

การพัฒนาระบบการมองเห็นแบบสเตอริโอโดยใช้ ESP32-CAM สำหรับ วัดระยะของวัตถุและสิ่งงานแขนหุ่นยนต์ขนาดเล็ก

สมชาย เตียเจริญ¹ และ สมพร เตียเจริญ^{2*}

¹วิศวกร ภาควิชาเทคโนโลยีวิศวกรรมอิเล็กทรอนิกส์ วิทยาลัยเทคโนโลยีอุตสาหกรรม

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

²อาจารย์ ภาควิชาเทคโนโลยีวิศวกรรมอิเล็กทรอนิกส์ วิทยาลัยเทคโนโลยีอุตสาหกรรม

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

*somporn.t@cit.kmutnb.ac.th

บทคัดย่อ

งานวิจัยนี้นำเสนอการพัฒนาระบบการมองเห็นแบบสเตอริโอโดยใช้ ESP32-CAM สำหรับวัดระยะของวัตถุและสิ่งงานแขนหุ่นยนต์ขนาดเล็กให้ทำการจับหรือจัดการวัตถุอัตโนมัติ ระบบประกอบด้วยกล้องคู่ที่ติดตั้งบนโครงสร้างที่สามารถปรับระยะฐาน (Baseline) ได้ โดยใช้หลักการคำนวณความแตกต่างของภาพ (Disparity Map) จากภาพซ้ายและขวา เพื่อประมาณระยะทางแบบเรียลไทม์ด้วยเอไอผ่านไลบรารีโอเพ่นซีวี นอกจากนี้ ระบบยังรวมการตรวจจับวัตถุด้วยโมเดล Deep Learning ที่ใช้เทคนิค Mask R-CNN-LSTM เพื่อระบุตำแหน่งและขอบเขตของวัตถุเป้าหมายก่อนนำไปคำนวณระยะทาง โดยใช้สมการเชิงเส้นจากการถดถอย (Linear Regression) เพื่อลดค่าคลาดเคลื่อนของการวัดระยะ ผลการทดลองแสดงให้เห็นว่าโมเดลเอไอมีค่าความแม่นยำในการตรวจจับวัตถุมากกว่าร้อยละ 80 สำหรับวัตถุขนาดเล็กและมากกว่าร้อยละ 90 สำหรับวัตถุขนาดกลาง ระบบสามารถวัดระยะทางของวัตถุในช่วง 20–70 เซนติเมตร ด้วยค่าคลาดเคลื่อนสัมพัทธ์เฉลี่ยต่ำกว่าร้อยละ 5 สำหรับวัตถุขนาดเล็กและต่ำกว่าร้อยละ 3 สำหรับวัตถุขนาดกลาง ข้อมูลระยะทางและตำแหน่งที่ได้จากกล้องสเตอริโอถูกแปลงฟีดไปยังระบบฟีดของแขนหุ่นยนต์ขนาดเล็กที่พัฒนาขึ้นเองจากเครื่องพิมพ์สามมิติ ที่ควบคุมสแต็ปเปอร์มอเตอร์ด้วยอาดูอินูโน มีกระบวนการ Calibration และการใช้เมทริกซ์การหมุน (Rotation Matrix) เพื่อสั่งงานการจับวัตถุอย่างแม่นยำ ระบบที่เสนอมีประโยชน์ในการประยุกต์ใช้กับหุ่นยนต์อุตสาหกรรมขนาดเล็ก หุ่นยนต์บริการ และระบบอัตโนมัติในงานประกอบหรือจัดเรียงชิ้นส่วนที่ต้องการความแม่นยำ

คำสำคัญ: อีเอสพี32 แขนหุ่นยนต์ โอเพ่นซีวี มาร์กอาร์-ซีเอ็นเอ็น-แอลเอสทีเอ็ม สเตอริโอ

Development of a Stereo Vision System Using ESP32-CAM for Object Distance Measurement and Small Robot Arm Control

Somchai Tiacharoen¹ and Somporn Tiacharoen^{2*}

¹Engineer, Electronics Engineering Technology, College of Industrial Technology,
King Mongkut's University of Technology North Bangkok

²Lecturer, Electronics Engineering Technology, College of Industrial Technology,
King Mongkut's University of Technology North Bangkok

*somporn.t@cit.kmutnb.ac.th

Abstract

This research presents the development of a stereo vision system using ESP32-CAM for object distance measurement and small robot arm control to automatically grasp or manipulate objects. The system consists of a dual camera mounted on a baseline-adjustable structure. It uses disparity mapping from the left and right images to estimate the distance in real-time using AI via an OpenCV library. Furthermore, the system incorporates object detection using a deep learning model with Mask R-CNN-LSTM techniques to identify the position and boundaries of the target object before calculating the distance. Linear regression is used to minimize measurement errors. Experimental results show that the AI model has an object detection accuracy of over 80% for small objects and over 90% for medium-sized objects. The system can measure the distance of objects in the range of 20–70 centimeters with a relative error of less than 5% for small objects and less than 3% for medium-sized objects. The distance and position data obtained from the stereo cameras are coordinate-transformed into the coordinate system of a custom-developed small robotic arm using a 3D printer. The Arduino Uno-controlled stepper motor system incorporates calibration and a rotation matrix for precise object gripping. This proposed system is useful for applications in small industrial robots, service robots, and automation in assembly or parts sorting requiring high precision.

Keywords: ESP32, Robot Arm, Open CV, Mask R-CNN-LSTM, Stereo

1. บทนำ

ระบบการมองเห็นด้วยคอมพิวเตอร์ (Computer Vision) เป็นส่วนสำคัญสำหรับหุ่นยนต์และระบบอัตโนมัติทำให้หุ่นยนต์ทำงานร่วมกับมนุษย์ได้อย่างมีประสิทธิภาพ [1-5] ระบบกล้องสเตอริโอวิชัน (Stereo Vision) เป็นการนำกล้องสองตัวถ่ายภาพวัตถุจากมุมมองที่แตกต่างกันคล้ายการทำงานของตามนุษย์ที่มีสองข้าง จากนั้นข้อมูลรูปภาพทั้งสองจะถูกนำมาวิเคราะห์หาตำแหน่งพิกเซล (Disparity) และคำนวณเป็นระยะของวัตถุ ระบบสเตอริโอวิชันมีราคาถูกเมื่อเทียบกับเซนเซอร์วัดระยะทางแบบไลดาร์ (LiDAR) และได้ข้อมูลที่มากกว่าด้วย เช่น สีและพื้นผิวของวัตถุ ด้วยข้อจำกัดของการทำงานใน

