บทที่ 2

หลักการและทฤษฎีที่เกี่ยวข้อง

หลักการและทฤษฎีที่เกี่ยวข้องกับงานวิจัยนี้ประกอบด้วย 2 ส่วน คือ การประมวลผลและการวิเคราะห์ภาพดิจิทัล (Digital image processing and analysis) และการจำแนกประเภทแบบรูป (Pattern classification) การหาตำแหน่งนาฬิกาตัวเลขในภาพวิดิทัศน์แบบอัตโนมัติจำเป็นต้องใช้การประมวลผลและวิเคราะห์ภาพดิจิทัล โดยใช้ลักษณะสำคัญของนาฬิกาที่แตกต่างจากส่วนอื่นในภาพช่วยระบุตำแหน่ง เมื่อทราบตำแหน่งของนาฬิกาตัวเลขแล้วจะใช้การจำแนกประเภทแบบรูปเพื่อรู้จำเวลาของนาฬิกาตัวเลขในภาพ รายละเอียดของหลักการและทฤษฎีของทั้ง 2 ส่วนมีดังต่อไปนี้

2.1 การประมวลผลและการวิเคราะห์ภาพดิจิทัล (Digital Image Processing)

ในงานวิจัยนี้ใช้การประมวลผลและการวิเคราะห์ภาพดิจิทัลเพื่อการทำงานหลายอย่าง เช่น เพื่อดึงลักษณะที่สำคัญของนาฬิกาตัวเลขในภาพออกมา หรือเพื่อกำจัดส่วนที่ไม่ใช่นาฬิกาตัวเลขออกจากภาพทำให้ตรวจหาตำแหน่งนาฬิกาตัวเลขได้ง่ายและถูกต้องยิ่งขึ้น รวมถึงใช้ในการประมวลผลข้อมูลเบื้องต้นก่อนนำนาฬิกาตัวเลขเข้าสู่กระบวนการรู้จำ

2.1.1 การตรวจหาขอบภาพ (Edge Detection)

ขอบภาพคือเซตของจุดภาพที่เชื่อมต่อกันซึ่งวางตัวอยู่บนขอบเขตระหว่างสองบริเวณที่แตกต่างกัน [11] การตรวจหาขอบภาพทำได้โดยการวัดอัตราการเปลี่ยนแปลงของค่าระดับเทาในภาพ ถ้าค่าระดับเทามีการเปลี่ยนแปลงทันทีในช่วงระยะทางสั้น ๆ แสดงว่าบริเวณนั้นมีแนวโน้มอย่างมากที่จะเป็นขอบภาพ วิธีการหนึ่งที่นิยมใช้กันคือ การคำนวณหาค่าเกรเดียนต์ ค่าเกรเดียนต์ G ของภาพ $f(x,y)$ ณ ตำแหน่ง (x,y) ใด ๆ นิยามโดยเวกเตอร์ซึ่งประกอบด้วยสมาชิก 2 ตัวคือ G_x และ G_y แสดงดังสมการที่ 2.1

$$\nabla\vec{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \qquad \ldots (2.1)$$

เมื่อ G_x และ G_y เป็นค่าแสดงอัตราการเปลี่ยนแปลงค่าระดับเทาในทิศตามแนวแกน x และ y ตามลำดับ ขนาดของเกรเดียนต์ ∇f คำนวณได้โดยสมการที่ 2.2

$$\nabla f = \left\|\nabla \vec{f}\right\| = \sqrt{G_x^2 + G_y^2} \quad \text{... (2.2)}$$

แต่ในทางปฏิบัติจะใช้ค่าประมาณขนาดของเกรเดียนต์ด้วยผลรวมของค่าสัมบูรณ์ของ G_x และ G_y เพื่อให้การประมวลผลภาพรวดเร็วและมีประสิทธิภาพมากยิ่งขึ้น แสดงดังสมการที่ 2.3

$$\nabla f \approx |G_x| + |G_y| \quad \text{... (2.3)}$$

ปริมาณที่สำคัญของเกรเดียนต์อีกปริมาณหนึ่งคือทิศของเวกเตอร์เกรเดียนต์ซึ่งหาได้จากสมการที่ 2.4

$$\theta(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad \text{... (2.4)}$$

เมื่อ $\theta(x, y)$ คือมุมแสดงทิศของเวกเตอร์ $\nabla \vec{f}$ ณ จุด (x, y) ซึ่งกระทำกับแกน x

ตัวอย่างตัวดำเนินการเกรเดียนต์ที่นิยมใช้ตรวจหาขอบภาพคือตัวดำเนินการ Sobel ซึ่งมีลักษณะเป็นหน้ากากขนาด 3x3 ดังแสดงในรูปที่ 2.1 การตรวจหาขอบภาพทำโดยนำหน้ากากสำหรับตรวจหาขอบภาพในแนวนอนและแนวตั้งของตัวดำเนินการ Sobel ทำคอนโวลูชันกับภาพซึ่งจะได้ผลลัพธ์เป็นค่าเกรเดียนต์หรือขอบภาพในทิศตามแนวแกน x และ y ตามลำดับ จากนั้นจึงสามารถหาขอบภาพและทิศทางของขอบภาพ ณ จุดใด ๆ ในภาพได้จากสมการที่ 2.3 และ 2.4 ตามลำดับ

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

(ก) หน้ากากสำหรับหาขอบในแนวนอน *(ข) หน้ากากสำหรับหาขอบแนวตั้ง*

รูปที่ 2.1 หน้ากากของตัวดำเนินการ Sobel

ภาพของขอบภาพที่หาได้จากการตรวจหาขอบภาพโดยใช้หน้ากากของตัวดำเนินการ Sobel เป็นภาพระดับเทา ขอบภาพที่ได้จะหนา ทำให้ไม่ทราบตำแหน่งของขอบภาพที่ถูกต้อง ดังนั้นหากต้องการทำให้ขอบภาพมีความหนาเพียง 1 จุดภาพจึงต้องทำขอบภาพให้บางด้วยวิธีการกำจัดค่าที่ไม่มากที่สุด (Non-maxima suppression) [12] ซึ่งเป็นการหาจุดภาพในภาพขอบที่มีค่ามากกว่าค่าขีดแบ่งค่าหนึ่ง และเป็นจุดที่มีค่ามากที่สุดเทียบกับจุดข้างเคียง 2 จุดในทิศทางของเกรเดียนต์ ณ จุดนั้น จุดภาพที่สอดคล้องกับเงื่อนไขดังกล่าวถือว่าเป็นขอบภาพใหม่หลังการทำขอบภาพให้บางโดยกำหนดค่าให้จุดภาพนั้นเป็นสีขาว ส่วนจุดภาพที่ไม่สอดคล้องกับเงื่อนไขถือว่าไม่เป็นขอบภาพ ให้กำจัดออกจากภาพขอบโดยกำหนดค่าให้จุดภาพนั้นเป็นสีดำ ดังนั้นจะได้ภาพผลลัพธ์เป็นภาพลักษณ์ฐานสอง (Binary image)

