

การแยกส่วนภาพวัตถุดีไอทีเลือกบริเวณที่สนใจได้บนพื้นฐานของสีประกอบสีและเทคนิคเชิงสัญญาณ



นายธีรยุทธ สว่างศรี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

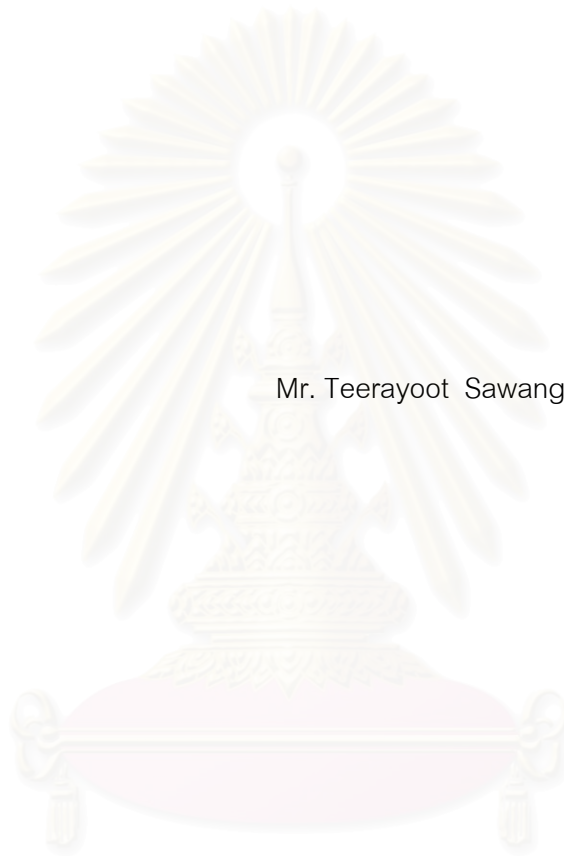
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-17-5609-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SELECTIVE REGION OF INTEREST VIDEO OBJECTS SEGMENTATION BASED ON COLOR
COMPONENT AND MORPHOLOGICAL TECHNIQUE



Mr. Teerayoot Sawangsri

สถาบันวิทยบริการ

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

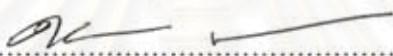
Chulalongkorn University

Academic Year 2005


ISBN 974-17-5609-7


หัวข้อวิทยานิพนธ์ การแยกส่วนภาพวัตถุวิดีโอที่เลือกบริเวณที่สนใจได้บนพื้นฐาน
องค์ประกอบสีและเทคนิคเชิงเส้นฐาน
โดย นายธีรยุทธ สว่างศรี
สาขาวิชา วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต


..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ติเรก สว่างศรี)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(ศาสตราจารย์ ดร.ประสิทธิ์ ประพัฒน์มงคล)


..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)

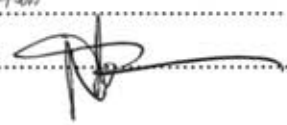

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุภาวดี อร่ามวิทย์)

สถาบันวิศวกรรมบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ธีรยุทธ สว่างศรี : การแยกส่วนภาพวัตถุวิดีโอที่เลือกบริเวณที่สนใจได้บนพื้นฐานองค์ประกอบสี และเทคนิคเชิงสีฐาน. (SELECTIVE REGION OF INTEREST VIDEO OBJECTS SEGMENTATION BASED ON COLOR COMPONENT AND MORPHOLOGICAL TECHNIQUE) อ. ที่ปรึกษา : รศ. ดร. สมชาย จิตะพันธ์กุล, 113 หน้า. ISBN 974-17-5609-7.

วิทยานิพนธ์นี้เสนอกรรมวิธีการแยกส่วนภาพบริเวณที่สนใจที่สามารถเลือกบริเวณที่สนใจได้เพื่อนำไปประยุกต์ใช้งานในด้านต่างๆ ที่ต้องการการแยกส่วนภาพเป็นบริเวณหรือวัตถุ โดยกำหนดว่ามีข้อมูลที่ใช้ในการเลือกบริเวณที่สนใจสำหรับการจัดสรรบิตอยู่แล้ว โดยการแยกส่วนบริเวณที่สนใจบริเวณใบหน้าประกอบไปด้วยการใช้องค์ประกอบสีและกรรมวิธีเชิงสีฐาน ส่วนการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุวิดีโอในลำดับภาพอยู่บนพื้นฐานของการรวมกันของคุณลักษณะสำคัญหลายประการของสัญญาณวิดีโอ โดยประกอบไปด้วยข้อมูลเชิงพื้นที่ ข้อมูลเชิงเวลา และข้อมูลส่วนบริเวณที่สนใจที่เป็นบริเวณใบหน้าของวัตถุ การแยกส่วนเชิงพื้นที่ประกอบด้วยขั้นตอน การตรวจรู้ขอบเขตของบริเวณ และการกำหนดบริเวณให้แก่อุณหภูมิโดยกรรมวิธีสันปันน้ำ และข้อมูลส่วนบริเวณใบหน้าของวัตถุซึ่งนำไปใช้ในการกำหนดขอบเขตบริเวณให้กับวัตถุ เพื่อแก้ไขข้อจำกัดของการแยกส่วนภาพเชิงพื้นที่เวลาที่ไม่สามารถแยกส่วนวัตถุได้อย่างสมบูรณ์ และข้อจำกัดที่เมื่อวัตถุไม่มีการเคลื่อนที่ หรือมีการเคลื่อนที่น้อยมากจนไม่สามารถตรวจรู้การเปลี่ยนแปลงได้อย่างมีประสิทธิภาพ โดยการใช้ข้อมูลบริเวณใบหน้าทำให้สามารถระบุขอบเขตของวัตถุได้แม้ว่าจะไม่มีการเคลื่อนที่ของวัตถุก็ตาม ซึ่งข้อมูลนี้ได้มาจากขั้นตอนการแยกส่วนบริเวณที่สนใจที่เป็น บริเวณใบหน้า ทำยสุดนำผลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อทำให้สามารถแยกส่วนภาพบริเวณที่สนใจที่เป็นวัตถุได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่อนิสิต.....*ส.ศ.น.ร. 15/12/68*
สาขาวิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2548.....

4470704121 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: SKIN COLOR MODELING / FACE SEGMENTATION / VIDEO OBJECT SEGMENTATION

TEERAYOOT SAWANGSRI : SELECTIVE REGION OF INTEREST VIDEO OBJECTS SEGMENTATION BASED ON COLOR COMPONENT AND MORPHOLOGICAL TECHNIQUE. THESIS ADVISOR : ASSOC.PROF.SOMCHAI JITAPUNKUL, Dr.Ing., 113 pp. ISBN 974-17-5609-7.

This thesis proposed an algorithm to dynamically define the region of interest videophone application. By assuming that there is information to select which region of interest are. In face segmentation stage, The algorithm uses the color information Hue and Cr to find the skin-color pixels and also use range of threshold obtained from red and blue components in normalized RGB color space to remove nonskin-color pixels because the human skin tends to have a predominance of red and nonpredominance of blue. Post-processing is used to remove such noises by a morphological operator. Moreover, the algorithm performs temporal filtering to remove skin-color pixels that immediately appear from frame to frame by using object tracking process to perform as memory for collecting skin-color objects obtained from previous frame to guide the next frame. The experimental results confirm the effectiveness of the proposed algorithm. In object segmentation, this thesis used an algorithm consisted of five processes: spatial segmentation process, temporal segmentation process, decision process, postprocessing and object tracking process. Spatial segmentation process segment image into many region by morphological operators. In temporal segmentation process, frame difference of intensity is used to segment roughly object. Postprocessing is the final process to segment and obtain fine object. Object tracking guides the next frame to locate the object.

Department.....Electrical Engineering.....Student's signature.....*Teerayot S*.....
 Field of study.....Electrical Engineering.....Advisor's signature.....*Somchai*.....
 Academic year.....2005.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี กระผมใคร่ขอกราบขอบพระคุณอย่างสูงสำหรับความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษา วิทยานิพนธ์ที่ได้ให้ความเมตตา ให้แนวทาง ข้อคิดเห็น และข้อเสนอแนะที่มีประโยชน์ในการทำ วิทยานิพนธ์และเป็นแบบอย่างที่ดี อีกทั้งให้โอกาสผู้เขียนได้ไปเสนอมผลงานวิชาการในต่างประเทศ ซึ่งนับว่าเป็นการสร้างประสบการณ์ที่ดีในชีวิตอีกครั้งหนึ่ง

ขอขอบคุณ อาจารย์ วรพจน์ พัฒนวิจิตร รุ่นพี่ปริญญาเอก ประจำกลุ่มวิจัยที่คอยกระตุ้น และผลักดันผู้เขียนจนมาถึงทุกวันนี้ได้ อีกทั้งคอยช่วยเหลือในทุกๆ ด้านด้วยความเมตตาและความปรารถนาดี ผู้เขียนไม่แน่ใจเหมือนกันว่าจะมาถึงจุดนี้ได้อย่างไรถ้าไม่มีรุ่นพี่ที่คอยเอาใจใส่ ท่านนี้ ผู้วิจัยขอระลึกไว้ในความทรงจำตลอดไป

ขอขอบคุณ นายสุนทร กนกวิลาวัฒน์ ที่ให้ความช่วยเหลือมาตลอดตั้งแต่ต้นจนจบ ทำให้ผู้เขียนได้ซาบซึ้งในน้ำใจของคำว่าเพื่อนและตระหนักรู้ว่าอย่างน้อยก็มีเพื่อนแท้คนหนึ่งในชีวิตที่คอยให้แนวคิดที่มีประโยชน์และว่ากล่าวอย่างตรงไปตรงมา อีกทั้งหนี้น้ำใจครั้งนี้ไม่รู้เมื่อใดจะชดใช้หมดสิ้นไป

ขอขอบคุณ นายศักดิ์ เหมชัย ที่ให้คำแนะนำและความช่วยเหลือตลอดมาในทุกเรื่องราว ผู้เขียนรู้สึกขอบคุณเป็นอย่างมากในทุกความช่วยเหลือและการเป็นตัวอย่างที่ดีในบางเรื่องราวของชีวิต

ขอขอบคุณ นายอภิชาติ บุญกิตติเจริญ ที่ให้ความช่วยเหลือตลอดมา ผู้เขียนรู้สึกขอบคุณเป็นอย่างมากในทุกความช่วยเหลือและจะระลึกถึงน้ำใจนี้ตลอดไป

ขอขอบคุณ นางสาวจินตนา เล้ารัตนารักษ์ ที่ให้ความช่วยเหลือเป็นอย่างมากจนกระทั่งผู้เขียนรู้สึกซาบซึ้งในน้ำใจในครั้งนี้เป็นอย่างมากและจะระลึกถึงน้ำใจนี้ตลอดไป

ขอขอบคุณโครงการวิจัยร่วมเสริมสร้างความเชื่อมโยงระหว่างภาควิชาชีพวิศวกรรมไฟฟ้าและภาคเอกชนทางด้านการวิจัยและพัฒนา ที่ให้การสนับสนุนเกี่ยวกับการทำงานวิจัยนี้เป็นอย่างสูง

ขอขอบคุณ รุ่นพี่ รุ่นน้อง เพื่อนๆ ผู้วิจัยทุกคน ไม่ว่าจะป็นภายในห้องปฏิบัติการวิจัยการประมวลผลสัญญาณดิจิทัลหรือที่ใดก็ตาม สำหรับความช่วยเหลือและกำลังใจในการทำวิจัยตลอดมาขอขอบคุณที่ทำให้วันและเวลาของผู้วิจัยผ่านไปอย่างมีความหมายแม้ว่าจะยากลำบากและต้องแลกมาด้วยหลายสิ่งหลายอย่างที่ต้องสูญเสียไปก็ตาม

ขอขอบคุณอาจารย์ทุกท่านที่ได้ให้ความรู้ในการศึกษา

สารบัญ

หน้า

บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฎ
ดัชนีคำศัพท์.....	ฒ
บทที่	
1. บทนำ.....	1
1.1 แนวเหตุผลและความเป็นมา.....	1
1.2 ภาพรวมของงานวิจัยที่เกี่ยวข้อง	2
1.2.1 งานวิจัยด้านการใช้ข้อมูลการแยกส่วนสัญญาณวิดีโอร่วมกับการ ควบคุมอัตราบิต	2
1.2.2 งานวิจัยด้านการแยกส่วนสัญญาณวิดีโอเฉพาะบริเวณใบหน้า.....	6
1.2.3 งานวิจัยด้านการแยกส่วนสัญญาณวิดีโอสำหรับวัตถุที่มีความหมาย	10
1.3 แนวทางของวิทยานิพนธ์	15
1.4 เป้าหมายและขอบเขตของวิทยานิพนธ์.....	16
1.4 ขั้นตอนและวิธีการดำเนินงาน	16
1.5 ประโยชน์ที่คาดว่าจะได้รับ	16
1.6เค้าโครงวิทยานิพนธ์.....	17
2. ทฤษฎีพื้นฐาน	18
2.1 แบบจำลองรูปร่างที่เคลื่อนที่ได้ (Active Shape Models)	18
2.1.1 สเนก (Snake).....	18
2.1.1.1 พลังงานของภาพ	21
2.1.1.1.1 สนามแรงต้นแบบ.....	22
2.1.1.1.2 สนามแรงคีย์ดาระยะทาง.....	25
2.1.1.1.3 สนามแรงแบบการไหลของเวกเตอร์เกรเดียน	26
2.1.2 เทมเพลตที่เปลี่ยนแปลงรูปร่างได้	28
2.2 พื้นฐานของสี.....	29

บทที่	หน้า
2.2.1 แบบจำลองสี	30
2.2.1.1 แบบจำลองสี RGB	31
2.2.1.2 แบบจำลองสี YIQ	33
2.2.1.3 แบบจำลองสี YCbCr	34
2.2.1.4 แบบจำลองสี HSI	35
2.2.1.5 แบบจำลองสีเนอร์แมลไลซ์ RGB	36
2.3 การประมวลผลภาพเชิงสัณฐาน (Morphological Image Processing).....	36
2.3.1 การทำไดเลชัน (Dilation) และ การทำอีโรชัน (Erosion)	37
2.3.2 ตัวดำเนินการแบบเปิดและปิด.....	38
2.3.3 การเติมเต็มบริเวณ (Region Filling)	42
2.4 การตรวจหาขอบ (Edge Detection)	42
2.4.1 เกรเดียนต์เชิงสัณฐาน (Morphological Gradient)	43
2.4.2 กรรมวิธีสันปันน้ำ.....	44
2.4.2.1 ขั้นตอนการสร้างเขื่อน	44
3. กรรมวิธีที่นำเสนอ.....	47
3.1 การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า.....	49
3.1.1 การแยกส่วนด้วยสี.....	50
3.1.2 การกำจัดพื้นหลัง.....	51
3.1.3 กระบวนการข้างหลังและการติดตามวัตถุ.....	51
3.2 การแยกส่วนบริเวณที่สนใจที่เป็นวัตถุที่มีความหมาย	51
3.2.3 รูปร่างก่อนหน้าของบริเวณที่สนใจ (Shape Prior of Region of Interest).....	52
3.2.2 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)	52
3.2.3 การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation)	53
3.2.4 กฎการตัดสินใจ (Decision Rule).....	55
3.2.5 กระบวนการข้างหลัง (Postprocessing)	56
3.2.6 การติดตามวัตถุ (Object Tracking)	57
3.2.7 การแยกส่วนภาพเชิงพื้นที่และเวลา	57
3.2.8 การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐาน (Hypothesis Testing).....	58
4. ผลการทดลอง	61

บทที่	หน้า
4.1 การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า.....	62
4.1.1 การแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าที่ได้มาจากกระบวนการ แยกส่วนด้วยองค์ประกอบสี	62
4.1.2 การแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าที่ได้จากกระบวนการแยกส่วน จากกรรมวิธีต่าง ๆ	65
4.1.3 การวัดผลของการแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าจากกระบวนการ แยกส่วนจากกรรมวิธีที่นำเสนอและกรรมวิธีอ้างอิงต่าง ๆ.....	73
4.2 การแยกส่วนบริเวณที่เป็นวัตถุที่มีความหมาย	78
4.2.1 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Claire	78
4.2.2 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Akiyo	79
4.2.3 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Grandmother	81
4.2.4 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Miss America.....	84
4.3 การวัดสมรรถนะของกรรมวิธีการแยกส่วนวัตถุ	86
4.3.1 การวัดผลการแยกส่วนภาพเชิงวัตถุ	86
4.4 การวิเคราะห์ความซับซ้อนของการแยกส่วนภาพทั้งสามวิธี	89
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	93
5.1 สรุปผลการวิจัย	93
5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต	96
รายการอ้างอิง.....	97
ภาคผนวก.....	96
บทความที่ได้รับการเผยแพร่	97
ประวัติผู้เขียนวิทยานิพนธ์	113

สารบัญตาราง

ญ

หน้า

ตารางที่ 4.1 ค่าอัตราความถูกต้องในแต่ละปริภูมิสี่ของแบบจำลองชนิดไม่มีตัวแปรเทียบ กับการแยกส่วนสี่ผิวด้วยสายตา.....	62
ตารางที่ 4.2 ค่าอัตราความผิดพลาดในแต่ละปริภูมิสี่ของแบบจำลองชนิดไม่มีตัวแปรเทียบ กับการแยกส่วนสี่ผิวด้วยสายตา.....	62
ตารางที่ 4.3 การเปรียบเทียบกระบวนการแยกส่วนภาพของทั้ง 3 กรรรมวิธี	89
ตารางที่ 4.4 เปรียบเทียบจำนวนเฟรมที่ใช้ในการแยกส่วนได้เต็มตัว	89
ตารางที่ 4.5 แสดงเวลาที่ใช้ในการประมวลผลในแต่ละวิธีโดยเฉลี่ยจาก 100 ลำดับภาพ	90



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

๘

ภาพประกอบ	หน้า
รูปที่ 2.1 แสดงค่าเกรเดียนและแสดงการหัดตัวเข้าหาค่าเกรเดียนย.....	23
รูปที่ 2.2 แสดงพฤติกรรมของสเนกต้นแบบ	24
รูปที่ 2.3 แสดงพฤติกรรมของสเนกแบบแรงค้ำยกัระยะทาง	26
รูปที่ 2.4 แสดงพฤติกรรมของสเนกแบบการไหลของเวกเตอร์เกรเดียน	28
รูปที่ 2.5 พิกัดแบบจำลองสี RGB.....	32
รูปที่ 2.6 ลักษณะค่าแกมมาที่ใช้ในการแสดงผล	32
รูปที่ 2.7 ความแตกต่างของพิกัด UV และพิกัด IQ	33
รูปที่ 2.8 ตำแหน่งของจุดตัวอย่างหรือจุดภาพของ YCbCr.....	34
รูปที่ 2.9 การทำไดเลชัน	38
รูปที่ 2.10 การทำอีโรชัน	38
รูปที่ 2.11 แสดงผลจากการกระทำตัวดำเนินการแบบเปิดและแบบปิด.....	40
รูปที่ 2.12 การกระทำตัวดำเนินการแบบเปิดต่อภาพไบนารี	41
รูปที่ 2.13 การกระทำตัวดำเนินการแบบเปิดต่อภาพไบนารี	41
รูปที่ 2.14 แสดงการขจัดสิ่งรบกวนออกจากวัตถุและพื้นหลังโดยตัวดำเนินการเชิงสัดส่วน	41
รูปที่ 2.15 การเติมเต็มบริเวณ	42
รูปที่ 2.16 แบบจำลองของขอบ	43
รูปที่ 2.17 กรรรมวิธีสันปันน้ำ	44
รูปที่ 2.18 การสร้างเขื่อน.....	45
รูปที่ 3.1 โครงสร้างของแนวคิดวิทยานิพนธ์การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า.....	48
รูปที่ 3.2 โครงสร้างของแนวคิดวิทยานิพนธ์การแยกส่วนวัตถุที่มีความหมาย	48
รูปที่ 3.3 ฮิสโตแกรมของโดเมนสีน้ำเงินและสีแดง ในปริภูมิสีนอร์มอลไลซ์แดงเขียวน้ำเงิน (ก) ฮิสโตแกรมของโดเมนสีน้ำเงิน (ข) ฮิสโตแกรมของโดเมนสีแดง	50
รูปที่ 3.4 แสดงตัวอย่างผลลัพธ์จากขั้นตอนการกำหนดขอบเขตของวัตถุจากใบหน้า และการแยกส่วนด้วยกรรรมวิธีสเนก	52
รูปที่ 3.5 โครงสร้างการแยกส่วนภาพเชิงพื้นที่.....	54
รูปที่ 3.6 โครงสร้างการจัดวางตำแหน่งไปต์.....	54
รูปที่ 3.7 แสดงตัวอย่างผลลัพธ์จากขั้นตอนการแยกส่วนเชิงพื้นที่.....	55

ภาพประกอบ	หน้า
รูปที่ 3.8 กรรมวิธีที่นำเสนอโดย ดัชกรณั ตันเจริญ และคณะ.....	58
รูปที่ 3.9 กรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee, Myoung Ho Lee, Chieteuk และ Yo-Sung Ho.....	59
รูปที่ 4.1 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสี ของลำดับภาพ Grandmother	63
รูปที่ 4.2 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสี ของลำดับภาพ Akiyo.....	63
รูปที่ 4.3 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสี ของลำดับภาพ Claire	64
รูปที่ 4.4 ผลการแยกส่วนภาพที่เปรียบเทียบกรรมของแต่ละปริภูมิสี ของลำดับภาพ MissAM.....	64
รูปที่ 4.5 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่นำเสนอของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6	65
รูปที่ 4.6 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6	65
รูปที่ 4.7 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6	66
รูปที่ 4.8 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่ [10] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6	66
รูปที่ 4.9 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่เสนอของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6	67
รูปที่ 4.10 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6	67
รูปที่ 4.11 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6	68
รูปที่ 4.12 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่ [10] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6	68

ภาพประกอบ	หน้า
รูปที่ 4.13 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่นำเสนอของลำดับภาพ Claire เฟรมที่ 1 ถึง 6	69
รูปที่ 4.14 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Claire เฟรมที่ 1 ถึง 6	69
รูปที่ 4.15 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Claire เฟรมที่ 1 ถึง 6	70
รูปที่ 4.16 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Claire เฟรมที่ 1 ถึง 6	70
รูปที่ 4.17 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีนำเสนอของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6	71
รูปที่ 4.18 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6	71
รูปที่ 4.19 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6	72
รูปที่ 4.20 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6	72
รูปที่ 4.21 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีที่นำเสนอเทียบกับ การแยกส่วนด้วยสายตา	73
รูปที่ 4.22 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีที่นำเสนอเทียบกับ การแยกส่วนด้วยสายตา	73
รูปที่ 4.23 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีที่ใช้องค์ประกอบสี CbCr [8] เทียบกับการแยกส่วนด้วยสายตา	74
รูปที่ 4.24 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีที่ใช้องค์ประกอบสี CbCr [8] เทียบกับการแยกส่วนด้วยสายตา.....	75
รูปที่ 4.25 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธี [22] เทียบกับการ แยกส่วนสีผิวด้วยสายตา	75
รูปที่ 4.26 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธี [22] เทียบกับการ การแยกส่วนสีผิวด้วยสายตา.....	75

ภาพประกอบ	หน้า
รูปที่ 4.27 รูปภาพแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธี [10] เทียบกับการแยกส่วนสีผิวด้วยสายตา.....	76
รูปที่ 4.28 รูปภาพแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธี [10] เทียบกับการแยกส่วนสีผิวด้วยสายตา	76
รูปที่ 4.29 ผลการแยกส่วนภาพโดยกรรมวิธีที่นำเสนอ เฟรมที่ 1 ถึง 6	77
รูปที่ 4.30 ผลการแยกส่วนภาพโดยกรรมวิธีทดสอบสมมติฐาน เฟรมที่ 1 ถึง 6	78
รูปที่ 4.31 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 6	78
รูปที่ 4.32 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอ เฟรมที่ 1 ถึง 9	79
รูปที่ 4.33 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9	80
รูปที่ 4.34 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9	80
รูปที่ 4.35 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอ เฟรมที่ 1 ถึง 9	81
รูปที่ 4.36 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9	82
รูปที่ 4.37 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9	82
รูปที่ 4.38 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอ เฟรมที่ 1 ถึง 9	84
รูปที่ 4.39 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9	84
รูปที่ 4.40 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9	85
รูปที่ 4.41 รูปภาพแสดงค่าอัตราความผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Akiyo	86
รูปที่ 4.42 รูปภาพแสดงค่าอัตราความผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Claire	86
รูปที่ 4.43 รูปภาพแสดงค่าอัตราความผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Grandmother ...	87
รูปที่ 4.44 รูปภาพแสดงค่าอัตราความผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Miss Am	87

บัญชีคำศัพท์

Algorithm	กรรมวิธี
Analog	แอนะล็อก
Background Elimination	การกำจัดพื้นหลัง
Bit Allocation	การจัดสรรบิต
Block	บล็อก
Calculus	แคลคูลัส
Characteristics	คุณลักษณะ
Clustering	การจัดกลุ่ม
Component	องค์ประกอบ
Color Segmentation	การแยกส่วนด้วยองค์ประกอบสี
Convolution	คอนโวลูชัน
Cyclic	ไซคลิก
Decision Rule	กฎการตัดสินใจ
Digital Video	วิดีโอดิจิทัล
Detection	การตรวจหา
Detection Rate	อัตราการตรวจหา
Dilation	ไดเลชัน
Distortion	ความเพี้ยน
Distance	ระยะทาง
Dynamic	พลวัต
Eigenface	ไอเกนเฟส
Elastic	ขยับขยายตัวได้
Energy	พลังงาน
Equilibrium	สมดุล
Erosion	อีโรชัน
Euclidean	ยูคลิเดียน
False Alarm Rate	อัตราการเตือนความผิดพลาด

False Negative	ความผิดพลาดแบบลบ
False Positive	ความผิดพลาดแบบบวก
Feature	คุณลักษณะ
Foreground	พื้นหน้า
Force	แรง
Frame Difference Mask	마스크ความต่างเฟรม
Field	สนาม
Fast Marching	ฟาสมาซิง
Gray Level	ระดับสีเทา
Gaussian	เกาส์เซียน
Gradient	เกรเดียน
Geometric	ทางเรขาคณิต
HSI	ปริภูมิสี HSI
Hue	สี
Head-and-Shoulders	ศีรษะและลำตัว
Histogram	ฮิสโตแกรม
Input	อินพุต
Internal	ภายใน
Information Technology (IT)	เทคโนโลยีสารสนเทศ
Intensity	ความเข้มแสง
Knowledge	ความรู้
Level Set	เลเวลเซต
Logic	ลอจิก
Low-Level	ระดับล่าง
Logarithm	ลอการิทึม
Macroblock	มาโครบล็อก
Map	แมป
Mask	마스크
Motion Vector	เวกเตอร์การเคลื่อนที่

