

บทที่ 2

ทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง

2.1 ปริภูมิสี

ปริภูมิสี มีมากมายหลายแบบ แต่ละแบบมีข้อดีข้อเสียแตกต่างกันไป และมีความเหมาะสมในการนำมาใช้งานลักษณะต่าง ๆ กัน ปริภูมิสีที่นิยมใช้และมักพบบ่อย คือ ปริภูมิสี RGB ซึ่งปริภูมิสีอื่น ๆ ส่วนมากจะได้จากการแปลง RGB โดยคำนึงถึงวัตถุประสงค์ที่นำมาใช้งาน เช่น ปริภูมิสี CMYK ซึ่งให้แหล่งกำเนิดสีอยู่ที่การซึมซับของหมึกพิมพ์บนกระดาษนำมาใช้สำหรับเครื่องพิมพ์ ปริภูมิสี L^*a^*b สร้างเพื่อใช้เป็นมาตรฐานในการวัดสี ครอบคลุมทุกสีใน RGB และ CMYK เป็นต้น

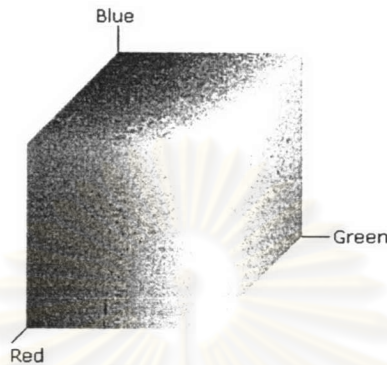
เมื่อพิจารณาระบบการมองเห็นของมนุษย์ (Human Vision System) กล่าวคือ มนุษย์มีความสามารถในการรับรู้ทางความส่องสว่างและสีได้ดีไม่เท่ากัน โดยสายตามนุษย์จะมีความไวต่อความส่องสว่างมากกว่าสี ในที่นี้จึงขอกล่าวรายละเอียดเฉพาะปริภูมิสีที่มีการแยกระหว่างความส่องสว่าง (Luminance) และสี (Chrominance) ออกจากกัน และปริภูมิสีพื้นฐาน RGB รวม 5 ปริภูมิสี คือ

- ปริภูมิสี RGB
- ปริภูมิสีอินอร์แมลไลซ์ RGB
- ปริภูมิสี HSI
- ปริภูมิสี YCbCr
- ปริภูมิสี L^*a^*b

2.1.1 ปริภูมิสี RGB

ปริภูมิสี RGB เป็นปริภูมิสีพื้นฐานหรือปริภูมิสีหลัก โดยค่าแต่ละจุดภาพจะเกิดจากการรวมกันของสเปกตรัมของแสงสีแดง (R) เขียว (G) และ น้ำเงิน (B) ในสัดส่วนความเข้มที่แตกต่างกัน โดยมีการกำหนดค่าความเข้มของสีแดง เขียว และน้ำเงิน เป็นค่าตั้งแต่ 0-255 (ทั้งหมด 256 ระดับ แทนด้วยค่า 8 บิต) ซึ่งที่ค่า 0 จะมีความเข้มมากที่สุด ไล่ลำดับไปจนถึง 255 ซึ่งมีความเข้มน้อยสุด ค่า RGB นี้เป็นค่าที่โทรทัศน์ กล้องวิดีโอ และจอคอมพิวเตอร์ใช้แสดงผล โดยจะเป็นปริภูมิเวกเตอร์บน R^3 ดังรูปที่ 2.1

ข้อเสียของปริภูมิสี RGB คือ ความสว่างและสี จะถูกรวมเข้าด้วยกันหรือเป็นสหสัมพันธ์กัน ซึ่งไม่สอดคล้องกับระบบการมองเห็นของสายตามนุษย์



รูปที่ 2.1 ปริภูมิสี RGB

2.1.2 ปริภูมิสีนอร์แมลไลซ์ RGB (rgb)

ปัญหาของปริภูมิสี RGB คือ การรวมกันของความสว่างและสี จึงเกิดเป็นปริภูมิสีนอร์แมลไลซ์ RGB ซึ่งมีวัตถุประสงค์เพื่อตัดความสนใจขององค์ประกอบความสว่างทิ้งไป โดยหากค่าจุดภาพในปริภูมิสี RGB นั้นมีสีเดียวกันแต่ความสว่างต่างกัน จะถูกกำหนดเป็นค่าเดียวกันในปริภูมิสีนอร์แมลไลซ์ RGB และแทนเป็นสัญลักษณ์ "rgb" แต่ละค่าสามารถหาได้จากสมการที่ (2.1)-(2.3)

$$r = \frac{R}{(R+G+B)} \quad (2.1)$$

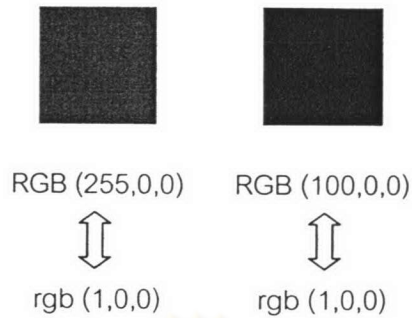
$$g = \frac{G}{(R+G+B)} \quad (2.2)$$

$$b = \frac{B}{(R+G+B)} \quad (2.3)$$

ตัวอย่างเช่น 1. ค่าในปริภูมิสี RGB (255,0,0) ทำการแปลงไปสู่ rgb จะได้ (1,0,0)

และ 2. ค่าในปริภูมิสี RGB (100,0,0) ทำการแปลงไปสู่ rgb จะได้ (1,0,0)

ซึ่งทั้งสองค่านี้ในปริภูมิสี RGB จะมีเฉพาะองค์ประกอบของสีแดง ที่มีค่าความเข้มของแสงแตกต่างกัน ส่วนในปริภูมิสี rgb องค์ประกอบทางแสงจะถูกละทิ้งไปทำให้ 2 ค่านี้ถือเป็นค่าเดียวกัน ดังรูปที่ 2.2



รูปที่ 2.2 ความแตกต่างของแสงในปริภูมิสี rgb จะถูกละทิ้ง

งานวิจัยที่เกี่ยวข้อง

J. Fritsch [6] เสนอขั้นตอนวิธีในการตรวจหาใบหน้า โดยใช้แบบจำลองสีผิวที่มีการปรับตัวได้ โดยพิจารณาการกระจายตัวของจุดภาพที่เป็นสีผิวในปริภูมิสีย่อย rg (นอร์แมลไลซ์ RG) แล้ว พบว่ามีรูปร่างคล้ายหอยเชลล์ (Shell shape) ดังแสดงในรูปที่ 2.4 จึงกำหนดขอบเขตล้อมรอบอาณาบริเวณนี้โดยใช้พาราโบลา 2 เส้น ตามสมการที่ (2.4) ขั้นตอนวิธีที่ใช้แบบจำลองสีจะถูกปรับตัวอยู่ตลอดเวลาโดยวิธีการ คือนำข้อมูลสีผิวที่ตรวจหาได้ ณ ปัจจุบันมาทำการปรับปรุงแบบจำลองสมการพาราโบลาที่นำมาใช้ ถูกแสดงในสมการที่ (2.4)

$$F_i = A_i r^2 + B_i r + C_i \quad \text{with } i \in \{u, d\} \quad (2.4)$$

เมื่อ u คือ พาราโบลาเส้นบน

d คือ พาราโบลาเส้นล่าง

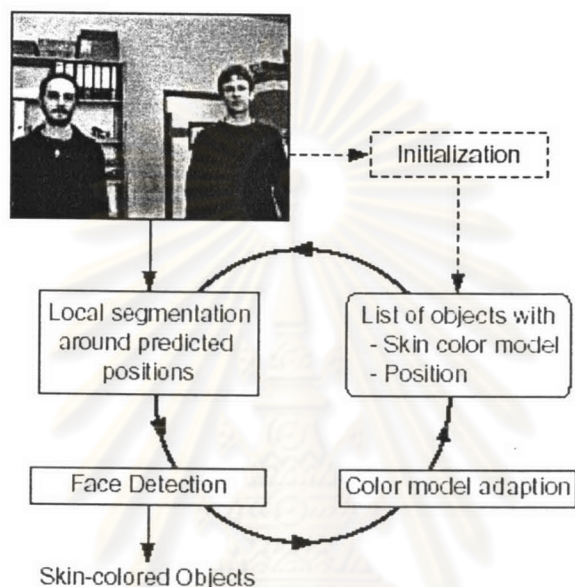
กำหนดค่าคงที่ A , B และ C ดังนี้

$$A_u = -5.05, A_d = -0.65, B_u = 3.71, B_d = 0.05, C_u = -0.32, C_d = 0.36 \quad (2.5)$$

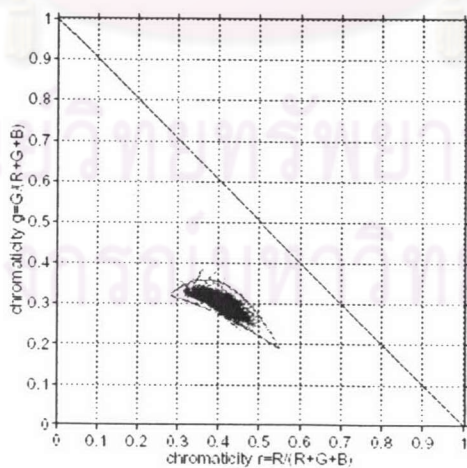
ข้อด้อยของวิธีนี้ คือ การรบกวนกันในปริภูมิสีนอร์แมลไลซ์ RGB เกิดขึ้นได้ง่ายกว่าปริภูมิสีอื่น ๆ โดยเฉพาะบริเวณตำแหน่งจุดภาพที่มีค่าใกล้เคียงกับสีดำ จะมีความแปรปรวนได้ง่าย และค่าของจุดภาพจำพวกนี้มีโอกาสที่จะเป็นค่าใดก็ได้ตั้งแต่ค่าต่ำสุดจนถึงสูงสุดในปริภูมิสีนอร์แมลไลซ์ RGB ทำให้แบบจำลองสีผิวที่สร้างขึ้นตีความหมายผิดพลาดได้

การจำแนกสีผิวทำโดยเลือกบริเวณที่ถูกล้อมด้วยพาราโบลา ในปริภูมิสีย่อย rg

$$Pixel \text{ is Skin} = \begin{cases} 1, & \text{if } (g < F_u) \& (g > F_d) \\ 0, & \text{else} \end{cases} \quad (2.6)$$



รูปที่ 2.3 ขั้นตอนวิธีการตรวจหาใบหน้าซึ่งมีการปรับตัวของแบบจำลองสีผิวเสนอโดย J. Fritsch [6]



รูปที่ 2.4 การกระจายตัวของจุดภาพที่เป็นสีผิวในปริภูมิสีย่อย rg ของ J. Fritsch [6]

2.1.3 ปริภูมิสี HSI

ค่าในปริภูมิสี HSI แยกเป็นค่า H และ S คือ องค์ประกอบของสี ส่วนค่า I คือ ความเข้มของแสง เป็นปริภูมิสีที่สร้างขึ้นตามพื้นฐานการมองเห็นสีด้วยสายตาของมนุษย์ โดยตัว H ย่อจาก Hue เป็นสีที่สะท้อนมาจากสีของวัตถุ ซึ่งแตกต่างกันตามความยาวของคลื่นแสงที่มากกระทบวัตถุและสะท้อนกลับมาที่ตาเรา จะถูกแทนด้วยค่า 0 ถึง 360 องศา ตัว S ย่อจาก Saturation เป็นค่าความอิ่มตัวของสี คือ สัดส่วนที่มีอยู่ในสีเทา วัดเป็นค่าเปอร์เซ็นต์ คือจาก 0 (สีเทา) ถึง 100 (สีมีความอิ่มตัวเต็มที่) และ ตัว I ย่อจาก Intensity เป็นค่าความเข้มของแสง วัดเป็นค่าเปอร์เซ็นต์ คือจาก 0 (ความมืด) ถึง 100 (ความสว่าง) ซึ่งการแปลงจาก RGB ไปสู่ HSI ทำได้โดยสมการที่ (2.7)-(2.10)