สภาพแวดล้อมจริง เมื่อพื้นผิววัตถุเรียบและมีเงาสะท้อน อัลกอริทึมแบบเดิมอาจจะไม่เหมาะสมเช่น การจับคู่บล็อก (Block Matching) หรือการจับคู่แบบกึ่งโกลบอล (Semi-Global Matching: SGM) จากความก้าวหน้าของเทคนิคการเรียนรู้เชิงลึก (Deep Learning) โดยเฉพาะการจดจำโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) ร่วมกับระบบเชื่อมโยงทางเวลา เข้ากับความจำระยะยาว (LSTM) จะทำให้การกระพริบของการตรวจจับ (Flickering) ลดลงอย่างมาก[6-9] ทั้งนี้เมื่อนำการถดถอยเชิงเส้น (Linear Regression) รวมเข้ามาในระบบจะช่วยลดค่าผิดพลาดได้ดีขึ้น[10]

2. วัตถุประสงค์การวิจัย

บทความนี้นำเสนอการพัฒนาระบบกล้องสเตอริโอวิชันที่ใช้เทคนิคการเรียนรู้ของเครื่องจักรด้วย Mask R-CNN-LSTM สำหรับการตรวจจับและใช้การถดถอยเชิงเส้นช่วยวัดระยะทาง โดยมีการส่งพิกัดตำแหน่งไปยังระบบแขนหุ่นยนต์ขนาดเล็ก ผ่านการสื่อสารแบบอนุกรมและมีการประเมินประสิทธิภาพของระบบ

3. ขอบเขตของการวิจัย

การพัฒนาต้นแบบโดยใช้ ESP 32 เป็นกล้องจำนวน 2 ตัว ความละเอียด 640×480 พิกเซล ระบายฐานไม่เกิน 12 เซนติเมตร ทดสอบในช่วงระยะการเอื้อมของแขนหุ่นยนต์ขนาดเล็กแบบ 4 แกน ส่งการเคลื่อนที่ผ่านการสื่อสารผ่านพอร์ตอนุกรมที่ความเร็ว 115,200 บิตต่อวินาที

4. กรอบแนวคิด

ระบบกล้องสเตอริโอวิชันที่ใช้ ESP 32 ส่งข้อมูลผ่านเครือข่ายไร้สาย สตรีมวิดีโอเข้าสู่ส่วนประมวลผล ซึ่งจะทำการจับรูปภาพ เทคนิคการเรียนรู้ของเครื่องจักรด้วย Mask R-CNN-LSTM ประมวลผลการถดถอยเชิงเส้น คำนวณระยะทาง และส่งข้อมูลพิกัดตำแหน่งผ่านพอร์ตอนุกรมไปยัง ดังแสดงในภาพที่ 1

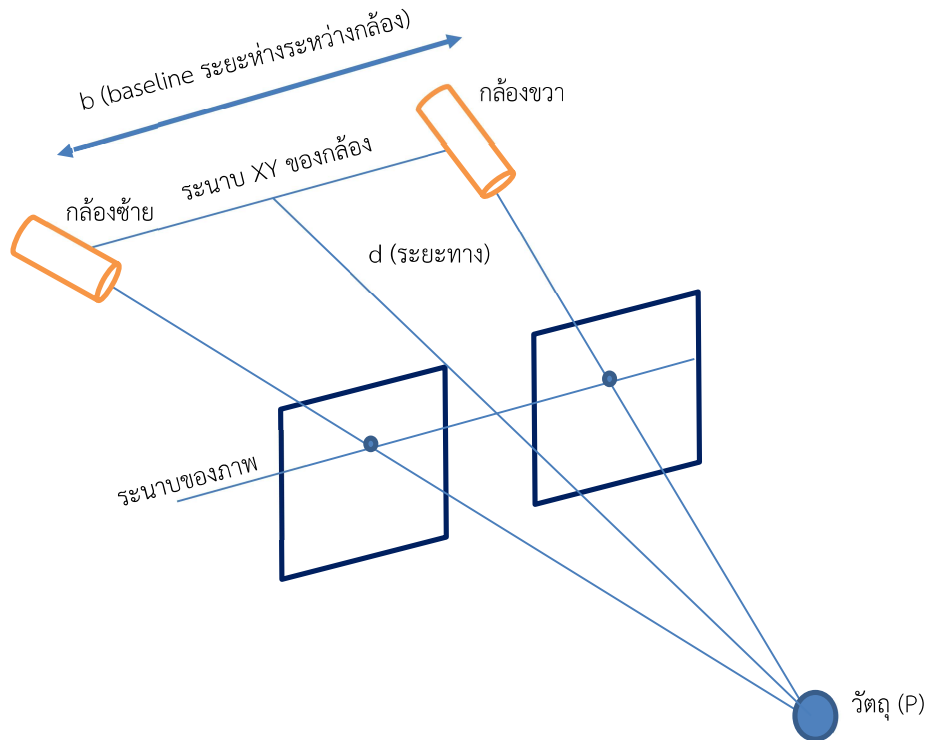


ภาพที่ 1 กรอบแนวคิด

5. หลักการและข้อมูลทั่วไป

งานวิจัยด้านสเตอริโอวิชัน (Stereo Vision) เน้นไปที่การประมวลผลภาพเพื่อเพิ่มความแม่นยำในการระบุตำแหน่ง อัลกอริทึมดั้งเดิมทำงานซ้ำในการหาค่าความลึก (Disparity Map) ระบบกล้องสเตอริโอเป็นเทคนิคการมองเห็นด้วยคอมพิวเตอร์ที่ใช้กล้องสองตัว (คล้ายกับตามนุษย์) เพื่อประมาณระยะทางของวัตถุในภาพ โดยอาศัยหลักการไตรยางศ์ (Triangulation) ในการคำนวณระยะทาง ภาพรวมการมองเห็นแบบสเตอริโอคือจับคู่ฉากจากภาพสองภาพแล้วค้นหาพิกเซลที่ตรงกันในภาพทั้งสองภาพ จากนั้นแปลงตำแหน่ง 2 มิติของการจับคู่ให้เป็นความลึก โดยใช้ฐานเส้น (สำหรับมนุษย์คือระยะห่างระหว่างดวงตา 60 มม.) เริ่มจากกล้องถ่ายภาพเดียวกันในสองตำแหน่งที่แตกต่างกัน มุมจากกล้องซ้ายและมุมจาก