Minimize	น้อยที่สุด
Mode	วิธี
Model	แบบจำลอง
Morphological	เชิงสัณฐาน
Multimedia	มัลติมีเดีย
Negative	ทางลบ
Non-face Region	บริเวณที่ไม่ใช่บริเวณใบหน้า
Normalized	นอร์มอลไลซ์
Operator	ตัวดำเนินการ
Object – based Coding	การเข้ารหัสเชิงวัตถุ
Parameter	พารามิเตอร์
Physical	เชิงฟิสิกส์
Pixel	จุดภาพ
Postprocessing	กระบวนการข้างหลัง
Projection	โปรเจคชัน
Prototype	ต้นแบบ
Potential	ศักยภาพ
Prior	ก่อนหน้า
Projection	โปรเจคชัน
Quantization Level	ระดับควอนไทซ์
Rate Control	การควบคุมอัตราบิต
Rate-and-Distortion	อัตราและความเพี้ยน
Real - time Multimedia	มัลติมีเดียแบบเวลาจริง
Region of Interest (ROI)	บริเวณที่สนใจ
Region Growing	การเติบโตของบริเวณ
Remaining Areas	พื้นที่ที่เหลืออยู่
RGB	แดง เขียว น้ำเงิน
rgb	แดง เขียว น้ำเงิน นอร์มอลไลซ์
Scalar	สเกลาร์

Shape	รูปร่าง
Skin Color Model	แบบจำลองสีผิว
Smooth	ความราบเรียบ
Spatial Domain	โดเมนปริภูมิเชิงพื้นที่
Spatial Segmentation	การแยกส่วนเชิงพื้นที่
Spatial Redundancy	ส่วนเกินในโดเมนปริภูมิ
Spatial- temporal Correlation	ความสัมพันธ์ในเชิงพื้นที่และเวลา
Standard Deviation	ค่าเบี่ยงเบนมาตรฐาน
Stereo	สเตอริโอ
System	ระบบ
Structure	โครงสร้าง
Template	เทมเพลต
Temporal Segmentation	การแยกส่วนเชิงเวลา
Threshold	ค่าขีดเริ่มเปลี่ยน
Theorem	ทฤษฎีบท
Test Sequence	ลำดับภาพมาตรฐาน
True Negative	ความถูกต้องแบบลบ
True Positive	ความถูกต้องแบบบวก
YCbCr	ปริภูมิสี YCbCr
Variation	การผันแปร
Vector Quantization	การควอนไทซ์แบบเวกเตอร์
Very Low Bit-rate	อัตราบิตต่ำมาก
Visual Quality	คุณภาพทางการมองเห็น
Video Coding	การเข้ารหัสสัญญาณวิดีโอ
Videoconference	การประชุมสัมมนาทางโทรทัศน์
Videophone	โทรศัพท์ภาพ
Video Segmentation	การแยกส่วนสัญญาณวิดีโอ

บทที่ 1

บทนำ

1.1 แนวเหตุผลและความเป็นมา

การสื่อสารโทรคมนาคมในปัจจุบันได้เข้ามามีบทบาทในสังคมมากขึ้นและมีแนวโน้มที่จะทวีความสำคัญมากขึ้นเรื่อยๆ ดังจะเห็นได้จากการขยายตัวของระบบการสื่อสารในประเทศ และความพยายามในการสร้างระบบสื่อสารที่มีประสิทธิภาพมากขึ้น นอกจากนี้ระบบที่พัฒนาขึ้นภายหลังจะมีประสิทธิภาพในการส่งผ่านข้อมูลได้สูงกว่าระบบที่พัฒนาขึ้นมาก่อน ระบบสื่อสารจะพยายามรองรับการส่งผ่านสัญญาณภาพและสัญญาณวิดีโอเพิ่มเติมดังนั้นจึงส่งผลให้ข่าวสารในการส่งผ่านมีปริมาณสูงขึ้นและจำเป็นต้องใช้ช่องสัญญาณที่มีอัตราเร็วสูงแต่เนื่องจากการส่งผ่านช่องสัญญาณที่มีอัตราเร็วสูงจะเป็นการเพิ่มต้นทุนของการส่งผ่านข้อมูลและทำให้อัตราของค่าบริการเพิ่มสูงขึ้นดังนั้นจึงมีความพยายามในการพัฒนาเทคโนโลยีของการเข้ารหัสข่าวสารแบบต่าง ๆ เพื่อลดอัตราบิดในการส่งสัญญาณวิดีโอให้ต่ำลงแต่มีคุณภาพของสัญญาณที่ดีพอซึ่งจะช่วยทำให้ลดการใช้แถบความถี่ในช่องสัญญาณให้น้อยลง ทำให้สามารถใช้ช่องสัญญาณส่งข่าวสารต่าง ๆ ได้หลายชนิดพร้อมกัน และทำให้ใช้ช่องสัญญาณได้อย่างมีประสิทธิภาพที่สุด มาตรฐานการเข้ารหัสสัญญาณวิดีโอทั่วไปมีอยู่ 2 มาตรฐานคือมาตรฐาน ISO/ IEC MPEG-4 [1] และ CCITT H.261/ ITU H.263 [2]

จำนวนบิตที่ได้จากการเข้ารหัสในกระบวนการส่งสัญญาณวิดีโอต้องสอดคล้องกับอัตราช่องสัญญาณดังนั้นการควบคุมอัตราบิต (rate control) จะกำหนดจำนวนบิตที่จะส่งผ่านในแต่ละเฟรมโดยจะนำไปใช้ในการกำหนดความละเอียดของระดับควอนไทซ์ (quantization level) เพื่อทำการควอนไทซ์สัญญาณวิดีโอ การเลือกระดับควอนไทซ์ของการควบคุมอัตราบิตจะต้องพิจารณาความเต็มของบัฟเฟอร์ (buffer fullness) และแบนด์วิดท์ (bandwidth) ของช่องสัญญาณ การควบคุมอัตราที่สามารถหาระดับควอนไทซ์ที่เหมาะสม แบ่งได้เป็น 2 วิธี คือ วิธีการโปรแกรมพลวัต (dynamic programming) และวิธีการลากรองจ์ (lagrange) แต่เนื่องจากวิธีทั้งสองไม่ได้พิจารณาถึงระดับของบริเวณที่สนใจ (region of interest) ดังนั้นบริเวณที่สนใจอาจมีความละเอียดน้อยเนื่องจากการได้รับการจัดสรรบิตที่เท่ากันกับบริเวณที่ไม่ได้สนใจและเป็นผลทำให้คุณภาพของภาพมีค่าที่ลดลง

มาตรฐานการเข้ารหัสสัญญาณวิดีโอแบบ MPEG-4 จะประกอบด้วยส่วนของการเข้ารหัสแบบวัตถุโดยการเข้ารหัสแบบนี้จะเข้ารหัสภาพหรือเสียงโดยตัวของแต่ละวัตถุเองดังนั้น นักวิจัยจำนวนมากนำเสนอแนวคิดที่ใช้การนำข้อมูลการแยกส่วนสัญญาณวิดีโอ (video segmentation) ร่วมกับการควบคุมอัตราบิตเพื่อพัฒนาสัญญาณวิดีโอในบริเวณที่สนใจ [6-14] ผู้ใช้จะสามารถกำหนดบริเวณที่สนใจอย่างเช่นบริเวณใบหน้า [6-13] หรือที่เป็นบริเวณวัตถุที่มีความหมาย (content-based) [14] หรือพิจารณาอัตราส่วนบริเวณใบหน้าและบริเวณพื้นหลังร่วมกับเวกเตอร์การเคลื่อนที่ของบริเวณดังกล่าวมาช่วยในการจัดสรรบิต [11] เป็นต้น

วิทยานิพนธ์ฉบับนี้นำเสนอกรรมวิธีการแยกส่วนภาพบริเวณที่สนใจที่เป็นบริเวณใบหน้าและบริเวณวัตถุที่มีความหมาย โดยกำหนดว่ามีข้อมูลที่ใช้ในการเลือกบริเวณที่สนใจสำหรับการจัดสรรบิตอยู่แล้ว เนื่องจากว่าการพัฒนาประสิทธิภาพการแยกส่วนบริเวณที่สนใจให้สูงขึ้นจะส่งผลให้ได้สัญญาณวิดีโอที่ผ่านการจัดสรรบิตในบริเวณที่สนใจมีคุณภาพที่สูงขึ้น

1.2 ภาพรวมของงานวิจัยที่เกี่ยวข้อง

หัวข้อนี้จะกล่าวถึงภาพรวมของงานวิจัยที่เกี่ยวข้อง ซึ่งแบ่งออกเป็น 3 ส่วนด้วยกันคือ งานวิจัยด้านการนำข้อมูลการแยกส่วนสัญญาณวิดีโอมาใช้ร่วมกับการควบคุมอัตราบิต งานวิจัยด้านการแยกส่วนสัญญาณวิดีโอเฉพาะบริเวณใบหน้า (face segmentation) และงานวิจัยด้านการแยกส่วนสัญญาณวิดีโอสำหรับวัตถุที่มีความหมาย (content-based segmentation)

1.2.1 งานวิจัยด้านการนำข้อมูลการแยกส่วนสัญญาณวิดีโอร่วมกับการควบคุมอัตราบิต

John Hartung, Arnaud Jacquin, James Pawlyk, Jonathan Rosenberg, Hiroyuki Okada และ Paul E. Crouch ได้นำเสนอพื้นฐานการเข้ารหัสสัญญาณวิดีโอที่สอดคล้องกับ H.263 เิงวัตถุสำหรับการประยุกต์ใช้งานทางด้านการประชุมสัมมนาในปี ค.ศ. 1998 [6] ซึ่งแนวคิดคือ การอธิบายพื้นฐานของการเข้ารหัสสัญญาณวิดีโอต้นแบบสำหรับการสร้างสรรค์และทดสอบผลิตภัณฑ์ทางมัลติมีเดีย (multimedia) ตามมาตรฐาน H.263 ด้วยการเพิ่มคุณลักษณะในการปรับปรุงคุณภาพทางการมองเห็นของวัตถุเพื่อความเหมาะสมมากยิ่งขึ้น สำหรับการเข้ารหัสที่อัตราบิตต่ำมาก (very low bit-rate) ของวิดีโอที่มีลักษณะศีรษะและลำตัว (head-and-shoulders) เช่นการประยุกต์ใช้งานทางมัลติมีเดียแบบเวลาจริง (real-time multimedia), โทรศัพท์ภาพ (videophone), การประชุมสัมมนาทางโทรทัศน (videoconference)

คุณลักษณะในการปรับปรุงคุณภาพที่เพิ่มเข้าไปประกอบด้วย 5 ขั้นตอนคือการแยกส่วน ออกเป็นวัตถุที่สนใจ, กระบวนการกรองก่อนหน้าการแยกส่วน, การควบคุมอัตราที่มีส่วนช่วยเหลือ แบบจำลองที่สนใจ (model-assisted rate control), การควอนไทซ์แบบเวกเตอร์ (vector quantization) ที่ปรับตัวได้ และท้ายสุดกระบวนการภายหลังการแยกส่วน

การแยกส่วนอยู่บนพื้นฐานของสีผิวและรูปร่างของศีรษะ โดยกรรมวิธีการแยกส่วนรูปร่าง ของศีรษะกระทำโดยการตรวจจับวัตถุพื้นหน้า (foreground) ด้วยเส้นโครงร่าง (outline) รูปร่าง ในภาพอินพุต (input) โดยขั้นตอนของกรรมวิธีมีการกระทำเป็นลำดับขั้น โดยขั้นตอนแรกทำ การหาเส้นโครงร่างโดยการค้นหาอัตราส่วนที่เหมาะสมที่สุด (fitness-ratio) ของเทมเพลต (template) บริเวณที่น่าจะเป็น (candidate) ไบหน้าตามแนวตั้ง (vertical) และแนวเฉียงเพื่อ ประมาณความสมมาตรของ ไบหน้า

มีการใช้การประมาณการเคลื่อนที่โดยรวม (global motion estimation) ในกรณีที่พื้น หลังมีความซับซ้อน ส่วนการแยกส่วนบนพื้นฐานของสีผิวนั้นจะใช้ภายหลังได้ผลลัพธ์จากการแยก ส่วนรูปร่างของศีรษะแล้ว

การควบคุมอัตราทำได้เพื่อรับรองว่าบริเวณไบหน้าจะได้รับค่าควอนไทซ์ที่ต่ำกว่าบริเวณอื่นๆ เช่น บริเวณพื้นหลัง มีการกระทำซ้ำเพื่อหาค่าควอนไทซ์ที่เหมาะสมของบริเวณไบหน้าและ บริเวณที่ไม่ใช่ไบหน้า โดยมีค่าต้นทุน (budget) จำนวนบิตสำหรับเฟรมที่ทำ ไว้เป็นตัวแปรที่ สำคัญในการเลือกค่าควอนไทซ์ของแต่ละบริเวณที่จะกำหนดจำนวนบิตให้ ทำให้มีความซับซ้อน ในการคำนวณค่อนข้างสูง

Douglas Chai และ King N. Ngan ได้นำเสนอการแยกส่วนไบหน้าโดยใช้สีผิวในการ ประยุกต์ใช้งานทางโทรศัพท์ภาพในปี ค.ศ. 1999 [8] ซึ่งแนวคิดคือการนำการแยกส่วนไบหน้ามา ใช้กับการเข้ารหัสสัญญาณวิดีโอในการประยุกต์ใช้งานทางด้านโทรศัพท์ภาพ โดยใช้ผลจากการ แยกส่วนบริเวณไบหน้ามาปรับปรุงคุณภาพทางการมองเห็นของลำดับภาพที่เข้ารหัสตาม มาตรฐาน H.261 โดยการเข้ารหัสบริเวณที่สนใจซึ่งก็คือบริเวณไบหน้าด้วยความละเอียดที่ มากกว่าบริเวณอื่นๆ อีกทั้งมีการใช้ควอนไทซ์เซอร์ 2 ตัวในการควบคุมอัตราบิตที่ใช้ แต่วิธีนี้มี ข้อกำหนดที่ว่าพื้นหลังต้องมีความสำคัญน้อยกว่าพื้นหน้าอย่างมากเพื่อใช้ระดับค่าควอนไทซ์ที่สูง กว่ามากๆ ไบบริเวณพื้นหลังเพื่อมีจำนวนบิตมากพอที่จะจัดสรรให้กับบริเวณที่สนใจได้อย่างเต็มที่ ทำให้มีข้อจำกัดถ้าพื้นหลังมีความสำคัญใกล้เคียงกันกับพื้นหน้าและ Douglas Chai และ Abdesselam Bouzerdoun ได้เสนอการเข้ารหัสลำดับภาพโทรศัพท์ภาพที่ให้คุณภาพการ มองเห็นที่ดีกว่าโดยการใช้การระบุตำแหน่งของไบหน้าและการกระจายจำนวนบิตใหม่ในปี ค.ศ. 2000 [25] โดยการใช้กรรมวิธีข้างต้น

Chia-Wen Lin, Yao-Jen Chang และ Yung-Chang Chen ได้เสนอการเข้ารหัสสัญญาณวิดีโอซึ่งมีส่วนช่วยเหลือบริเวณใบหน้าด้วยความซับซ้อนต่ำในปี ค.ศ. 2000 [9] ซึ่งมีแนวความคิดคือ การนำเสนอโครงสร้างการเข้ารหัสสัญญาณวิดีโอที่มีส่วนช่วยเหลือบริเวณใบหน้าเพื่อช่วยปรับปรุงคุณภาพทางการมองเห็น (visual quality) ของบริเวณใบหน้าในการประยุกต์ใช้งานทางด้านโทรศัพท์ภาพ บนพื้นฐานของสถิติ การตรวจจับและติดตามใบหน้าได้รับการใช้งานเพื่อระบุตำแหน่งของบริเวณใบหน้าโดยใช้แบบจำลองเกาส์เซียน (gaussian model) ในการแยกประเภทของจุดภาพ (pixel) ว่าเป็นส่วนของสถิติหรือไม่แล้วจึงใช้กรรมวิธีโปรเจกชัน (projection) แบบอินทิกรัล (integral) สองครั้งรูปแบบใหม่ในการแยกส่วนของบล็อก (block) ใบหน้าออกมา

นอกจากนี้ บนพื้นฐานของโครงสร้างการควบคุมอัตราบิต (rate control) ของ H.263 TMN8 บทความได้เสนอโครงสร้างการปรับค่าถ่วงน้ำหนักความเพี้ยนพลวัต (dynamic distortion weighting adjustment : DDWA) เพื่อปรับปรุงคุณภาพของบริเวณใบหน้าโดยการทิ้งมาโครบล็อก (macroblock) ที่ไม่ใช่บริเวณใบหน้า (non-face region) ที่คงที่ด้วยการใช้ค่าถ่วงน้ำหนักที่แตกต่างไปจากค่าถ่วงน้ำหนักที่เป็นบริเวณใบหน้าซึ่งค้ำค่าที่จะเสียลดคุณภาพของการมองเห็นของบริเวณที่มีความสำคัญน้อยและใช้บิตที่ได้มานั้นไปกับการปรับปรุงคุณภาพของบริเวณที่มีความสำคัญมากกว่า เช่น บริเวณใบหน้าในการประยุกต์ใช้งานทางด้านโทรศัพท์ภาพ แต่งานวิจัยยังมีข้อจำกัดในการใช้งานบางอย่าง คือ ผู้ใช้ต้องกำหนดพารามิเตอร์การถ่วงน้ำหนักเพื่อการจัดสรรบิตเอง แต่ทว่าสัญญาณวิดีโอที่มีความซับซ้อนต่างกันไม่สามารถกำหนดค่าพารามิเตอร์ดังกล่าวเป็นค่าใดค่าหนึ่งได้ซึ่งทำให้ขาดความยืดหยุ่นในการใช้งาน

ในปี ค.ศ. 2003 Mei-Juan Chen, Ming-Chieh Chi, Ching-Ting Hsu และ Jeng-Wei Chen ได้เสนอการเข้ารหัสวิดีโอบริเวณที่สนใจ (region of interest) บนพื้นฐาน H.263+ ด้วยเทคนิคการตรวจจับสถิติที่มั่นคง [10] ซึ่งมีแนวคิดคือ การพัฒนากรรมวิธีการตรวจจับสถิติที่มั่นคงเพื่อตรวจจับใบหน้าเพื่อที่นำไปใช้ร่วมกับการควบคุมอัตราในตัวเข้ารหัส (encoder) สัญญาณวิดีโอตามมาตรฐาน H.263+ ที่เพิ่มวิธี (mode) การปรับปรุงการควอนไทซ์ โดยการปรับพารามิเตอร์ค่าถ่วงน้ำหนักความเพี้ยนและพารามิเตอร์ค่าความแปรปรวน (variance) ที่ระดับมาโครบล็อกเพื่อควบคุมคุณภาพทางการมองเห็นที่บริเวณต่างๆ โดยการเข้ารหัสสัญญาณวิดีโอในบริเวณที่สนใจนั้นจะมีการจัดสรรบิตจำนวนมากให้กับบริเวณใบหน้าโดยการลดจำนวนบิตในบริเวณที่ไม่ใช่บริเวณใบหน้าให้น้อยลง คุณภาพทางการมองเห็นโดยภาพรวมที่อัตราบิตเรตต่ำมีการพัฒนาให้ดีขึ้นด้วยการเข้ารหัสบริเวณใบหน้าด้วยอัตราบิตที่

มากกว่าพื้นหลังเมื่อเปรียบเทียบกับกรรมวิธีที่เสนอใน [8] อย่างไรก็ตาม ค่าพารามิเตอร์ที่ปรับให้ผลในการจัดสรรบิตที่ไม่ต่างกันมากนักแม้ว่าค่าที่ปรับจะเปลี่ยนแปลงไปอย่างมากก็ตาม จากผลการทดลองพบว่าคุณภาพไม่สามารถปรับปรุงได้เรื่อยไป การเลือกค่าพารามิเตอร์ให้มีค่ามากกว่าค่าเทรชโฮลด์ (threshold) มีผลเพียงทำให้บริเวณที่สนใจมีคุณภาพที่ดีขึ้นกว่าบริเวณที่ไม่ได้สนใจเท่านั้น และพารามิเตอร์ให้ผลลัพธ์ที่ซ้ำซ้อนกันทำให้ต้องหาค่าที่เหมาะสมหลายค่าทั้งที่ไม่ได้มีผลต่อการปรับปรุงคุณภาพก็ตามและเมื่อลำดับภาพมีความซับซ้อนขึ้นอย่างทันทีทันใดการกำหนดค่าพารามิเตอร์ที่คงที่ทำให้ผลลัพธ์ที่ได้ไม่ได้ผลที่ดีเท่าที่ควร

ในปี ค.ศ. 2004 Ming-Chieh Chi, Mei-Juan Chen และ Ching-Ting Hsu ได้เสนอการเข้ารหัสวิดีโอบริเวณที่สนใจ (region of interest) โดยการควบคุมแบบฟัซซี (fuzzy) สำหรับมาตรฐาน H.263+ [11] ซึ่งมีแนวคิดคือ การใช้การควบคุมแบบฟัซซีลอจิก (logic) กับตัวควบคุมอัตราบิตมาตรฐาน (TMN8) เพื่อให้ได้พารามิเตอร์ค่าถ่วงน้ำหนักที่ปรับตัวได้สำหรับแต่ละมาโครบล็อกและจัดสรรจำนวนบิตที่แตกต่างกันในแต่ละมาโครบล็อกตามความซับซ้อนและการผันแปรของค่าความแปรปรวนในรูปแบบของอัตราและความเพี้ยน (rate-and-distortion) เนื่องด้วยกรรมวิธีที่เสนอมีการเพิ่มความซับซ้อนเข้าไปในกระบวนการควบคุมอัตราเพื่อให้การเลือกค่าพารามิเตอร์ปรับตัวตามความซับซ้อนของลำดับภาพได้ทำให้มีความซับซ้อนเพิ่มขึ้นและเมื่อเปรียบเทียบผลลัพธ์จากค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนของกรรมวิธีที่ใช้กับกรรมวิธี [10] ที่เป็นการกำหนดค่าพารามิเตอร์คงที่มีค่าแตกต่างกันไม่มากอย่างมีนัยสำคัญ

ในปี ค.ศ. 2004 Kiran Challapali, Tomas Brodsky, Yun-Ting Lin, Yong Yan และ Richard Yi Chen ได้เสนอการแยกส่วนวัตถุแบบเวลาจริงและการเข้ารหัสสำหรับการสื่อสารวิดีโอที่สามารถเลือกคุณภาพของสัญญาณได้ [41] โดยระบบประกอบด้วยสองส่วน คือ กรรมวิธีในการแยกส่วน และวิธีการควบคุมอัตราแบบหลายวัตถุ มี 2 วิธีการในการแยกวัตถุ คือ การแยกส่วนพื้นหน้าบนพื้นฐานของกล้องสองตัว และการแยกส่วนใบหน้าบนพื้นฐานของกล้องเพียงตัวเดียว โดยการแยกส่วนบริเวณใบหน้าประกอบไปด้วยการแปลงปริภูมิสี การลดขนาดของภาพ การแยกส่วนด้วยสีผิว การเข้าสู่คู่มือการใช้โอแกนเฟส (eigenface) การตัดสินใจร่วมกันระหว่างข้อมูลสีและข้อมูลโครงสร้าง โดยกรรมวิธีที่ใช้อยู่บนปริภูมิสี HSV มีการตัดสินใจทำให้มีความซับซ้อนเพิ่มขึ้นมาในกระบวนการ ภายใต้การจัดกล้องแบบสเตอริโอ (stereo) ได้มีการปรับปรุงกรรมวิธีการประมาณความแตกต่างเพื่อให้ได้ขอบของวัตถุพื้นหน้าที่ราบเรียบและเป็นลอนไม่ขรุขระ ส่วนสำหรับกล้องเดียวนั้นจะเป็นกรรมวิธีใหม่สำหรับการตรวจจับใบหน้าและติดตาม การใช้ข้อมูลสีของใบหน้าร่วมกับข้อมูลเชิงโครงสร้างของใบหน้า และได้มีการพัฒนากรรมวิธีการควบคุมอัตราบิตที่แปรผันได้ที่ให้ค่าคุณภาพที่คงที่ (CQ-VBR) ซึ่งรับรองคุณภาพที่ระบุไว้สำหรับแต่ละวัตถุที่ได้จากวิธีการแยกส่วนวัตถุทั้งสองแบบ กรรมวิธีทั้งคู่งานในระบบเวลา

จริงด้วยตัวประมวลผลราคาต่ำและได้ทดสอบในสภาพแวดล้อมภายในสำนักงานที่หลากหลาย
กรรมวิธีที่ใช้ควบคุมอัตราบิดที่แปรผันได้ที่ให้ค่าคุณภาพที่คงที่ เป็นเครื่องมือที่มีประโยชน์สำหรับ
คุณค่าของการเข้ารหัสเชิงวัตถุ สำหรับการประยุกต์ใช้งานทางด้านอัตราบิดต่ำ โดยสามารถ
ลดบิตเรตลงอย่างมีนัยสำคัญในอัตราบิดโดยรวมทั้งหมด ในขณะที่ยังคงมีคุณภาพทางการ
มองเห็นเดียวกันกับวัตถุพื้นหน้าและใบหน้าเมื่อเปรียบเทียบกับกับการเข้ารหัสเชิงเฟรมตามธรรมเนียม
ปกติ จากการทดสอบกับบางลำดับภาพที่ระดับความซับซ้อนหลายระดับ สามารถ
ประหยัดอัตราบิดประหยัดได้ถึงร้อยละ 48 การแยกส่วนพื้นหน้าที่พอใช้ได้เปิดโอกาสให้ส่งผ่าน
วัตถุพื้นหน้าปัจจุบันไปสู่จากวิดีโอเพื่อสร้างวิดีโอที่ประกอบไปด้วยวัตถุพื้นหน้าที่ปรากฏขึ้นมา
อย่างทันทีทันใด แต่การแยกส่วนพื้นหน้าต้องประกอบไปด้วยกล่องสองตัวทำให้มีข้อจำกัดและ
การแยกส่วนบริเวณใบหน้ามีความซับซ้อนเพิ่มขึ้นที่เกิดจากกระบวนการตัดสินใจ

1.2.2 งานวิจัยด้านการแยกส่วนสัญญาณวิดีโอเฉพาะบริเวณใบหน้า

เทคโนโลยีการแยกส่วนบริเวณใบหน้ามีบทบาทที่สำคัญในการเข้ารหัสสัญญาณวิดีโอที่
อัตราบิดต่ำเป็นอย่างมากสำหรับการประยุกต์ใช้งานทางด้านโทรศัพท์ภาพ เนื่องจากบริเวณที่
สำคัญที่สุดในโทรศัพท์ภาพคือบริเวณใบหน้าที่นั่นจึงได้มีการนำเอากรรมวิธีการแยกส่วนบริเวณ
ใบหน้ามาใช้ประโยชน์เพื่อกำหนดจำนวนบิตให้กับบริเวณส่วนของใบหน้าสูงกว่าบริเวณส่วนอื่น
[13] และ [14] นอกจากนี้แล้วยังมีการนำเอาขั้นตอนการตรวจจับใบหน้าไปประยุกต์ใช้ใน
งานในด้านต่าง ๆ [15] เช่น ระบบการระบุตัวบุคคลที่อ้างอิงกับการจดจำใบหน้า ระบบการค้นคืน
ภาพที่อ้างอิงแบบคอนเทนต์โดยใช้การตรวจจับใบหน้าเป็นต้น เนื้อหาในส่วนนี้จะกล่าวถึงภาพรวม
ของงานวิจัยทางการแยกส่วนบริเวณใบหน้าที่เกี่ยวข้องกับการเข้ารหัสสัญญาณวิดีโอเป็นหลัก