2.1.2 การแปลงภาพสีเป็นภาพระดับเทา

เนื่องจากการประมวลผลภาพสีค่อนข้างยุ่งยาก ดังนั้นจึงนิยมที่จะแปลงภาพสีเป็นภาพระดับเทาก่อนการประมวลผลภาพ โมเดลสีที่นิยมใช้กันในการแทนข้อมูลรูปภาพได้แก่ โมเดลสี RGB ซึ่งประกอบด้วยระนาบข้อมูลของสีแดง เขียว และน้ำเงิน นอกจากนี้ยังมีโมเดลสีชนิดอื่นซึ่งพยายามแยกส่วนประกอบของสีกับส่วนประกอบของความเข้มแสงออกจากกัน และส่วนประกอบของความเข้มแสงนี้เองที่เป็นส่วนแสดงค่าระดับเทาของภาพสีแต่ละภาพ โมเดลสีที่ใช้แปลงภาพสีเป็นภาพระดับเทาที่ใช้กันอย่างแพร่หลายได้แก่ โมเดลสี YIQ ซึ่งมีค่า Y เป็นค่าแสดงความเข้มแสง คำนวณได้จากสมการที่ 2.5 [12]

$$Y = 0.30R + 0.59G + 0.11B \qquad \text{... (2.5)}$$

จากสมการจะเห็นได้ว่าการแปลงภาพสีเป็นภาพระดับเทาโดยใช้โมเดลสี YIQ จะให้ค่าน้ำหนักในแต่ละสีไม่เท่ากัน โดยจะให้ค่าน้ำหนักใกล้เคียงกับความไวแสงที่สายตามนุษย์รับรู้ได้ นั่นคือสายตามนุษย์จะมีความไวต่อแสงสีเขียวมากกว่าสีแดง และมากกว่าสีน้ำเงินตามลำดับ

2.1.3 การสเกลภาพ (Image Scaling)

การสเกลภาพเป็นการขยายหรือหดขนาดของภาพโดยการชักตัวอย่าง (Sampling) จากภาพต้นฉบับ ทำให้ได้ภาพผลลัพธ์ที่มีขนาดใหญ่ขึ้นหรือเล็กลงตามต้องการ โดยที่การขยายขนาดของภาพเป็นการชักตัวอย่างเกิน (Over sampling) และการหดขนาดของภาพเป็นการชักตัวอย่างขาด (Under sampling) การสเกลภาพประกอบด้วย 2 ขั้นตอนคือ การสร้างตำแหน่งจุดภาพใหม่ และการกำหนดค่าระดับเทาให้กับจุดภาพ ณ ตำแหน่งใหม่เหล่านั้น [11]

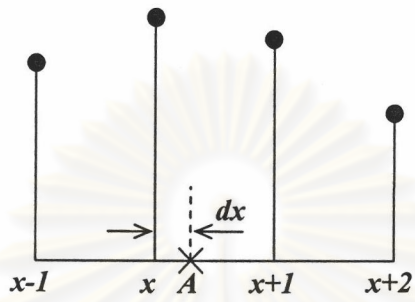
ยกตัวอย่างเช่น มีภาพต้นฉบับขนาด 100x100 จุดภาพ ต้องการขยายภาพให้มีขนาดใหญ่ขึ้นจากเดิมเป็น 200x200 จุดภาพ ขั้นตอนแรกคือการสร้างภาพเปล่าที่มีขนาด 200x200 จุดภาพขึ้นมาก่อน จากนั้นจึงกำหนดค่าระดับเทาให้กับจุดภาพแต่ละจุดซึ่งมีหลายวิธีการ วิธีการหนึ่งที่นิยมใช้กันได้แก่การประมาณค่าในช่วงแบบเชิงเส้นคู่ (Bilinear interpolation) [11] โดยจะใช้จุด 4 จุดใกล้เคียงกับจุดที่สนใจมาช่วยในการประมาณค่า สมมติให้ (x', y') เป็นพิกัดของจุดในภาพที่ทำการสเกล และให้ $v(x', y')$ เป็นค่าระดับเทาที่กำหนดให้จุดนั้น ๆ การประมาณค่าในช่วงแบบเชิงเส้นคู่แสดงได้ดังสมการที่ 2.6

$$v(x', y') = ax' + by' + cx'y' + d \qquad \text{... (2.6)}$$

เมื่อ a, b, c และ d เป็นค่าสัมประสิทธิ์ซึ่งหาได้จากการแทนค่าจุด 4 จุดบริเวณใกล้เคียงกับจุด (x', y') เมื่อทราบค่า a, b, c และ d แล้วจึงสามารถประมาณค่าระดับเทาของจุดที่สนใจได้

การสเกลภาพโดยใช้การประมาณค่าในช่วงแบบเชิงเส้นคู่นั้นบางครั้งผลลัพธ์ที่ได้อาจยังเห็นลักษณะเป็นตารางหมากรุก (Checker board) วิธีการอีกวิธีหนึ่งที่ใช้เพื่อให้การสเกลภาพได้ผลลัพธ์ที่

ราบเรียบมากยิ่งขึ้นคือการประมาณค่าในช่วงแบบ Bicubic (Bicubic interpolation) ซึ่งเป็นการใช้จุด 16 จุดใกล้เคียงมาช่วยประมาณค่า เพื่อให้เข้าใจหลักการการประมาณค่าในช่วงแบบ Bicubic ได้ง่ายขึ้นจึงแสดงการประมาณค่าในช่วงแบบเชิงกำลังสาม (Cubic interpolation) ใน 1 มิติ ดังรูปที่ 2.2 สังเกตได้ว่าจะใช้ 4 จุดใกล้เคียงมาประมาณค่า



รูปที่ 2.2 การประมาณค่าในช่วงแบบเชิงกำลังสามใน 1 มิติ

สมมติว่าความสูงในแต่ละตำแหน่งแทนด้วยฟังก์ชัน f ในที่นี้ต้องการหาความสูงของตำแหน่ง A ซึ่งอยู่ห่างจากตำแหน่ง x เป็นระยะทาง dx โดยใช้การประมาณค่าในช่วงแบบเชิงกำลังสาม ซึ่งสามารถหาได้จากสมการที่ 2.7

$$f(A) = a_0 f(x-1) + a_1 f(x) + a_2 f(x+1) + a_3 f(x+2) \quad \text{... (2.7)}$$

โดยที่ สัมประสิทธิ์ a_0 a_1 a_2 และ a_3 หาได้จากสมการที่ 2.8 ถึง 2.11 ตามลำดับ [13]

$$a_0 = -\frac{1}{2}(dx+1)^3 + \frac{5}{2}(dx+1)^2 - 4(dx) + 2 \quad \text{... (2.8)}$$

$$a_1 = \frac{3}{2}(dx)^3 - \frac{5}{2}(dx)^2 + 1 \quad \text{... (2.9)}$$

$$a_2 = \frac{3}{2}(dx-1)^3 - \frac{5}{2}(dx-1)^2 + 1 \quad \text{... (2.10)}$$

$$a_3 = -\frac{1}{2}(dx-2)^3 + \frac{5}{2}(dx-2)^2 - 4(dx-2) + 2 \quad \text{... (2.11)}$$

