต์พุตของวงจรความต้านทานแบบโค้งงอ แล้วส่งต่อไปให้เทคนิค YOLOv2 ซึ่งมีประสิทธิภาพสูงในการตรวจจับและจดจำวัตถุในภาพ โดยการทดลองจะใช้ Pretrained CNN ทั้งหมด 3 โมเดล คือ VGG16, ResNet 50 และ

GoogLeNet หลังจากได้ผลการตรวจจับและจดจำท่ามือ จะถูกประเมินผลการจดจำโดยใช้ค่า IoU เพื่อนำไปคำนวณค่า Accuracy, Precision, Recall และ Average Precision ของระบบ



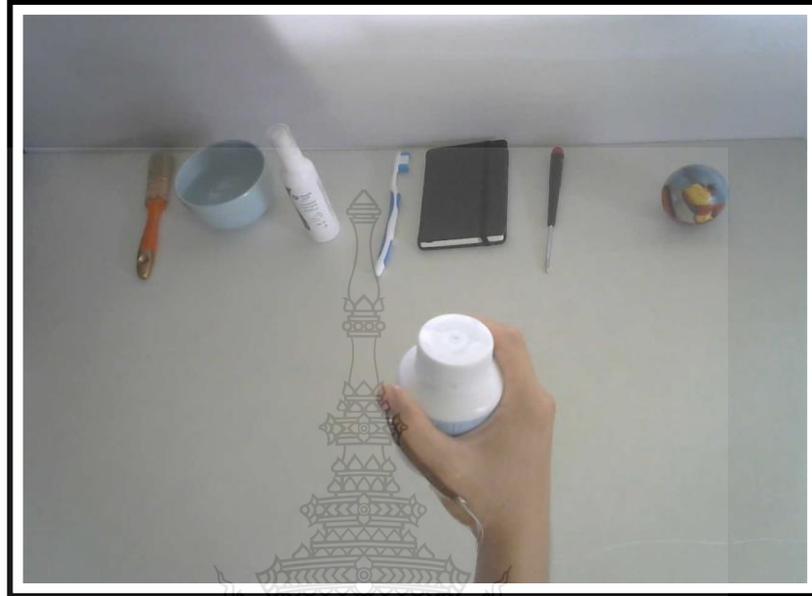
รูปที่ 3.12 ขั้นตอนการทำงานระบบตรวจจับและจดจำท่ามือจับวัตถุในสภาวะความเข้มแสงสว่างต่ำ

3.4 การเก็บรวบรวมฐานข้อมูลภาพและสัญญาณวงจรความต้านทานแบบโค้งงอ

3.3.1 การเก็บรวบรวมฐานข้อมูลภาพ

รูปแบบของท่าจับมือขึ้นอยู่กับรูปร่างของวัตถุที่ต้องการจับและการนำไปใช้งาน ดังนั้นเพื่อให้สอดคล้องกับท่ามือจับที่ใช้ในการทดสอบจึงได้จัดวัตถุที่เหมาะสมกับการใช้มือจับวัตถุในแต่ละท่ามือจับ โดยมีการติดความต้านทานแบบโค้งงอที่นิ้วโป้งเพื่อนำการเปลี่ยนแปลงของสัญญาณเอาต์พุตไปเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุ ซึ่งสัญญาณมีการเปลี่ยนแปลงขณะใช้มือจับวัตถุ แตกต่างกันจากการโค้งงอของนิ้วโป้งขณะจับวัตถุเพื่อจัดการวัตถุในท่ามือต่าง ๆ ดังตัวอย่างภาพการจับขวดด้วยท่ามือ Large Diameter ในรูปที่ 3.13 โดยทำการจับวัตถุแต่ละชนิดพร้อมกับบันทึกภาพในมุมมองต่าง ๆ ด้วยกล้อง Logitech รุ่น C270 ที่ติดตั้งบริเวณศีรษะและดำเนินการบันทึกในเวลากลางวันที่มีความละเอียดของภาพขนาด 640 × 480 พิกเซล โดยในแต่ละท่ามือจับใช้วัตถุ 5 ชิ้น จำนวน 40 ภาพต่อวัตถุ รวมหนึ่งท่ามือจับจะมีจำนวน 200 ภาพ ดังนั้นในการทดลองได้ใช้ 18 ท่ามือ รวมเป็นทั้งหมด 3600 ข้อมูล ซึ่งถูกกำหนดเป็นชุดข้อมูลที่นำไปทำ Label ภาพเพื่อเป็นชุดฐานข้อมูลสำหรับฝึกการ

เรียนรู้ของเทคนิค YOLOv2 และทดสอบในสภาวะความเข้มแสงสว่างของสภาพแวดล้อมต่าง ๆ ต่อไป
ซึ่งทำมือจับวัตถุและชนิดวัตถุที่ใช้ในงานวิจัยนี้เป็นดังตารางที่ 3.2



รูปที่ 3.13 ภาพตัวอย่างที่บันทึกขณะใช้มือจับวัตถุพร้อมความต้านทานแบบโค้งงอ

ตารางที่ 3.2 ทำมือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยเพิ่มประสิทธิภาพตรวจจับทำมือ

ทำมือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Large Diameter	Can	Cup	Box	Bottle 1	Bottle 2
Medium Wrap	Can	Remote TV	Knife	Phone	Bottle

ตารางที่ 3.2 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยเพิ่มประสิทธิภาพตรวจจับจำท่ามือ (ต่อ)

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Small Diameter	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Fixed Hook	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Index Finger Extension	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Thumb Index Finger	Key 	Spoon 	Credit Card 	Bill 	Paper 
Thumb 2 Finger	Ladle 	Pen 	Wrench 	Toothbrush 	Spoon 
Writing Tripod	Pen 1 	Pen 2 	Wrench 	Toothbrush 	Spoon 

ตารางที่ 3.2 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยเพิ่มประสิทธิภาพตรวจจับจำท่ามือ (ต่อ)

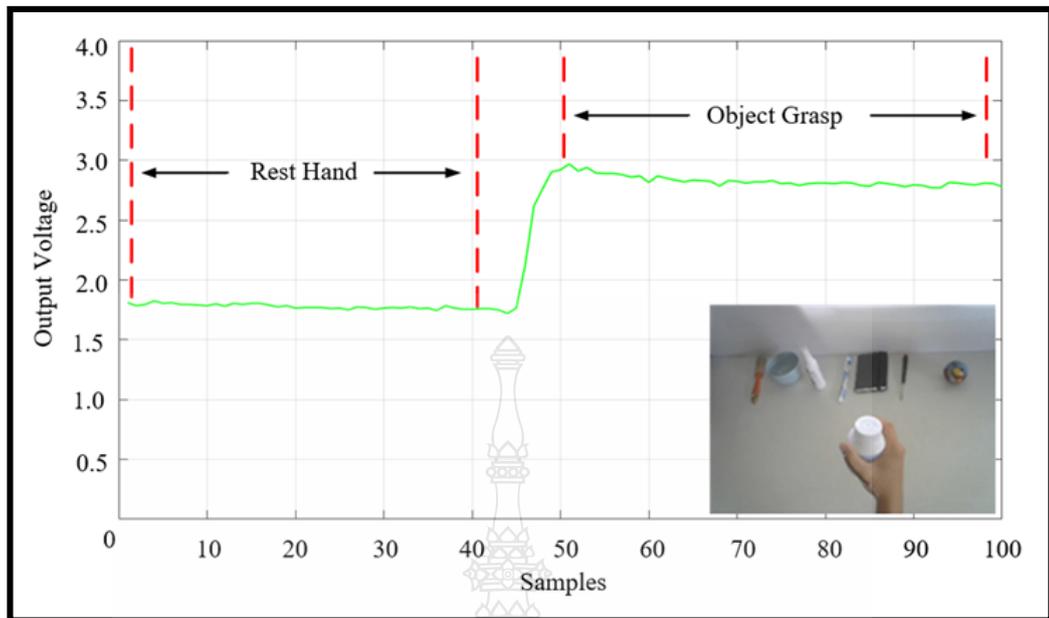
ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Tip Pinch	Key 	Spoon 	Credit Card 	Bill 	Paper 
Power Sphere	Ball 1 	Ball 2 	Ball 3 	Ball 4 	ball 5 
Power Disk	Can 	Cup 	Disk 	Tap 1 	Tap 2 
Precision Disk	Can 1 	Cup 	Disk 	Tap 1 	Tap 2 
Tripod	Watch 	Lock 	Bottle Cap 1 	Bottle Cap 2 	Bottle Cap 3 
Lateral Tripod	Watch 	Lock 	Bottle Cap 1 	Bottle Cap 2 	Bottle Cap 3 

ตารางที่ 3.2 ท่ามือจับและวัตถุแต่ละชนิดที่ใช้ในงานวิจัยเพิ่มประสิทธิภาพตรวจจับจำท่ามือ (ต่อ)

ท่ามือจับ	วัตถุที่ 1	วัตถุที่ 2	วัตถุที่ 3	วัตถุที่ 4	วัตถุที่ 5
Parallel Extension	Bowl 	Smart Phone 	Book 	Wallet 	Disk 
Extension Type	Bowl 	Smart Phone 	Book 	Wallet 	Disk 
Literal Pinch	Key 	Spoon 	Credit Card 	Bill 	Paper 
Ring	Can 1 	Can 2 	Can 3 	Bottle 1 	Bottle 2 

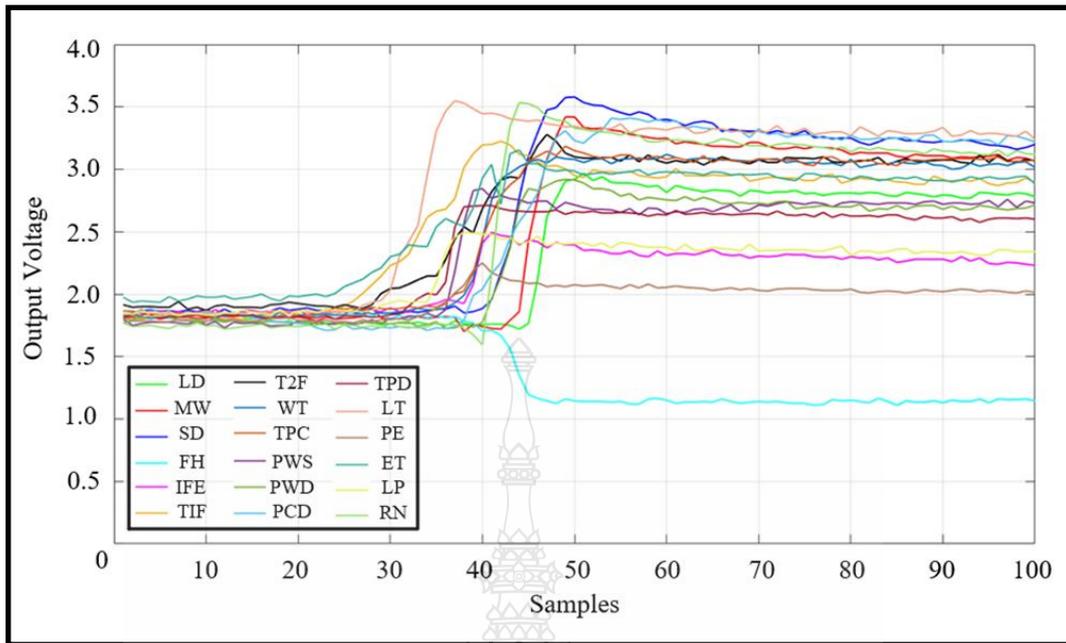
3.3.2 สัญญาณวงจรความต้านทานแบบโค้งงอ

ความต้านทานแบบโค้งงอที่ติดอยู่กับนิ้วโป้งเพื่อให้ข้อมูลเกี่ยวกับการเคลื่อนไหวและตำแหน่งของนิ้ว จากนั้นจึงทำการเก็บข้อมูลการเปลี่ยนแปลงของสัญญาณเอาต์พุต ซึ่งจะแปลงเป็นข้อมูลดิจิทัลด้วยไมโครคอนโทรลเลอร์ Arduino ข้อมูลที่ได้นี้ใช้เพื่อกำหนดค่าพารามิเตอร์การถ่วงน้ำหนักของเทคนิค AGCWD ในส่วนการปรับปรุงรูปภาพอินพุตที่มีความเข้มแสงสว่างในค่าต่าง ๆ สัญญาณความต้านทานแบบโค้งงอที่ได้เป็นการตอบสนองของเซนเซอร์ดังรูปที่ 3.14 โดยการเริ่มต้นด้วยท่ามือที่พร้อมสำหรับจับวัตถุและเมื่อจับวัตถุนิ้วมือจะมีการจัดเรียงขนวัตถุจนถึงและแรงดันเอาต์พุตของวงจรคงที่ หลังจากนั้นจะต้องคำนวณค่าเฉลี่ยของแรงดันเอาต์พุต โดยสัญญาณแรงดันเอาต์พุตในการจับอาจแตกต่างกันไปขึ้นอยู่กับท่ามือจับและลักษณะรูปร่างของวัตถุ



รูปที่ 3.14 ตัวอย่างสัญญาณเอาต์พุตวงจรความต้านทานแบบโค้งงอขณะใช้มือจับวัตถุ

การใช้ความต้านทานแบบโค้งงอที่นิ้วโป้งเพื่อนำการเปลี่ยนแปลงของสัญญาณเอาต์พุตไปเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุ ซึ่งใช้สัญญาณที่มีการเปลี่ยนแปลงขณะมือจับวัตถุ โดยการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุใช้ความแตกต่างกันจากการโค้งงอของนิ้วโป้งขณะจับวัตถุเพื่อจัดการวัตถุในท่ามือต่าง ๆ ซึ่งสัญญาณเอาต์พุตที่ได้จากเซนเซอร์ที่ติดตั้งไว้ที่นิ้วโป้งขณะจับวัตถุในท่ามือที่แตกต่างกัน งานวิจัยนี้ได้ใช้ท่ามือในการทดสอบ 18 ท่ามือ ได้แก่ Large Diameter (LD) , Medium Wrap (MW), Small Diameter (SM), Fixed Hook (FH), Index Finger Extension (IFE), Thump Index Finger, Thump 2Finger (T2F), Writing Tripod (WT), Tip Pinch (TPC), Power Sphere (PWS), Power Disk (PWD), Precision Disk (PCD), Tripod (TPD), Lateral Tripod (LT), Parallel Extension (PE), Extension Type (ET), Literal Pinch (LP) และ Ring (RN) ในแต่ละท่าจับมีลักษณะการจัดวางและงอนิ้วโป้งที่ต่างกัน ซึ่งจุดหมุนของข้อนิ้วในแต่ละจุดที่ติดตั้งเซนเซอร์ไว้จะทำให้ค่าความต้านทานเปลี่ยนแปลงไปตามด้วย โดยจะเพิ่มขึ้นเมื่อมีการงอนิ้วมากขึ้น ลักษณะความเปลี่ยนแปลงนี้ทำให้เราสามารถนำไปใช้จดจำรูปแบบของท่ามือจับวัตถุในแต่ละท่าได้ ซึ่งลักษณะสัญญาณจะมีขนาดแรงดันเอาต์พุตที่ต่างกันดังตัวอย่างสัญญาณในรูปที่ 3.15 โดยสัญญาณการเปลี่ยนแปลงที่ได้จากเซนเซอร์จะถูกทำการเปลี่ยนจากสัญญาณ Analog เป็น Digital ด้วยไมโครคอนโทรลเลอร์ Arduino แล้วนำเข้าระบบคอมพิวเตอร์และบันทึกด้วยโปรแกรม MATLAB 2021a



รูปที่ 3.15 สัญญาณเอาต์พุตที่ได้จากเซนเซอร์ที่ติดที่นิ้วโป่งขณะจับวัตถุในท่ามือที่แตกต่างกัน

3.5 การออกแบบอัลกอริทึม

การทำงานของระบบเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุในระบบอ็อกซ์เซ็นทริกซ์ในสภาพแวดล้อมที่มีความเข้มแสงสว่างต่ำจะนำค่าเฉลี่ยของแรงดันเอาต์พุตของวงจรเซนเซอร์ที่ติดบริเวณนิ้วโป่งเพื่อใช้กำหนดค่าพารามิเตอร์การถ่วงน้ำหนักของเทคนิค AGCWD [13] ในส่วนการปรับปรุงรูปภาพด้วยการคำนวณค่า Grasp Factor (gf) ดังสมการที่

$$gf = \frac{V_{avg}}{V_{fist}} \quad (3.14)$$

โดยที่ V_{avg} เป็นค่าเฉลี่ยแรงดันในขณะที่ใช้มือการจับวัตถุและ V_{fist} เป็นค่าเฉลี่ยแรงดันไฟฟ้าที่ขณะกำมือเป็นท่ากำปั้นแน่น ซึ่งทำให้งอนิ้วได้มากที่สุดเนื่องจากไม่มีวัตถุใดรองรับนิ้ว โดยค่าเกมมาจะค่อยๆ ปรับเพิ่มช่วงความเข้มแสงสว่างต่ำและหลีกเลี่ยงการลดลงอย่างมีนัยสำคัญในช่วงความเข้มแสงสว่างสูง ส่วนค่า Weight Distribution (WD) ยังใช้เพื่อปรับเปลี่ยนฮิสโตแกรมทางสถิติและลดผลกระทบให้เหลือน้อยที่สุด ฟังก์ชัน WD สามารถคำนวณได้ตามสมการที่ 3.15

$$pdf_w(l) = pdf_{max} \left(\frac{pdf(l) - pdf_{min}}{pdf_{max} - pdf_{min}} \right)^{gf} \quad (3.15)$$

โดยที่ gf คือพารามิเตอร์ที่ปรับแกมมา, pdf_{max} คือฟังก์ชันการแจกแจงความน่าจะเป็นสูงสุดของฮิสโตแกรมทางสถิติ และ pdf_{min} คือฟังก์ชันการแจกแจงความน่าจะเป็นต่ำสุด ดังสมการที่ 3.15 จะได้ค่า cdf ดังสมการที่ 3.16

$$cdf_w(l) = \sum_{l=0}^{l_{max}} pdf_w(l) / \Sigma pdf_w \quad (3.16)$$

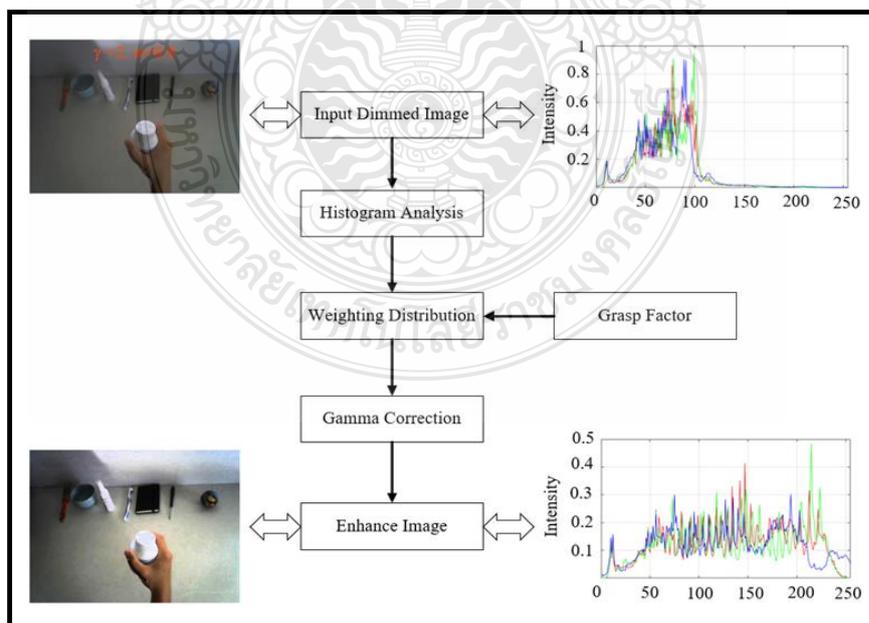
โดยผลรวมของ pdf_w จะสามารถคำนวณได้จาก

$$\Sigma pdf_w = \sum_{l=0}^{l_{max}} \Sigma pdf_w(l) \quad (3.17)$$

สุดท้ายจะได้พารามิเตอร์สำหรับปรับค่าแกมมาของภาพดังสมการที่ 3.18

$$\gamma = 1 - cdf_w(l) \quad (3.18)$$

จากนั้นภาพจะถูกดำเนินการปรับแกมมาด้วยเทคนิค AGCWD โดยน้ำหนักพารามิเตอร์ WD ที่ใช้ปรับแกมมาจะขึ้นอยู่กับค่าเฉลี่ยของแรงดันเอาต์พุตของวงจรเซนเซอร์สำหรับแต่ละท่ามือจับและทำ Normalization ด้วยแรงดันเอาต์พุตของท่าก้ามือจะได้ค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งเทคนิค AGCWD เป็นการแปลงภาพอัตโนมัติที่มีประสิทธิภาพเพิ่มความสว่างของภาพที่มีดด้วยการแก้ไขค่าแกมมาและการกระจายความน่าจะเป็นของพิกเซลความสว่าง ซึ่งเทคนิคนี้ใช้ข้อมูลจากความแตกต่างระหว่างข้อมูลภาพเพื่อทำให้การคำนวณง่ายขึ้นในการปรับปรุงภาพอินพุต โดยอะแกรมของวิธีการเป็นดังในรูปที่ 3.16