กล้องขวา ส่วนคำคัญคือหาความลึกของแต่ละพิกเซล จากกล้องสองตัวภาพที่ได้จะแตกต่างกันจากการเลื่อนของพิกเซลเป็นผลจากตำแหน่งใหม่ที่เกิดขึ้นต่างกันในแต่ละกล้อง จากวัตถุเดียวกันที่จับรูปภาพได้ภาพทั้งสองจะมีพิกเซลที่สอดคล้องกัน ดังนั้นต้องจับคู่พิกเซลที่ตรงกันระหว่างภาพสองภาพนี้ พิจารณารูปสามเหลี่ยมที่เกิดจากฐาน(baseline)กับเส้นตรงสองเส้นที่ลากจากวัตถุในรูปไปยังจุดศูนย์กลางของกล้องทั้งสองตัว บนระนาบพิกัด XY ตามแนวของ baseline ระยะ x คือผลต่างโดยตัวเลขนี้ทำให้จุด 2 จุดอยู่บนแนวเดียวกันในทั้งสองภาพ (ซึ่งต้องปรับเทียบความสัมพันธ์ก่อน) กำหนดให้มีกล้องสองตัวที่มีแกนวางขนานกัน กล้องแต่ละจับภาพวัตถุตามแนวแกน x กล้องวางห่างกันด้วยค่า baseline ดังแสดงในภาพที่ 2



ภาพที่ 2 ระบบกล้องสเตอริโอ

พิจารณาเฉพาะระนาบที่เกิดจากแกน Y ดังแสดงในภาพที่ 3

หากกล้องสองตัวถ่ายภาพวัตถุเดียวกันจากตำแหน่งที่ห่างกันระยะฐาน (Baseline) b เมตร ระยะทาง d จากกล้องถึงวัตถุสามารถคำนวณได้จาก

$$\frac{d}{f} = \frac{x - b}{xR} \quad (1)$$

$$\frac{x}{xL} = \frac{x - b}{xR} \quad (2)$$

$$b = (xL - xR) \frac{x}{xL} \quad (3)$$

$$d = f \frac{x}{xL} = \frac{f \cdot b}{xL - xR} = \frac{f \cdot b}{disparity} \quad (4)$$

จากสามเหลี่ยมคล้าย เมื่อกำหนดให้

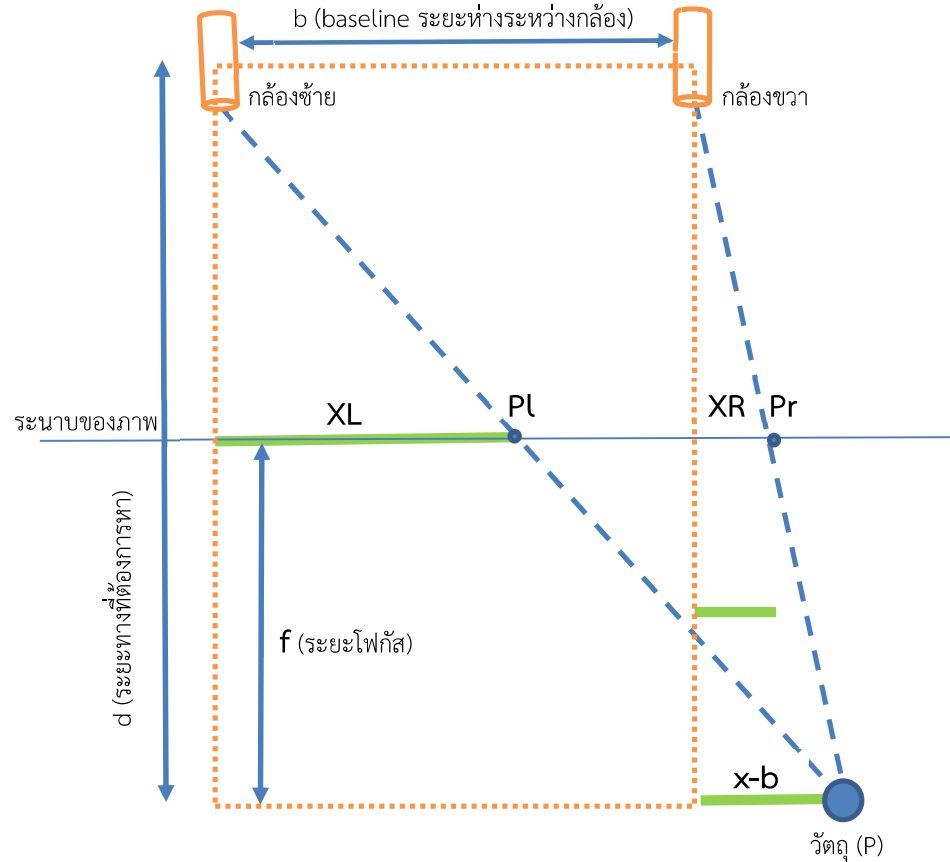
d คือ ระยะทางจากกล้องถึงวัตถุ

f คือ ระยะโฟกัส (focal length) ของเลนส์เป็นหน่วยพิกเซล

b คือ ระยะห่างระหว่างกล้องสองตัว (baseline)

x_L และ x_R คือตำแหน่งแนวนอนของวัตถุในภาพซ้ายและขวาตามลำดับ

disparity คือ ความแตกต่างของตำแหน่งพิกเซลของวัตถุในภาพซ้าย-ขวา (ความแตกต่างพิกเซล)



ภาพที่ 3 ระบบกล้องสเตอริโอพิจารณาในระนาบแกน Y

จากสมการระยะทาง d มีความสัมพันธ์กับค่า disparity นั่นคือยิ่งวัตถุอยู่ใกล้กล้องค่า disparity ยิ่งมากและในทางกลับกันยิ่งวัตถุอยู่ไกล ค่า disparity ยิ่งน้อย ความละเอียดในการวัดระยะทาง (Depth Resolution) สามารถคำนวณได้จากสมการ

$$\Delta d = \frac{d^2}{f \cdot b} \cdot \Delta \text{disparity} \quad (5)$$

โดยที่ระยะฐาน b ที่สามารถปรับได้ซึ่งส่งผลต่อช่วงการวัดและความละเอียดโดยตรง

$\Delta \text{disparity}$ ความละเอียดในการจับคู่พิกเซล (Disparity Resolution) ปกติมีค่าเท่ากับ 1 พิกเซล

เทคนิคการจับคู่แบบคลาสสิกการหาว่ากลุ่มพิกเซลสองกลุ่มเหมือนกันหรือไม่

1. Local Methods (วิธีหาในระดับพื้นที่)

ใช้เทคนิคของฟังก์ชันความคล้ายคลึง (Similarity/Cost Functions) ที่ใช้เปรียบเทียบพื้นที่พิกเซลในภาพระหว่างภาพซ้ายและขวาใน Stereo Vision เพื่อหาจุดที่ตรงกัน (Correspondence)