Shi Kewei, Fu Xitian, Cai Anni และ Sun Jingao เสนอการแยกส่วนใบหน้าแบบ
อัตราบิดในภาพ YCrCb [16] โดยการตรวจจับใบหน้าที่มนุษย์อย่างอัตราบิดในโทรศัพท์ภาพด้วย
รูปแบบ YCrCb โดยการใช้ช่วงค่าของฮิสโตแกรม (histogram) Cr-Cb ในการกำจัดพื้นหลังและ
การใช้กรรมวิธีการเติบโตของบริเวณ (region growing) ในพื้นที่ที่เหลืออยู่ (remaining areas)
ดังนั้นเทคนิคนี้จะสามารถตรวจจับใบหน้าได้อย่างแม่นยำจากภาพพื้นหลังด้วยวิธีที่เสนอนี้โดยการ
เติบโตของบริเวณจะกระทำกับแกน Y และท้ายสุดบริเวณใบหน้าจะได้รับการตรวจจับโดยการใช้
มาตรฐานที่ขึ้นกับคุณลักษณะ (characteristics) ของใบหน้าที่มนุษย์ วิธีนี้มีข้อดี คือ เนื่องจาก
รูปแบบของข้อมูลภาพในลำดับภาพวิดีโอมีฟอร์แมตเป็น YCrCb ตาม CCIR Recommendation
601 ดังนั้นปริภูมิสีแบบ YCrCb จึงสะดวกต่อการประยุกต์ใช้อย่างมากที่สุดและยังมีการคำนวณ
ที่ไม่ซับซ้อนโดยรายละเอียดของเทคนิคนี้คือการนำ Cr-Cb ในการแยกส่วนของสี Y ใช้ในการจับ

แชมปัง แล้วเข้าสู่การกำจัดพื้นหลังแล้วจึงใช้กรรมวิธีการเติบโตของบริเวณ เนื่องจากวิธีนี้ใช้เพียงค่าฮิสโตแกรมสองมิติของ Cr-Cb ในการแยกบริเวณใบหน้าจากพื้นหลังดังนั้นจึงมีประสิทธิภาพจำกัดและอาจเกิดความผิดพลาดได้และนักวิจัยจำนวนมากได้นำเสนอว่าปริภูมิสีแบบ HSI จะเหมาะสมในการแยกส่วนใบหน้าอย่างอัตโนมัติมากกว่าแต่ใช้ได้กับภาพส่วนบริเวณที่มีผิวหนังปรากฏ เช่น ศีรษะและไหล่ (head-and-shoulder video) เท่านั้น และถึงแม้ว่าภายหลังการแยกส่วนด้วยสีแล้วจะเหลือพื้นที่ในแกน Y เล็กน้อยแต่วิธีการเติบโตของบริเวณมีความซับซ้อนในการคำนวณสูง

Chi Zhan และ P. Wang ได้นำเสนอกรรมวิธีการแยกส่วนภาพสีชนิดใหม่บนพื้นฐานของความเข้มแสงและกลุ่มของข้อมูลสี [18] ซึ่งมีแนวคิดคือการใช้กรรมวิธีการจัดกลุ่มในปริภูมิสีแบบ HSI และใช้ประโยชน์ได้ครอบคลุมกับปริภูมิสี RGB อีกด้วย นอกจากนี้ยังมีการใช้ประโยชน์ของทั้งองค์ประกอบสี (hue component) และองค์ประกอบความเข้มแสง (intensity component) ในกระบวนการจัดกลุ่ม (clustering) คุณสมบัติพิเศษทางไซคลิก (cyclic) ขององค์ประกอบสี (hue) ได้ถูกนำมาพิจารณาและมีการสร้างกรรมวิธีการจัดกลุ่มสีขึ้นมาใช้

P. Sharma และ R. Reilly นำเสนอการประยุกต์ใช้กรรมวิธีมาซซิงอย่างรวดเร็วกับการระบุตำแหน่งใบหน้าในการประยุกต์ใช้งานทางโทรศัพท์ภาพโดยการใช้ข้อมูลสี [19] ซึ่งแนวคิดคือวิธีการแยกส่วนใบหน้ามนุษย์อย่างอัตโนมัติจากลำดับภาพที่ประกอบไปด้วยลักษณะของศีรษะและไหล่ โดยใช้วิธีการของฟาสมาซซิงเลเวลเซต (fast marching level set) วิธีการนี้เป็นกรรมวิธีที่มีความเร็วในการคำนวณสูง มีความน่าเชื่อถือสูงและมีประสิทธิภาพในการคำนวณสูงเมื่อนำมาใช้กับข้อมูลสีของภาพ ในภาพที่จะทำการแยกส่วนบริเวณใบหน้านั้นข้อมูลสีจะได้รับการนำมาใช้แยกประเภทจุดภาพภายในภาพเป็นประเภทต่าง ๆ โดยขึ้นอยู่กับค่าของข้อมูลสีสำหรับจุดภาพนั้น ๆ ซึ่งทำได้โดยการใช้กรรมวิธีเลเวลเซต (level set) แบบฟาสมาซซิงที่กำหนดจลนภาพไว้ก่อนล่วงหน้า (priority) ในการระบุบริเวณใบหน้ามีการใช้ความสว่างในการแยกระหว่างพื้นหลังกับบริเวณใบหน้า เนื่องจากบริเวณพื้นหลังจะมีความสว่างสม่ำเสมอแต่บริเวณใบหน้ามีความสว่างไม่สม่ำเสมอดังนั้นถึงแม้ว่าพื้นหลังจะมีสีเดียวกับสีผิว เทคนิคนี้ก็จะสามารถกำจัดได้เพราะสีพื้นหลังจะมีความสว่างที่สม่ำเสมอกว่าส่วนบริเวณใบหน้า

Douglas Chai และ King N. Ngan ได้นำเสนอการแยกส่วนใบหน้าโดยใช้สีผิวในการประยุกต์ใช้งานทางโทรศัพท์ภาพ [8] ซึ่งแนวคิดคือ การนำเสนอการใช้แบบจำลองสีผิว YCbCr เพื่อใช้กับภาพวิดีโอมาแยกข้อมูลที่เป็นสีผิวออกจากกลุ่มข้อมูลทั้งหมด และนำการตรวจสอบความถูกต้องทางรูปทรงเชิงเรขาคณิตของใบหน้ากับเส้นโครงร่างมากำจัดขอบของข้อมูลภาพใบหน้ามนุษย์ที่ไม่สม่ำเสมอออกจากภาพ เนื่องจากวิธีการที่ใช้สนใจเฉพาะข้อมูลที่เป็นสีผิวเท่านั้น

ดังนั้นวิธีนี้จึงมีปัญหในเรื่องของการแยกสีผิวและให้ผลลัพธ์ไม่ดีเท่าที่ควรเมื่อจำนวนใบหน้ามนุษย์ในภาพมีจำนวนมาก

Sobottka และ Pitas ได้นำเสนอขั้นตอนการค้นหาใบหน้ามนุษย์จากภาพสี [22] โดยประกอบด้วย 2 ขั้นตอนคือขั้นตอนแรกนำแบบจำลองสีแบบ HSV มาแยกส่วนสีผิวออกจากภาพ โดยจะนำเฉพาะกลุ่มข้อมูล HS มาวิเคราะห์เพื่อสร้างตัวกรองสีผิวและขั้นตอนที่สองจะนำกลุ่มข้อมูลสีผิวภาพมาเปรียบเทียบกับภาพวงรีโดยอาศัยแนวคิดที่ว่าโครงร่างใบหน้าของมนุษย์มีลักษณะคล้ายกับวงรีเพื่อใช้ค้นหาตำแหน่งภาพใบหน้ามนุษย์ แต่เนื่องจากส่วนขององค์ประกอบสี Hue จะขาดความเชื่อถือได้ในการแยกส่วนจุดภาพในกรณีที่ส่วนขององค์ประกอบความอิ่มตัวของสี (saturation) มีค่าต่ำ ๆ

Mei-Juan Chen, Ming-Chieh Chi, Ching-Ting Hsu และ Jeng-Wei Chen ได้นำเสนอโครงสร้างการตรวจจับบริเวณสีผิวและการปรับปรุงคุณภาพของบริเวณที่สนใจ [10] โดยการใช้ข้อมูลสีในการกำหนดบริเวณที่สนใจโดยการใช้แผนที่แสดงบริเวณสีผิวในการแยกว่าบริเวณใดเป็นบริเวณที่สนใจและบริเวณที่ไม่สนใจ การใช้แบบจำลองแบบเกาส์เซียนในการแสดงรูปแบบของสีผิวด้วยค่าความแปรปรวนและค่าเฉลี่ยในการตรวจจับสีผิวนั้นใช้ปริภูมิสี YCbCr แต่ที่ไม่สนใจขององค์ประกอบสี Cb โดยใช้เพียงองค์ประกอบสี Cr เท่านั้นในการนิยามสีผิว ผลจากการใช้ขององค์ประกอบสีองค์ประกอบเดียวทำให้ผลการแยกส่วนในบางลำดับภาพให้ผลลัพธ์ที่ไม่ดีนักจึงทำให้ต้องมีการเพิ่มกระบวนการขจัดส่วนที่ไม่ต้องการโดยการใช้วิธีการขจัดสัญญาณรบกวนอย่างง่ายซึ่งใช้ค่าขีดเริ่มเปลี่ยน (threshold) ของค่าความแปรปรวนของปริภูมิสี RGB ในการขจัดบริเวณที่ไม่ต้องการที่มีช่วงค่าเดียวกับสีผิว ทำให้มีความซับซ้อนเพิ่มขึ้นเพราะต้องทำการคำนวณค่าความแปรปรวนของภาพที่ได้จากการใช้ช่วงค่าขององค์ประกอบสี Cr เพียงองค์ประกอบเดียว

M. M. Fleck, D. Forsyth และ C. Bregler ได้นำเสนอการสร้างตัวกรองสีผิวเพื่อใช้ในงานวิจัยสำหรับค้นหาภาพเปลือย [23] ซึ่งใช้ปริภูมิสีแบบ HSV ร่วมกับสมการทางคณิตศาสตร์ที่มีฟังก์ชันลอการิทึม (logarithm) เข้ามาช่วยเพื่อตัดส่วนของสีที่มีลักษณะความเข้มต่างกันออกไป เนื่องจากคุณสมบัติของฟังก์ชันลอการิทึมจะทำให้ความแตกต่างของความเข้มของสีให้ชัดเจนยิ่งขึ้น ถึงแม้ว่าตัวกรองสีผิวให้ผลลัพธ์ที่ค่อนข้างดีแต่มีข้อเสียคือต้องใช้เวลาในการคำนวณสูงเนื่องจากมีการคำนวณค่าฟังก์ชันลอการิทึม

J. Fritsch, S. Lang, M. Kleinhagenbrock, G. A. Fink และ G. Sagerer นำเสนอการปรับปรุงการแยกส่วนสีผิวที่ปรับตัวได้ [24] โดยขั้นตอนวิธีในการตรวจหาใบหน้าจะใช้แบบจำลองสีผิวที่มีการปรับตัวได้โดยพิจารณาการกระจายตัวของจุดภาพที่เป็นสีผิวในปริภูมิย่อย rgb (ซึ่งเป็นค่าแอมพลิจูด RGB) โดยจะมีรูปร่างคล้ายหอยเชลล์ (shell shape) จึงกำหนดขอบเขตล้อมรอบอาณาบริเวณนี้โดยใช้สมการเชิงพาราโบลา 2 เส้นและกรรมวิธีที่ใช้แบบจำลอง

สีจะถูกปรับตัวอยู่ตลอดเวลาโดยวิธีการคือการนำข้อมูลสีผิวที่ตรวจหาได้ ณ ปัจจุบันมาทำการปรับปรุงแบบจำลอง ข้อเสียของวิธีการนี้คือการรบกวนกันในปริภูมิสี นอร์แมลไลซ์ RGB เกิดขึ้นได้ง่ายกว่าปริภูมิสีอื่นๆ โดยเฉพาะบริเวณตำแหน่งจุดภาพที่มีค่าใกล้เคียงกับสีดำนั้นมีความแปรปรวนได้อย่างง่ายและเนื่องจากค่าของจุดภาพจำพวกนี้มีโอกาสที่จะเป็นค่าใดก็ได้ตั้งแต่ค่าต่ำสุดจนถึงสูงสุดในปริภูมิสีนอร์แมลไลซ์ RGB ดังนั้นจึงส่งผลให้แบบจำลองสีผิวที่สร้างขึ้นตีความหมายผิดพลาดได้

Son Lam Phung, Douglas Chai และ Abdesselam Bouzerdoum ได้เสนอการแยกส่วนสีผิวแบบปรับตัวได้ในภาพสี [25] โดยเทคนิคที่นำเสนอจะใช้แบบจำลองสีผิวมนุษย์บนพื้นฐานการตัดสินใจแบบเบย์ที่ใช้ในการแยกส่วนที่เป็นสีผิวและส่วนที่ไม่ใช่สีผิวเนื่องจากการใช้ปริภูมิสีแบบ RGB ในการดำเนินการและใช้คุณสมบัติความเป็นเนื้อเดียวกันของพื้นผิวที่เป็นสีผิวของมนุษย์ในการเลือกค่าขีดเริ่มเปลี่ยนที่สามารถแยกแต่ละบริเวณสีผิวจากพื้นหลังเมื่อพื้นหลังนั้นมีสีใกล้เคียงสีผิวของมนุษย์ซึ่งค่าขีดเริ่มเปลี่ยนนี้จะปรับตัวตามบริเวณภาพใดๆ โดยที่ในการใช้กฎการตัดสินใจแบบเบย์นั้นจะใช้กับวิธีการที่ไม่ใช่พารามิเตอร์ คือ การใช้ฮิสโตแกรมเพราะว่าวิธีการฮิสโตแกรมมีการประมาณค่าของเงื่อนไขที่ใช้ในการตัดสินใจที่แม่นยำ ในกรรมวิธีการเลือกค่าขีดเริ่มเปลี่ยนนั้นอยู่บนพื้นฐานของข้อสังเกตที่ว่าสีผิวของมนุษย์นั้นมีพื้นผิวที่ราบเรียบ ในขณะที่บริเวณของภาพที่มีสีผิวซึ่งปนกันระหว่างบริเวณที่เป็นผิวมนุษย์กับพื้นหลังมีแนวโน้มที่จะไม่เป็นเนื้อเดียวกัน ดังนั้นค่าขีดเริ่มเปลี่ยนจึงเลือกจากการเพิ่มค่าขีดเริ่มเปลี่ยนจากค่าเริ่มต้นของค่าขีดเริ่มเปลี่ยนทีละค่าจนกระทั่งบริเวณสีผิวที่แยกได้จะกลายเป็นเนื้อเดียวกัน ซึ่งการกระทำดังนี้ทำให้เวลาที่ใช้ในการแยกส่วนมีค่าเพิ่มขึ้นตามจำนวนการปรับค่าขีดเริ่มเปลี่ยน

Q. Huynh-Thu, M. Meguro และ M. Kaneko เสนอวิธีการแยกสีผิวจากรูปภาพซึ่งมีฉากหลังที่ซับซ้อนและมีการเปลี่ยนแปลงความสว่าง [26] โดยเทคนิคที่นำเสนอจะใช้แบบจำลองแบบเกาส์ผสมรวมกับเทคนิคการปรับปรุงค่าทดสอบหลายๆ ค่าแบบอัตโนมัติโดยคำนวณบนปริภูมิสี HS ข้อด้อยของเทคนิคคือมีความซับซ้อนในการคำนวณสูงและใช้เวลาในการประมวลผลนานดังนั้นจึงอาจมีผลกระทบกับระบบที่ต้องการให้ทำงานตามเวลาจริงเมื่อสัญญาณวิดีโอที่รับเข้ามีอัตราเฟรมที่สูง

Nugroho H., Takahashi S. K., Ooi Y. และ Ozawa S. เสนอการตรวจจับใบหน้ามนุษย์จากลำดับภาพโดยการใช้กรรมวิธีเชิงพันธุกรรม [42] โดยการค้นหาภาพใบหน้าจากภาพที่ได้จากกล้องวิดีโอ และนำมาแปลงเป็นภาพในแต่ละเฟรมขนาด 256x240 จุดภาพ แล้วนำภาพไปผ่านการหาขอบ และใช้แบบจำลองวงรีในการค้นหา โดยเริ่มจากการค้นหาส่วนหัวที่อยู่ภายในภาพซึ่งใช้กรรมวิธีเชิงพันธุกรรมในการเลือกค่าต่างๆ ของวงรีที่เหมาะสม การทำงานจะประกอบไปด้วย 3 ส่วนหลักคือ ค้นหาเฉพาะส่วนหัว หาอัตราส่วนของสีเข้มกับสีอ่อนภายในภาพใบหน้า และ

สุดท้ายวัดความสมมาตรของตาซ้าย ตาขวา และปาก การใช้กรรมวิธีเชิงพันธุกรรมในการหาตำแหน่ง ขนาดและการเอียง ของภาพใบหน้าที่เหมาะสมอาจจะใช้เวลาในการค้นหานาน เพราะมีลักษณะการค้นหาคล้ายกับการสุ่มค่าจนกว่าจะได้ค่าตำแหน่งของภาพใบหน้าออกมา

1.2.3 งานวิจัยด้านการแยกส่วนสัญญาณวิดีโอสำหรับวัตถุที่มีความหมาย

งานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุเริ่มขึ้นประมาณช่วงปี ค.ศ. 1980 เป็นต้นมา เทคนิคในช่วงเริ่มแรกจะเป็นการนำเทคนิคที่ใช้กับภาพนิ่งมาประยุกต์ใช้กับภาพวิดีโอในแต่ละเฟรมแต่เนื่องจากว่าเป็นปี ค.ศ.1994 มีการร่างมาตรฐานการเข้ารหัสวิดีโอ MPEG-4 โดยมาตรฐาน MPEG-4 เป็นมาตรฐานที่ครอบคลุมถึงการเข้ารหัสแบบวัตถุด้วยการแยกส่วนภาพวิดีโอออกเป็นวัตถุที่มีความหมายจะเริ่มวิจัยกันอย่างแพร่หลายตั้งแต่ปีค.ศ.1994 เป็นต้นมาและเป็นผลทำให้งานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุที่มีความหมายแพร่หลายมากขึ้น งานวิจัยด้านการแยกส่วนวิดีโอสามารถแบ่งออกเป็นแนวทางหลักๆ ได้ 2 ประเภทคือ

- การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติ : การนำภาพวิดีโอมาทำการแยกส่วน โดยที่ไม่ต้องใช้ผู้ใช้ช่วยกำหนดข้อมูลบางอย่าง โดยกรรมวิธีของแต่ละบทความนั้น อาจจะทำที่ละเฟรม สองเฟรมติดกัน หรือทำทุกเฟรมซ้ำหลายรอบก็ได้
- การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบกึ่งอัตโนมัติ : การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบกึ่งอัตโนมัติ เป็นการนำภาพวิดีโอมาทำการแยกส่วน โดยที่ต้องใช้ผู้ใช้ช่วยกำหนดข้อมูลพื้นฐานบางอย่างเช่น สี ตำแหน่ง หรือรูปร่างอย่างคร่าวๆของวัตถุ เป็นต้น

ส่วนถัดไป จะกล่าวถึงงานวิจัยในด้านนี้และกรรมวิธีของแต่ละงานที่นำเสนอโดยสรุป

Hotter และ Thoma ได้เสนอแบบแผนการแยกส่วนภาพโดยอยู่บนพื้นฐานของการประมาณพารามิเตอร์การแมปปิงเชิงวัตถุ [37] กรรมวิธีนี้แบ่งเป็น 4 ขั้นตอน

- ขั้นแรก การเคลื่อนที่ของกล้องจับภาพจะถูกประมาณและชดเชยการเคลื่อนที่โดยใช้แบบจำลองการเคลื่อนที่ 8 พารามิเตอร์
- ขั้นที่สอง การตัดฉาก (scene cut) และการเลื่อนของกล้องจับภาพ (camera panning) จะถูกตรวจรู้โดยการประเมินค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนระหว่างสองเฟรมที่ต่อเนื่องกัน แล้วจะนำเฉพาะบริเวณพื้นหลังของเฟรมก่อนหน้ามาพิจารณา ส่วนกรณีที่เกิดการตัดฉาก ขั้นตอนนี้จะทำใหม่อีก
- ขั้นที่สาม มาสก์ของการเปลี่ยนแปลงจะถูกสร้างขึ้นโดยใช้ข้อมูลระหว่างเฟรมสองเฟรมที่ต่อเนื่องกัน โดยใช้เทคนิครีแลกเซชัน (relaxation technique) และค่าจุดเปลี่ยนท้องถิ่น

(local threshold) ซึ่งจะพิจารณาสถานะของจุดภาพข้างเคียง ส่วนหน่วยความจำจะถูกนำมาใช้สำหรับเก็บมาส์กการเปลี่ยนแปลง เพื่อสร้างบริเวณภาพวัตถุที่มีเสถียรภาพ

- **ขั้นที่สี่** มาส์กของวัตถุเคลื่อนที่จะได้จากการกำจัดพื้นหลังที่ไม่ถูกครอบคลุม (uncovered background) จากมาส์กการเปลี่ยนแปลงแล้วปรับแต่งบริเวณขอบของภาพ การประมาณการเคลื่อนที่และการสร้างมาส์กของวัตถุ จะอยู่บนพื้นฐานของการประมาณพารามิเตอร์การแมปปีงเชิงวัตถุ ส่วนการตรวจรู้การเปลี่ยนแปลงถูกปรับปรุงมาจากกรรมวิธีของ Aach [38] ซึ่งมีแนวคิดสำคัญ คือ การสร้างมาส์กของวัตถุเริ่มแรกจากการตรวจรู้การเปลี่ยนแปลง และนำมาสก็นไปใช้ในการติดตามการเคลื่อนที่ของวัตถุที่สนใจด้วย

การสร้างมาส์กของวัตถุเริ่มแรก จะใช้การขดเขยการเคลื่อนที่โดยรวม เพื่อขดเขยการเคลื่อนที่ของกล้องจับภาพ หลังจากนั้นจึงใช้การตรวจรู้การเปลี่ยนแปลง กรรมวิธีนี้สามารถให้กล้องจับภาพเลื่อน (panning) และขยายภาพ (zooming) ได้ ในการประมาณการเคลื่อนที่ทั้งเฟรม จุดภาพที่ใกล้ขอบภาพทั้งซ้ายและขวาน้อยกว่า 10 จุดภาพจะใช้เป็นจุดอ้างอิง ซึ่งจะสมมุติว่าไม่มีการเคลื่อนที่ที่ใกล้กับขอบภาพนั้น

Chen Sipei, Li Zongjie และ Zhang Guilin ได้เสนอกรรมวิธีการสกัดวัตถุเคลื่อนที่ออกจากพื้นหลังในระบบติดตามวิถีโอบแบบเวลาจริง [27] ซึ่งกรรมวิธีนี้มีแนวคิดคือ การใช้ทฤษฎีการตัดสินใจแบบเบย์ โดยไม่จำเป็นต้องรู้ความน่าจะเป็นก่อน (prior probability) ซึ่งมีเงื่อนไขการตัดสินใจว่าเป็นวัตถุเคลื่อนที่เมื่อ

$$\frac{h(x | T)}{h(x | B)} > \text{Threshold} \quad (1.1)$$

โดยกำหนดค่า Threshold เท่ากับ 8 $h(x|T)$ คือ ความน่าจะเป็น (likelihood) ของวัตถุ และ $h(x|B)$ คือ ความน่าจะเป็นของพื้นหลัง x คือ ความเข้มแสง T คือ คลาสของวัตถุ B คือ คลาสของพื้นหลัง ผลลัพธ์จากแนวคิดดังกล่าวใช้ได้กับภาพขาวดำ

A. Neri, S. Colonnese, G. Russo และ P. Ralone เสนอการแยกส่วนของวัตถุที่เคลื่อนที่และพื้นหลังอย่างอัตโนมัติ [39] โดยสร้างการแยกส่วนของแต่ละเฟรมในลำดับภาพด้วยการประมวลผลเป็นกลุ่มของเฟรม $\{f_{i-1}, i = 0, \dots, n\}$ จำนวนของเฟรม (n) ขึ้นกับความเร็วของวัตถุเคลื่อนที่ โดยหลังจากการขดเขยการเคลื่อนที่โดยรวมแล้ว จะมีกระบวนการ 3 ขั้นตอน คือ การตรวจรู้พื้นที่ที่เปลี่ยนแปลง การตรวจรู้การเคลื่อนที่ และการปฏิบัติตามข้อบังคับ (regularization) โดยจะทำการตรวจรู้การเปลี่ยนแปลงกลุ่มของเฟรมด้วย Higher Order Statistics เพื่อหลีกเลี่ยงความแปรปรวนของความเข้มแสงเนื่องจากจุดสัญญาณรบกวน มาส์กของวัตถุเคลื่อนที่ที่สร้างจากการประมาณการเคลื่อนที่บน HOS map ที่ไม่รวมพื้นหลังที่ไม่ถูกครอบคลุม (uncovered background) และในขั้นตอนของการปฏิบัติตามข้อบังคับ บริเวณที่ระบุ

ว่าอยู่หนึ่งที่อยู่ภายในบริเวณที่ระบุว่าเคลื่อนที่จะถูกกำหนดเป็นบริเวณที่เคลื่อนที่ หรือฉากหน้า (foreground) กรรมวิธีนี้ไม่จำเป็นต้องใช้การตัดฉากและการติดตามการเคลื่อนที่

Fabric Heitz และ Patrick Bouthmy เสนอการประมาณการเคลื่อนที่ และการแยกส่วนภาพ วิดีโอออกเป็นวัตถุโดยใช้วิธีโกลบอลเบย์เซียน [28] ซึ่งกรรมวิธีดังกล่าวนี้เป็นการหาขอบเขตของ วัตถุที่เคลื่อนไหวภายในเฟรม โดยใช้ข้อมูลจากการตรวจจับขอบของภาพนิ่งและข้อมูลจากการ ประมาณการเคลื่อนที่ ซึ่งการหาข้อมูลเหล่านั้นจะกระทำผ่านทางแบบจำลองเชิงสถิติมาร์คอฟ แรนดอมฟิลด์ (markov random field)

กรรมวิธีนี้ให้ผลลัพธ์และความเที่ยงตรงสูงเมื่อเป็นลำดับภาพที่สังเคราะห์ขึ้นเองโดยเทียบกับวิธีหาขอบโดยการไหลของแสง (optical flow) วิธีอื่นๆ แต่เมื่อนำมาใช้กับลำดับภาพจริงยังคง ให้ผลที่ดีเฉพาะตรงบริเวณที่เคลื่อนไหวเท่านั้น

Chiou-Shann Fuh และ Petros Maragos เสนอการตรวจจับการเคลื่อนที่ของภาพโดยใช้ แบบจำลองแอฟฟายน์ (affine models) [29] ซึ่งเป็นแบบจำลองการเคลื่อนที่ของวัตถุ (ซึ่งจะ นำมาใช้กับทั้งภาพก็ได้) โดยวัตถุจะเคลื่อนที่ไปในทิศทางแนวราบ แนวตั้ง เคลื่อนที่แบบหมุน ขยายวัตถุ หรือหดวัตถุได้ โดยสามารถเขียนเป็นสมการได้ดังนี้