รูปที่ 3.16 ไดอะแกรมการทำงานของอัลกอริทึมที่นำเสนอสำหรับใช้ปรับปรุงภาพ

3.6 วิธีการทดสอบงานวิจัย

3.5.1 การเตรียมภาพสำหรับทดสอบ

การทดสอบประสิทธิภาพของเทคนิคที่นำเสนอในงานวิจัยนี้จะใช้ภาพที่มีความเข้มแสงสว่างในสภาพแวดล้อมโดยการจำลองค่าพารามิเตอร์ของภาพด้วยค่า Gamma (γ) และ Contrast (α) [46] โดยภาพชุดข้อมูลถูกรวบรวมภาพด้วยกล้องติดที่ศีรษะและบันทึกในเวลากลางวันที่มีความละเอียดของภาพขนาด 640 × 480 พิกเซล รายละเอียดของทั้งสองวิธีมีดังนี้

3.5.1.1. การปรับแกมมา (Gamma Transform) ค่าการปรับแกมมาเป็นความสัมพันธ์ระหว่างความไวของเซนเซอร์รับแสงของกล้องแล้วแปลงเป็นภาพดิจิทัลและความไวต่อดวงตาของมนุษย์ ซึ่งการปรับภาพเอาต์พุต I_{out} กำหนดด้วยสมการที่ 3.19

$$I_{out} = A \cdot I_{in}^{\gamma} \quad (3.19)$$

โดยที่ A คือ ค่าคงที่ที่กำหนดโดยความเข้มแสงสว่างของพิกเซลสูงสุดในภาพอินพุต โดยเมื่อค่า $\gamma > 1$ จะถ่วงน้ำหนักกับค่าความเข้มแสงสว่างในพิกเซลระดับสีเทาที่ต่ำกว่า (สีเข้มกว่า) วิธีการนี้เป็นการเพิ่มค่า γ เพื่อให้ค่าของภาพในเวลากลางวันมีลดลงจากระดับ 1 ถึง 6 โดยภาพที่มีความมืดนั้นเกิดจากค่า γ ที่แตกต่างกัน ยังมีค่าสูงก็จะทำให้ภาพมีดีขึ้น

3.5.1.2. การปรับความคมชัด (Contrast Adjustment) เป็นการปรับความสัมพันธ์ของพิกเซลระหว่างส่วนที่สว่างและส่วนที่มืดของรูปภาพ ทำให้วัตถุหรือรายละเอียดภายในรูปภาพชัดเจนยิ่งขึ้น การลด Contrast จะทำให้ส่วนที่สว่างและมีค่าพิกเซลระดับสีเทาใกล้เคียงกัน โดยเอาต์พุต I_{out} กำหนดโดยสมการที่ 3.20 โดยที่ α ในงานวิจัยนี้จะปรับในช่วงระหว่าง 0.5 ถึง 1

$$I_{out} = \alpha \cdot I_{in} \quad (3.20)$$

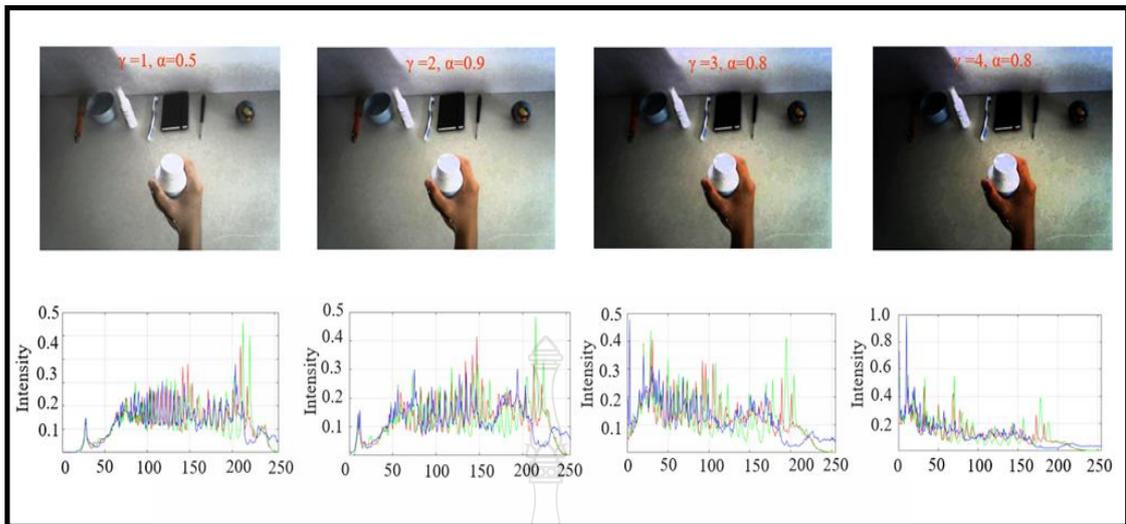
การทดลองจะกำหนดสภาพแวดล้อมให้มีความเข้มแสงสว่างออกเป็น 3 รูปแบบ คือ 1. การปรับค่าแกมมา (γ) จากระดับ 1 ถึง 6, 2. การปรับความคมชัด (α) จากระดับ 0.5 ถึง 1, และ 3. การปรับผสมทั้งแกมมาและความคมชัดพร้อมกัน ตัวอย่างภาพที่ได้ปรับความสว่างของภาพในแต่ละรูปแบบและระดับต่าง ๆ เป็นดังรูปที่ 3.17 จากนั้นภาพจะถูกปรับปรุงด้วยอัลกอริทึม AGCWD เพื่อใช้ทดสอบประสิทธิภาพในการตรวจจับและจดจำการจับวัตถุของมือในระบบอ็γκไ้เซ็นทริกวิชั่น



รูปที่ 3.17 ตัวอย่างภาพที่ได้ปรับความเข้มแสงสว่างให้ลดต่ำลงในแต่ละรูปแบบ

3.5.2 การปรับปรุงภาพด้วยอัลกอริทึม AGCWD

เทคนิค AGCWD ใช้สำหรับปรับภาพที่ระดับความเข้มแสงสว่างต่าง ๆ ก่อนเข้าเป็นอินพุตของเทคนิค YOLOv2 เพื่อตรวจจับและจดจำท่ามือขณะจัดการวัตถุในระบบอีโก้เซ็นเซอร์วิชั่น ซึ่งจะได้รับการปรับปรุงภาพโดยใช้ค่า Grasp Factor ที่ได้จากค่าเฉลี่ยของแรงดันเอาต์พุตของวงจรเซนเซอร์ที่ติดตั้งไว้เพื่อกำหนดค่าพารามิเตอร์การถ่วงน้ำหนัก โดยขึ้นอยู่กับแรงดันเอาต์พุตของวงจรความต้านทานแบบโค้งงอที่แตกต่างกันสำหรับแต่ละท่าทางมือ ซึ่งจะสร้างรูปแบบคุณลักษณะเฉพาะตัวที่ไม่เหมือนกันในแต่ละท่ามือ เป็นผลให้การฝึกการเรียนรู้ของโครงข่ายคอนโวลูชันแบบ YOLOv2 ที่มีประสิทธิภาพสูงในการตรวจจับและจดจำวัตถุในภาพสามารถเรียนรู้ความแตกต่างของคุณลักษณะดังกล่าวได้ หลักการทำงานสำหรับเทคนิค AGCWD ที่เสนอเป็นการคำนวณจากพารามิเตอร์แกมมาผ่านความหนาแน่นของความน่าจะเป็นเพื่อรวมรูปแบบง่าย ๆ ของการปรับค่าแกมมา ร่วมกับการทำ Histogram Equalization แบบดั้งเดิม แม้ว่า Histogram ดั้งเดิมจะไม่ถูกใช้โดยตรงเพื่อสร้างฟังก์ชันการแปลงภาพและปรับปรุงความคมชัดของภาพ แต่ข้อดีของเทคนิค AGCWD ที่นำมาใช้ในงานวิจัยนี้สามารถปรับปรุงความสว่างและให้ภาพชัดเจนขึ้นที่สามารถยอมรับได้ ดังในรูปที่ 3.18



รูปที่ 3.18 ตัวอย่างภาพและ Histogram ที่ผ่านการปรับความสว่างด้วยเทคนิค AGCWD

3.5.3 การตรวจจับและจดจำด้วย YOLOv2

การทดลองนี้ใช้โครงข่ายประสาทคอนโวลูชันเชิงลึกแบบ YOLOv2 ร่วมกับ Pertained CNN จำนวน 3 โมเดล ในการฝึกการเรียนรู้ตรวจจับและจดจำ โดย Pertained CNN ใช้สำหรับการสกัดคุณลักษณะของภาพ (Feature Extraction) ได้แก่ VGG16 [49], ResNet 50 [50] และ GoogLeNet [51] โดยมีพารามิเตอร์ของโครงข่ายประสาท CNN ที่ใช้ในการทำงานเป็นดังในตารางที่ 3.3 โดยกระบวนการทำงานเริ่มด้วยนำชุดภาพฐานข้อมูลอินพุตป้อนเข้าสู่ระบบ ซึ่งจะมีการแบ่งข้อมูลภาพออกเป็น 3 ส่วน คือ ส่วนแรกใช้สำหรับฝึกการเรียนรู้ของระบบร้อยละ 60 ซึ่งเป็นภาพปกติที่ยังไม่ได้มีการปรับระดับความเข้มแสงสว่างให้มีติดลง ส่วนที่สองสำหรับใช้ทดสอบผลการฝึกอบรมร้อยละ 10 เป็นภาพที่มีคุณสมบัติเดียวกับภาพที่ใช้ในชุดฝึกการเรียนรู้และส่วนสุดท้ายที่เหลือร้อยละ 30 ใช้สำหรับการทดสอบประสิทธิภาพการตรวจจับและจดจำภาพท่ามือจับในระบบอีโก้เซ็นทริกวิชั่น โดยภาพที่ใช้ในการทดสอบจะถูกทำให้ระดับความเข้มของแสงสว่างมีติดลงในระดับการปรับค่าแกมมา (γ) และค่าความคมชัด (α) ตามวิธีที่ได้กล่าวมาแล้วในหัวข้อที่ 3.5.1 การทำงานระบบจะสกัดคุณลักษณะภาพอินพุตที่เป็นภาพท่ามือจับในส่วนเฉพาะที่ได้ทำ Label ตำแหน่งมือจับไว้เพื่อให้โครงข่ายประสาท CNN เรียนรู้ในส่วนที่สนใจเท่านั้น ซึ่งจะจำแนกและเรียนรู้ความแตกต่างของภาพในแต่ละท่ามือจับโดยละเอียด เช่น Color, Edges, Texture Shape เป็นต้น จากนั้นปรับค่าน้ำหนักการคำนวณในแต่ละส่วนของโครงข่ายประสาท CNN ให้มีความถูกต้องมากที่สุด โดยมีการคำนวณเปรียบเทียบกับ Training Loss ตลอดเวลาในขณะทำการเรียนรู้

ตารางที่ 3.3 พารามิเตอร์ของโครงข่ายประสาทคอนโวลูชันแบบเชิงลึกที่ใช้สำหรับฝึกการเรียนรู้

พารามิเตอร์	ค่า
Class	18
Image	3600
Initial Learning Rate	0.001
Epochs	20
Stochastic Gradient Descent with Momentum	0.9
Mini Batch Size	16



บทที่ 4

ผลการทดสอบและการวิเคราะห์

ในบทนี้จะกล่าวถึงผลการทดสอบงานวิจัยในบทที่ 3 และการวิเคราะห์ผล ซึ่งประกอบด้วย ผลการทดสอบการประมวลผลภาพเบื้องต้นในหัวข้อย่อยต่าง ๆ ผลการทดสอบจับวัตถุของถุงมือ เซนเซอร์ ผลการทดสอบตรวจจับและจดจำท่ามือจับวัตถุในสภาวะความเข้มแสงสว่างต่ำและการวิเคราะห์ผลการทดลองของอัลกอริทึมที่นำเสนอในงานวิจัยนี้

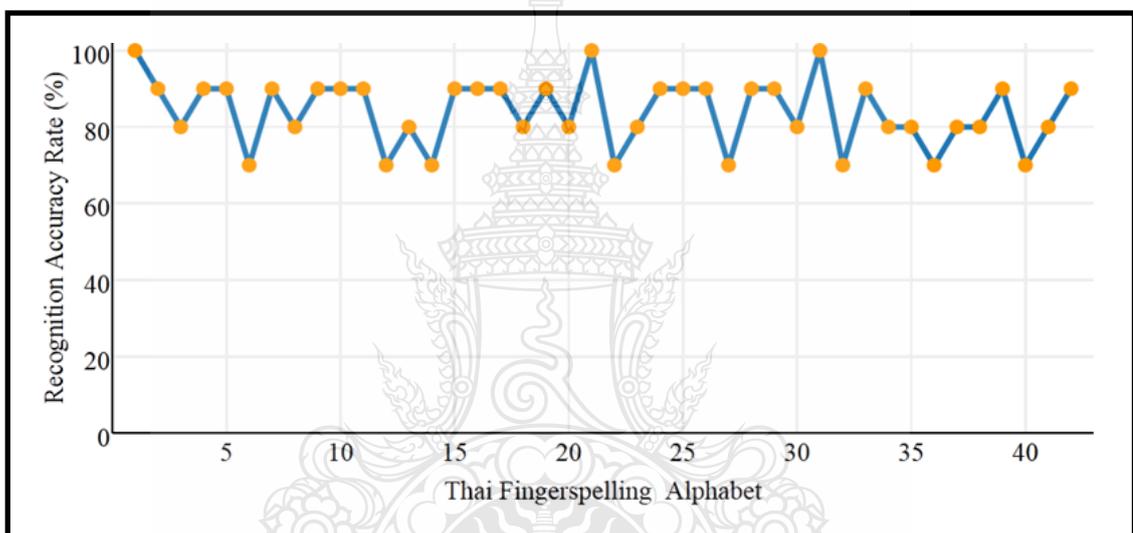
4.1 ผลการทดสอบการประมวลผลภาพเบื้องต้น

ผลการทดสอบของการประมวลผลภาพเบื้องต้น ซึ่งได้ทดลองตรวจจับและจดจำท่าทางภาษา มือด้วยกล้องดิจิทัล โดยได้มีหัวข้อที่วิจัยจำนวน 3 เรื่อง คือ การตรวจจับและจดจำท่าภาษามือ การทดสอบความถูกต้องและนำเชื่อถือของกล้อง Kinect และการควบคุมอุปกรณ์ด้วยท่าทางมือ โดยได้ผลการทดลองเป็นดังต่อไปนี้

4.1.1 การตรวจจับและจดจำท่าภาษามือ

การทดลองนี้ใช้ท่ามือสะกดนิ้วของตัวอักษรไทยประกอบด้วยตัวอักษร 42 ตัว โดยมีท่าทางมือ 24 แบบ ชุดข้อมูลภาพที่ใช้ในการทดสอบถูกบันทึกในสภาพแวดล้อมที่แตกต่างกัน เช่น ห้องทำงาน ห้องนั่งเล่น เป็นต้น อาสาสมัครจะสวมเสื้อแขนสั้นและยืนอยู่หน้ากล้อง Kinect ระยะห่าง 1 เมตร และจะแสดงท่าภาษามือการสะกดคำไทยโดยแต่ละท่ามือมีจำนวน 10 ภาพ รวมทั้งหมด 420 ภาพ สำหรับใช้ประเมินประสิทธิภาพของระบบ โดยการทำงานมี 2 ขั้นตอนหลัก คือ ขั้นตอนแรกเป็นการแบ่งส่วนมือที่ปรากฏในภาพเพื่อแยกภาพมือออกจากพื้นหลังของฉาก โดยการประมวลผลระหว่างภาพสี RGB และภาพความลึก จากนั้นจึงแยกคุณลักษณะของท่ามือที่ตรวจจับได้ด้วยเทคนิค HOG แล้วส่งต่อไปยังขั้นตอนที่สองซึ่งใช้เทคนิคโครงข่ายประสาทเทียมสำหรับฝึกการเรียนรู้จดจำท่าทางมือ ในงานนี้ใช้ฝึกการเรียนรู้แบบ Back propagation โดยมีเลเยอร์อินพุตจำนวน 81 neurons ซึ่งเป็นข้อมูลอินพุตจากการสกัดคุณลักษณะของภาพท่ามือด้วยเทคนิค HOG สำหรับ Hidden Layer จะมีจำนวน 20 neurons ใช้ในการควบคุมผลการคำนวณความถูกต้องของการเรียนรู้ของโครงข่ายและเลเยอร์เอาต์พุตมีจำนวน 24 คลาส หลังจากได้ฝึกการเรียนรู้จากฐานข้อมูลภาพทั้งหมด พบว่า มีประสิทธิภาพการตรวจสอบความถูกต้องระหว่างทำการวนซ้ำและ Mean Squared Error (MSE) โดยมีค่าประสิทธิภาพความถูกต้องที่ดีที่สุดคือ $1.6217e-5$ ที่ 38 Epoch ซึ่งเป็นผลลัพธ์ที่สามารถยอมรับได้เนื่องจากค่า MSE น้อยมากสามารถนำไปใช้ทดสอบภาพตัวอย่างได้ ในส่วนการทดสอบประสิทธิภาพของระบบ ขั้นตอนการสกัด

คุณลักษณะของภาพท่ามือจะถูกดึงออกมาในลักษณะเดียวกับส่วนการฝึกการเรียนรู้ โดยระบบจะตรวจจับท่าทางมือในแต่ละเฟรมเพื่อคำนวณหาสัญญาณมือที่ปรากฏในฉาก ซึ่งแบบสะกดนิ้วมือไทยประกอบด้วยท่าทางมือตั้งแต่ 1 ถึง 3 ท่ามือต่อหนึ่งอักษร ระบบสามารถตรวจจับสัญญาณมือที่ปรากฏในฉากแต่ละเฟรมเพื่อใช้ประมวลผลการจดจำท่าทางมือ ผลการจดจำของตัวอักษรในแบบสะกดนิ้วมือไทยแต่ละตัวเป็นดังรูปที่ 4.1 แสดงให้เห็นว่าระบบสามารถจดจำท่าทางมือได้ดีในการใช้สัญญาณมือไม่เกิน 2 ท่ามือต่อตัวอักษรเพราะข้อผิดพลาดจะเกิดขึ้น เมื่อมีการเปลี่ยนท่าทางของมือหลายครั้งเนื่องจากภาพมือจะเบลอจากการเคลื่อนที่ของมือระหว่างการเปลี่ยนท่า ซึ่งการประมวลผลจะมีความผิดพลาดเกิดขึ้นในเวลานี้และจากการประเมินประสิทธิภาพการจดจำของระบบได้ผลความถูกต้องร้อยละ 84.05

