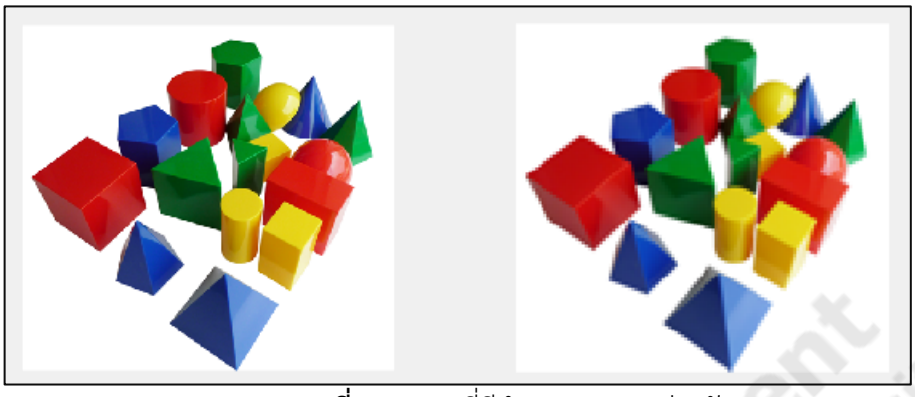
บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

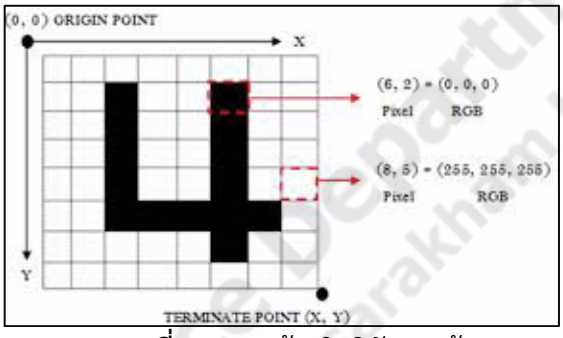
2.1.1 การประมวลผลภาพ (Image Processing)

การนำสัญญาณภาพที่ได้จากแหล่งกำเนิดสัญญาณภาพ อาจเป็นสัญญาณจากกล้องดิจิตอล (Digital camera) ข้อมูลรูปภาพ (Digital Image) หรือฟิล์ม ภาพยนตร์ต่าง ๆ นำมากระทำการปรับปรุง เปลี่ยนแปลงคุณสมบัติต่าง ๆ ของภาพหรืออาจจะนำมากระทำการคำนวณหาค่าต่างของภาพ ขึ้นอยู่กับความต้องการและจุดประสงค์ การประมวลผลภาพสามารถนำไปประยุกต์ใช้งานในสาขาต่าง ๆ มากมาย โดยเฉพาะการหาตำแหน่งวัตถุและการนับจำนวนวัตถุ เทคนิคการประมวลผลภาพเบื้องต้น จะอ้างอิงถึงแสงที่มนุษย์สามารถมองเห็นหรือรับรู้ได้โดยการที่จะระบุหรือแยกแยะข้อมูลต่าง ๆ ที่เกิดขึ้นในภาพ จำเป็นที่จะต้องทำการคัดกรองข้อมูลที่ได้จากภาพดิจิตอล (Digital Image) ทีละขั้นตอน และ เนื่องจากข้อมูลที่จะนำมาประมวลผลต้องเป็นข้อมูลภาพแบบดิจิตอลข้อมูลของภาพจะอยู่ในรูปแบบ ของจุดสีต่าง ๆ โดยเมื่อนำจุดสีต่าง ๆ มาประกอบกันอย่างต่อเนื่องก็จะทำให้เกิดภาพขึ้นมา การแปลงภาพที่มองเห็นหรือตรวจจับจากอุปกรณ์รับภาพจะถูกทำการแปลงจากข้อมูลที่ได้ ให้เป็นข้อมูลเชิงดิจิตอล โดยการสุ่มเลือกทางจุดตำแหน่ง หรือ Spatially Sampling โดยการสุ่มเลือกเฉพาะบางตำแหน่งในภาพ ซึ่งหากทำการสุ่มเลือกโดยหากมีจำนวนการสุ่ม (Sample size) ที่มากความละเอียดของภาพที่ได้รับในรูปแบบของภาพดิจิตอลก็จะมีความละเอียดสูง หน่วยของการสุ่มเลือกคือจุดภาพ (Pixel) โดยที่การที่จะระบุถึงตำแหน่งหรืออ้างอิงถึงจุดภาพ (Pixel) ในตำแหน่งต่าง ๆ จะทำการอ้างอิงผ่านระบบพิกัด โดยอ้างอิงในพิกัด ของ X และพิกัด ของ Y เพื่อการเข้าถึงค่าของจุดสี เหตุผลของการสุ่มเลือกทางจุดตำแหน่ง หรือ Spatially Sampling นั้นก็คือในการประมวลผลของคอมพิวเตอร์หรือ หน่วยประมวลผล (Processing Unit) จะประมวลผลข้อมูลได้ ข้อมูลจะต้องอยู่ในรูปของดิจิตอล (Digital) เท่านั้น ในขณะที่ภาพที่สามารถมองเห็นด้วยตาเปล่าของมนุษย์จะเป็นข้อมูลเชิง Analog เมื่อต้องกระทำการประมวลผลภาพโดยคอมพิวเตอร์หรือหน่วยประมวลผล จะต้องทำการสุ่มเลือกทางจุดตำแหน่งภาพ (Image Sampling) เพื่อแปลงข้อมูลจากรูปแบบของ Analog เป็นข้อมูลในรูปของดิจิตอล (Digital) จากเหตุผลดัง กล่าวแสดงให้เห็นว่าหากทำการสุ่มเลือกจุดตำแหน่งโดยมีปริมาณการสุ่มมากเท่าใด คุณภาพหรือรายละเอียดของภาพที่ได้รับก็จะสูงขึ้นและใช้เนื้อที่ในการเก็บข้อมูลภาพสูงขึ้น รวมถึงขนาดของภาพก็จะสูงขึ้นเช่นกัน ดังตัวอย่างที่แสดงถึงการสุ่มเลือกจุดตำแหน่งในปริมาณที่ต่างกัน โดยจะสังเกตเห็นว่าภาพประกอบที่ 2.1 มีความละเอียดของภาพสูงสุด และภาพประกอบที่ 2.1มีความละเอียดของภาพต่ำสุด

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.1 ภาพที่มีจำนวนจุดภาพต่างกัน



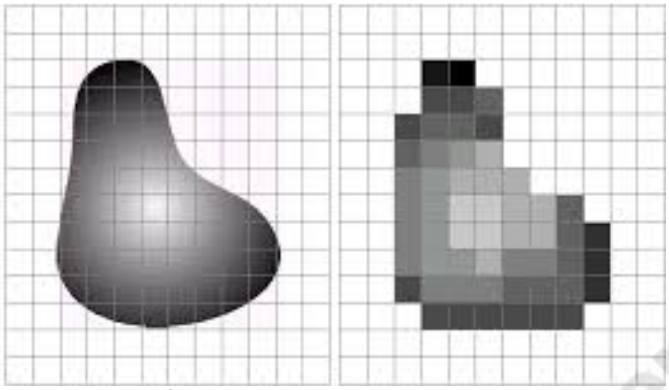
ภาพประกอบที่ 2.2 การอ้างอิงพิกัดของข้อมูลภาพ

การสุ่มเลือกทางจุดตำแหน่งในพิกัด ของ X จะหมายถึงพิกัดในแนวแกนนอน โดยเริ่มจากซ้ายที่พิกัด 0 ไปจนถึงพิกัด ที่ X และพิกัด ของ Y หมายถึงพิกัด ในแนวแกนตั้ง โดยเริ่มจากบนที่พิกัด 0 ไปจนถึงพิกัด ที่ Y ดัง ตัวอย่างภาพดิจิตอลภาพหนึ่งมีขนาด 320x270 จุดภาพ (Pixel) นน้ั คือภาพนี้มีพิกัด ซ้ายสุดที่ตำแหน่ง0 และตำแหน่งขวาสุดที่ 320 และมีพิกัด บนสุดที่ตำแหน่ง 0 และตำแหน่งล่างสุดที่ 270 เช่นตัวอย่างดัง ภาพประกอบที่ 2.2 ในการสุ่มเลือกทางจุดแต่ละจุดภาพจะถูกแทนด้วยค่าของสี สำหรับภาพสี

2.1.2 การแทนภาพด้วยภาพแบบดิจิตอล

ภาพแบบดิจิตอล (Digital Image) เป็นภาพที่ถูกแปลงมาจากแอนะล็อกให้อยู่ในรูปของตัวเลข โดยภาพแอนะล็อกถูกแบ่งเป็นพื้นที่สี่เหลี่ยมเล็ก ๆ ที่เรียกว่าพิกเซลในแต่ละพิกเซลจะถูกระบุตำแหน่งด้วยคู่โคออดิเนต (x, y) และค่าระดับความเข้มแสงของพิกเซลนั้น ๆ โดยเราสามารถแปลงภาพเป็นภาพแบบดิจิตอลได้โดยมีขั้นตอนและวิธีการดังนี้เมื่อเรานำสัญญาณแอนะล็อกที่ต้องการประมวลผลผ่านส่วนที่เรียกว่าดิจิไตเซอร์ (Digitizor) ซึ่งจะมีหน้าที่ในการเปลี่ยนสัญญาณแอนะล็อกใหเป็นสัญญาณดิจิตอล จากนั้น ทำการควอนไตซ์ (Quantizing) เพื่อแปลงค่าความเข้มแสงให้เป็นตัวเลขฟังก์ชั่นของภาพ f(x, y) จะถูกทำให้เป็นสัญญาณไม่ต่อเนื่องทั้งระนาบของภาพซึ่งเรียกว่า การสุ่มภาพ (Image Sampling) ของฟังก์ชั่นที่ได้เรียกว่าการควอนไตซ์ระดับ ความเข้ม แสง (Grey LevelQuantization)ก็จะได้ข้อมูลที่เป็นดิจิตอลดัง ภาพประกอบที่ 2.3

