บทที่ 2

ทฤษฎี งานวิจัยและระบบงานที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ปัญญาประดิษฐ์

ปัญญาประดิษฐ์ (AI : Artificial Intelligence) คือ เครื่องจักร(machine) ที่มีฟังก์ชันที่มีความสามารถในการทำความเข้าใจ เรียนรู้องค์ความรู้ต่าง ๆ อาทิเช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหาต่าง ๆ เครื่องจักรที่มีความสามารถเหล่านี้ก็ถือว่าเป็น ปัญญาประดิษฐ์ (AI : Artificial Intelligence) นั่นเอง

เพราะฉะนั้นจึงสามารถกล่าวได้ว่า AI ถือกำเนิดขึ้นเมื่อเครื่องจักรมีความสามารถที่จะเรียนรู้นั่นเอง ซึ่ง AI ก็ถูกแบ่งออกเป็นหลายระดับตามความสามารถหรือความฉลาด โดยจะวัดจากความสามารถในการ ให้เหตุผล การพูด และทัศนคติของ AI ตัวนั้น ๆ เมื่อเปรียบเทียบกับมนุษย์อย่างเราๆ

AI ถูกจำแนกเป็น 3 ระดับตามความสามารถหรือความฉลาดดังนี้

1) ปัญญาประดิษฐ์เชิงแคบ (Narrow AI) หรือ ปัญญาประดิษฐ์แบบอ่อน (Weak AI) : คือ AI ที่มีความสามารถเฉพาะทางได้ดีกว่ามนุษย์(เป็นที่มาของคำว่า Narrow(แคบ) ก็คือ AI ที่เก่งในเรื่องแคบๆ หรือเรื่องเฉพาะทางนั่นเอง) อาทิ เช่น AI ที่ช่วยในการผ่าตัด(AI-assisted robotic surgery) ที่อาจจะเชี่ยวชาญเรื่องการผ่าตัดกว่าคุณหมอยุคปัจจุบัน แต่แน่นอนว่า AIตัวนี้ไม่สามารถที่จะทำอาหาร ร้องเพลง หรือทำสิ่งอื่นที่นอกเหนือจากการผ่าตัดได้นั่นเอง ซึ่งผลงานวิจัยด้าน AI ณ ปัจจุบัน ยังอยู่ที่ระดับนี้

2) ปัญญาประดิษฐ์ทั่วไป (General AI) : คือ AI ที่มีความสามารถระดับเดียวกับมนุษย์ สามารถทำทุก ๆอย่างที่มนุษย์ทำได้และได้ประสิทธิภาพที่ใกล้เคียงกับมนุษย์

3) ปัญญาประดิษฐ์แบบเข้ม (Strong AI) : คือ AI ที่มีความสามารถเหนือมนุษย์ในหลายๆด้าน

โดยสรุป , ปัญญาประดิษฐ์หรือ AI เป็นเทคโนโลยีที่ล้ำสมัยที่สามารถรับมือกับปัญหาที่ซับซ้อนเกินกว่าที่มนุษย์จะสามารถรับมือได้ และ AI ยังเป็นเครื่องมือที่สามารถทำงานที่ซ้ำซากน่าเบื่อแทนมนุษย์ได้อย่างดีเยี่ยม ช่วยให้เราสามารถมีเวลาไปโฟกัสงานที่สำคัญและสามารถสร้างมูลค่าได้มากกว่า นอกจากนี้การประยุกต์ใช้ AI ในระดับอุตสาหกรรม ยังช่วยลดต้นทุนและเพิ่มรายได้มหาศาล

ประวัติย่อ AI เป็นคำยอดฮิตที่ปัจจุบัน แม้ว่ามันไม่มันจะไม่ใช่คำที่เพิ่งถูกบัญญัติขึ้นมาใหม่แต่อย่างใด ในปี 1956 , กลุ่มของผู้เชี่ยวชาญแนวหน้าจากหลายๆวงการได้ร่วมกันทำงานวิจัยเกี่ยวกับ AI มีผู้นำทีมได้แก่ John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (IBM) และ Claude Shannon (Bell Telephone Laboratories) โดยมีจุดประสงค์หลักของงานวิจัย คือ การค้นหามุมมองและหลักการต่าง ๆที่ใช้การเรียนรู้อย่างครอบคลุมเพื่อที่จะนำมาประยุกต์ใช้ให้เครื่องจักรสามารถเรียนรู้ได้เช่นกัน โดยมีเนื้อหาของโครงการมีดังนี้

1.) คอมพิวเตอร์อัตโนมัติ (Automatic Computers)

2.) จะสามารถเขียนโปรแกรมคอมพิวเตอร์โดยใช้ภาษาคอมพิวเตอร์ได้อย่างไร (How Can a Computer Be Programmed to Use a Language?)

3.) โครงข่ายประสาทเทียม (Neural Nets)

4.) การพัฒนาด้วยตนเอง (Self-improvement)

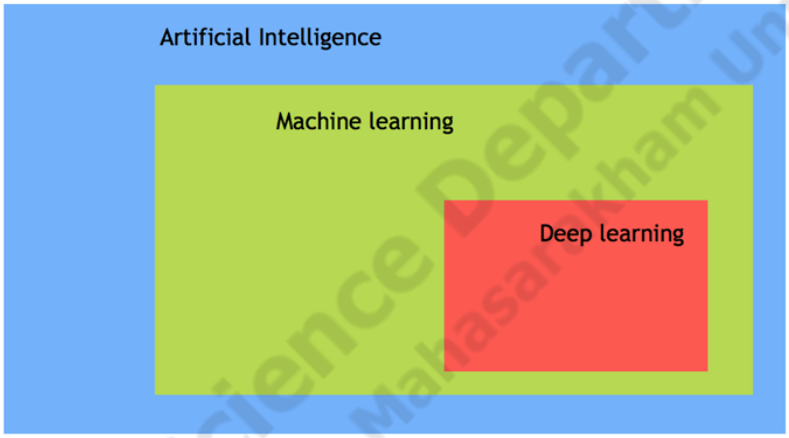
องค์ความรู้เหล่านี้เป็นองค์ความรู้พื้นฐานที่ทำให้คอมพิวเตอร์มีความฉลาดมากขึ้น และยังทำให้ความคิดที่จะการสร้าง AI มีความเป็นไปได้มากยิ่งขึ้น

ชนิดของ AI (Type of Artificial Intelligence) ถูกแบ่งออกเป็น 3 sub field ได้แก่

1) ปัญญาประดิษฐ์ (Artificial Intelligence)

2) การเรียนรู้ของเครื่อง (Machine learning)

3) การเรียนรู้เชิงลึก (Deep Learning)