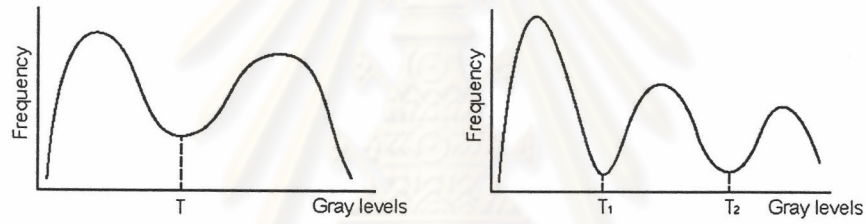
สำหรับการประมาณค่าในช่วงแบบ Bicubic ซึ่งเป็นการประมาณค่าในช่วงใน 2 มิติ หรือเป็นข้อมูลภาพนั่นเอง สามารถทำได้โดยประยุกต์ใช้การประมาณค่าในช่วงใน 1 มิติมาคำนวณทีละแถว แล้วนำค่าที่ได้ในแต่ละแถวมาคำนวณตามแนวคอลัมน์อีกครั้งหนึ่ง

2.1.4 การหาค่าขีดแบ่ง (Thresholding)

การหาค่าขีดแบ่งให้กับภาพเป็นวิธีที่นิยมใช้กันอย่างแพร่หลายเมื่อต้องการแบ่งส่วนวัตถุออกจากภาพ เนื่องจากสามารถเข้าใจและนำไปใช้ได้ง่าย

รูปที่ 2.3(ก) เป็นตัวอย่างการหาค่าขีดแบ่งแบบเชิงเดี่ยว (Single threshold) สมมติว่าเป็นรูปที่แสดงฮิสโทแกรมของภาพ $f(x,y)$ ซึ่งประกอบด้วยวัตถุสว่างอยู่ในพื้นหลังที่มืด นั่นคือจุดภาพที่เป็นวัตถุและพื้นหลังจะมีกลุ่มของค่าระดับเทา 2 กลุ่มหลัก วิธีหนึ่งในการแบ่งแยกวัตถุออกจากพื้นหลังทำโดยเลือกค่าขีดแบ่ง T ที่สามารถแบ่ง 2 กลุ่มดังกล่าวออกจากกัน จากนั้นกำหนดให้จุดภาพ (x,y) ที่มีค่าระดับเทา $f(x,y) > T$ เป็นจุดภาพของวัตถุ ส่วนจุดอื่นที่เหลือเป็นจุดภาพพื้นหลัง

ในรูปที่ 2.3(ข) แสดงกรณีการหาค่าขีดแบ่งเมื่อรูปภาพมีฮิสโทแกรมแบบ 3 ฐานนิยม เช่นภาพที่มีวัตถุสว่าง 2 อันอยู่ในพื้นหลังที่มืด การหาค่าขีดแบ่งแบบหลายเชิง (Multiple threshold) จะจำแนกจุดภาพ (x,y) ว่าเป็นจุดภาพของวัตถุแรกถ้า $T_1 < f(x,y) < T_2$ และเป็นจุดภาพของวัตถุที่สองถ้า $f(x,y) > T_2$ และเป็นจุดภาพพื้นหลังถ้า $f(x,y) \le T_1$



(ก) การหาค่าขีดแบ่งแบบเชิงเดี่ยว *(ข) การหาค่าขีดแบ่งแบบหลายเชิง*

รูปที่ 2.3 การหาค่าขีดแบ่งจากฮิสโทแกรมของค่าระดับเทา

การหาค่าขีดแบ่งอาจเขียนให้อยู่ในรูปแบบของฟังก์ชันได้ดังสมการที่ 2.12 [11]

$$T = T[x, y, p(x,y), f(x,y)] \qquad \text{... (2.12)}$$

เมื่อ $f(x,y)$ เป็นค่าระดับเทาของภาพ ณ จุด (x,y) และ $p(x,y)$ แสดงสมบัติเชิงท้องถิ่น (Local property) ของจุดดังกล่าว เช่น ค่าระดับเทาเฉลี่ยของจุดภาพใกล้เคียงที่มีจุดศูนย์กลาง ณ ตำแหน่ง (x,y) โดยทั่วไปภาพลักษณ์ฐานสอง $g(x,y)$ ที่ได้หลังจากการทำขีดแบ่งกับภาพระดับเทา $f(x,y)$ ด้วยค่าขีดแบ่ง T มีค่าดังสมการที่ 2.13

$$g(x,y) = \begin{cases} 1 & if\ f(x,y) > T \\ 0 & if\ f(x,y) \le T \end{cases} \qquad \text{... (2.13)}$$

จุดภาพที่มีค่าเท่ากับ 1 คือจุดภาพที่เป็นวัตถุ ขณะที่จุดภาพที่มีค่าเท่ากับ 0 คือจุดภาพที่เป็นพื้นหลัง

เมื่อ T ขึ้นอยู่กับ $f(x,y)$ หรือค่าระดับเทาเพียงอย่างเดียว จะเรียกการหาค่าขีดแบ่งนี้ว่าการ

หาค่าขีดแบ่งแบบครอบคลุม (Global thresholding) ถ้า T ขึ้นอยู่กับทั้ง $f(x,y)$ และ $p(x,y)$ จะเรียกว่าเป็นการหาค่าขีดแบ่งแบบท้องถิ่น (Local thresholding) และถ้า T ขึ้นกับพิกัด (x,y) ด้วย จะเรียกว่าเป็นการหาค่าขีดแบ่งแบบพลวัต (Dynamic thresholding)

2.1.5 การหาค่าขีดแบ่งด้วยวิธี Otsu (Otsu thresholding method)

การหาค่าขีดแบ่งด้วยวิธี Otsu [12] เป็นการหาค่าขีดแบ่งแบบอัตโนมัติโดยการพิจารณาฮิสโทแกรมของภาพซึ่งเหมาะกับภาพที่มีฮิสโทแกรมของค่าระดับเทาที่กระจายตัวแบบทวิฐานนิยม (Bimodal distribution) วิธี Otsu มีหลักการคือ พยายามหาค่าขีดแบ่งที่ทำให้ค่าความแปรปรวนภายในกลุ่ม (Within-group variance) ของทั้งสองกลุ่มในฮิสโทแกรมมีค่าต่ำที่สุด โดยที่กลุ่มทั้งสองกลุ่ม ได้แก่ กลุ่มที่มีค่าน้อยกว่าค่าขีดแบ่ง และกลุ่มที่มีค่ามากกว่าค่าขีดแบ่ง ค่าความแปรปรวนรวม σ^2 มีค่าดังสมการที่ 2.14

$$\sigma^2 = \sigma_W^2(t) + \sigma_B^2(t) \qquad \text{... (2.14)}$$

เมื่อ $\sigma_W^2(t)$ และ $\sigma_B^2(t)$ คือค่าความแปรปรวนภายในกลุ่มและระหว่างกลุ่มตามลำดับ และ t คือค่าขีดแบ่งที่แบ่งฮิสโทแกรมออกเป็นสองกลุ่ม เมื่อพิจารณาสมการสังเกตได้ว่าค่าความแปรปรวนภายในกลุ่มมีค่าน้อยที่สุดก็ต่อเมื่อค่าความแปรปรวนระหว่างกลุ่มมีค่าสูงที่สุด เพราะว่าค่าความแปรปรวนรวมมีค่าคงที่ Shapiro และ Stockman [12] ได้พิสูจน์ค่า $\sigma_B^2(t)$ แสดงดังสมการที่ 2.15

$$\sigma_B^2(t) = q_1(t)[1-q_1(t)][\mu_1(t)-\mu_2(t)]^2 \qquad \text{... (2.15)}$$