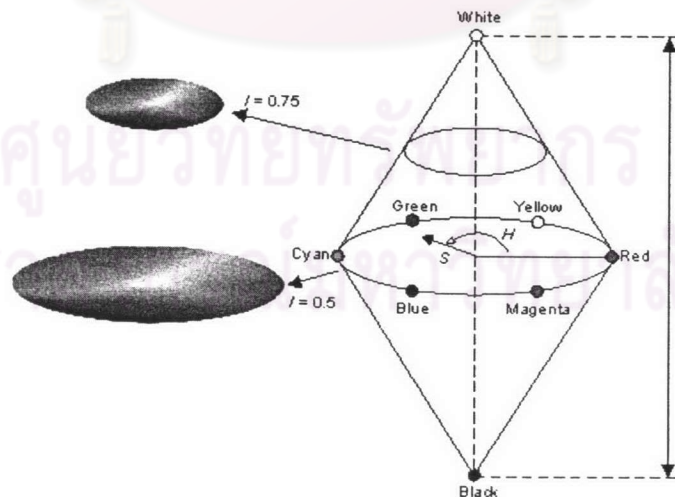
$$H = \begin{cases} \theta & ; B \leq G \\ 360^\circ - \theta & ; B > G \end{cases} \quad (2.7)$$

ซึ่งหา θ ได้จาก

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\} \quad (2.8)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \quad (2.9)$$

$$I = \frac{1}{3}(R+G+B) \quad (2.10)$$



รูปที่ 2.5 ปริภูมิสี HSI

การแปลงกลับจากปริภูมิสี HSI ไปสู่ RGB พิจารณาจากสมการที่ (2.11)-(2.21)

- ถ้า H มีค่าระหว่าง $0 \leq H < 120^\circ$

หาองค์ประกอบ RGB ได้ดังแสดงในสมการที่ (2.11)-(2.13)

$$B = I(1 - S) \quad (2.11)$$

$$R = I \left[1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.12)$$

$$G = 1 - (R + B) \quad (2.13)$$

- ถ้า H มีค่าระหว่าง $120^\circ \leq H < 240^\circ$ ต้องทำการลบค่า H ด้วย 120°

$$H = H - 120^\circ \quad (2.14)$$

และหาองค์ประกอบ RGB ได้ดังแสดงในสมการที่ (2.15)-(2.17)

$$R = I(1 - S) \quad (2.15)$$

$$G = I \left[1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.16)$$

$$B = 1 - (R + G) \quad (2.17)$$

- ถ้า H มีค่าระหว่าง $240^\circ \leq H < 360^\circ$ ต้องทำการลบค่า H ด้วย 240°

$$H = H - 240^\circ \quad (2.18)$$

และหาองค์ประกอบ RGB ได้ดังแสดงในสมการที่ (2.19)-(2.21)

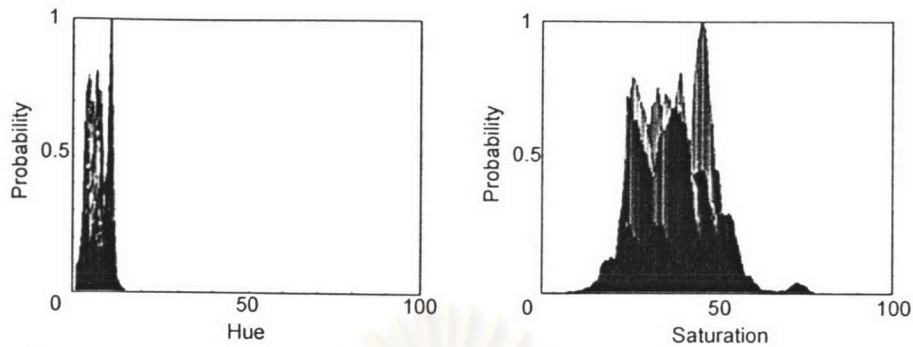
$$G = I(1 - S) \quad (2.19)$$

$$B = I \left[1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.20)$$

$$R = 1 - (G + B) \quad (2.21)$$

งานวิจัยที่เกี่ยวข้อง

Q. Huynh-Thu [8] เสนอวิธีการแยกสีผิวจากรูปภาพซึ่งมีฉากหลังที่ซับซ้อนและมีการเปลี่ยนแปลงความสว่าง สำหรับแบบจำลองที่ใช้เป็นแบบเกาส์ผสม ร่วมกับเทคนิคการปรับปรุงค่าทดสอบหลาย ๆ ค่า โดยอัตโนมัติ ซึ่งทำการแยกบนปริภูมิสีย่อย HS ข้อดีที่เกิดขึ้น คือ หากพิจารณาฮิสโทแกรมของสีผิวในปริภูมิสีย่อย HS ในรูปที่ 2.6 จำเป็นต้องใช้แบบจำลองเกาส์ผสมในการแยกส่วนนี้ออกมา ซึ่งส่งผลให้การคำนวณมีความซับซ้อนและใช้เวลาในการประมวลผลนาน อาจมีผลกระทบกับระบบที่ต้องการให้ทำงานตามเวลาจริงเมื่อวีดิทัศน์รับเข้ามีอัตราเฟรมสูง



รูปที่ 2.6 ฮิสโทแกรมของจุดภาพที่เป็นสีผิวในปริภูมิสีย่อย HS ของ Q. Huynh-Thu [8]

2.1.4 ปริภูมิสี YCbCr

ปริภูมิสี YCbCr เป็นปริภูมิสีดิจิทัล และมีการกำหนดเป็นมาตรฐานโดย BT.601 ได้จากการแปลงองค์ประกอบ YUV แอนะล็อก ซึ่งจะแบ่งข้อมูลของสีออกในลักษณะที่สอดคล้องกับการรับรู้ของมนุษย์เช่นเดียวกับปริภูมิสี HSI โดยมีค่า Y เป็นองค์ประกอบความสว่างมีค่าตั้งแต่ 16-235 ส่วนค่า Cb และ Cr คือ องค์ประกอบของสี ซึ่งเป็นค่าความแตกต่างของ สีน้ำเงินกับค่าความสว่าง (B-Y) และสีแดงกับค่าความสว่าง (R-Y) ตามลำดับ โดยที่ Cb และ Cr มีค่าตั้งแต่ 16-240

ในการแปลงปริภูมิสี RGB ไปสู่ YCbCr ทำได้โดยใช้สมการที่ (2.22)

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0980 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (2.22)$$

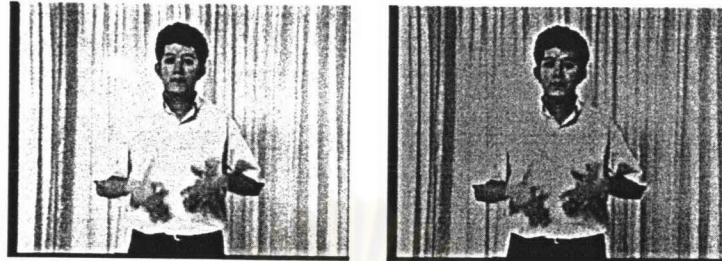
และการแปลงกลับไปจากปริภูมิสี YCbCr ไปสู่ RGB ทำได้โดยสมการที่ (2.23)

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.1643 & -0.0018 & 1.5958 \\ 1.1643 & -0.3914 & -0.8135 \\ 1.1643 & 2.0178 & -0.0012 \end{bmatrix} \begin{bmatrix} Y - 16 \\ Cb - 128 \\ Cr - 128 \end{bmatrix} \quad (2.23)$$

ค่าสูงสุดและต่ำสุดของ Cb และ Cr เมื่อเปรียบเทียบกับค่าในปริภูมิสี RGB

- ค่าสูงสุดของ Cb (240) คือ ค่า R=G=0, B=255 ค่าต่ำสุดของ Cb (16) คือ ค่า R=G=255, B=0
- ค่าสูงสุดของ Cr (240) คือ ค่า R=255, G=B=0 ค่าต่ำสุดของ Cr (16) คือ ค่า R=0, G=B=255

ตัวอย่าง การแยกกันระหว่างความสว่างและสีในปริภูมิสี YCbCr

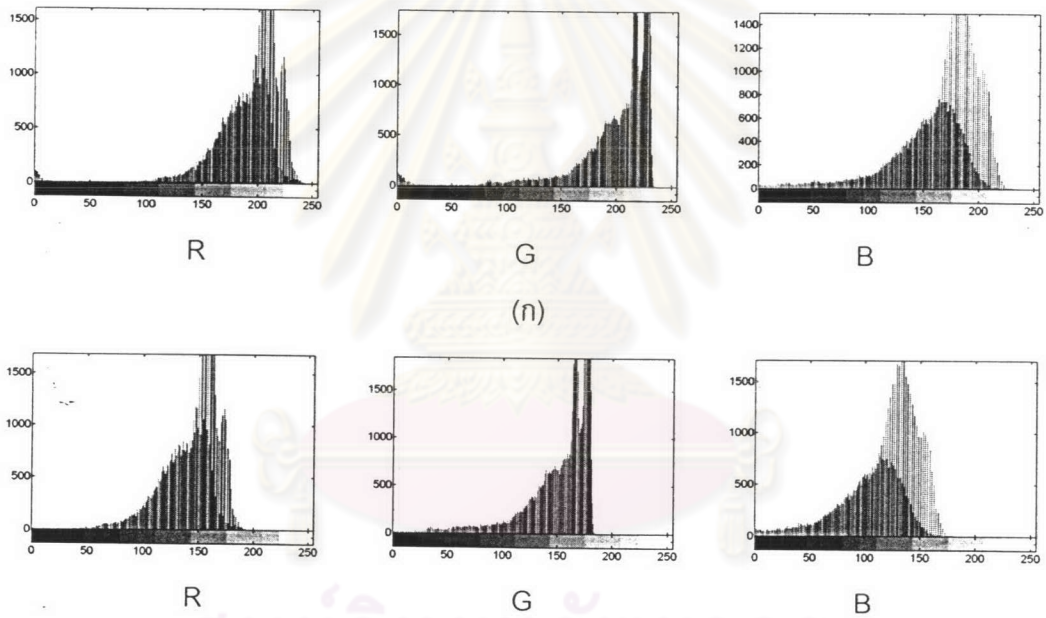


(ก)

(ข)

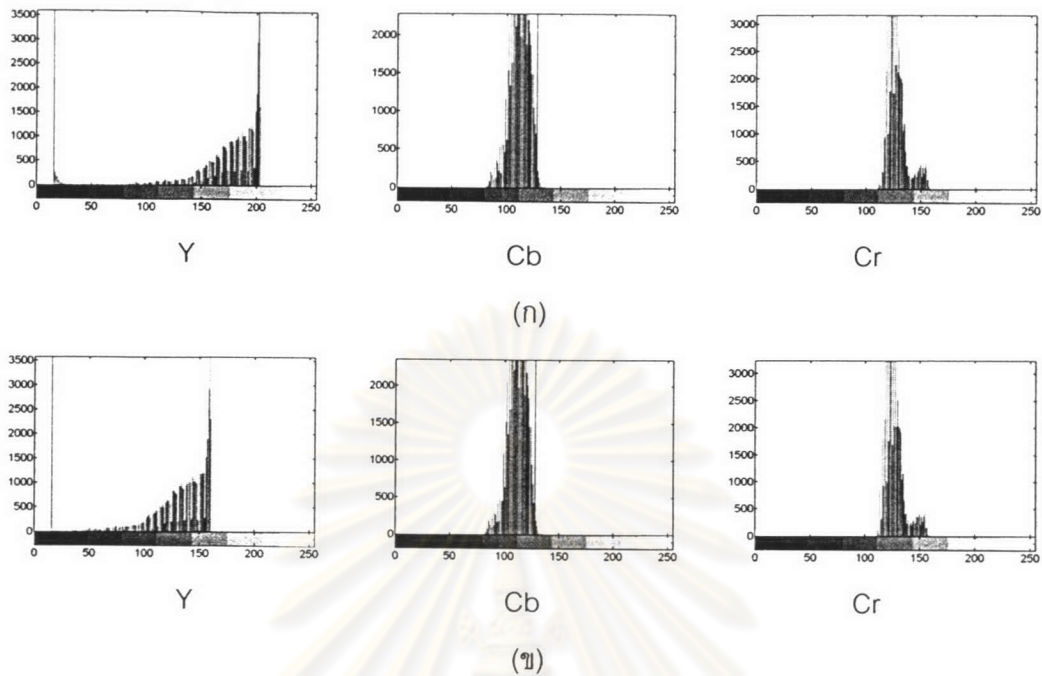
รูปที่ 2.7 ภาพขนาด 320x240 จุดภาพ ที่มีค่าความสว่างต่างกัน