ภาพประกอบที่ 2.1 ชนิดของ AI

AI คือ คอมพิวเตอร์ที่มีคุณสมบัติและความสามารถคล้ายมนุษย์อีกทั้งยังสามารถทำงานได้อย่างลงตัว (seamlessly) หรือ อาจเรียกได้ว่า AI ก็คือ วิทยาศาสตร์ของการเลียนแบบทักษะของมนุษย์

Machine Learning คือ ศาสตร์ของการศึกษา วิธีการคิด (algorithm) ที่ใช้ในการเรียนรู้ (learn) จากตัวอย่าง (example) และ ประสบการณ์ (experience) โดยมีพื้นฐานมาจากหลักการที่เชื่อว่า ทุกสิ่งอย่างมีรูปแบบหรือแบบแผน (pattern) ที่สามารถบ่งบอกความเป็นไปของสิ่งนั้นๆ ซึ่งเราสามารถที่จะนำแบบแผนนี้ มาประยุกต์ใช้เพื่อทำการคาดเดาถึงความเป็นไปในอนาคตได้ (prediction) อาทิเช่น การใช้ machine learning ในการคาดเดาราคาหุ้นในอนาคต จากข้อมูลกราฟในอดีตและปัจจุบัน

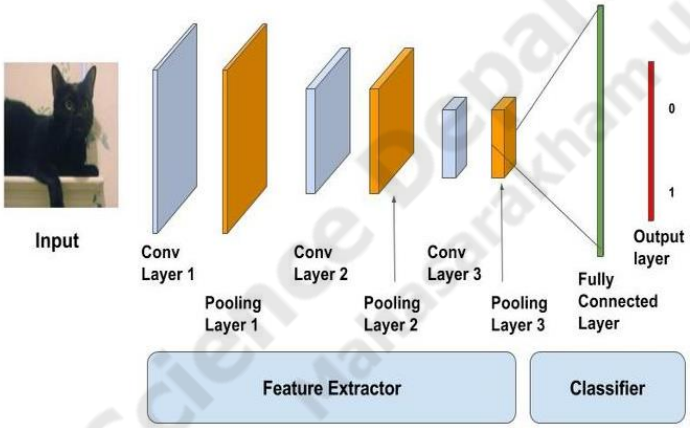
Deep Learning เป็นซับเซตของ Machine Learning โดย Deep Learning นั้นไม่ได้หมายความว่า มันคือการทำความเข้าใจองค์ความรู้ (knowledge) ในเชิงลึก แต่หมายถึงการที่เครื่องจักร (machine)ใช้หลายๆเลเยอร์(layer)ที่แตกต่างกันในการทำความเข้าใจหรือเรียนรู้ข้อมูล โดยความซับซ้อนของโมเดล (model) ก็แปรผันตามจำนวนของเลเยอร์ (layer) ยกตัวอย่างเช่น บริษัทกูเกิล ใช้ LeNet model ในการวิเคราะห์และทำความเข้าใจภาพ (Image recognition) โดยมีการใช้เลเยอร์ทั้งหมด 22 เลเยอร์ โดยใน Deep learning , จะมีเฟสของการเรียนรู้ (learning phase) ที่ถูกสร้างขึ้นโดย Neural Network ซึ่งอาจเรียกได้ว่าเป็นสถาปัตยกรรมของ layer ที่แต่ละ layer ซ้อนทับกันอยู่ (stack)

Computer Science Department Faculty of Informatics, Mahasarakham University

2.1.2 Object Detection

Object Detection คือ การคาดเดาตำแหน่งของวัตถุพร้อมกับประเภทสัญลักษณ์เรียกว่าการตรวจจับวัตถุ แทนที่การคาดเดาคลาสของวัตถุจากรูปภาพตอนนี้เราต้องคาดเดาคลาสรวมถึงสี่เหลี่ยม (เรียกว่ากล่องขอบเขต) ที่บรรจุวัตถุนั้น ใช้ตัวแปร 4 ตัวเพื่อระบุรูปสี่เหลี่ยมผืนผ้าโดยเฉพาะ Object Detection คือการรวมกันของ Classification + Localization

Classification หรือ การจัดหมวดหมู่ คือ การจำแนกภาพเป็นงานหนึ่งในการค้นหาการแปลระหว่างภาพและประเภทในการจำแนกภาพ เราจะส่งภาพเป็นสัญญาณเข้าสู่เครือข่ายและคาดการณ์ผลลัพธ์ขึ้นอยู่กับฉลากที่ผ่านการฝึกอบรมมาแล้วและจัดหมวดหมู่ของวัตถุในภาพ ตัวอย่างเช่นเมื่อเราสร้างตัวจำแนกแมว - สุนัขเราถ่ายภาพของแมวหรือสุนัขและคาดการณ์ระดับของพวกเขา CNNs จะใช้สำหรับงานการจำแนกรูปภาพ

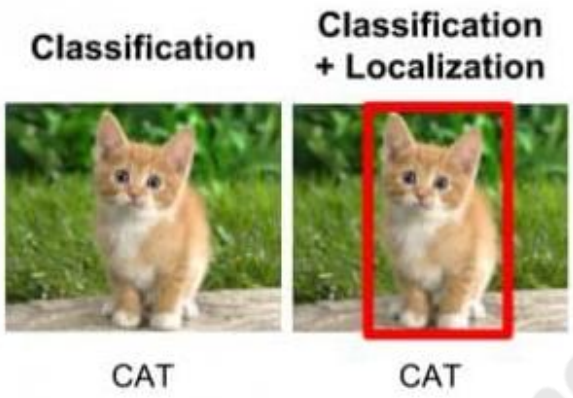


ภาพประกอบที่ 2.2 เครือข่าย CNN สำหรับงานการจัดประเภทรูปภาพใน Deep Learning

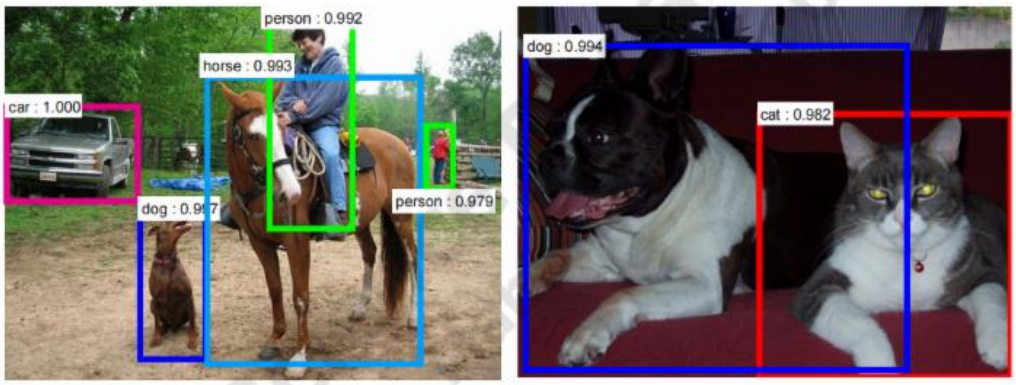
ภาพประกอบที่ 2.3 ผลจากการจำแนกแมวและสุนัข