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.3 การทำ Sampling และ Quantization

สมมุติว่าสัญญาณภาพต่อเนื่อง f(x, y) ถูกดิจิไตซ์ในระนาบ x และ y เป็นช่วงเท่า ๆ กันเราสามารถจดั f(x, y) ให้อยู่ในรูปของเมตริกซ์ขนาด N x N ได้ดังสมการที่ 2.1

$$f(x,y) = \begin{matrix} f(0,0) & f(0,1) & f(0,2) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & f(1,2) & \dots & f(1,N-1) \\ . & . & . & & . \\ . & . & . & & . \\ f(N-1,0) & f(N-1,1) & f(N-1,2) & \dots & f(N-1, N-1) \end{matrix} \quad (2.1)$$

ซึ่งทางขวาของสมการจะเรียกว่าข้อมูลภาพดิจิตอลและทุก ๆ สมาชิกของเมตริกซ์จะ เรียกว่า พิกเซลจากขบวนการสร้างภาพดิจิตอลดังข้างต้น จะเห็นได้ว่าเราสามารถทราบขนาดของความ ละเอียดของภาพ N x N พิกเซลและจำนวนวนระดับของความเข้ม แสงในทางปฏิบัติการทำควอนไตเซชั่นในระบบภาพดิจิตอลจะมีค่าดัง สมการที่ 2.2

$$B=N\times N\times M \quad (2.2)$$

เมื่อ B=ขนาดของข้อมูลภาพที่เป็นดิจิตอล

$$G = 2M \quad (2.3)$$

โดย M = จำนวนวนบิตที่ใช้ในการแทนข้อมูลภาพ 1 พิกเซล

G = จำนวนวนความเข้ม แสงที่ต้องการใช้ในนการเก็บข้อมูลภาพ

2.1.3 ลักษณะการจัดเก็บข้อมูลภาพแบบดิจิตอล

โดยทั่วไปแล้วภาพจะมีความเข้ม ตั้งแต่ 2 ระดับ ขึ้นไปแต่ที่นิยมใช้กันมากคือค่าระดับความเข้มของพิกเซลที่เท่ากับ 256 ระดับ ซึ่งจะทำให้ค่าของพิกเซลอยู่ในช่วง (0-255) โดยใช้เนื้อที่ใน การเก็บข้อมูลขนาด 1 ไบต์หรือ 8 บิตสำหรับข้อมูล 1 พิกเซล (256) ในกรณีที่ต้องการภาพที่มีความเข้มสูงอาจจะต้องการจำนวนวนบิตสำหรับการเก็บข้อมูลมากกว่า 8 บิตคืออาจจะเป็น 16 หรือ 24 บิตโดยจะแยกความแตกต่างของภาพแต่ละประเภทให้เห็นอย่างชัดเจนได้ดังนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University

- ภาพ 2 ระดับ คือมีพิกเซลสีขาวกับ สีดำเท่านั้น โดยแต่ละพิกเซลจะมีขนาดของขอ้มูล เท่ากับ 1บิต
- ภาพ 16 ระดับ คือในแต่ละพิกเซลจะมีขนาดของข้อมูล 4 บิตซึ่งทำ ให้ สามารถแสดงได้16 ระดับ สี
- ภาพ 256 ระดับ คือในแต่ละพิกเซลจะมีขนาดของข้อมูล 8 บิตซึ่งทำ ให้ สามารถแสดง ภาพได้ความเข้ม ถึง256ระดับ
- ภาพทรูคัลเลอร์ (True Color) คือในแต่ละพิกเซลจะมีขนาดของข้อมูล 24 บิต ทำให้ สามารถแสดงภาพออกมาได้เหมือนจริงมากที่สุดเพราะสามารถแสดงสีได้ถึง 16,777,216 สีภาพ ทรูคัลเลอร์สามารถแสดงได้เฉพาะภาพสีเท่านั้น ไม่สามารถแสดงเป็นภาพขาวดำได้ ดังภาพประกอบที่ 2.4

ภาพประกอบที่ 2.4 ลักษณะของภาพ ภาพทรูคัลเลอร์

โดยทั่วไปวิธีการประมวลผลภาพเชิงตัวเลขที่ทำให้คอมพิวเตอร์สามารถรู้จักวัตถุในภาพได้นั้นแบ่งออกได้เป็นสองระดับด้วยกัน คือการประมวลผลภาพในระดับต่ำ (Low-Level Image Processing) และการประมวลผลภาพในระดับสูง (High-Level Image Processing) การประมวลผลภาพในระดับต่ำจะเป็นการประมวลผลเชิงตัวเลขเกือบทั้งหมดเพื่อหาตัวแปรต่าง ๆ มาอธิบาย ข้อมูลภาพโดยมีจุดประสงค์ที่จะนำตัวแปรเหล่านั้น ไปใช้ในการประมวลผลภาพระดับสูงต่อไป โดยทั่วไปแล้วการประมวลผลภาพระดับต่ำ จะประกอบด้วยการประมวลผลภาพก่อน (Pre Processing) เช่นการกำจัดสัญญาณรบกวนหรือการทำ ให้ภาพคมชัดการหาขอบภาพเป็นต้น การประมวลผลระดับสูงเป็นการนำผลลัพธ์หรือสัญลักษณ์ที่ได้จากการประมวลผลภาพระดับต่ำ มาตีความหรือประมวลเพื่อให้คอมพิวเตอร์สามารถรู้จักและเข้าใจภาพได้ สำหรับความแตกต่างของการประมวลผล ภาพทั้ง 2 ประเภทนั้น คือการประมวลผลภาพระดับต่ำ จะใช้ค่าความสว่างของจุดภาพ (พิกเซล) ส่วน การประมวลภาพระดับสูงนั้น ข้อมูลภาพที่นำมาประมวลผลจะถูกแสดงในรูปสัญลักษณ์ซึ่งสัญลักษณ์ เหล่านี้จะแสดงถึงสิ่งต่าง ๆ ที่อยู่ในนภาพเช่นขนาดของวัตถุรูปร่างและความสัมพันธ์กัน ระหว่างวัตถุภาพ

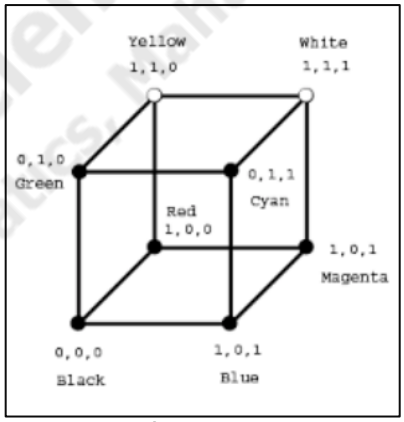
Computer Department Faculty of Informatics Rakhan University

2.1.4 แบบจำลองสี (ColorModel)

แบบจำลองสี (Color Model) เป็นสิ่งที่ใช้อ้างอิงถึงสีต่าง ๆ สำหรับคอมพิวเตอร์แล้ว เราจะไม่ใช้แบบจำลองที่เป็น Analytical Model เหมือนกับ ที่ใช้ในทางวิทยาศาสตร์ซึ่งใช้วิธีการวัดที่อยู่ในรูปของพลังงานตลอดช่วงของสเปคตรัม (Spectrum) แต่จะเป็น Empirical Model ที่ได้รับ ความสัมพันธ์ของค่าที่ใช้อ้างอิงกับสีใด ๆ จากการทดลองการศึกษาแบบ Psychophysical ที่มีการรับรู้ของมนุษย์เข้ามาเกี่ยวข้องแบบจำลองสีมีหลายแบบด้วยกันเช่นแบบจำลองสี RGB แบบจำลองสี CMY แบบจำลองสี CMYK แบบจำลองสี HSV แบบจำลองสี HIS แบบจำลองสี HLS แบบจำลองสี YIQ และแบบจำลองสี YUV (แบบจำลอง YcbCr) เป็นต้น

2.1.5 แบบจำลองสี RGB (RGB Color Model)

เป็นแบบจำลองที่เฉพาะเจาะจงกับ จอภาพคอมพิวเตอร์เนื่องจาก RGB Model ได้ทำการสร้างสีต่าง ๆ ขึ้นโดยการใช้แหล่งกำเนิดแสงจำนวนวนสามสีได้แก่ สีแดง (Red) สีเขียว (Green) และ สี น้ำเงิน (Blue) ที่เกิดจากสารเรืองแสงที่มีคุณสมบัติที่แตกต่างกัน ตามลำดับ ซึ่งแสงทั้ง สามสีจะไม่เท่ากันในแต่ละอุปกรณ์นอกเสียจากว่าจะมีคุณสมบัติของสารเรืองแสง การตั้งค่าจอภาพและ สภาพแวดล้อมที่จอภาพคอมพิวเตอร์เหมือนกัน ทุกประการ ซึ่งโดยปกติแล้ว จะมีค่าที่แตกต่างกัน ออกไปดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.5 แบบจำลองสี RGB

แบบจำลองสี RGB ประกอบด้วยข้อมูลจำนวนวนสามส่วนคือค่า Intensity ของสีทั้ง สาม ซึ่งได้แก่สีแดง สีเขียวและสีน้ำเงิน

2.1.6 แบบจำลองสี HSV (HSV Color Model)