(ก) ความสว่างที่ระดับปกติ (ข) ความสว่างที่ถูกลดระดับลงครึ่งหนึ่ง



รูปที่ 2.8 ค่าฮิสโตแกรมในปริภูมิ RGB ของภาพ 2.7 (ก) และ 2.7 (ข) ตามลำดับ

ศูนย์วิทยุทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.9 ค่าฮิสโทแกรมในปริภูมิ YCbCr ของภาพ 2.7 (ก) และ 2.7 (ข) ตามลำดับ

จากตัวอย่างจะเห็นว่า เมื่อค่าความสว่างเปลี่ยนแปลงไป ในปริภูมิสี RGB ค่าฮิสโทแกรมใน ทุก ๆ องค์ประกอบจะมีการเปลี่ยนแปลงตาม ส่วนปริภูมิสี YCbCr มีเพียงค่าฮิสโทแกรมของ องค์ประกอบ Y เท่านั้นที่เปลี่ยนแปลงอย่างเห็นได้ชัด ส่วนค่าฮิสโทแกรมขององค์ประกอบ Cb และ Cr มีผลกระทบเพียงเล็กน้อยเท่านั้น ฉะนั้นจึงสามารถพิจารณาเฉพาะค่าความสว่างหรือสี แยกจากกันได้ ในปริภูมิสี YCbCr

งานวิจัยที่เกี่ยวข้อง

S.L. Pung [13] ได้นำเสนอแบบจำลองที่ใช้ในการแยกส่วนสีผิวบนปริภูมิสีย่อย CbCr เพื่อ งานประยุกต์สำหรับการตรวจหาใบหน้า โดยแบบจำลองที่สร้างขึ้นอาศัยสมมติฐานของเกาส์หลังจาก เก็บข้อมูลในการสร้างแบบจำลองแล้ว เพื่อความรวดเร็วสามารถจำแนกสีผิวได้จากช่วงของ Cb และ Cr ดังนี้

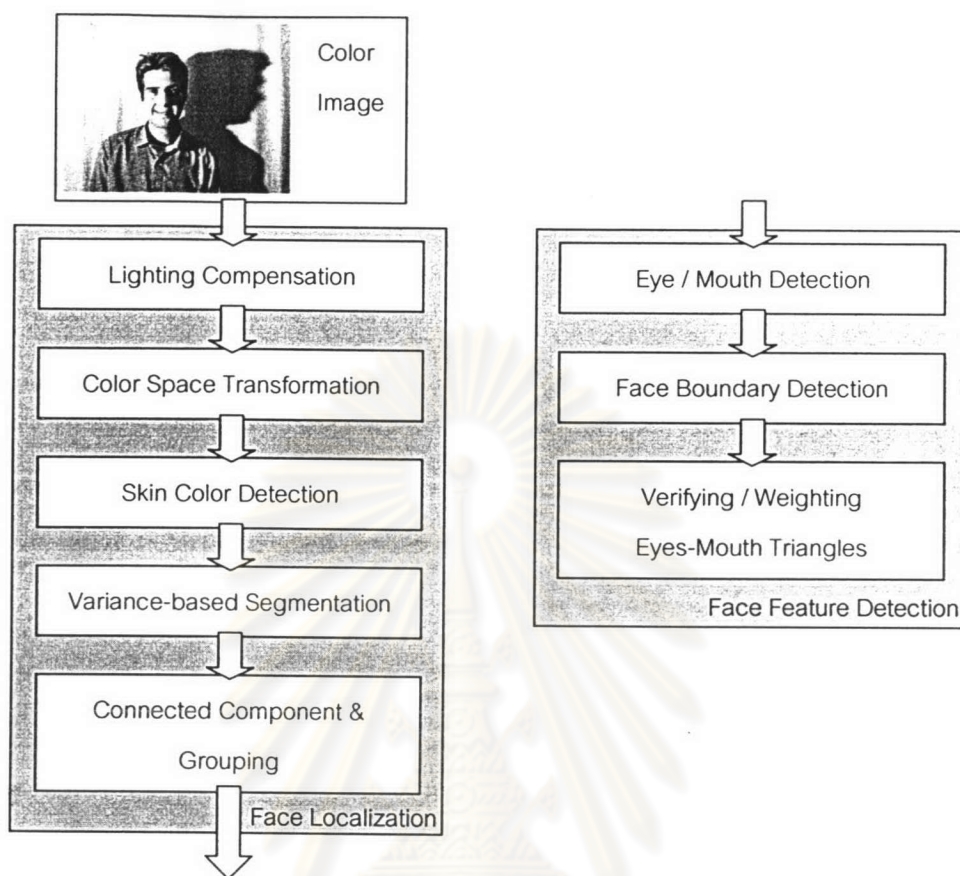
- Cb มีค่าอยู่ระหว่าง 75 ถึง 135
- Cr มีค่าอยู่ระหว่าง 130 ถึง 180

การกำจัดสัญญาณรบกวนที่เกิดขึ้น กรณีที่แบบจำลองตีความส่วนฉากหลังเป็นสีผิวเกินมา สำหรับงานวิจัยชิ้นนี้จะอาศัยคุณลักษณะของขนาด และความกระชับ (Compactness) ของแต่ละบริเวณร่วมด้วย ข้อดีของแบบจำลองสีผิวที่นำเสนอ คือ มีความซับซ้อนน้อยมาก หากพิจารณาบน ปริภูมิสีย่อย CbCr แบบจำลองที่ได้จะเป็นรูปสี่เหลี่ยม การทำงานของระบบ คือ พิจารณาเพียงว่า จุดภาพนั้นอยู่ในช่วงที่กำหนดหรือไม่ ถ้าใช่ก็จะตีความเป็นสีผิว แต่มีข้อด้อย คือ แบบจำลองจะให้ค่า ความผิดพลาดค่อนข้างมาก เนื่องจากค่าจุดภาพบริเวณสีผิวในปริภูมิสีย่อย CbCr จะมีการเกาะกลุ่ม กันเป็นลักษณะเฉพาะตัว ซึ่งไม่คล้ายกับสี่เหลี่ยมดังที่แบบจำลองกำหนด

R.L. Hsu [12] ได้เสนอการตรวจหาใบหน้าจากรูปภาพดิจิทัลสี ซึ่งขั้นตอนวิธีที่ใช้ในการ กำหนดตำแหน่งและตรวจหาคุณลักษณะสำคัญบนใบหน้าของ R.L. Hsu แสดงในรูปที่ 2.11 ชั้นแรก รูปภาพรับเข้าจะผ่านกระบวนการขจัดแสงและแปลงไปสู่ปริภูมิสีย่อย CbCr จากนั้นทำการแยก สีผิวโดยใช้ แบบจำลองวงรีบนระนาบ 2 มิติ (แบบเกาส์) ถัดมาทำการกำจัดสัญญาณรบกวนและ วิเคราะห์บริเวณเชื่อมต่อกัน ซึ่งผลลัพธ์ที่ได้ควรจะเป็นตำแหน่งบริเวณใบหน้า เพื่อให้แน่ใจว่าบริเวณที่ ได้มาเป็นใบหน้าจริงจะทำการหาคุณลักษณะที่สำคัญบนใบหน้าเพิ่มเติม โดยคำนึงถึงองค์ประกอบ หลัก 2 ส่วน คือ ปากและตา อาศัยสมมติฐานที่ว่า ส่วนตาจะเป็นส่วนที่มีค่า Cb สูง แต่ Cr ต่ำ และ ส่วนปากจะเป็นส่วนที่มีค่าความเข้มของสีแดงที่มากกว่าบริเวณผิวอื่น ๆ



รูปที่ 2.10 ผลการแยกสีผิวและตรวจหาใบหน้าของ R.L. Hsu [12]

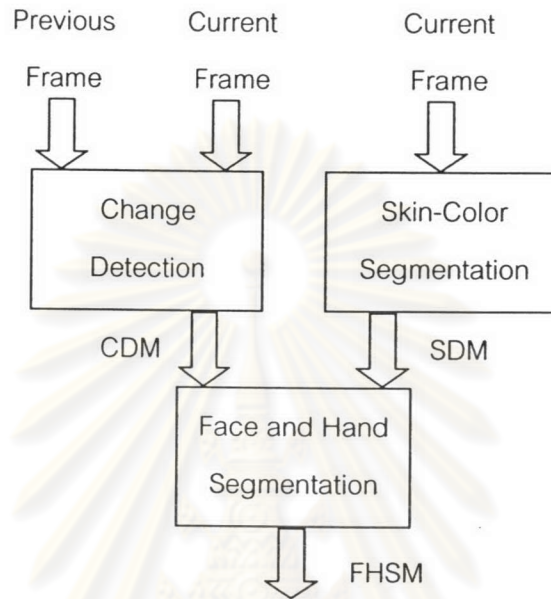


รูปที่ 2.11 ขั้นตอนวิธีการตรวจหาใบหน้าที่เสนอโดย R.L. Hsu [12]

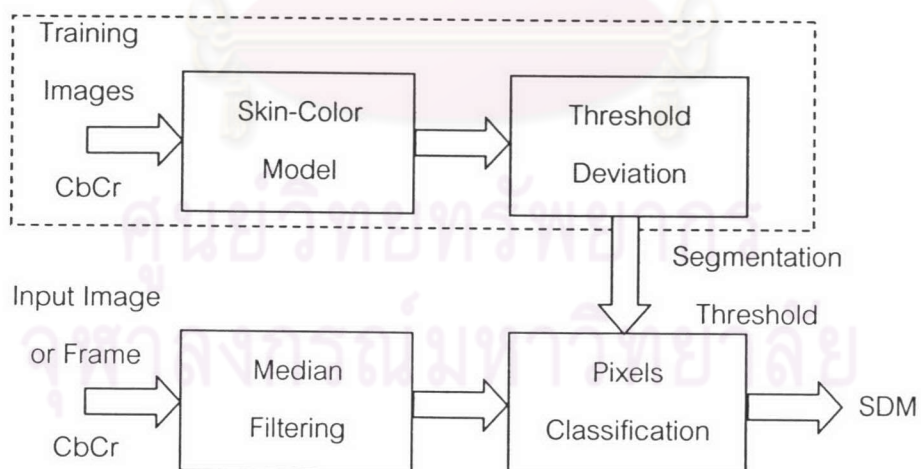
N. Habili [11] เสนอการแยกใบหน้าและมือจากวิดีโอทัศนภาษามือ โดยใช้คุณลักษณะของสีและเวกเตอร์การเคลื่อนที่ ซึ่งขั้นตอนวิธีที่นำมาใช้แสดงในรูปที่ 2.12 ในส่วนคุณลักษณะของสี ใช้การแยกสีผิวเมื่อมีรูปภาพเฟรมปัจจุบันเข้ามา กระบวนการนี้จะสร้างมาสก์ที่ได้จากการแยกสีผิวขึ้น (SDM: Skin Detection Mask) สำหรับการพิจารณาเวกเตอร์การเคลื่อนที่ที่จะกระทำระหว่างสองเฟรมที่อยู่ติดกันและสร้างมาสก์ตรวจหาการเปลี่ยนแปลง (CDM: Change Detection Mask) โดยขั้นตอนวิธีที่นำมาใช้ จะพิจารณาสองส่วนนี้ร่วมกัน เพื่อสร้างเป็นมาสก์ของการแยกใบหน้าและมือ (FHSM: Face and Hand Segmentation Mask)

ขั้นตอนวิธีในการแยกส่วนสีผิวดังแสดงในรูปที่ 2.13 ขั้นตอนแรกรูปภาพในฐานะข้อมูลจะถูกนำมาใช้ในการสร้างแบบจำลองเพื่อกำหนดขีดแบ่ง (threshold) ที่เหมาะสมสำหรับการแยกสีผิว โดยกระทำในปริภูมิสีย่อย CbCr จากนั้นจะทำการบันทึกค่าต่าง ๆ ที่จำเป็นของแบบจำลองไว้ เมื่อมีรูปภาพรับเข้าใด ๆ ที่ต้องการแยกส่วนสีผิว รูปภาพนี้จะถูกแปลงไปสู่ค่า CbCr ก่อน จากนั้นจึงนำมา

ผ่านตัวกรองความถี่แถบกลาง (Median Filter) แล้วถึงทำการกำหนดว่าจุดภาพใดเป็นสีผิว โดยอ้างอิงจากค่าต่าง ๆ ของแบบจำลองที่บันทึกไว้



รูปที่ 2.12 ขั้นตอนวิธีในการแยกใบหน้าและมือที่เสนอโดย N. Habili [11]



รูปที่ 2.13 ขั้นตอนวิธีในการแยกสีผิวที่เสนอโดย N. Habili [11]