Localization หรือ การจำกัดขอบเขต คือการคาดเดาวัตถุในภาพรวมถึงขอบเขตในการแปลภาษาของวัตถุมีจุดมุ่งหมายเพื่อค้นหาวัตถุหลัก (หรือที่มองเห็นได้มากที่สุด) ในภาพในขณะที่การตรวจจับวัตถุพยายามค้นหาวัตถุและขอบเขตทั้งหมด

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.4 ไดอะแกรมเกี่ยวกับการจำกัดขอบเขต



ภาพประกอบที่ 2.5 ผลลัพธ์การคัดแยกโดยใช้อัลกอริทึมที่แตกต่างกันใน Deep leaning

ความแตกต่างระหว่างอัลกอริทึมการตรวจจับวัตถุและอัลกอริทึมการจำแนก คือ ในอัลกอริทึมการตรวจจับ เราพยายามที่จะวาดกล่องล้อมรอบวัตถุที่น่าสนใจเพื่อค้นหาภายในภาพ นอกจากนี้คุณอาจไม่จำเป็นต้องวาดเพียงแค่หนึ่งกล่องที่มีขอบเขตในกรณีการตรวจจับวัตถุอาจมีหลายกล่องที่เป็นตัวแทนของวัตถุที่น่าสนใจในภาพและคุณจะไม่ทราบจำนวนล่วงหน้า

2.1.3 การประมวลผลภาพ

2.1.3.1 การประมวลผลภาพ คืออะไร

การประมวลผลภาพ (Image Processing) คือ การจัดการหรือประมวลผลภาพถ่ายดิจิตอล ซึ่งจะดำเนินการโดยใช้ การคำนวณและประมวลผลทางคอมพิวเตอร์ การประมวลผลข้อมูลภาพในที่ นี้จำแนกตาม วัตถุประสงค์ในการใช้งาน

2.1.3.2 การรับภาพและการได้มาของภาพ (Image Sensing and Acquisition)

การรับภาพเป็นการนำเข้าภาพจากเซนเซอร์ และแปลงสัญญาณภาพให้เป็นภาพถ่ายดิจิตอล โดยขั้นตอนการแปลงสัญญาณภาพจะมีการสุ่มตัวอย่าง (Sampling) สัญญาณภาพที่ ได้รับจากนั้นนำค่าที่ได้ในแต่ละจุดภาพหรือพิกเซล (Pixel) มาจัดระดับข้อมูล (Quantize) เพื่อ แปลงค่าจำนวนจริงที่เป็นค่าไม่ต่อเนื่อง (Discrete) ที่ได้จากการสุ่มตัวอย่าง ให้เป็นค่าจำนวน เต็ม ซึ่งจำนวนเต็ม

นี้จะถูกนำมาแปลงเป็นเลขฐานสองด้วยขั้นตอนการแทนเลขไบนารี (Binary Representation) ขั้นตอนการสุ่มตัวอย่างและจัดระดับข้อมูลนี้จะต้องมีการดำเนินการใน ลักษณะที่ลดความผิดเพี้ยนที่เกิดจากเซนเซอร์ เพื่อให้ได้ภาพที่มีคุณภาพ

จากที่กล่าวมาแล้วว่าคลื่นแม่เหล็กไฟฟ้าที่มีความยาวคลื่นต่าง ๆ นั้นสามารถนำไปใช้สร้าง ภาพได้ แต่การที่จะรับภาพเหล่านั้นจำเป็นจะต้องมีเซนเซอร์ที่เหมาะสมเพื่อตรวจรับสัญญาณตอบสนอง (Image Sensing) ที่ได้รับมาและแปลงให้เป็นภาพสองมิติหรือที่เรียกว่า “เทคนิคการ ได้มาของภาพ (Image Acquisition)” รูปแบบของภาพที่เราสนใจนั้นเป็นส่วนผสมระหว่างแหล่งกำเนิดความเข้มแสงและการสะท้อนหรือการดูดซับพลังงานจากแหล่งกำเนิดพลังงานของ สสารที่อยู่ในสถานที่ ที่ต้องการถ่ายภาพนั้น การถ่ายภาพสามารถจำแนกได้เป็น 2 ประเภทคือ

ภาพแพสซีฟ (Passive Image) ภาพที่ได้จากการใช้แหล่งพลังงานที่มีอยู่ในสถานที่นั้นแล้ว

ภาพแอ๊กทีฟ (Active Image) เป็นภาพที่ได้จาก การใช้แหล่งพลังงานเสมือน คุณภาพของภาพแพสซีฟขึ้นอยู่กับแหล่งกำเนิดของพลังงาน ส่วนคุณภาพของภาพแอ๊กทีฟนั้นขึ้นอยู่กับความซับซ้อนและค่าใช้จ่ายในขั้นตอนการได้มาของภาพ เนื่องจากจำเป็นต้องมีการควบคุมการแผ่รังสีของแหล่งพลังงานนอกเหนือไปจากการใช้อุปกรณ์ในการถ่ายภาพ ตัวอย่างของภาพแอ๊กทีฟ ได้แก่ ภาพที่ใช้ในทางการแพทย์ ภาพจากการรับภาพระยะไกล (Remote Sensing) เช่น ภาพดาวเทียม เป็นต้น ในการถ่ายภาพนอกจากจะต้องมีแหล่งกำเนิดพลังงานแล้วยังต้องมีอุปกรณ์เพื่อรับสัญญาณที่ตอบสนองจากการสะท้อนหรือดูดซับพลังงานหรือสัญญาณตอบสนองที่ได้รับให้เป็นความต่างศักย์ของสัญญาณ ซึ่งค่าความต่าง

2.1.3.3 การปรับปรุงคุณภาพของภาพ (Image Enhancement and Restoration)

การปรับปรุงคุณภาพของภาพเป็นการปรับปรุงหรือซ่อมแซมให้ข้อมูลภาพที่มีอยู่นั้นมีคุณภาพดีขึ้น เช่น ภาพที่ได้มาอาจมีความคมชัด (Contrast) น้อยหรือเบลอ ไม่คมชัด เรา สามารถปรับภาพให้คมชัดได้ด้วยเทคนิค เช่น การปรับค่าความคมชัด (Contrast Enhancement) หรือการปรับเน้นเส้นขอบภาพ (Edge Enhancement) หรือในกรณีที่ภาพที่มี อยู่มีความไม่สมบูรณ์ เช่น มีสัญญาณรบกวน (Noise) เราสามารถใช้เทคนิคการกรองสัญญาณ ภาพ (Image Filtering) เพื่อกำจัดสัญญาณรบกวนได้

2.1.3.4 การบีบอัดภาพ (Image Compression)