SAD (Sum of Absolute Differences) เป็นเปรียบเทียบผลต่างของค่าความสว่าง (IL) ในแต่ละพื้นที่ของแต่ละภาพ โดยทำการเลื่อนพื้นที่ของแต่ละภาพ ไปทางตามจำนวนพิกเซลที่กำหนด แล้วหาค่าแล้วหาผลรวมของค่าความสว่างในแต่ละพื้นที่ของภาพซ้ายและขวามาเปรียบเทียบกัน

$$SAD(x, y, d) = \sum_{i=-n}^n \sum_{j=-m}^m |I_L(x+i, y+j) - I_R(x+i-d, y+j)| \quad (6)$$

SSD (Sum of Squared Differences) เป็นเทคนิคเหมือนกับ SAD แต่ยกกำลังสองผลต่างก่อนรวมค่า

$$SSD = \sum (A_i - B_i)^2 \quad (7)$$

$$SSD(x, y, d) = \sum_{i=-n}^n \sum_{j=-m}^m [I_L(x+i, y+j) - I_R(x+i-d, y+j)]^2 \quad (8)$$

NCC (Normalized Cross-Correlation) เป็นการหาค่าเฉลี่ย (Expectation) เทียบกับส่วนเบี่ยงเบนมาตรฐานของค่าพิกเซลในพื้นที่ที่กำหนด

$$NCC(x, y, d) = \frac{\sum_{i,j} (I_L(x+i, y+j) - \bar{I}_L)(I_R(x+i-d, y+j) - \bar{I}_R)}{\sigma_L \cdot \sigma_R} \quad (9)$$

เมื่อ

$\bar{I}_L, \bar{I}_R$ คือค่าเฉลี่ยความเข้มแสงในพื้นที่

σ_L, σ_R คือส่วนเบี่ยงเบนมาตรฐานของพื้นที่

การเปรียบเทียบแต่ละเทคนิค แสดงในตารางที่ 1

2. Global Methods (วิธีหาในระดับภาพรวม)

วิธีการคือแปลงเป็นปัญหาการหาค่าที่เหมาะสมที่สุด (optimizing) ภาพ โดยสร้างฟังก์ชันพลังงาน (Energy Function) และทำการปรับให้เหลือน้อยที่สุด (Energy Minimization) วิธีนี้ให้ความแม่นยำสูง ขอบวัตถุชัดเจน จัดการกับจุดที่มองไม่เห็น (Occlusion) ได้ดี แต่ใช้เวลาประมวลผลนานมาก เช่น เทคนิค Dynamic Programming (DP) Graph Cuts และ Belief Propagation

โมเดลดีพ์เลิร์นนิงสำหรับสเตอริโอแมตชิ่ง (Stereo Matching) โดยมีจุดประสงค์เพื่อประเมินความลึก (Depth) โดยนำโครงข่ายประสาทเทียม (CNN) มาทดแทนการทำงานบางส่วนของกระบวนการทำสเตอริโอแบบดั้งเดิม (Traditional Pipeline) เช่น ใช้ CNN คำนวณหาต้นทุนในการจับคู่ (Matching Cost) แล้วใช้ขั้นตอนวิธีแบบเดิมในการประมวลผลขั้นต่อไป MC-CNN (Matching Cost Convolutional Neural Network): ผลงานบุกเบิก ที่ใช้เครือข่ายแบบ Siamese ในการเปรียบเทียบความคล้ายคลึงของแพตช์ภาพ แทนการใช้สูตรคณิตศาสตร์แบบดั้งเดิม ข้อดีคือมีความแม่นยำสูง แต่ไม่สามารถปรับค่าพารามิเตอร์ทั้งระบบพร้อมกันได้ และใช้เวลาในการคำนวณนาน

ตารางที่ 1 การเปรียบเทียบ matching แต่ละเทคนิค

รายละเอียด	วิธีการ		
	SAD	SSD	NCC
คณิตศาสตร์	L ₁ Norm (Manhattan Distance)	L ₂ ² Norm (Squared Euclidean Distance)	Cosine Similarity (มุมระหว่างเวกเตอร์)
ความหมาย	ค่า SAD ต่ำ = บล็อกคล้ายกันมาก	ค่า SSD ต่ำ = บล็อกคล้ายกันมาก	$-1 \leq NCC \leq 1$ NCC = 1 เหมือนกันพอดี NCC = -1 ตรงกันข้าม
ความเร็ว	เร็วมาก (ใช้การลบและ absolute value)	ปานกลาง (ต้องคูณเพิ่ม ขึ้น)	ช้า (ต้องคำนวณค่าเฉลี่ย, variance, square root)
ความทนทานต่อ noise	ปานกลาง (outlier มีผลแบบ linear)	ต่ำ (outlier มีผลแบบ quadratic)	สูงมาก (หักค่าเฉลี่ย + normalize)
ข้อเสีย	ไม่ทนต่อการเปลี่ยนแปลง ความสว่าง (Illumination Changes)	ไม่ทนต่อ noise รุนแรง; ไม่ทนต่อ illumination changes	ช้ากว่า SAD/SSD; ไม่ทนต่อ geometric distortion
ข้อดี	ให้ความสำคัญกับความต่าง	ให้ความสำคัญกับความต่างเล็กน้อยน้อยกว่า SAD	ทนต่อการเปลี่ยนแปลง ความสว่างและคอนทราสต์

Mask R-CNN (Mask Region-Based Convolutional Neural Network) คือ โมเดลปัญญาประดิษฐ์ประเภท Deep Learning การทำงานของโมเดลมีสองขั้นตอนคือ การตรวจจับวัตถุในภาพ (Object Detection) และการแบ่งส่วนพิกเซล (Pixel-level Segmentation) โดยมาสก์ (Mask) ขอบของวัตถุแยกแต่ละอันออกจากกันอย่างชัดเจน ระบุวัตถุ บอกตำแหน่ง และติกรอบเข้ารูปวัตถุนั้นๆ แบบละเอียดระดับพิกเซล ทำให้ตรวจจับวัตถุ บอกตำแหน่ง และติกรอบวัตถุที่สนใจได้

R-CNN (Region-based Convolutional Neural Network) แทนการพิจารณาภาพทั้งหมดจะใช้อัลกอริทึม Image Segmentation (Selective Search) สร้างพื้นที่เป้าหมายย่อยหลายๆพื้นที่ จากนั้นจะติดป้ายกำกับพิกเซลว่าเป็นอะไรโดยใช้ CNN ในภายหลัง ขั้นแรกสร้างชุดกรอบขอบเขตที่สนใจขนาดคงที่ โดยใช้อัลกอริทึม Selective Search (ประมาณ 2000 ขอบเขต) ซึ่งอาจมีวัตถุอยู่ในพื้นที่นั้น จากนั้นการหาคุณลักษณะเด่น (Feature Extraction) ด้วย CNN ที่ได้รับการฝึกฝนล่วงหน้า จากนั้นจำแนกประเภท (Classification) ด้วย SVM เพื่อหาคลาสของวัตถุภายในกรอบขอบเขตที่สนใจ