2.1.5 ปริภูมิสี L*a*b

ในปริภูมิสี L*a*b ค่า L คือความองค์ประกอบความสว่าง ซึ่งวัดค่าเป็นเปอร์เซ็นต์จาก 0 (ความมืด) ถึง 100 (ความสว่าง) a แสดงความแตกต่างระหว่างสีเขียว (+a) และสีแดง (-a) ส่วน b แสดงความแตกต่างระหว่างสีเหลือง (+b) และสีน้ำเงิน (-b)

การแปลงค่าไปจากปริภูมิสีนอร์มัลไลซ์ RGB ไปสู่ปริภูมิสี L*a*b สามารถคำนวณค่า a และ b ได้จากสมการที่ (2.24) และ (2.25)

$$a = r + \frac{g}{2} \quad (2.24)$$

$$b = \frac{\sqrt{3}}{2} g \quad (2.25)$$

โดยที่ r และ g ได้จากการนอร์มัลไลซ์ ค่า R และ G จากปริภูมิสี RGB

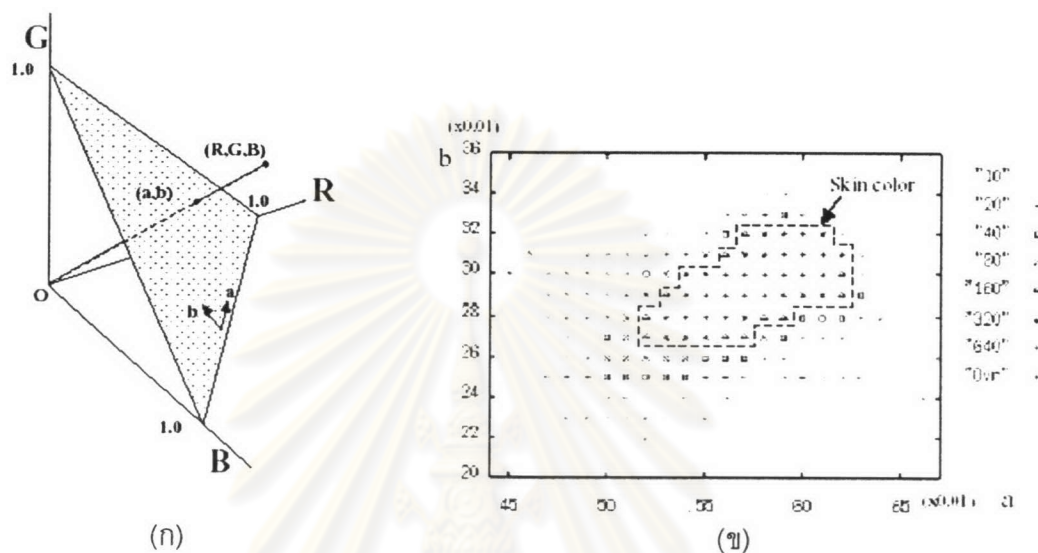


รูปที่ 2.14 ปริภูมิสี L*a*b

งานวิจัยที่เกี่ยวข้อง

S. Kawato [16] เสนอวิธีแยกสีผิวอัตโนมัติสำหรับการตรวจหาและติดตามใบหน้า ซึ่งนำเสนอปริภูมิสีสำหรับการแยกส่วนสีผิวที่แตกต่างออกไป โดยปริภูมิสีที่นำมาใช้ คือ ปริภูมิสี L*a*b โดยละทิ้งองค์ประกอบความสว่าง L พิจารณาเฉพาะค่า a และ b ข้อด้อยที่เกิดขึ้น คือ เนื่องจากปริภูมิสี L*a*b

ค่า a และ b ได้จากการแปลงนอร์มัลไลซ์ RG ดังนั้นจึงมีค่ารบกวนที่เกิดขึ้นจากการตีความผิดพลาด บริเวณตำแหน่งจุดภาพที่มีค่าใกล้เคียงกับสีดำเช่นกัน ในรูปที่ 2.15 (ก) แสดงขอบเขตของปริภูมิสีย่อย ab และ ในรูปที่ 2.15 (ข) แสดงอาณาบริเวณที่ถูกกำหนดให้เป็นสีผิว



รูปที่ 2.15 (ก) ขอบเขตของปริภูมิสีย่อย ab เมื่อเทียบกับปริภูมิสี RGB

(ข) การกระจายตัวของจุดภาพที่เป็นสีผิวในปริภูมิสีย่อย ab [16]

หัวข้อที่ผ่านมาได้แสดงวิธีที่ใช้ในการแปลงค่าจากปริภูมิสีพื้นฐาน RGB ไปสู่ RGB นอร์มัลไลซ์, HSI, YCbCr และ L^*a^*b ตามลำดับ พร้อมทั้งอธิบายความหมายของอักษรย่อแต่ละตัว ในปริภูมิสีเหล่านั้น ซึ่งในความเป็นจริงแล้วยังมีปริภูมิสีอีกมากมายที่ยังคงเป็นข้อถกเถียงในงานวิจัยแต่ไม่นำมากล่าวถึง เนื่องจากคัดเลือกมาเฉพาะปริภูมิสีที่ได้รับความนิยมในงานวิจัยและมีการยอมรับกันอย่างแพร่หลายเท่านั้น

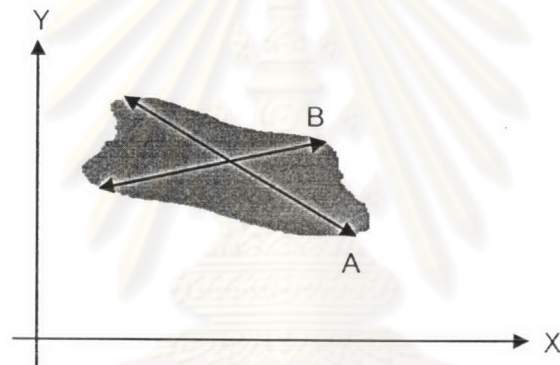
งานวิจัยที่เปรียบเทียบประสิทธิภาพการแยกสีผิวในแต่ละปริภูมิสี

B. D. Zarit [17] เสนอกรรมวิธีในการเปรียบเทียบผลการแยกส่วนสีผิวใน 5 ปริภูมิสี คือ CIE- L^*a^*b , Fleck HS, HSV, Normalized RGB และ YCbCr โดยกรรมวิธีในการแยกสีผิวที่นำมาใช้มี 2 แบบวิธีที่หนึ่งทำการเปรียบเทียบค่าจากตาราง (Lookup Table Method) และ วิธีที่สองคือวิธีของเบย์ (Bayesian Method) ส่วน J.C. Terrillon [18] ทำการทดสอบใน 9 ปริภูมิสี คือ TSL, rg, CIE-xy, CIE-DSH, HSV, YIQ, YES, CIE- L^*u^*v และ CIE- L^*a^*b โดยแบบจำลองที่ใช้เป็นแบบเกาส์

สำหรับในส่วนถัดไป ตั้งแต่หัวข้อ 2.2 – 2.4 จะกล่าวถึงทฤษฎีของการประมวลผลภาพดิจิทัลที่นำมาใช้ในงานวิจัยนี้

2.2 การหาความเยื้องศูนย์กลาง (Eccentricity)

กรรมวิธีการหาความเยื้องศูนย์กลางนี้ จะถูกนำไปใช้ในส่วนการสร้างแบบจำลองวงรีเพื่อแยกสีผิว หลักการของกรรมวิธี คือ หากมีรูปภาพลักษณะที่ไม่สมมาตรรูปร่างใด ๆ ก็ตาม การหาจุด A ที่มีความยาวสูงสุด และ จุด B ซึ่งตั้งฉากกัน สามารถทำได้โดยพิจารณาความเยื้องศูนย์กลาง ซึ่งมีวิธีการคือใช้ค่าตำแหน่งในระนาบ XY ของรูปร่างที่ต้องการพิจารณาแปลงไปสู่ระนาบ $\alpha\beta$ ดังต่อไปนี้



รูปที่ 2.16 รูปร่างที่จะหาความเยื้องศูนย์กลาง

การแปลงแต่ละคู่อันดับ (x, y) ไปสู่ (α, β) ทำโดย

$$\begin{bmatrix} \alpha(x, y) \\ \beta(x, y) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.26)$$

โดยที่ (α, β) คือ คู่อันดับใหม่

θ คือ ทิศทางของคอร์ดที่ยาวที่สุด

ในการหาค่าต่ำสุดและสูงสุดของ α และ β จะทำเพื่อกำหนดขอบเขตของความยาว (I_1) และความกว้าง (I_2) ที่ล้อมรอบรูปร่างนั้น จะใช้สมการที่ (2.27) และ (2.28)

$$l_1 = \alpha_{\max} - \alpha_{\min} \quad (2.27)$$

$$l_2 = \beta_{\max} - \beta_{\min} \quad (2.28)$$

การหาทิศทางของคอร์ดที่ยาวที่สุด (θ) สามารถหาได้จากสมการที่ (2.29)

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{11}}{\mu_{20} + \mu_{02}} \right) \quad (2.29)$$

ซึ่งสามารถหา μ_{pq} ได้จากสมการที่ (2.30)

$$\mu_{pq} = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (i - \bar{x})^p (j - \bar{y})^q f(i, j) \quad (2.30)$$

โดยที่

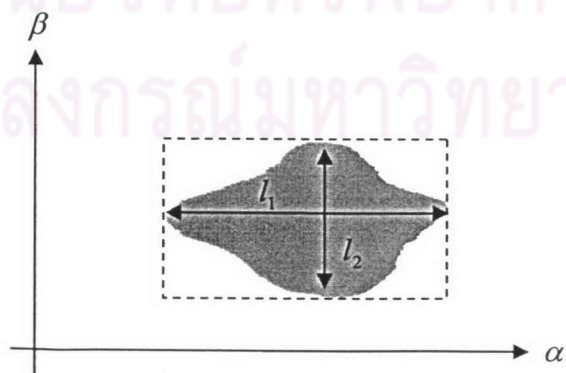
$$\bar{x} = \frac{m_{10}}{m_{00}} \quad (2.31)$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \quad (2.32)$$

และสามารถคำนวณ m_{pq} ได้จากสมการที่ (2.33)

$$m_{pq} = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} i^p j^q f(i, j) \quad (2.33)$$

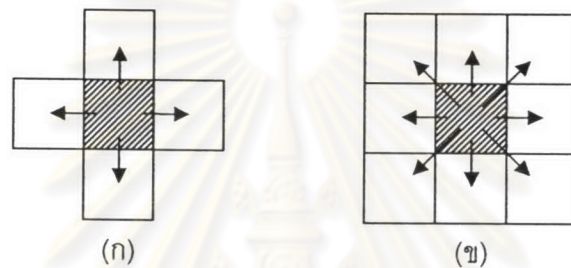
หากพิจารณาคู่อันดับ (\bar{x}, \bar{y}) แท้จริงแล้วก็คือจุดศูนย์กลางของรูปร่างนั้นนั่นเอง ในรูปที่ 2.17 คือ ค่าที่ถูกแปลงไปสู่แกน α และ β แล้ว และสามารถกำหนดขอบเขตล้อมรอบจากค่า l_1 และ l_2



รูปที่ 2.17 แปลงไปสู่ระนาบ $\alpha\beta$

2.3 การวิเคราะห์ส่วนประกอบที่เชื่อมต่อกัน (Connected Component Analysis)

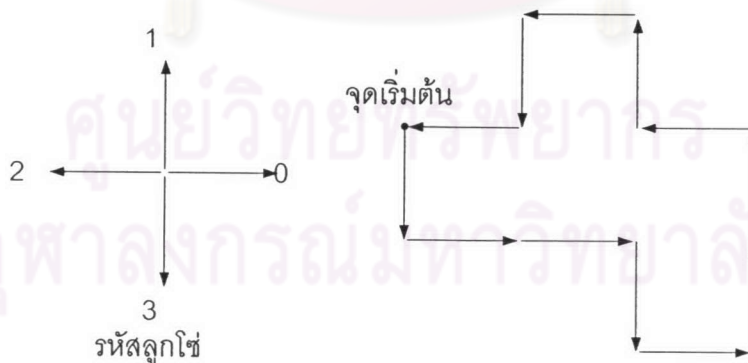
ในการกำหนดขอบเขตและอธิบายรูปร่างของแต่ละวัตถุในรูปภาพฐานสอง วิธีการหนึ่งที่ใช้คือวิธีแบบรหัสลูกโซ่ (Chain code) ซึ่งจะมีการกำหนดรหัสไว้ก่อนล่วงหน้า โดยแต่ละรหัสจะแสดงถึงการเปลี่ยนแปลงทิศทางของจุดภาพซึ่งเป็นค่าเดียวกันในบริเวณข้างเคียง และเมื่อเราทำการเข้ารหัสรูปภาพฐานสอง จะสามารถพิจารณาได้ว่ามีจุดภาพกลุ่มใดบ้างที่เป็นวัตถุเดียวกันจากความสัมพันธ์ของรหัส การกำหนดรหัสจะมีทั้งแบบพิจารณาจุดภาพข้างเคียง 4 จุดเชื่อมต่อ และ แบบ 8 จุดเชื่อมต่อ