เทคนิคการประมวลผลภาพในรูปแบบของการบีบอัดเป็นหัวใจสำคัญในการจัดการข้อมูลภาพเพื่อการรับส่งและการจัดเก็บข้อมูลภาพ เนื่องจากข้อมูลภาพนั้นมีขนาดใหญ่มาก โดยเฉพาะภาพสี การจัดเก็บหรือรับส่งไฟล์ข้อมูลขนาดใหญ่จะเสียเวลามาก ดังนั้นเพื่อการจัดการเก็บและรับส่งข้อมูลภาพให้รวดเร็วขึ้น จึงทำการบีบอัดข้อมูลภาพก่อนเพื่อให้ขนาดของ ข้อมูลภาพเล็กลง ตัวอย่างการบีบอัดข้อมูลภาพ เช่น JPEG หรือ GIF และตัวอย่างการบีบอัด ข้อมูลวีดีโอ เช่น MPEG หรือ AVI เป็นต้น

2.1.3.5 การวิเคราะห์ข้อมูลภาพ (Image Analysis)

การวิเคราะห์ข้อมูลภาพเป็นการวัดหรือวิเคราะห์ข้อมูลภาพเชิงปริมาณ เพื่อสร้างคำจำกัด ความให้กับภาพแล้วนำไปใช้งาน ตัวอย่างเช่น การนำภาพถ่ายของเซลล์เม็ดเลือดมาวิเคราะห์หาขนาดและมุมเอียงแล้วนำข้อมูลที่ได้ไปตรวจสอบความผิดปกติของเม็ดเลือดในผู้ป่วย หรือ การนำภาพถ่ายมาวิเคราะห์เพื่อนำไปควบคุมการทำงานของหุ่นยนต์ เช่น ควบคุมให้หุ่นยนต์ เคลื่อนย้ายวัตถุที่

มีลักษณะตามที่กำหนด ในการวิเคราะห์ข้อมูลภาพจำเป็นต้องมีการแยกแยะ คุณลักษณะของวัตถุต่าง ๆ ในภาพเพื่อช่วยระบุชนิดของวัตถุนั้น นอกจากนี้ยังต้องมีการแยก ส่วนวัตถุ (Segmentation) ที่สนใจออกมาจากรูปภาพหรือฉากนั้น ๆ เพื่อหาคุณลักษณะของวัตถุ

2.1.4 รูปร่างของภาพ

วัตถุที่มีอยู่ตามธรรมชาติและที่มนุษย์สร้างขึ้นมีรูปร่างที่แตกต่างกันไป ทั้งที่เป็นรูปทรงเรขาคณิตและไม่เป็นรูปทรงเรขาคณิต ในศาสตร์ของการประมวลผลภาพนั้น การกำหนดขอบเขตของภาพทุกภาพให้อยู่ในรูปสี่เหลี่ยม เป็นวิธีที่นิยมใช้กันมากที่สุด เนื่องจากทำให้การอ่านภาพ การจัดเก็บข้อมูลภายในหน่วยความจำ และการแสดงภาพออกทางอุปกรณ์ต่าง ๆ เป็นไปได้อย่างมีประสิทธิภาพ

การเก็บข้อมูลภาพลงหน่วยความจำของคอมพิวเตอร์สามารถทำได้โดยการจองหน่วยความจำของเครื่องไว้ในรูปของตัวแปรอาเรย์ โดยค่าในแต่ละช่องของอาเรย์แสดงถึงคุณสมบัติของจุดภาพ และตำแหน่งของช่องอาเรย์เป็นตัวกำหนดตำแหน่งของจุดภาพ

สมมุติให้ Image เป็นตัวแปรแบบอาเรย์ขนาด M X N (M แถว และ N คอลัมน์) ที่ใช้เก็บภาพขนาด M x N จุด (M จุดในแนวนอน และ N จุดในแนวตั้ง) ค่าสี (หรือความสว่าง ในกรณีที่เป็นภาพ grey level) ของจุดภาพในแถวที่ 5 คอลัมน์ที่ 4 จะตรงกับค่าของ Image(5,4) จะเห็นว่าเราใช้ตำแหน่งของจุดภาพทั้งสองแกนเป็นตัวชี้ค่าข้อมูลในอาเรย์

จากการใช้หน่วยความจำเพื่อการเก็บภาพในลักษณะที่กล่าวมา เนื้อที่ในการเก็บภาพสามารถคำนวณได้จาก M x N x g เมื่อ g เป็นจำนวนเต็มที่แทนจำนวนบิตของข้อมูลในแต่ละจุดภาพ ตัวอย่างถ้า g มีค่าเท่ากับ 8 บิต เราจะสามารถเก็บความแตกต่างของระดับสีที่เป็นไปสูงสุด 256 ระดับ ค่า M และ N จะเป็นตัวบอกถึงความละเอียดของภาพ สำหรับคอมพิวเตอร์ทั่วไปในระบบ VGA (Video Graphic Array) จะมีขนาด 640x480, 800x600 และ 1024x768 จุด เป็นต้น การกำหนดความละเอียดจะขึ้นอยู่กับงานที่จะใช้ ในงานบางอย่างใช้ความละเอียดแค่ 30 x 50 จุด ก็พอแล้วแต่ในงานบางชนิด ใช้ความละเอียดถึง 1000 x 1000 จุด ก็ยังไม่พอ

ปกติแล้วในการเก็บข้อมูลภาพโดยเครื่องมือต่าง ๆ จะเก็บตามมาตรฐานของโทรทัศน์ซึ่งมีอัตราส่วน x ต่อ y เท่ากับ 4:3 สำหรับเครื่องมือเก็บข้อมูลภาพที่ไม่เป็นไปตามอัตราส่วน 4:3 เมื่อนำภาพนี้ไปแสดงในจอภาพมาตรฐานจะทำให้ภาพที่แสดงนั้นมีขนาดของจุดภาพไม่เป็นสีเหลี่ยมจัตุรัสเช่นในบางระบบอาจจะใช้ความละเอียดในการแสดงเท่ากับ 640 x 512 ซึ่งจะทำให้ขนาดของจุดภาพที่ได้มีขนาดของด้านกว้างมีความยาวมากกว่าด้านสูง ซึ่งลักษณะดังกล่าวนี้เป็นหัวข้อที่ต้องสนใจสำหรับการเขียนโปรแกรมทางด้านกราฟิกและการจัดการข้อมูล

2.1.5 ระบบสี

มาตรฐานของสีที่ใช้อยู่ในปัจจุบันทีอยู่หลายระบบด้วยกันทั้งนี้จะขึ้นอยู่กับการนำไปใช้ แต่โดยทั่วไปแล้วทุกมาตรฐานจะมีแนวคิดเดียวกันคือ การแทนจุดสีด้วยจุดที่อยู่ภายในพื้นที่ 3 มิติ โดยจะมีแกนอ้างอิงสำหรับจุดสีนั้นในพื้นที่ซึ่งแต่ละแกนจะมีความเป็นอิสระต่อกัน ตัวอย่างเช่นในระบบ RGB จะมีแกนสีคือ แกนสีแดง เขียว และน้ำเงิน ในระบบ HSV จะมีแกนเป็นค่าสี ความสว่าง และความบริสุทธิ์ของสี