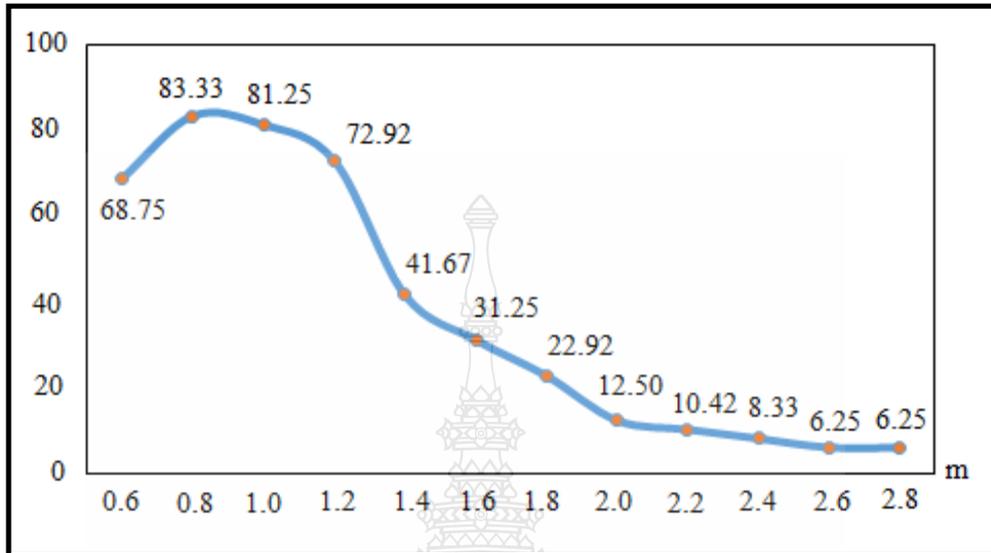


รูปที่ 4.1 ผลการจดจำท่าทางมือของแต่ละตัวอักษรในแบบสะกดนิ้วมือไทยจำนวน 42 ตัว

4.1.2 ความถูกต้องและน่าเชื่อถือของกล้อง Kinect

การทดลองจะดำเนินการโดยใช้กล้อง Kinect ในโหมดทำงานระยะใกล้ด้วยระบบปฏิบัติการ Windows เพื่อประเมินความถูกต้องของการรู้จำภาษามือไทย โดยอาสาสมัครที่ทดลองจะสวมเสื้อยืดแขนสั้นและยืนอยู่หน้ากล้อง ทำท่าแบบสะกดนิ้วมือไทยที่ใช้ท่ามือเดียวในพื้นที่ฉากที่ซับซ้อน โดยใช้ท่าภาษามือ 16 ท่ามือ ระหว่างการทดลองกล้อง Kinect และผู้ทำภาษามือ มีระยะห่างเริ่มตั้งแต่ 0.6 ถึง 2.8 เมตร โดยมีขั้น Step ละ 0.2 เมตร รวมระยะทางทั้งหมด 12 ช่วง เทคนิคที่ใช้ประกอบด้วยสองขั้นตอนหลัก คือ ขั้นตอนแรกเป็นการแบ่งส่วนมือที่ปรากฏในภาพเพื่อแยกมือออกจากพื้นหลังฉาก โดยการประมวลผลระหว่างภาพสี RGB และภาพความลึก จากนั้นจึงสกัดคุณลักษณะของท่ามือด้วยเทคนิค HOG เพื่อส่งต่อให้กับขั้นตอนที่สอง ซึ่งใช้เทคนิคโครงข่ายประสาทเทียมสำหรับฝึกการเรียนรู้

จดจำท่าทางมือและหลังจากทดสอบระบบตรวจจับท่ามือของแบบสะกดนิ้วมือไทยในแต่ละระยะทางแล้ว ได้ผลการทดลองดังรูปที่ 4.2



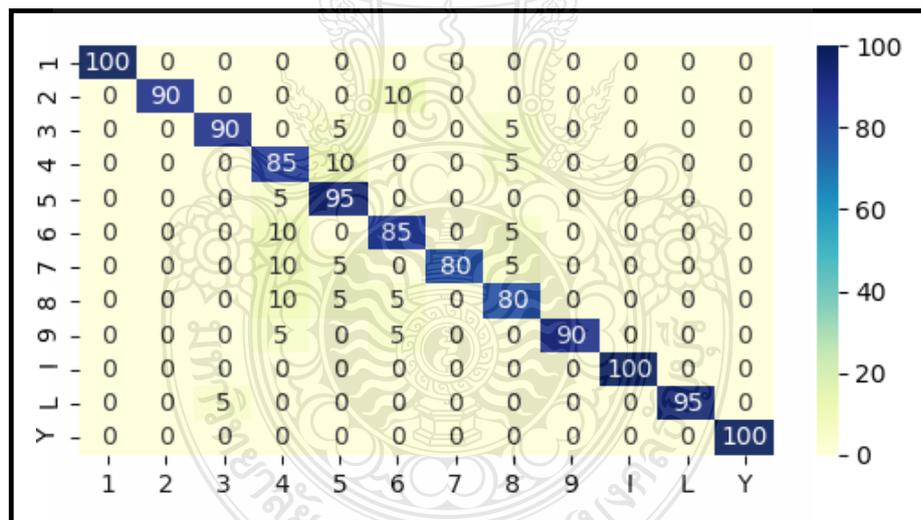
รูปที่ 4.2 ผลการทดลองตรวจจับท่ามือแบบสะกดนิ้วมือไทยในแต่ละระยะทางตั้งแต่ 0.6 ถึง 2.8 เมตร

ผลการทดลองพบว่าการตรวจจับและรู้จำภาษามือไทยด้วยกล้อง Kinect สามารถทำงานได้ดีและมีประสิทธิภาพในระยะ 0.8 - 1.2 เมตร มีความถูกต้องเฉลี่ยมากที่สุดในระยะทาง 0.8 เมตร ที่ร้อยละ 83.33 รองลงมา ร้อยละ 81.25 ที่ 1.0 เมตร และร้อยละ 72.92% ที่ 1.2 เมตร ตามลำดับ ผลการทดสอบแสดงให้เห็นว่า ระยะห่างระหว่างผู้แสดงท่ามือและกล้อง Kinect มากขึ้นจะทำให้ลดประสิทธิภาพของการตรวจจับท่ามือออกจากพื้นหลังของภาพ เนื่องจากความสว่างของภาพความลึกจะลดลงทำให้ยากต่อการตรวจจับมือในภาพ ยิ่งไปกว่านั้น ระยะห่างมากขึ้นจะทำให้ขนาดมือเล็กลง ซึ่งจะทำให้คุณสมบัติต่าง ๆ ของภาพด้อยลง เช่น Gradients, Texture, Edge, Contrast และ Contour เป็นต้น สิ่งเหล่านี้จะส่งผลกระทบต่อการสกัดคุณลักษณะของภาพในระยะทางที่ไกลขึ้น โดยเฉพาะอย่างยิ่งในระยะ 1.4 เมตร ขึ้นไป แต่อย่างไรก็ตาม ที่ระยะ 0.6 เมตร เป็นจุดที่มีความสว่างสูงสุดแต่ก็ให้ผลลัพธ์ที่ไม่น่าพอใจเช่นกัน เนื่องจากอยู่ใกล้ Blind spot ของกล้องในโหมดการทำงานระยะใกล้เกินไป ซึ่งทำให้ระบบทำงานผิดพลาดในบางครั้ง

4.1.3 การควบคุมอุปกรณ์ด้วยท่าทางมือ

ในการทดลองนี้นำเสนอเทคนิคการใช้ท่ามือสำหรับระบบควบคุมสั่งงานอุปกรณ์โดยใช้ในท่าภาษามือแบบอเมริกัน (American Sign Language) จำนวน 12 ท่ามือ ซึ่งประกอบด้วยตัวเลข 1-9 และตัวอักษรคือ I, L และ Y โดยแต่ละท่ามือจะมีการใช้ปลายนิ้วแสดงสัญลักษณ์ ซึ่งการทดลองใช้สัญญาณมือจำนวน 12 ท่ามือ โดยยื่นห่างจากกล้อง RGB ในระยะ 0.5-1 เมตร อยู่ในสภาพแวดล้อมการใช้งาน

จริง โดยทดลอง 20 ครั้งต่อท่ามือ ซึ่งจะแปลงภาพจากระบบสี RGB ของกล้อง USB Webcam ให้เป็นภาพระบบสี YCbCr และ ภาพ Gray Scale โดยภาพสี YCbCr ใช้เพื่อหาสีผิวมนุษย์ภายในเฟรม ส่วนภาพ Gray Scale ใช้สำหรับคำนวณหาความเข้มแสงที่มีการเคลื่อนไหวภายในเฟรมด้วยเทคนิค Optical flow เมื่อนำทั้งสองส่วนมาประมวลผลรวมกันก็จะได้ภาพไบนารีของภาพมือที่อยู่ในเฟรม จากนั้นจะคำนวณหา Contours และ Convexity Defects เพื่อใช้คำนวณการรู้จำท่ามือที่ปรากฏในภาพ โดยมีพารามิเตอร์จำนวน 3 ตัว ได้แก่ จำนวนปลายนิ้วมือ (n), ระยะทาง Normalize (D_{nor}) และมุมของปลายนิ้ว (θ) โดยระบบเริ่มจากการคำนวณหาจำนวนและตำแหน่งปลายนิ้วมือ จากนั้นคำนวณหาจุดศูนย์กลางและรัศมีของพื้นที่ฝ่ามือเพื่อหาระยะทาง Normalize โดยแต่ละนิ้วมือที่ตรวจจับได้ต้องมีค่า D_{nor} มากกว่า 1.9 จากนั้นจะคำนวณหาตำแหน่งข้อมือและจุดกึ่งกลางเพื่อใช้หามุมความเอียงของมือ α_{rad} โดยลากเส้นตรงระหว่างจุดศูนย์กลางของฝ่ามือกับจุดกึ่งกลางข้อมือ แล้วนำค่า α_{rad} ที่ได้บวกด้วยค่ามุมของแต่ละปลายนิ้วมือเทียบกับจุดศูนย์กลางของฝ่ามือจะได้ค่ามุม θ_{rad} ซึ่งผลการทดลองตรวจจับและจดจำท่าสะกดนิ้วมืออเมริกาจำนวน 12 ท่ามือ ได้ผลการทดลองเขียนเป็น Confuse Matrix ได้ดังรูปที่ 4.3



รูปที่ 4.3 ผลการทดลอง Confuse Matrix ของการจดจำท่าสะกดนิ้วมืออเมริกาจำนวน 12 ท่ามือ

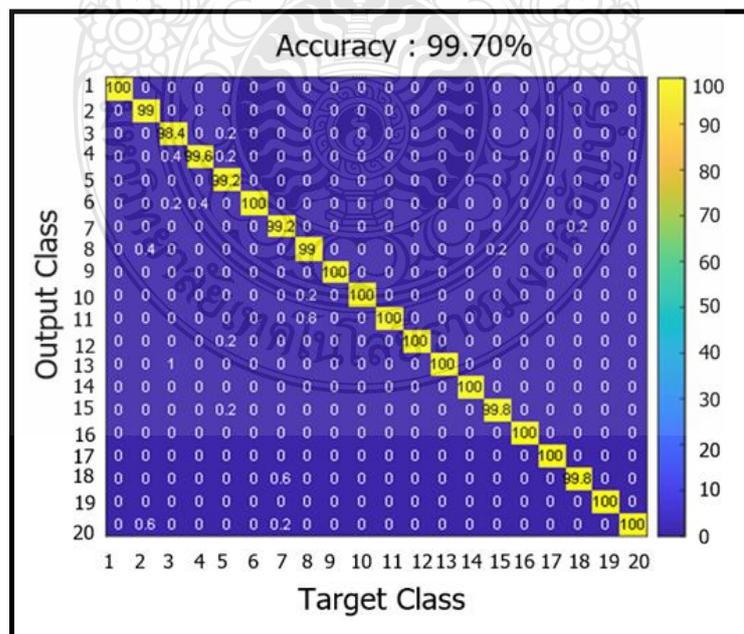
จะเห็นได้ว่าการตรวจจับและจดจำท่ามือด้วยเทคนิคการวิเคราะห์มุมเรเดียลของปลายนิ้วมือสามารถทำงานได้เป็นอย่างดีกับท่ามือสะกดนิ้วของอเมริกาทั้ง 12 ท่า โดยเฉพาะท่ามือเลข 1, ตัวอักษร I และ Y เนื่องจากมีจำนวนนิ้วที่ใช้้น้อยและมีค่ามุมเรเดียลอยู่ในช่วงที่แตกต่างกันมาก จึงทำให้ง่ายต่อการตรวจจับและจดจำ แต่ในท่ามือที่มีการใช้นิ้วจำนวน 3 นิ้ว ดังเช่น ท่ามือเลข 6, 7, 8 จะเริ่มมีความผิดพลาดมากขึ้น เนื่องจากนิ้วที่ใช้้อยู่ใกล้กัน ซึ่งจะมีค่ามุมเรเดียลของปลายนิ้วใกล้เคียงกันด้วย

ประกอบกับรูปร่างของนิ้วมือในภาพไบนารีที่ตรวจจับได้มีลักษณะไม่เรียบ มีส่วนโค้งขนาดเล็กเพิ่มขึ้น มาบริเวณก่อนถึงปลายนิ้วมือทำให้การคำนวณค่า Convexity Defects ผิดพลาดและได้ผลการคำนวณ จำนวนนิ้วมือมากกว่าจำนวนจริงที่ปรากฏในภาพและพารามิเตอร์สำคัญอีกตัวหนึ่งที่มีผลต่อความ ถูกต้องของการจดจำท่ามือก็คือ ค่ามุม α_{rad} เนื่องจากต้องใช้ค่ามุมเอียงของฝ่ามือรวมกับมุมของปลาย นิ้วมือได้เป็นค่ามุม θ_{rad} ใช้ในการจดจำท่ามือ จากการทดลองพบว่า ความคลาดเคลื่อนของการคำนวณ ค่ามุม α_{rad} มีสาเหตุมาจากการหาตำแหน่งตรงกึ่งกลางข้อมือไม่ถูกต้อง โดยผลการทดลองโดยรวมของ การใช้เทคนิคการวิเคราะห์มุมเรเดียนของปลายนิ้วมือจดจำท่าสะกดนิ้วมืออเมริกาจำนวน 12 ท่ามือ ได้ผลการตรวจจับและจดจำท่ามือที่ร้อยละ 90.83

4.2 ผลการทดสอบถุงมือเซนเซอร์แบบ 16 DoF

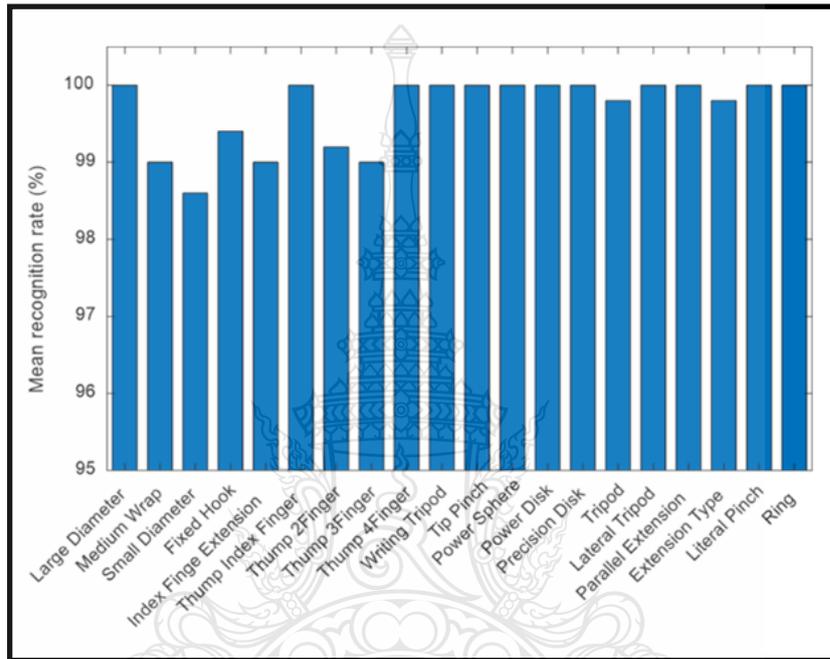
4.2.1 ผลการทดลองวิธีการ Nearest interpolation

การทดสอบประสิทธิภาพการใช้งานจริงของถุงมือเซนเซอร์แบบ 16 DoF ที่พัฒนาขึ้นจาก ความต้านทานแบบโค้งงอ ได้ทดสอบการจดจำท่ามือจากการจับวัตถุ 20 ท่ามือ โดยใช้โครงข่ายประสาท CNN แบบ Alexnet [48] สำหรับการจดจำท่ามือจับวัตถุ ซึ่งได้ฝึกเรียนรู้และตรวจสอบโดยใช้ข้อมูลจาก การจับวัตถุท่ามือละ 5 ชนิด ทำการจับชนิดละ 40 ครั้ง รวมเป็นท่ามือละ 200 ครั้ง ทำให้มีข้อมูลการจับ วัตถุทั้งหมด 4000 ข้อมูล ซึ่งนำมาแบ่งออกเป็น 50% สำหรับสอนให้ระบบเรียนรู้และอีก 50% สำหรับ ทดสอบประสิทธิภาพ โดยทดสอบ 10 ครั้ง และได้ผลการทดสอบเป็น Confuse Matrix ดังรูปที่ 4.4



รูปที่ 4.4 ผลการทดสอบถุงมือเซนเซอร์แบบ 16 DoF ด้วยวิธีการ Nearest interpolation

โดยมีความถูกต้องเฉลี่ยเท่ากับ 99.70% โดยท่ามือที่มีความถูกต้อง 100 % มี 12 ท่า คือ Large Diameter, Thump Index Finger, Thump 4Finger, Writing Tripod, Tip Pinch, Power Sphere, Power Disk, Precision Disk, Lateral Tripod, Parallel Extension, Literal Pinch และ Ring ส่วนท่ามือที่ค่าความถูกต้องในการทำนายต่ำสุดคือ Small Diameter มีค่าความถูกต้อง 98.4 % ดังกราฟแสดงความถูกต้องในการจดจำท่ามือจับทั้ง 20 ท่ามือ ในรูปที่ 4.5



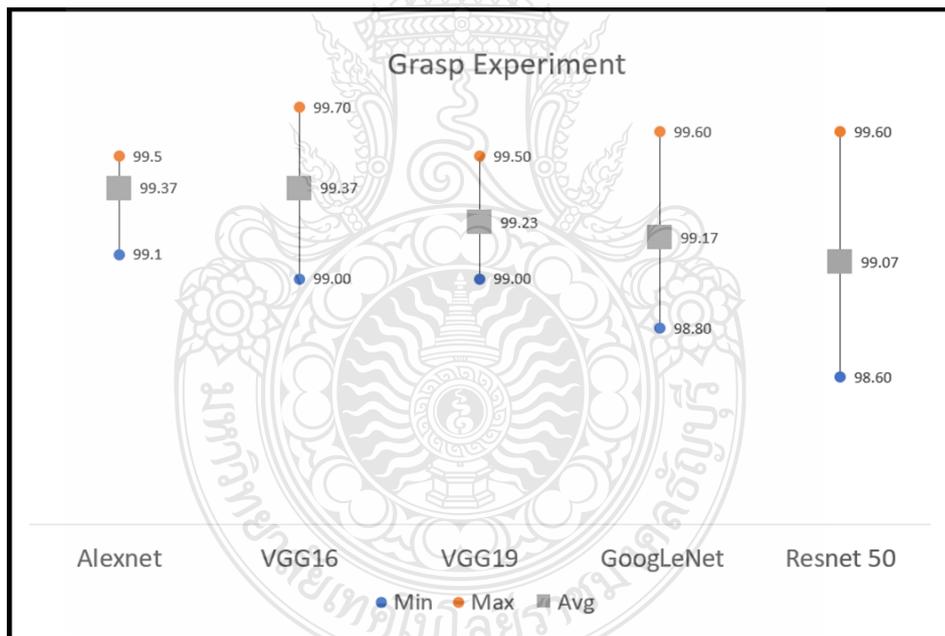
รูปที่ 4.5 กราฟแสดงความถูกต้องในการจดจำท่ามือจับวัตถุของถุงมือเซนเซอร์ทั้ง 20 ท่ามือ