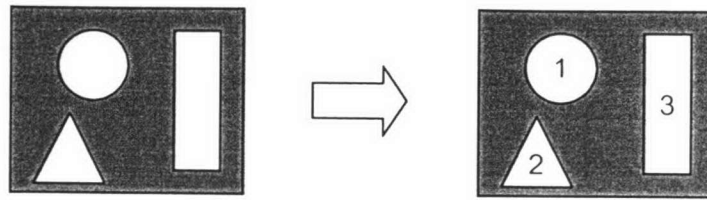
รูปที่ 2.18 การพิจารณาจุดภาพข้างเคียง (ก) แบบ 4 จุดเชื่อมต่อ (ข) แบบ 8 จุดเชื่อมต่อ

สำหรับค่าอนุพันธ์จะนำมาใช้ในการอธิบายรูปร่างของวัตถุ ซึ่งหาได้จากการเปลี่ยนแปลงการหมุนของลูกศรเป็นทิศ 90 หรือ 45 องศา และ ตามเข็มนาฬิกา หรือ ทวนเข็มนาฬิกา ดังแสดงในรูปที่ 2.19

ตัวอย่าง การเข้ารหัสลูกโซ่แบบ 4 จุดเชื่อมต่อ



รูปที่ 2.19 การเข้ารหัสลูกโซ่แบบ 4 จุดเชื่อมต่อ โดยมีรหัสคือ 3,0,0,3,0,1,1,2,1,2,3,2 และค่าอนุพันธ์เป็น 1,0,3,1,1,0,1,3,1,1,3,1



รูปที่ 2.20 กำหนดกลุ่มของจุดภาพที่เป็นวัตถุเดียวกันหรือการทำป้าย (Labeling)

การวิเคราะห์ส่วนประกอบที่เชื่อมต่อกันนี้ จะกระทำหลังจากแยกส่วนสีผิวและกำจัดสัญญาณรบกวนในลักษณะหลุมขนาดเล็กแล้ว โดยกรรมวิธีนี้จะทำให้ทราบถึงตำแหน่งและขนาดของแต่ละบริเวณ ซึ่งสามารถนำข้อมูลเหล่านี้ไปพิจารณาได้ว่ากลุ่มจุดภาพบริเวณนั้นเป็นใบหน้า มือ หรือไม่ใช่ทั้งสองอย่าง

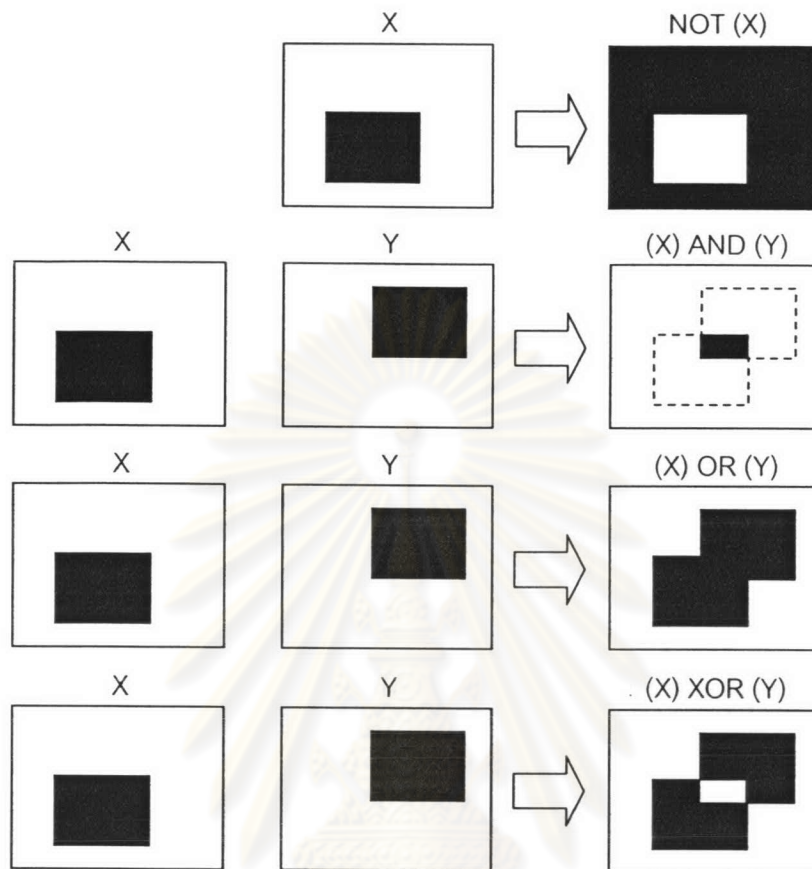
2.4 การประมวลผลภาพเชิงสัณฐาน (Morphological Image Processing)

2.4.1 การดำเนินการตรรกะกับรูปภาพฐานสอง

หลักของการดำเนินการตรรกะที่ถูกนำมาใช้ในการประมวลผลภาพ คือ AND, OR, XOR และ NOT ซึ่งจะกระทำบนฐานหลักของจุดภาพระหว่าง 2 รูปภาพหรือมากกว่านั้น (ยกเว้น NOT ที่เป็นตัวดำเนินการของรูปภาพเพียงรูปเดียว) สำหรับการประมวลผลที่ซับซ้อนกว่านี้ เมื่อพิจารณาอย่างถ่องแท้แล้วจะพบว่าเกิดจากตัวดำเนินการหลัก 4 ตัวนี้เรียงลำดับอยู่ในขั้นตอนเช่นกัน

ตารางที่ 2.1 การดำเนินการตรรกะพื้นฐาน 4 ตัวดำเนินการ

P	Q	P AND Q	P OR Q	P XOR Q	NOT(P)
0	0	0	0	0	1
0	1	0	1	1	1
1	0	0	1	1	0
1	1	1	1	0	0



รูปที่ 2. 21 ตัวอย่างของการดำเนินการตรรกะระหว่างรูปภาพฐานสอง

2.4.2 การทำไดเลชัน (Dilation) และ การทำอีโรชัน (Erosion)

รูปภาพหรือกลุ่มของจุดภาพสีดำและขาว ซึ่งมีลักษณะเป็นรูปภาพฐานสอง (Binary image) การทำไดเลชันและอีโรชัน ถือเป็น ตัวดำเนินการการเชิงสัญญาณเบื้องต้น และเป็นพื้นฐานในการนำไปใช้กับกระบวนการที่มีความซับซ้อนมากขึ้น ได้แก่ ตัวดำเนินการแบบเปิด (Opening operator), ตัวดำเนินการแบบปิด (Closing operator) และการแยกรูปร่าง (Shape decomposition)

การทำไดเลชัน

นิยาม คือ เซตของจุด p ที่เกิดจาก $x+b$ โดยที่ x เป็นสมาชิกของ X และ b เป็นสมาชิกของ B ดังสมการที่ (2.34)

$$X \oplus B = \{p \in \varepsilon^2 : p = x + b, x \in X \& b \in B\} \quad (2.34)$$

โดยที่ X คือ รูปภาพที่ทำการไคเลชัน

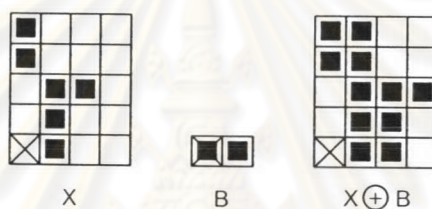
B คือ โครงสร้างส่วนย่อยในการทำไคเลชัน

ตัวอย่าง การทำไคเลชัน รูปที่ 2.22

กำหนดให้ $X = \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4)\}$

$$B = \{(0,0), (1,0)\}$$

$$X \oplus B = \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4), (2,0), (2,1), (2,2), (3,2), (1,3), (1,4)\}$$



รูปที่ 2.22 การทำไคเลชัน

การทำอีโรชัน

นิยาม คือ เซตของจุด p ทุกจุดจากรูปภาพทดสอบ ผลลัพธ์ของการทำอีโรชันคือ จุด p เหล่านั้นที่ทำให้ $p+b$ เป็นสมาชิกของ X สำหรับทุก b ที่เป็นสมาชิกของ B แสดงในสมการที่ (2.35)

$$X \ominus B = \{p \in \varepsilon^2 : p + b \in X \text{ for every } b \in B\} \quad (2.35)$$

โดยที่ X คือ รูปภาพที่ทำการอีโรชัน

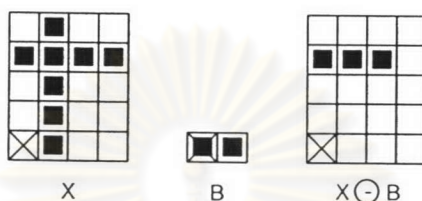
B คือ โครงสร้างส่วนย่อยในการทำอีโรชัน

ตัวอย่าง การทำอิโรชัน รูปที่ 2.23

กำหนดให้ $X = \{(1,0), (1,1), (1,2), (0,3), (1,3), (2,3), (3,3), (1,4)\}$

$$B = \{(0,0), (1,0)\}$$

$$X \ominus B = \{(0,3), (1,3), (2,3)\}$$



รูปที่ 2.23 การทำอิโรชัน

2.4.3 ตัวดำเนินการแบบเปิดและปิด

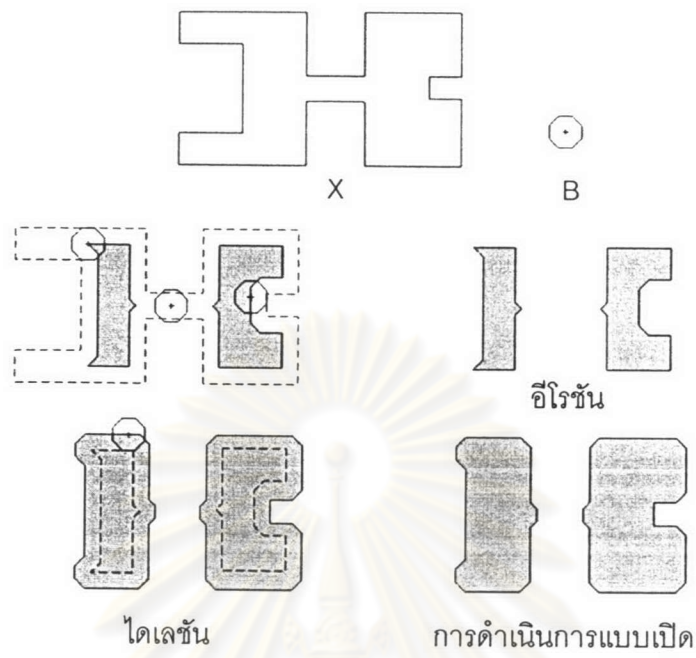
การทำอิโรชันและไดเลชันไม่ถือเป็นการแปลงผกผัน (Inverse transformation) ระหว่างกัน ดังนั้น รูปภาพที่นำมาผ่านการอิโรชันและไดเลชัน ไม่จำเป็นต้องเท่ากับรูปภาพดั้งเดิม แต่จะได้ผลลัพธ์ซึ่งมีรายละเอียดซับซ้อนน้อยกว่ารูปภาพดั้งเดิมแทน

กระบวนการอิโรชันแล้วตามด้วยไดเลชัน เรียกว่า ตัวดำเนินการแบบเปิด ดังแสดงในสมการที่ (2.36)