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} A_5 \\ A_6 \end{bmatrix} \quad (1.2)$$

โดย (x', y') คือตำแหน่งของจุดภาพในเฟรมปัจจุบัน (x, y) เป็นตำแหน่งของจุดภาพใน เฟรมอ้างอิง A_1 ถึง A_6 เป็นค่าพารามิเตอร์ที่จะทำให้เกิดการเคลื่อนที่ตามแบบจำลองแอฟฟายน์ การหาเวกเตอร์การเคลื่อนที่ของแบบจำลองแอฟฟายน์ทำได้โดยทำให้ฟังก์ชันความ ผิดพลาดมีค่าน้อยที่สุดคือ

$$\text{Min } E(A_1, \dots, A_6) = \sum_{x, y \in R} |I_i(x', y') - I_{i+1}(x, y)|^2 \quad (1.3)$$

โดย I_{i+1} คือความเข้มแสงของภาพเฟรมที่ i และ I_i คือความเข้มแสงของภาพเฟรมที่ $i+1$ จากกรรมวิธีที่นำเสนอได้ทำการทดลองกับภาพที่สังเคราะห์การเคลื่อนไหวขึ้น ซึ่งผลลัพธ์ ที่ได้มีความราบเรียบกว่าการประมาณด้วยการไหลของแสงโดยวิธีอื่นๆ และยังสามารถนำไป ประยุกต์ใช้กับภาพอินฟราเรดแสดงการเคลื่อนไหวของเมฆด้วย ส่วนข้อด้อยของกรรมวิธีนี้คือ ต้องการการคำนวณสูงมาก และยังคงสอดคล้องกับข้อสมมติที่ว่า ภาพ (หรือวัตถุ) ต้องหมุนไปใน ทิศทางเดียวกัน

Raynard O. Hinds และ Thrasyvoulos N. Pappas เสนอกรรมวิธีการแยกส่วนภาพวิดีโอ ออกเป็นวัตถุโดยสามารถปรับเปลี่ยนการแบ่งกลุ่มได้ [30] แนวคิดของวิธีนี้คือ นำกรรมวิธีที่ใช้กับ ภาพใบนารี 2 มิติ ไปใช้กับภาพวิดีโอทุกภาพ โดยกรรมวิธีที่ใช้มีทั้งเชิงพื้นที่และเชิงเวลา

ผลลัพธ์ของกรรมวิธีที่เสนอได้ออกมาเป็นภาพใบนาวิสเก็คซ์ ในการกระทำแต่ละครั้ง จะต้องใช้ภาพทั้งหมดทุกเฟรม เพราะฉะนั้นไม่สามารถเป็นเวลาจริงได้และต้องการการคำนวณสูงมากแต่กรรมวิธีที่เสนอเป็นแบบอัตโนมัติ ซึ่งไม่ใช่คนเกี่ยวข้องเลย

Jean-Mare Odobez และ Patrick Bouthemy เสนอการใช้การเคลื่อนที่ช่วยในการแยก ส่วนภาพวิดีโอออกเป็นวัตถุ โดยสมมติว่าพื้นหลังเป็นไปตามแบบจำลองมาร์คอฟแรนดอมฟิลด์ ซึ่งแบบจำลองของการประมาณการเคลื่อนที่นั้นจะใช้แบบจำลอง 2 มิติ และประมาณการเคลื่อนที่แบบทนทาน (robust estimation) [31] ผลลัพธ์ที่ได้จากการกระทำโดยแนวความคิดนี้ สามารถแยกส่วนตัวคนหรือวัตถุเคลื่อนที่อื่นๆ ออกจากพื้นหลังได้ดีพอสมควร แต่การคำนวณค่อนข้างซับซ้อนยุ่งยาก และมีค่าตัวแปรบางค่าต้องหาจากการทดลอง ซึ่งไม่ตรงกันตามแต่ละชนิดของวิดีโอ

Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee Myoung Ho Lee, Chieteuk และ Yo-Sung Ho เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติ โดยพิจารณาข้อมูลเชิงพื้นที่และเชิงเวลาร่วมกัน [32]

โครงสร้างของกรรมวิธีนี้คือการนำภาพเฟรมที่ $k+1$ และเฟรมที่ k มาประมาณการเคลื่อนที่แบบโกลบอล และขดเขยการเคลื่อนที่แบบโกลบอล จากนั้นนำไปตรวจจับหาการเปลี่ยนแปลง แล้วผ่านกระบวนการแยกส่วนภาพเชิงเวลา เพื่อนำมาตัดสินใจว่าพื้นที่ใดเป็นวัตถุ พื้นที่ใดเป็นพื้นหลัง โดยพิจารณาเทียบกับภาพที่ได้จากการแยกส่วนภาพทางพื้นที่

การประมาณการเคลื่อนที่แบบโกลบอลทำได้โดยประมาณการเคลื่อนที่โดยใช้แบบจำลอง แอฟฟายน์ ซึ่งหาเวกเตอร์การเคลื่อนที่โดยใช้วิธีของกำลังสองน้อยที่สุด การแยกส่วนทางเวลาทำได้โดยใช้การทดสอบสมมติฐาน ซึ่งจำเป็นต้องรู้ข้อมูลของพื้นหลังก่อนด้วย การแยกส่วนภาพเชิงพื้นที่ใช้กรรมวิธีสันปันน้ำ (watershed algorithm) การตัดสินใจก็พิจารณาว่าภาพที่ได้จากการแยกส่วนเชิงพื้นที่ พื้นที่ใดอินเตอร์เซกกับพื้นที่ที่ได้จากการแยกส่วนเชิงเวลา ถ้าครอบคลุมมากกว่าร้อยละ 90 ก็ให้ตัดสินใจว่าเป็นวัตถุ

ผลลัพธ์จากการทดลองสามารถแยกส่วนได้ แต่ที่นำมาเสนอในบทความคือภาพที่เป็นวิดีโอคอนเฟอร์เรนซ์เพียง 2 ลำดับภาพวิดีโอมาตรฐาน

Candemir Toklu, A. Murat Tekalp และ A. Tanju Erden เสนอการแยกส่วนวิดีโอแบบกึ่งอัตโนมัติ (semi-automatic) ในกรณีที่เกิดการบดบังกันของวัตถุ [33] แนวคิดคือ ใช้โครงข่ายเมช (mesh) สี และตำแหน่งเข้าช่วยในการแยกส่วนซึ่งผู้ใช้ต้องช่วยแยกส่วนวัตถุทุกๆ 10 เฟรม

Yaakov Tsaig และ Amir Averbuch เสนอการแยกส่วนวัตถุเคลื่อนที่แบบอัตโนมัติ โดยใช้เทคนิคการกำหนดชื่อพื้นที่ [34] กรรมวิธีเริ่มจากนำภาพเฟรมที่ k และ $k+1$ มาประมาณการเคลื่อนที่โดยวิธีเกรเดียนต์แล้วหาเวกเตอร์การเคลื่อนที่ตามแบบจำลอง Perspective ซึ่งมี 8

พารามิเตอร์ โดยใช้วิธี Levenberg Marguardt (LM) Nonlinear minimization algorithm จากนั้นนำไปตรวจจับการเปลี่ยนแปลงของฉาก โดยหาความแตกต่างของภาพเฟรมที่ k และ $k+1$ แล้วรวมผลของค่าสัมบูรณ์ของความแตกต่างนั้น ถ้ามากกว่าค่าที่ตั้งไว้ ก็ให้ตัดสินว่าเป็นการเปลี่ยนแปลง จากนั้นตรวจสอบว่ามีการเคลื่อนไหวของกล้องหรือเปล่า ถ้ามีให้ทำการชดเชยแบบโกลบอล จากนั้นนำมาหาเกรเดียนต์ แล้วรวมพื้นที่ย่อยโดยใช้ข้อมูลเชิงพื้นที่และเชิงเวลามาพิจารณาร่วมกัน เพื่อแยกส่วนอย่างคร่าวๆ แล้วนำไปแบ่งส่วน โดยวิธีทางความน่าจะเป็นและสถิติ เริ่มด้วยแบ่งส่วนโดยใช้วิธีทดสอบทางสถิติ แบ่งว่าพื้นที่ใดคือพื้นหน้า พื้นที่ใดคือพื้นหลัง จากนั้นก็ใช้แบบจำลองมาร์คอฟแรนดอมฟิลด์ ในการแบ่งส่วนอย่างละเอียด แล้วทำการมิไนไมซ์ (minimize) ด้วยวิธี HCF (highest confident first)

ผลลัพธ์ที่ได้ใช้ได้กับลำดับวิดีโอมาตรฐานถึง 5 ลำดับวิดีโอ ซึ่งการแยกส่วนเป็นไปอย่างอัตโนมัติ แต่จากการวิเคราะห์หัดแล้ว กรรมวิธีจะให้ผลลัพธ์ที่ดีสำหรับลำดับภาพที่มีการเคลื่อนไหวปานกลางถึงมาก ลำดับภาพที่มีการเคลื่อนไหวน้อยยังไม่นำมาแสดงผล ทั้งนี้กรรมวิธีที่เสนอยังมีความซับซ้อนในการคำนวณสูงมากเนื่องจากต้องมีการประมาณการเคลื่อนที่ และยังคงทำการมิไนไมซ์ (minimize) ฟังก์ชันความน่าจะเป็นอีกด้วย

Hui zhu และ Zaiming Li เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยใช้ข้อมูลเชิงพื้นที่และเวลาพิจารณาร่วมกัน [35] แนวคิดก็ไม่ต่างจากหลายบทความที่กล่าวมา ที่แตกต่างคือในส่วนของการประมาณการเคลื่อนที่แบบโกลบอลนั้นใช้การค้นหาแบบเต็ม (full search) และหาพารามิเตอร์ของแบบจำลองแอฟฟายน์ ด้วยวิธีกำลังสองน้อยที่สุด

Jong Bae Kim และ Hang Joon Kim เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยใช้การแปลงเวฟเลทร่วมกับกรรมวิธีสันปันน้ำ [36] แนวคิดคือ ทำการแปลงเวฟเลทกับภาพก่อนแล้วนำส่วนความละเอียดต่ำ (low resolution) ของชั้นที่ 2 มาทำด้วยอัลกอริทึมสันปันน้ำเพื่อแยกส่วนภาพ กรรมวิธีมีความรวดเร็วในการประมาณผล แต่ผลลัพธ์ยังคงจำกัดอยู่เฉพาะบางลำดับภาพวิดีโอ

Daehee Kim และ Yo-Sung Ho เสนอการเสนอเชิงรูปร่างสำหรับการแยกส่วนวิดีโอ [40] เป็นการเสนอกรรมวิธีการแยกส่วนแบบผู้มีส่วนร่วมเพราะว่าการแยกส่วนวิดีโอแบบอัตโนมัติไม่สามารถแยกส่วนได้อย่างแม่นยำ ดังนั้นการแยกส่วนแบบกึ่งอัตโนมัติโดยการให้ข้อมูลรูปร่างเพื่อที่จะลดผลกระทบที่ไม่ต้องการที่เกิดจากการประมวลผลภาพระดับล่าง จึงได้ใช้เส้นโครงร่างแบบเคลื่อนที่ได้ในการแยกส่วนภาพ โดยการให้ผู้ใช้งานกำหนดจุดเริ่มต้นของเส้นโครงร่างรอบบริเวณวัตถุที่สนใจเพื่อทำการแยกส่วนภาพ แต่วิธีการยังต้องมีการกำหนดเส้นโครงร่างใกล้ตัววัตถุและไม่ได้เป็นระบบอัตโนมัติ

Changick Kim และ Jenq-Neng Hwang เสนอการแยกส่วนและติดตามวัตถุวิดีโออย่างรวดเร็วและอย่างอัตโนมัติสำหรับการประยุกต์ใช้งานเชิงวัตถุ [43] กรรมวิธีที่ใช้เป็นการหาขอบภาพจากความแตกต่างระหว่างเฟรมและการหาความแตกต่างระหว่างขอบของพื้นหลังกับขอบภาพปัจจุบันที่ได้จากขั้นตอนความแตกต่างระหว่างเฟรม แล้วจึงหาขอบของวัตถุที่เคลื่อนที่โดยเทียบกับเฟรมก่อนหน้า อีกครั้งทำให้สามารถแยกวัตถุออกมาจากภาพได้ กรรมวิธีที่เสนอเหมาะสำหรับการประยุกต์ใช้งานทางด้านการรักษาความปลอดภัยด้วยวิดีโอที่ซึ่งใช้สำหรับตรวจจับวัตถุที่ปรากฏขึ้นมาใหม่อย่างอัตโนมัติและติดตามวัตถุที่สนใจนั้นไป แต่ทว่าสำหรับลำดับภาพประเภทวีดิโอและไฟล์ต้องให้การกำหนดค่าเริ่มต้นด้วยผู้ใช้ก่อนการประมวลผล

1.3 แนวทางของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้มีแนวคิดที่จะเสนอกรรมวิธีการแยกส่วนภาพบริเวณที่สนใจที่เป็นบริเวณใบหน้า และบริเวณตัววัตถุที่สนใจ โดยกำหนดว่ามีข้อมูลที่ใช้ในการเลือกบริเวณที่สนใจสำหรับการจัดสรรบิตอยู่แล้ว โดยจะนำไปประยุกต์ใช้กับวิดีโอที่มีลักษณะเป็นวิดีโอเฟรมหรือวิดีโอคอนเฟอเรนซ์ โดยการแยกส่วนบริเวณที่สนใจบริเวณใบหน้า ประกอบไปด้วย การใช้องค์ประกอบสี เพื่อทำการแยกส่วนที่เป็นบริเวณใบหน้าออกมา แต่อย่างไรก็ตามบริเวณของใบหน้าที่ได้จากการแยกส่วนยังคงเป็นการแยกส่วนที่ไม่สมบูรณ์จึงใช้กับกรรมวิธีเชิงสัญญาณซึ่งเป็นวิธีที่สามารถกำจัดบริเวณที่ไม่ใช่บริเวณใบหน้าออกไปได้ ทำให้สามารถแยกส่วนบริเวณที่สนใจ คือ เพียงบริเวณใบหน้าออกมาได้อย่างสมบูรณ์ ส่วนการแยกส่วนบริเวณที่สนใจที่เป็น วัตถุวิดีโอในเฟรมภาพอยู่บนพื้นฐานของการรวมกันของคุณลักษณะสำคัญหลายอย่างของสัญญาณวิดีโอ โดยประกอบด้วยข้อมูลเชิงพื้นที่ ข้อมูลเชิงเวลาและข้อมูลบริเวณใบหน้าของวัตถุ การแยกส่วนเชิงพื้นที่ประกอบด้วยขั้นตอนการสร้างความเรียบง่ายให้กับภาพ การตรวจรู้ขอบเขตของบริเวณ และการกำหนดบริเวณให้แก่จุดภาพโดยกรรมวิธีสันปันน้ำ ส่วนข้อมูลเชิงเวลาเป็นการใช้ข้อมูลความแตกต่างระหว่างเฟรมและข้อมูลส่วนบริเวณใบหน้าของวัตถุนำไปใช้ในการกำหนดขอบเขตบริเวณให้กับวัตถุเพื่อเป็นข้อมูลรูปร่างก่อนหน้า (shape prior) เพื่อแก้ไขข้อจำกัดของการแยกส่วนภาพเชิงพื้นที่เวลา เมื่อวัตถุไม่มีการเคลื่อนที่หรือมีการเคลื่อนที่น้อยมากจนกระทั่งไม่สามารถตรวจรู้การเปลี่ยนแปลงได้ โดยการใช้ข้อมูลบริเวณใบหน้าจะทำให้สามารถระบุขอบเขตของวัตถุได้แม้ว่าจะไม่มีการเคลื่อนที่ของวัตถุก็ตาม ซึ่งข้อมูลนี้จะได้มาจากขั้นตอนการแยกส่วนบริเวณที่สนใจที่เป็น บริเวณใบหน้า จากนั้นนำผลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อทำให้สามารถแยกส่วนภาพบริเวณที่สนใจที่เป็นวัตถุได้

1.4 เป้าหมายและขอบเขตของวิทยานิพนธ์

1. สามารถแยกส่วนภาพบริเวณที่สนใจที่เป็นบริเวณใบหน้าและที่เป็นวัตถุได้ เมื่อวัตถุเหล่านั้นมีส่วนของใบหน้าปรากฏอยู่
2. การแยกส่วนภาพบริเวณที่สนใจแบบอัตโนมัติ

1.5 ขั้นตอนและวิธีการดำเนินงาน

1. ศึกษาบทความและงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์
 - 1.1) ศึกษาบทความและงานวิจัยที่เกี่ยวข้องกับกรรมวิธีการแยกส่วนภาพทั้งแบบอัตโนมัติ และกึ่งอัตโนมัติ
 - 1.2) ศึกษาบทความและงานวิจัยที่เกี่ยวข้องกับกรรมวิธีการตรวจจับบริเวณใบหน้า
 - 1.3) ศึกษาบทความและงานวิจัยที่เกี่ยวข้องกับวิธีการควบคุมอัตรา
2. ทดลองกรรมวิธีต่าง ๆ ที่ศึกษามาเพื่อพิจารณาถึงข้อดีและข้อเสียของแต่ละกรรมวิธี
3. นำเสนอกรรมวิธีการแยกส่วนภาพที่คาดว่าจะให้ประสิทธิภาพที่สูงและทดสอบกรรมวิธีนั้น
4. วิเคราะห์ผลของกรรมวิธีที่เสนอ โดยพิจารณาถึงข้อดี ข้อเสีย และข้อที่ควรปรับปรุง
5. สรุปผลการทดลองและเขียนวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงกระบวนการและกรรมวิธีที่เหมาะสมสำหรับการแยกส่วนภาพวิดีโอสำหรับบริเวณที่สนใจ ซึ่งนำไปใช้เป็นบริเวณที่ใช้ในการจัดสรรบิตที่เลือกบริเวณที่สนใจได้
2. สร้างแบบจำลองที่เหมาะสมในการตรวจหาบริเวณสีผิวของมนุษย์
3. สามารถนำเทคนิคที่เสนอไปประยุกต์ใช้กับอุปกรณ์สื่อสารต่อไปในอนาคตได้

จุฬาลงกรณ์มหาวิทยาลัย

1.7 คำโครงวิทยานิพนธ์

สำหรับเนื้อหาของวิทยานิพนธ์ฉบับนี้แบ่งรายละเอียดออกเป็น 5 บทดังต่อไปนี้

บทที่ 1 บทนำ กล่าวถึง แนวเหตุผล ภาพรวมของงานวิจัยที่เกี่ยวข้อง แนวทางของวิทยานิพนธ์ เป้าหมายและขอบเขตของวิทยานิพนธ์ ขั้นตอนและวิธีการดำเนินงาน ประโยชน์ที่คาดว่าจะได้รับ และคำโครงวิทยานิพนธ์

บทที่ 2 ทฤษฎีพื้นฐาน เนื้อหาในบทนี้จะกล่าวถึง แบบจำลองรูปร่างที่เคลื่อนที่ได้ เช่น สเนก เทมเพลตที่เปลี่ยนแปลงรูปร่างได้ พลังงานรูปแบบต่างของสเนก ปริภูมิสีต่างๆ การประมวลผลภาพดิจิทัลที่นำมาใช้ในงานวิจัย การประมวลผลภาพเชิงสัญญาณ (morphological image processing) การตรวจหาขอบ (edge detection)

บทที่ 3 กรรมวิธีที่นำเสนอ เนื้อหาในบทนี้กล่าวถึง โครงสร้างของระบบการแยกส่วนบริเวณใบหน้าและการแยกส่วนภาพบริเวณที่สนใจที่เป็นวัตถุ ในรายละเอียดประกอบไปด้วยสองส่วนหลัก คือ ส่วนการแยกส่วนบริเวณใบหน้า จะสร้างแบบจำลองที่ใช้ในการแยกส่วนสีผิวมนุษย์โดยใช้การใช้องค์ประกอบของสองปริภูมิสีร่วมกัน คือ องค์กรประกอบสี Cr ในปริภูมิสี YCbCr และองค์กรประกอบสี Hue ในปริภูมิสี HSI การเติมเต็มในส่วนที่ขาดหายและกำจัดหุลุมขนาดเล็กของมาส์กด้วยกรรมวิธีเชิงสัญญาณ และส่วนการติดตามบริเวณใบหน้าเพื่อขจัดพื้นหลังที่ไม่ต้องการโดยมีสมมุติฐานว่าใบหน้ามีตำแหน่งของการเคลื่อนที่ไม่ต่างไปจากบริเวณเดิมมากนัก

ส่วนการแยกส่วนภาพบริเวณที่สนใจที่เป็นวัตถุ กล่าวถึงการแยกส่วนภาพทางพื้นที่ การแยกส่วนภาพเชิงเวลา การใช้ข้อมูลบริเวณใบหน้าของวัตถุเพื่อกำหนดขอบเขตบริเวณให้กับวัตถุเพื่อเป็นข้อมูลรูปร่างก่อนหน้า กระบวนการตัดสินใจและการติดตามวัตถุ

บทที่ 4 ผลการวิจัย กล่าวถึง ผลของเทคนิคที่นำเสนอในการแยกส่วนบริเวณใบหน้า ผลการแยกส่วนสีผิวโดยใช้แบบจำลองชนิดไม่มีตัวแปร เปรียบเทียบใน 3 ปริภูมิสี คือ CbCr, HS และ HCr ผลของการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุที่เป็นผลของการแยกส่วนภาพเชิงพื้นที่ การแยกส่วนภาพเชิงเวลา การติดตามวัตถุเพื่อช่วยในการแยกส่วนภาพ สุดท้ายเป็นการวิเคราะห์ผลการแยกส่วนภาพ

บทที่ 5 บทสรุป กล่าวถึง ข้อสรุปผลการวิจัยและข้อเสนอแนะ

บทที่ 2

ทฤษฎีพื้นฐาน

2.1 แบบจำลองรูปร่างที่เคลื่อนที่ (Active shape models)

แบบจำลองรูปร่างที่เคลื่อนที่ประกอบไปด้วยรูปแบบที่หลากหลาย เช่น สเนก (snake) เทมเพลตที่เปลี่ยนแปลงรูปร่างได้ (deformable template) และโครงร่างพลวัต (dynamic contour) สเนกเป็นกลไกการใช้ความรู้ก่อนหน้า (prior knowledge) เช่น รูปร่างพื้นผิว สี หรือตำแหน่งของวัตถุที่นำไปสู่การแปลความหมายของภาพในระดับล่าง (low-level) แทนที่จะคาดหวังถึงคุณสมบัติที่ต้องการ อาทิเช่น ความต่อเนื่องและความราบเรียบที่จะได้จากข้อมูลภาพกลับกลายเป็นว่าคุณสมบัติเหล่านั้นจะได้รับการกำหนดไว้ตั้งแต่ต้น โดยเฉพาะอย่างยิ่ง แบบจำลองของความต่อเนื่องที่ขยับขยายตัวได้ (elastic) และเส้นโค้งที่ยืดหยุ่นได้มีการกำหนดให้โดยขึ้นกับความเหมาะสมและเข้ากันได้กับภาพนั้นๆ โดยการแปรเปลี่ยนค่าพารามิเตอร์ (parameter) หรือการควบคุมข้อสมมติฐานเบื้องต้นไว้อย่างถาวร การออกแบบเบื้องต้นจะทำให้เกิดลักษณะเฉพาะด้านเกิดขึ้นซึ่งกระทำได้โดยการสร้างส่วนประกอบของเส้นโค้งที่ยืดหยุ่นได้ที่ซึ่งกลุ่มของค่าพารามิเตอร์จะคอยควบคุมตัวแปรที่เกี่ยวข้องกับการเคลื่อนไหว เช่น ขนาดของส่วนย่อย (subpart) ต่างๆ ที่หลากหลายและมุมของส่วนเชื่อมต่อที่เชื่อมส่วนย่อยต่างๆ เข้าด้วยกัน ตัวอย่างของแบบจำลองที่เป็นที่รู้จักก็เช่น เทมเพลตที่เปลี่ยนแปลงรูปร่างได้ซึ่งเป็นกลไกที่มีประสิทธิภาพอย่างมากในการกำหนดตำแหน่งทางโครงสร้างในภาพ

2.1.1 สเนก (Snake)

ศิลปะของการตรวจจับคุณลักษณะต่างๆ ของภาพนั้นได้มีการศึกษามาเป็นอย่างมาก หลักการของการใช้มาสก์ (mask) หรือ ตัวดำเนินการ (operator) ที่ได้รับการออกแบบมาเพื่อให้สัญญาณเอาต์พุตที่มีค่าที่ดีที่สุดในบริเวณตำแหน่งที่มีค่าปรากฏออกมามากที่สุดตามประเภทของคุณลักษณะเฉพาะที่ได้เลือกไว้ ผลลัพธ์ที่ได้คือภาพใหม่หรือแมปตามคุณลักษณะ (feature map) ซึ่งจะให้ผลลัพธ์ที่ตอบสนองตามประเภทของคุณลักษณะที่ได้เลือกไว้ที่แต่ละจุดภาพ ถึงตอนนี้ก็กล่าวได้ว่าตัวดำเนินการคือภาพขนาดย่อยซึ่งใช้กราดครอบคลุมไปตลอดทั้งภาพโดยใช้การเทียบเคียงเชิงคณิตศาสตร์หรือคอนโวลูชัน (convolution) มาสก์เป็นภาพต้นแบบ (prototype) ของคุณลักษณะที่ต้องการค้นหาซึ่งโดยทั่วไปแล้วมีขนาดเล็ก สำหรับ

คุณลักษณะที่เป็นหุบเขานั้นมาส์ที่ใช้ควรเป็นฟังก์ชันความเข้มแสงรูปร่างเช่นรูปตัววี ผลลัพธ์ที่ได้จากกระบวนการเทียบเคียง (correlation) คือค่าของความสอดคล้องของภาพต้นแบบกับภาพในแต่ละตำแหน่งของภาพที่ทำการหาค่า

อย่างไรก็ตาม แมปตามคุณลักษณะเป็นเพียงการเริ่มต้นเท่านั้นเพราะว่าเป็นเพียงการทำให้ประเภทของคุณลักษณะที่ต้องการมีความโดดเด่นขึ้นแต่ไม่ได้กระทำการตรวจจับคุณลักษณะนั้นๆ การตรวจจับต้องการการตัดสินใจที่กระทำกับแต่ละจุดภาพ กฎการตัดสินใจอย่างง่ายที่สุดคือการทำให้คุณลักษณะปรากฏชัดเจนออกมาเมื่อค่าของคุณลักษณะเกินค่าขีดเริ่มเปลี่ยน (threshold) ค่าขีดเริ่มเปลี่ยนที่มีค่าคงที่มักจะให้ผลลัพธ์ที่ไม่ค่อยดียกเว้นในกรณีอย่างง่ายของบางสถานการณ์ เช่น วัตถุทึบแสงบนโต๊ะสีดำ ซึ่งเป็นสภาพการณ์ทั่วไป ๆ ในศาสตร์ของระบบการมองเห็นของเครื่องจักร (machine vision)