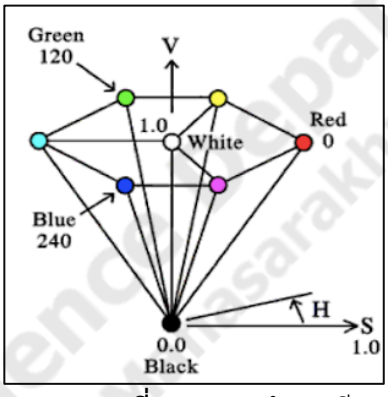
เป็นแบบจำลองสีที่ถูกสร้างขึ้นเพื่อเป็นทางเลือกดัง รูปที่ 2.6 ซึ่งจะให้ความหมายที่ ดีกว่าเมื่อกล่าวถึงสีต่าง ๆ ในเชิงศิลปะเช่นเมื่อพดูถึงสีเหลืองในทางศิลปะจะมีความแตกต่างกัน เมื่อ พิจารณาสีเหลืองอ่อนสีเหลืองแก่หรือสีน้ำ ตาลว่ามีความแตกต่างกัน อย่างไร จะพบว่าทุกสีคือสีเหลือง นั้น เองที่มีระดับ ความเข้ม หรือมีความอิ่มตัว ที่แตกต่างกัน ดังนั้นสีในแบบจำลองสี HSV จึงให้ ความรู้สึกที่เข้าใจได้มากกว่าสำหรับมนุษย์ ซึ่งจำลองสี HSV ประกอบด้วยสามส่วนคือ

Computer Science Department
Faculty of Informatics, Mahasarakham University

H หมายถึง Hue หรือสีที่มีค่าที่แตกต่างออกไปตามความถี่ของแสง เช่น แดง เหลือง เขียว น้ำ เงิน หรือม่วง เป็นต้น

S หมายถึง Saturation หรือความอิ่มตัว ของ Hue นั้น ๆ เช่น สีแดงและสีชมพู ก็คือสี แดงเพียงแต่สีชมพูมีความอิ่มตัวน้อยกว่า

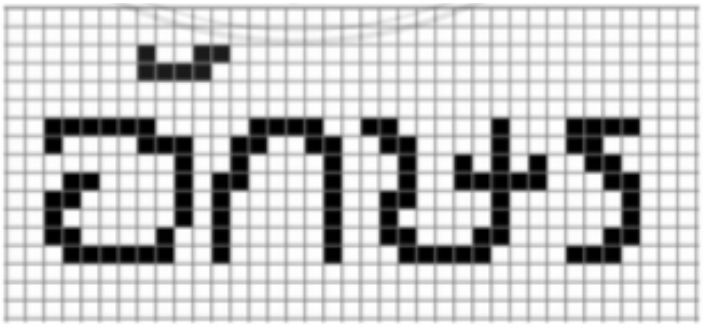
V หมายถึง Value หรือค่าความสว่างของสีโดยที่ค่า Value ต่ำ สุดหมายถึงสีดำไม่ว่า จะมี Hue หรือ Saturation เท่าใดและค่า Value สูงสุดหมายถึงสีขาวซึ่งเป็นสีที่สว่างที่สุดของ Hue และ Saturation นั้น ๆ เช่น Hue ใด ๆ มีค่า Saturation มีค่าเท่ากับ 0 เมื่อ Value สูงสุดก็คือสีขาวและ Value ต่ำสุดคือสีดำหรือHueสีเหลืองที่Saturationมีค่าเท่ากับ 100เมื่อValueสูงสุดคือสีเหลืองและ Value ต่ำ สุดคือสีดำ ดัง ภาพประกอบที่ 2.6



ภาพประกอบที่ 2.6 แบบจำลองสี HSV

2.1.7 ภาพไบนารี (Binary Image)

ภาพไบนารีก็คือภาพซึ่งในแต่ละพิกเซลจะมีความเข้ม ของแสงได้สองแบบเท่านั้น คือสว่าง กับ มืด โดยอาจเขียนแทนได้ด้วยเลข 0 และ 1 ด้วยความเข้ม ของแสงเพียงสองระดับ ภาพไบนารีจึงมี ข้อจำกัดที่จะนำมาใช้แสดงภาพทั่ว ๆ ไป แต่อย่างไรก็ตามการมีความเข้ม ของแสงเพียงสองระดับ ก็ทำ ให้การประมวลผลทำ ได้อย่างมีประสิทธิภาพ การประมวลผลภาพไบนารีนำไปใช้มากในการ ประมวลผลเอกสาร การประมวลผลภาพในอุตสาหกรรมที่ต้องการความเร็วสูง เช่น การนับจำนวนวน ชิ้นส่วนที่อยู่บนสายพานที่กำลังเคลื่อนที่เป็นตอนการมีความเข้ม เพียงสองระดับ ในภาพไบนารีทำ ให้ สามารถที่จะเลือกพิจารณาให้ความเข้ม ระดับหนึ่งแทนภาพของสิ่งที่เราสนใจ โดยเราจะขอเรียก พิกเซลที่มีความเข้มระดับนี้ว่าพิกเซลภาพและความเข้ม อีกระดับ แทนพื้นหลังโดยพิกเซลที่มีความเข้มระดับนี้จะถูกเรียกว่าพิกเซลพื้นหลังในทางปฏิบัติในการพิจารณาภาพไบนารีเราจะไม่สนใจแต่ละพิกเซลแยกกัน ไปแต่เราจะสนใจกลุ่มของพิกเซลที่อยู่ติดกัน หรือที่เรียกว่าพิกเซลเพื่อนบ้าน เช่น กลุ่ม ของพิกเซลที่เรียงกันเป็นตัว อักษรดัง ภาพประกอบที่ 2.7



ภาพประกอบที่ 2.7 ลักษณะภาพไบนารีที่แสดงกลุ่มของพิกเซลภาพที่ประกอบเป็นตัว อักษร

ภาพการสร้างภาพไบนารี่สามารถทำ ได้โดยใช้เทคนิคการทำเทรชโฮล (Thresholding Technique)โดยพิจารณาว่าพิกเซลใดเป็นสีขาวหรือสีดำ จะกระทำโดยการเปรียบเทียบระหว่าง พิกเซล ของภาพเริ่มต้นกับค่าคงที่ค่าหนึ่งที่เรียกว่า “ค่าเทรชโฮล” (Threshold Value) ข้อมูลภาพมีลักษณะที่ ต่างกัน ระหว่างวัตถุ(Object) และพื้นหลัง (Background) โดยค่าของพิกเซลของภาพใด ๆ ที่มีค่า มากกว่าหรือเท่ากับค่าเทรชโฮลจะถูกเปลี่ยนเป็น 0 (จุดดำ) ในการสร้างภาพไบนารี่โดยใช้เทคนิค เทรชโฮลเพื่อให้ได้ผลลัพธ์ที่เหมาะสมและคมชัด สิ่งที่สำคัญที่สุดคือค่าเทรชโฮลเนื่องจากถ้าเลือกค่า เทรชโฮลที่ไม่เหมาะสม (ค่าเทรชโฮลที่มีค่าน้อยเกินไปหรือมากเกินไป) ภาพที่ได้อาจจะมืดเกินไป หรือสว่างมากเกินไปหรือภาพที่ได้มีสิ่งรบกวน (Noise) เกิดขึ้นอันเป็นผลทำทำให้ภาพไม่สวยงามเท่าที่ควรดังนั้น ปัญหาของการสร้างภาพไบนารี่คือวิธีการกำหนดค่าเทรชโฮลที่เหมาะสมสำหรับแต่ละภาพที่จะนำมาทำ การสร้างภาพไบนารี่ซึ่งมีวิธีการคำนวณหาค่าเทรชโฮลหลายวิธีโดยแต่ละวิธี เหมาะสมกับ ลักษณะการทำงานที่แตกต่างกัน ไปเช่นการหาค่าเทรชโฮลโดยกำหนดค่าล่วงหน้า (Pre- Assigned Threshold Value) การหาค่าเทรชโฮลจากค่ากลาง (Mid-Range Threshold Value) ซึ่งแต่ละวิธีอธิบายได้ดังนี้

การหาค่าเทรชโฮลโดยการกำหนดค่าล่วงหน้าเป็นการกำหนดค่าเทรชโฮลโดยการกำหนด เองจากผู้ใช้ซึ่งการกำหนดนี้จะขึ้นอยู่กับประสบการณ์ของผู้ใช้คนนั้น ๆ โดยการเลือกค่าคงที่ค่าหนึ่ง ซึ่งเรียกว่าค่าเทรชโฮล โดยค่าที่เลือกมานี้จะเป็นค่าที่อยู่ระหว่างค่าต่ำ สุดและค่าสูงสุดของระดับ ความเข้ม แสงของภาพ เช่น ภาพอินพุทมีระดับ ความเข้ม แสง 256 ระดับ จะมีค่าได้ตั้งแต่ 0-255 เมื่อเลือกค่า เทรชโฮลได้แล้วสามารถสร้างภาพไบนารี่ได้

การหาค่าเทรชโฮลจากค่ากลางเป็นการหาเทรชโฮลที่แตกต่างจาการหาค่าเทรชโฮลวิธีแรก สำหรับวิธีนี้จะเป็นการคำนวณหาค่าเทรชโฮลโดยอัตโนมัติโดยไม่ต้องให้ผู้ใช้เป็นผู้กำหนด โดยการหาค่าเทรชโฮลนี้ใช้วิธีทางสถิตในเรื่องการหาค่ากลางหรือค่าเฉลี่ย (Mean) ค่าเทรชโฮลที่คำนวณได้จากค่ากึ่งกลางที่อยู่ระหว่างค่าระดับ ความเข้ม สูงสุด (Maximum Level) และระดับ ความเข้ม ต่ำสุด (Minimum Level) ของภาพเมื่อทำ การคำนวณค่าเทรชโฮลได้แล้วก็สามารถสร้างภาพไบนารี่ได้โดยนำค่าเทรชโฮลที่ได้มาใช้