เมื่อพิจารณาผลการทดลองที่ได้พบว่าการทำนายของ Small Diameter บางครั้งถูกทำนายเป็นท่า Fixed Hook สาเหตุเนื่องจาก 2 ท่านี้มีการจัดเรียงนิ้วมือขณะจับวัตถุที่คล้ายกันมากและท่าที่ถูกทำนายผลบวกสูงมากที่สุดคือท่า Power Disk และท่า Tip Pinch ซึ่งท่ามือจับมีการจัดเรียงนิ้วและอง นิ้วมือบางครั้งคล้ายกับท่า Small Diameter และ Thumb 3 Finger ตามลำดับ โดยสัญญาณที่ได้จาก ถุงมือเซนเซอร์ซึ่งอ่านผ่านวงจรมัลติเพล็กซ์และวงจรบัฟเฟอร์แบบปรับได้ วงจรได้ออกแบบมาให้มีความไวอยู่ในช่วง 0-5 โวลต์ เพื่อให้เหมาะกับวงจร A/D ในไมโครคอนโทรลเลอร์ ช่วงไดนามิกของสัญญาณ จากการอินพุต 0-90 องศาอยู่ที่ 1.5 โวลต์ ค่าความผิดพลาดของค่าความต้านทานเริ่มต้นของเซนเซอร์ แต่ละตัวและมุมเริ่มต้นของเซนเซอร์ที่ถูกติดตั้งเข้ากับนิ้วมือแต่ละตัวไม่ได้อยู่ที่ 0 องศา ส่งผลให้ค่า แรงดันขาออก V_{out} เริ่มต้นจึงไม่อยู่ในตำแหน่งเดียวกัน อีกสิ่งหนึ่งคือค่าความไม่เชิงเส้นและความไม่แน่นอนของค่าความต้านทานของเซนเซอร์อันเกิดมาจากกระบวนการผลิต จึงส่งผลให้เกิดเหตุการณ์ ดังกล่าว ซึ่งไม่มีผลกับกระบวนการวิเคราะห์ของงานวิจัยนี้ เนื่องจากการใช้โครงข่ายประสาท CNN เป็น

ส่วนรู้จำ ซึ่งข้อมูลภาพอินพุตที่ป้อนเข้าจากเซนเซอร์แต่ละตัวมีความอิสระต่อกันและการจับวัตถุแต่ละท่ามือมีการจัดเรียงนิ้วและมุมองของนิ้วมือแตกต่างกัน ทำให้สัญญาณที่ได้จากความต้านทานแบบโค้งงอจึงแตกต่างกัน เมื่อแปลงเป็นข้อมูลภาพทำให้ผลที่ได้มีความชัดเจนในด้านความแตกต่างของภาพ ซึ่งประสิทธิภาพในการจดจำในงานวิจัยนี้อยู่ที่ 99.7% จากการทดสอบด้วยท่ามือจับวัตถุ 20 ท่ามือ

4.2.2 ผลการทดลองวิธีการ Gramian Angular Fields

การทดสอบประสิทธิภาพของมือเซนเซอร์แบบ 16 DoF ในการแปลงภาพจากข้อมูลอนุกรมเวลาหนึ่งมิติเป็นภาพสองมิติด้วยเทคนิค GAF ในหัวข้อนี้ใช้ข้อมูลที่น่ามาทดสอบเหมือนกับการทดลองในหัวข้อ 4.2.1 แต่มีการเพิ่มการทดลองอีกสองหัวข้อ คือ การรู้จำสัญลักษณ์มือและการจับวัตถุตามลักษณะรูปทรง โดยใช้โมเดลของโครงข่ายประสาท CNN เพิ่มเติมจากใช้เพียง Alexnet ในการทดลองหัวข้อ 4.2.1 ในการทดลองนี้ได้เพิ่มอีก 4 โมเดล คือ VGG 16, VGG19, GoogLeNet และ Resnet 50 รวมทั้งสิ้น 5 โมเดล โดยผลทดสอบประสิทธิภาพการจดจำท่ามือจับวัตถุ 20 ท่ามือ ซึ่งใช้ข้อมูลการทดลองเดียวกันกับวิธีการ Nearest interpolation เป็นดังรูปที่ 4.6



รูปที่ 4.6 กราฟความถูกต้องในการใช้ถูงมือเซนเซอร์จดจำท่ามือจับวัตถุ 20 ท่ามือ

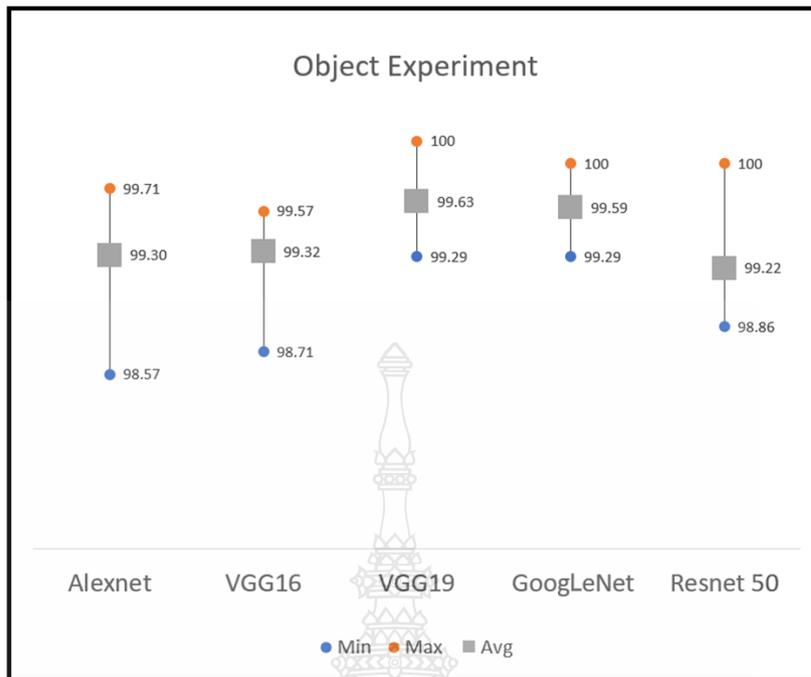
ส่วนการทดลองที่สองเป็นการจดจำท่าสัญลักษณ์ภาษามือขณะทำการสวมถูงมือเซนเซอร์ โดยการทดลองจะใช้ท่าสัญลักษณ์มือจำนวน 12 ท่ามือ [47] ดังนี้ ท่ามือพร้อม (Hand Rest), ท่ากำมือ (Hand Close), ท่าเลขศูนย์ (Num 0), ท่าเลขหนึ่ง (Num 1), ท่าเลขสอง (Num 2), ท่าเลขสาม (Num 3), ท่าเลขสี่ (Num 4), ท่าเลขห้า (Num 5), ท่าเลขหก (Num 6), ท่าเลขเจ็ด (Num 7), ท่าเลขแปด

(Num 8) และท่าเลขเก้า (Num 9) หลังจากใช้โมเดลของโครงข่ายประสาทแบบ CNN ทั้ง 5 โมเดล เรียนรู้และจดจำท่าภาษามือทั้ง 12 ท่ามือแล้ว ได้ผลการทดลองดังรูปที่ 4.7



รูปที่ 4.7 กราฟแสดงความถูกต้องในการใช้ข้อมือเซนเซอร์จดจำท่าภาษามือจำนวน 12 ท่ามือ

สุดท้ายเป็นการทดลองที่ 3 ในหัวข้อการจับตามลักษณะรูปทรงวัตถุ ซึ่งการทดลองนี้ได้แบ่งรูปทรงของวัตถุออกเป็น 4 ประเภท คือ รูปทรง Prismatic, รูปทรง Round, รูปทรง Flat และ รูปทรง Thin โดยการกำหนดลักษณะรูปทรงแต่ละประเภทได้กำหนดขนาดในมิติต่าง ๆ ดังตารางที่ 2.1 การจับวัตถุแต่ละรูปทรงมีการใช้ท่า มือแตกต่างกันไปอยู่ที่การจัดเรียงนิ้วมือเพื่อรองรับการจับวัตถุและการนำไปใช้งาน จากการจัดแบ่งลักษณะของวัตถุจะได้ 4 Class ตามท่ามือนี้นี้ Class 1 รูปทรง Prismatic ประกอบด้วยท่าจับวัตถุ 5 ท่ามือจับ คือ Large Diameter, Medium Wrap, Fixed Hook, Index Finger Extension และ Small Diameter ส่วน Class 2 รูปทรง Round ประกอบด้วยท่าจับวัตถุ 4 ท่ามือจับ คือ Power Sphere, Power Disk, Tripod และ Lateral Tripod ส่วน Class 3 รูปทรง Flat ประกอบด้วยท่าจับวัตถุ 2 ท่ามือจับ คือ Parallel Extension และ Extension Type และสุดท้าย Class 4 รูปทรง Thin ประกอบด้วยท่าจับวัตถุ 3 ท่ามือจับ คือ Thumb Index Finger, Tip Pinch และ Literal Pinch หลังจากใช้โมเดลของโครงข่ายประสาทแบบ CNN ทั้ง 5 โมเดล เรียนรู้และจดจำลักษณะการใช้มือจับตามลักษณะรูปทรงวัตถุได้ผลการทดลองดังรูปที่ 4.8



รูปที่ 4.8 กราฟแสดงความถูกต้องในการใช้ถุงมือเซนเซอร์จดจำการจับวัตถุตามลักษณะรูปทรง

4.3 ผลการทดสอบตรวจจับและจดจำท่ามือจับวัตถุ

อัลกอริทึมที่เสนอดำเนินการโดยใช้ AGCWD และความต้านทานแบบโค้งงอเพื่อเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นทริกซ์ที่มีสภาพแวดล้อมความเข้มแสงสว่างต่ำ โดยฝึกการเรียนรู้ของระบบและทดสอบกับแบบจำลองสภาพแวดล้อมในระดับความเข้มแสงสว่างต่าง ๆ การทดลองนี้ใช้โครงข่ายประสาทคอนโวลูชันเชิงลึกแบบ YOLOv2 ร่วมกับ Pertained CNN จำนวน 3 โมเดล ในการฝึกการเรียนรู้ตรวจจับและจดจำ โดย Pertained CNN ใช้สำหรับการสกัดคุณลักษณะของภาพ ได้แก่ VGG16 [49], ResNet 50 [50] และ GooLeNet [51] หลังจากกระบวนการสกัดคุณลักษณะของภาพของแต่ละโมเดล Pertained CNN แล้วจะส่ง YOLOv2 สร้าง Anchor boxes หรือ กล่องสี่เหลี่ยมหลายๆ ขนาดที่ครอบคลุม Feature map ในตำแหน่งต่าง ๆ ของภาพที่เรา Label เป็น Ground Truth โดยจะ 3 พารามิเตอร์ในการทำนาย คือ Intersection over union, Anchor box offsets การปรับตำแหน่งของ Anchor boxes และ Class probability คาดการณ์ของการ Label ของคลาสที่กำหนดให้กับแต่ละ Anchor boxes ที่กำหนดไว้ล่วงหน้าในแต่ละตำแหน่งใน Feature map และตำแหน่งที่ได้รับการปรับ ระบบก็จะตีกรอบเพื่อที่จับคู่กับคลาสที่เรา Label ไว้ โดย Feature map จะได้จาก Pretrained CNN การทำงานเริ่มด้วยนำชุดภาพฐานข้อมูลอินพุตป้อนเข้าสู่ระบบ ซึ่งจะมีการแบ่งข้อมูลภาพออกเป็น 3 ส่วน คือ ส่วนแรกใช้สำหรับฝึกการเรียนรู้ของระบบร้อยละ 60 ส่วนที่สองสำหรับใช้ทดสอบผลการฝึกอบรมร้อยละ 10 และส่วนสุดท้ายที่เหลือร้อยละ 30 ใช้สำหรับการทดสอบ

ประสิทธิภาพการตรวจจับและจดจำภาพท่ามือจับในระบบอีโก้เซ็นเซอร์ขึ้น โดยภาพที่ใช้ในการทดสอบ จะถูกทำให้ระดับความเข้มของแสงสว่างมีดลงโดยการปรับค่าแกมมาและค่าความคมชัด ซึ่งการทดสอบ ประสิทธิภาพการทำงานของเทคนิคที่นำเสนอจะปรับสภาพแวดล้อมให้มีความเข้มแสงสว่างลดลง ออกเป็น 3 รูปแบบ คือ การปรับค่าแกมมา (γ) จากระดับ 1 ถึง 6, การปรับความคมชัด (α) จากระดับ 0.5 ถึง 1, และการปรับผสมทั้งแกมมาและความคมชัดพร้อมกัน โดยระบบเริ่มทำงานจากการสกัดคุณลักษณะภาพอินพุตที่เป็นภาพท่ามือจับในส่วนเฉพาะที่ได้ทำ Label ตำแหน่งมือจับไว้ได้เป็น Feature map เพื่อให้โครงข่ายประสาท CNN เรียนรู้ในส่วนที่สนใจเท่านั้น

4.3.1 การวัดประสิทธิภาพการตรวจจับและจดจำ

หลังจากฝึกการเรียนรู้ของโครงข่าย CNN ทำให้ระบบสามารถจำแนกความแตกต่างของคุณลักษณะของแต่ละท่ามือจับและสามารถตรวจจับตำแหน่งมือที่ปรากฏในฉากออกจากพื้นหลังของภาพที่ใช้สำหรับทดสอบ ซึ่งผลที่ได้จากการตรวจจับและจดจำจะถูกคำนวณหาพื้นที่ท่ามือที่ตรวจจับได้ว่ามืออยู่จริงหรือไม่และขนาดมากน้อยเพียงไรจากความน่าจะเป็นของท่ามือที่ตรวจจับได้กับท่ามือจริงในภาพที่ได้ทำ Label ไว้ ด้วยกระบวนการ IoU [52] ซึ่งคำนวณจากอัตราส่วนของพื้นที่ซ้อนทับของพื้นที่ที่ทำนายกับพื้นที่จริงต่อพื้นที่รวมของพื้นที่ที่ทำนายกับพื้นที่จริง หลังจากนั้นจะคำนวณประสิทธิภาพการตรวจจับและจดจำ [53] โดยใช้ตัวชี้วัดจำนวน 5 พารามิเตอร์ ได้แก่ ค่า Intersection over Union, Precision, Recall, Average Precision (AP) และ Mean Average Precision (mAP) ดังนี้

4.3.1.1. Intersection over Union ซึ่งสามารถหาได้จากอัตราส่วนของพื้นที่ที่ทับซ้อน (Area of Overlap) ระหว่างบริเวณที่ทำนาย (Prediction Area) กับบริเวณค่าความจริง (Ground Truth Area) ที่ได้ทำ Label ไว้หารด้วยพื้นที่รวมทั้งหมด (Area of Union) ระหว่างบริเวณที่ทำนายกับบริเวณค่าความจริง ดังแสดงในสมการที่ 4.1

$$IoU = \frac{\text{Ground Truth} \cap \text{Prediction}}{\text{Ground Truth} \cup \text{Prediction}} \quad (4.1)$$

ในกรณีที่ท่ามือจับที่ตรวจจับได้มีขนาดพื้นที่เท่ากับขนาดที่ทำการ Label ท่ามือจับไว้พอดีจะได้ค่า IoU เท่ากับ 1 โดยในงานวิจัยนี้ได้กำหนดให้กรณีที่ ค่า $IoU \geq 0.5$ เป็นผลการตรวจจับท่ามือจับถูกต้อง หรือ True Positive (TP), ในกรณีที่ค่า $IoU < 0.5$ ให้เป็นผลการตรวจจับท่ามือผิดพลาด หรือ False Positive (FP) และถ้ากรณีที่ท่ามือจับที่ Label ปรากฏในภาพแต่ระบบไม่สามารถตรวจจับได้จะได้เป็น False Negative (FN) สำหรับ True Negative (TN) จะเป็นทุกส่วนของภาพที่ไม่ได้ทำ Label ไว้ ซึ่งในงานด้านการตรวจจับวัตถุไม่ได้สนใจในค่า TN

4.3.1.2. Precision คือ อัตราส่วนของภาพท่ามือจับที่ตรวจจับได้ถูกต้องหารด้วยภาพท่ามือจับที่ตรวจจับได้ถูกต้องรวมกับภาพท่ามือจับที่ตรวจจับผิดพลาด สามารถหาได้ตามสมการที่ 4.2

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4.2)$$

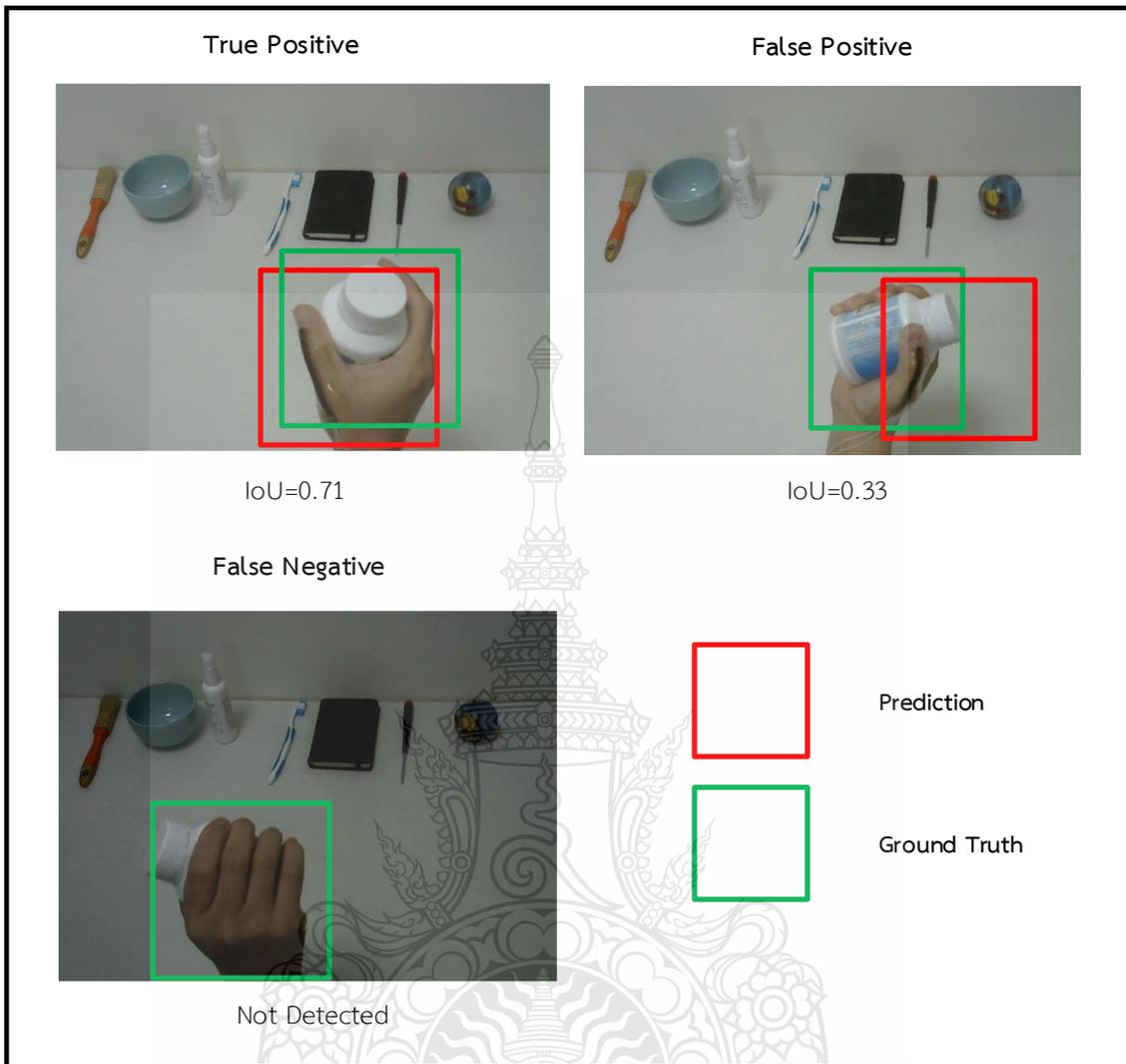
4.3.1.3. Recall คือ อัตราส่วนของภาพท่ามือจับที่ตรวจจับได้ถูกต้องหารด้วยภาพท่ามือจับที่ตรวจจับได้ถูกต้องรวมกับภาพท่ามือจับที่ระบบไม่สามารถตรวจจับได้ สามารถหาได้ตามสมการที่ 4.3

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4.3)$$

4.3.1.4. Average Precision (AP) คือ ค่าเฉลี่ยการตรวจจับของแต่ละภาพมือจับวัตถุโดยในการประเมินประสิทธิภาพการตรวจจับโดยทั่วไปจะใช้เส้น Precision-Recall curve ซึ่งค่า AP เป็นพื้นที่ใต้กราฟของแต่ละท่ามือจับ การคำนวณค่า AP ทำให้ง่ายต่อการเปรียบเทียบประสิทธิภาพกับแต่ละท่ามือจับที่ใช้ในการทดลอง