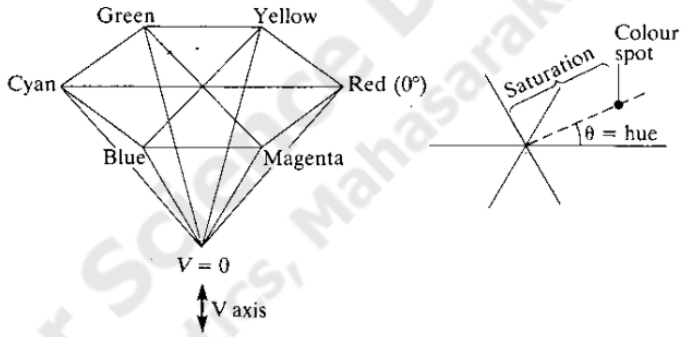
ตัวอย่างระบบสีที่นิยมใช้ได้แก่ ระบบ RGB HSV (Hue Saturation Value) HLS (Hue Lightness Saturation)

2.1.5.1 ระบบสี RGB

ระบบสี RGB เป็นระบบสีที่เกิดจากการรวมกันของแสงสีแดง เขียวและน้ำเงินโดยมีการรวมกันแบบ Additive ซึ่งโดยปกติจะนำไปใช้ในจอภาพแบบ CRT (Cathode ray tube) ในการใช้งานระบบสี RGB ยังมีการสร้างมาตรฐานที่แตกต่างกันออกไปที่นิยมใช้งานแต่ RGB_{CIE} และ RGB_{NTSC}

2.1.5.2 ระบบสี HSV

ระบบสี HSV (Hue Saturation Value) เป็นการพิจารณาสีโดยใช้ Hue Saturation และ Value ซึ่ง Hue คือค่าสีของสีหลัก (แดง เขียว และน้ำเงิน) ในทางปฏิบัติจะอยู่ระหว่าง 0 และ 255 ซึ่งถ้า Hue มีค่าเท่ากับ 0 จะแทนสีแดงและเมื่อ Hue มีค่าเพิ่มขึ้นเรื่อย ๆ สีก็จะเปลี่ยนแปลงไปตามสเปกตรัมของสีจนถึง 256 จึงจะกลับมาเป็นสีแดงอีกครั้ง ซึ่งสามารถแทนให้อยู่ในรูปขององศาได้ ดังนี้คือ สีแดงเท่ากับ 0 องศา สีเขียวเท่ากับ 120 องศา สีน้ำเงินเท่ากับ 240 องศา โดยมีรูปภาพแสดงระบบสี HSV ดังภาพประกอบที่ 2.6



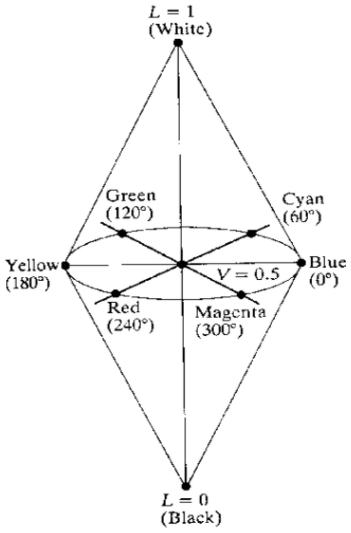
ภาพประกอบที่ 2.6 การแสดงระบบสี HSV

ที่มา http://fivedots.coe.psu.ac.th/~montri/Teaching/image/chap1.htm

2.1.5.3 ระบบสี HLS

ระบบสี HLS (Hue lightness saturation) พัฒนาโดย Teletromix Incorporated จะมีลักษณะคล้ายกับ HSV ดังนี้คือสีของระบบจะขึ้นอยู่กับ Hue Lightness และ Saturation โดยมีรูปภาพแสดงระบบสี HLS ดังภาพประกอบที่ 2.7

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.7 การแสดงระบบสี HLS
ที่มา http://fivedots.coe.psu.ac.th/~montri/Teaching/image/chap1.htm

2.1.6 OpenCV

OpenCV (Open source Computer Vision) เป็นไลบรารีฟังก์ชันการเขียนโปรแกรม (Library of Programming Functions) โดยส่วนใหญ่จะมุ่งเป้าไปที่การแสดงผลด้วยคอมพิวเตอร์แบบเรียลไทม์ (Real-Time Computer Vision) เดิมทีแล้วถูกพัฒนาโดย Intel แต่ภายหลังได้รับการสนับสนุนโดย Willow Garage ตามมาด้วย Itseez (ซึ่งต่อมาถูกเข้าซื้อโดย Intel) OpenCV เป็นไลบรารีแบบข้ามแพลตฟอร์ม (Cross-Platform) และใช้งานได้ฟรีภายใต้ลิขสิทธิ์ของ BSD แบบโอเพ่นซอร์ส (Open-Source BSD License)

OpenCV ยังสนับสนุนเฟรมเวิร์กการเรียนรู้เชิงลึก (Deep Learning Frameworks) ได้แก่ TensorFlow, Torch/PyTorch และ Caffe
ตัวอย่างการประยุกต์ใช้งาน OpenCV มีดังนี้

- ชุดเครื่องมือคุณลักษณะ 2 มิติและ 3 มิติ (2D and 3D feature toolkits)
- การประมาณระยะในขณะเคลื่อนที่ (Egomotion Estimation)
- ระบบรู้จำใบหน้า (Facial recognition system)
- การจดจำท่าทาง (Gesture recognition)
- ปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ (Human-Computer interaction; HCI)

ภาษาการเขียนโปรแกรม

OpenCV ถูกเขียนขึ้นด้วยภาษา C++ มีการรองรับ Python, Java และ MATLAB/OCTAVE — API สำหรับอินเทอร์เฟสเหล่านี้สามารถพบได้ในเอกสารออนไลน์ ซึ่งมีการรวมไว้หลากหลายภาษา เช่น C#, Perl, Ch, Haskell และ Ruby ได้รับการพัฒนาเพื่อส่งเสริมการนำมาใช้งานโดยผู้ใช้ที่เพิ่มขึ้น

Computer Science Department
Faculty of Informatics Mahasarakham University

2.1.7 กล้องเว็บแคม

2.1.7.1 เว็บแคม คืออะไร ?