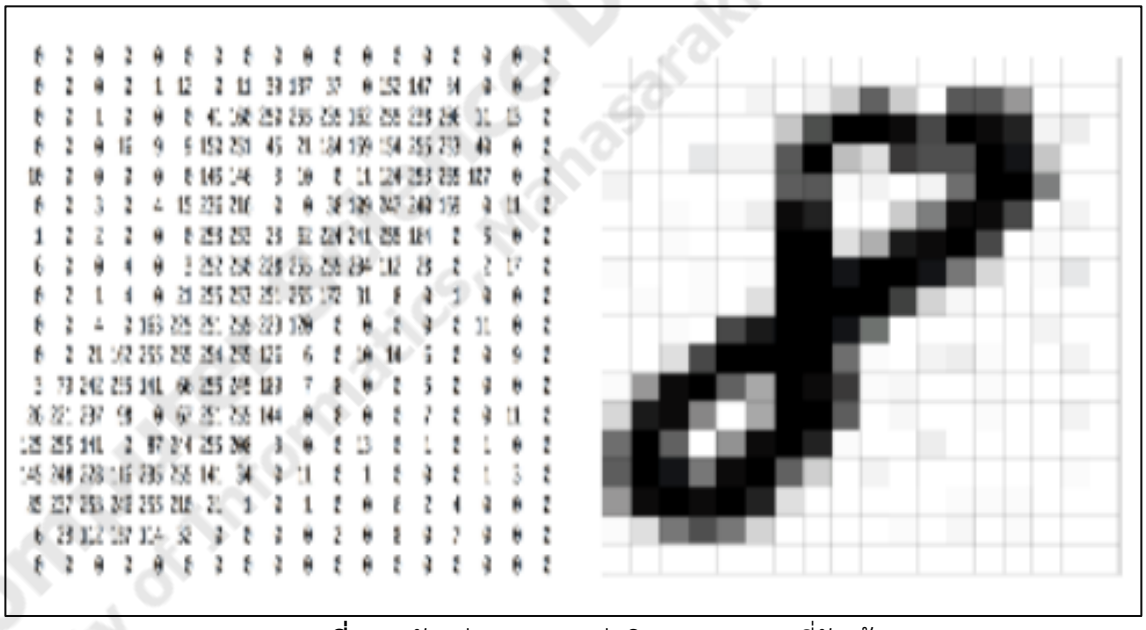
Computer Science Department
Faculty of Informatics Mahasarakham University

2.1.8 การเรียนรู้เชิงลึก (Deep Learning)

โดยอัลกอริทึมที่เลือกใช้คือ Convolutional Neural Network (CNN) เนื่องจากอัลกอริทึมที่มีประสิทธิภาพใกล้เคียงกับมนุษย์

Convolutional Neural Network เป็นโครงข่ายประสาทที่มีไอเดียในการแก้ปัญหาทางด้านรูปภาพ ซึ่งในเวลานั้นถือเป็นการปฏิวัติวงการ Image Recognition การทำงานแบบคร่าวๆ มีรูปมาใช้ นำรูปมาเข้ากระบวนการ Convolution เพื่อเพิ่มรายละเอียด เพิ่ม Noise ให้กับรูป Input จากนั้นก็ใส่ Polling เข้าไป (มีหลายแบบ Average, Max)

ความรู้เกี่ยวกับการเรียนรู้เชิงลึก (Deep Learning) เป็นแนวคิดในการให้คอมพิวเตอร์เรียนรู้เข้าใจข้อมูลที่ได้รับ โดยมีสถาปัตยกรรมการเรียนรู้ข้อมูลของคอมพิวเตอร์มากมายโดยจะกล่าวถึงโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network ; CNN) ที่นำมาใช้เพื่อการสกัดเอาลักษณะเด่น จนกระทั่งทราบถึงผลลัพธ์หรือวัตถุที่ตรวจจับ การทำงานของโครงข่ายประสาทเทียมแบบสังวัตนาการมี 4 กระบวนด้วยกัน

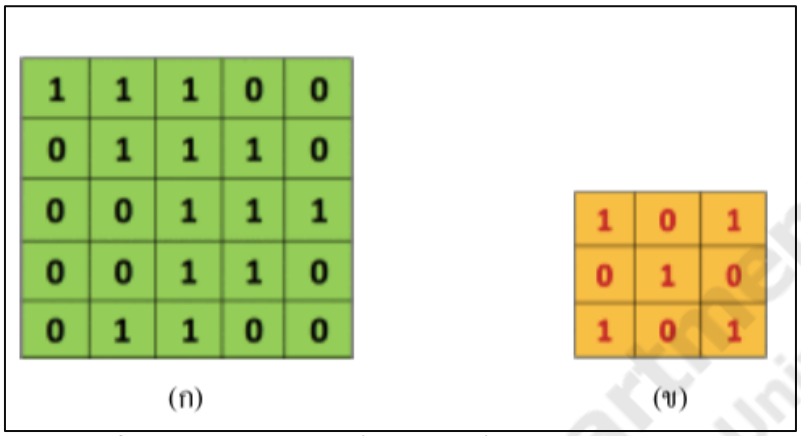


ภาพประกอบที่ 2.8 ตัวอย่างการแทนค่าพิกเซลลงบนรูปที่รับเข้ามา

2.1.9 คอนโวลูชั่น (Convolution)

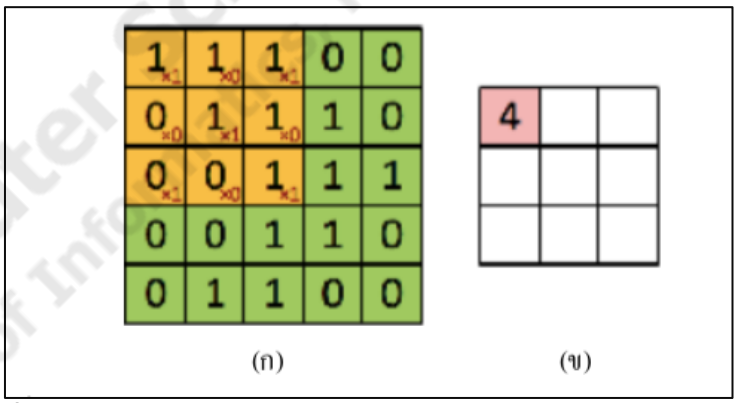
คอนโวลูชั้นเป็นกระบวนการที่ทาเพื่อสกัด เอาลักษณะสำคัญ จากภาพออกมา โดยการใช้ค่า พิกเซล โดยค่าพิกเซลได้มาจากการมองของกล้องทั่ว ไปนั้นมีด้วยกันสามแชนแนล (Channel) แบ่งเป็นสี ได้แก่ สีแดง, น้ำเงิน และเขียว โดยแต่ละจุดสามารถแทนค่าด้วยตัว เลขเพื่อบอกความเขม้ ของสีนั้น ๆ โดยมีค่าตั้งแต่ 0 ถึง 255 จากความเข้มน้อยไปหามากโดยในการทาภาพขาวดาแชนแนล ของภาพนั้นจะมีเพียงหนึ่งแชนแนลเท่านั้นคือแชนแนลของสีดาซึ่งค่าตัว เลข 0 นั้น คือสีขาวไล่ไป จนถึง 255 ซึ่งเป็นสีดา

สนิทตามภาพประกอบที่ 2.8 ในแต่ละชุดทาการคำนวณแล้ว เก็บไว้ตัว อย่างการทำงาน ของคอนโวลูชั่น จะมีดังภาพประกอบที่ 2.9 (ก)



ภาพประกอบที่ 2.9 จำลองเมทริกซ์ที่ได้จากรูปที่รับเข้ามาและเมทริกซ์ตัวกรองค่า

ในแต่ละรูปจะมีชุดค่าเมทริกซ์ที่ต่างกัน จากรูปที่ 2.9 (ก) เป็นรูปภาพขนาด 5x5 พิกเซลที่ เป็นภาพขาวดำโดยที่นี้กำหนดค่าให้ค่า 0 พิกเซลเป็นสีขาวและ 1 พิกเซลเป็นสีดำและมีการกำหนด เมทริกซ์อีกหนึ่งชุดขึ้นมา โดยเราจะใช้เมทริกซ์ชุดนี้เป็นตัว กรองค่าไปเก็บไว้ในเมทริกซ์ชุดที่เล็กกว่า เราเรียกเมทริกซ์ชุดนี้ว่าตัว กรองค่า (Filter), เคอเนล (Kernel) หรือตัว ตรวจจับ ลักษณะสำคัญ (FeatureDetector) ดังที่ปรากฏในภาพประกอบที่ 2.9 (ข)