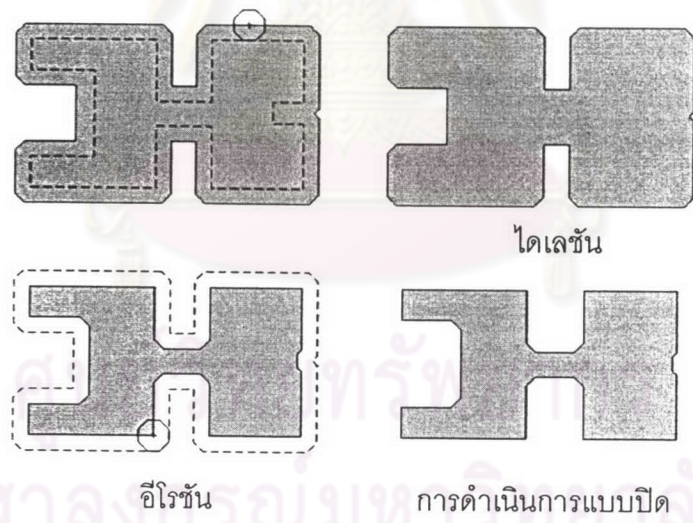
$$X \circ B = (X \ominus B) \oplus B \quad (2.36)$$

กระบวนการไดเลชันแล้วตามด้วยอิโรชัน เรียกว่า ตัวดำเนินการแบบปิด ดังแสดงในสมการที่ (2.37)

$$X \bullet B = (X \oplus B) \ominus B \quad (2.37)$$



รูปที่ 2.24 ตัวดำเนินการแบบเปิด



รูปที่ 2.25 ตัวดำเนินการแบบปิด

2.4.4 ความเป็นวงกลม (Circularity)

การวัดความเป็นวงกลมของแต่ละบริเวณในรูปภาพฐานสองใด ๆ สามารถพิจารณาได้จากอัตราส่วนระหว่างค่าระยะทางรัศมีเฉลี่ย (mean radial distance : μ_R) กับค่าเบี่ยงเบนมาตรฐานของระยะทางรัศมี (standard deviation of radial distance : σ_R) ตามสมการที่ (2.38)

$$C = \frac{\mu_R}{\sigma_R} \quad (2.38)$$

ซึ่งสามารถหา μ_R และ σ_R ได้จากสมการที่ (2.39) และ (2.40) ตามลำดับ

$$\mu_R = \frac{1}{K} \sum_{k=0}^{K-1} \|(x_k, y_k) - (\bar{x}, \bar{y})\| \quad (2.39)$$

$$\sigma_R = \left(\frac{1}{K} \sum_{k=0}^{K-1} \left[\|(x_k, y_k) - (\bar{x}, \bar{y})\| - \mu_R \right]^2 \right)^{1/2} \quad (2.40)$$

เมื่อ (\bar{x}, \bar{y}) คือ จุดศูนย์กลาง (centre of gravity) ของรูปร่าง

(x_k, y_k) คือ ตำแหน่งของจุดภาพบริเวณที่ต้องการวัดความเป็นวงกลม

ตัวอย่าง จากการคำนวณหาค่า μ_R และ σ_R^2 ของแต่ละบริเวณในรูปที่ 2.26 พบว่าได้ค่าต่าง ๆ ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 คุณสมบัติความเป็นวงกลมของทั้ง 3 บริเวณ

บริเวณที่	\bar{x}	\bar{y}	μ_R	σ_R^2	C
1	6	11.5	3.33	0.05	15.4
2	9	1.5	3.80	2.28	2.5
2	13	7	1.2	0.04	5.8

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0
2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0
2	2	2	2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

รูปที่ 2.26 รูปภาพฐานสองที่ต้องการวัดความเป็นวงกลมในแต่ละบริเวณ

2.4.5 การแปลงแบบฮิตหรือมิส (Hit-or-miss transformation)

การแปลงแบบฮิตหรือมิส ถือเป็นตัวกระทำเชิงสัจฐานที่ใช้ในการตรวจหารูปร่าง เพื่อใช้ในการหารูปแบบกลุ่มของจุดภาพเฉพาะบริเวณ ซึ่งจะมีการเปลี่ยนแปลง เช่น สนวนมุม หรือ จุดที่เป็นขอบ โดยที่สัญลักษณ์ของการแปลงแบบฮิตหรือมิส จะแทนด้วย \odot ดังแสดงในสมการที่ (2.41)

$$X \odot B = (X \ominus B_1) \cap (X \ominus B_2) \quad (2.41)$$

การแปลงแบบฮิตหรือมิต คือ การนำเอารูปภาพฐานสองทำอีโรชันกับโครงสร้างส่วนย่อย B_1 อินเตอร์เซกชันกับ รูปภาพซึ่งผ่านการคอมพลีเมนต์ (Complement) ทำอีโรชันกับโครงสร้างส่วนย่อย B_2 โดยที่ $B_1 \cap B_2 = \phi$

ตัวอย่าง กำหนดให้ $B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, $B_1 = \begin{bmatrix} - & - & - \\ - & 1 & - \\ - & - & - \end{bmatrix}$ และ $B_2 = \begin{bmatrix} - & 1 & - \\ 1 & - & 1 \\ - & 1 & - \end{bmatrix}$

โดยที่ตำแหน่งที่มีเครื่องหมาย "-" เป็นค่าใดก็ได้ จะได้ผลการแปลงแบบฮิตหรือมิต ดังรูปที่ 2.27

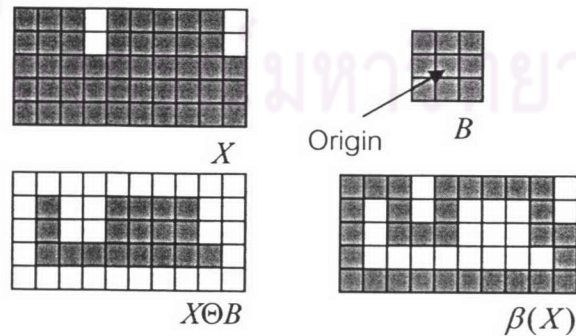


รูปที่ 2.27 การแปลงแบบฮิตหรือมิต

2.4.6 การสกัดขอบ (Boundary Extraction)

ขอบของรูปภาพ X จะถูกแสดงด้วย $\beta(X)$ สามารถหาได้จากการทำอีโรชัน X ด้วย B และ หาผลต่างระหว่าง X กับผลลัพธ์ที่ได้จากการอีโรชัน ดังแสดงในสมการที่ (2.42)

$$\beta(X) = X - X \ominus B \tag{2.42}$$

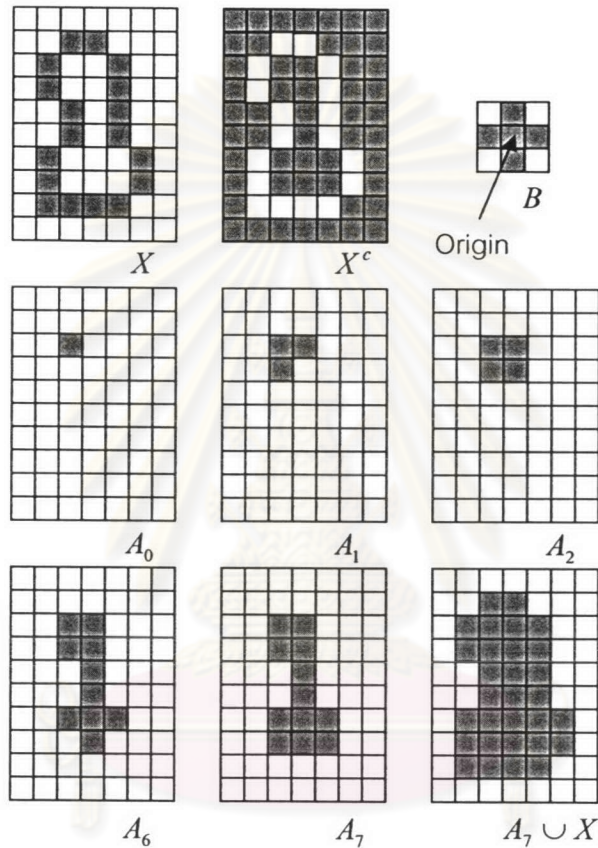


รูปที่ 2.28 การสกัดขอบ

2.4.7 การเติมเต็มลงในบริเวณที่ถูกล้อมรอบ (Region filling)

วิธีการในการเติมเต็มจุดภาพลงในบริเวณที่ถูกล้อมรอบจะกระทำโดยอาศัยตัวดำเนินการไคเลชัน, คอมพลีเมนต์ และ อินเตอร์เซกชัน ตามสมการที่ (2.43)

$$A_k = (A_{k-1} \oplus B) \cap X^c \quad k = 1, 2, 3, \dots \tag{2.43}$$



รูปที่ 2.29 การเติมเต็มบริเวณที่ถูกล้อมรอบ

2.4.8 กระบวนการธinning (Thinning)

คือ การนำรูปภาพไปลบกับรูปที่ผ่านการแปลงแบบฮิตหรือมิต ตัวดำเนินการในการธinningจะแทนด้วย \otimes โดยมีสมการในการธinningแสดงดังสมการที่ (2.44)

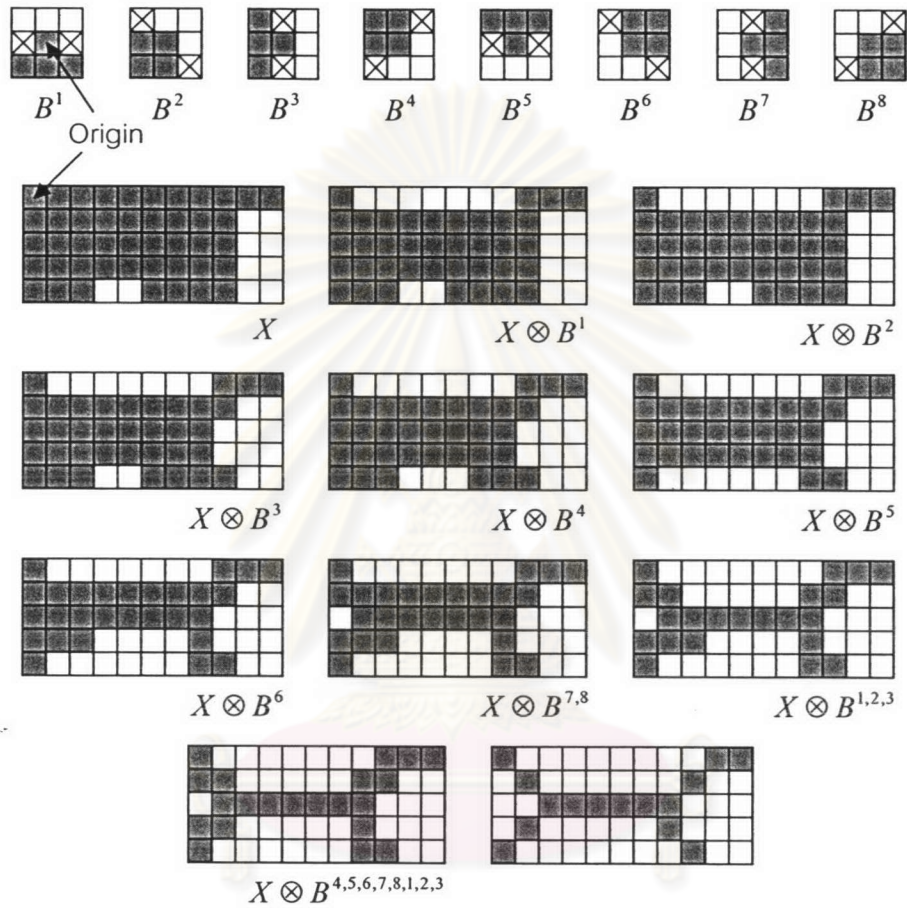
$$X \otimes B = X - (X \circledast B) \tag{2.44}$$

โดยที่ B คือ

$$\{B\} = \{B^1, B^2, B^3, \dots, B^n\} \tag{2.45}$$

และสามารถหา B^i ได้จากการหมุนจาก B^{i-1} ดังรูปที่ 2.30

$$X \otimes \{B\} = (((((X \otimes B^1) \otimes B^2) \dots) \otimes B^n) \quad (2.46)$$



รูปที่ 2.30 กระบวนการอินนิง

2.5 การตรวจหาขอบ (Edge Detection)

การตรวจหาขอบ ถือเป็นกระบวนการประมวลผลภาพที่สำคัญมากอันหนึ่ง ถูกนำมาใช้ในการวิเคราะห์รูปภาพเพื่อกำหนดขอบเขตของบริเวณที่ต้องการ ซึ่งกระบวนการนี้จะกระทำโดยตรวจสอบว่าจุดภาพจุดใดที่มีการเปลี่ยนแปลงขององค์ประกอบทางความสว่างอย่างทันทีทันใดเมื่อเปรียบเทียบกับบริเวณใกล้เคียงจะกำหนดจุดภาพนั้นเป็นขอบ