เว็บแคม (Webcam) หรือ ชื่อเรียกเต็มๆว่า Web Camera แต่ในบางครั้งก็มีคนเรียกว่า Video Camera หรือ Video Conference ก็แล้วแค่ความเข้าใจแต่ละคน เว็บแคมเป็นอุปกรณ์อินพุตที่สามารถจับภาพเคลื่อนไหวของเราไปปรากฏในหน้าจอมอนิเตอร์ และสามารถส่งภาพเคลื่อนไหวนี้ผ่านระบบเครือข่ายเพื่อให้คนอีกฟากหนึ่งสามารถเห็นตัวเราเคลื่อนไหว ได้เหมือนอยู่ต่อหน้า ถือว่าเป็นอุปกรณ์ที่มีประโยชน์อีกตัวหนึ่ง และเริ่มมีความจำเป็นมากขึ้นเรื่อยๆ

2.1.7.2 ประเภทของเว็บแคม

อุปกรณ์อย่างกล้องเว็บแคมไม่ใช่ว่าจะเหมือนกันหมดทุกตัว แต่ละรุ่น แต่ละยี่ห้อจะมีลักษณะและคุณสมบัติที่แตกต่างกันไปตามแต่ผู้ผลิตจะคิดค้นและออกแบบมาให้เหมาะสมกับการใช้งานอย่างไร ซึ่งสามารถแบบประเภทของเว็บแคมได้ดังนี้

1) แบ่งตามรูปทรงของกล้อง

โดยปกติกล้องเว็บแคมส่วนใหญ่จะเป็นทรงกลม เนื่องจากเป็นรูปทรงต้นแบบที่ทำกันมานานและก็ทำให้รู้ได้ทันทีว่านี้คืออุปกรณ์ เว็บแคม แต่ไม่จำเป็นที่กล้องเว็บแคมต้องเป็นทรงกลมเสมอไป เพราะบางครั้ง กล้องเว็บแคม ก็จำเป็นต้องมีรูปทรงอื่นๆ เพื่อให้เข้ากับการใช้งานในบางลักษณะ ดังนั้น การเลือกรูปทรงให้เหมาะสมนั้น ก็จะขึ้นอยู่กับลักษณะการใช้งานของเรามากกว่า

2) แบ่งตามประเภทของขาตั้งกล้อง

โดยส่วนใหญ่ลักษณะของฐานตั้งกล้องจะเป็นแบบตั้งพื้นเสียส่วนใหญ่ โดยแบบแรก คือ แบบมีขาสำหรับวางบนพื้น อาจจะมีขา 3 ขา หรือ 4 ขา ก็แล้วแต่การออกแบบ แต่ฐานแบบ 3 ขา จะมีปัญหาตรงที่ วางแล้วยังไม่มั่นคงดีนัก และไม่สามารถหมุนตัวกล้องได้สะดวกนัด ดังนั้น ถ้าต้องการเว็บแคมที่มีฐานมั่นคงและสามารถหมุนได้ง่ายๆ ก็ต้องเลือกแบบฐานทรงกลมขนาดใหญ่ ซึ่งแบบนี้จะมีข้อดีตรงที่ วางได้มั่นคงและยังสามารถหมุนแกน ของตัวกล้องได้ไม่จำเป็นต้องยกตัวกล้องหมุนไปมาให้เสียเวลา

3) แบ่งตามชนิดของเซ็นเซอร์

สำหรับเซ็นเซอร์ที่กล้องเว็บแคมใช้นั้นจะมีหลักๆอยู่ 2 ชนิด คือ CCD และ CMOS แต่ที่นิยมใช้กันมากที่สุดในตอนนี้ก็คือ CMOS เนื่องจากเหตุผลหลายๆประการและตัวเซ็นเซอร์ แบบ CMOS เองก็สามารถแบบออกได้ถึง 2 ชนิดด้วยกันคือ

CLF Color CMOS Censor ที่มีความละเอียดของพิกเซลแค่ 110,000 พิกเซล (367 x 291) เท่านั้น ในขณะที่ VGA Color CMOS Censor ให้ความละเอียดที่สูงกว่าที่ 350,000พิกเซล (655 x 493) ดังนั้น เวลาเลือกซื้อกล้องเว็บแคมก็ดูได้ทั้งความละเอียดที่ระบุไว้ หรือชนิดของ CMOS สำหรับเซ็นเซอร์แบบ CCD จะเป็นเซ็นเซอร์ที่นิยมใช้ในกล้องดิจิตอล เพราะให้ความละเอียดที่สูงกว่าและก็มี noise ไม่มากเหมือนกับเซ็นเซอร์แบบ CMOS

4) แบ่งตามรูปแบบการเชื่อมต่อ

สำหรับการเชื่อมต่อของกล้องเว็บแคมในปัจจุบันส่วนใหญ่ จะเป็นอินเทอร์เฟซแบบ USB แทบทั้งสิ้นโดย USB ที่ใช้ก็จะเป็นเวอร์ชัน 1.1 เสียส่วนมาก แต่ก็จะมีเวอร์ชัน 2.0 ในบางรุ่นกล้อง

Computer Science Department Faculty of Informatics Mahasarakham University

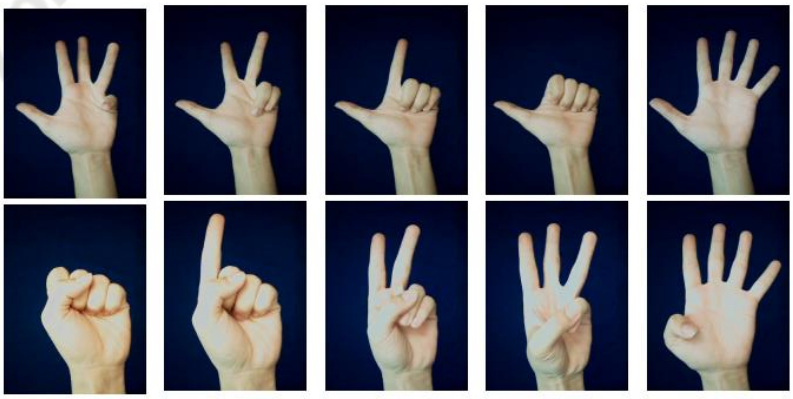
เว็บแคมแบบไร้สายจะใช้การเชื่อมต่อในแบบ WiFi หรือ Wireless lan นั่นเองทำให้สามารถเคลื่อนย้ายไปได้ทุกที่โดยไม่ต้องคำนึงถึงสายให้วุ่นวาย แต่เว็บแคมที่เป็น Wireless ตอนนี้ก็ยังมีราคาค่อนข้างแพง

ภาพประกอบที่ 2.8 กล้องเว็บแคม EGA TYPE-W1
ที่มา http://egazone.com/type-w1/

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 A Simple Shape-Based Approach to Hand Gesture Recognition การจดจำท่าทางมือตามรูปร่างพื้นฐาน