ภาพประกอบที่ 2.10 การทำงานของตัวกรองค่าและการกำเนิดเมทริกซ์ชุดใหม่หรือฟีเจอร์แมพ

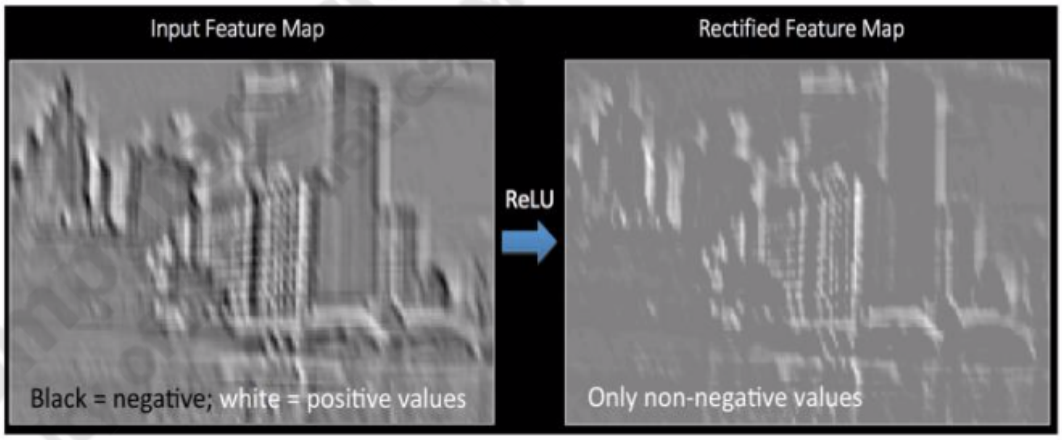
เมทริกซ์ที่ทำหน้าที่เป็นตัวกรองค่าจะเคลื่อนไปทั่ว ภาพและคูณค่าเก็บไว้ในเมทริกซ์ชุด ใหม่ดังภาพประกอบที่ 2.10 (ก) ซึ่งเราเรียกเมทริกซ์ชุดใหม่นี้ว่า คอนโวลฟีเจอร์ (Convolved Feature) หรือฟีเจอร์แมพ (FeatureMap) ดังที่ปรากฏในภาพประกอบที่ 2.10 (ข) โดยผลลัพธ์ที่ได้เมื่อรูปผ่านการทำคอนโวลูชั่น มีตัวอย่างดังภาพประกอบที่ 2.11

Computer Science Department
Faculty of Informatics
Mahasarakham University

ภาพประกอบที่ 2.11 ภาพขาวดำดั้งเดิมเมื่อผ่านการทำคอนโวลูชั่นกลายเป็นฟีเจอร์แมพ

2.1.10 การขจัดความเป็นเชิงเส้น (ReLU)

หลังจากได้ทำการคอนโวลูชั่นรูปภาพและได้ฟีเจอร์แมพมาแล้ว เราจะนาปรับแต่งให้ ฟีเจอร์แมพไม่เป็นลักษณะเชิงเส้น ด้วยวิธีการ ReLU ทำการแทนที่ผลของพิกเซลที่มีค่าเป็นเชิงลบในฟีเจอร์แมพด้วยค่า 0 จุดประสงค์ของการทำ ReLU นั้นเพื่อให้โครงข่ายประสาทเทียมแบบสังวัตนาการ เรียนรู้ข้อมูลที่ไม่เป็นเชิงเส้นจากภาพ เมื่อนำภาพเข้า สู่กระบวนการที่เป็นฟีเจอร์แมพเข้ามาทำ ReLU โดยสีดำในภาพเป็นค่าเชิงลบและสีขาวในภาพเป็นค่าเชิงบวกเมื่อทำการ ReLU ค่าที่ได้ จะมีเพียงค่าที่เป็นเชิงบวกเท่านั้น ดังภาพประกอบที่ 2.12

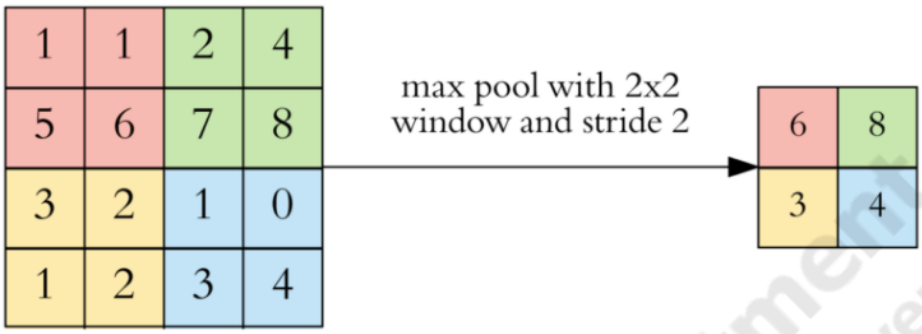


ภาพประกอบที่ 2.12 ตัวอย่างผลลัพธ์หลังการทำ ReLU

2.1.11 การพูลลิ่ง (Pooling)

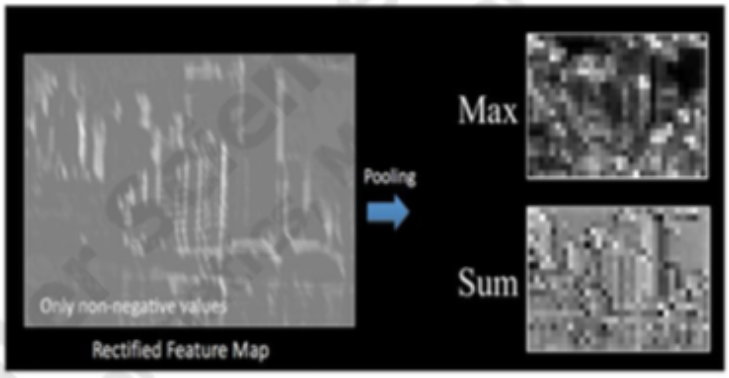
Pooling คือความสามารถในการย่อรูปแบบหนึ่ง ซึ่งมีสองประเภทหลักที่นิยมกันคือ max pooling และ mean pooling Max Pooling เป็นตัวกรองแบบหนึ่งที่หาค่าสูงสุดในบริเวณที่ตัวกรองทาบอยู่มาเป็นผลลัพธ์ โดยเราจะเตรียมตัวกรองในลักษณะเดียวกับการทำ Feature Extraction ของ CNN มาทาบบนข้อมูลแล้วเลือกค่าที่สูง ที่สุดบนตัวกรองนั้นมาเป็นผลลัพธ์ใหม่ และจะเลื่อนตัวกรองไปตาม

Stride ที่กำหนดไว้ โดยขนาดตัวกรองของการทำ max pooling จะนิยมเรียกกันว่า pool size ดังภาพประกอบที่ 2– 13



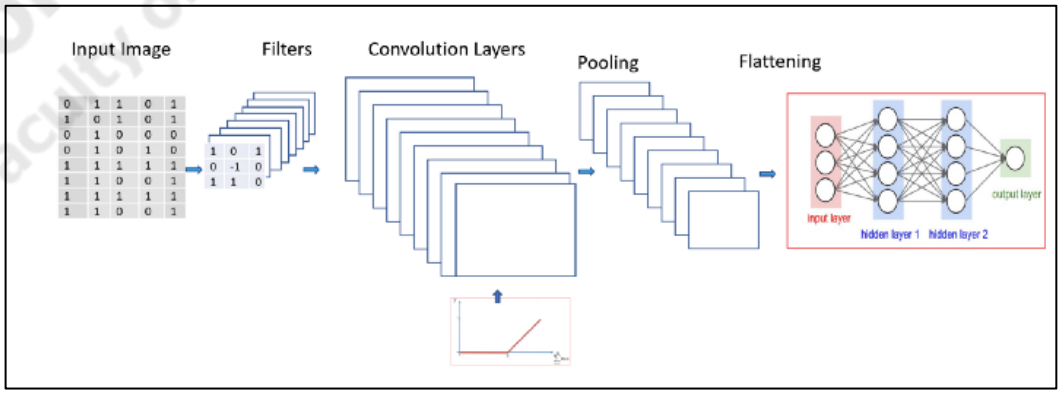
ภาพประกอบที่ 2.13 การพูลลิ่งค่ามากที่สุด

นอกจากนี้ การพูลลิ่งทาตามจำนวนแชนแนลของภาพซึ่งในกรณีภาพสีเราจะได้รับผลลัพธ์ออกมาถึงสามผลลัพธ์ เมื่อฟีเจอร์แมพที่ผ่านการทำ ReLU เข้ามาทำการพูลลิ่งจะมีผลลัพธ์ ดังภาพประกอบที่ 2.14



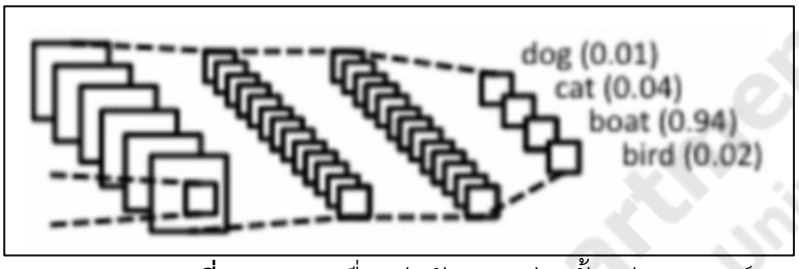
ภาพประกอบที่ 2.14 ภาพผลลัพธ์ที่ได้หลังการทำพูลลิ่ง

การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer)



ภาพประกอบที่ 2.15 ภาพผังการทำงานของระบบโครงข่ายประสาทเทียมแบบสังวัตนาการ

กระบวนการคอโวลูชั่น , ReLU และการพูลลิ่งกระบวนการทั้งสามกระบวนการต้องมีการทำซ้ำจนกว่าจะมีการเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer) ผลลัพธ์จากคอนโวลูชั่นและ พูลลิ่ง นั้นให้ลักษณะเด่นในระดับสูง (High-level features) ของรูปที่รับเข้ามา จุดประสงค์ของการทำให้เชื่อมต่อกันแต่ละชั้นโดยสมบูรณ์นั้นเพื่อนำลักษณะเด่นไปทำการคัดกรองรูปที่รับเข้า มาให้อยู่ในรูปของคลาส (Classes)



ภาพประกอบที่ 2.16 การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์

โดยผลลัพธ์ที่ได้จะแสดงความค่าความมั่นใจ (Confident) ออกมา ตัวอย่างของการคัดกรองที่มีการเทรนข้อมูลไว้สี่ประเภท เมื่อรูปถูกนำเข้าสู่กระบวนการทั้งหมดจะแสดงค่าดังภาพประกอบที่ 2.16