4.3.1.5. mean Average Precision (mAP) คือ ค่าเฉลี่ยการตรวจจับของภาพทั้งหมดในทุกท่ามือจับวัตถุที่ใช้ในการทดลอง

ตัวอย่างการตรวจจับและจดจำท่ามือจับวัตถุที่ผ่านการทดสอบประสิทธิภาพเป็นดังรูปที่ 4.9 โดยเป็นภาพท่ามือ Large Diameter ขณะจับขวดในมุมมองการมองผ่านระบบอีโก้เซ็นเซอร์ ซึ่งจากรูปแสดงให้เห็นการวัดประสิทธิภาพในพารามิเตอร์ True Positive คือ ภาพที่ระบบสามารถตรวจจับท่ามือจับที่ใช้จับขวดได้ถูกต้องและอัตราส่วนของพื้นที่ที่ทับซ้อนของพื้นที่ที่ทำนายกับพื้นที่จริงที่ได้ทำ Label ไว้ต่อพื้นที่รวมของพื้นที่ทั้งสองมีค่ามากกว่าหรือเท่ากับ 0.5 ส่วน False Positive คือ ภาพที่ระบบสามารถตรวจจับท่ามือจับที่ใช้จับขวดได้ถูกต้องและอัตราส่วนของพื้นที่ที่ทับซ้อนของพื้นที่ที่ทำนายกับพื้นที่จริงที่ได้ทำ Label ไว้ต่อพื้นที่รวมของพื้นที่ทั้งสองมีค่าน้อยกว่า 0.5 และสุดท้าย False Negative คือ มีท่ามือจับที่ใช้จับขวดปรากฏในภาพและได้ทำ Label ไว้ แต่ระบบไม่สามารถ ตรวจจับท่ามือนี้ได้ อันเนื่องจากมุมมองของภาพที่ทดสอบเห็นรายละเอียดท่ามือน้อยหรือภาพมีความเข้มแสงสว่างต่ำมากจนไม่สามารถตรวจจับท่ามือจับได้ ตัวอย่างลักษณะภาพที่ตรวจจับและผลการทำนายท่ามือจับวัตถุที่ได้ในการใช้ค่า IoU เท่ากับ 0.5 เป็นดังรูปที่ 4.9



รูปที่ 4.9 ตัวอย่างภาพที่ตรวจจับและผลการทำนายท่ามือจับวัตถุที่ค่า IoU เท่ากับ 0.5

4.3.2 ผลการทดสอบตรวจจับและจดจำ

หลังนำภาพทดสอบที่มีการปรับความเข้มแสงสว่างในระดับการปรับค่าแกมมา (γ) และค่าความคมชัด (α) ต่าง ๆ ระบบจะทำนายท่ามือจับวัตถุที่ปรากฏในภาพแล้วคำนวณพื้นที่ทับซ้อนของพื้นที่ที่ทำนายกับพื้นที่จริงที่ได้ทำ Label ไว้ จากนั้นจะคำนวณค่า IoU กับ ค่า Threshold ที่ตั้งไว้ ซึ่งงานวิจัยนี้ใช้ค่า IoU เท่ากับ 0.5 เมื่อได้ตรวจจับและทำนายผลในทุกภาพของแต่ละท่ามือแล้วก็จะคำนวณค่า Precision และ Average Precision ในแต่ละภาพมือจับวัตถุ โดยผลการตรวจจับและจดจำท่ามือจับวัตถุที่ค่า IoU เท่ากับ 0.5 ด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN โมเดลทั้งสาม คือ VGG16, Resnet 50 และ GoogLeNet ในสภาวะความเข้มแสงสว่างที่ระดับค่า Gamma (γ) และ Contrast (α) ที่มีนัยสำคัญเป็นดังตารางที่ 4.2-4.3 ตามลำดับ

ตารางที่ 4.1 ผล Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ
Pertained CNN โมเดล VGG16

Grasp	Illumination Condition						
	$\gamma=1$	$\gamma=1$	$\gamma=1$	$\gamma=2$	$\gamma=2$	$\gamma=3$	$\gamma=3$
	$\alpha=1$	$\alpha=0.8$	$\alpha=0.6$	$\alpha=1$	$\alpha=0.8$	$\alpha=1$	$\alpha=0.8$
Large Diameter	0.89	0.88	0.84	0.83	0.87	0.57	0.55
Medium Wrap	0.77	0.78	0.80	0.77	0.73	0.59	0.38
Small Diameter	0.80	0.81	0.80	0.82	0.63	0.59	0.35
Fixed Hook	0.88	0.83	0.80	0.70	0.74	0.43	0.36
Index Finger Extension	0.93	0.93	0.86	0.41	0.34	0.05	0.11
Thumb Index Finger	0.84	0.84	0.83	0.78	0.59	0.46	0.22
Thumb 2 Finger	0.64	0.59	0.58	0.74	0.51	0.58	0.10
Writing Tripod	0.90	0.88	0.82	0.84	0.72	0.49	0.31
Tip Pinch	0.89	0.91	0.87	0.85	0.75	0.68	0.23
Power Sphere	1.00	1.00	1.00	1.00	0.97	0.89	0.66
Power Disk	0.79	0.79	0.78	0.67	0.62	0.53	0.48
Precision Disk	0.77	0.75	0.71	0.73	0.81	0.59	0.61
Tripod	0.92	0.94	0.91	0.85	0.83	0.53	0.41
Lateral Tripod	0.90	0.87	0.87	0.85	0.73	0.53	0.32
Parallel Extension	0.79	0.83	0.85	0.21	0.28	0.02	0.00
Extension Type	0.89	0.90	0.88	0.75	0.79	0.45	0.45
Literal Pinch	0.93	0.93	0.92	0.81	0.47	0.28	0.15
Ring	0.88	0.88	0.84	0.80	0.86	0.59	0.53

ตารางที่ 4.2 ผล Average Precision ของการตรวจจับและ ละเอียดด้วยเทคนิค YOLOv2 ร่วมกับ
Pertained CNN โมเดล ResNet50

Grasp	Illumination Condition						
	$\gamma=1$	$\gamma=1$	$\gamma=1$	$\gamma=2$	$\gamma=2$	$\gamma=3$	$\gamma=3$
	$\alpha=1$	$\alpha=0.8$	$\alpha=0.6$	$\alpha=1$	$\alpha=0.8$	$\alpha=1$	$\alpha=0.8$
Large Diameter	0.76	0.76	0.76	0.64	0.61	0.57	0.47
Medium Wrap	0.62	0.71	0.70	0.63	0.63	0.59	0.39
Small Diameter	0.61	0.76	0.74	0.62	0.66	0.59	0.42
Fixed Hook	0.87	0.78	0.78	0.52	0.66	0.43	0.25
Index Finger Extension	0.95	0.89	0.88	0.63	0.73	0.05	0.02
Thumb Index Finger	0.73	0.32	0.32	0.16	0.63	0.46	0.32
Thumb 2 Finger	0.78	0.75	0.73	0.54	0.44	0.58	0.07
Writing Tripod	0.84	0.85	0.85	0.59	0.75	0.49	0.29
Tip Pinch	0.94	0.91	0.91	0.81	0.75	0.68	0.57
Power Sphere	1.00	1.00	1.00	1.00	0.97	0.89	0.89
Power Disk	0.61	0.78	0.77	0.65	0.51	0.53	0.37
Precision Disk	0.80	0.83	0.83	0.76	0.72	0.59	0.32
Tripod	0.85	0.67	0.68	0.54	0.71	0.53	0.46
Lateral Tripod	0.65	0.89	0.88	0.86	0.61	0.53	0.33
Parallel Extension	0.81	0.83	0.83	0.74	0.65	0.02	0.00
Extension Type	0.77	0.87	0.87	0.73	0.45	0.45	0.19
Literal Pinch	0.75	0.57	0.52	0.22	0.55	0.28	0.33
Ring	0.89	0.88	0.86	0.75	0.83	0.59	0.45

ตารางที่ 4.3 ผล Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN โมเดล GoogLeNet

Grasp	Illumination Condition							
	$\gamma=1$	$\gamma=1$	$\gamma=1$	$\gamma=2$	$\gamma=2$	$\gamma=3$	$\gamma=3$	
	$\alpha=1$	$\alpha=0.8$	$\alpha=0.6$	$\alpha=1$	$\alpha=0.8$	$\alpha=1$	$\alpha=0.8$	
Large Diameter	0.77	0.77	0.70	0.62	0.56	0.46	0.45	
Medium Wrap	0.54	0.54	0.55	0.52	0.51	0.40	0.33	
Small Diameter	0.67	0.67	0.65	0.54	0.53	0.41	0.43	
Fixed Hook	0.92	0.92	0.90	0.84	0.82	0.60	0.49	
Index Finger Extension	0.92	0.92	0.82	0.63	0.50	0.30	0.36	
Thumb Index Finger	0.47	0.48	0.49	0.56	0.57	0.57	0.48	
Thumb 2 Finger	0.29	0.24	0.26	0.27	0.25	0.14	0.12	
Writing Tripod	0.76	0.73	0.72	0.70	0.69	0.44	0.40	
Tip Pinch	0.68	0.66	0.62	0.70	0.64	0.57	0.47	
Power Sphere	1.00	1.00	1.00	1.00	1.00	0.97	0.95	
Power Disk	0.65	0.64	0.64	0.53	0.52	0.54	0.55	
Precision Disk	0.66	0.65	0.63	0.63	0.63	0.62	0.59	
Tripod	0.81	0.83	0.82	0.85	0.85	0.83	0.72	
Lateral Tripod	0.76	0.73	0.70	0.69	0.66	0.53	0.45	
Parallel Extension	0.83	0.85	0.72	0.30	0.33	0.12	0.10	
Extension Type	0.82	0.82	0.80	0.79	0.75	0.59	0.51	
Literal Pinch	0.85	0.86	0.84	0.77	0.69	0.46	0.42	
Ring	0.84	0.82	0.83	0.83	0.83	0.68	0.57	

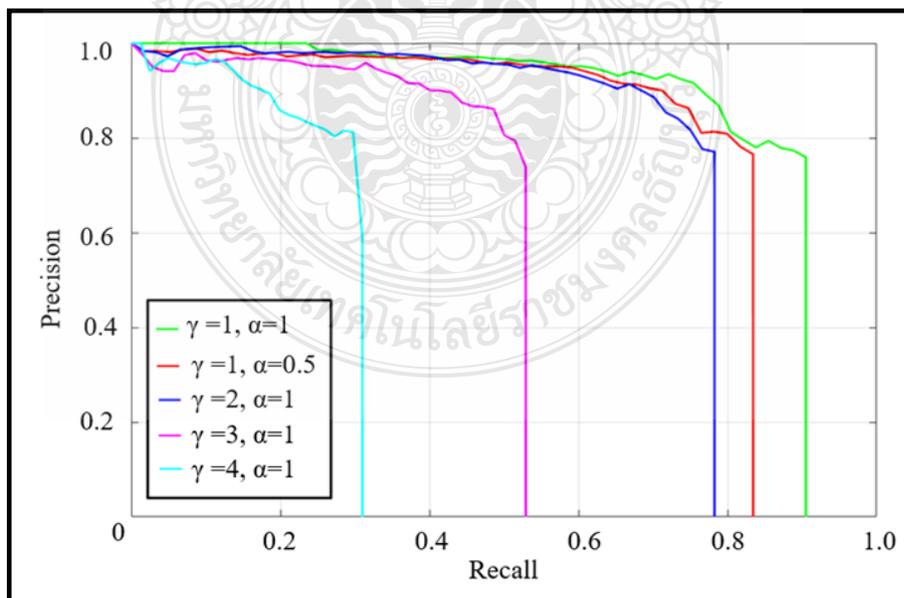
จากผลการทดลองพบว่า Pertained CNN โมเดล VGG16 ให้ผลค่า Average Precision โดยส่วนใหญ่ดีกว่าโมเดล Resnet 50 และ GoogLeNet และเพื่อให้ง่ายต่อการพิจารณาเปรียบเทียบประสิทธิภาพผลการตรวจจับและจดจำท่ามือจับวัตถุที่ได้ในแต่ละโมเดลจะใช้ค่า mAP ซึ่งเป็นค่าเฉลี่ยรวมในทุกท่ามือจับ ซึ่งให้ผลในสภาวะความเข้มแสงสว่างที่ระดับค่า Gamma (γ) และ Contrast (α) ที่มีนัยสำคัญเป็นดังตารางที่ 4.5

ตารางที่ 4.4 ผล Mean Average Precision ของการตรวจจับและจดจำด้วยเทคนิค YOLOv2 ร่วมกับ Pertained CNN ในทุกท่ามือจับวัตถุ

Pertained CNN Model	Illumination Condition							
	$\gamma=1$	$\gamma=1$	$\gamma=1$	$\gamma=2$	$\gamma=2$	$\gamma=3$	$\gamma=3$	
	$\alpha=1$	$\alpha=0.8$	$\alpha=0.6$	$\alpha=1$	$\alpha=0.8$	$\alpha=1$	$\alpha=0.8$	
VGG16	0.86	0.85	0.83	0.75	0.68	0.49	0.35	
Resnet 50	0.79	0.78	0.77	0.63	0.66	0.49	0.34	
GoogLeNet	0.74	0.73	0.71	0.65	0.63	0.51	0.47	

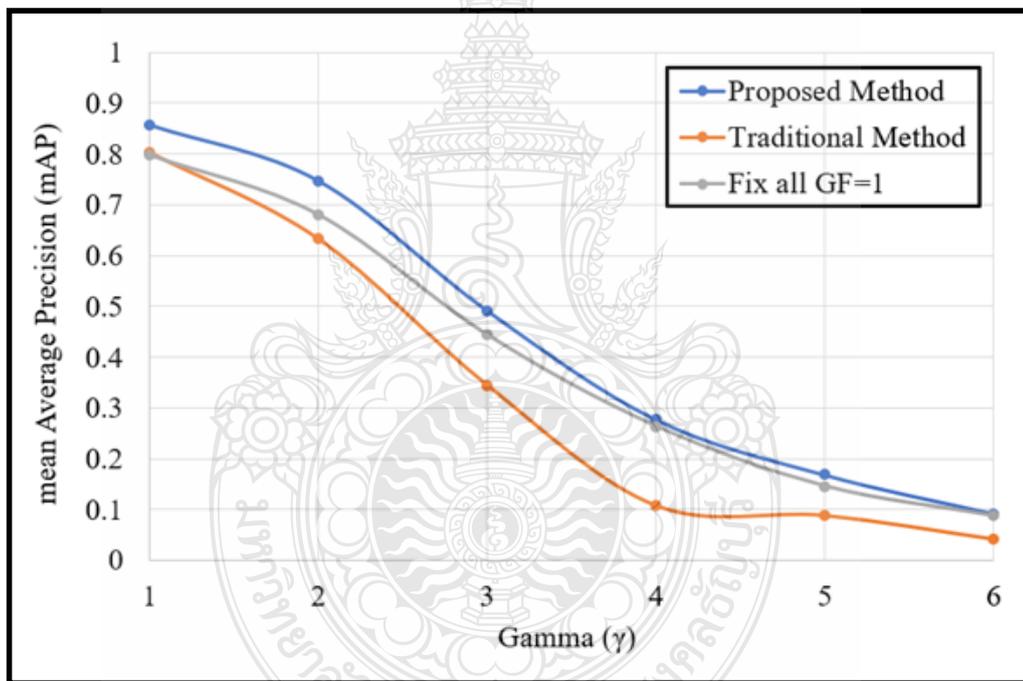
4.3.3 การประเมินประสิทธิภาพ

การประเมินประสิทธิภาพการทำงานของเทคนิคที่นำเสนอเพื่อให้ง่ายต่อการพิจารณาผลการทดสอบจะใช้กราฟ Precision-Recall curve ซึ่งค่า AP เป็นพื้นที่ใต้กราฟของแต่ละท่ามือจับเป็นอีกหนึ่งวิธีที่ให้การแสดงผลเข้าใจง่ายต่อการประเมินประสิทธิภาพการตรวจจับและจดจำในความเข้มแสงสว่างสถานะต่าง ๆ โดยการเขียน Precision-Recall curve จะใช้ค่า Precision อยู่ในแกน Y และ ค่า Recall อยู่ในแกน X การประเมินจะพิจารณา Area Under the Curve (AUC) หรือพื้นที่ใต้กราฟของการตรวจจับและจดจำ ถ้ามีประสิทธิภาพที่ดีจะได้ค่า AUC มากกว่า โดยมีค่าสูงสุดเท่ากับ 1 ดังในรูปที่ 4.10 เป็น Precision-Recall curve ของโมเดล VGG 16 ที่ให้ผลการทดสอบที่ดีที่สุดในงานวิจัยนี้



รูปที่ 4.10 Precision-Recall curve of VGG 16 ที่สภาวะระดับความสว่างต่าง ๆ

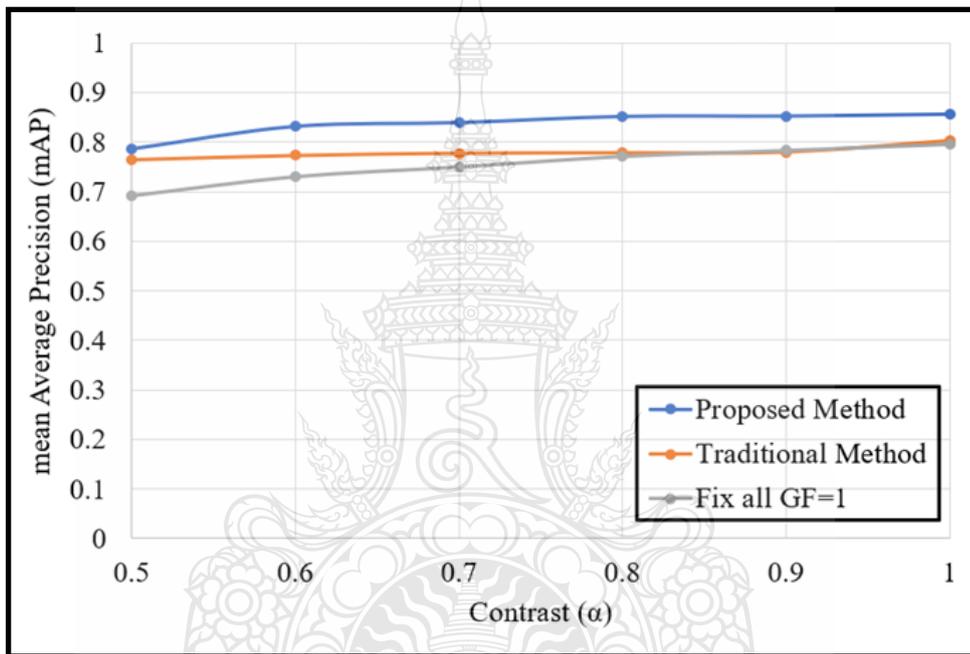
ซึ่งการทดสอบประสิทธิภาพของเทคนิคที่นำเสนอได้จำลองค่าพารามิเตอร์ของภาพด้วยค่า Gamma (γ) และ Contrast (α) ที่สภาพความเข้มแสงสว่างออกเป็น 3 รูปแบบ คือ 1. การปรับค่าแกมมา (γ) จากระดับ 1 ถึง 6, 2. การปรับความคมชัด (α) จากระดับ 0.5 ถึง 1 และ 3. การปรับผสมทั้งแกมมาและความคมชัดพร้อมกันและได้ผลการทดลองในความเข้มแสงสว่างที่มีนัยสำคัญดังรูปที่ 4.10 และเมื่อพิจารณาปรับค่าพารามิเตอร์เพียงตัวเดียว ไม่ว่าจะเป็นค่าแกมมาหรือความคมชัด โดยคงที่อีกพารามิเตอร์หนึ่งไว้เพื่อสังเกตความเปลี่ยนแปลงในประสิทธิภาพการตรวจจับและจดจำโดยแบ่งการทดสอบเป็น 3 กรณี ได้แก่ การใช้เทคนิคที่เสนอ (Proposed Method) การใช้เทคนิคการเรียนรู้แบบเดิม (Traditional Method) ซึ่งไม่ได้ใช้พารามิเตอร์ Grasp Factor ที่ได้จากความต้านทานแบบโค้งงอขณะจับวัตถุและสุดท้ายให้ค่าคงที่ $gf = 1$ ของภาพทั้งหมดที่ใช้ในการทดสอบ (Fix all GF) ซึ่งในส่วนของการปรับค่าแกมมาได้ผลดังรูปที่ 4.11