Amornched และคณะ ได้ศึกษางานวิจัยเรื่อง A simple shape-Based Approach to Hand Gesture Recognition โดยงานวิจัยดังกล่าวได้วิจัยเกี่ยวกับการรู้จำท่าทางของมือบนพื้นฐานของรูปร่างท่าทางของมือประกอบด้วย 10 ท่าทาง ดังภาพประกอบที่ 2.9 งานวิจัยนี้มี 3 ขั้นตอนในการรู้จำท่าทางของมือ ประกอบด้วย การแยกมือออกจากภาพ (segmentation) การคำนวณคุณลักษณะเด่นของภาพ (Feature calculation) และ การรู้จำท่าทางมือ (hand gesture) ได้นำภาพ 400 ภาพมาทดลอง โดยแบ่งภาพเป็น 2 ส่วนคือ ส่วนของการเรียนรู้และส่วนของการทดสอบอย่างละ 200 ภาพ ผลการทดลองของงานวิจัยนี้มีความถูกต้อง 91%



ภาพประกอบที่ 2.9 พื้นฐานของรูปร่างท่าทางของมือทั้ง 10 ท่าทาง
ที่มา https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5491589

Computer Science Department
Faculty of Informatics Mahasarakham University

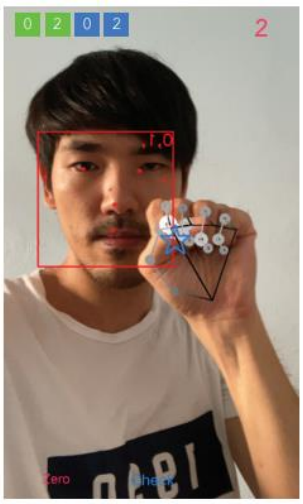
2.2.2 การรู้จำท่ามือภาษาไทยโดยใช้การจำลองแบบอาณาจักรมดร่วมกับซัพพอร์ตเวกเตอร์แมชชีน

นายอนุสรณ์ อุ่นท้าว ได้ศึกษางานวิจัยเรื่องการรู้จำท่ามือภาษาไทยโดยใช้การจำลองแบบอาณาจักรมดร่วมกับซัพพอร์ตเวกเตอร์แมชชีน โดยจากการศึกษางานวิจัยเกี่ยวกับการประมวลผลรู้จำภาพด้วยซัพพอร์ตเวกเตอร์แมชชีนแบบเรเดียลเบสิสฟังก์ชันของ สมปอง เวฬุวนาธร และสุพจน์ สีบุตร ที่กล่าวถึงปัญหาการรู้จำการกำหนดค่าพารามิเตอร์ที่สำคัญ จึงได้นำเสนอขั้นตอนวิธีพาทิเคิลสวอร์มออพทิไมเซชันซึ่งเป็นวิธีเชิงวิวัฒนาการมาแก้ปัญหาในการปรับค่าพารามิเตอร์ร่วมกับซัพพอร์ตเวกเตอร์แมชชีน และด้วยวิธีพาทิเคิลสวอร์มออพทิไมเซชันมีประสิทธิภาพสูงกว่าการรู้จำด้วยซัพพอร์ตเวกเตอร์แมชชีนเพียงอย่างเดียว การจำลองแบบอาณาจักรมดซึ่งเป็นวิธีเชิงวิวัฒนาการอีกวิธีหนึ่งที่สามารถค้นหาค่าพารามิเตอร์ ร่วมกับซัพพอร์ตเวกเตอร์แมชชีนได้อย่างมีประสิทธิภาพ ตัวอย่างการแสดงภาษามือไทยดังภาพประกอบที่ 2.10

ภาพประกอบที่ 2.10 การแสดงภาษามือไทย
ที่มา http://www.esanpedia.oar.ubu.ac.th/e-research/sites/default/files/Anusorn_Untao.pdf

2.2.3 การรู้จำท่าทางมือสำหรับตรวจจับความมีชีวิตของผู้ใช้แบบทันทีทันใดในแอปพลิเคชันโทรศัพท์เคลื่อนที่โดยใช้การเรียนรู้เชิงลึก

วรเมธ เลิศศิวนนท์ และฐิติรัตน์ ศิริบวรรัตนกุลได้ศึกษางานวิจัยเรื่องการรู้จำท่าทางมือสำหรับตรวจจับความมีชีวิตของผู้ใช้แบบทันทีทันใดในแอปพลิเคชันโทรศัพท์เคลื่อนที่โดยใช้การเรียนรู้เชิงลึก มีจุดประสงค์เพื่อทดสอบความเป็นไปได้และประสบการณ์ของผู้ใช้ในสถานะการณ์ที่สัญลักษณ์มือถูกนำมาใช้เพื่อเพิ่มความปลอดภัยให้กับระบบระบุตัวตนด้วยใบหน้าบทสมาร์ทโฟน ในงานนี้ผู้วิจัยได้พัฒนาแบบจำลองการเรียนรู้เชิงลึกที่ผสานการใช้งาน MediaPipe ของกูเกิลร่วมกับแบบจำลองโครงข่ายประสาทเทียมแบบลึกที่ผู้วิจัยพัฒนาเอง ทำให้สามารถแยกแยะสัญลักษณ์มือของการนับเลขศูนย์ถึงห้าได้ โดยรูปหน้าจอสมาร์ทโฟนระหว่างการใช้งานระบบต้นแบบ ดังภาพประกอบที่ 2.11

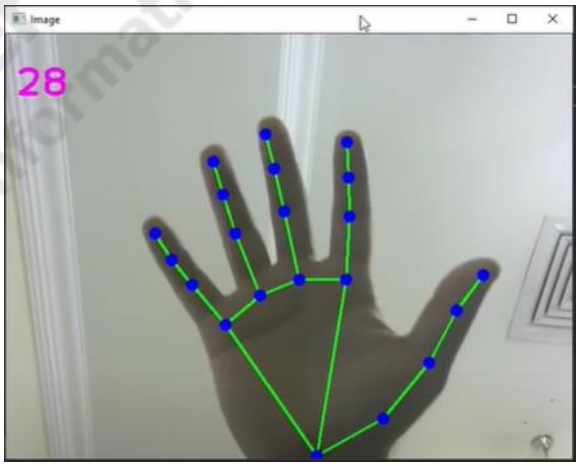


ภาพประกอบที่ 2.11 หน้าจอสมาร์ทโฟนระหว่างการใช้งานระบบต้นแบบ
ที่มา http://ojs.kmutnb.ac.th/index.php/kjournal/article/view/4255/3615

2.3 ระบบงานที่เกี่ยวข้อง

2.3.1 Hand Tracking

เป็นโปรแกรมที่จะคอยตรวจจับมือ ที่จะรับเข้ามาเป็นข้อมูลจากกล้องเว็บแคม ที่ติดตั้งบนเครื่องคอมพิวเตอร์ แบบเรียลไทม์ โดยพัฒนาโปรแกรมนี้ขึ้นมาด้วยภาษาไพทอน เป็นโปรแกรมที่มีการทำงานที่ไม่ซับซ้อน ในบริเวณที่มีแสงสว่างไม่เพียงพอ โปรแกรมอาจไม่สามารถติดตามมือได้เต็มประสิทธิภาพ ดังภาพประกอบที่ 2.12