2.1.12 แคปท์ช่า

แคปท์ช่าเป็นเครื่องมือบนเว็บไซต์ที่ใช้แยกแยะผู้ใช้งานในระบบออนไลน์ว่าเป็นมนุษย์หรือเป็นโปรแกรม อัตโนมัติที่เรียกว่าบอท ซึ่งสามารถแฝงตัวในระบบเป็นเหมือนผู้ใช้ทั่วไป เพื่อรบกวนระบบออนไลน์นั้นทั้งทางตรง และทางอ้อมนับตั้งแต่เริ่มพบบอทในระบบออนไลน์ ก็เริ่มมีผู้คิดค้นเครื่องมือป้องกันระบบจากบอทโดยต้องการให้ ผู้ใช้ที่เป็นมนุษย์เข้าไปใช้งานในระบบได้เท่านั้น หนึ่งในเครื่องมือนั้นคือแคปท์ช่า แต่องค์ความรู้ด้านต่าง ๆ ที่ทำให้ คอมพิวเตอร์มีความฉลาดได้ถูกพัฒนาขึ้นไปมากขึ้นทุกวันตามความเจริญก้าวหน้าทางเทคโนโลยี ส่งผลให้บอทมี ความฉลาดมากขึ้นจนสามารถผ่านระบบแคปท์ช่าได้สำเร็จ จึงจำเป็นต้องมีผู้พัฒนาระบบแคปท์ช่า โดยอาศัยจุดอ่อน ด้านความสามารถของคอมพิวเตอร์ที่ไม่สามารถทำได้เหมือนมนุษย์หรือดีเทียบเท่ามาสร้างเพื่อป้องกันบอท เช่นเดียวกัน ดังนั้นความสามารถของบอทและความซับซ้อนของแคปท์ช่าจึงพัฒนาขึ้นตามกันไปตลอด มีนักวิจัยและผู้พัฒนาระบบได้วิจัยและพัฒนาแคปท์ช่ารูปแบบต่าง ๆ จำนวนมาก ในงานวิจัยนี้จึงได้รวบรวมงานวิจัยที่เกี่ยวข้อง ด้านแคปท์ช่าที่สำคัญไว้ และนำมาวิเคราะห์ถึงจุดแข็งและจุดอ่อนของแคปท์ช่ารูปแบบหลัก พร้อมทั้งแนวโน้มของ แคปท์ช่าที่มีประสิทธิภาพในอนาคตว่าควรเป็นไปในทิศทางใด

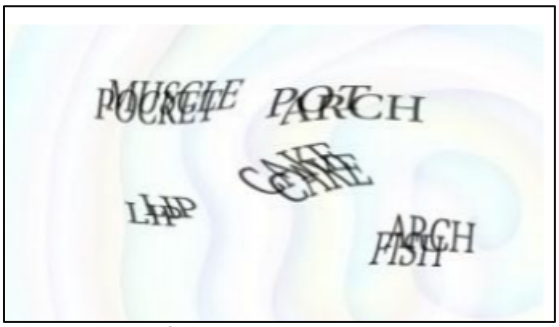
ประเภทของ แคปท์ช่า

แบ่งเป็น 3 ประเภท ดังนี้

2.1.12.1 แบบอักษรข้อความ (text-based CAPTCHA)

แคปท์ช่ารูปแบบดั้งเดิมนับตั้งแต่แคปท์ช่าตัวแรกถูกสร้างขึ้นมา และยังคงใช้กันอย่างแพร่หลายในระบบออนไลน์ในปัจจุบัน รูปแบบอักษรข้อความแบบดั้งเดิมชื่อ Gimpy ดังรูปที่ 2.17 ที่แสดงอักษรข้อความเป็นศัพท์แบบสุ่ม 7 คำแล้วให้ผู้ใช้ใส่เพียง 3 คำเท่านั้น โดยคำที่แสดงถูกทำให้บิดเบี้ยวและ

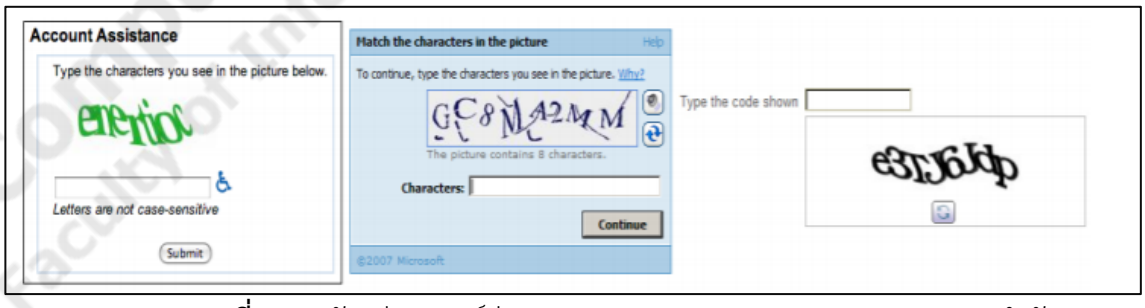
ซ้อนทับกัน แต่เนื่องจากความยากในการจำแนกของผู้ใช้จึง พัฒนาต่อมาเป็น EZ-Gimpy ที่แสดงคำมาเพียงคำเดียวดังภาพประกอบที่ 2.18



ภาพประกอบที่ 2.17 ตัวอย่าง Gimpy CAPTCHA

ภาพประกอบที่ 2.18 EZ-Gimpy CAPTCHA

แต่เนื่องด้วยการพัฒนาความสามารถที่ทำให้คอมพิวเตอร์สามารถรู้และจำแนกอักษรข้อความที่เป็นภาพให้เป็นอักษรได้ (optical character recognition) ทำให้บอทสามารถอ่านอักษรข้อความในแคปท์ช่าให้เป็นตัวอักษรได้เช่นกัน จึงทำให้แคปท์ช่าแบบอักษรข้อความต้องทำข้อความให้บิดเบี้ยวมากขึ้น ใส่สีพื้นหลัง หรือสุ่มคำที่ไม่มีความหมาย เพื่อให้บอทจำแนกได้ยากขึ้น แคปท์ช่ารูปแบบนี้ในปัจจุบันยังคงใช้ หลักการนี้อยู่ เช่น Google, Microsoft หรือ Yahoo ดังภาพประกอบที่ 2.19



ภาพประกอบที่ 2.19 ตัวอย่างแคปท์ช่าของ Google, Microsoft และ Yahoo ตามลำดับ

สิ่งที่น่าสนใจสำหรับแคปท์ช่าแบบอักษรข้อความที่ดีในปัจจุบันคือ reCAPTCHA ที่พัฒนาโดย Ahn และคณะ ดังรูปที่ 4 เป็นแคปท์ช่าที่นิยมใช้แพร่หลายในระบบออนไลน์ โดยแคปท์ช่านี้แสดงข้อความมาสองคำเป็นคำที่มีความหมายที่สุ่มมาจากข้อความในหนังสือหรือนวนิยายที่ทำให้ผู้ใช้สามารถเดาถึงความหมายและการสะกดของคำที่ปรากฏได้ และอีกคำเป็นอักษรสุ่มที่ไม่มีความหมายหรือ reCAPTCHA ทำการเปลี่ยนตัวอักษรในข้อความให้ผิดเป็นการป้องกันบอทคาดเดาคำ จากนั้นจะทำให้

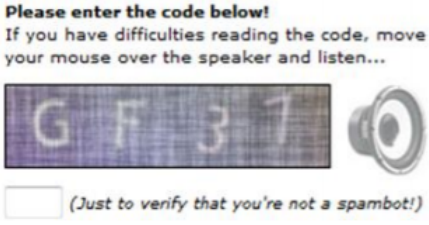
ข้อความทั้งสองบิดเบี้ยวและไม่ชัดเจน ผู้ใช้จะต้องพิมพ์คำที่ปรากฏ ทั้งสองคำให้ถูกต้อง นอกจากนี้ reCAPTCHA ยังสามารถช่วยผู้ใช้ที่พิมพ์ผิดเล็กน้อยให้กลายเป็นถูกต้องได้ และเหมาะสำหรับผู้มีปัญหาทางสายตา โดยมีปุ่มฟังเสียงคำที่ปรากฏ รวมไปถึงการสร้างข้อความที่ไม่สามารถพิมพ์ได้บนแป้นพิมพ์ปกติ เช่น ภาษาต่าง ประเทศ หรือสัญลักษณ์ทางคณิตศาสตร์ที่ไม่ปรากฏบนแป้นพิมพ์ โดยผู้ใช้ต้องไม่พิมพ์ข้อความนั้น ในขณะที่บอทจะพยายามจำแนกเพื่อให้ได้คำตอบ

ภาพประกอบที่ 2.20 ตัวอย่าง Fedora CAPTCHA

แคปท์ช่าแบบอักษรข้อความนอกจากจะแสดงอักษรบิดเบี้ยวให้ผู้ใช้พิมพ์ ยังสามารถพลิกแพลงให้ผู้ใช้ทำอย่างอื่นได้นอกจากพิมพ์คำที่เห็น เช่น Fedora ที่แสดงตัวเลขบิดเบี้ยวสองตัวเลข เครื่องหมายบวก เครื่องหมายเท่ากับ และเส้นจำนวนหนึ่งที่เป็นสิ่งรบกวนลงไปในภาพ โดยสุ่มตำแหน่งและรูปร่างของเส้น และให้ผู้ใช้พิจารณาแล้วทำการบวกเลขทั้งสองตัว ใส่ผลลัพธ์ที่ถูกต้องดังรูปที่ 2.15 พัฒนาโดย Fedora Security System หรือแนวคิดในการทำแคปท์ช่าแบบอักษรที่ใช้เพียงตัวเลขของ PaPPy ที่ชื่อ Clickable CAPTCHA เป็นการแสดงกลุ่มตัวเลขให้ผู้ใช้กดเลือกตัวเลขที่ตรงกับคำถามแทนที่การพิมพ์คำตอบดังภาพประกอบที่ 2.20