รูปที่ 4.11 เปรียบเทียบผลการตรวจจับและจดจำท่ามือจับวัตถุขณะปรับค่าแกมมา

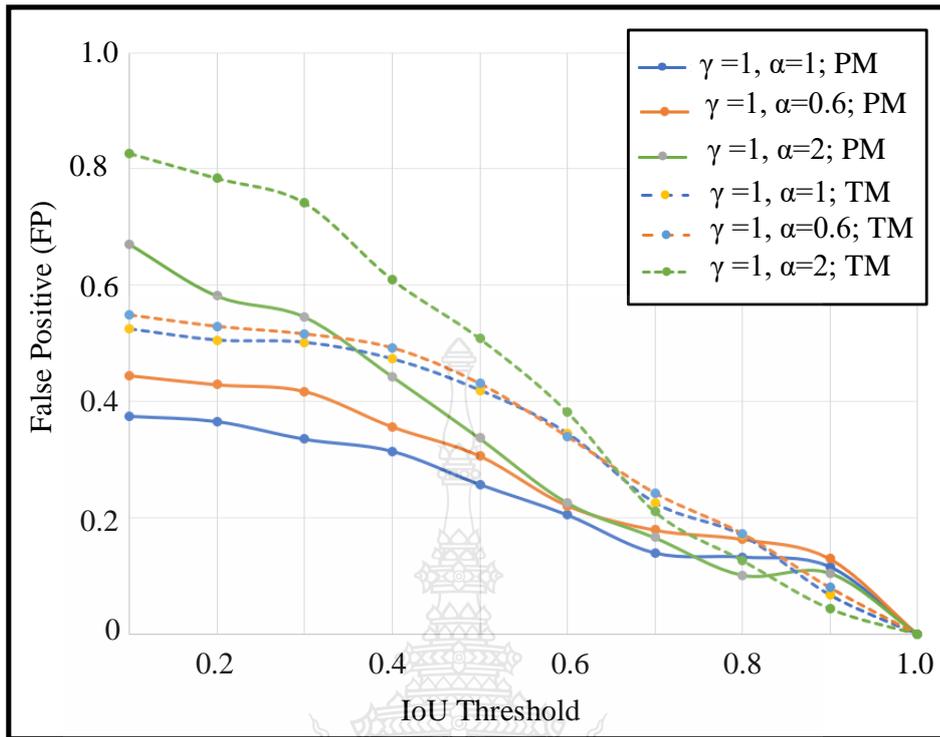
จากรูปที่ 4.11 เปรียบเทียบผลการทดลองแสดงให้เห็นว่าการปรับแกมมา γ โดยการคงที่ความคมชัด α ให้เป็น 1 ซึ่งผลการทดลองแสดงให้เห็นอย่างมีนัยสำคัญต่อประสิทธิภาพการตรวจจับและจดจำภาพท่ามือจับมากกว่าการปรับค่าความคมชัด α ในรูปที่ 4.12 โดยประสิทธิภาพการทำงานลดลงแบบเชิงเส้นตรงต่อการปรับในแต่ละ step ตามการเพิ่มค่าแกมมา γ เนื่องจากการปรับค่าแกมมาส่งผลต่อคุณลักษณะของภาพ ตามระดับความสว่างของค่าพิกเซลในระบบสี RGB ซึ่งมีการเปลี่ยนแปลงมากกว่าการปรับค่าความคมชัด α อย่างไรก็ตาม ตัวพารามิเตอร์ท่ามือจับ gf ช่วยกำหนดการปรับ

น้ำหนักของเทคนิค AGCWD ในการสร้างคุณลักษณะเฉพาะของภาพการจับแต่ละท่ามือ ซึ่งจะทำให้โครงข่ายประสาทคอนโวลูชัน YOLOv2 สามารถยังคงประสิทธิภาพการตรวจจับและจดจำได้อย่างมีนัยสำคัญ ซึ่งสังเกตได้เมื่อเปรียบเทียบกับการคงที่ค่าตัวพารามิเตอร์ท่ามือจับ gf เท่ากับ 1 ในทุกท่าทางของมือ โดยเทคนิคที่เสนอมีประสิทธิภาพการตรวจจับและจดจำที่สูงกว่า นอกจากนี้ ภาพที่ผ่านการปรับแก้มาและความคมชัดยังคงรักษาประสิทธิภาพการตรวจจับและจดจำได้ใกล้เคียงกับภาพในสถานะความเข้มแสงสว่างในระดับปกติที่ $\gamma=1$, $\alpha=1$ แม้กระทั่งทดสอบด้วยการลด $\alpha=0.5$ ดังผลการทดลองปรับค่าความคมชัดในช่วง 0.5 ถึง 1 และคงที่ค่าเกมมาในรูปแบบที่ 4.12

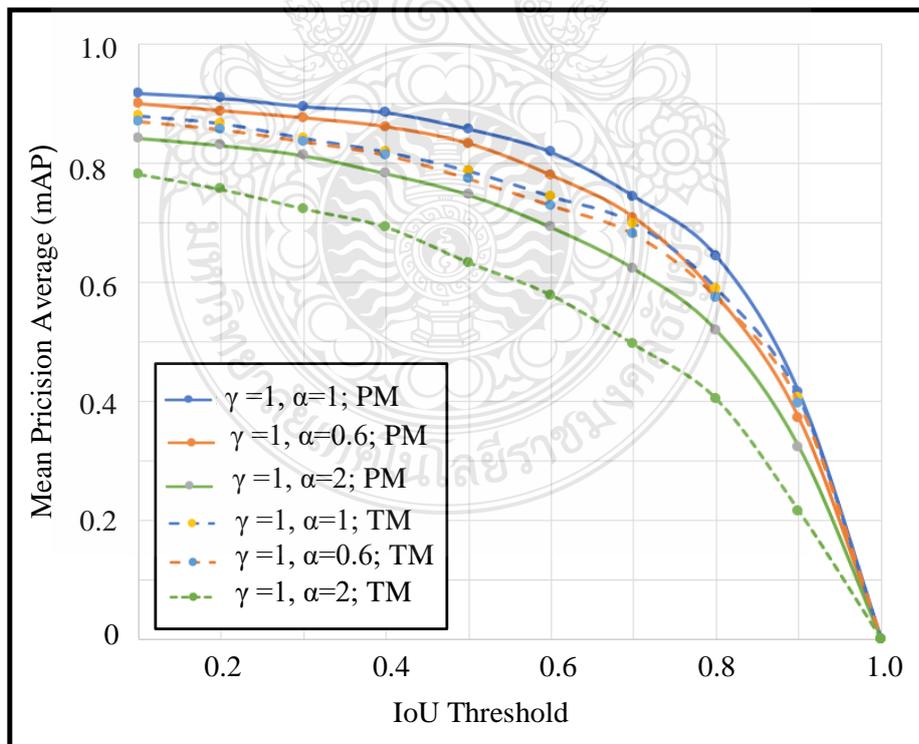


รูปที่ 4.12 เปรียบเทียบผลการตรวจจับและจดจำท่ามือจับวัตถุขณะปรับค่าความคมชัด

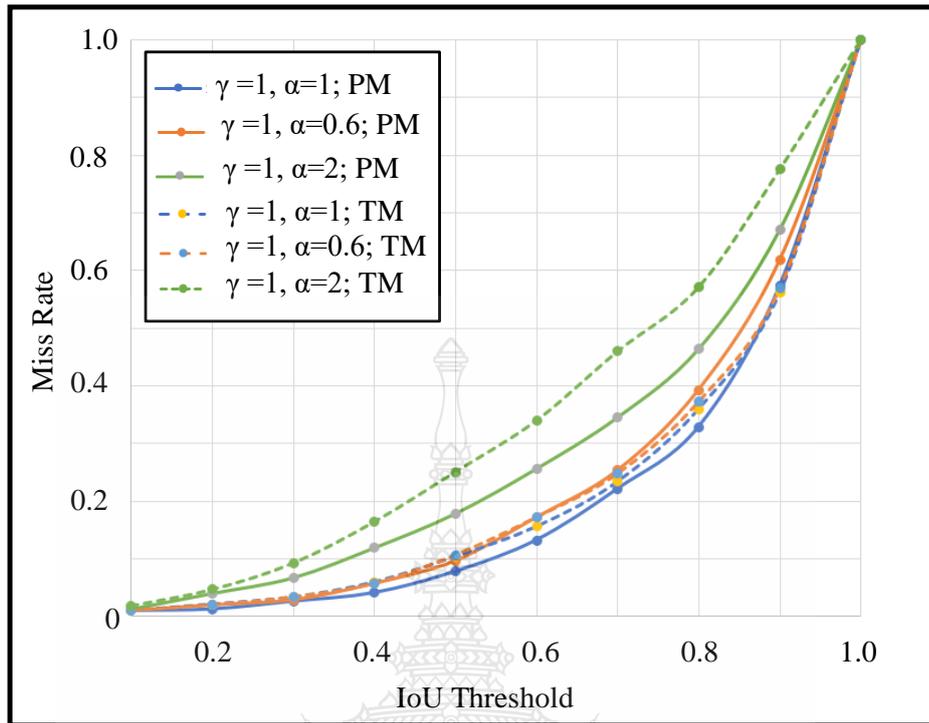
จากผลการทดลองที่ผ่านมาเป็นการใช้ค่า $IoU \geq 0.5$ ซึ่งให้เป็นผลการตรวจจับท่ามือถูกต้อง หรือ True Positive (TP) ในการตรวจจับท่ามือจับ แต่ถ้าพิจารณาตัวแปร IoU ที่มีการปรับเปลี่ยนตั้งแต่ 0.1 ถึง 1.0 เพื่อดูประสิทธิภาพของเทคนิคที่นำเสนอ เมื่อค่า IoU ที่ใช้ทดสอบมีค่าเพิ่มสูงขึ้น โดยจะใช้พารามิเตอร์ False Positive, Overlap Matching และ Miss Rate ซึ่ง False Positive เป็นการตรวจจับท่ามือถูกต้องแต่ได้ค่า $IoU < 0.5$ ส่วน Overlap Matching การซ้อนทับกันของพื้นที่ท่ามือที่ตรวจจับได้ต่อค่า mAP และสุดท้าย Miss Rate ผลการตรวจจับท่ามือที่ผิดพลาด เมื่อมีการเพิ่มขนาดพื้นที่ IoU เพิ่มขึ้น ซึ่งทั้ง 3 พารามิเตอร์ให้ผลการทดลองดังในรูปที่ 4.13-4.15 โดยได้เปรียบเทียบเทคนิคที่นำเสนอ (Proposed Method, PM) และการใช้เทคนิคแบบเดิม (Traditional Method, TM)



รูปที่ 4.13 ค่า False Positive ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับค่า IoU



รูปที่ 4.14 ผล Overlap Matching ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับค่า IoU



รูปที่ 4.15 ค่า Miss Rate ของการตรวจจับและจดจำท่ามือจับวัตถุ เมื่อปรับเปลี่ยนค่า IoU

จากผลการทดลองที่ได้ปรับเปลี่ยนค่า IoU โดยมีการเปรียบเทียบเทคนิค PM และ TM พบว่าเมื่อเพิ่มค่า IoU ที่สูงขึ้นจะทำให้ค่า Miss Rate สูงขึ้นตามไปด้วยหรือแปรผันตามกับค่า IoU แต่ค่า False Positive และ Overlap Matching จะแปรผกผันกับค่า IoU หรือ จะทำให้ผลของทั้งสองค่าลดลงเมื่อมีการเพิ่มขนาดพื้นที่ IoU ส่วนในระดับความเข้มแสงสว่างที่ $\gamma=1, \alpha=1$ เทคนิคที่นำเสนอให้ผลการทดลองที่ดีที่สุด ซึ่งจากผลการทดสอบทั้ง 3 พารามิเตอร์จะเห็นได้ว่าในระดับความเข้มแสงสว่างที่ $\gamma=1, \alpha=2$ มีประสิทธิภาพต่ำที่สุด แต่เมื่อมีการปรับปรุงภาพที่ใช้ทดสอบด้วยเทคนิคที่นำเสนอจะเห็นได้ว่าการทดสอบที่ $\gamma=1, \alpha=2$ ให้ผลการทดลองที่ดีขึ้นใกล้เคียงกับระดับ $\gamma=1, \alpha=1$ ของเทคนิคเดิม โดยเฉพาะอย่างยิ่ง ค่า False Positive ในรูปที่ 4.13 ที่ระดับ $\gamma=1, \alpha=2$ ผลการทดลองของเทคนิคที่นำเสนอให้ผลการทดลองที่ดีกว่าระดับ $\gamma=1, \alpha=1$ ของเทคนิคเดิม ในช่วงค่า IoU ที่ 0.4-0.8 ถึงแม้จะมีการเพิ่มระดับค่าแกมมาขึ้นอีก 1 Step ซึ่งผลการประเมินเปรียบเทียบทั้ง 3 พารามิเตอร์ แสดงให้เห็นว่าเทคนิคที่นำเสนอสามารถเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุได้สูงขึ้น

นอกจากนี้เมื่อนำผลการตรวจจับและจดจำทดสอบกับภาพที่ได้จำลองค่าความเข้มแสงสว่างในพารามิเตอร์ต่าง ๆ ทั้งการปรับค่า γ และ α ซึ่งหลังจากระบบรับภาพอินพุตจะทำการปรับปรุงภาพให้มีความสว่างมากขึ้นโดยใช้เทคนิค AGCWD โดยใช้น้ำหนักของพารามิเตอร์ GF ที่ได้จากค่าเฉลี่ยสัญญาณการเปลี่ยนแปลงแรงดันไฟฟ้าเอาต์พุตของวงจรความต้านทานแบบโค้งงอ แล้วส่งต่อให้เทคนิค

YOLOv2 ซึ่งมีประสิทธิภาพสูงในการตรวจจับและจดจำวัตถุในภาพ ซึ่งการทดลองจะใช้ Pretrained CNN ทั้งหมด 3 โมเดล คือ VGG16, ResNet 50 และ GooleNet เมื่อนำผลการตรวจจับและจดจำทำมือจับวัตถุประเมินผลโดยการคำนวณค่าร้อยละความถูกต้องได้ดังตารางที่ 4.6 ซึ่งเป็นผลเปรียบเทียบระหว่างเทคนิคที่นำเสนอกับเทคนิคเดิมในระดับความเข้มแสงสว่างที่มีนัยสำคัญและอัตราร้อยละความถูกต้องของการตรวจจับ (Accuracy rate) เป็นอีกหนึ่งพารามิเตอร์ที่แสดงให้เห็นว่าเทคนิคที่นำเสนอมีประสิทธิภาพดีกว่าการใช้เทคนิคเดิมเช่นกัน โดยสามารถคำนวณได้ตามสมการที่ 4.4

$$\% \text{ Accuracy rate} = \frac{\text{จำนวนภาพทำมือจับที่ตรวจจับได้ถูกต้อง}}{\text{จำนวนภาพที่ทดสอบทั้งหมด}} \times 100 \quad (4.4)$$

จำนวนภาพทำมือจับที่ตรวจจับได้ถูกต้อง คือ ภาพที่ระบบสามารถตรวจจับทำมือจับที่ใช้จับวัตถุชนิดต่าง ๆ ในทุกท่ามือที่ทดสอบได้ถูกต้องและมีอัตราส่วนของพื้นที่ที่ทับซ้อนของพื้นที่ที่ทำนายกับพื้นที่จริงที่ได้ทำ Label ไว้ต่อพื้นที่รวมของพื้นที่ทั้งสองมีค่ามากกว่าหรือเท่ากับ 0.5 ส่วนจำนวนภาพที่ทดสอบทั้งหมดเป็นจำนวนภาพที่ระบบได้แบ่งมาเพื่อใช้สำหรับการทดสอบประสิทธิภาพการตรวจจับและจดจำ ซึ่งเป็นส่วนที่เหลือจากใช้ในการฝึกเรียนรู้ของระบบ

ตารางที่ 4.5 เปรียบเทียบอัตราความถูกต้องของผลการตรวจจับและจดจำท่ามือจับระหว่างเทคนิคที่นำเสนอกับเทคนิคเดิม

Condition	Proposed Method			Traditional learning		
	VGG16	ResNet50	GoogLeNet	VGG16	ResNet50	GoogLeNet
$\gamma = 1, \alpha = 1$	86.28	80.84	76.51	81.21	77.99	70.95
$\gamma = 1, \alpha = 0.8$	85.78	79.35	76.14	80.84	77.87	70.70
$\gamma = 1, \alpha = 0.6$	83.44	76.76	74.04	79.72	76.63	70.45
$\gamma = 2, \alpha = 1$	75.23	64.03	69.47	70.33	70.95	66.01
$\gamma = 2, \alpha = 0.8$	69.84	60.69	67.49	69.34	69.96	65.76
$\gamma = 2, \alpha = 0.6$	58.22	53.64	59.46	69.22	69.83	65.51
$\gamma = 3, \alpha = 1$	50.43	40.42	55.37	34.73	45.61	49.20
$\gamma = 3, \alpha = 0.8$	35.72	35.11	49.94	33.99	44.87	48.70
$\gamma = 3, \alpha = 0.6$	27.69	28.43	43.26	33.49	44.01	48.45

4.4 การวิเคราะห์ผลการทดสอบของอัลกอริทึมที่นำเสนอ

ผลการทดสอบการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นเซอร์ที่สร้างขึ้น ซึ่งใช้เทคนิคการทำงานร่วมกันในการปรับปรุงภาพระหว่างเทคนิค AGCWD กับความต้านทานแบบโค้งงอที่ติดอยู่กับนิ้วโป้งในการปรับความสว่างของภาพและมีท่ามือจับวัตถุที่ใช้ในการทดลองจำนวน 18 ท่ามือ ส่วนการตรวจจับและจดจำใช้โมเดล Pretrained CNN ทั้งหมด 3 โมเดล คือ VGG16, ResNet 50 และ GooLeNet ร่วมกับโครงข่ายประสาทคอนโวลูชัน YOLOv2 ซึ่งได้ผลการทดลองในระดับความเข้มแสงสว่างที่มีนัยสำคัญดังตารางที่ 4.6 พบว่าผลการทดลองโมเดล VGG16 ให้ผลค่า Accuracy มากที่สุดร้อยละ 86.28 ที่ $\gamma = 1$, $\alpha = 1$ และรองลงมาเป็นโมเดล ResNet50 ร้อยละ 80.84 เมื่อพิจารณาค่า Accuracy ของผลการทดลองในแต่ละท่ามือจับพบว่า ท่ามือที่ให้ความถูกต้องสูงสุด ได้แก่ Power Sphere, Index Finger Extension และ Fixed Hook เนื่องจากการจัดเรียงนิ้วรองรับรูปร่างของวัตถุที่จับมีความแตกต่างกันมาก ซึ่งผลการทดลองสามารถตรวจจับและจดจำท่ามือเหล่านี้ได้มากกว่าร้อยละ 90 โดยมีผลตรวจจับและจดจำได้อย่างดีเยี่ยมสำหรับโครงข่าย CNN ที่ใช้สำหรับการสกัดคุณลักษณะทั้งสามโมเดล นอกจากนี้พบว่า ท่ามือจับซึ่งประกอบด้วย Thump 2 Finger, Precision Disk, Power Disk และ Lateral Tripod เป็นท่ามือที่ยากต่อการตรวจจับและจดจำเพราะการจัดเรียงของนิ้วและรูปร่างของวัตถุที่ยึดจับนั้นเหมือนกันในบางมุมมองผ่านกล้องอีโก้เซ็นเซอร์ที่สร้างขึ้น แต่ด้วยพารามิเตอร์ท่ามือจับ GF สามารถช่วยแก้ปัญหานี้ได้ ทำให้มีความถูกต้องในการตรวจจับและจดจำเป็นที่น่าพอใจ เนื่องจากความต้านทานแบบโค้งงอที่ติดอยู่ที่นิ้วโป้งได้เพิ่มข้อมูลเกี่ยวกับการเคลื่อนไหวของนิ้วที่ใช้อำนาจในการจับวัตถุ ซึ่งเป็นอีกหนึ่งพารามิเตอร์ที่ช่วยเพิ่มความสามารถในการแยกความแตกต่างของการจับมือวัตถุแต่ละท่ามือ จะเห็นได้จากประสิทธิภาพการตรวจจับและจดจำที่เพิ่มขึ้นดังในรูปที่ 4.11 และรูปที่ 4.12 โดยผลการทดลองในสภาวะความเข้มแสงสว่างปกติที่ $\gamma = 1$ และ $\alpha = 1$ มีผลการตรวจจับและจดจำ โดยใช้ YOLOv2 ให้ประสิทธิภาพสูงมากกว่าร้อยละ 80 และมีประสิทธิภาพเพิ่มขึ้นประมาณร้อยละ 6 เมื่อเทียบกับเทคนิคการเรียนรู้แบบเดิม ซึ่งไม่ได้ใช้พารามิเตอร์ GF โดยเฉพาะอย่างยิ่ง Pretrained CNN โมเดล VGG16 ซึ่งให้ผลการตรวจจับและจดจำท่ามือจับที่สูงสุตร้อยละ 86.28