เมื่อ $q_1(t)$ คือ ค่าความน่าจะเป็นรวมของกลุ่มที่มีค่าน้อยกว่าหรือเท่ากับ t กำหนดให้เป็นกลุ่มที่ 1

$q_2(t)$ คือ ค่าความน่าจะเป็นรวมของกลุ่มที่มีค่ามากกว่า t กำหนดให้เป็นกลุ่มที่ 2

$\mu_1(t)$ และ $\mu_2(t)$ คือ ค่าเฉลี่ยของกลุ่มที่ 1 และ 2 ตามลำดับ

การคำนวณค่า t ที่ทำให้ $\sigma_B^2(t)$ มีค่ามากที่สุดทำได้อย่างง่ายและรวดเร็วโดยใช้ความสัมพันธ์ระหว่างค่าที่คำนวณสำหรับค่า t และ $t+1$ ซึ่งสามารถเขียนความสัมพันธ์เวียนเกิด (Recursive relationship) ได้ดังสมการที่ 2.16 2.17 และ 2.18

$$q_1(t+1) = q_1(t) + P(t+1) \qquad \text{... (2.16)}$$

เมื่อ $P(t)$ คือ ค่าความน่าจะเป็นเมื่อจุดภาพมีค่าระดับเทาเท่ากับ t และกำหนดค่าเริ่มต้นให้ $q_1(1) = P(1)$ และสามารถหา μ_1 จาก

$$\mu_1(t+1) = \frac{q_1(t)\mu_1(t) + (t+1)P(t+1)}{q_1(t+1)} \qquad \text{... (2.17)}$$

โดยกำหนดค่าเริ่มต้นให้ $\mu_1(0) = 0$ และสามารถหา μ_2 จาก

$$\mu_2(t+1) = \frac{\mu - q_1(t+1)\mu_1(t+1)}{1 - q_1(t+1)} \quad \text{... (2.18)}$$

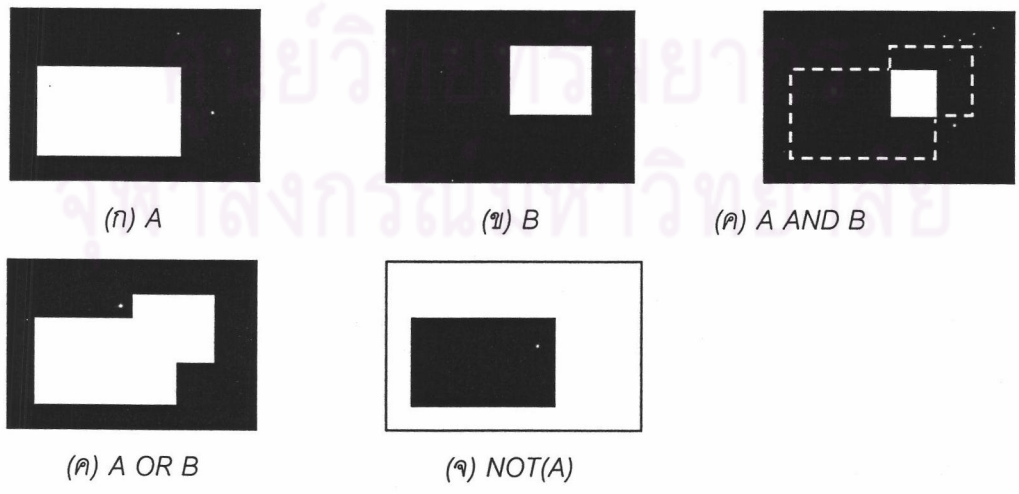
เมื่อ μ คือ ค่าเฉลี่ยรวมของฮิสโทแกรม

2.1.6 การประมวลผลภาพเชิงสัณฐานวิทยา (Morphological Image Processing)

การประมวลผลภาพเชิงสัณฐานวิทยาเป็นการนำสัณฐานวิทยาทางคณิตศาสตร์ (Mathematical morphology) มาใช้เพื่อช่วยสกัด (Extracting) ส่วนประกอบของภาพที่เป็นประโยชน์ในการแสดงและการอธิบายรูปร่างของพื้นที่ภายในภาพ เช่น บริเวณ โครงร่าง (Skeleton) ส่วนโค้ง และส่วนเว้า เป็นต้น สัณฐานวิทยาทางคณิตศาสตร์อาศัยทฤษฎีเซตในการแสดงวัตถุภายในภาพ ยกตัวอย่างเช่น เซตของจุดภาพสีขาวในภาพลักษณ์ฐานสองเป็นการอธิบายวัตถุเชิงสัณฐานวิทยาในภาพ การประมวลผลภาพเชิงสัณฐานวิทยาสามารถประมวลผลได้ทั้งภาพลักษณ์ฐานสอง ภาพระดับเทา และภาพสี อย่างไรก็ตามงานวิจัยนี้ใช้การประมวลผลภาพเชิงสัณฐานวิทยากับภาพลักษณ์ฐานสองเท่านั้น เทคนิคที่งานวิจัยนี้ใช้มีดังนี้

1) การดำเนินการทางตรรกะ (Logic Operation)

การดำเนินการทางตรรกะ [11] ที่ใช้เป็นพื้นฐานในการประมวลผลภาพได้แก่ ตรรกะแอนด์ (AND) ออร์ (OR) และนอต (NOT, COMPLEMENT) การดำเนินการทางตรรกะจะกระทำกับจุดภาพต่อจุดภาพระหว่างภาพ 2 ภาพหรือมากกว่านั้น (ยกเว้น ตรรกะนอต ซึ่งดำเนินการกับจุดภาพของภาพ 1 ภาพ) ตัวอย่างของตัวดำเนินการแต่ละตัวแสดงไว้ดังรูปที่ 2.4 ซึ่งสามารถเข้าได้โดยง่าย โดยที่สีขาวแทนเลข 1 และสีดำแทนเลข 0



รูปที่ 2.4 ตัวอย่างการดำเนินการทางตรรกะแอนด์ ออร์ และนอต

2) **การขยายขนาด** (Dilation)