2.1.12.2 แบบเสียง (audio-based CAPTCHA)

ถูกสร้างขึ้นมาเพื่อรองรับผู้ใช้ที่มีปัญหาทางสายตา โดยใช้เสียงเป็นหลักให้ผู้ใช้ฟังในขณะที่ผู้ใช้สายตาปกติสามารถทำการทดสอบได้เช่นกัน แคปท์ช่าประเภทนี้โดยมากเป็นการทดสอบให้ผู้ใช้พิมพ์ตามคำที่ได้ยินโดยการสะกดทีละตัวอักษร ซึ่งจะอยู่รวมกับรูปแบบอักษร ข้อความ เช่น reCAPTCHA หรือ audio and visual CAPTCHA ที่ผู้ใช้สามารถกดไอคอนลำโพงเพื่อฟังเสียงการสะกดทีละอักษรตามภาพที่ปรากฏดังภาพประกอบที่ 2.21



ภาพประกอบที่ 2.21 audio and visual CAPTCHA

แต่องค์ความรู้ด้านการรับรู้และจำแนกเสียงเป็นข้อความได้พัฒนาให้คอมพิวเตอร์สามารถแปลงเสียงเป็นข้อความได้ แคปท์ช่ารูปแบบเสียงจึงมีการใส่เสียงรบกวนลงไปเพื่อให้บอทจำแนกเสียงยาก ขึ้น โดยเสียงนั้นอาจเป็นเสียงที่ไม่เกี่ยวข้อง หรือเสียงดนตรีประกอบพื้นหลัง จนกระทั่งทำโทนเสียงให้เพี้ยนไปจากปกติ เช่น reCAPTCHA หรือ SimpleCAPTCHA หรือ SimpleCAPTCHA แคปท์ช่ารูปแบบนี้ยัง

มีความหลากหลายด้านการใช้เสียงให้ผู้ใช้ได้ฟัง เช่น ใช้เสียงเครื่องดนตรี เสียงยานพาหนะ เสียงจากธรรมชาติ หรือเสียงสัตว์ เพื่อมีความเป็นสากลกับผู้ใช้ที่ไม่ใช้ภาษาอังกฤษเป็นหลักให้ผู้ใช้ฟังแล้วพิมพ์ชื่อวัตถุของสิ่งที่ได้ยิน อีกทั้งบอทไม่สามารถแปลงเสียงเหล่านี้ให้เป็นข้อความได้

แคปท์ช่ารูปแบบเสียงที่น่าสนใจอีกชนิดหนึ่งที่ ที่มีหลักการใช้โปรแกรมแปลงข้อความให้เป็นเสียงและออกเสียงจากข้อความนั้นให้ผู้ใช้ฟังโดยนำข้อความจากในหนังสือเป็นฐานข้อมูล แคปท์ช่านี้มีความต่างจากแคปท์ช่ารูปแบบเสียงทั่วไปคือ ผู้ใช้ต้องออกเสียงตามที่ได้ยินให้ถูกต้อง โดยข้อความมีประโยคเดียวความยาวประมาณ 8-20 คำ เป็นภาษาอังกฤษระบบจะบันทึกเสียงของผู้ใช้มาเปรียบเทียบกับเสียงที่ได้จากการแปลงข้อความเป็นเสียง หากมีความตรงกันตามเงื่อนไขในระบบจะถือว่าทดสอบผ่าน ซึ่งน่าสนใจที่ผู้ใช้ไม่ต้องพิมพ์ข้อความเพียงแค่ออกเสียงผ่านอุปกรณ์รับเสียง ดังภาพประกอบที่ 2.22

ภาพประกอบที่ 2.22 แคปท์ช่ารูปแบบเสียงที่ให้ผู้ใช้ออกเสียงตามประโยคที่ได้ยิน

2.1.12.3 แบบภาพ (image-based CAPTCHA)

จากความสามารถของบอทที่ถูกพัฒนาให้จำแนกข้อความในภาพแปลงเป็นอักษรได้สำเร็จ ทำให้เริ่มมีการคิดค้นแคปท์ช่าแบบรูปภาพอย่างเดียวขึ้นมา โดยองค์ความรู้ด้านการจำแนกภาพของบอทยังถูกพัฒนายังไม่สูงมากนัก ลักษณะของรูปแบบนี้เริ่มแรกเป็นตัวเลือกภาพให้ผู้ใช้เลือกภาพตามคำสั่งหรือเลือก ภาพที่ต่างจากพวก เมื่อเลือกภาพที่ถูกต้องจะผ่านการทดสอบ โดยผู้ใช้ไม่ต้องพิมพ์และไม่มีการบิดเบือนรูปภาพแต่อย่างใด เช่น ESP-PIX CAPTCHA สุ่มรูปภาพจากฐานข้อมูลที่เป็นภาพที่มีเนื้อหาเหมือนกัน ขึ้นมาสี่ภาพ ผู้ใช้ต้องพิจารณาหาความสัมพันธ์ที่เหมือนกันของภาพทั้งสี่แล้วเลือกคำตอบเป็นวัตถุที่เกี่ยวข้องที่มีอยู่ในภาพทั้งสี่ ตัวอย่างในภาพประกอบที่ 2.23 คำตอบ คือ Volcano

ภาพประกอบที่ 2.23 ESP-PIX CAPTCHA

แต่คำตอบที่เป็นตัวเลือกทำให้บอทสามารถเดาคำตอบได้โดยการสุ่มเลือก มีความน่าจะเป็นที่สุ่ม ตัวเลือกตอบถูกสูง จึงมีการพัฒนาแคปท์ช่าให้ตัวเลือกคำตอบที่ถูกต้องมีหลายตัวไม่ใช่เพียงตัวเดียว และผู้ใช้ ต้องเลือกคำตอบที่ถูกต้องให้ครบจึงผ่านการทดสอบ เช่น Asirra ของ Microsoft ที่มีภาพสุนัขและแมวปนกันรวมทั้งหมด 12 ภาพ แล้วให้ผู้ใช้เลือกภาพแมวทั้งหมด 6 ภาพ ดังภาพประกอบที่ 2.24 ภาพถ่ายสุนัขและแมว ทั้งหมดมาจากฐานข้อมูลของเว็บไซต์ Petfinder.com ที่สนับสนุนให้รับสัตว์เลี้ยงที่ไม่มีเจ้าของ ผู้ใช้ที่ทำการทดสอบสามารถเลื่อนเมาส์ชี้ตำแหน่งที่ภาพเพื่อขยายภาพให้ใหญ่ขึ้นได้ การให้คำตอบมีมากกว่าหนึ่งคำตอบและต้องเลือกทุกคำตอบช่วยลดโอกาสที่บอทคาดเดาคำตอบได้ อีกทั้งการจำแนกสุนัขและแมวไม่ยากสำหรับมนุษย์

เนื่องจากรูปภาพสามารถนำมาพลิกแพลงได้หลากหลายกว่าอักษรข้อความหรือเสียง ทำให้มีผู้นำแคปท์ช่าแบบรูปภาพสร้าง แคปท์ช่าที่มีลักษณะแปลกซับซ้อน และหลากหลายมากขึ้นสุดแล้วแต่จินตนาการของผู้พัฒนา หรือนำแคปท์ช่ารูปแบบต่าง ๆ มาผสมกัน ทั้งอักษร ภาพ และเสียง ตัวอย่าง เช่น Four-panel cartoon CAPTCHA แสดงรูปภาพการ์ตูนสี่ช่องที่มีคำสนทนาของตัวการ์ตูน โดยที่รูปภาพทั้งสี่ถูกสลับที่แบบสุ่มให้ผู้ใช้พิจารณาทำการเรียงลำดับเหตุการณ์ให้ถูกต้องจึงผ่านการทดสอบ ดังภาพประกอบที่ 2.25

Computer Science Department
Faculty of Informatics, Mahasarakham University

ภาพประกอบที่ 2.24 Asirra CAPTCHA

ภาพประกอบที่ 2.25 Four-panel catoon CAPTCHA

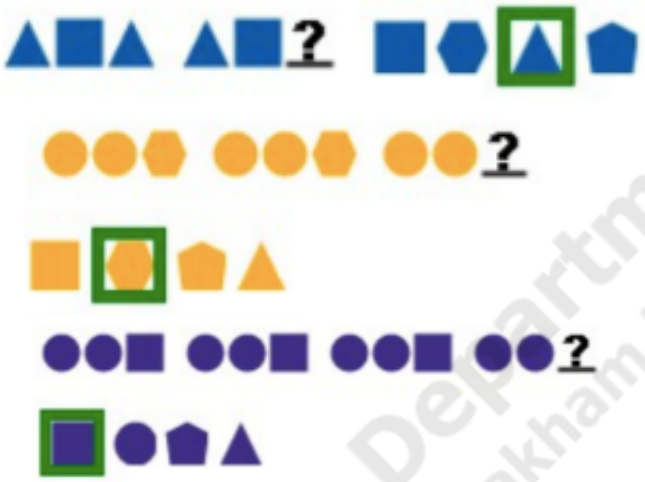
ภาพประกอบที่ 2.26 NuCAPTCHA

NuCAPTCHA เป็นแคปท์ช่ารูปภาพผสมอักษรข้อความแสดงภาพเคลื่อนไหวเป็นวิดีโอที่เป็นพื้นหลัง และแสดงข้อความยาวเป็นประโยคเคลื่อนที่ผ่านไปโดยทุกอักษรของข้อความนั้นจะมีการเคลื่อนไหวไปมา แต่จะมีตัวอักษรเป็นสีแดงจำนวนหนึ่ง ผู้ใช้ต้องพิมพ์ตัวอักษรสีแดงที่ปรากฏทั้งหมดให้ถูกต้อง ตามลำดับจึงผ่านการทดสอบ ดังภาพประกอบที่ 2.26