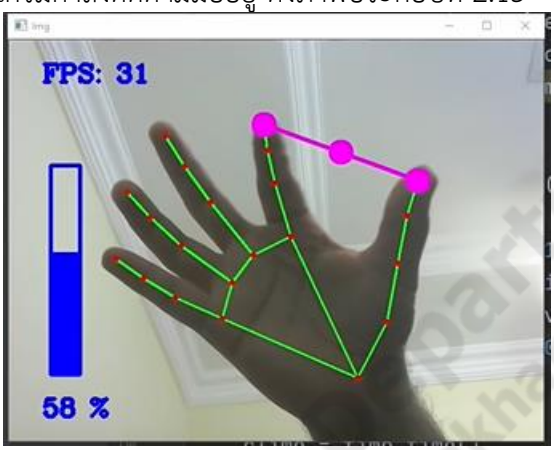


ภาพประกอบที่ 2.12 ผลลัพธ์การติดตามมือด้วยโปรแกรม Hand Tracking
ที่มา https://www.computervision.zone/courses/hand-tracking/

Computer Science Department
Faculty of Informatics, Mahasarakham University

2.3.2 Gesture Volume Control

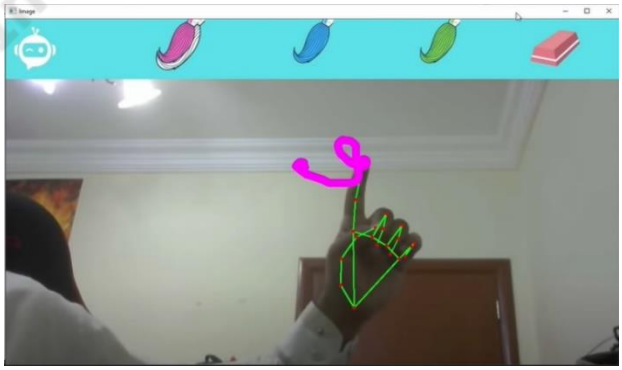
เป็นโปรแกรมที่จะคอยตรวจจับมือ จากกล้องเว็บแคมและหลังจากนั้นจะใช้ตำแหน่งที่ตั้งของนิ้วมือมาคำนวณจากท่าทาง เพื่อที่จะนำไปปรับเพิ่ม หรือลดความดังของเสียง โดยจะมีโปรแกรม Hand Tracking เป็นฐานในการพัฒนาโปรแกรมนี้ขึ้นมา ซึ่งโปรแกรมนี้ไม่สามารถยืนยันขนาดความดังของเสียงที่ต้องการได้ในขณะที่โปรแกรมกำลังติดตามมืออยู่ ดังภาพประกอบที่ 2.13



ภาพประกอบที่ 2.13 ผลลัพธ์ของโปรแกรม Gesture Volume Control
ที่มา https://www.computervision.zone/courses/gesture-volume-control/

2.3.3 AI Virtual Painter

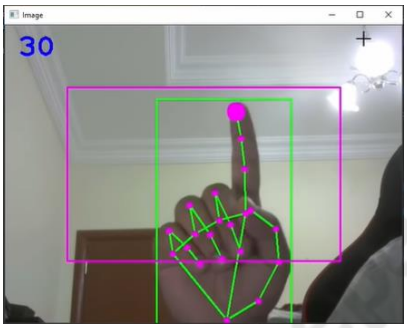
เป็นโปรแกรมวาดรูปบนอากาศ ที่จะคอยตรวจจับมือที่จะรับเข้ามาจากกล้องเว็บแคม ใช้นิ้วชี้ในการวาดรูป และใช้สองนิ้วในการเลือกสีหรือสลับโหมดระหว่างโหมดการวาดรูปกับโหมดการลบ ซึ่งสามารถทดแทนการวาดด้วยเมาส์ หรือการวาดด้วยเมาส์ปากกาได้ โดยจะมีโปรแกรม Hand Tracking เป็นฐานในการพัฒนาโปรแกรมนี้ขึ้นมา ซึ่งโปรแกรมนี้ไม่สามารถบันทึกภาพที่วาดไว้ได้ และมีสีที่จำกัด ดังภาพประกอบที่ 2.14



ภาพประกอบที่ 2.14 แสดงการวาดภาพจากโปรแกรม AI Virtual Painter
ที่มา https://www.computervision.zone/courses/ai-virtual-painter/

2.3.4 AI Virtual Mouse

เป็นโปรแกรมควบคุมตัวชี้เมาส์ ที่จะคอยตรวจจับมือที่จะรับเข้ามาจากกล้องเว็บแคม โดยการใช้นิ้วชี้เป็นจุดชี้สำหรับตัวชี้เมาส์เพื่อแสดงบนตำแหน่งต่าง ๆ บนหน้าจอ และควบคุมการคลิกซ้ายด้วยนิ้วชี้และนิ้วกลางจรดกัน กับการคลิกซ้ายด้วยเมาส์ซึ่งช่วยลดโอกาสเกิดการอักเสบของเอ็นบริเวณข้อมือบริเวณนิ้วมือ หรือพังผืดเส้นประสาทบริเวณข้อมือเมื่อมีการใช้งานเมาส์ในระยะเวลานาน ซึ่งโปรแกรมนี้ครอบคลุมเพียงการควบคุมตัวชี้เมาส์ กับการคลิกซ้ายเท่านั้น ดังภาพประกอบที่ 2.15



ภาพประกอบที่ 2.15 แสดงการควบคุมเคอร์เซอร์ด้วยโปรแกรม AI Virtual Mouse
ที่มา https://www.computervision.zone/courses/ai-virtual-mouse/

ตารางที่ 2.1 ตารางสรุปการทำงานของระบบงานที่เกี่ยวข้อง

หัวเรื่อง	รายละเอียด	หมายเหตุ
Hand Tracking 30 FPS using CPU	ตรวจจับมือที่รับเข้ามาจากกล้องเว็บแคม แล้วแสดงผลลัพธ์ออกมาเป็นโครงสร้างมือ	
Gesture Volume Control	ตรวจจับมือที่รับเข้ามาจากกล้องเว็บแคม แล้วนำมาควบคุมการเพิ่ม ลดขนาดเสียง	
AI Virtual Painter	ตรวจจับมือที่รับเข้ามาจากกล้องเว็บแคม แล้วนำมาควบคุมการวาดรูป	
AI Virtual Mouse	ตรวจจับมือที่รับเข้ามาจากกล้องเว็บแคม แล้วนำมาควบคุมการเคลื่อนไหวของตัวชี้เมาส์ กับการคลิกซ้าย	

Computer Science Department
Faculty of Informatics, Mahasarakham University