สำหรับการปรับความสว่างของภาพให้มีผลโดยการเพิ่มค่า γ ขึ้น 1 step จะทำให้ความสามารถในการตรวจจับและจดจำของระบบลดลงอย่างรวดเร็วขึ้น เมื่อเทียบกับการปรับค่า α ตัวอย่างเช่น การทดสอบที่สภาวะ $\gamma = 2$, $\alpha = 1$ ให้ผลการทดลองลดลงประมาณ 10% ในการเรียนรู้ของเทคนิคแบบเดิม แต่เทคนิคที่เสนอสามารถยังคงให้ผลการทดลองเป็นที่น่าพึงพอใจ โดยให้ค่า Accuracy ได้ถึงร้อยละ 75 เมื่อเปรียบเทียบการลดลงของประสิทธิภาพการตรวจจับและจดจำระหว่างการปรับ γ และ α จากการพิจารณา Histogram ของภาพที่ผ่านการปรับลดความคมชัดพบว่าจะมีลักษณะขนาดความสูงของแต่ละความสว่างใกล้เคียงกันกับภาพต้นฉบับ แต่จะเลื่อนไปทางซ้ายมากขึ้น ทำให้ภาพมืด

ลงเท่านั้น อย่างไรก็ตาม คุณลักษณะเฉพาะของภาพนั้นยังคงคล้ายกันมากกว่าการปรับค่าแกมมา โดยตัวอย่างที่เห็นได้ชัดเจน คือ ที่สภาวะ $\gamma = 1, \alpha = 0.6$ ผลการทดลองยังคงใกล้เคียงกับการทดลองที่สภาวะ $\gamma = 1, \alpha = 1$ นอกจากนี้ เมื่อทดลองปรับพารามิเตอร์ทั้งสองพร้อม เช่น ที่สภาวะ $\gamma = 2, \alpha = 0.8$ ผลการจดจำของโมเดล VGG 16 ได้ประมาณร้อยละ 70 ยังคงสูงกว่าการใช้เทคนิคเรียนรู้แบบเดิม อย่างไรก็ตาม เมื่อ γ เพิ่มขึ้น 1 step เป็น $\gamma = 3, \alpha = 0.8$ ก็พบว่า ผลกระทบของอัตราการจดจำลดลงอย่างรวดเร็ว จากที่กล่าวมาแสดงให้เห็นว่าเทคนิคที่เสนอให้ประสิทธิภาพการตรวจจับและจดจำเพิ่มขึ้นประมาณ 6% และทำงานได้ดีที่สภาวะ $\gamma < 2$ และ $\alpha > 0.5$ โดย Pretrained CNN โมเดล VGG16 ให้ผลการจดจำที่ดีที่สุดเมื่อเทียบกับ GoogLeNet และ ResNet 50 โดยให้ผลดีที่สุดที่ร้อยละ 86.28



บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงสรุปผลการวิจัยที่ทำมาทั้งหมด รวมไปถึงข้อเสนอแนะแนวทางการพัฒนาระบบการตรวจจับและจดจำท่ามือจับวัตถุในระบบอ็อกซ์เซ็นทริกวิชั่นไปประยุกต์ใช้ในอนาคตต่อไป

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอระบบตรวจจับและจดจำท่ามือจับวัตถุในระบบอ็อกซ์เซ็นทริกวิชั่นในสภาวะความเข้มแสงสว่างต่ำ เนื่องจากมือของมนุษย์มีการเปลี่ยนแปลงรูปร่างและเคลื่อนไหวตลอดเวลาทำให้เป็นสิ่งที่ยากต่อการตรวจจับ เมื่อต้องตรวจจับและจดจำท่ามือด้วยกล้อง FPV ซึ่งกล้องจะติดที่บริเวณศีรษะของผู้สวมใส่ทำให้ภาพค่อนข้างมีการเคลื่อนไหวที่รวดเร็วและการเปลี่ยนแปลงความเข้มแสงสว่างของสภาพแวดล้อมทำให้ประสิทธิภาพการตรวจจับมือที่ปรากฏในภาพลดลง ซึ่งขณะใช้มือจับวัตถุมีการจัดเรียงนิ้วมือในการจัดการวัตถุรูปแบบต่าง ๆ ทำให้การตรวจจับและจดจำท่ามือมีความยากลำบากมากขึ้น ในงานวิจัยนี้จึงได้นำเสนอเทคนิคใหม่ในการเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือในสภาวะที่มีความเข้มแสงสว่างต่ำ โดยนำเสนอการทำงานร่วมกันของการปรับปรุงภาพด้วยเทคนิค AGCWD และความต้านทานแบบโค้งงอ โดยการนำความต้านทานแบบโค้งงอไปติดบริเวณนิ้วโป้ง ซึ่งในแต่ละท่ามือจับมีการจัดวางนิ้วโป้งในแต่ละตำแหน่งที่แตกต่างกัน จึงสามารถนำมาใช้เป็นพารามิเตอร์ในการกำหนดน้ำหนักปรับปรุงภาพของเทคนิค AGCWD ซึ่งแต่ละท่ามือมีน้ำหนักที่แตกต่างกันทำให้แต่ละภาพของท่ามือจับวัตถุมีคุณลักษณะเฉพาะ หลังจากนั้นใช้การตรวจจับและจดจำท่ามือที่ปรากฏในภาพด้วยเทคนิคการเรียนรู้เชิงลึก YOLOv2 ซึ่งมีความสามารถในการทำงานที่รวดเร็วและให้ประสิทธิภาพสูงในการตรวจจับและจดจำวัตถุในงานคอมพิวเตอร์วิทัศน์

ผลการทดลองพบว่า เทคนิคที่นำเสนอสามารถเพิ่มประสิทธิภาพการตรวจจับและจดจำท่ามือจับวัตถุภายใต้แบบจำลองสภาวะความเข้มแสงสว่างในช่วงการปรับค่าแกมมาและความคมชัดแต่ละระดับได้ดี ซึ่งผลการทดลองเมื่อเทียบกับเทคนิคเรียนรู้แบบเดิมที่ใช้เฉพาะภาพในการฝึกเรียนรู้อย่างเดียวพบว่า สามารถเพิ่มประสิทธิภาพการตรวจจับและจดจำได้ประมาณ 6% และเทคนิคที่นำเสนอสามารถทำงานได้ดีที่ค่า $\gamma < 2$ และ $\alpha > 0.5$ ซึ่งการใช้ Pertained CNN ทั้ง 3 โมเดล VGG16, GoogLeNet และ ResNet 50 ให้ผลการทดลองที่ดี โดย VGG16 ให้ผลดีที่สุทธ้อยู่ที่ 86.28

5.2 ข้อเสนอแนะ

เทคนิคที่นำเสนอนี้เป็นงานวิจัยแรกที่ได้นำความต้านทานแบบโค้งมาประยุกต์ใช้สำหรับการตรวจจับและจดจำท่ามือจับวัตถุในระบบอีโก้เซ็นเซอร์ควิซัน ซึ่งได้ผลการทดลองเป็นอย่างดี แต่ยังมีช่วงการทำงานของ γ และ α ไม่กว้างมากนัก ซึ่งต้องมีการปรับปรุงวิธีการที่จะช่วยเพิ่มประสิทธิภาพการตรวจจับและจดจำเพื่อขยายช่วงการทำงานในระดับความเข้มแสงสว่างให้กว้างขึ้น เพื่อสามารถนำไปสู่การพัฒนาระบบไปใช้งานจริงในด้านๆ ได้ต่อไป



บรรณานุกรม

- [1] T. Feix, J. Romero, H. Schmiedmayer, A. M. Dollar, and D. Kragic, "The GRASP Taxonomy of Human Grasp Types," *IEEE Transactions on Human-Machine System*, vol. 46, No.1, pp. 66-77, Feb 2016. doi:10.1109/THMS.2015.2470657
- [2] M. Cai, K.M.Kitani, and Y.Sato, "Understanding hand-object manipulation by modeling the contextual relationship between actions, grasp types and object attributes," in *2016 Robotics: Science and Systems Conference (RSS) XII*, Ann Arbor, Michigan, USA, 2016, pp. 1-10.
- [3] I. M. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, "Grasp Frequency and Usagw in Daily Household and Machine Shop Tasks," *IEEE Transactions on Haptics*, vol. 6, No. 3, pp. 296-308, Jul 2013. doi:10.1109/TOH.2013.6
- [4] J. Z. Zheng, S. De La Rosa, and A. M. Dollar, "An Investigation of Grasp Type and Frequency in daily Household and Machine Shop Tasks," in *2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011 pp. 4169-4175.
- [5] T. Feix, I. M. Bullock, Student Member, and A. M. Dollar, "Analysis of Human Grasping Behavior: Object Characteristics and Grasp Type," *IEEE Transactions on Haptics*, vol. 7, No.3, pp. 311-323, Jul 2014, doi:10.1109/TOH.2014.2326871
- [6] K. Gupta, D. Burschka, and A. Bhavsan, "Classifying Object Manipulation Actions based on Grasp-types and Motion-Constraints," *The Computing Research Repository (CoRR)*, pp 1-10, Jun 2018.
- [7] A. Betancourt, P. Morerio, C.S. Regazzoni, and M. Rauterberg, "The Evolution of First Person Vision Methods: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, No. 5, pp. 744 - 760, Apr 2015. doi:10.1109/TCSVT.2015.2409731

บรรณานุกรม (ต่อ)

- [8] A. Bandini, and J. Zariffa, "Analysis of the hands in egocentric vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, pp.1-19, Apr 2020. doi:10.1109/TPAMI.2020.2986648
- [9] E. Talavera, N. Petkov, and P. Radeva, "Chapter 15 - Egocentric vision for lifestyle Understanding," *Wearable Sensors (Second Edition)*, pp 415-433, 2021.
- [10] G. Bleser, A. Behera, A. G. Cohn, and D. Damen, "Cognitive Workflow Capturing and Rendering with On-Body Sensor Networks (COGNITO)," *Research Report RR-13-02*, pp. 1-39, 2013.
- [11] S.A.A.S.M.Ali, N.S.Ahmad, and P.Goh, "Flex Sensor Compensator via Hammerstein–Wiener Modeling Approach for Improved Dynamic Goniometry and Constrained Control of a Bionic Hand," *Sensors*, vol. 19, pp. 3896, Sep 2019. doi:10.3390/s19183896.
- [12] Interfacing Flex Sensor with Arduino (*online*), 2021, Available: <https://lastminuteengineers.com/flex-sensor-arduino-tutorial> (18 July 2021).
- [13] S. C. Huang, F. C. Cheng, and Y. S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, No. 3, pp. 1032–1041, Mar 2013. doi:10.1109/TIP.2012.2226047
- [14] S. Rahman, MD. M. Rahman, M. Abdullah-Al-Wadud, G. Dastagir Al-Quaderi, and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP Journal on Image and Video Processing*, vol. 35, pp. 1-13, 2016. doi:10.1186/s13640-016-0138-1
- [15] Gamma Correction (*online*), 2013, Available: <https://alwayyours.blogspot.com/2013/02/gamma-correction-22.html> (18 September 2021).

บรรณานุกรม (ต่อ)

- [16] X. Zhand, H. Shen, X. Xu, and Z. Zhang, "A method of continuous nonlinear gamma correction," *Advances in Engineering Research (AER)*, vol.117, pp. 807-815, 2016.
- [17] Deep Learning แบบฉบับคนสามัญชน EP 1: Neural Network History, (online), 2019, Available: <https://medium.com/mmp-li/deep-learning-ep-1-neural-network-history-f7789236a9a3> (16 May 2020).
- [18] จักรินทร์ สนุกแสน, และ โอลาริก สุรินดี, "โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกสำหรับการจำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ," *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, vol. 38, No. 2, pp. 113-124, 2019.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1-10. doi:10.1109/CVPR.2016.91
- [20] H. Nakahara, F. Tomoya, and S. Sato, "A Lightweight YOLOv2: A Binarized CNN with A Parallel Support Vector Regression for an FPGA," in *FPGA 2018*, Monterey, CA, USA, 2018, pp 1-11. doi:10.1145/3174243.3174266
- [21] J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, and B. Cai, "An Improved YOLOv2 for Vehicle Detection," *Sensors*, vol. 18, pp. 4272, Dec 2018. doi:10.3390/s18124272
- [22] S. Seong, J. Song, D. Yoon, J. Kim, and J. Choi, "Determination of Vehicle Trajectory through Optimization of Vehicle Bounding Boxes Using a Convolutional Neural Network," *Sensors*, vol. 19, pp. 4263, Sep 2019. doi:10.3390/s19194263
- [23] Z. Z. Bufan, *Rereading yolov2 for introduction to target detection (online)*, 2018, Available: <https://chowdera.com/2021/12/202112181241102191.html> (4 May 2021).
- [24] S. Mann, "Wearable computing: a first step toward personal imaging," *Computer*, Vol. 30, No. 2, pp. 25 – 32, Feb. 1997. doi:10.1109/2.566147

บรรณานุกรม (ต่อ)

- [25] T. Stamer B. Schiele, and A. Pentland, "Visual Contextual Awareness in Wearable Computing," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, Pittsburgh, PA, USA, 1998, pp. 1-9. doi:10.1109/ISWC.1998.729529
- [26] W. W. Mayol-Cuevas, B.J. Tordoff, and D.W. Murray, "Wearable Visual Robots," in *Digest of Papers. Fourth International Symposium on Wearable Computers*, Atlanta, GA, USA, 2000, pp.95-102. doi:10.1109/ISWC.2000.888470
- [27] M. F. Land, and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?," *Vision Research*, vol. 41, 2001. pp. 3559-3565.
- [28] R. Tenmoku, M. Kanbara, and N. Yokoya, "Annotating User-Viewed Objects for Wearable AR Systems," in *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, Vienna, Austria, 2005, pp 1-2. doi:10.1109/ISMAR.2005.10
- [29] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur and K. Wood, "SenseCam: A Retrospective Memory Aid," *International Conference on Ubiquitous Computing*, Orange County, CA, USA, 2006, pp. 177-193. doi:10.1007/11853565_11
- [30] N. Alves, and T. Chau, "Vision-Based Segmentation of Continuous Mechanomyographic Grasping Sequences," *IEEE Transactions on Biomedical Engineering*, vol. 55, No. 2, pp.765-773, Feb 2008. doi:10.1109/TBME.2007.902223
- [31] Y. Motai, and A. Kosaka, "Hand -Eye Calibration Applied to Viewpoint Selection for Robotic Vision," *IEEE Transactions on Industrial Electronics*, vol. 55, No. 10, pp. 3731-3741, Oct 2008. doi:10.1109/TIE.2008.921255

บรรณานุกรม (ต่อ)

- [32] S. Sundaram, and W. W. M. Cuevas, “High Level Activity Recognition using Low Resolution Wearable Vision,” in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, FL, USA, 2009, pp. 25-32. doi:10.1109/CVPRW.2009.5204355
- [33] A. Fathi, X. Ren, and J. M. Rehg, “Learning to Recognize Objects in Egocentric Activities,” in *2011 IEEE Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 2011, pp. 3281-3288. doi:10.1109/CVPR.2011.5995444
- [34] H. Pirsiavash, and D. Ramanan, “Detecting Activities of Daily Living in First-Person Camera Views, in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, 2012, pp. 2847-2854. doi:10.1109/CVPR.2012.6248010
- [35] G. Serra, M. Camurri, L. Baraldi, M. Benedetti, and R. Cucchiara, “Hand Segmentation for Gesture Recognition in EGO-Vision,” in *2013 ACM Multimedia Conference & co-located workshops (IMMPD13)*, Barcelona, Spain, 2013, pp 31-36. doi:10.1145/2505483.2505490
- [36] M. Cai, K. M. Kitani, and Y. Sato. “A scalable approach for understanding the visual structures of hand grasps,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, 2015, pp. 1-8. doi:10.1109/ICRA.2015.7139367
- [37] S. Singh, C. Arora, and C.V. Jawahar, “First Person Action Recognition Using Deep Learned Descriptors,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2620-2628. doi:10.1109/CVPR.2016.287

บรรณานุกรม (ต่อ)

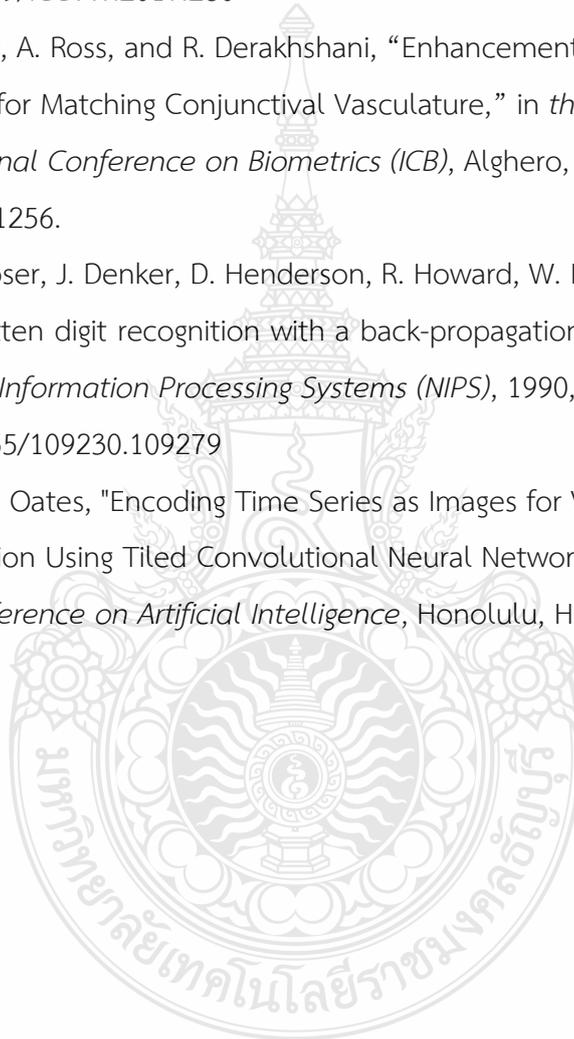
- [38] M. Cai, K. M. Kitani, and Y. Sato, "An Ego-Vision System for Hand Grasp Analysis, *IEEE Transactions on Human-Machine System*," vol. 47, No. 4, pp. 524-535, Aug 2017, doi:10.1109/THMS.2017.2681423
- [39] M. Wang, C. Luo, B. Ni, and J. Yuan, "First-Person Daily Activity Recognition With Manipulated Object Proposals and Non-Linear Feature Fusion," *IEEE Transactions on Circuits and System for Video Technology*, vol. 28, No. 10, pp. 2946-2955, Oct 2018. doi:10.1109/TCSVT.2017.2716819
- [40] G. Cotugno, K. Althoefer, and T. Nanayakkara, "The role of the thumb: Study of finger motion in grasping and reachability space in human and robotics hands," *IEEE Transaction on Systems Man and Cybernetics: Systems*, vol. 47 No. 7, pp. 1061-1070, Jul 2017. doi:10.1109/TSMC.2016.2531679
- [41] A. Sreejan, and Y. S. Narayan, "A review on applications of flex sensors," *International Journal of Engineering Technology and Advanced Engineering*, vol. 7, No. 7, pp. 97-100, Jul 2017.
- [42] พจนานุกรมสารสนเทศภาษาไทย, (online), 2013, Available: http://pirun.ku.ac.th/~fhumalt/THSL/THSL/html/nav_th/THSL_intro_th.htm (16 May 2016).
- [43] J. A. M. Basilio, G. A. Torres, G. S. Perez, L. T. Medina, and H. M. P. Meana, "Explicit Image Detection using YCbCr Space Color Model as Skin Detection," in *the 5th WSEAS International Conference on Computer Engineering and Applications (CEA'11)*, Puerto Morelos, Mexico, 2011, pp. 123-128.
- [44] A. Satpathy, X. Jiang, and H. L. Eng, "Human detection by quadratic classification on subspace of extended histogram of gradients," *IEEE Transactions on Image Processing*, vol. 23, No. 1, pp. 287-297, Jan 2014. doi:10.1109/TIP.2013.2264677

บรรณานุกรม (ต่อ)

- [45] E. Memin, and P. Perez, “Dense Estimation and Object-Based Segmentation of the Optical Flow with Robust Techniques,” *IEEE Transactions on Image Processing*, vol. 7, No. 5, May 1998. pp. 703-719. doi:10.1109/83.668027
- [46] G. Li, Y. Yang, X. Qu, D. Cao and K. Li, “A deep learning based image enhancement approach for autonomous driving at night,” *Knowledge-Based Systems*, vol. 213, pp. 1–14, 2021. doi:10.1016/j.knosys.2020.106617
- [47] S. Shin, H. U. Yoon, and B. Yoo, “Hand Gesture Recognition Using EGaln-Silicone Soft Sensors,” *Sensors*,” vol. 21, pp. 3204, May 2021. doi:10.3390/s21093204
- [48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, No. 6, pp. 84-90, Jun 2017. doi:10.1145/3065386
- [49] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations 2015 (ICLR 2015)*, San Diego, CA, USA, 2015, pp. 1–14.
- [50] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, Las Vegas, NV, USA, 2015, pp. 770–778. doi:10.1109/CVPR.2016.90
- [51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, Boston, MA, USA, 2015, pp. 1-12. doi:10.1109/CVPR.2015.7298594
- [52] R. Padilla, S. L. Netto, and E. A. B. da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Niteroi, Brazil, 2020, pp. 237-242. doi:10.1109/IWSSIP48289.2020.9145130

บรรณานุกรม (ต่อ)

- [53] A. Furnari, S. Battiato, and G.M. Farinella, "How shall we evaluate egocentric action recognition?," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, 2017, pp. 2373–2382.
doi:10.1109/ICCVW.2017.280
- [54] S. Crialmeanu, A. Ross, and R. Derakhshani, "Enhancement and Registration Schemes for Matching Conjunctival Vasculature," in *the 3rd IAPR/IEEE International Conference on Biometrics (ICB)*, Alghero, Italy, 2009, pp. 1247-1256.
- [55] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems (NIPS)*, 1990, pp. 396-404
doi:10.5555/109230.109279
- [56] Z. Wang and T. Oates, "Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks," in *the Workshops at AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 2019, pp. 40–46.