Fast R-CNN ปรับปรุงวิธีการประมวลผลพื้นที่ที่สนใจ (Region Proposal หรือ RoI) โดยแทนที่จะตัดแต่ละพื้นที่และรัน CNN แต่ละพื้นที่แยกกัน จะทำการรัน Deep CNN (ResNet) กับภาพทั้งหมด ดึงคุณลักษณะเด่นของภาพทั้งหมด แล้วถึงจะใช้คุณลักษณะนี้ร่วมกัน (RoI pooling) คุณลักษณะเด่นจะถูกส่งต่อไปยัง fully connected layers เป็นการให้โครงข่ายประสาทเทียมแทนการค้นหาแบบเลือกสรร (Selective Search)

Mask R-CNN เป็นส่วนขยายของ Faster R-CNN ที่ไม่เพียงแต่ตรวจจับวัตถุในภาพเท่านั้น แต่ยังให้การแบ่งส่วนระดับพิกเซลสำหรับแต่ละวัตถุอีกด้วย ดังแสดงในตารางที่ 2 เปรียบเทียบการเรียนรู้เชิงลึก

ตารางที่ 2 เปรียบเทียบการเรียนรู้เชิงลึก

RCNN	FasterRCNN	MaskRCNN
1. อินพุตภาพ 2. แบ่งภาพออกเป็นส่วนย่อยๆ 2. จัดกลุ่มสี่เหลี่ยมที่มีลักษณะคล้ายกัน 3. รวมกลุ่มเข้าด้วยกันเป็นพื้นที่ใหญ่ 4. ส่งออกกรอบขอบเขตของพื้นที่ที่สนใจประมาณ 2000 รูป	1. อินพุตภาพ 2. หาคุณลักษณะเด่น CNN 3. คุณลักษณะเด่นและขอบเขต 4. เลเยอร์รวม ROI 5. เลเยอร์เชื่อมต่อแบบเต็ม 6. Softmax และหากรอบขอบเขต	1. อินพุตภาพ 2. หาคุณลักษณะเด่น CNN (ResNet) 3. เครือข่ายการเสนอภูมิภาค (RPN) 4. การจัดตำแหน่ง RoI 5. ขั้นตอนการทำงานขนานกัน - Softmax เพื่อหาคาส - Regression หากรอบขอบเขต - Mask การทำนายใหม่

การเพิ่มระบบความจำระยะยาว (LSTM) เข้าไปในระบบตรวจจับวัตถุมีวัตถุประสงค์เพื่อเพิ่มความแม่นยำของการตรวจจับผ่านเวลา LSTM สามารถจดจำวัตถุจากเฟรมก่อนหน้าได้ ทำให้ลดปัญหาการตรวจจับผิดพลาดที่เกิดจากการบดบัง (Occlusion) หรือการเปลี่ยนแปลงของแสงชั่วขณะ (Transient Errors)

การถดถอยเชิงเส้น (Linear Regression) เป็นเทคนิคสถิติพื้นฐานที่ใช้หาความสัมพันธ์เชิงเส้นระหว่างตัวแปรต้น (Independent Variable) และตัวแปรตาม (Dependent Variable) สามารถใช้สมการเชิงเส้นเพื่อปรับแก้ระยะทางที่คำนวณได้จากสมการทฤษฎีให้ใกล้เคียงกับระยะทางจริงมากขึ้น

แขนหุ่นยนต์ขนาดเล็กเป็นแบบกลไกที่สามารถสร้างได้เอง ความยาวของข้อต่อแขนหุ่นยนต์สามารถดูได้จากตารางที่ 6 ด้านล่าง และโครงสร้างแขนหุ่นยนต์ แสดงดังภาพที่ 2 โดยแกนแขนแรกยาว 140 มิลลิเมตร แกนแขนที่สองยาว 220 มิลลิเมตร ความยาวข้อมือ 100 มิลลิเมตร แสดงดังภาพที่ 4 โครงสร้างแขนหุ่นยนต์



ภาพที่ 4 โครงสร้างแขนหุ่นยนต์

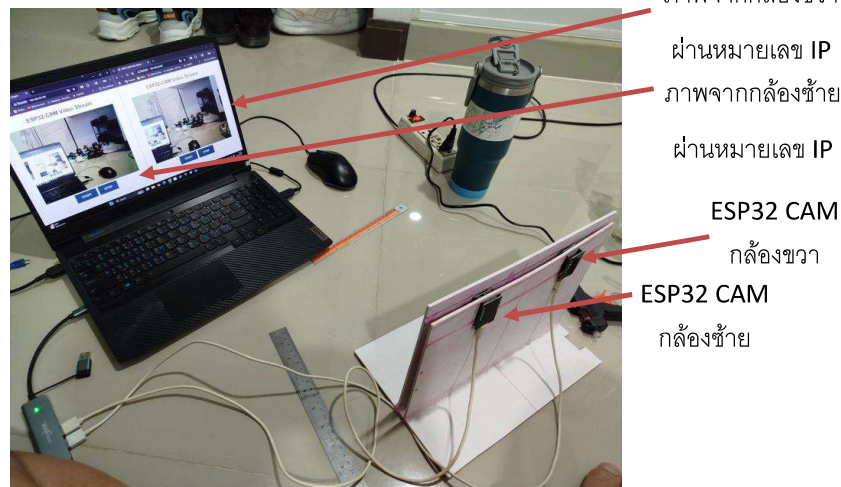
แขนหุ่นยนต์ที่สร้างใช้สเต็ปเปอร์มอเตอร์ วงจรไดรเวอร์อิเล็กทรอนิกส์ A4988 และใช้สเต็ปเปอร์มอเตอร์เป็น NEMA17 การหมุนขั้นละ $1.8^{\circ} \pm 0.09^{\circ}$ รับน้ำหนักได้ 700 มิลลินิวตัน/เมตร สำหรับส่วนข้อมือใช้ MG90 ซึ่งเป็นเซอร์โวมอเตอร์ขนาดเล็ก อาตุอิโนเป็นส่วนใหญ่ควบคุมหลักสั่งให้มอเตอร์หมุนไปที่ละขั้นตามทิศทางที่ต้องการ โดยโปรแกรมควบคุมหลักจะรับค่าตำแหน่งจากพอร์ตอนุกรม