วิธีการที่ใช้หากการเปลี่ยนแปลงของฟังก์ชันต่อเนื่อง คือ การหาอนุพันธ์ แต่ในฟังก์ชันของรูปภาพจะขึ้นกับ 2 ตัวแปรในระบบพิกัดระนาบรูปภาพ ดังนั้นวิธีการที่นำมาใช้ คือ การหาอนุพันธ์ย่อย โดยการเปลี่ยนแปลงของฟังก์ชันรูปภาพสามารถอธิบายด้วยค่าเกรเดียนต์ในทิศทางสูงสุด

ขอบ คือ คุณสมบัติที่เกี่ยวกับจุดภาพใดจุดภาพหนึ่งซึ่งนำไปคำนวณร่วมกับพฤติกรรมของจุดภาพรอบข้าง และประกอบไปด้วยตัวแปรชนิดเวกเตอร์ 2 ตัวแปร คือ ขนาดและทิศทาง ขนาดของขอบสามารถหาได้จากขนาดของการเกรเดียนต์ และทิศทางของขอบถูกหมุนในส่วนทิศเกรเดียนต์ ซึ่งมีค่าสูงสุด จากจุดภาพสีดำ [$f(i,j) = 0$] ไปจุดภาพสีขาว [$f(i,j) = 255$] ในรูปที่ 2.31 แสดงแบบจำลองขอบในอุดมคติซึ่งมีการเปลี่ยนแปลงอย่างทันทีทันใด และ ขอบที่เปลี่ยนลักษณะทางลาดซึ่งจะมีการเปลี่ยนแปลงในลักษณะไล่ระดับ



รูปที่ 2.31 แบบจำลองขอบ

การตรวจหาขอบด้วยค่าเกรเดียนต์

อนุพันธ์อันดับหนึ่งของรูปภาพดิจิทัล จะหาได้จากการประมาณของเกรเดียนต์ใน 2 มิติ ค่าเกรเดียนต์ของรูปภาพ $f(x,y)$ ที่ตำแหน่ง (x,y) ถูกกำหนดด้วยเวกเตอร์ ∇f ดังแสดงในสมการที่ (2.47)

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.47)$$

ในการวิเคราะห์เวกเตอร์ก็คือจุดของเวกเตอร์เกรเดียนต์ ซึ่งมีอัตราการเปลี่ยนแปลงสูงสุดของทิศทางของ f ที่พิกัด (x,y) โดยมีค่าสำคัญที่นำมาใช้ตรวจหาขอบ คือ ค่าขนาดของเวกเตอร์ แทนด้วย ∇f และสามารถคำนวณได้จากสมการที่ (2.48)

$$\nabla f = \text{mag}(\nabla f) = [G_x^2 + G_y^2]^{1/2} \quad (2.48)$$

ส่วนค่าทิศทางคำนวณได้จาก

$$\psi = \tan^{-1} \left(\frac{G_x}{G_y} \right) \quad (2.49)$$

ψ คือ มุมที่วัดจากแกน x ไปยังจุด $f(x,y)$ มีหน่วยเป็นเรเดียน

การคำนวณค่าเกรเดียนต์ของรูปภาพด้วยอนุพันธ์ย่อยในทุก ๆ จุดภาพ ทำได้โดย ให้อัปเดตขนาด 3×3 (รูปที่ 2.32) แสดงค่าระดับสีเทาที่จุด Z_5 ของจุดภาพและบริเวณใกล้เคียง อนุพันธ์อันดับหนึ่งที่จุด Z_5 สามารถหาได้ด้วยตัวดำเนินการที่ต่าง ๆ กัน ดังแสดงในสมการที่ (2.50)-(2.55)

- ตัวดำเนินการไขว้เกรเดียนต์แบบโรเบิร์ต (Roberts cross-gradient operators)

$$G_x = (z_9 - z_5) \quad (2.50)$$

$$G_y = (z_8 - z_6) \quad (2.51)$$

- ตัวดำเนินการแบบพรีวิตต์ (Prewitt operators)

$$G_x = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3) \quad (2.52)$$

$$G_y = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7) \quad (2.53)$$

- ตัวดำเนินการแบบโซเบล (Sobel operators)

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \quad (2.54)$$

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7) \quad (2.55)$$

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6
Z_7	Z_8	Z_9

บล็อกขนาด 3x3

$$G_x = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Roberts

$$G_x = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

Prewitt

$$G_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Sobel

รูปที่ 2.32 มาสก์ต่าง ๆ ที่ใช้ในการคำนวณหาค่าเกรเดียนต์ที่จุด Z_5

(2.56) การหาด้วยวิธีไขว้ศูนย์ (Zero crossing) หรืออนุพันธ์อันดับสอง สามารถทำได้โดยใช้สมการที่

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2.56)$$

2.6 การวัดประสิทธิภาพของแบบจำลอง

ตัวแปรสำคัญที่นำมาใช้สำหรับการวัดประสิทธิภาพแบบจำลองเกี่ยวกับงานด้านคอมพิวเตอร์วิทัศน์ (Computer vision) จะมีอยู่ด้วยกัน 4 ตัวแปร การอธิบายความหมายของแต่ละตัวแปรแสดงในรูปที่ 2.34 โดยสมมติว่า ณ ที่นี้ เราสนใจใบหน้ามนุษย์ จะกำหนดให้เป็นฉากหน้า (Foreground) ดังนั้นส่วนที่ไม่ใช่ใบหน้าจะถูกกำหนดเป็นฉากหลัง (Background) ทั้งหมด

- TP ย่อจาก "True Positive" มีหน่วยเป็น จุดภาพ
- TN ย่อจาก "True Negative" มีหน่วยเป็น จุดภาพ
- FP ย่อจาก "False Positive" มีหน่วยเป็น จุดภาพ
- FN ย่อจาก "False Negative" มีหน่วยเป็น จุดภาพ

ค่า False Positive Rate คือ อัตราส่วนระหว่าง FP และ TN รวมกับ FP สามารถคำนวณจากสมการที่ (2.57)

$$\text{False Positive Rate} = \frac{FP}{TN + FP} \quad (2.57)$$

ค่า False Negative Rate คือ อัตราส่วนระหว่าง FN และ FN รวมกับ TP สามารถคำนวณจากสมการที่ (2.58)

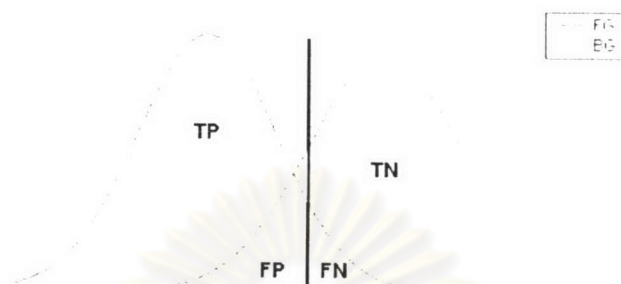
$$\text{False Negative Rate} = \frac{FN}{FN + TP} \quad (2.58)$$

ค่า DR ย่อจาก "Detection Rate" หรือบางครั้งเรียกว่า True Positive rate คือ อัตราความถูกต้องในการตรวจหา มีหน่วยเป็น (%) เปอร์เซนต์ ซึ่งแบบจำลองที่ให้ค่า DR มากกว่าแสดงถึงความถูกต้องในการตรวจหาดีกว่า สามารถหาได้จากสมการที่ (2.59)

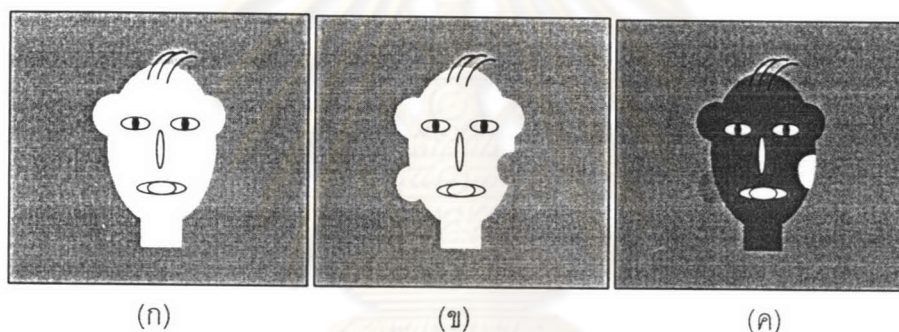
$$DR = \frac{TP}{TP + FN} \times 100\% \quad (2.59)$$

ค่า FAR ย่อจาก "False Alarm Rate" คือ อัตราการเกิดความผิดพลาด มีหน่วยเป็น (%) เปอร์เซนต์ โดยแบบจำลองที่มีประสิทธิภาพมากกว่าจะมีค่า FAR ต่ำกว่า สามารถหาได้จากสมการที่ (2.60)

$$FAR = \frac{FP}{TP + FP} \times 100\% \quad (2.60)$$



รูปที่ 2.33 เกณฑ์ในการแยกค่าตัวแปรทั้ง 4



รูปที่ 2.34 ความหมายของตัวแปรทั้ง 4 ตัวแปร สามารถอธิบายได้ดังนี้

- (ก) ภาพอ้างอิงซึ่งได้จากการแยกส่วนใบหน้าด้วยสายตามนุษย์
- (ข) ผลลัพธ์ที่ได้จากแบบจำลองตรวจหาใบหน้า

(ค) ■ TP (ส่วนบริเวณใบหน้าที่ตรวจหาได้ถูกต้อง)

■ TN (ส่วนฉากหลังที่ตรวจหาได้ถูกต้อง)

■ FP (ส่วนฉากหลังแต่แบบจำลองให้ผลเป็นค่าบริเวณใบหน้า)

■ FN (ส่วนบริเวณใบหน้าแต่แบบจำลองให้ค่าเป็นฉากหลัง)

การวัดประสิทธิภาพโดยพิจารณาจากกราฟ ROC (Receiver Operating Characteristic) แสดงในรูปที่ 2.35 โดยพิจารณาค่าในแกนนอน คือ ค่า False Positive Rate เทียบกับค่าในแกนตั้ง คือ ค่า True Positive Rate หากแนวของเส้นโค้งนี้ทำมุมใกล้เคียงกับมุมฉากแสดงว่าแบบจำลองที่ใช้ให้ประสิทธิภาพที่ดี



รูปที่ 2.35 การแสดงประสิทธิภาพของแบบจำลองด้วยกราฟ ROC

2.7 การเคลื่อนที่ของวัตถุในแนวเส้นตรง

การเคลื่อนที่ของวัตถุในแนวเส้นตรงมีสมการการเคลื่อนที่ของวัตถุอยู่หลายสมการ แต่ในหัวข้อนี้จะแสดงเพียงสมการที่นำมาใช้ในงานวิจัย โดยขั้นตอนการหาความเร็วและความเร่งในการเคลื่อนที่ของวัตถุสามารถคำนวณได้จากสมการที่ (2.61) และ (2.62)

$$s = vt \quad (2.61)$$

$$v = at \quad (2.62)$$

การหาระยะทางเมื่อทราบความเร็ว, ความเร่ง และเวลา ทำได้ตามสมการที่ (2.63)

$$s = ut + \frac{1}{2}at^2 \quad (2.63)$$

เมื่อ s คือ ระยะทาง มีหน่วยเป็น เมตร

t คือ เวลา มีหน่วยเป็น วินาที

v คือ ความเร็ว มีหน่วยเป็น เมตรต่อวินาที

a คือ ความเร่ง มีหน่วยเป็น เมตรต่อวินาที²

สามารถความรู้พื้นฐานเรื่องการเคลื่อนที่แนวตรงนี้ไปประยุกต์ใช้ในการประมาณค่าการเคลื่อนที่ของวัตถุในเฟรมถัดไป

2.8 งานวิจัยที่เกี่ยวข้องด้านการติดตามและภาษามือ

A. Shamaie [15] ได้เสนอขั้นตอนวิธีที่ใช้ในการตรวจหาและติดตามมือ สามารถนำมาประยุกต์ใช้ในงานต่าง ๆ เช่น พิมพ์คีย์บอร์ด มือกลอง แนะนำนักบินสำหรับนำเครื่องลงจอด แสดงขนาดของสิ่งของ เป็นต้น เสนอแก้ปัญหากรณีที่มีการเคลื่อนที่มาซ้อนทับกัน โดยใช้แบบจำลองพลวัต (Dynamic model) บนพื้นฐานของวงจรรองแบบกาลแมน ในเชิงพื้นที่ (Spatial) จะทำการหาตำแหน่งของมือทั้งสอง ส่วนในเชิงเวลา (Temporal) จะนำตำแหน่งที่ได้มาค้นหาตำแหน่งในเฟรมถัดไป โดยอาศัยการคำนวณจากความเร็วและความเร่งของแต่ละมือ การแก้ปัญหากรณีที่เกิดการบังกัน (Occlude) ของมือ ทำโดยเก็บข้อมูลล่าสุดของมือในเฟรมที่ก่อนจะเกิดการบังกัน