การขยายขนาด [14] เป็นการขยายขนาดของวัตถุในภาพเพื่อจุดประสงค์บางประการ เช่น เพื่อปิดรูเล็ก ๆ ในวัตถุ หรือใช้เพื่อช่วยให้วัตถุ 2 วัตถุที่ไม่มีสมาชิกร่วมกันแต่อยู่ใกล้กันสามารถต่อเชื่อมกันได้ เป็นต้น การขยายขนาดทำได้โดยวางสมาชิกโครงสร้าง (Structure element) ลงบนภาพแล้วเลื่อนสมาชิกโครงสร้างในลักษณะเดียวกับการทำคอนโวลูชัน แต่ต่างกันที่ขั้นตอนการดำเนินการ การขยายขนาดมีขั้นตอนดังนี้

1. ถ้าจุดศูนย์กลางของสมาชิกโครงสร้างตรงกับค่า '0' ในภาพ ไม่ต้องดำเนินการใด ๆ และให้เลื่อนสมาชิกโครงสร้างไปยังจุดภาพถัดไป
2. ถ้าจุดศูนย์กลางของสมาชิกโครงสร้างตรงกับค่า '1' ในภาพ ให้ดำเนินการด้วยตัวดำเนินการทางตรรกะออร์ (or) ระหว่างภาพกับสมาชิกโครงสร้าง

รูปที่ 2.5(ค) แสดงภาพหลังจากการทำการขยายขนาดรูปที่ 2.5(ก) ด้วยสมาชิกโครงสร้างดังรูปที่ 2.5(ข) จะเห็นได้ว่าวัตถุในภาพซึ่งมีค่า '1' มีการขยายขนาดเพิ่มขึ้น

1	1	1	1	1	1	1	
			1	1	1	1	
			1	1	1	1	
		1	1	1	1	1	
			1	1	1	1	
		1	1				

(ก) ภาพลักษณ์ฐานสอง

1	1	1
1	1	1
1	1	1

(ข) สมาชิกโครงสร้าง

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1
	1	1	1	1	1	1	1
	1	1	1	1	1	1	1
	1	1	1	1	1	1	1
	1	1	1	1			

(ค) การขยายขนาด

				1	1		
				1	1		
				1	1		

(ง) การกร่อน

รูปที่ 2.5 การดำเนินการทางสัณฐานวิทยากับภาพลักษณ์ฐานสอง (ก) ด้วยสมาชิกโครงสร้าง (ข) โดย (ค)เป็นการขยายขนาด และ (ง)เป็นการกร่อน [14]

3) **การกร่อน** (Erosion)

การกร่อนเป็นการดำเนินการทางสัณฐานวิทยาเช่นเดียวกับการขยายขนาด แต่ให้ผลลัพธ์ในลักษณะที่ตรงกันข้าม นั่นคือการกร่อนจะลดขนาดของวัตถุให้เล็กลง แสดงดังรูปที่ 2.5(ง) ขั้นตอนการประมวลผลทำเช่นเดียวกับการขยายขนาด แต่ดำเนินการต่างกันดังนี้

1. ถ้าจุดศูนย์กลางของสมาชิกโครงสร้างตรงกับค่า '0' ในภาพ ไม่ต้องดำเนินการใด ๆ และให้เลื่อนสมาชิกโครงสร้างไปยังจุดภาพถัดไป
2. ถ้าจุดศูนย์กลางของสมาชิกโครงสร้างตรงกับค่า '1' ในภาพ ให้พิจารณาว่าถ้ามีจุดภาพค่า '1' ในสมาชิกโครงสร้างเลยออกมาจากวัตถุซึ่งมีค่า '1' ในภาพ ให้เปลี่ยนค่า '1' ณ ตำแหน่งที่เป็นจุดศูนย์กลางในภาพเป็นค่า '0'

2.1.7 **การปิดป้ายส่วนประกอบที่เชื่อมต่อกัน** (Connected Component Labeling)

สมมติภาพ B เป็นภาพลักษณ์ฐานสอง และกำหนดให้ $B[r,c] = B[r',c'] = v$ เมื่อ v มีค่าเท่ากับ 0 หรือ 1 จุดภาพ $[r,c]$ จะเชื่อมต่อกับจุดภาพ $[r',c']$ ด้วยค่า v ถ้ามีลำดับของจุดภาพ $[r,c] = [r_0,c_0],[r_1,c_1],...,[r_n,c_n] = [r',c']$ ซึ่ง $B[r_i,c_i] = v$ โดยที่ $i = 0,...,n$ และ $[r_i,c_i]$ เป็นจุดภาพเพื่อนบ้านของ $[r_{i-1},c_{i-1}]$ สำหรับแต่ละ $i = 1,...,n$ ลำดับของจุดภาพ $[r_0,c_0],...,[r_n,c_n]$ สร้างเส้นทางเชื่อมต่อกันจาก $[r,c]$ ถึง $[r',c']$ ส่วนประกอบที่เชื่อมต่อกันของค่า v คือเซตของจุดภาพที่มีค่าเท่ากับ v และคู่ของจุดภาพทุกคู่ภายในเซตเชื่อมต่อกันด้วยค่า v [12] รูปที่ 2.6(ก) แสดงภาพลักษณ์ฐานสองที่มีส่วนประกอบที่เชื่อมต่อกันของค่า 1

ส่วนการปิดป้ายส่วนประกอบที่เชื่อมต่อกันของภาพลักษณ์ฐานสอง B คือการปิดป้ายให้กับแต่ละส่วนประกอบที่เชื่อมต่อกันในภาพลักษณ์ฐานสอง B โดยที่ป้ายก็คือสัญลักษณ์ที่ตั้งชื่อให้กับแต่ละส่วนประกอบซึ่งอาจจะตั้งเป็นอักขระก็ได้ อย่างไรก็ตามการตั้งชื่อเป็นเลขจำนวนเต็มบวกจะสะดวกและนิยมใช้กันมากกว่า รูปที่ 2.6(ข) แสดงการปิดป้ายส่วนประกอบที่เชื่อมต่อกันของภาพลักษณ์ฐานสองในรูปที่ 2.6(ก)

2.1.8 **โพรไฟล์การฉาย** (Projection Profile)

การฉายของภาพลักษณ์ฐานสองลงบนเส้นตรงสามารถทำได้โดยแบ่งเส้นตรงออกเป็นช่อง ๆ และหาจำนวนของจุดภาพที่มีค่าเท่ากับ 1 ซึ่งอยู่ในแนวตั้งฉากกับเส้นตรงในแต่ละช่องนั้น โพรไฟล์การฉายเป็นการแสดงข้อมูลของภาพที่มีประโยชน์อย่างยิ่ง อย่างไรก็ตามโพรไฟล์การฉายมีข้อเสียอันเนื่องมาจากเป็นข้อมูลที่ไม่เป็นเอกลักษณ์ของภาพ ภาพมากกว่าหนึ่งภาพอาจมีโพรไฟล์การฉายที่เหมือนกัน โพรไฟล์การฉายตามแนวนอนและแนวตั้งหาได้อย่างง่าย ๆ จากการนับจำนวนจุดภาพที่มีค่าเท่ากับ 1 ในทิศทางตามแนวนอนและแนวตั้งของช่องแต่ละช่องในเส้นตรงที่แบ่งไว้ตามลำดับ ดังแสดงในรูปที่ 2.7 โดยที่จุดภาพสีดำมีค่าเท่ากับ 1 และจุดภาพสีขาวมีค่าเท่ากับ 0