6. งานวิจัยที่เกี่ยวข้อง

กิตติศักดิ์, ส. (2564) วิเคราะห์เทคโนโลยีการมองเห็นภาพแบบสเตอริโอ (Stereo Vision) ในระบบอัตโนมัติ ทำให้หุ่นยนต์คำนวณระยะทาง และสร้างผังความลึก (Depth Map) ใช้ในการจัดเรียง ตรวจสอบ และการนำทางหุ่นยนต์ไปยังชิ้นงานในโรงงาน ญัฐพล วงศ์สุวรรณ (2564) งานด้าน Stereo Matching Computer Vision มีการจัดหมวดหมู่และจำแนกและเปรียบเทียบสถาปัตยกรรมโมเดล Deep Learning ที่เพื่อคำนวณความลึก (Depth Map) ประเสริฐ, ส. และคณะ. (2564) ใช้ปัญญาประดิษฐ์ในการรักษาความปลอดภัย ด้วยการสกัดคุณลักษณะของภาพ (CNN) และการจดจำลำดับเวลา (LSTM) ใช้ช่วยลดความผิดพลาดในการแจ้งเตือนจากกล้องวงจรปิด สมเกียรติ น. (2563) ระบบสเตอริโอวิชัน (Stereo Vision) จะเกิดความผิดพลาดแบบไม่เป็นเชิงเส้น (Non-linear error) เมื่อจับภาพระยะไกล นำเสนอการแก้ปัญหาโดยใช้สมการเส้นตรงเพื่อสร้างแบบจำลองทำให้ผลที่ได้ถูกต้องมากขึ้น สุรศักดิ์ พ. (2562) การคำนวณแบบการถดถอยเชิงเส้น (Linear Regression) งานวิจัยนี้มุ่งเน้นการแก้ปัญหาค่าเคลื่อนที่ของหุ่นยนต์หลบหลีกสิ่งกีดขวางในสภาวะแวดล้อมจริง โดยใช้ข้อมูลภาพจากกล้องแบบ Stereo Vision ซึ่งสามารถประเมินระยะทางและความลึกของวัตถุ จากนั้นนำหลักการทางสถิติและ มาใช้ในการหาความสัมพันธ์ของข้อมูล เพื่อช่วยประเมินทิศทางและตัดสินใจควบคุมการหลบหลีกได้อย่างมีประสิทธิภาพ สมพร เตียเจริญ และคณะ (2566) ในบทความนี้ได้พัฒนาหุ่นยนต์นำทางอัตโนมัติใช้ระบบปฏิบัติการหุ่นยนต์ ROS ร่วมกับ Lidar โดยสร้างแผนที่ด้วย SLAM ในการหลบสิ่งกีดขวางโดยการใช้ AMCL ในการนำทางอัตโนมัติ

7. การพัฒนาระบบ

การสร้างและพัฒนาต้นแบบแยกเป็น 2 ส่วนคือการพัฒนากล้องจับภาพด้วย ESP32 CAM และการพัฒนาโปรแกรมในการประมวลผลบนคอมพิวเตอร์ ดังแสดงระบบในภาพที่ 5

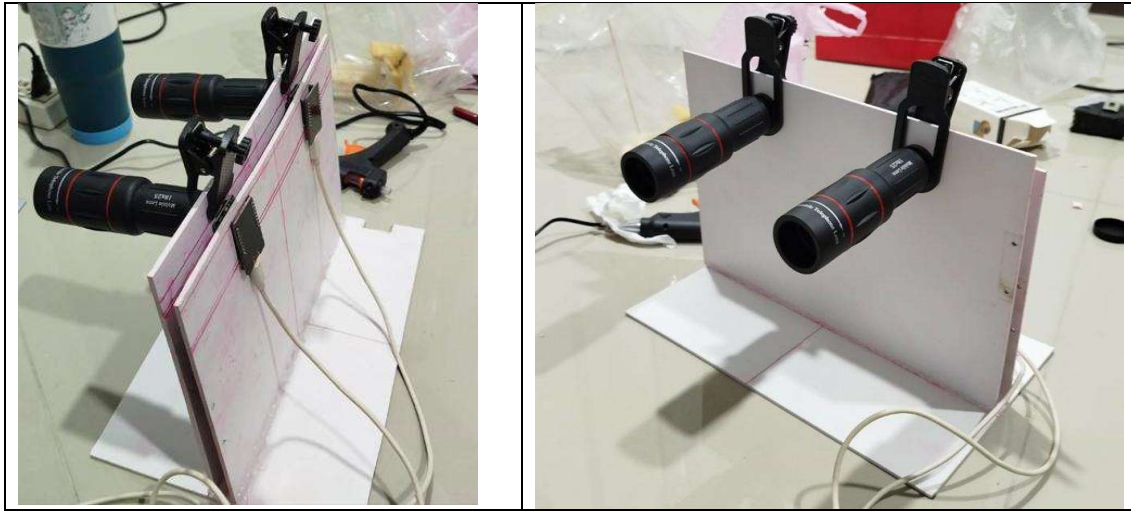


ภาพที่ 5 โปรแกรมบนคอมพิวเตอร์รับภาพจากกล้อง ESP32 CAM 2 ตัว

7.1 การพัฒนากล้องจับภาพด้วย ESP32 CAM

โปรแกรมพัฒนาด้วย Arduino IDE เป็นภาษา C โดยการกำหนดการทำงานเบื้องต้นให้กับ ESP32 CAM มีการกำหนดขนาดของภาพความละเอียด 640×480 พิกเซล ความเร็วในการสตรีมภาพ การเชื่อมต่อ WiFi กำหนด ssid และรหัสผ่าน จากนั้นจึงส่งข้อมูลภาพมายัง WiFi การดูสตรีมภาพและจับภาพสามารถทำได้ผ่านทางหมายเลข IP เนื่องจากในการทดลองครั้งแรกต้องการให้ได้ระยะทางในการตรวจจับภาพไกลที่สุดจึงได้ซื้อเลนส์ขยายภาพมาประกอบกับ ESP32 CAM ดังภาพที่ 2 สาย USB สีขาวที่เห็นในภาพเป็นสายจ่ายพลังงานไฟฟ้าให้กับบอร์ดเอสพี ESP32 CAM โดยคุณลักษณะของเลนส์

โทรศัพท์มือถือคือมีรูรับแสงของเลนส์ 25 มิลลิเมตร เส้นผ่าศูนย์กลางเคลมป์ 18/28 มิลลิเมตร ขนาดของเลนส์ 13.75 x 6.8 x 7.3 เซนติเมตร ดังแสดงในภาพที่ 6