อย่างไรก็ตาม คุณลักษณะของใบหน้าไม่ได้มีลักษณะเช่นนั้นและถ้าค่าขีดเริ่มเปลี่ยนถูกกำหนดค่าไว้สูงมากจะทำให้ปรากฏช่องว่างที่ขอบขึ้น ถ้าค่าขีดเริ่มเปลี่ยนถูกกำหนดค่าไว้ต่ำมากจะทำให้ปรากฏขอบที่ไม่ใช่ขอบที่แท้จริงขึ้นได้ กระบวนการตรวจจับคุณลักษณะระดับล่างนั้นมีประสิทธิภาพแต่ไม่สามารถคาดหวังได้ว่าจะสามารถให้ผลลัพธ์เป็นโครงสร้าง (structure) ทางเรขาคณิต (geometric)

สเนกคือกรรมวิธีใหม่ที่จัดการกับข้อจำกัดของการประมวลผลระดับล่างเหล่านี้ แนวความคิดที่สำคัญคือการนำเอาแมปทางคุณลักษณะ $F(r)$ และใช้ $(-F(r))$ เป็นเหมือนเส้นทางที่สเนกซึ่งเป็นเส้นโค้งที่เปลี่ยนรูปร่างได้ $r(s)$, $0 \leq s \leq 1$, สามารถเคลื่อนที่โดยมีลักษณะเหมือนเช่นการเลื้อยได้

สเนกเป็นเส้นโครงร่างเชิงพารามิเตอร์ถูกนิยามขึ้นภายในโดเมนของภาพ คุณสมบัติและพฤติกรรมทั้งหมดของสเนกถูกระบุผ่านฟังก์ชันที่เรียกว่าฟังก์ชันของพลังงานโดยอุปมาเหมือนกับระบบเชิงฟิสิกส์ (physical system) เพื่อที่จะเข้าใจรูปแบบของสเนกนั้นมีความเหมาะสมอย่างมากที่ต้องเปรียบเทียบมันกับรูปแบบเชิงฟิสิกส์จริง ๆ โดยการอุปมาเหมือนระบบเชิงฟิสิกส์ที่สามารถขยายและมีการเคลื่อนที่ที่มองดูคล้ายการเคลื่อนที่ของงูนั้นเป็นเพราะมีแรงเสมือนมากระทำบนวัตถุนั้น ๆ ทำให้เกิดปรากฏการณ์ดังกล่าวข้างต้น

พิจารณาภาพขอบของรูปภาพเป็นเหมือนเส้นทางที่เส้นโครงร่างเชิงพารามิเตอร์สามารถเคลื่อนที่ในลักษณะเหมือนการเคลื่อนที่แบบงูเลื้อยนั้นจะต้องมีแรงที่จะกระทำไปบนเส้นโครงร่างและเคลื่อนที่ไปตลอดเส้นทางบริเวณขอบภาพเพื่อให้เส้นโครงร่างเข้าสู่ตำแหน่งที่สมดุลพลังงาน (energy equilibrium) การที่จะมีรูปแบบที่กระทำอย่างนี้ได้มันต้องประกอบไปด้วยองค์ประกอบ 2 องค์ประกอบ คือ อย่างแรก คือ การบอกให้สเนกรู้ว่าควรจะมีพฤติกรรมอย่างไร เช่น รู้ว่ารักษาตำแหน่งเริ่มต้นอย่างไร รู้ว่าสร้างมุมอย่างไร เป็นต้น อย่างสุดท้าย คือ การบอก

ให้สเนกรู้ว่าควรเคลื่อนที่ไปที่ใด เมื่อสิ่งที่ต้องการคือการทำให้เส้นโครงร่างเลื่อนไปตามขอบของวัตถุที่ต้องการในภาพ โดยที่บริเวณขอบคือบริเวณที่ค่าทางลบ (negative) ของภาพขอบมีค่าน้อย ดังนั้นสมการสมมูลควรสร้างขึ้นในแนวทางที่เส้นโครงร่างมีแนวโน้มที่จะทำให้พจน์ที่เกี่ยวข้องกับค่าทางลบของภาพขอบมีค่าน้อยลงที่สุด (minimize)

เมื่อเข้าใจแนวความคิดแล้วจึงเป็นการออกแบบระบบเชิงคณิตศาสตร์เพื่อให้สามารถสร้างสเนกขึ้นมาได้จริง ๆ แบบจำลองสเนกแบบจำลองแรกเสนอโดย Kass [19] ในปี ค.ศ. 1987 โดยที่ฟังก์ชันของพลังงานนี้เรียกว่า พลังงานสเนก (snake energy) ซึ่งได้รับการนิยามเป็น

$$E_{\text{snake}} = \int_0^1 \{ E_{\text{int}}(x(s)) + E_{\text{image}}(x(s)) \} ds \quad (2.1)$$

ที่ซึ่งตำแหน่งของสเนกบนภาพ คือ พารามิเตอร์ของเส้นโค้งบนระนาบสองมิติ $x(s) = (x(s), y(s))$

E_{int} เป็นพลังงานภายใน (internal energy) ของเส้นโค้งที่เนื่องมาจากการบิดงอ เป็นพลังงานที่ได้มาจากพารามิเตอร์ของเส้นโค้งเพื่อบังคับให้เส้นโครงร่างมีความราบเรียบ (smooth)

E_{image} เป็นแรงของภาพที่ผลักดันให้สเนกมุ่งไปสู่วัตถุที่ต้องการ บางบทความเรียกพจน์นี้ว่า พลังงานภายนอก (external energy)

รูปแบบพลังงานภายในที่เสนอได้รับการนิยามเป็น

$$E_{\text{int}} = \left(\alpha |x'(s)|^2 + \beta |x''(s)|^2 \right) / 2, \quad s \in [0,1]$$

(2.2)

ที่ซึ่ง $x'(s)$ เป็นอนุพันธ์อันดับที่ 1 และ $x''(s)$ เป็นอนุพันธ์อันดับที่ 2 ของ $x(s)$ เทียบกับ s

E_{image} กำหนดจากภาพ ดังนั้นเมื่อใช้สเนกแยกแยะวัตถุที่สนใจโดยพิจารณากลุ่มของค่าที่มีค่าน้อยในค่าทางลบของภาพขอบหรือในคุณลักษณะที่สนใจเช่น ขอบภาพ เมื่อพิจารณาภาพระดับสีเทา (gray level) $I(x, y)$ ซึ่งเป็นเหมือนฟังก์ชันของตัวแปรตำแหน่ง (x, y) ที่ต่อเนื่องแล้ว แบบจำลองสำหรับพลังงานภาพหรือพลังงานภายนอกนิยามได้เป็น

$$E_{\text{image}}^{(1)}(x, y) = -|\nabla I(x, y)|^2 \quad (2.3)$$

$$E_{\text{image}}^{(2)}(x, y) = -|\nabla [G_{\sigma}(x, y) * I(x, y)]|^2 \quad (2.4)$$

ที่ซึ่ง $G_{\sigma}(x, y)$ เป็นฟังก์ชันเกาส์เซียนสองมิติด้วยส่วนเบี่ยงเบนมาตรฐาน σ และ ∇ เป็นตัวดำเนินการเกรเดียนต์ (gradient operator) ถ้าภาพเป็นลายเส้นขาวบนพื้นดำ พจน์ $I(x, y)$ แทนค่าระดับสีเทาของภาพทำให้ประมาณพลังงานภาพหรือพลังงานภายนอกได้เป็น

$$E_{\text{image}}^{(1)}(x, y) = I(x, y) \quad (2.5)$$

$$E_{\text{image}}^{(2)}(x, y) = G_{\sigma}(x, y) * I(x, y) \quad (2.6)$$

พจน์อนุพันธ์อันดับที่หนึ่งของ $V(s)$ เทียบกับ s ทำให้ทราบถึงอัตราของการเปลี่ยนแปลงของความยาวของเส้นโครงร่างซึ่งหมายถึงการหาค่าความยาวของเส้นโครงร่าง ส่วนพารามิเตอร์

α เป็นพารามิเตอร์ถ่วงน้ำหนักซึ่งควบคุมความตึงของสเนก โดยกำหนดให้เส้นโครงร่างมีค่าน้อยกว่าหรือมากกว่าระดับของการยืดหรือหดตัวซึ่งทำให้สเนกมีพฤติกรรมเหมือนเส้นลวดที่ยืดหยุ่นได้ ค่าที่มากของ α หมายถึงการหดที่มากของสเนกตามทิศทางของแรง ดังนั้น α ได้รับการนิยามเป็นเหมือนค่าคงที่ความยืดหยุ่น

พจน์อนุพันธ์อันดับที่สองของ $x(s)$ แสดงถึงความสามารถในการโค้งมนหรือเว้า พารามิเตอร์ β เป็นพารามิเตอร์ถ่วงน้ำหนักซึ่งควบคุมความแข็งแกร่งของสเนก พารามิเตอร์ β มีอำนาจมากกว่าอัตราการเปลี่ยนแปลงของความยาวของเส้นโครงร่างในทิศทางปกติที่มุ่งไปยังขอบของภาพ พจน์นี้ทำให้สเนกมีพฤติกรรมคล้ายเส้นลวดแข็งแกร่งซึ่งหมายถึงว่าเส้นโครงร่างจะมีความราบเรียบและรูปร่างที่เป็นเส้นตรงแต่ไม่ยืดหยุ่น ถ้าค่าของ β มีค่าสูงเส้นโครงร่างมีความแข็งแรงและยากต่อการโค้งงอ ในขณะที่ถ้า β มีค่าน้อยทำให้เส้นโครงร่างสามารถทำเป็นมุมได้ โดยการปรับค่าพารามิเตอร์ 2 ค่านี้ ทำให้เส้นโครงร่างมีความยืดหยุ่นที่เหมาะสมซึ่งจะทำให้สเนกสามารถล้อมรอบวัตถุที่สนใจได้

2.1.1.1 พลังงานของภาพ (Image Energy)

พลังงานของภาพเป็นพจน์ที่สำคัญ ค่าเกรเดียนต์ของพลังงานของภาพเป็นแรงขับเคลื่อนให้สเนกมีการเคลื่อนที่ กล่าวได้ว่าผลของการแยกส่วนนั้นขึ้นกับการนิยามพลังงานของภาพนั้นได้ดีเท่าใด

จากนิยาม สเนกซึ่งหาค่าที่ต่ำที่สุดของพลังงานสเนกต้องเป็นไปตามสมการออยเลอร์ (euler)

$$\alpha x''(s) - \beta x'''(s) - \nabla E_{image} = 0 \quad (2.7)$$

ซึ่งสามารถพิจารณาเป็นเหมือนสมการสมดุลแรง

$$F_{int} + F_{image} = 0 \quad (2.8)$$

ที่ซึ่ง F_{int} คือ $\alpha x''(s) - \beta x'''(s)$ และ F_{image} คือ $-\nabla E_{image}$ F_{int} คือแรงภายในซึ่งแสดงถึงการยืดหดและการโค้งงอในขณะที่ F_{image} คือแรงภายนอกหรือแรงจากภาพที่ดึงสเนกให้มุ่งไปยังขอบของภาพที่ต้องการ ซึ่งรูปแบบของสมการสเนกนี้เหมาะสมอย่างมากสำหรับการศึกษาในทางทฤษฎี

ในการหาคำตอบของสมการที่ (2.7) นั้นสเนกได้รับการทำให้เป็นพลวัตโดยการให้ เป็นฟังก์ชันของเวลาพร้อมกับ s นั่นคือ $x(s, t)$ ดังนั้นอนุพันธ์แบบพาเซี่ยลของ x เทียบกับ t ได้รับการกำหนดให้เท่ากับด้านซ้ายมือของสมการที่ (2.7)

$$x_t(s, t) = \alpha x''(s, t) - \beta x'''(s, t) - \nabla E_{image} \quad (2.9)$$

2.1.1.1.1 สนามแรงต้นแบบ (The Original Force Field)

สนามต้นแบบที่เสนอโดย Kass นั้นพลังงานของภาพนิยามเป็นดังสมการที่ 2.5

$$E^{(1)}_{image}(x, y) = -I(x, y)$$

การหาพลังงานของภาพอาศัยข้อมูลจากค่าเกรเดียนต์ของภาพ โดยบริเวณที่มีค่าเกรเดียนต์สูงจึงมีความน่าจะเป็นบริเวณขอบของภาพสูง สำหรับภาพที่ประกอบไปด้วยวัตถุหลายเส้นนั้นสนามเวกเตอร์เกรเดียนต์ซึ่งเมื่อขยายไปที่บริเวณขอบนั้นพบว่าสนามแรงจะกระทำบนเส้นโครงร่างอย่างเป็นระเบียบและมีลักษณะที่เข้าและชี้ออกจากบริเวณขอบอย่างเห็นได้ชัดเจน แต่อย่างไรก็ตาม ถ้าวัตถุที่สนใจมีลักษณะพื้นผิวที่กลมกลืนกัน รูปแบบของพลังงานจะไม่ให้สนามแรงที่เหมาะสมเท่าที่ควรเมื่อขยายไปที่บริเวณขอบของสนามแรงนั้นพบว่าสนามแรงที่อยู่ภายนอกขอบของวัตถุที่พุ่งไปยังขอบของวัตถุแต่สนามแรงที่อยู่บริเวณภายในขอบของวัตถุจะชี้พุ่งออกไปจากขอบของวัตถุ ทำให้เส้นโครงร่างที่ถูกดึงด้วยสนามแรงนี้จะไม่หยุดอยู่ที่ขอบของวัตถุแต่จะเคลื่อนที่ต่อเนื่องไปยังจุดศูนย์กลางของวัตถุ ดังนั้นการคำนวณหาภาพขอบของภาพและใช้เป็นพลังงานของภาพดังสมการที่ (2.3)

$$E^{(1)}_{image}(x, y) = -|\nabla I(x, y)|^2$$

เมื่อขยายไปที่บริเวณขอบของวัตถุในสนามแรงแล้วจะพบว่าทั้งสนามแรงที่อยู่ทั้งภายในและภายนอกขอบของวัตถุจะมีทิศทางพุ่งเข้าหาขอบของวัตถุทำให้เส้นโครงร่างไม่ถูกดึงเข้าสู่จุดศูนย์กลางของวัตถุ แต่อย่างไรก็ตามจากขอบเขตของสนามเวกเตอร์เกรเดียนต์พบว่าเส้นโครงร่างอาจไม่ได้แตะขอบของวัตถุนั้นถ้าเส้นโครงร่างที่กำหนดไว้ในตอนเริ่มต้นไม่ได้ถูกวางไว้ ใกล้กับวัตถุ แนวทางหนึ่งที่จะขยายขอบเขตของสนามเพื่อเพิ่มระยะการตรวจจับก็คือการพยายามทำให้ขอบภาพมีความราบเรียบเพิ่มขึ้น ดังนั้น รูปแบบพลังงานที่ได้ปรับปรุงแล้วจะเป็นดังสมการที่ (2.4) ที่ซึ่ง $G_\sigma(x, y)$ เป็นฟังก์ชันเกาส์เซียนสองมิติด้วยส่วนเบี่ยงเบนมาตรฐาน σ สังเกตว่าค่าส่วนเบี่ยงเบนมาตรฐานที่มากจะทำให้ส่วนขอบของภาพมีลักษณะเบลอแต่สิ่งนี้ผลทั้งด้านดีและด้านลบ นั่นคือ ขอบเขตการตรวจจับจะเพิ่มขึ้นแต่ตำแหน่งของขอบจะลดความแม่นยำและความชัดเจนลงโดยเฉพาะเมื่อค่าส่วนเบี่ยงเบนมาตรฐานมีค่ามาก

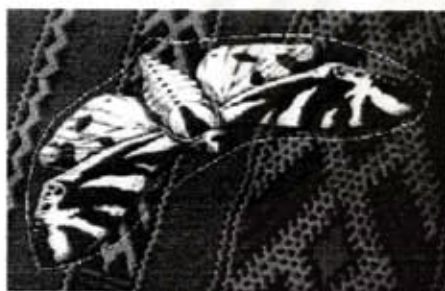
ในรูปที่ 2.1 (ก) แสดงค่าเกรเดียนต์ของภาพที่ได้รับการคำนวณเพื่อสนามจะได้ไปพุ่งเข้าไปยังบริเวณที่มีค่าเกรเดียนต์สูง รูปที่ 2.1 (ข) และ (ค) แสดงการหดตัวของสนามไปบนขอบของผีเสื้อ อย่างไรก็ตามบางจุดภาพไม่ปรากฏขอบที่ชัดเจนเพียงพอทำให้การหดตัวของสนามทำได้ไม่ค่อยดีนัก

ตัวอย่างของพฤติกรรมของสนามต้นแบบข้างต้นนั้นแสดงดังรูปที่ 2.2 รูปที่ 2.2 (ก) แสดงภาพวาดเส้นรูปร่างด้วยซึ่งมีขอบส่วนเว้าที่ด้านบน ลำดับของเส้นโค้งแสดงให้เห็นถึงการทำซ้ำตามลำดับของสนามต้นแบบที่มีการกำหนดจุดเริ่มต้นภายนอกวัตถุแต่อยู่ภายในย่านการตรวจจับ

ของสนามแรง ศักดา (potential force field) สนามแรงศักดา $F_{image} = -\nabla E^{(2)}_{image}$ แสดงดังรูปที่ 2.2 (ข) ผลลัพธ์สุดท้ายของภาพที่ 2.2 (ก) คือผลลัพธ์ที่ได้จากการแก้สมการออยส์เลอร์ของสเนกแต่ยังคงมีการแยกส่วนที่ข้ามส่วนเว้าของวัตถุ สาเหตุของการหลุดตัวของสเนกที่บกพร่องของสเนกแบบนี้แสดงให้เห็นในรูปที่ 2.2 (ค) ซึ่งแสดงภาพขยายของสนามแรงภายนอกภายในขอบบริเวณส่วนโค้งเว้า โดยถึงแม้ว่าแรงภายนอกซีไปยังขอบของวัตถุอย่างถูกต้องก็ตามแต่ว่าภายในส่วนขอบของบริเวณส่วนโค้งเว้าแรงซีไปในแนวนอนในทิศทางตรงกันข้าม ดังนั้นเส้นโครงร่างจึงถูกดึงไปยังส่วนที่ยื่นออกมาของวัตถุรูปร่างตัวยูแต่ไม่ได้ถูกดึงลงไปจนถึงภายในส่วนโค้งเว้า



(ก)



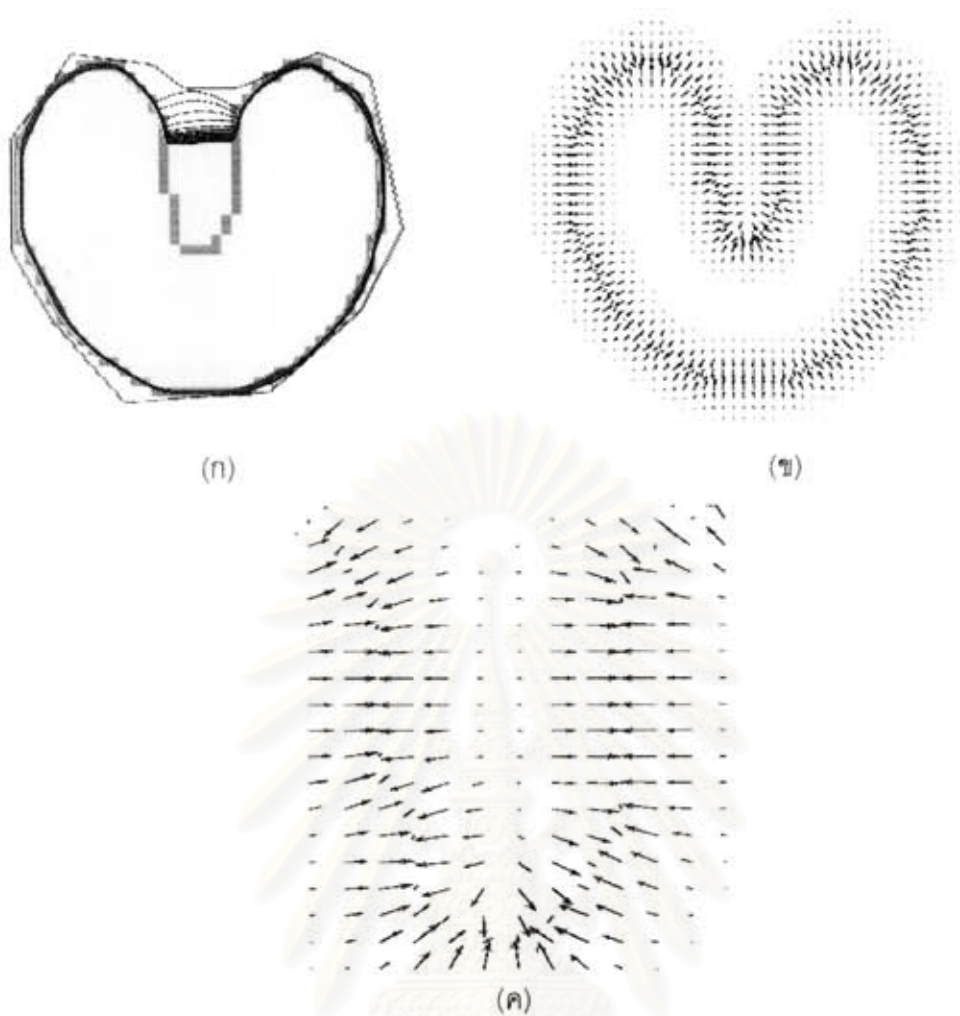
(ข)



(ค)

รูปที่ 2.1 (ก) แสดงค่าเกรเดียนต์ (ข) และ (ค) แสดงการหลุดตัวเข้าหาค่าเกรเดียนต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.2 แสดงพฤติกรรมของสเนกตันแบบ

(ก) แสดงภาพวาดเส้นรูปร่างด้วยซึ่งมีขอบส่วนเว้าที่ด้านบน

(ข) สนามแรง ศักดา (potential force field)

(ค) แสดงภาพขยายของสนามแรงภายนอกภายในขอบบริเวณส่วนโค้งเว้า

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

2.1.1.1.2 สนามแรงศักกักระยะทาง (The Distance Potential Force Field)

Cohen [20] ได้เสนอรูปแบบของแรงภายนอกซึ่งเพิ่มขอบเขตของการตรวจจับที่มากขึ้นอย่างมีนัยสำคัญกับสเนกตันแบบ แรงภายนอกเหล่านี้คือค่าเกรเดียนต์ทางลบของฟังก์ชันศักกัที่คำนวณมาจากการใช้แมประยะทางยูคลิเดียน (euclidean distance map) กล่าวอ้างได้ว่าแรงเหล่านี้เป็นเหมือนแรง ศักกัระยะทางซึ่งทำให้แยกประเภทของแรงนี้ออกจากแรง ศักกัแบบดั้งเดิม Cohen ได้ปรับปรุงสนามแรงภายนอกหรือสนามแรงของภาพ F_{image} โดยการนอร์มอลไลซ์สนามแรงภายนอก โดยให้

$$F_{image} = -k \nabla E_{image} / \|\nabla E_{image}\| \quad (2.10)$$

ที่ซึ่งผลคูณของขั้นเวลากับค่าคงที่ k จะอยู่บนอันดับของขนาดภาพ เพราะว่าขนาดของ F_{image} มีค่าประมาณหนึ่งจุดภาพ ดังนั้นเมื่อจุดของเส้นโค้งอยู่ใกล้กับจุดขอบภาพ มันถูกดึงดูไปยังขอบภาพและหยุดคงที่อยู่ ณ ที่ตรงนั้น ดังนั้นเกรเดียนต์ทั้งที่มีค่ามากและน้อยจะได้รับผลกระทบเช่นเดียวกันบนเส้นโค้ง

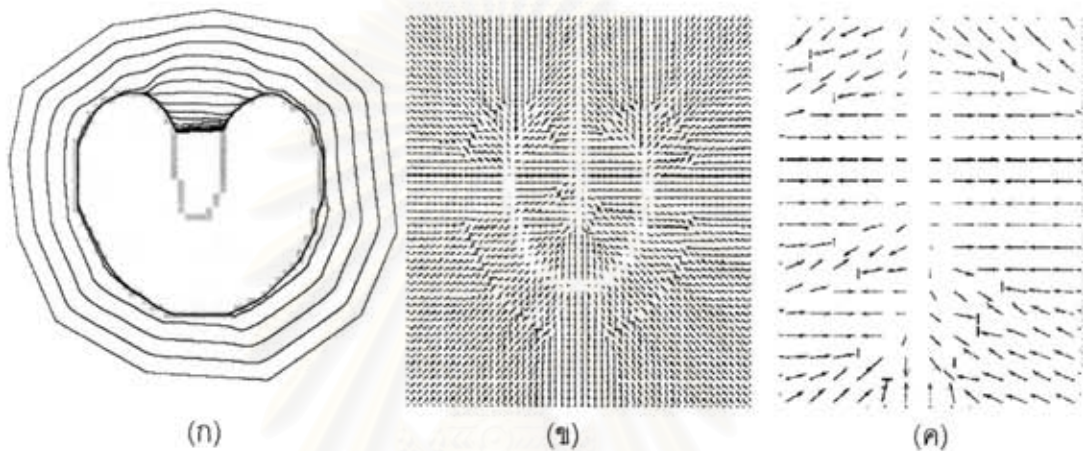
ผลกระทบจากสัญญาณรบกวนอาจทำให้บางจุดที่อยู่ห่างออกไปจากวัตถุและมีค่าเกรเดียนต์สูงไปทำให้เส้นโครงร่างหยุดอยู่ที่ตรงจุดนั้นเมื่อเส้นโครงร่างผ่านจุด ๆ นั้น เพื่อที่จะแก้ปัญหานี้ ได้มีการเพิ่มแรงอีกแรงหนึ่งขึ้นที่ทำให้เส้นโครงร่างมีพฤติกรรมพลวัตมากขึ้น โดยที่เส้นโค้งได้รับการพิจารณาเป็นเหมือนบอลลูในสองมิติที่ใส่ลมเข้าไป โดยการเพิ่มแรงส่วนหน้าเป็นแรงกดดันที่ผลักออกไปด้านนอกหรือเข้าไปด้านในขึ้นอยู่กับข้อกำหนดทิศทางของเวกเตอร์หนึ่งหน่วยไว้ในตอนแรก ดังนั้นแรงกลายเป็น

$$F = k_1 n(s) - k_2 \frac{\nabla E_{image}}{\|\nabla E_{image}\|} \quad (2.11)$$

ที่ซึ่ง $n(s)$ เป็นเวกเตอร์ที่ตั้งฉากกับเส้นโค้งที่จุด $v(s)$ และ k_1 เป็นขนาดของแรง ดังนั้นเมื่อเส้นโค้งขยายตัวและถูกดึงดูและหยุดลงโดยชอบเหมือนเช่นก่อนหน้านั้นแต่ถ้ามีแรงกดดันแล้วนั้นในกรณีที่ว่าหากว่าชอบมีค่าอ่อนมากเส้นโค้งก็จะสามารถผ่านขอบนี้ไปได้ หรือในกรณีที่ถ้าเส้นโค้งปะทะกับจุดบางจุดที่อยู่ห่างออกไปจากวัตถุแต่มีค่าเกรเดียนต์มากนั้นมีแนวโน้มว่าจะมีความไม่ต่อเนื่องเกิดขึ้นที่จุดนี้แรงนี้จะช่วยทำให้มีความราบเรียบเกิดขึ้นโดยการกำจัดความไม่ต่อเนื่องทิ้งไปและเส้นโค้งสามารถผ่านไปยังขอบภาพได้