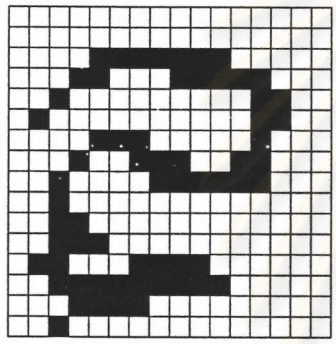
1	1	0	1	1	1	0	1
1	1	0	1	0	1	0	1
1	1	1	1	0	0	0	1
0	0	0	0	0	0	0	1
1	1	1	1	0	1	0	1
0	0	0	1	0	1	0	1
1	1	0	1	0	0	0	1
1	1	0	1	0	1	1	1

(ก) ภาพลักษณ์ฐานสอง

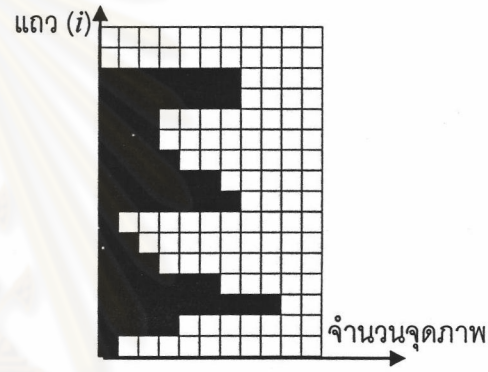
1	1	0	1	1	1	0	2
1	1	0	1	0	1	0	2
1	1	1	1	0	0	0	2
0	0	0	0	0	0	0	2
3	3	3	3	0	4	0	2
0	0	0	3	0	4	0	2
5	5	0	3	0	0	0	2
5	5	0	3	0	2	2	2

(ข) การปิดป้ายส่วนประกอบที่เชื่อมต่อกัน

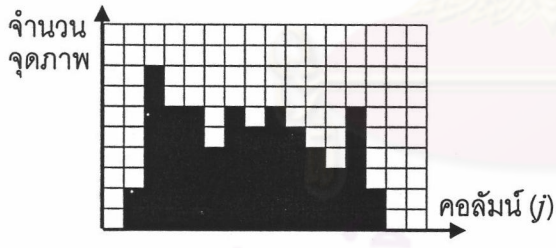
รูปที่ 2.6 ภาพลักษณ์ฐานสองและส่วนประกอบที่เชื่อมต่อกัน 5 ส่วนของค่า 1



(ก) ภาพลักษณ์ฐานสอง



(ข) โพรไฟล์การฉายตามแนวนอน



(ค) โพรไฟล์การฉายตามแนวตั้ง

รูปที่ 2.7 ตัวอย่างโพรไฟล์การฉายของภาพภาพลักษณ์ฐานสอง

โพรไฟล์การฉายตามแนวนอน $H[i]$ และแนวตั้ง $V[j]$ ของภาพลักษณ์ฐานสองกำหนดโดยสมการที่ 2.19 และ 2.20 ตามลำดับ [15]

$$H[i] = \sum_{j=1}^{m} B[i,j] \qquad \text{... (2.19)}$$

$$V[j] = \sum_{i=1}^{n} B[i,j] \qquad \text{... (2.20)}$$

โดยที่ B เป็นภาพลักษณ์ฐานสองซึ่งมีความสูงเท่ากับ m จุดภาพ และมีความกว้างเท่ากับ n จุดภาพ

2.1.9 การเข้ารหัสแบบลดความยาว (Run-Length Encoding)

การแสดงข้อมูลภาพลักษณ์ฐานสองวิธีหนึ่งก็คือการเข้ารหัสข้อมูลภาพแบบลดความยาว [15] ซึ่งเป็นการบีบอัดข้อมูลวิธีหนึ่ง จะใช้ในการส่งผ่านข้อมูลภาพ และยังสามารถใช้คำนวณสมบัติของภาพบางประการได้ เช่น พื้นที่ของวัตถุ การแสดงข้อมูลภาพนั้นจะแทนด้วยความยาวของจำนวนจุดภาพที่มีค่าเท่ากับ 1 ที่เรียงติดต่อกันในแต่ละแถว การเข้ารหัสที่นิยมใช้กันมี 2 วิธีหลัก ได้แก่ การใช้ตำแหน่งเริ่มต้น และความยาวที่ติดต่อกันของจุดภาพที่มีค่าเท่ากับ 1 ในแต่ละแถว และอีกวิธีจะใช้เฉพาะความยาวเท่านั้น โดยเริ่มต้นด้วยความยาวที่ติดต่อกันของจุดภาพที่มีค่าเท่ากับ 1 และค่าถัดไปจะเป็นความยาวที่ติดต่อกันของจุดภาพที่มีค่าเท่ากับ 0 สลับกันไป ดังแสดงในรูปที่ 2.8

1	1	1	0	0	0	1	1	0	0	0	1	1	1
0	0	0	0	1	1	1	1	1	1	1	1	0	0
1	0	0	0	0	0	0	0	0	1	1	1	1	0

(ก) ภาพลักษณ์ฐานสอง

ตำแหน่งเริ่มต้นและความยาวที่ติดต่อกันของจุดภาพที่มีค่าเท่ากับ 1 : (1, 3) (7, 2) (12, 3)
(5,8)
(1, 1) (10, 4)

(ข) การเข้ารหัสแบบลดความยาววิธีแรก

ความยาวที่ติดต่อกันของจุดภาพที่มีค่าเท่ากับ 1 และ 0 : 3, 3, 2, 3, 3
0, 4, 8, 2
1, 8, 4, 1

(ค) การเข้ารหัสแบบลดความยาววิธีที่สอง

รูปที่ 2.8 การเข้ารหัสแบบลดความยาวสำหรับภาพลักษณ์ฐานสอง

2.2 การจำแนกประเภทแบบรูป (Pattern Classification)

การจำแนกประเภทแบบรูปเป็นกระบวนการที่ใช้ระบุว่าวัตถุที่ไม่รู้จัก (Unknown object) ในภาพจัดอยู่ในกลุ่มวัตถุที่รู้จักกลุ่มใด คำว่า *แบบรูป* (Pattern) หมายถึงกลุ่มของตัวอธิบาย (Descriptor) หรือลักษณะต่าง ๆ ที่ได้มาจากวัตถุ ส่วนคำว่าการจำแนกประเภท (Classification) หรือการรู้จำ (Recognition) คือกระบวนการที่จัดแบบรูปที่ไม่รู้จักให้อยู่ในกลุ่มของแบบรูปอย่างถูกต้อง