แบบจำลองพลวัตที่ใช้ติดตามมือแสดงตามสมการที่ (2.64) และ (2.65)

$$x_{j,k+1}^i = x_{j,k}^i + hx_{j,k}^{*i} + \frac{1}{2}h^2x_{j,k}^{**i} \quad (2.64)$$

$$x_{j,k+1}^{*i} = x_{j,k}^{*i} + hx_{j,k}^{**i} \quad (2.65)$$

x คือ ตำแหน่งของวัตถุ ณ เฟรมใด ๆ

x^* คือ อนุพันธ์อันดับหนึ่งของ x (ความเร็ว)

x^{**} คือ อนุพันธ์อันดับสองของ x (ความเร่ง)

k คือ ดรรชนีของเวลา (Index time)

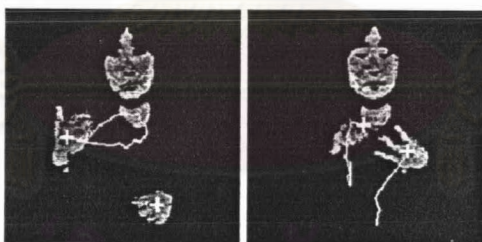
h คือ เวลาในการสุ่ม (Sampling time)

J. Yang [7] ได้เสนอวิธีการติดตามใบหน้าในเวลาจริง ซึ่งประกอบไปด้วย การสร้างแบบจำลองทางสถิติของข้อมูลส่วนสี่ผิวนในปริภูมิสี่ย่อย rg ส่วนการติดตามใบหน้าจะใช้การทำนายและการประมาณการเคลื่อนที่ ด้วยวิธีแบบกาลแมน (Kalman) เพื่อทราบถึงตำแหน่งของใบหน้าในทุก ๆ เฟรม

T. Starner [5] เสนอการนำแบบจำลองมาร์คอฟแฝง (Hidden Markov Model) มาช่วยในการแก้ปัญหาเรื่องการแปลความหมายภาษามืออเมริกัน ซึ่งถือเป็นงานวิจัยชิ้นแรก ๆ เกี่ยวกับภาษามือที่มีผู้ให้ความสนใจมาก แต่ในเบื้องต้นระบบที่ทำขึ้นยังมีฐานข้อมูลอยู่เป็นส่วนน้อย

J.S. Kim [11] เสนอระบบซึ่งใช้ในการแปลความหมายภาษามือเกาหลี โดยอุปกรณ์ที่ใช้ในการตรวจหาตำแหน่งของมือและนิ้ว คือ ถุงมือไซเบอร์ เชื่อมต่อกับคอมพิวเตอร์ส่วนบุคคล วิธีการที่นำมาใช้แก้ปัญหาเพื่อจำแนกความหมายของแต่ละคำ คือ โครงข่ายประสาทเทียม (Neural network) จากนั้นความหมายที่ได้จะถูกแสดงผลยังหน้าจอคอมพิวเตอร์เป็นลักษณะข้อความ ข้อเด่น คือ ตำแหน่งของมือที่ได้มีความแม่นยำมาก ส่วนข้อด้อย คือ ถุงมือไซเบอร์มีราคาค่อนข้างสูง และ ขั้นตอนวิธีนี้ไม่ได้นำข้อมูลบริเวณใบหน้าซึ่งถือเป็นข้อมูลสำคัญอีกอันหนึ่งมาร่วมพิจารณาด้วย

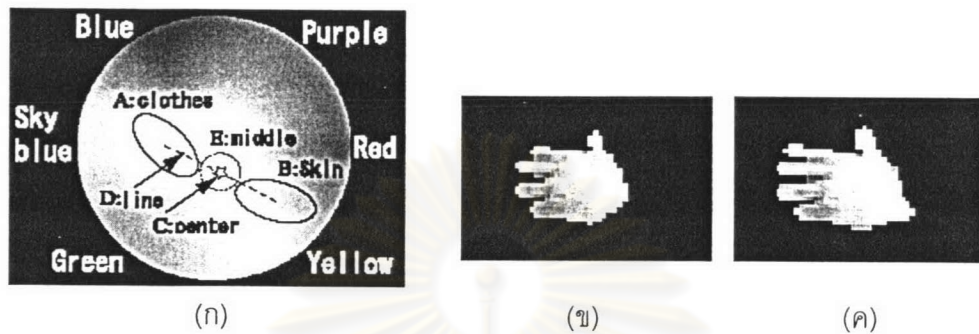
K. Imagawa [2] เสนอระบบติดตามมือของผู้แปลภาษามือตามเวลาจริง โดยการทำงานของระบบจะใช้การแยกสีผิวในการหาบริเวณใบหน้าและมือก่อน จากนั้นทำการประมวลผลหาตำแหน่งจุดศูนย์กลางบริเวณมือและติดตาม โดยใช้ตัวกรองแบบคาลแมน (Kalman Filter) ข้อเด่นของงานวิจัยนี้คือ การประมวลผลมีความซับซ้อนค่อนข้างน้อย สามารถงานตามเวลาจริงได้ เพราะใช้การติดตามแบบคาลแมนถือเป็นพื้นฐานที่สุด และ เก็บข้อมูลเพียงแนววิถีการเคลื่อนที่ของมือเท่านั้น สำหรับข้อด้อย คือ ไม่ได้คำนึงถึงใบหน้าของผู้แสดงภาษามือ ซึ่งถือเป็นข้อมูลส่วนที่มีความจำเป็นเช่นกัน



รูปที่ 2.36 ผลการติดตามมือและแนววิถี (Trajectory) ที่นำเสนอโดย K. Imagawa [2]

N. Tanibata [3] เสนอวิธีแยกคุณลักษณะเด่นของมือรวมถึงการจำแนกความหมายภาษามือจากวิดีโอคนภาษามือ ซึ่งแสดงด้วยผู้แปลภาษามือญี่ปุ่น ในเฟรมแรก ระบบจะทำการแยกสีผิวในปริภูมิสีย่อย HS และทำการหาตำแหน่งของใบหน้า มือ และข้อศอก ซึ่งบริเวณข้อศอกนี้ หาโดยอ้างอิงจากตำแหน่งของข้อมือ สำหรับเฟรมถัด ๆ ไปจะทำการติดตามใบหน้าและมือด้วยวิธีตรวจหาสีผิว และการติดตามข้อศอกโดยแผ่นแบบจับคู่ (Matching Template) ของรูปร่างข้อศอก เมื่อมีปัญหาเรื่องการซ้อนกันระหว่างบลิบทำโดยใช้ลายผิวจับคู่ (Matching Texture) ของใบหน้าและมือในเฟรมก่อนหน้า

โดยอาศัยคุณลักษณะเด่นของมือ มาช่วยในการตัดสินใจด้วย เช่น ทิศทางการเคลื่อนที่ จำนวนของนิ้วมือ ฯลฯ

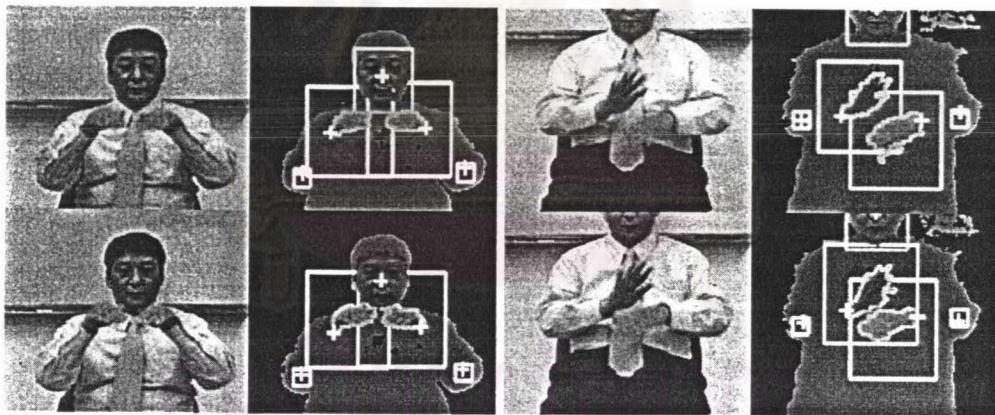


รูปที่ 2.37 แบบจำลองและผลการแยกสีผิวของ N. Tanibata [3]

(ก) บริเวณที่ถูกกำหนดเป็นสีผิวบนปริภูมิย่อย HS

(ข) ผลลัพธ์เมื่อเลือกเฉพาะขอบเขต B

(ค) ผลลัพธ์เมื่อเลือกขอบเขต B และ C

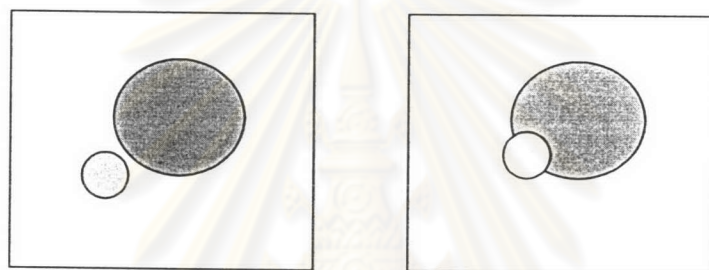


รูปที่ 2.38 ผลการติดตามใบหน้าและมือที่เสนอโดย N. Tanibata [3]

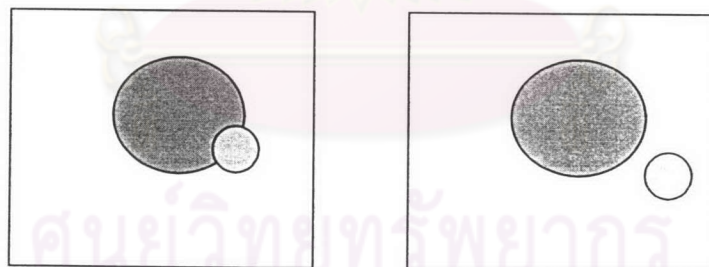
D.M. Saxe [19] นำเสนอการเข้ารหัสบริเวณที่สนใจเพื่อใช้สำหรับการสื่อสารภาษามือ ซึ่งเน้นเข้ารหัสเฉพาะบล็อกบริเวณใบหน้าและมือ งานวิจัยนี้แม้จะไม่เกี่ยวข้องกับการติดตามใบหน้าและมือ แต่ก็แสดงให้เห็นว่าการแปลความหมายภาษามือได้สมบูรณ์นั้น ต้องอาศัยข้อมูลทั้งในส่วน of ใบหน้าและมือ

2.9 แนวทางแก้ปัญหาคำถามการผสานและแยกกันของวัตถุ

ในหัวข้อนี้จะกล่าวถึงการพิจารณาสถานะที่เกิดขึ้นกรณีที่มีการผสาน (Merging) และแยก (Splitting) กันระหว่าง 2 วัตถุขึ้นไป กรณีการผสานกันดังแสดงในรูปที่ 2.39 กล่าวคือ ถ้าวัตถุที่ชนเฟรมปัจจุบันมีการค้นพบวัตถุ 2 วัตถุ แต่ในเฟรมถัดไป พบกลุ่มของวัตถุเพียงกลุ่มเดียวแสดงว่ามีการผสานกันระหว่าง 2 วัตถุนั้น (อินเตอร์เซกชันไม่เท่ากับเซตว่าง) กรณีแยกกันดังแสดงในรูปที่ 2.40 กล่าวคือ วัตถุที่ชนเฟรมปัจจุบันมีการค้นพบวัตถุเพียงอันเดียว แต่ในเฟรมถัดไปมี พบกลุ่มวัตถุ 2 กลุ่มแสดงว่ามีการแยกกันของวัตถุ

เฟรมที่ t เฟรมที่ $t+1$

รูปที่ 2.39 การผสานกันระหว่าง 2 วัตถุ

เฟรมที่ t เฟรมที่ $t+1$

รูปที่ 2.40 การแยกกันระหว่าง 2 วัตถุ