รูปที่ 2.3 แสดงถึงประสิทธิภาพของสเนกที่ใช้แรงศักกัระยะทาง รูปที่ 2.3 (ก) แสดงวัตถุรูปตัวยูและลำดับของเส้นโครงร่างที่แสดงให้เห็นถึงการเคลื่อนที่ของสเนกจากจุดเริ่มต้นที่ไกลจากวัตถุค่อยๆ หดเข้าไปยังวัตถุ แรงศักกัระยะทางแสดงให้เห็นดังรูปที่ 2.3 (ข) ซึ่งมีเวกเตอร์จำนวนมากห่างไปจากวัตถุซึ่งแสดงให้เห็นว่าทำไมขอบเขตการตรวจจับสำหรับรูปแบบของแรง

ภายนอกประเภนี้ถึงมีขนาดกว้างมาก จากรูปที่ 2.3 (ก) แสดงให้เห็นว่าสเนกประเภทนี้จะทำงานได้ไม่ค่อยดีนักในกรณีที่หดตัวเข้าไปยังบริเวณส่วนเว้าของวัตถุ ซึ่งเห็นได้จากรูปที่ 2.3 (ข) ที่ขยายส่วนของแรง คัดดาระยะทางซึ่งพบว่ามีลักษณะเหมือนแรง คัดดาแบบสเนกดันแบบที่แรงมีการพุ่งไปในแนวนอนในทิศทางที่ตรงข้ามกันซึ่งทำให้มีการดึงสเนกลงมาแต่ไม่หดเข้าไปยังส่วนล่างสุดของขอบภายในส่วนเว้า เนื่องจากว่าแรง คัดดาระยะทางนี้ไม่ได้เปลี่ยนทิศทางของแรง แต่ที่เปลี่ยนไปมีเพียงขนาดของแรงเท่านั้น ดังนั้นปัญหาของการหดตัวเข้าหาส่วนเว้าของวัตถุจึงยังคงอยู่แม้ว่าจะใช้แรง คัดดาแบบระยะทางแล้วก็ตาม



รูปที่ 2.3 แสดงพฤติกรรมของสเนกแบบแรง คัดดาระยะทาง

- (ก) แสดงการหดตัวของสเนกที่ใช้ คัดดาระยะทาง
- (ข) สนามแรง คัดดาระยะทาง
- (ค) แสดงภาพขยายของสนามแรงภายในขอบบริเวณส่วนโค้งเว้า

2.1.1.1.3 สนามแรงแบบการไหลของเวกเตอร์เกรเดียนต์ (The Gradient Vector Flow Force Field)

รูปแบบของพลังงานที่เสนอโดย Cohen [20] นั้นแก้ปัญหบางประการของรูปแบบพลังงานของสเนก Xu [21] ได้เสนอแรงภายนอกแบบไม่พลวัตรูปแบบใหม่ซึ่งไม่เปลี่ยนแปลงไปตามเวลาหรือขึ้นกับตำแหน่งของสเนกเท่านั้น การอ้างอิงเชิงคณิตศาสตร์ของแรงชนิดใหม่นี้มาจากทฤษฎีบทเฮลโฮลท์ (helmholtz theorem) ที่กล่าวว่า สนามเวกเตอร์ที่คงที่ทั่วไปส่วนใหญ่สามารถแยกออกเป็นสององค์ประกอบ คือ องค์ประกอบที่ไม่หมุน ซึ่งมีอิสระในการโค้งงอหรือพันเป็นวง กับองค์ประกอบที่เป็นขด ซึ่งมีอิสระในการขยายตัวออก แรง คัดดาของสเนกดันแบบเป็นเหมือนสนามที่ไม่หมุนอย่างคงที่เมื่อมันเป็นเกรเดียนต์ของฟังก์ชัน คัดดาเชิงสเกลาร์ (scalar potential function) Xu ได้เสนอการสร้างสนามแรงภายนอกที่ใช้งานได้ทั่วไปมากยิ่งขึ้นโดย

การอนุญาตให้มีความเป็นไปได้ที่มีทั้งองค์ประกอบที่ไม่หมุน กับองค์ประกอบที่เป็นขดภายใน สนามแรงภายนอก สนามแรงภายนอกที่ได้รับการออกแบบนี้จึงมีคุณสมบัติที่ต้องการทั้ง ขอบเขตการค้นหาที่กว้างและให้แรงที่พุ่งไปยังส่วนเว้าของขอบวัตถุที่ดีซึ่งสนามแรงภายนอกนี้จะมี ทั้งองค์ประกอบที่ไม่หมุนและเป็นขดอยู่ทั้งสององค์ประกอบ

จากสมการที่ (2.9) ของสเนกตันแบบ สนามแรงภายนอกหรือสนามแรงของภาพ $-\nabla E_{image}$ ได้รับการแทนที่ด้วยสนามแรงภายนอกรูปแบบใหม่ซึ่งเรียกว่า สนามการไหลของ เวกเตอร์เกรเดียนต์ (gradient vector flow field) $v(x,y)$ ได้ผลเป็น

$$x_t(s,t) = \alpha x''(s,t) - \beta x'''(s,t) + v \quad (2.11)$$

เมื่อ $v(x,y) = [u(x,y), v(x,y)]$ เป็นสนามเวกเตอร์ที่หาค่าที่น้อยที่สุดของฟังก์ชันพลังงาน

$$\mathcal{E} = \iint \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |v - \nabla f|^2 dx dy \quad (2.12)$$

เมื่อ $f(x,y)$ เป็นแมปขอบภาพที่นิยามเป็น $f(x,y) = |\nabla I(x,y)|$ หรือ สำหรับภาพระดับสีเทาจะ กลายเป็น $f(x,y) = |\nabla(G_\sigma(x,y) * I(x,y))|$

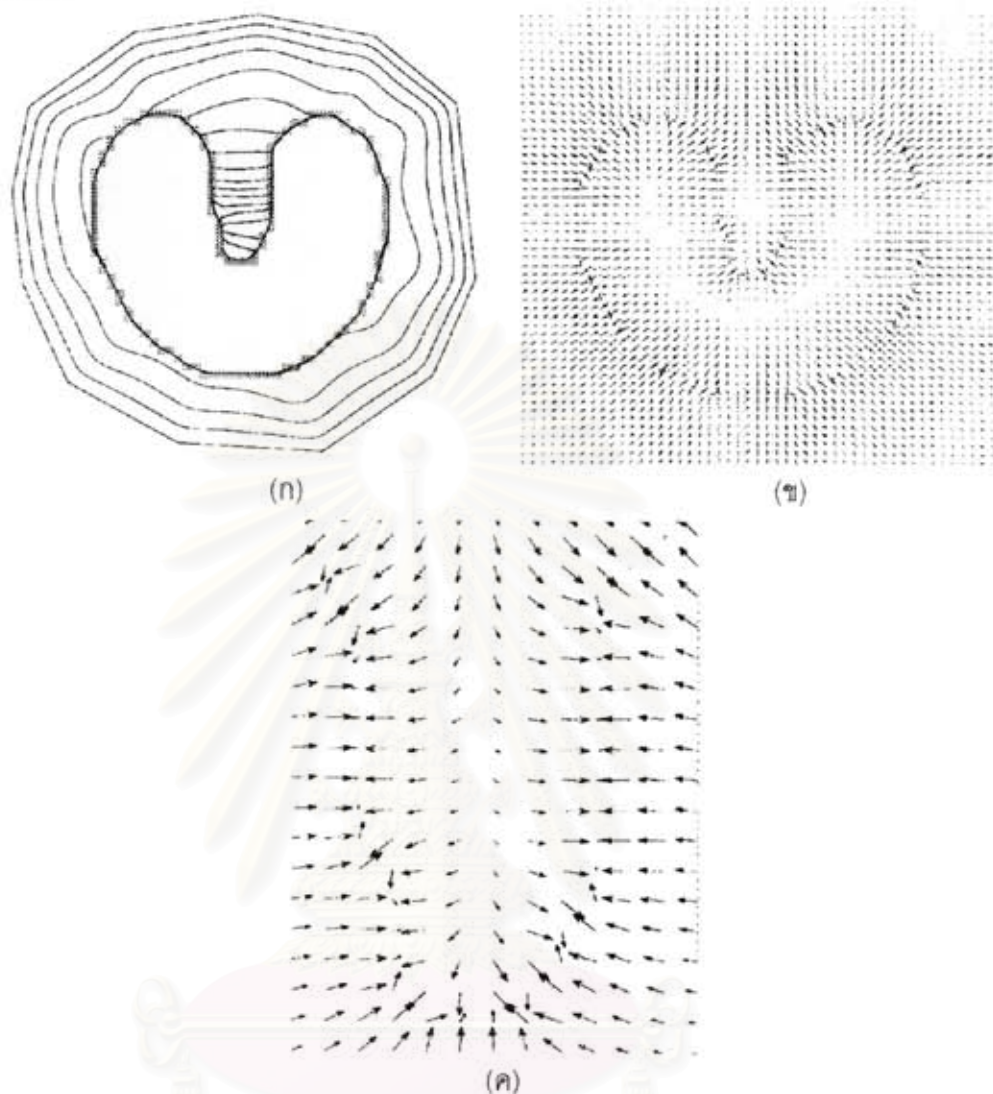
พบว่าเมื่อ $|\nabla f|$ มีค่าน้อย พลังงานจะตกอยู่ภายใต้ผลลัพท์ของผลรวมของกำลังสองของอนุพันธ์พาร์เชียลของสนามเวกเตอร์ ก่อให้เกิดสนามที่มีการเปลี่ยนแปลงอย่างช้า ๆ ในมุมกลับกันก็คือเมื่อ $|\nabla f|$ มีค่ามาก พจน์ที่สองจะมีอิทธิพลกับผลลัพท์ของสมการที่ (2.12) และสามารถหาค่าที่น้อยที่สุดได้โดยการกำหนดให้ $v = |\nabla f|$ ซึ่งสิ่งนี้ทำให้สามารถรักษาค่า v ไว้ให้ใกล้เคียงเท่ากับค่าเกรเดียนต์ของแมปขอบภาพไว้ได้ โดยที่ค่าพารามิเตอร์ μ เป็นพารามิเตอร์ประจำที่ใช้บังคับการถ่วงดุลระหว่างพจน์แรกกับพจน์ที่สองในสมการที่ (2.12) โดยที่พารามิเตอร์นี้จะถูกกำหนดตามลักษณะของสัญญาณรบกวนที่มีในภาพเมื่อมีสัญญาณรบกวนมากค่าพารามิเตอร์นี้ต้องเพิ่มค่าขึ้น

พจน์ที่ใช้ในการสร้างความราบเรียบนั้นคือพจน์แรกในสมการที่ (2.12) ซึ่งมีลักษณะเช่นเดียวกับการถ่วงน้ำหนักในการขยายตัวและการโค้งตัวของสนามเวกเตอร์

สนามเวกเตอร์ที่ต้องการจะคำนวณได้จากฟังก์ชันพลังงานโดยใช้แคลคูลัสของการผันแปร (calculus of variations)

รูปที่ 2.4 แสดงสนามการไหลของเวกเตอร์เกรเดียนต์ อย่างแรกพบว่าสนามการไหลของเวกเตอร์เกรเดียนต์เป็นเช่นเดียวกับสนามของแรงศักย์ดาระยะทางที่มีขอบเขตการค้นหาที่ค่อนข้างกว้างและกว้างกว่าสเนกตันแบบ อย่างที่สองพบว่าเมื่อขยายเวกเตอร์การไหลของเวกเตอร์เกรเดียนต์ที่อยู่ภายในบริเวณส่วนเว้าของวัตถุนั้นจะมีองค์ประกอบที่มีทิศทางพุ่งชี้ลงไป

ยังส่วนเว้าของวัตถุซึ่งแตกต่างกับสเนกตันแบบทำให้สามารถจัดปัญหาการแยกส่วนบริเวณส่วนเว้าของวัตถุได้



รูปที่ 2.4 แสดงพฤติกรรมของสเนกแบบการไหลของเวกเตอร์เกรเดียนต์

(ก) แสดงการเริ่มต้น การเคลื่อนตัว และเส้นโครงร่างสุดท้ายของสเนก

(ข) แสดงสนามการไหลของเวกเตอร์เกรเดียนต์

(ค) แสดงภาพขยายของสนามการไหลของเวกเตอร์เกรเดียนต์บริเวณส่วนเว้า

2.1.2 เทมเพลตที่เปลี่ยนแปลงรูปร่างได้

เทมเพลตที่เปลี่ยนแปลงรูปร่างได้เป็นวิธีการอีกวิธีการหนึ่งในการตรวจจับวัตถุ พื้นฐานทางคณิตศาสตร์ของเทมเพลตที่เปลี่ยนแปลงรูปร่างได้นี้มาจากทฤษฎีของการจัดจำรูปแบบ วิธีการนี้ให้ความรู้ก่อนหน้าเกี่ยวกับรูปร่างของวัตถุเป็นหลักในการแก้ปัญหา ข้อมูลรูปร่างก่อน

หน้านี้ได้รับการระบุลงไปเป็นเหมือนเทมเพลตโบนารีหรือต้นแบบเชิงพารามิเตอร์ (parametric prototype) ความแตกต่างระหว่างสเนกกับเทมเพลตที่เปลี่ยนแปลงรูปร่างได้คือสเนกเป็นฟังก์ชันการหาค่าที่น้อยที่สุดของพลังงานที่มีรูปร่างที่เป็นอิสระ สเนกนั้นไม่มีโครงสร้างของเส้นโค้งแบบครอบครัวกาลวันแต่ข้อจำกัดทั่ว ๆ ไปเช่น ความราบเรียบและความต่อเนื่องของขอบภาพ

ในทางตรงกันข้ามเทมเพลตที่เปลี่ยนแปลงรูปร่างได้จะควบคุมการเปลี่ยนแปลงรูปร่างโดยใช้กลุ่มของค่าพารามิเตอร์ซึ่งมีความสามารถในการเข้ารหัสรูปร่างเฉพาะเจาะจงได้ วิธีการนี้จึงจะได้รับการใช้งานก็ต่อเมื่อสามารถหาข้อมูลรูปร่างที่เฉพาะเจาะจงได้โดยการใช้เทมเพลตโบนารีหรือกลุ่มของพารามิเตอร์

เทมเพลตต้นแบบส่วนใหญ่แล้วจะเหมือนรูปร่างของวัตถุ โดยการใช้การแปลงเชิงพารามิเตอร์ของต้นแบบและเปลี่ยนแปลงขอบเขตของมันโดยการค่อย ๆ เปลี่ยนพารามิเตอร์การเปลี่ยนรูปร่างเพื่อที่จะตรวจจับวัตถุที่มีการเปลี่ยนแปลงรูปร่างได้ การเปลี่ยนแปลงเล็กน้อยอาจจะถูกตรวจจับได้ดีกว่าถ้าใช้การเปลี่ยนแปลงรูปร่างแบบสุ่มของต้นแบบซึ่งเทมเพลตที่เปลี่ยนแปลงรูปร่างได้นี้อาจจะสอดคล้องกับวัตถุที่สนใจได้ดีกว่าเทมเพลตอันแรก ในกรณีที่วัตถุมีสัญญาณรบกวนหรือถูกลดคุณภาพลงซึ่งทำให้รูปร่างตอนต้นสูญหายไปนั้นเทมเพลตที่เปลี่ยนแปลงรูปร่างได้วิธีการนี้อาจจะทำงานได้ดีกว่าเทมเพลตอันแรก

2.2 พื้นฐานของสี

ในปี ค.ศ. 1666 เซอร์ไอแซค นิวตัน พบว่าเมื่อแสงขาวฉายผ่านปริซึมจะเกิดสเปกตรัมของสีซึ่งมีช่วงตั้งแต่สีม่วงจนถึงสีแดงแต่ละสเปกตรัมของสีก็มีช่วงความยาวคลื่นที่ไม่เท่ากัน เช่นสีเขียวมีความยาวคลื่นในช่วง 500 ถึง 570 นาโนเมตร การเห็นสีของวัตถุเกิดจากวัตถุนั้นได้สะท้อนความถี่ในช่วงนั้นออกมาและได้ดูดซับความถี่ของสีอื่นๆไว้

แสงที่ไม่มีสีคือแสงที่มีสมบัติเดียวคือความเข้มแสง ความเข้มแสงจะเป็นตัวแสดงถึงระดับสี ซึ่งมีอยู่สองระดับสีคือระดับสีขาวและระดับสีดำ ค่าของความเข้มแสงจะอยู่ในช่วงสีดำ สีเทา ไปจนถึงสีขาว ถ้าจุดสีในภาพใดๆ มีการชักตัวอย่างและควอนไทซ์ที่ระดับ 8 บิตแล้ว ระดับความเข้มแสงจะมี 256 ระดับ (2^8) มีค่าอยู่ในช่วง 0 ถึง 255

แสงที่มีสี คือ แสงที่มีคลื่นแม่เหล็กไฟฟ้าอยู่ในช่วง 400 ถึง 700 นาโนเมตร ปริมาณพื้นฐาน 3 อย่างที่ใช้เพื่ออธิบายคุณภาพของแหล่งกำเนิดแสงมีสี คือ การแผ่แสง (radiance), ระดับความเข้มของแสงสว่าง (luminance), และความสว่าง (brightness) การแผ่แสง คือ ปริมาณแสงทั้งหมดที่ไหลจากแหล่งกำเนิดแสง วัตถุปริมาณได้ออกมาเป็นหน่วยของวัตต์ ระดับความเข้มของแสงสว่าง คือปริมาณพลังงานที่ผู้สังเกตได้รับจากแหล่งกำเนิดแสง มีหน่วยคือลูเมนส์

(lumens) เช่นแสงอินฟราเรด (infrared) มีพลังงานที่ถูกปล่อยออกมาจากแหล่งกำเนิดแสงเป็นปริมาณพลังงานค่าหนึ่ง แต่ผู้สังเกตได้รับพลังงานเหล่านั้นได้น้อยหรืออาจจะไม่ได้รับเลย นั้นหมายความว่ามีการแผ่แต่ไม่มีค่าระดับความเข้มของแสงสว่าง ความสว่าง คือตัวที่แสดงถึงการมองเห็นซึ่งไม่สามารถวัดค่าออกมาได้ในเชิงปฏิบัติ และเป็นตัวอธิบายถึงแสงที่ไม่มีสีในเชิงความเข้มสีและเป็นปัจจัยหนึ่งที่สำคัญในการอธิบายสี

ตามโครงสร้างของตามนุษย์จะมองเห็นสีได้จากการรวมของสีพื้นฐานสามสีคือ สีแดง, สีเขียวและสีน้ำเงิน ตามมาตรฐานของ CIE (Commission Internationale de l'Eclairage – the International Commission on Illumination) บ่งไว้ว่าสีน้ำเงิน มีความยาวคลื่น 435.8 นาโนเมตร สีเขียวมีความยาวคลื่น 546.1 นาโนเมตร และสีแดงมีความยาวคลื่น 700 นาโนเมตร แต่ไม่ได้หมายความว่าสีทั้งสามจะผลิตสีอื่นได้ครบ หากจะต้องประกอบด้วย ช่วงสเปกตรัมของสีเหล่านั้น จึงจะสร้างสีอื่นได้ครบ สีพื้นฐานสามสีแรกสามารถผลิตสีพื้นฐานที่สองได้ โดยการรวมตัวกัน สีม่วงอมแดงเกิดจากสีแดงรวมกับสีน้ำเงิน สีฟ้าเกิดจากสีเขียวรวมกับสีน้ำเงิน สีเหลืองเกิดจากสีแดงและสีเขียว การรวมกันของสีพื้นฐานแรกกับสีพื้นฐานที่สองด้วยความหนาแน่นที่เหมาะสม ผลที่ได้จะออกมาเป็นสีขาว

คุณลักษณะโดยทั่วไปที่ใช้แยกสีต่างๆออกจากกันคือความสว่าง เจดสี (hue) และความอิ่มตัวของสี (saturation) ความสว่างแสดงถึงความเข้มแสง เจดสีเป็นสมบัติที่แสดงถึงความยาวคลื่นของที่ผสมกันอยู่ ดังนั้น เจดสีคือ ลักษณะเฉพาะของสีที่แลเห็นโดยผู้สังเกต เช่น เมื่อเรียกสีแดง ส้ม หรือเหลือง นั่นคือกำลังมองเห็นเจดสีของมัน ความอิ่มตัวของสีแสดงถึงปริมาณของแสงขาวที่เจืออยู่ในสีนั้นๆ หรือความบริสุทธิ์ของสีนั่นเอง สเปกตรัมของสีที่บริสุทธิ์จะมีค่าความอิ่มตัวมาก สีชมพูและสีม่วงอ่อนมีความอิ่มตัวน้อยเพราะมีแสงขาวปนอยู่มาก นั่นคือ ค่าความอิ่มตัวของสีจะแปรผกผันกับปริมาณแสงขาวที่ผสมอยู่

การรวมเจดสีและความอิ่มตัวของสีไว้ด้วยกันจะเรียกว่าความมีสี (chromaticity) เพราะฉะนั้น สีอาจจะแบ่งตามคุณลักษณะได้สองแบบคือ ความสว่างและความมีสี สีต่างๆเกิดจากการประกอบกันของสีแดง สีเขียว และสีน้ำเงิน

2.2.1 แบบจำลองสี (color model)

จุดประสงค์ของการมีแบบจำลองสี คือใช้ในการอ้างอิงและเป็นมาตรฐานเดียวกัน โดยในที่นี้จะกล่าวถึง 5 แบบจำลองสี คือ แบบจำลองสี RGB แบบจำลองสีนอร์มอลไลซ์ RGB แบบจำลองสี YIQ แบบจำลองสี YUV แบบจำลองสี YCbCr และแบบจำลองสี HSI

2.2.1.1 แบบจำลองสี RGB (RGB color model)

ระบบสี RGB (แดง น้ำเงิน เขียว) เป็นระบบพื้นฐานของสีที่ใช้ในคอมพิวเตอร์กราฟิก และหน่วยเก็บความจำของภาพ เพราะว่าหลอดภาพของจอคอมพิวเตอร์นั้นใช้สารเรืองแสงสีแดง สีเขียว และสีน้ำเงิน ในการผลิตสีอื่นๆออกมา นั่นคือสีแดง สีเขียว และสีน้ำเงิน เป็นองค์ประกอบของสีทุกๆสี ซึ่งในการประกอบกันเป็นสีต่างๆนั้นเกิดจากการรวมกันของสามองค์ประกอบดังแสดงในรูปที่ 2.5

ทั้งสามองค์ประกอบสีนั้นต้องใช้แบนด์วิดท์ (bandwidth) ที่เท่ากันในการแสดงแทนภาพๆหนึ่ง ฉะนั้นจึงไม่ค่อยมีประสิทธิภาพที่ตีมากนัก เพราะภาพหนึ่งภาพต้องใช้แบนด์วิดท์ขององค์ประกอบของสีทั้งสามสีเท่าๆกัน แต่สายตามนุษย์นั้นจะตอบสนองต่อองค์ประกอบของความเข้มแสงมากกว่าองค์ประกอบของความเข้มสี ด้วยเหตุผลนี้ มาตรฐานการเข้ารหัสภาพหลายๆมาตรฐาน และระบบการแพร่กระจายข้อมูลต่างๆ จึงเลือกใช้สัญญาณองค์ประกอบความเข้มแสง และสัญญาณความแตกต่างกันของสี (color-difference) เช่น YUV YIQ YCbCr และ HSI เป็นต้น

2.2.1.1 การแก้ไขแกมมา (Gamma correction)

องค์ประกอบของสีนั้นจะเกิดขึ้นในรูปแบบของสัญญาณแรงดันทางไฟฟ้า ซึ่งแปรผันตรงตามความสว่างที่ตกลงบนจุดในหลอดภาพของกล้องถ่าย สัญญาณนี้ไม่เหมาะสมที่จะนำมาแสดงโดยตรงกับหลอด CRT หรือ เครื่องรับโทรทัศน์ เพราะหลอดภาพรังสีแคโทด (cathode-ray) เป็นอุปกรณ์ที่ไม่เป็นเชิงเส้น ดังนั้นฟังก์ชันถ่ายโอนของหลอดภาพ CRT จะผลิตความเข้มแสงเป็นสัดส่วนโดยตรงกับค่ากำลังสัญญาณแรงดันไฟฟ้าบางค่าเท่านั้น ซึ่งสามารถเขียนได้ดังสมการ

$$I_{\text{แสดงหน้าจอ}} \propto V^\gamma \quad (2.13)$$

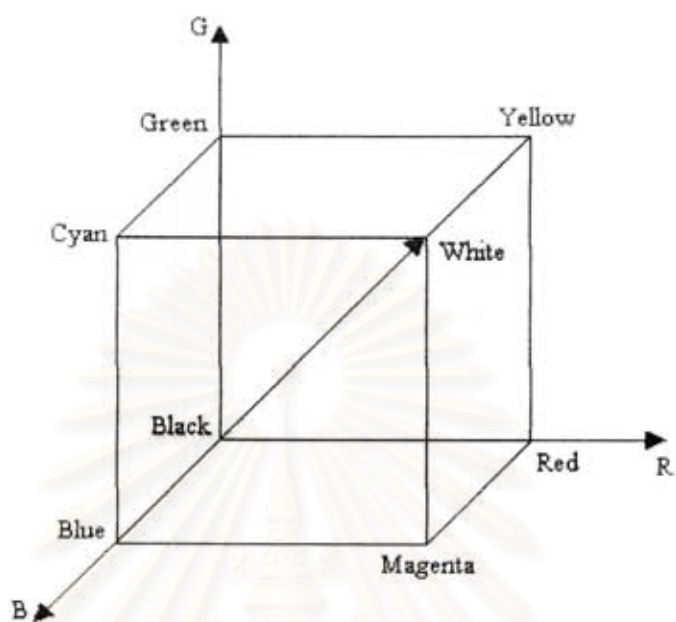
โดย $I_{\text{แสดงหน้าจอ}}$ คือความเข้มแสงที่แสดงออกบนหน้าจอ

V คือ แรงดันทางไฟฟ้าของสัญญาณภาพที่ได้รับ

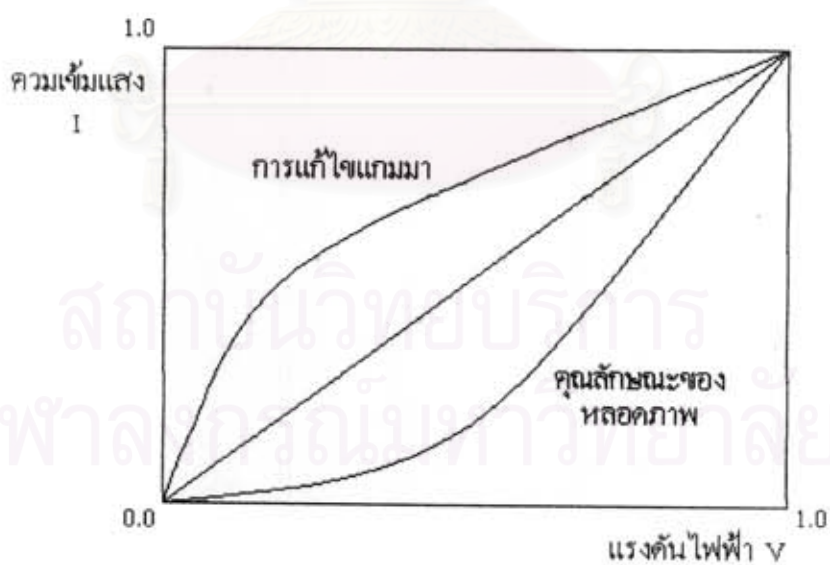
γ คือ ค่าแกมมาที่ใช้ในการแสดงผล มีค่าอยู่ในช่วง 2.0 ถึง 3.0 ซึ่งขึ้นกับชนิดของหลอดภาพ โดยทั่วไปจะมีค่า 2.2 สำหรับระบบ NTSC ดังแสดงในรูปที่ 2.2 และมีค่า 2.8 สำหรับระบบ PAL/SECAM ซึ่งค่าสัญญาณแรงดันอินพุตจะถูกทำการนอร์มอลไลซ์ (normalize) ให้อยู่ในช่วง 0 ถึง 1