โดยทั่วไปการจำแนกประเภทแบบรูปแบ่งออกเป็น 3 วิธี [16] ได้แก่ วิธีเชิงสถิติ (Statistical method) วิธีเชิงวากยสัมพันธ์ (Syntactic method) และวิธีที่ใช้สถาปัตยกรรม วิธีแรกเป็นวิธีเชิงสถิติซึ่งจะนิยามเซตของกฎการตัดสินใจโดยใช้ข้อมูลทางสถิติที่ได้จากการวิเคราะห์ลักษณะเด่นของแบบรูปเป็น

ฐาน วิธีที่สองคือวิธีเชิงวากยสัมพันธ์หรือมีอีกชื่อหนึ่งว่าวิธีการวิเคราะห์โครงสร้าง (Structural analysis method) ซึ่งเป็นการวิเคราะห์โครงสร้างของลักษณะเด่นของแบบรูป บางครั้งข้อมูลที่สำคัญนอกจากจะอยู่ในรูปแบบของลักษณะเด่นของแบบรูปแล้ว ยังอยู่ในรูปแบบของความสัมพันธ์ระหว่างลักษณะเด่นเหล่านั้นอีกด้วย และวิธีที่สามเป็นวิธีที่ใช้สถาปัตยกรรมซึ่งสามารถสอนให้ตอบสนองกับข้อมูลในส่วนนำเข้าได้อย่างถูกต้อง ตัวอย่างของวิธีนี้ เช่น โครงข่ายประสาทเทียม เป็นต้น

ในงานวิจัยนี้ใช้วิธีเชิงสถิติในการจำแนกประเภทแบบรูป เนื่องจากวัตถุที่จะนำมารู้จำในงานวิจัยนี้คือตัวเลขนาฬิกาซึ่งมีทั้งหมด 10 ตัว คือเลข 0 ถึงเลข 9 อีกทั้งตัวเลขมีแบบอักษรที่ตายตัว (Fixed font) ทำให้ไม่ต้องการระบบในการรู้จำที่ซับซ้อนมากนัก ดังนั้นจึงขอกล่าวถึงทฤษฎีที่เกี่ยวข้องในการจำแนกแบบรูปเฉพาะวิธีเชิงสถิติเท่านั้น

2.2.1 ต้นไม้การตัดสินใจ (Decision Trees)

สำหรับงานทางด้านการรู้จำแบบรูปที่เกี่ยวข้องกับลักษณะเด่นที่แตกต่างกันเป็นจำนวนมาก การเปรียบเทียบแบบรูปที่ไม่รู้จักกับแบบรูปอื่น ๆ ที่แตกต่างกันเป็นจำนวนมากด้วยเวกเตอร์ของลักษณะเด่นเป็นการเสียเวลาอย่างมาก การใช้ต้นไม้การตัดสินใจ [12] ทำให้ขั้นตอนการหาลักษณะเด่นและการจำแนกประเภทสามารถทำควบคู่กันไป กล่าวคือเป็นการเลือกลักษณะเด่นที่มีความสำคัญที่ทำให้สามารถจำแนกประเภทแต่ละประเภทออกจากกันได้ ต้นไม้การตัดสินใจเป็นโครงสร้างที่กระชับซึ่งใช้ลักษณะเด่น 1 อย่างต่อการแยกการค้นหาในแบบรูปที่เป็นไปได้ทั้งหมด ต้นไม้จะมีโหนด (Node) ซึ่งแสดงลักษณะเด่นที่แตกต่างกัน และแต่ละโหนดจะมีกิ่ง (Branch) มากกว่า 1 กิ่งซึ่งแต่ละกิ่งจะแทนค่าที่เป็นไปได้ทั้งหมดของลักษณะในโหนดนั้น ขั้นตอนการตัดสินใจว่าวัตถุที่ไม่รู้จักถูกจำแนกอยู่ในประเภทใดจะเริ่มจากการพิจารณาตั้งแต่โหนดราก (Root node) แล้วตัดสินว่าค่าของลักษณะเด่นของวัตถุที่ไม่รู้จักนั้นสอดคล้องกับเงื่อนไขในกิ่งใด แล้วจึงไล่ไปตามกิ่งนั้น จนกระทั่งพบโหนดใบ (Leaf node) ซึ่งแสดงประเภทของวัตถุ สมมติว่ามีข้อมูลดังตารางที่ 2.1 ซึ่งแสดงลักษณะต่าง ๆ ของลูกปิงปอง ลูกกอล์ฟ ลูกวอลเล่ย์บอล ลูกเทนนิส และลูกบาสเกตบอล จะสามารถสร้างต้นไม้การตัดสินใจได้โดยใช้ลักษณะเด่นทั้ง 3 อย่างได้แก่ สี ขนาด และผิวของวัตถุ

ตารางที่ 2.1 ตัวอย่างข้อมูลที่แสดงลักษณะเด่นของวัตถุทรงกลมประเภทต่าง ๆ

สี	ขนาด	ผิว	ประเภท
ขาว	เล็ก	เรียบ	ลูกปิงปอง
ขาว	เล็ก	ขรุขระ	ลูกกอล์ฟ
ขาว	ใหญ่	เรียบ	ลูกวอลเล่ย์บอล
เขียว	เล็ก	เรียบ	ลูกเทนนิส
ส้ม	ใหญ่	ขรุขระ	ลูกบาสเกตบอล

หลักการในการเลือกลักษณะเด่นเพื่อสร้างต้นไม้การตัดสินใจวิธีหนึ่งคือพยายามเลือกลักษณะเด่นที่แยกข้อมูลออกเป็นเซตย่อย โดยทำให้สมาชิกส่วนใหญ่ในแต่ละเซตย่อยเป็นประเภทเดียวกันมากที่สุด สิ่งที่ใช้วัดความสามารถในการแยกตัวอย่าง (Gain) ของแต่ละโหนดแสดงได้ดังสมการที่2.21 [12] ซึ่งเป็นการประยุกต์ใช้ทฤษฎีข่าวสาร (Information theory)

$$Gain(node) = \sum_b ((\frac{n_b}{n_t})(\sum_c -\frac{n_{bc}}{n_b}\log_2\frac{n_{bc}}{n_b})) \quad \text{... (2.21)}$$

โดยที่ n_b คือ จำนวนตัวอย่างของกิ่ง b

n_t คือ จำนวนตัวอย่างทั้งหมด

n_{bc} คือ จำนวนตัวอย่างของวัตถุประเภท c ที่กิ่ง b

ค่าความสามารถในการแยกตัวอย่างยิ่งต่ำยิ่งดี หมายถึงข้อมูลในกิ่งเดียวกันโดยส่วนใหญ่เป็นข้อมูลประเภทเดียวกัน ตัวอย่างการคำนวณความสามารถในการแยกตัวอย่างข้อมูลในตารางที่ 2.1 ของโหนดสีแสดงได้ดังนี้

$$Gain(color) = \frac{3}{5}(-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3})+\frac{1}{5}(-\log_2 1)+\frac{1}{5}(-\log_2 1) = 0.951$$