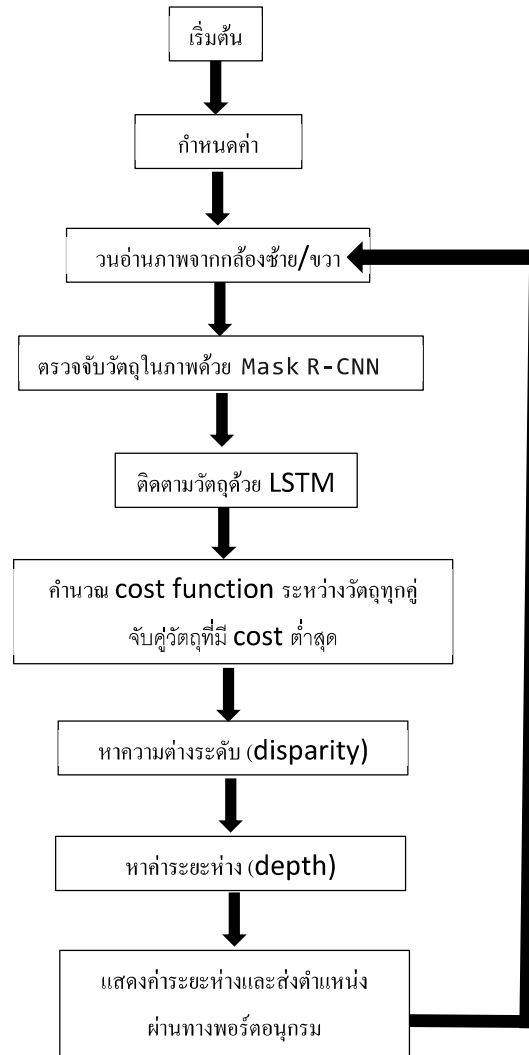


ภาพที่ 6 การติดตั้งเลนส์ขยายภาพร่วมกับ ESP32 CAM

7.2 การพัฒนาโปรแกรมในการประมวลผลบนคอมพิวเตอร์

โปรแกรมพัฒนาด้วยภาษา python โดยเครื่องคอมพิวเตอร์ต้องเชื่อมต่อ WiFi เดียวกันกับบอร์ด ESP32 CAM ทั้ง 2 ตัว การ inference ทำบนเครื่องคอมพิวเตอร์และ ESP32 เป็นเพียง image source โดยโปรแกรมจะเริ่มจากการอ่านค่าสตรีมภาพจากกล้องทั้ง 2 ตัวผ่านหมายเลข IP จากนั้นทำการจับภาพเพื่อนำไปประเมินผลตามหลักการของสเตอริโอวิชั่นต่อไป ดังแสดงในภาพที่ 3

การทำงานของโปรแกรมจะเริ่มจากการกำหนดการเชื่อมต่อ WiFi ด้วย ssid และรหัสผ่านเดียวกันกับกล้องทั้งสองตัว จากนั้นจะทำการเชื่อมต่อกล้องผ่านหมายเลข IP เพื่อ Streaming video เมื่อได้ภาพจากกล้องจะทำการ Capture ภาพจากกล้องแต่ละตัว ภาพจะผ่านขั้นตอนการตรวจจับวัตถุโดยใช้เทคนิค Mask R-CNN ซึ่งการติดตามวัตถุในภาพวิดีโอ โดยจำนวนภาพทดสอบ 120 ภาพ ซึ่งมีความแม่นยำ 81.7% และมีการกระพริบบางเฟรมเมื่อมีการหายไปของข้อมูลจากอินเทอร์เน็ต โดยจะปรับปรุงให้ดีขึ้นด้วยเทคนิค LSTM ที่จะลดความผิดพลาดเนื่องจากการหายไปของข้อมูลบางส่วน สำหรับวัตถุขนาดกลางทำให้ค่าความแม่นยำเป็น 93.3% ขั้นตอนต่อไปคือทำการ matching โดยใช้หลักการหาค่าที่เหมาะสมที่สุด โดยการหาค่า cost function ของการจับภาพวัตถุแต่ละคู่ เมื่อได้ตำแหน่งภาพของวัตถุที่เป็นวัตถุอันเดียวกันนั้นแล้วก็จะนำมาหาค่าความต่างระดับ คำนวณหาความระยะห่างจากวัตถุถึงระนาบกล้อง และสุดท้ายเมื่อได้ค่าระยะห่างที่ต้องการแล้วโปรแกรมจะส่งค่าผ่านทางพอร์ตอนุกรมไปยังเซนเซอร์หุ่นยนต์ เพื่อให้เซนเซอร์หุ่นยนต์ทำงานหยิบวัตถุต่อไป ดังแสดงในภาพที่ 7



ภาพที่ 7 การทำงานของโปรแกรมในการประมวลผลบนคอมพิวเตอร์

8. ผลการทดลอง

ผลการทดลอง ESP32 CAM หลังจากโปรแกรมลงบอร์ดเรียบร้อยแล้ว จะสังเกตการเชื่อมต่อ WiFi เมื่อบอร์ดสามารถเชื่อมต่อ WiFi ได้จะแสดงหมายเลข IP บนพอร์ตอนุกรม โดยจากการทดลองกล้องซ้ายมือจะได้รับหมายเลข IP เป็น 192.168.0.143 และกล้องขวามือจะได้รับหมายเลข IP เป็น 192.168.0.199 โดยภาพจะถูกจับภาพเป็น Format JPEG และการอ่านข้อมูลของ Steam Video จะใช้คำสั่ง HTTP_GET ผลการทำงานทดลองการจับภาพแบบสเตอริโอ แสดงในภาพที่ 8



ภาพที่ 8 ผลการทดสอบการจับภาพแบบสเตอริโอ

ระบบการประมวลผลใช้คอมพิวเตอร์ส่วนบุคคล ติดต่อกับกล้องคู่ ใช้ไลบรารีโอเพนซีวี (OpenCV) สำหรับการประมวลผลภาพ และใช้เพทอร์ช (PyTorch) สำหรับโมเดลการเรียนรู้ของเครื่องจักร โดยการตรวจจับวัตถุสถาปัตยกรรม Mask R-CNN ร่วมกับ LSTM เพื่อจับลักษณะเฉพาะจากวิดีโอ(เฟรมติดต่อกัน)

ตัวอย่างผลการทดลองโดยพื้นที่ในการทดสอบเป็นโต๊ะขนาด 100 x 100 cm วางชิดกับผนังพื้นสีขาว แสดงในตารางที่

3

ตารางที่ 3 ผลการตรวจจับวัตถุ

	ระยะจริง (ซม.)	20	30	40	50	60	70
วัตถุ ขนาดเล็ก	ค่าเฉลี่ยที่วัดได้ (ซม.)	20.42	30.85	41.12	51.48	62.15	72.90
วัตถุ ขนาดกลาง	ค่าเฉลี่ยที่วัดได้ (ซม.)	20.28	30.55	40.72	50.95	61.30	71.85