การชดเชยสำหรับกระบวนการที่ไม่เป็นเชิงเส้นจะทำให้ด้านส่ง ซึ่งข้อมูลเชิงเส้น RGB นั้น จะเรียกว่า การแก้ไขแกมมา (gamma-corrected) การแก้ไขทำได้โดยนำสัญญาณจริงมายกกำลัง 0.45 เช่น $V^{0.45}$ เพราะฉะนั้น เมื่อส่งไปและแสดงผล จะทำให้ค่า $I_{\text{แสดงหน้าจอ}}$ มีค่าเท่ากับ $V^{0.45\gamma}$ ถ้าให้ค่าแกมมาเป็น 2.2 จะได้ $I_{\text{แสดงหน้าจอ}} = V$ เพราะฉะนั้นภาพที่ได้ก็จะเป็นเชิงเส้นต่อการรับชม

ประโยชน์ของการแก้ไขแกมมาที่ด้านส่งคือทำให้ต้นทุนของเครื่องรับโทรทัศน์ลดลงและส่วนของภาพที่มีมืดจะถูกขยายขึ้น



รูปที่ 2.5 พิกัดแบบจำลองสี RGB



รูปที่ 2.6 ลักษณะค่าแกมมาที่ใช้ในการแสดงผล

ในระบบโทรทัศน์สีเช่น HDTV นั้น สัญญาณ YUV จะหาได้โดยใช้เมตริกซ์การแปลง RGB เป็น YUV ซึ่งโดยทั่วไปแล้ว สัญญาณความแตกต่างกันของสี (U และ V) จะถูกซัดตัวอย่างเป็นครึ่งหนึ่งของ Y ในมิติทางขนาดของภาพ เพราะระบบการรับรู้ของมนุษย์ต่อความเข้มแสงจะมากกว่าความเข้มสี ซึ่งเป็นประโยชน์อย่างยิ่งในการที่จะช่วยลดอัตราบิตของเทคนิคการเข้ารหัสสัญญาณวิดีโอ

2.2.1.2 แบบจำลองสี YIQ (YIQ color model)

แบบจำลองสี YIQ สืบทอดมาจาก YUV และเป็นส่วนเสริมของระบบโทรทัศน์แบบ NTSC ซึ่ง I นั้นหมายถึง เฟสตรงกัน (in-phase) Q หมายถึง เฟสต่างกัน 90 องศา (quadrature-phase) โดยมีสมการการแปลงคือ

$$Y = 0.299R' + 0.587G' + 0.114B' \quad (2.14)$$

$$I = 0.596R' - 0.275G' - 0.321B' \quad (2.15)$$

$$= 0.736(R' - Y) - 0.268(B' - Y) \quad (2.16)$$

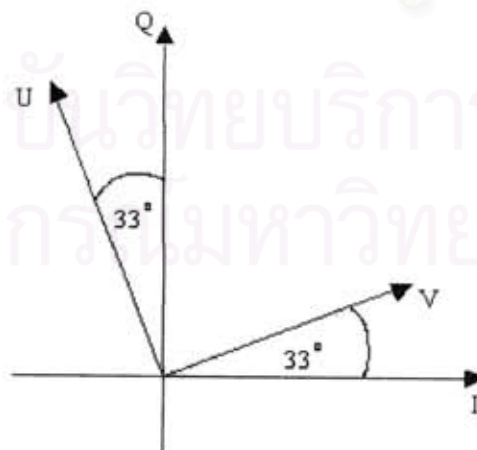
$$Q = 0.212R' - 0.523G' + 0.311B' \quad (2.17)$$

$$= 0.478(R' - Y) + 0.413(B' - Y) \quad (2.18)$$

ซึ่งสามารถเขียนได้อีกรูปแบบคือ

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos 33^\circ & \sin 33^\circ \\ -\sin 33^\circ & \cos 33^\circ \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \quad (2.19)$$

จากรูปที่ 2.7 จะเห็นได้ว่า พิกัด UV จะต่างจากพิกัด IQ อยู่ 33 องศา

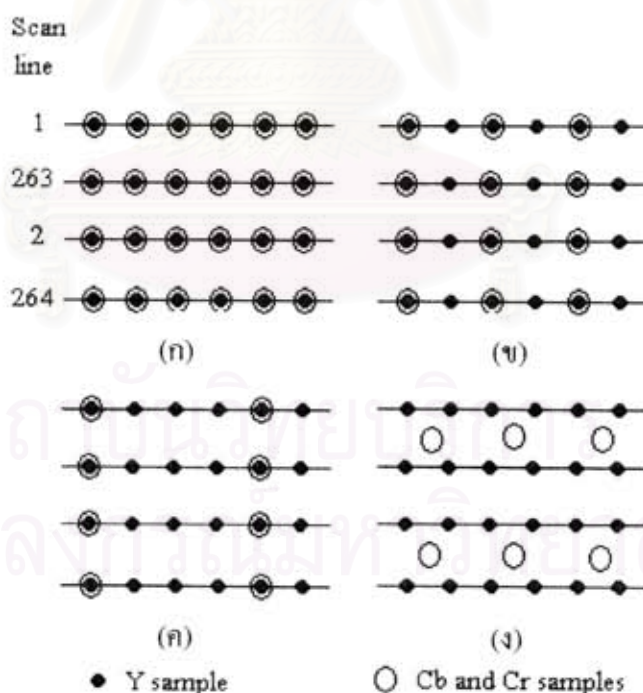


รูปที่ 2.7 ความแตกต่างของพิกัด UV และพิกัด IQ

2.2.1.3 แบบจำลองสี YCbCr (YCbCr color model)

มาตรฐาน ITU-R BT.601 ได้มีการพัฒนาแบบจำลองสี YCbCr ขึ้นมา มาตรฐานกำหนดถึงความเข้ากันได้ระหว่างระบบ 2 ระบบ (ระบบ 525 เส้น และ ระบบ 625 เส้น) สัญญาณ YCbCr คือสัญญาณ YUV ที่ได้รับการปรับขนาดและออฟเซต (offset) แล้ว Y จะอยู่ในช่วง 16 ถึง 235 (220 ระดับการควอนไทซ์) Cb และ Cr จะอยู่ระหว่าง 16 ถึง 240 (225 ระดับการการควอนไทซ์) ความถี่การชักตัวอย่างความเข้มแสงจะมีค่า 13.5 เมกะเฮิร์ตซ์ (hertz) ซึ่งมีค่าประมาณ 4 เท่าของ 3.5 เมกะเฮิร์ตซ์ ซึ่งเป็นความถี่ของคลื่นพหุย่อยสำหรับระบบ NTSC

ตำแหน่งของจุดตัวอย่างหรือจุดภาพของ YCbCr ในรูปแบบ 4:4:4, 4:2:2, 4:1:1 และ 4:2:0 แสดงดังรูปที่ 2.4 แต่ละองค์ประกอบถูกควอนไทซ์ขนาด 8 บิต สำหรับในรูปแบบ 4:4:4 ใน 1 จุดภาพจะใช้บิตเป็นจำนวน 24 บิต $((4+4+4)*8/4)$ รูปแบบ 4:2:2 ใน 1 จุดภาพจะใช้บิตเป็นจำนวน 16 บิต $((4+2+2)*8/4)$ รูปแบบ 4:1:1 ใน 1 จุดภาพจะใช้บิตเป็นจำนวน 12 บิต $((4+1+1)*8/4)$ ในรูปแบบ 4:2:0 ก็ใช้บิตเป็นจำนวน 12 บิต โดยจะเปลี่ยนเพียงแต่ตำแหน่งของ Y ในการชักตัวอย่างร่วมกับ Cb และ Cr เช่น 4 จุดภาพของ Y จะมี Cb และ Cr อย่างละ 1 จุดภาพในแนวทแยงมุม (2 เส้น 2 คอลัมน์) แต่ในรูปแบบ 4:1:1 4 จุดภาพของ Y จะมี Cb และ Cr อย่างละ 1 จุดภาพในแนวนอน (1 เส้น 4 คอลัมน์)



รูปที่ 2.8 ตำแหน่งของจุดตัวอย่างหรือจุดภาพของ YCbCr

(ก) 4:4:4 (ข) 4:2:2 (ค) 4:1:1 (ง) 4:2:0

มาตรฐานการเข้ารหัสบีบอัดสัญญาณวิดีโอ เช่น MPEG หรือ H.261 นั้นมีรูปแบบการชักตัวอย่างแตกต่างกันเพียงเล็กน้อย คือเป็น 4:2:0 เหมือนกัน แต่ต่างกันตำแหน่งขององค์ประกอบความเข้มแสงเท่านั้น

สัญญาณ YCbCr นั้นหาได้จากการแปลงดังนี้

$$Y = 0.299R' + 0.587G' + 0.114B' \quad (2.20)$$

$$Cb = -0.169R' - 0.331G' + 0.500B' \quad (2.21)$$

$$Cr = 0.500R' - 0.419G' - 0.081B' \quad (2.21)$$

สัญญาณความแตกต่างกันของสีเป็นไปตามสมการ

$$(B - Y) = -0.299R' - 0.587G' + 0.886B' \quad (2.23)$$

$$(R - Y) = 0.701R' - 0.587G' - 0.114B' \quad (2.24)$$

โดยที่ค่า (B-Y) จะอยู่ในช่วง -0.866 ถึง +0.866 และ (R-Y) อยู่ในช่วง +0.701 ถึง -0.701 Y อยู่ในช่วง 0 ถึง 1 ค่าความแตกต่างสีจะถูกปรับให้อยู่ในช่วง +0.5 ถึง -0.5 โดยจะคูณ 0.564 (มาจาก 0.5/0.866) กับ (B-Y) และคูณ 0.713 (มาจาก 0.5/0.701) กับ (R-Y) เพราะฉะนั้น จะได้ค่า Y อยู่ในช่วง 0 ถึง 1 Cb และ Cr อยู่ในช่วง +0.5 ถึง -0.5

Y มีระดับของการควอนไทซ์ 220 ระดับ (16 ถึง 235) ส่วน Cb และ Cr มีระดับของการควอนไทซ์ 225 ระดับ (16 ถึง 240) ค่า YCbCr สามารถเปลี่ยนจากทศนิยมเป็นจำนวนฐานสิบได้ดังต่อไปนี้

$$\bar{Y} = 219Y + 16 \quad (2.25)$$

$$\bar{Cb} = 224[0.564(B-Y)] + 128 = 126(B-Y) + 128 \quad (2.26)$$

$$\bar{Cr} = 224[0.713(R-Y)] + 128 = 160(B-Y) + 128 \quad (2.27)$$

ถ้าค่าที่ออกมาเป็นทศนิยมแล้วจะปัดเป็นค่าจำนวนเต็มที่ใกล้ที่สุด สังเกตว่าถ้าค่า Cb เป็น 0 จะแปลงเป็นจำนวนเต็มคือ 128 หรือสามารถเขียนการแปลงได้อีกแบบคือ

$$\bar{Y} = 219Y + 16 \quad (2.28)$$

$$\bar{Cb} = 224Cb + 128 \quad (2.29)$$

$$\bar{Cr} = 224Cr + 128 \quad (2.30)$$

2.2.1.4 แบบจำลองสี HSI (HSI color model)

HSI (hue saturation intensity) ประกอบด้วย 3 องค์ประกอบเช่นเดียวกับแบบจำลองสีแบบอื่น โดย Hue คือ คุณสมบัติของสีที่อธิบายถึงความบริสุทธิ์ของสี (สีเหลือง สีส้ม หรือสีแดง)

ขณะที่ความอิ่มตัวของสี (saturation) เป็นการวัดองค์ประกอบความบริสุทธิ์ของสีที่ถูกเจือโดยแสงสีขาว เช่นถ้ามีแสงสีขาวอยู่มากจะมีค่าความอิ่มตัวน้อย แต่ถ้ามีแสงสีขาวเจืออยู่น้อยแสดงว่ามีความอิ่มตัวมาก แบบจำลองสี HSI นั้นเป็นที่ใช้กันอย่างแพร่หลายเนื่องจากความสอดคล้องกับกลไกในการรับรู้ของดวงตามนุษย์ที่มีเส้นประสาท 2 ชุดคือ rod และ cone ที่มีการตอบสนองที่ดีต่อความเข้มแสงและสีตามลำดับ กล่าวคือ ระบบ HSI มี 1 ซึ่งเป็นความเข้มแสงเป็นองค์ประกอบหนึ่งในแบบจำลอง องค์ประกอบ Hue และความอิ่มตัวนั้นมีความสัมพันธ์กับการรับรู้สีของมนุษย์ ทำให้แบบจำลอง HSI เป็นเครื่องมือที่สำคัญในการพัฒนากรรมวิธีทางการประมวลผลภาพ ซึ่งการพัฒนานั้นขึ้นกับสมบัติการรับรู้สีของมนุษย์เป็นสำคัญ

2.2.1.5 แบบจำลองสีนอร์มัลไลซ์ RGB (normalized RGB)

ปัญหาของปริภูมิสี RGB คือ การรวมกันของความสว่างและสี จึงเกิดเป็นปริภูมินอร์มัลไลซ์ RGB ซึ่งมีวัตถุประสงค์เพื่อตัดความสนใจขององค์ประกอบความสว่างทิ้งไป โดยหากค่าจุดภาพในปริภูมิสี RGB นั้นมีสีเดียวกันแต่ความสว่างต่างกัน จะถูกกำหนดเป็นค่าเดียวกันในปริภูมินอร์มัลไลซ์ RGB และแทนเป็นสัญลักษณ์ "rgb" แต่ละค่าสามารถหาได้จากสมการที่ (2.31)- (2.33)

$$r = \frac{R}{(R + G + B)} \quad (2.31)$$

$$g = \frac{G}{(R + G + B)} \quad (2.32)$$

$$b = \frac{B}{(R + G + B)} \quad (2.33)$$

เช่น ค่าในปริภูมิสี RGB (255,0,0) ทำการแปลงไปสู่ rgb จะได้(1,0,0) และค่าในปริภูมิสี RGB (100,0,0) ทำการแปลงไปสู่ rgb จะได้(1,0,0) ซึ่งทั้งสองค่านี้ในปริภูมิสี RGB จะมีเฉพาะองค์ประกอบของสีแดง ที่มีค่าความเข้มของแสงแตกต่างกัน ส่วนในปริภูมิสี rgb องค์ประกอบทางแสงจะถูกละทิ้งไปทำให้ 2 ค่านี้ถือเป็นค่าเดียวกัน

2.3 การประมวลผลภาพเชิงสัณฐาน (Morphological Image Processing)

การประมวลผลภาพเชิงสัณฐานนั้นเป็นการใช้ตัวดำเนินการทางคณิตศาสตร์ ซึ่งเป็นเครื่องมือที่ใช้ในการสกัดองค์ประกอบของภาพ ซึ่งองค์ประกอบนั้นอาจจะเป็นวัตถุหรือตัวแสดง

หรือตัวอธิบายรูปร่างของวัตถุ เช่น ขอบเขต, โค้งกระดูก และเลือกผิวโค้ง ตัวดำเนินการเชิงเส้นฐานนั้นอาจจะใช้ในกระบวนการก่อน (pre-processing), กระบวนการหลัง (post-processing) ซึ่งอาจจะมีหลายเทคนิคได้เช่น การกรองเชิงเส้นฐาน, การทำให้บาง และการตกแต่งให้สวยงาม

ตัวดำเนินการเชิงเส้นฐานเป็นตัวดำเนินการทางคณิตศาสตร์ที่อธิบายโดยทฤษฎีเซต โดยเซตคือกลุ่มข้อมูลที่เก็บลักษณะหรือรูปร่างของวัตถุในภาพ ในภาพไบนารี เซตคือกลุ่มข้อมูลเวกเตอร์ 2 มิติ (x, y) ในภาพสีเทา เซตคือกลุ่มข้อมูลเวกเตอร์ 3 มิติ (x, y, I) โดย x คือ ตำแหน่งแนวของจุดภาพ y คือ ตำแหน่งคอลัมน์ของจุดภาพ และ I คือ ความเข้มแสงที่ตำแหน่ง x, y

2.3.1 การทำโตเลชัน (Dilation) และ การทำอีโรชัน (Erosion)

รูปภาพหรือกลุ่มของจุดภาพสีดำและขาว ซึ่งมีลักษณะเป็นรูปภาพไบนารี (binary image) การทำโตเลชันและอีโรชัน ถือเป็น ตัวดำเนินการการเชิงเส้นฐานเบื้องต้น และเป็นพื้นฐานในการนำไปใช้กับกระบวนการที่มีความซับซ้อนมากขึ้น ได้แก่ ตัวดำเนินการแบบเปิด (opening operator), ตัวดำเนินการแบบปิด (closing operator) และการแยกรูปร่าง (shape decomposition)

การทำโตเลชัน

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

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

โดยที่ X คือ รูปภาพที่ทำการโตเลชัน

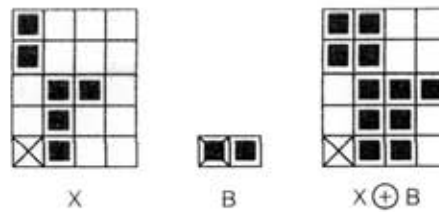
B คือ โครงสร้างส่วนย่อยในการทำโตเลชัน

การทำโตเลชันแสดงดัง รูปที่ 2.9

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 2.9 การทำไดเลชัน

การทำอีโรชัน

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

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

โดยที่ X คือ รูปภาพที่ทำการอีโรชัน

B คือ โครงสร้างส่วนย่อยในการทำอีโรชัน

การทำอีโรชันแสดงดัง รูปที่ 2.10

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



รูปที่ 2.10 การทำอีโรชัน

2.3.2 ตัวดำเนินการแบบเปิดและปิด

ไดเลชันเป็นการขยายภาพ ส่วนอีโรชันเป็นการหดภาพ ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดเป็นตัวดำเนินการเชิงสัจฐานชนิดหนึ่ง ตัวดำเนินการแบบเปิดจะทำให้เส้นขอบของวัตถุราบเรียบหรือทำให้เหลี่ยมมนกลมลง ทั้งยังทำให้ส่วนแหลมและส่วนยื่นของวัตถุหดหรือเล็กลง ตัวดำเนินการแบบปิดก็เป็นการทำให้เส้นขอบของวัตถุราบเรียบแต่ทำให้เหลี่ยมมนกลมลง

ทั้งยังทำให้ส่วนแหลมและส่วนยื่นของวัตถุใหญ่ขึ้นด้วย การทำอิโรชันและไดเลชันไม่ถือเป็น การแปลงผกผัน (inverse transformation) ระหว่างกัน ดังนั้น รูปภาพที่นำมาผ่านการอิโรชันและ ไดเลชัน ไม่จำเป็นต้องเท่ากับรูปภาพดั้งเดิม แต่จะได้ผลลัพธ์ซึ่งมีรายละเอียดซับซ้อนน้อยกว่า รูปภาพดั้งเดิมแทน

การกระทำแบบเปิดต่อเซต A โดยเซตโครงสร้างย่อย B คือ

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

ถ้าพิจารณาจากรูปสมการจะพบว่าเป็นการกระทำอิโรชันต่อ A โดย B แล้วนำผลที่ได้มากระทำ ไดเลชันโดย B อีกครั้ง

การกระทำแบบปิดต่อเซต A โดยเซตโครงสร้างย่อย B คือ

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

ถ้าพิจารณาจากรูปสมการจะพบว่าเป็นการกระทำไดเลชันต่อ A โดย B แล้วนำผลที่ได้มากระทำ อิโรชันโดย B อีกครั้ง รูปที่ 2.11 แสดงผลตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิด ตัว ดำเนินการแบบเปิดสามารถเขียนได้ในอีกรูปแบบหนึ่งคือ

$$A \circ B = \cup \{ (B)_x \mid (B)_x \subset A \} \quad (2.38)$$

ตัวดำเนินการแบบปิดสามารถเขียนได้ในอีกรูปแบบหนึ่งคือ

$$A \bullet B = \{ x \mid (B)_x \cap A \neq \emptyset \} \quad (2.39)$$

รูปที่ 2.12 และรูปที่ 2.13 แสดงถึงตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดตามสมการ (2.38) และ (2.39) คุณสมบัติความเป็นคู่ของตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิด พิจารณาได้จากสมการดังต่อไปนี้

$$(A \bullet B)^c = [(A \oplus B) \ominus B]^c \quad (2.40)$$

$$= (A \oplus B)^c \oplus \bar{B} \quad (2.41)$$

$$= (A^c \ominus \bar{B}) \oplus \bar{B} \quad (2.42)$$

$$= A^c \circ \bar{B} \quad (2.43)$$

ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดมีสมบัติดังนี้

ก) $A \circ B \subset A$

ค) $(A \circ B) \circ B = A \circ B$

จ) ถ้า $C \subset D$ แล้ว $(C \bullet B) \subset (D \bullet B)$

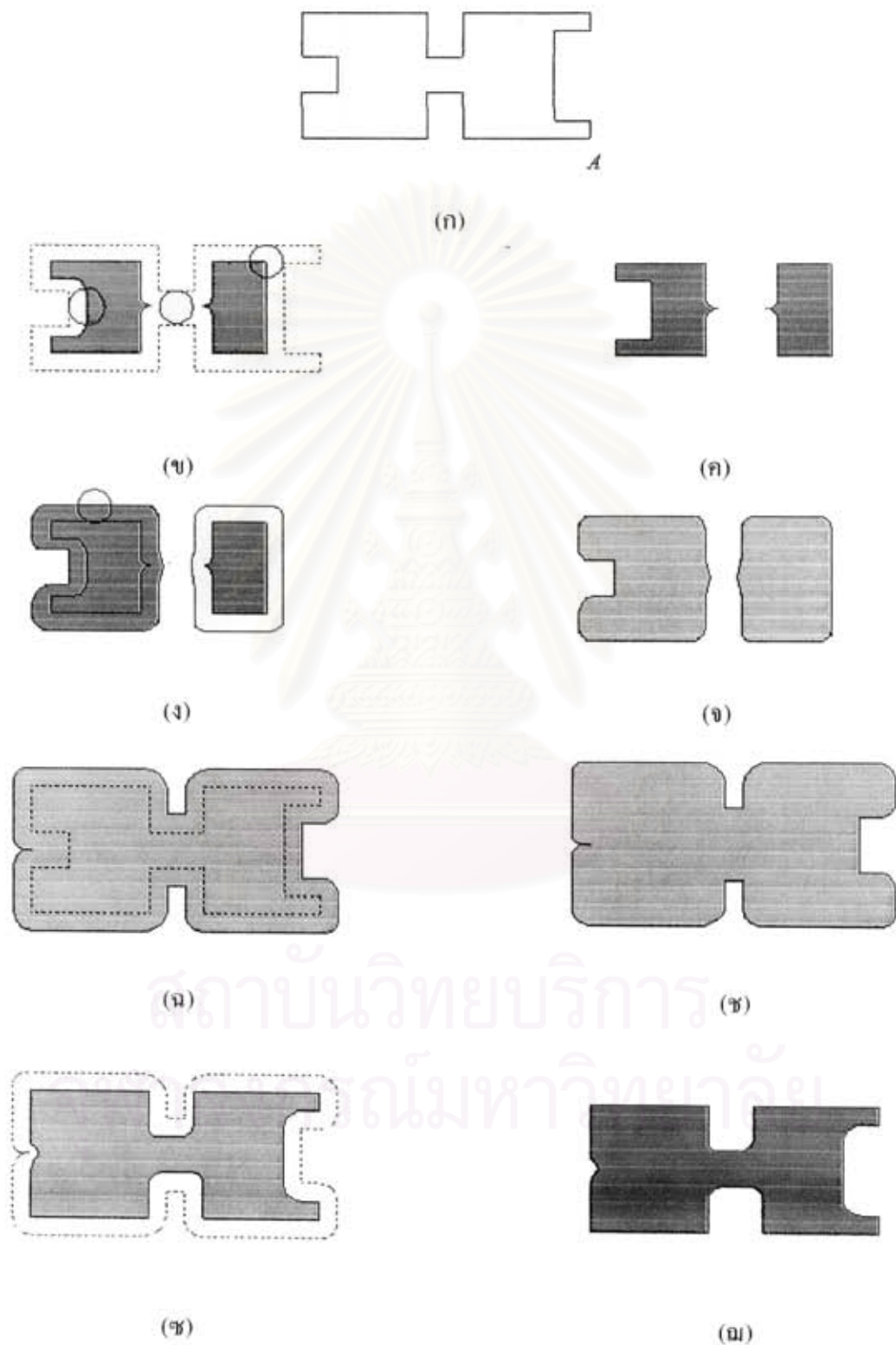
ข) ถ้า $C \subset D$ แล้ว $(C \circ B) \subset (D \circ B)$

ง) $A \subset A \bullet B$

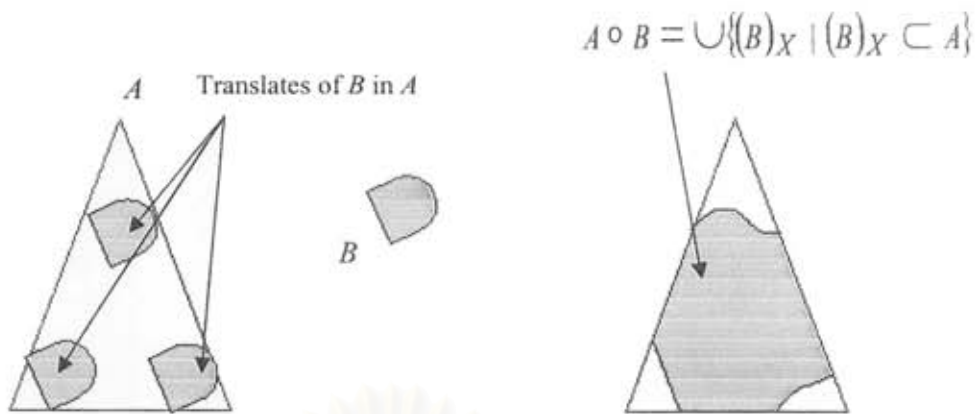
ฉ) $(A \bullet B) \bullet B = A \bullet B$

สมบัติทั้งสามข้อของตัวดำเนินการแบบปิดอธิบายได้โดย ก) ผลลัพธ์จากการทำแบบเปิด ยังเป็นสับเซตของตัวอินพุต ข) ยังคงมีสมบัติการเพิ่มขึ้น ค) ถ้ากระทำตัวดำเนินการแบบเปิดเป็น ครั้งที่สองโดยใช้โครงสร้างย่อยตัวเดิมจะให้ผลเท่าเดิม รูปที่ 2.14 แสดงวัตถุรูปสี่เหลี่ยมผืนผ้าที่ถูก