Computer Science Department
Faculty of Informatics, Mahasarakham University

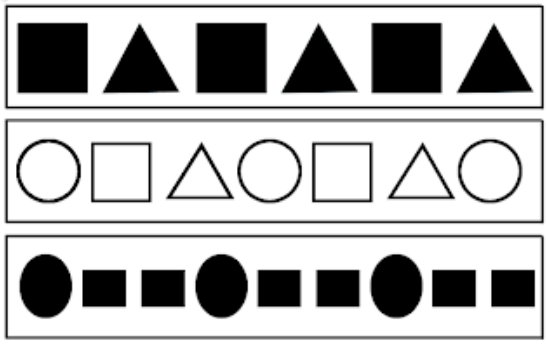
2.2 งานวิจัยที่เกี่ยวข้อง

แคปท์ช่าที่ถูกพัฒนาโดย อาจารย์ ธวัชวงศ์ ลาวัลย์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยสารคาม โดยงานวิจัยที่เสนอให้เห็นว่า แคปท์ช่ารูปแบบใหม่หรือแคปท์ช่ารูปเรขาคณิตนั้นมีความเข้าใจง่ายต่อมนุษย์และมีความปลอดภัยจากโปรแกรมอัตโนมัติ[1] ดังภาพประกอบที่ 2.27



ภาพประกอบที่ 2.27 แคปท์ช่ารูปร่าง เรขาคณิต

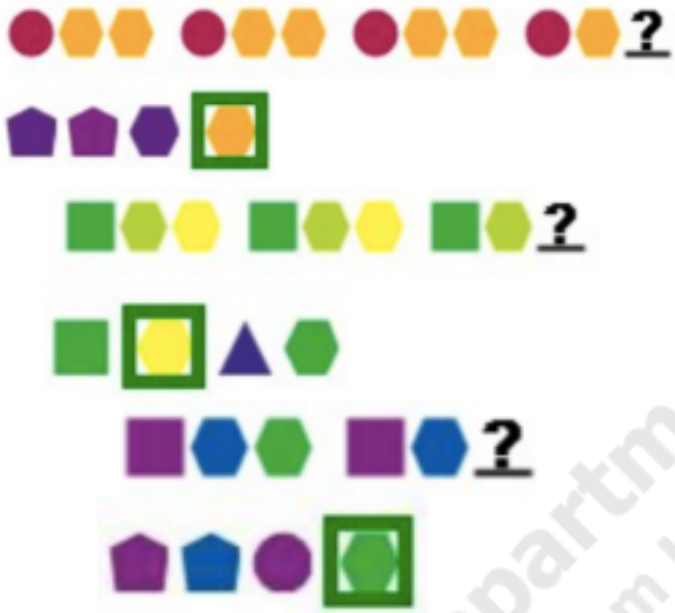
จากภาพประกอบที่ 2.27 เป็นนำรูปทรงเรขาคณิตและรูปแบบ(Patterns)หรือที่เรียงว่า “อนุกรม” คือชุดของตัวเลขหรือรูปภาพที่มีความสัมพันธ์กันอย่างใดอย่างหนึ่ง ในลักษณะของจำนวน รูปร่าง สี หรือขนาด ตามกฎเกณฑ์ที่กำหนด ซึ่งเมื่อทราบกฎเกณฑ์หรือความสัมพันธ์ที่กำหนดในแต่ละแบบรูป เราก็สามารถบอกคาดเดาหรือคาดการณ์ได้ว่าสิ่งต่าง ๆ รูปเรขาคณิต รูปอื่น ๆ หรือจำนวนที่หายไปคืออะไรได้ จากรูป จะเห็นได้ว่ารูปแบบที่ใช้คือ แบบรูปเรขาคณิต (Geometric Patterns) เป็นแบบรูปที่แสดงชุดของรูปเรขาคณิตที่มีความสัมพันธ์กันในลักษณะใดลักณะหนึ่ง เป็นแบบรูปที่มีความสัมพันธ์ในลักษณะของรูปร่าง ตัวอย่าง



ภาพประกอบที่ 2.28 แบบรูปที่มีความสัมพันธ์ในลักษณะของรูปร่าง

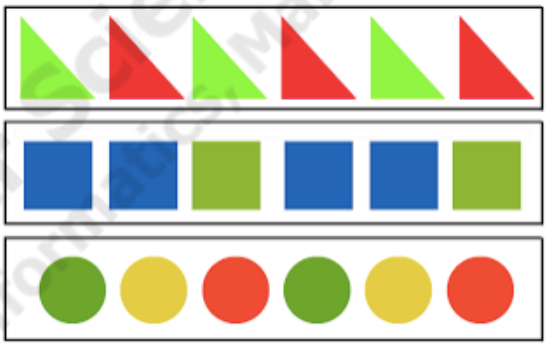
จากการสังเกตุเราจะเห็นว่าแบบรูปแต่ละแถวเป็นแบบรูปที่ประกอบด้วยรูปเรขาคณิตรูปทรงเดียวกันแตกต่างกันเพียงรูปทรงที่เรียงสลับตามลำดับซ้ำกันไปเรื่อย ๆ

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.29 แคปท์ช่ารูปร่าง เรขาคณิต และสี

จากภาพประกอบที่ 2.29 จะเห็นได้ว่ารูปแบบที่ใช้คือ แบบรูปเรขาคณิต (Geometric Patterns) เป็นแบบรูปที่แสดงชุดของรูปเรขาคณิตที่มีความสัมพันธ์กันในลักษณะใดลักณะหนึ่ง เป็นแบบรูปที่มีความสัมพันธ์ในลักษณะของสี ตัวอย่าง



ภาพประกอบที่ 2.30 แบบรูปที่มีความสัมพันธ์ในลักษณะของสี

จากการสังเกตุเราจะเห็นว่าแบบรูปแต่ละแถวเป็นแบบรูปที่ประกอบด้วยรูปเรขาคณิตรูปทรงเดียวกันแตกต่างกันเพียงสีที่เรียงสลับตามลำดับซ้ำกันไปเรื่อย ๆ

โดยแคปท์ช่านี้เป็นการผสมผสานระหว่าง แบบรูปที่มีความสัมพันธ์ในลักษณะของรูปร่างและแบบรูปที่มีความสัมพันธ์ในลักษณะของสี

2.3 การประเมินประสิทธิภาพ

ในการประเมินในงานวิจัยฉบับนี้จะมีผู้เชี่ยวชาญในการสร้างผลเฉลยเพื่อใช้ในการตรวจสอบภายใต้การใช้ตารางความคลาดเคลื่อน (Confusion Matrix) ซึ่ง เป็นรูปแบบตารางที่เฉพาะเจาะจงที่ช่วยให้เห็นผลการทำงานของแคปท์ช่ารูปเรขาคณิตได้

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 2.1 ตารางความคลาดเคลื่อน

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

จากตารางที่ 2.1 จะแสดงตาราง Confusion Matrix คือการประเมินผลลัพธ์การทานาย หรือผลลัพธ์ จากโปรแกรมเปรียบเทียบกับผลลัพธ์จริง ๆ ที่เป็นจริง โดยเปรียบเทียบกับผลเฉลยด้วยมนุษย์ โดยที่

TP (True Positive) หมายถึงสิ่งที่โปรแกรมทานายว่าจริงและคนบอกว่าจริง

TN (True Negative) หมายถึงสิ่งที่โปรแกรมทานายว่าจริงและคนบอกว่าไม่จริง

FP (False Positive) หมายถึงสิ่งที่โปรแกรมทานายว่าไม่จริงและคนบอกว่าจริง และ

FN (False Negative) หมายถึงสิ่งที่โปรแกรมทานายว่าไม่จริงและคนบอกไม่ว่าจริง

ขั้นตอนการประเมินการค้นคืนเอกสารก่อนไปใช้งานจริง ซึ่งโดยทั่วไปจะใช้เทคนิคมาตรฐานที่เรียกว่าการวัดค่าความระลึก (Recall) , การวัดค่าความแม่นยา (Precision) , และ F- measure

2.3.1 ค่าความระลึก (Recall) คืออัตราส่วนของเอกสารที่สืบค้นได้ จากเอกสารทั้งหมดที่มีอยู่ สามารถหาได้จากสมการดังต่อไปนี้

Recall = TP / (TP + FN) (2.10)

2.3.2 ค่าความแม่นยา (Precision) คืออัตราส่วนของเอกสารที่สืบค้นได้และถูกต้องตามความต้องการส่วนด้วยจานวนของเอกสารที่สืบค้นได้ สามารถแสดงเป็นสมการได้ดังนี้

Precision = TP / (TP + FP) (2.11)

2.3.3 ค่าประสิทธิภาพ (F-measure) เป็นการวัดประสิทธิภาพคือการวัดค่าความถูกต้องของระบบ โดยใช้ค่าเฉลี่ยถ่วงน้าหนักของค่าความแม่นยาและค่าความระลึก โดยมีประสิทธิภาพสูงสุดที่ 1 และ ต่ำสุดที่ 0 โดยสามารถหาค่าประสิทธิภาพได้ดังสมการต่อไปนี้

F – measure = 2 * ((Precision * Recall) / (Precision + Recall)) (2.12)

Computer Science Department
Faculty of Informatics, Mahasarakham University