ภาคผนวก



ภาคผนวก ก

วารสารทางวิชาการและบทความการประชุมวิชาการ

- [1] C. Chansri, and J. Srinonchat, "Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video," in *The 2016 International Electrical Engineering Congress (iEECON)*, Chiang Mai, Thailand, March 2 2016-March 4 2016, pp. 257-260. doi:10.1016/j.procs.2016.05.113
- [2] C. Chansri, and J. Srinonchat, "Reliability and Accuracy of Thai Sign Language Recognition with Kinect Sensor," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Mai, Thailand, June 28 2016-July 1 2016, pp. 1-4. doi:10.1109/ECTICon.2016.7561403
- [3] C. Chansri, and J. Srinonchat, "Low Cost Hand Gesture Control in Complex Environment Using Raspberry Pi," in *2019 International SoC Design Conference (ISOCC)*, Jeju, Korea (South), October 6 2019-October 9 2019, pp. 186-187. doi:10.1109/ISOCC47750.2019.9027669
- [4] ชนะ จันทร์ศรี, และ จักริ ศรีนนท์ฉัตร, "การพัฒนาการจดจำท่ามือด้วยถุงมือเซนเซอร์แบบ 16 จุด หมุนอิสระร่วมกับการโครงข่ายประสาทแบบคอนโวลูชัน," *วารสารวิศวกรรมศาสตร์, มหาวิทยาลัยเชียงใหม่*, vol. 28, No. 3, pp. 43-56, Dec 2021.
- [5] C. Chansri, and J. Srinonchat, "Enhance Egocentric Grasp Recognition Based Flex Sensor Under Low Illumination," *CMC-Computers, Materials & Continua*, vol. 71, No. 3, pp.4377-4389, Jan 2022. doi:10.32604/cmc.2022.024026



2016 International Electrical Engineering Congress, iEECON2016, 2-4 March 2016, Chiang Mai, Thailand

Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video

Chana Chansri^a, Jakkree Srinonchat^{b*}

^{a,b}Signal Processing Research Laboratory, Department of Electronics and Telecommunication Engineering, Faculty of Engineering Rajamangala University of Technology Thanyaburi, 39 Moo1, Klong6, Thanyaburi, Patumthani, 12110 Thailand

Abstract

Hand detection and gesture recognition are the active research area in the computer vision. The main purpose to develop the sign language recognition and Human Computer Interaction (HCI). This article investigates and develops the technique to recognize hand posture of Thai sign language in a complex background using fusion of depth and color video. The new technology of sensors, such as the Microsoft Kinect, recently provides the depth video which helps researchers to find the hand position in the scene. This advantage is used to segment the hand sign in the color video without the environment interference such as skin color background. The histograms of oriented gradients are used to extract the image features of hand sign. These features are then pass to the artificial neural network for training and recognition. The result showed that the proposed method is robust to detect the hand gestures in the complex background. It provides the accuracy recognition for the Thai fingerspelling of 84.05%

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of iEECON2016

Keywords: Thai Sign Language; Hand Detection; Depth Camera; Artificial Neural Network.

1. Introduction

The sensors technology are recently developed to operate with the real time system. This is very effective for the researchers and machine design. The hand detection and gesture recognition are also exploited the new technology of sensors to operate which based on the machine vision and human-computer interaction. The human body is non rigid object, it make hand shape varied with different time. This makes the challenges to a lot of research works for

* Corresponding author. Tel.: +66 2549 3588.
E-mail address: jakkree.s@en.rmutt.ac.th.

Reliability and Accuracy of Thai Sign Language Recognition with Kinect Sensor

Chana Chansri

Signal Processing Research Laboratory, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Patumthani, 12110 Thailand
chanachansri@hotmail.com

Jakkree Srinonchat

Signal Processing Research Laboratory, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Patumthani, 12110 Thailand
jakkree.s@en.rmutt.ac.th

Abstract— Hand detection is widely used for recognizing hand gesture which is a primary step for sign language recognition. With larger quantities of sensors, it can provide real-time depth measurements such as Kinect sensor and help researchers find the hand position in the scene without the environment interference such as skin color background. The hand segmentation is based on depth image which is particular correspond to the distance from the signer's hand to Kinect sensor. This article presents a study of the performance for Thai sign language recognition obtained by Kinect sensor in various distance to find the most suitable distance for the accuracy and reliability. After the hand sign is detected, the histograms of oriented gradients will extract the image features of hand sign and proceed to the artificial neural network for recognizing the hand gestures. The result found that the recognition accuracy of Thai finger-spelling of Kinect sensor can work effectively in the distance range of 0.8 - 1.2 meter. The accuracies of recognition for each distance are 83.33% at 0.8m, 81.25% at 1.0m and 72.92% at 1.2m respectively. This distance range can be generated the wide range of brightness in the depth image.

Keywords—kinect sensor; depth image; hand detection; Thai sign language; histograms of oriented gradients.

I. INTRODUCTION

The depth image contains information corresponded to the distance between depth of the camera and scene of the objects. The closer distance, the brighter the image will be. On the other hand, the greater distance will make the image less bright. The depth cameras are developed in the computer vision that can be applicable in sign language recognition and Human Computer Interaction (HCI). The monochrome CMOS camera and infrared structured light projector in Kinect Sensor is a feature for computing depth of the scene which help researchers to find the hand position in the scene and also able to locate the hand in color image due to the hand is always in front of the scene.

The objective of this article is to find the most suitable distance for the best accuracy and reliability to recognize the Thai sign language with Kinect sensor. There are previous researches which propose various techniques such as data glove and motion tracker [1], upright speed-up robust feature and the dynamic time warping [2]. This experiment designed

to operate in front of the depth camera for detecting the hand gesture in the scene. An interesting feature of this work is the total absence of a learning phase in this section without the training data which reduce computational time comparing with other techniques [3-5] such as Viola-Jones detector. The detection is also based on the combination of color and depth information which used for obtaining a specific threshold that defines the regions of interest on the scene. The signer is always in foreground of the image. When the depth information is processed with color information, it will eliminate the background of color image and limbs such as face, arm, elbow etc. This system is durable against complex environment. After the signer's hand is detected, the Histograms of Oriented Gradients (HOG) are used to extract the hand images features that algorithm provides good results with human and hand detection [6-7]. These features will be used as input for neural network to perform the database training and gesture recognition.

II. HAND DETECTION

There are still some difficulties for hand detection because human body is flexible object, thus the body actions make hand shape varied and deformation under different angle of views, varieties with the spaces and timing. The lighting conditions and the complex background are sensitive to skin color and color data. This problem has challenged a lot of research works for achievement of hand detection [8]. The paper presents hand detection method for Thai fingerspelling in the complicated background base on depth and color image.

A. Kinect Sensor

The researcher uses the Microsoft Kinect camera, with internal diffractive optical device and infrared laser diode, for generating an irregular dot pattern. The color and two mega pixel grayscale chip are incorporated with an infrared (IR) filter which determining the disparities between the emitted light dots and their observed position. After that, the depth of an object in the scene is triangulated in the condition that identity of an observed dot on the object must be known. The irregular pattern obtains better performance than uniform pattern. This

Low Cost Hand Gesture Control in Complex Environment Using Raspberry Pi

Chana Chansri and Jakkree Srinonchat

Faculty of Engineering, Rajamangala University of
Technology Thanyaburi, Patumthani, 12110 Thailand
chanachansri@mail.rmutt.ac.th, jakkree.s@en.rmutt.ac.th

Eng Gee Lim and Ka Lok Man

Xi'an Jiaotong-Liverpool University
AI University Research Centre

Abstract— This article focuses on implementation in an embedded system with Raspberry Pi to a standalone machine for controlling electronic devices which wirelessly controlled by a hand gesture in the complex environment background. This system uses the RGB camera in combination with Raspberry Pi, a popular device today due to the inexpensive price and reliable performance. The hand gesture detection in each frame uses the radian fingertip analysis technique, a new technique presented which does not require any data training. This technique provides a good robust for light effect and complex environment. The experiment had been tested with the America Sign Language fingerspelling 12 gestures, the results found that 90.83%.

Keywords; hand gesture; raspberry pi; radian fingertip analysis

I. INTRODUCTION

Daily life of human beings is communicating with many electronic devices such as computers, mobile phones, Game player etc. There is communication through various channels. According to each device such as a mouse, keyboard, voice even acting. One of them is a hand gesture control that has received great attention due to its potential applications in contactless Human Computer Interaction (HCI). There is now one of the most important application fields of computer vision, the main goal to develop practical and efficient ways to interact with the computer without using traditional input devices. The human hand is a complex articulated object consisting of many connected parts and joints, the human hand motion has roughly 27 degrees of freedom [1] involved by the human hand detection is a challenging problem due to the huge intra class variation stemming from color, clothing, pose and appearance variations. This study aims at investigate to develop a real time gesture control system that enables hand signers to interact with electrotonic devices. This work used a Raspberry Pi that is a System on Chip (SoC), a design where a single board carries all the essential circuits like computer main board but the availability of features such as the General Purpose Input Output (GPIO) [2] pins make to programming to control hardware as well as driving electronic circuitry. Moreover, a Raspberry are small size and cheaper than using a depth camera with personal computer when using hand gestures detection.

II. PROPOSED WORK

The hand gesture control system works by using the RGB camera as the master and Raspberry pi as hardware tool. The Raspberry Pi and camera module are used with a software written to detect hand of a person who show hand gestures in front of camera then interpret the detected command sent out to GPIO port. The system diagrams and algorithm of program working were created as Fig. 1

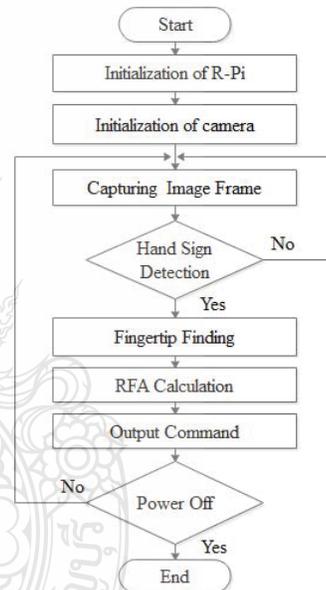


Figure. 1 Graphic diagram of the main algorithm of the system

After starting work the Raspberry Pi 3 model B is turned on and the initialization to work with its protocol and data transmission. The camera sends start readiness messages to Raspberry Pi. Once the module is started, the image acquisition is done using Raspberry Pi Camera V2 over CSI port interface, the program will start capturing the video which image size is 640x480 pixels. This work uses the human skin color detection techniques along with optical flow to find the position and hand sign that appear in each frame. The skin color detection must convert images from the RGB obtained from the camera into the YCbCr. This space is suitable for finding human skin color [3] because it is resistant to changing the light in the surrounding environment. The optical flow, the calculation of light intensity changes between two consecutive



การพัฒนาการจดจำท่ามือด้วยถุงมือเซนเซอร์แบบ 16 จุดหมุน อิสระร่วมกับการโครงข่ายประสาทแบบคอนโวลูชัน Grasp Recognition Development Using 16 Degree of Freedom Sensors and Convolutional Neural Network

ชนะ จันทร์ศรี และ จักกรี ศรีนนท์ฉัตร*

Chana Chansri and Jakkree Srinonchat*

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

39 หมู่ 1 ถนนรังสิต-นครนายก อำเภอธัญบุรี จังหวัดปทุมธานี 12110

Faculty of Engineering, Rajamangala University of Technology Thanyaburi

39 Moo 1, Rangsit-Nakhornnayok Road, Thanyaburi, Pathum Thani, 12110

*E-mail: jakkree.s@en.rmutt.ac.th, Tel.: +668-9777-5038

บทคัดย่อ

ในปัจจุบันการนำระบบเซนเซอร์ต่างๆ ไปใช้เชื่อมต่อข้อมูลระหว่างมือของมนุษย์กับคอมพิวเตอร์สำหรับพัฒนาระบบรู้จำท่ามือยังเป็นสิ่งที่ท้าทาย เนื่องจากระบบเซนเซอร์ยังไม่สามารถตอบสนองความแม่นยำในการจดจำท่ามือที่ดีพอสำหรับระบบการมองเห็นของระบบคอมพิวเตอร์เพื่อนำไปใช้งานควบคุมระยะไกลและการสื่อสาร ดังนั้น งานวิจัยนี้นำเสนอการพัฒนาการจดจำท่ามือด้วยถุงมือเซนเซอร์แบบ 16 จุดหมุนอิสระร่วมกับการโครงข่ายประสาทแบบคอนโวลูชัน เซนเซอร์วัดค่าการโค้งงอของวัตถุถูกนำไปติดตั้งตามจุดหมุนของมือมนุษย์บนถุงมือจำนวน 16 จุดหมุนอิสระ เพื่อให้สามารถตรวจวัดการงอของข้อนิ้วมือแต่ละจุดในขณะที่ใช้มือจับวัตถุ วงจรสำหรับอ่านค่าเซนเซอร์แบบ 16 จุดหมุนอิสระและวงจรรับเฟิร์มแวร์แบบปรับค่าถูกพัฒนาขึ้นใช้ในงานวิจัยนี้ เพื่อทำงานร่วมกับไมโครคอนโทรลเลอร์ Arduino Nano สำหรับบันทึกผลข้อมูลสัญญาณของเซนเซอร์แต่ละจุดเข้าสู่คอมพิวเตอร์ จากนั้นสัญญาณเหล่านี้จะถูกแปลงเป็นข้อมูลภาพด้วยวิธี Nearest Interpolation และ Color Mapping หลังจากนั้นใช้เทคนิคโครงข่ายประสาทแบบคอนโวลูชันเพื่อจดจำท่ามือจับ การจับวัตถุจริงด้วยท่ามือจับที่ใช้ในงานในชีวิตประจำวันด้วย 20 ท่ามือ จำนวน 4,000 ครั้ง ถูกนำมาใช้ในการทดสอบถุงมือเซนเซอร์ที่ออกแบบ ซึ่งผลการทดลองพบว่าระบบที่นำเสนอสามารถทำงานได้เป็นอย่างดีและมีประสิทธิภาพในการจดจำ 99.70%

คำสำคัญ: ถุงมือเซนเซอร์; การจดจำท่ามือจับ; ความต้านทานแบบโค้งงอ; โครงข่ายประสาทแบบคอนโวลูชัน

ABSTRACT

It is recently challenging to use various sensors to interface data in Human-Computer Interaction (HCI) for developing hand gesture recognition systems. However, the previous sensor system has not met the hand gesture recognition accuracy for the machine vision. Therefore, this article presents the

Enhance Egocentric Grasp Recognition Based Flex Sensor Under Low Illumination

Chana Chansri and Jakkree Srinonchat*

Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand

*Corresponding Author: Jakkree Srinonchat. Email: jakkree.s@en.rmutt.ac.th

Received: 30 September 2021; Accepted: 01 November 2021

Abstract: Egocentric recognition is exciting computer vision research by acquiring images and video from the first-person overview. However, an image becomes noisy and dark under low illumination conditions, making subsequent hand detection tasks difficult. Thus, image enhancement is necessary to make buried detail more visible. This article addresses the challenge of egocentric hand grasp recognition in low light conditions by utilizing the flex sensor and image enhancement algorithm based on adaptive gamma correction with weighting distribution. Initially, a flex sensor is installed to the thumb for object manipulation. The thumb placement that holds in a different position on the object of each grasp affects the voltage changing of the flex sensor circuit. The average voltages are used to configure the weighting parameter to improve images in the image enhancement stage. Moreover, the contrast and gamma function are used to adjust varies the low light condition. These grasp images are then separated to be training and testing with pre-trained deep neural networks as the feature extractor in YOLOv2 detection network for the grasp recognition system. The proposed of using a flex sensor significantly improves the grasp recognition rate in low light conditions.

Keywords: Egocentric vision; hand grasp; flex sensor; low light enhancement

1 Introduction

Hands are the priority for humans that allow us to collaborate with the matters and the surroundings, correspond with others and carry out daily activities like dining, cleaning and dressing. Focused on their significance, the computer vision researchers have attempted to analyze hands from various aspects: determine the position of the hand in the image [1], analyze the hands from multiple perspectives: localizing them in the images are investigated in any types of actions [2–4], as well as interact with the computer and the robot [5–7]. Wearable cameras allow hands to be examined from a first-person perspective, known as egocentric or First-Person Vision (FPV) in computer vision [8–12], to challenge object detection and identifying activities. The essential characteristic of egocentric vision is providing a first-person perspective of the scene by laying a forward-facing wearable camera on the chest or head. This wearable camera offers a person-centric view and is optimally set to capture



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ประวัติผู้เขียน

ชื่อ-สกุล	นายชนะ จันทร์ศรี
วัน เดือน ปีเกิด	วันที่ 21 เดือน พฤษภาคม พ.ศ. 2522
ที่อยู่	78/667 หมู่บ้านชื่อตรงบางใหญ่ เฟส 3 หมู่ที่ 9 ตำบลละหาร อำเภอบางบัวทอง จังหวัดนนทบุรี 11110
การศึกษา	- ปริญญาตรี วศ.บ.(วิศวกรรมไฟฟ้า - โทecomนาคม) คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี - ปริญญาตรี ศ.บ.(เศรษฐศาสตร์) คณะเศรษฐศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช - ปริญญาโท วศ.ม.(วิศวกรรมไฟฟ้า) คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ประสบการณ์ทำงาน	วิศวกร บริษัท ไทยยูนิค จำกัด ตั้งแต่ พ.ศ. 2545 จนถึงปัจจุบัน
เบอร์โทรศัพท์	08-6688-4330
อีเมล	chanachansri@mail.rmUTT.ac.th