ซึ่งเมื่อเทียบกับผลระยะทางจริงที่วัดด้วยตลับเมตรพบว่ามีความคลาดเคลื่อนเฉลี่ยอยู่ที่ประมาณ 1.5 เซนติเมตร ค่าความแม่นยำร้อยละ 90 จากตารางที่ 3 พบว่าค่าคลาดเคลื่อนสัมพัทธ์มีแนวโน้มเพิ่มขึ้นเมื่อระยะทางมากขึ้นซึ่งสอดคล้องกับทฤษฎีของความละเอียดเชิงความลึก และเมื่อทดสอบกับวัตถุขนาดเล็ก 2x2x2 เซนติเมตร พบว่าเมื่อระยะทางน้อยและวัตถุมีขนาดเล็กจะทำให้ผลการคำนวณมีความแม่นยำมากขึ้น ค่าความแม่นยำมากกว่าร้อยละ 80 การวัดระยะทางจะทำได้ในช่วง 20-70 เซนติเมตร ด้วยค่าคลาดเคลื่อนสัมพัทธ์ร้อยละ 5 สำหรับวัตถุขนาดเล็กและต่ำกว่าร้อยละ 3 สำหรับวัตถุขนาดกลาง 4x4x4 เซนติเมตร ผลการทดสอบโดยการจับของแขนหุ่นยนต์ โดยแขนมีความเร็วเฉลี่ย 6 เซนติเมตรต่อวินาทีเมื่อทดสอบที่ระยะ 50 เซนติเมตร ค่าความผิดพลาด 0.5 เซนติเมตรเมื่อกำหนดตำแหน่งเป้าหมายแล้วส่งงานแขนหุ่นยนต์ผ่านพอร์ตอนุกรมจะพบว่าแขนหุ่นยนต์สามารถจับวัตถุขนาดเล็กที่ระยะใกล้ได้ดีแต่ที่ระยะไกลเกินกว่าระยะเอื้อมของแขนหุ่นยนต์ไม่สามารถจับได้ขณะเดียวกันการใช้วัตถุขนาดใหญ่ (ขวดน้ำอัดลม) วางไว้ในระยะทางใกล้กับกล้องสเตอริโอแขนหุ่นยนต์ไม่สามารถเอื้อมไปจับวัตถุได้เนื่องจากจะชนกับแขนของหุ่นยนต์เอง

9. สรุปและข้อเสนอแนะ

ระบบกล้องสเตอริโอขึ้นเนื่องจากในบทความนี้ใช้ระยะ baseline ของกล้องคู่ไม่มากทำให้สามารถวัดระยะความลึกได้ดีสำหรับวัตถุที่อยู่ใกล้ ในอนาคตหากต้องการวัดความลึกของวัตถุที่มีระยะไกลออกไปจำเป็นต้องเปลี่ยนระยะ baseline ของกล้องคู่ให้มากขึ้น การพัฒนาระบบที่ผสมผสานเทคนิค Mask R-CNN ร่วมกับ LSTM สำหรับการตรวจจับวัตถุช่วยเพิ่มความแม่นยำในการตรวจจับจากวิดีโอสามมิติ ลดการกระพริบ การใช้การถดถอยเชิงเส้นช่วยลดค่าผิดพลาดของการวัดได้ดี และระบบสามารถส่งงานแขนหุ่นยนต์ ให้หยิบวัตถุในระยะใกล้ได้ ข้อจำกัดของระบบคือยังต้องทำการรันบนเครื่องคอมพิวเตอร์ส่วนบุคคลซึ่งในอนาคตสามารถใช้อุปกรณ์อย่างอื่นช่วยในการรันได้เช่น บอร์ดราสเบอร์รี่พาย [11] สำหรับการส่งงานแขนหุ่นยนต์นั้นหากต้องการจะหยิบวัตถุในระยะที่ไกลมากขึ้นอาจจะต้องใช้รูปแบบของแขนหุ่นยนต์เป็นแบบ cartesian robot ซึ่งสามารถกำหนดระยะความยาวของแขนได้จากแกน x และ y ที่มากขึ้น จากการพัฒนาที่สามารถนำประยุกต์ใช้ได้กับหลากหลายงานในอนาคตเช่นการเพาะชำต้นกล้าทางการเกษตรหรือการหยิบจับวัตถุหรือวัสดุในอุตสาหกรรมด้วยหุ่นยนต์นำทางอัตโนมัติ [12]



10. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากวิทยาลัยเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ตามสัญญาเลขที่ Res-CIT 534/2023

11. เอกสารอ้างอิง

- [1] Kittisak, S. (2021). A survey of research on stereo vision for industrial robots. *Journal of Electrical and Computer Engineering*, 15(2), 45–58. (In Thai)
- [2] Somkiat, N. (2020). The use of piecewise regression for distance correction in stereo vision systems. *Journal of Electrical and Electronics Engineering*, 12(4), 201–212. (In Thai)
- [3] Somchai, R. (2018). *Principles of stereo vision and applications*. Information Technology Publishing. (In Thai)
- [4] Nirut, M. (2020). Analysis of measurement resolution for distance measurement in stereo vision systems. *Journal of Industrial Engineering*, 28(1), 67–79. (In Thai)
- [5] Wirot, K. (2022). A study of measurement errors in distance measurement using stereo vision. *Journal of Mechanical Engineering*, 33(1), 15–28. (In Thai)
- [6] Prasert, S., et al. (2021). The use of CNN-LSTM for object detection in intelligent surveillance systems. *Journal of Computer Science*, 18(2), 89–102. (In Thai)
- [7] Nattapon, W. (2021). Classification of deep learning models for stereo matching. *Journal of Science and Technology*, 29(3), 112–125. (In Thai)
- [8] Weerachai, P. (2022). Application of LSTM for real-time object detection. In *Proceedings of the 12th Thailand Conference on Artificial Intelligence and Machine Learning* (pp. 156–163). Chiang Mai University. (In Thai)
- [9] Atthaphon, S. (2021). Classification of deep learning models for stereo image matching. *Journal of Computer Science*, 18(1), 34–47. (In Thai)
- [10] Surasak, P. (2019). *Development of an obstacle avoidance system for mobile robots using stereo vision combined with linear regression* [Master's thesis, King Mongkut's University of Technology Thonburi]. (In Thai)
- [11] Somchai Tiacharoen, & Somporn Tiacharoen. (2025). Development of a small robotic arm controlled by Edge AI for route recognition using an object-tracking camera. *Engineering Journal Research and Development*, 36(4), 9–19. (In Thai)
- [12] Somporn Tiacharoen, Wasana Prempramote, & Jirayu Chanduang. (2023). Autonomous navigation robot. In *Proceedings of the 15th Electrical Engineering Network Conference* (May 1–3, 2023). Fortune River View Hotel, Nakhon Phanom. (In Thai)