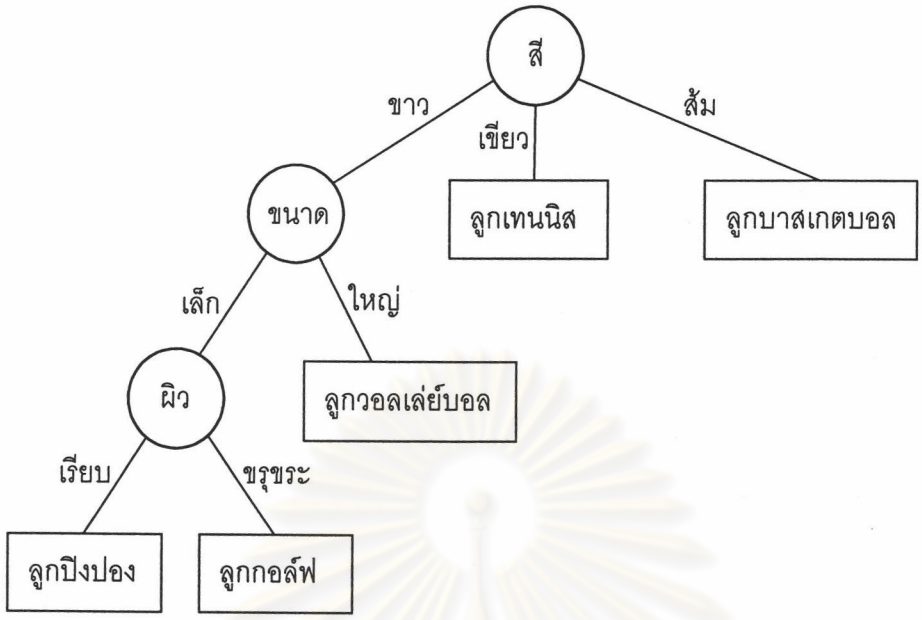
โดยที่ $\frac{3}{5}$ ในพจน์แรกมาจากจำนวนตัวอย่างในกิ่งสีขาวเท่ากับ 3 ตัวอย่างหารด้วยจำนวนตัวอย่างทั้งหมด 5 ตัวอย่าง และ $\frac{1}{3}$ ในพจน์แรกมาจากจำนวนตัวอย่างในแต่ละประเภทของกิ่งสีขาวมีประเภทละ 1 ตัวอย่าง โดยจำนวนตัวอย่างทั้งหมดในกิ่งสีขาวมี 3 ตัวอย่าง คำนวณกิ่งสีเขียวและสีส้มในทำนองเดียวกันซึ่งเป็นพจน์ที่สองและสามตามลำดับ

ความสามารถในการแยกตัวอย่างของโหนดขนาดและโหนดผิวมีค่าเท่ากัน คำนวณได้ดังนี้

$$Gain(size) = Gain(surface) = \frac{3}{5}(-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{3}\log_2\frac{1}{3})+$$
$$\frac{2}{5}(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}) = 1.351$$

ความสามารถในการแยกตัวอย่างของโหนดขนาดและผิวพบว่ามีค่ามากกว่าของโหนดสี ดังนั้นจึงเลือกโหนดสีเป็นโหนดราก จากนั้นจึงคำนวณในลักษณะเดียวกันกับโหนดที่ยังไม่สามารถจำแนกประเภทของวัตถุได้จนครบทั้งหมด

รูปที่ 2.9 แสดงต้นไม้การตัดสินใจที่สร้างได้หลังจากการคำนวณความสามารถในการแยกตัวอย่างของโหนดทั้งหมดแล้ว ทำให้สามารถจำแนกลูกเทนนิสและลูกบาสเกตบอลซึ่งมีสีเขียวและสีส้มตามลำดับออกมาได้ ส่วนลูกปิงปอง ลูกกอล์ฟ และลูกวอลเล่ย์บอลซึ่งมีสีขาวยังอยู่กลุ่มเดียวกัน จากนั้นใช้ขนาดเป็นตัวจำแนกทำให้แยกลูกวอลเล่ย์บอลออกมาได้ ส่วนลูกปิงปองและลูกกอล์ฟจะจำแนกด้วยลักษณะผิว



รูปที่ 2.9 ต้นไม้การตัดสินใจเพื่อจำแนกประเภทของวัตถุทรงกลม

2.2.2 ดัชนีความคล้าย (Similarity Index)

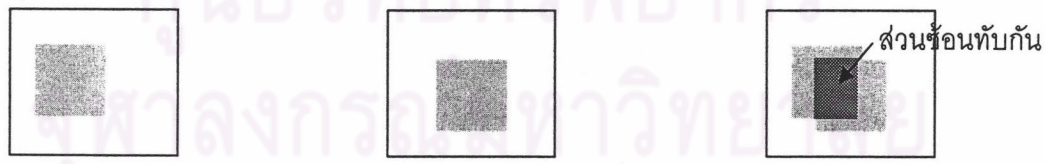
ดัชนีความคล้าย [17] ใช้เปรียบเทียบภาพ 2 ภาพว่ามีความคล้ายคลึงกันมากน้อยเพียงใด โดยคำนึงถึงทุก ๆ ส่วนของภาพทั้งบริเวณที่เหมือนและแตกต่าง สมมติว่ามีภาพ A และภาพ B แสดงดังรูปที่ 2.10(ก) และ (ข) ตามลำดับ ดัชนีความคล้าย S ของภาพ A กับภาพ B นิยามดังสมการที่ 2.22

$$S = \frac{2n\{A \cap B\}}{n\{A\} + n\{B\}} \qquad \text{... (2.22)}$$

โดยที่ $n\{A \cap B\}$ คือ จำนวนจุดภาพที่เป็นวัตถุที่ภาพ A และ B ซ้อนทับกัน แสดงดังรูปที่ 2.10(ค)

$n\{A\}$ คือ จำนวนจุดภาพที่เป็นวัตถุของภาพ A

$n\{B\}$ คือ จำนวนจุดภาพที่เป็นวัตถุของภาพ B



(ก) ภาพ A *(ข) ภาพ B* *(ค) ภาพ A ซ้อนทับกับภาพ B*

รูปที่ 2.10 ภาพประกอบการหาค่าดัชนีความคล้าย

ดัชนีความคล้าย S จะมีค่าอยู่ในช่วง 0 ถึง 1 ถ้า S มีค่าเท่ากับ 0 หมายถึงบริเวณที่เป็นวัตถุของทั้ง 2 ภาพไม่มีส่วนซ้อนทับกันเลย ถ้า S มีค่าเท่ากับ 1 หมายถึงบริเวณที่เป็นวัตถุของทั้งสองภาพซ้อนทับกันสนิทพอดี ดังนั้นยิ่งภาพมีความคล้ายกันค่า S จะมีค่าเข้าใกล้ 1