ทำลายโดยสิ่งรบกวน การขจัดสิ่งรบกวนออกจากวัตถุและพื้นหลังทำได้โดยตัวดำเนินการเชิง
 ล้วนฐาน



รูปที่ 2.11 (ข)-(จ) แสดงผลจากการกระทำตัวดำเนินการแบบเปิด (ค)-(ง) ตัวดำเนินการแบบปิด

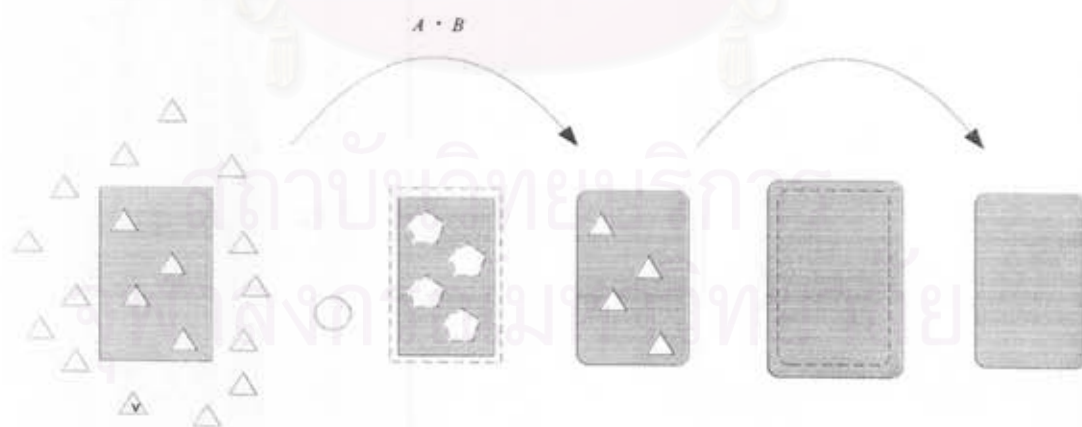


รูปที่ 2.12 การกระทำตัวดำเนินการแบบเปิดต่อภาพไมนารี

Translate of B containing z . Each intersects A and, therefore, $z \in A \cdot B$



รูปที่ 2.13 การกระทำตัวดำเนินการแบบปิดต่อภาพไมนารี

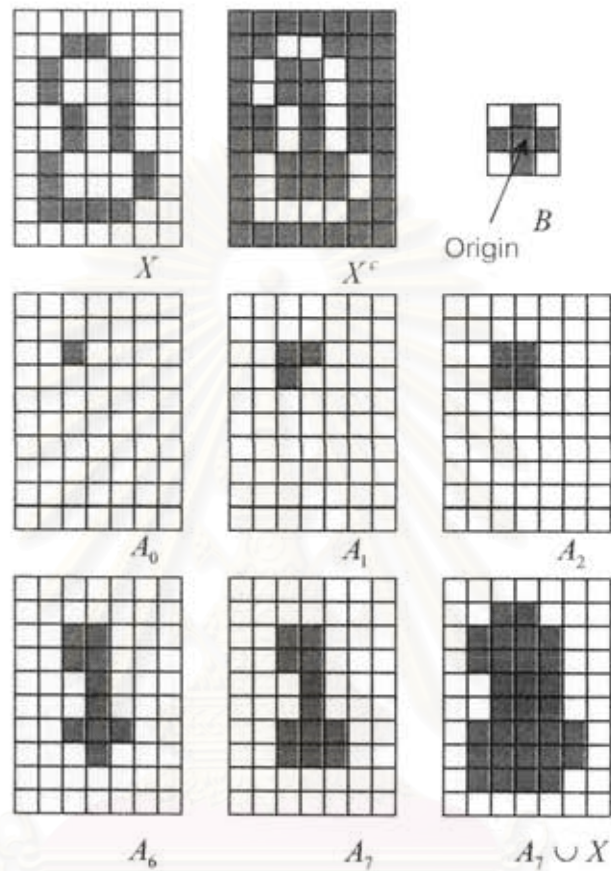


รูปที่ 2.14 แสดงการจัดสิ่งรบกวนออกจากวัตถุและพื้นหลังทำได้โดยตัวดำเนินการเชิงสัญญาณ

2.3.3 การเติมเต็มบริเวณ (Region filling)

วิธีการในการเติมเต็มจุดภาพลงในบริเวณกระทำโดยอาศัยตัวดำเนินการไต่เลขัน, คอมพลีเมนต์ และ อินเตอร์เซกชัน ตามสมการที่ (2.44)

$$A_k = (A_{k-1} \oplus B) \cap X^c \quad k = 1, 2, 3, \dots \quad (2.44)$$



รูปที่ 2.15 การเติมเต็มบริเวณ

2.4 การตรวจหาขอบ (Edge Detection)

การตรวจหาขอบ ถือเป็นกระบวนการประมวลผลภาพที่สำคัญมากอันหนึ่ง ถูกนำมาใช้ในการวิเคราะห์รูปภาพเพื่อกำหนดขอบเขตของบริเวณที่ต้องการ ซึ่งกระบวนการนี้จะกระทำโดยตรวจสอบว่า จุดภาพจุดใดที่มีการเปลี่ยนแปลงขององค์ประกอบทางความสว่างอย่างทันทีทันใด เมื่อเปรียบเทียบกับบริเวณใกล้เคียงจะกำหนดจุดภาพนั้นเป็นขอบ

ขอบ คือ คุณสมบัติที่เกี่ยวกับจุดภาพใดจุดภาพหนึ่งซึ่งนำไปคำนวณร่วมกับพฤติกรรมของจุดภาพรอบข้าง และประกอบไปด้วยตัวแปรชนิดเวกเตอร์ 2 ตัวแปร คือ ขนาดและทิศทาง

ขนาดของขอบสามารถหาได้จากขนาดของเกรเดียนต์และทิศทางของขอบถูกหมุนในส่วนทิศเกรเดียนต์ซึ่งมีค่าสูงสุด จากจุดภาพสีดำ [$f(i,j) = 0$] ไปจุดภาพสีขาว [$f(i,j) = 255$]



รูปที่ 2.16 แบบจำลองของขอบ

2.4.1 เกรเดียนต์เชิงสัณฐาน (Morphological Gradient)

ในกระบวนการตรวจจับขอบมีตัวดำเนินการที่สำคัญคือเกรเดียนต์เชิงสัณฐาน มีนิยามดังนี้

$$G(f) = \delta_n(f) - \epsilon_n(f) \quad (2.45)$$

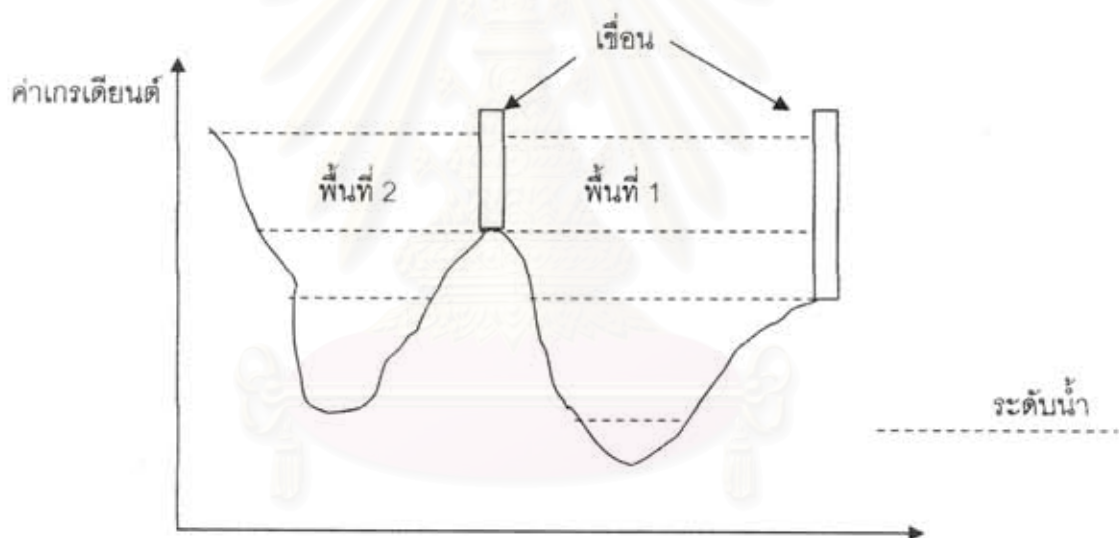
เกรเดียนต์เชิงสัณฐาน คือการตรวจจับขอบของภาพวิธีหนึ่ง ภาพที่ได้จะแสดงถึงขอบของวัตถุรวมถึงการเปลี่ยนแปลงทางสีภายในภาพอย่างชัดเจน โดยทั่วไปแล้วจะใช้ค่าเกรเดียนต์ขององค์ประกอบความเข้มแสงของภาพ บางครั้งควรนำองค์ประกอบของความเข้มสีมาช่วยในการคิดด้วย เพราะว่าในบางกรณี องค์ประกอบของความเข้มสีนั้นให้ผลที่ดีกว่า แต่การนำมาใช้ จะพิจารณาจากตัวถ่วงน้ำหนักคุณกับค่าเกรเดียนต์ในแต่ละองค์ประกอบแล้วตัดสินใจว่าค่าใดให้ค่าเกรเดียนต์มากที่สุดก็ใช้ค่านั้น

$$\max \{ \omega_1 G_r, \omega_2 G_{cb}, \omega_3 G_{c\alpha} \} \quad (2.46)$$

โดย ω_1 , ω_2 , ω_3 คือตัวถ่วงน้ำหนักขององค์ประกอบความเข้มสีและความเข้มแสงตามลำดับ ตัวดำเนินการเกรเดียนต์นั้นแม้ทำให้เกิดขอบในภาพ แต่ยังไม่สามารถแยกภาพออกเป็นบริเวณหลายๆ ส่วนได้ หรือไม่สามารแบ่งได้ว่าที่ใดในภาพคือบริเวณที่มีค่าเกรเดียนต์ต่ำสุด การแยกภาพออกเป็นหลายๆบริเวณจะใช้กรรมวิธีสันปันน้ำ (watershed algorithm)

2.4.2 กรรมวิธีสันปันน้ำ

กรรมวิธีสันปันน้ำ เป็นการดำเนินการเชิงสัญฐานชนิดหนึ่งซึ่งสามารถระบุค่าเกรเดียนต์ต่ำที่สุดออกมาเป็นพื้นที่ที่สามารถเห็นได้ รูปที่ 2.17 เป็นรูปที่แสดงความเข้าใจพื้นฐานของกรรมวิธีสันปันน้ำ โดยแสดงอยู่ในรูป 1 มิติ แกนตั้งแสดง ค่าเกรเดียนต์ของภาพ จะเห็นว่า มีลักษณะเป็นความชัน และมีจุดต่ำสุด วิธีการค้นหาว่า จุดต่ำสุดอยู่ที่ใดคือ เริ่มแรกให้น้ำเข้าไปในแอ่งที่ระดับหนึ่งแล้วค่อยๆเพิ่มขึ้นไปเรื่อยๆ ถ้าระดับน้ำของสองแอ่งใดๆเชื่อมติดกันแล้ว จะมีการสร้างเขื่อนขึ้นตรงนั้น ความสูงของเขื่อนจะต้องมีค่ามากกว่าหรืออย่างน้อยเท่ากับค่าเกรเดียนต์ที่มากที่สุด นั่นหมายความว่า ตัวเขื่อนจะเป็นตัวแบ่งว่าที่ตรงนั้นมีจุดต่ำสุดของค่าเกรเดียนต์ภายในบริเวณนั้นๆ นั่นคือ ถ้านำภาพเกรเดียนต์มาทำกระบวนการตามกรรมวิธีสันปันน้ำแล้วจะได้ภาพที่แบ่งเขตออกเป็นจุดที่เกรเดียนต์ต่ำสุดได้ ซึ่งนั่นก็หมายถึง สามารถแยกขอบของวัตถุภายในภาพออกมาได้อย่างชัดเจนด้วย



รูปที่ 2.17 กรรมวิธีสันปันน้ำ 1 มิติ

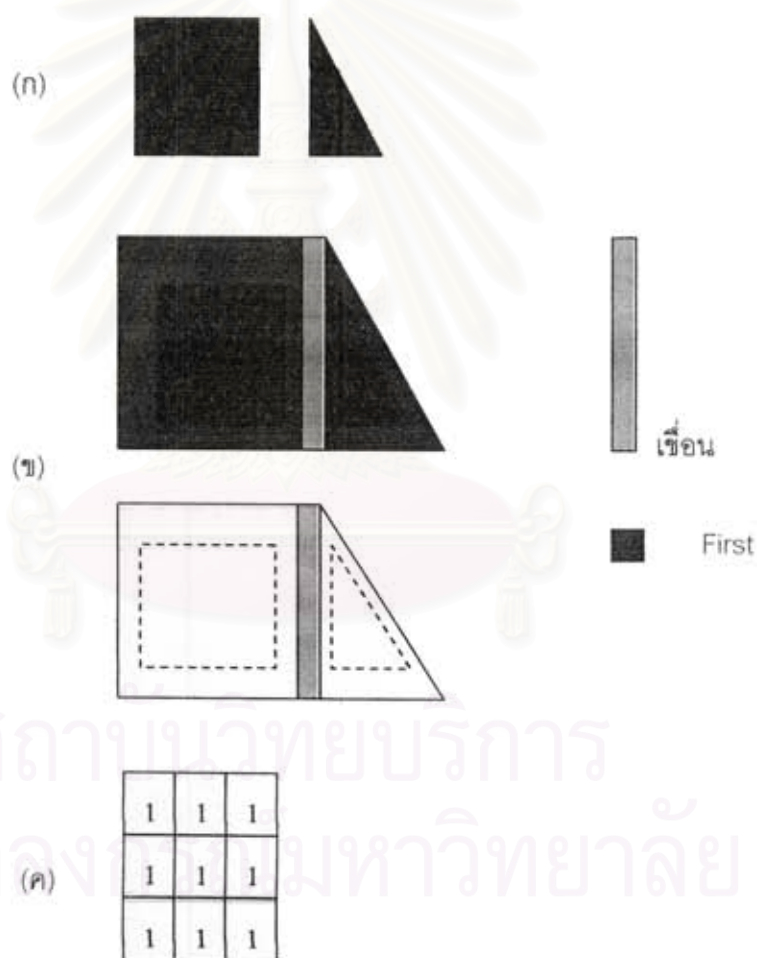
2.4.2.1 ขั้นตอนการสร้างเขื่อน

พิจารณารูปที่ 2.18(ก) เป็นภาพของแอ่งน้ำสองแอ่งที่ได้จากการทำเกรเดียนต์ ให้ M_1 และ M_2 เป็นเซตของจุดตำแหน่งของ 2 พื้นที่นั้น และให้ $C_{n-1}[M_1]$ และ $C_{n-1}[M_2]$ เป็นเซตที่บอกกว่า 2 พื้นที่นั้นอยู่ในระดับ $n-1$ รูปที่ 2.18(ข) เป็นระดับ n กำหนดสัญลักษณ์คือ q จะเห็นว่า ทั้ง 2 แอ่งเกิดการรวมกัน ถ้ากำหนดการยุบเนียนกันของ M_1 และ M_2 คือ $C[n-1]$ แล้ว การที่จะแยกพื้นที่ 2 พื้นที่ใน ระดับ n นั้น สามารถทำได้โดยใช้การกระทำและ (AND) หรือเขียนได้ว่า $q \cap C[n-1]$

เมื่อเกิดการรวมกันของ 2 แอ่งเช่นนี้จะต้องมีการสร้างเขื่อน วิธีการสร้างคือ ใช้มาสก์ขนาด 3×3 ดังรูปที่ 2.18(ค) ทำไต่เลขชั้นกับ 2 แอ่งนั้นผลลัพธ์ที่ได้จะทำให้ 2 แอ่งขยายตัวเองออกโดยการขยายต้องสอดคล้องกับเงื่อนไขคือ

- 1) ทำไต่เลขชั้นแล้วผลลัพธ์ที่ออกมาต้องอยู่ในพื้นที่ q
- 2) ผลลัพธ์ที่ออกมาจากการทำไต่เลขชั้นจะต้องไม่ซ้อนทับกัน

จากเงื่อนไข 2 ข้อนี้ จะเกิดจุดที่ไม่มีการทำไต่เลขชั้นและอยู่ภายในพื้นที่ q ซึ่งจุดนั้นก็คือจุดที่กำหนดเป็นเขื่อนนั่นเอง ในรูป 2.18(ข) จะเห็นว่าต้องทำการไต่เลขชั้นถึง 2 ครั้งจึงจะได้จุดที่เป็นเขื่อน ความสูงของเขื่อนก็ต้องกำหนดให้มีความมากกว่าค่าเกรเดียนต์สูงสุด เมื่อทำเช่นนี้จนครบทุกระดับความสูงของเกรเดียนต์แล้วจะได้เขื่อนที่แยกภาพออกเป็นส่วนๆออกมาได้



รูปที่ 2.18 การสร้างเขื่อน (ก) แอ่ง 2 แอ่ง ในระดับ $n-1$ (ข) การรวมของ 2 แอ่งในระดับ n
(ค) มาสก์ที่ใช้ทำไต่เลขชั้น

ข้อดีของกรรมวิธีนี้คือส่วนของภาพที่แยกออกมาได้จะมีมากเกินไป ดังนั้นจะต้องมีกระบวนการบางอย่างที่ใช้สำหรับรวมพื้นที่บางส่วนเข้าด้วยกัน พื้นที่ที่แยกออกมานั้นมีหลายส่วนที่ยังชี้ชัดไม่ได้ว่าเป็นส่วนพื้นหลัง (background) หรือ ส่วนพื้นหน้า (foreground)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

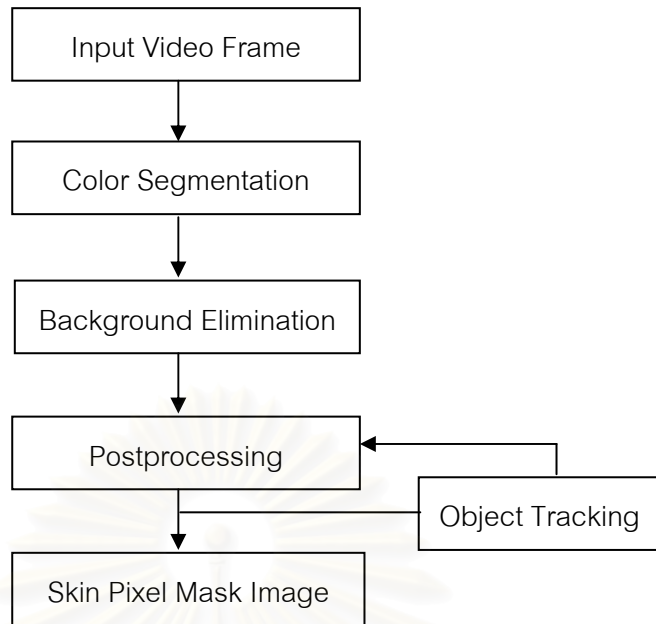
บทที่ 3

กรรมวิธีที่นำเสนอ

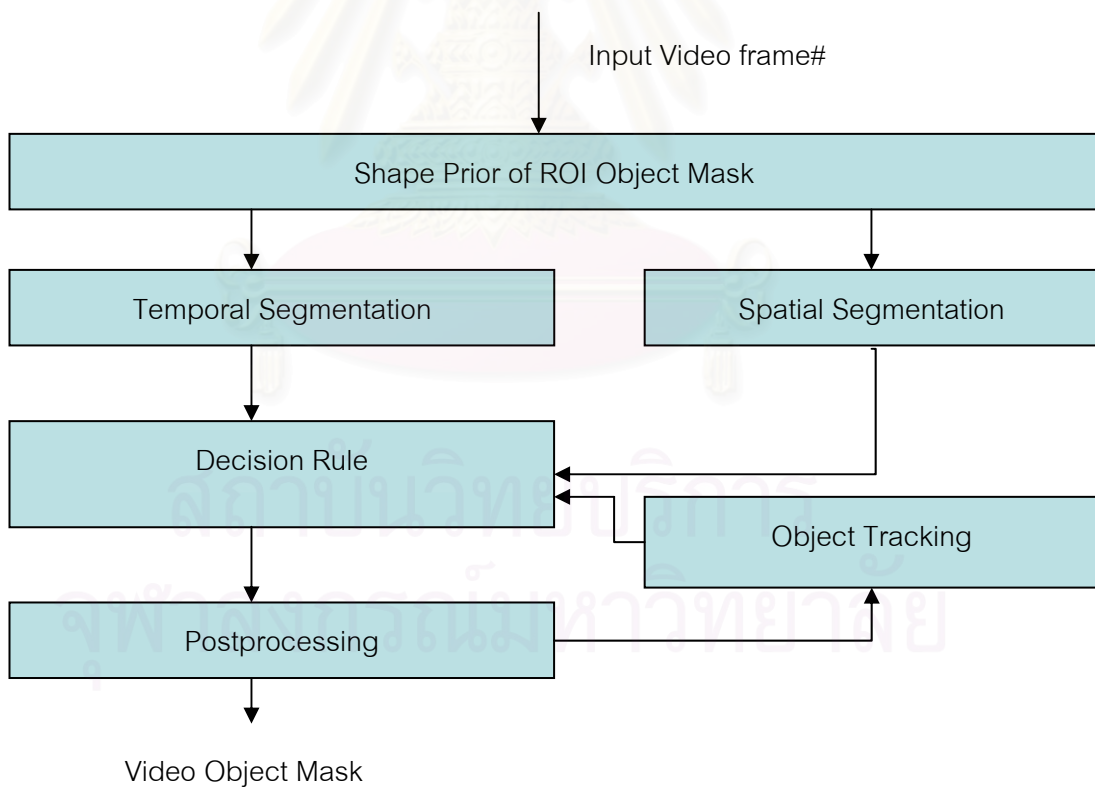
แนวทางในการวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นบริเวณที่สนใจและเป็นวัตถุที่มีความหมายมีทั้งแบบอัตโนมัติและกึ่งอัตโนมัติ โดยใช้เทคนิคต่างๆ ดังได้นำเสนอในบทที่ 1 นั้น มีหลากหลายวิธี ในวิทยานิพนธ์นี้ เสนอการแยกส่วนภาพวิดีโอบริเวณที่สนใจที่เป็นบริเวณใบหน้าและบริเวณวัตถุที่มีความหมายอย่างอัตโนมัติ โดยการแยกส่วนบริเวณที่สนใจบริเวณใบหน้าประกอบไปด้วย การใช้องค์ประกอบสี เพื่อทำการแยกส่วนที่เป็นบริเวณใบหน้าออกมา และกรรมวิธีเชิงสัญญาณเพื่อกำจัดบริเวณที่ไม่ใช่บริเวณใบหน้าออกไป ส่วนการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุวิดีโอในเฟรมภาพอยู่บนพื้นฐานของการรวมกันของคุณลักษณะสำคัญหลายอย่างของสัญญาณวิดีโอ โดยประกอบด้วยข้อมูลเชิงพื้นที่ ข้อมูลเชิงเวลาและข้อมูลบริเวณใบหน้าของวัตถุเพื่อเป็นข้อมูลรูปร่างก่อนหน้า (shape prior)

การแยกส่วนเชิงพื้นที่ประกอบด้วยขั้นตอนการตรวจจับขอบเขตของบริเวณ และการกำหนดบริเวณให้แก่จุดภาพโดยกรรมวิธีสันปันน้ำ ส่วนข้อมูลเชิงเวลาเป็นการใช้ข้อมูลความแตกต่างระหว่างเฟรมและข้อมูลส่วนบริเวณใบหน้าของวัตถุนำไปใช้เพื่อกำหนดขอบเขตบริเวณให้กับวัตถุเพื่อแก้ไขข้อจำกัดของการแยกส่วนภาพเชิงพื้นที่เวลา เมื่อวัตถุไม่มีการเคลื่อนที่หรือมีการเคลื่อนที่น้อยมากจนกระทั่งไม่สามารถตรวจจับการเปลี่ยนแปลงได้ โดยการใช้ข้อมูลบริเวณใบหน้าจะทำให้สามารถระบุขอบเขตของวัตถุได้แม้ว่าจะไม่มีการเคลื่อนที่ของวัตถุก็ตาม ซึ่งข้อมูลนี้จะได้มาจากขั้นตอนการแยกส่วนบริเวณที่สนใจที่เป็น บริเวณใบหน้า จากนั้นนำผลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อให้สามารถแยกส่วนภาพบริเวณที่สนใจที่เป็นวัตถุได้

รูปที่ 3.1 เป็นโครงสร้างแนวคิดของวิทยานิพนธ์ในการแยกส่วนบริเวณที่สนใจบริเวณใบหน้า รูปที่ 3.2 เป็นโครงสร้างในการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุที่มีความหมาย โดยทางด้านอินพุตเป็นภาพ 2 เฟรมที่ติดต่อกัน โดยผลลัพธ์ที่ได้คือมาสก์ของวัตถุ ในขอบเขตของวิทยานิพนธ์นี้ วัตถุคือตัวคนที่เคลื่อนไหวภายในภาพโดยลักษณะของวิดีโอจะเป็นวิดีโอคอนเฟอเรนซ์ ขั้นตอนของแต่ละส่วนแสดงในรูปที่ 3.1 และรูปที่ 3.2 ดังนี้



รูปที่ 3.1 โครงสร้างของแนวคิดวิทยานิพนธ์การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า



รูปที่ 3.2 โครงสร้างของแนวคิดวิทยานิพนธ์การแยกส่วนวัตถุที่มีความหมาย

3.1 การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า

การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า ประกอบไปด้วย ขั้นตอนการแยกส่วนด้วยสี การกำจัดพื้นหลัง กระบวนการข้างหลัง และการติดตาม

3.1.1 การแยกส่วนด้วยสี (Color Segmentation)

ขั้นตอนแรกของกรรมวิธีการแยกส่วนด้วยสีนั้นเกี่ยวข้องกับการใช้ข้อมูลสีในกระบวนการแยกส่วนบริเวณในระดับล่างเพื่อกำหนดบริเวณที่สนใจ จุดประสงค์ในการทำกระบวนการนี้คือการจำแนกจุดภาพของลำดับภาพที่เข้ามาว่าจุดภาพใดเป็นสีผิวและส่วนใดไม่ใช่บริเวณสีผิว

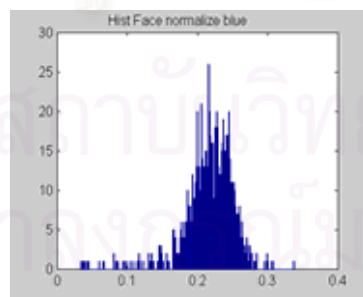
จากงานวิจัย Jones [41] พบว่าแบบจำลองสีฮิสโตแกรมชนิดไม่มีตัวแปรนั้นมีความเที่ยงตรงและใช้การประมวลผลต่ำกว่าวิธีการแบบที่มีตัวแปร ดังนั้นแบบจำลองสีผิวที่ใช้ในวิทยานิพนธ์นี้จึงเป็นแบบจำลองสีผิวฮิสโตแกรมชนิดไม่มีตัวแปร Sobottka และ Pitas [22] ได้ใช้องค์ประกอบสี HS ในปริภูมิสี HSI ในการแยกส่วนสีผิวแต่ Choi [43] แสดงให้เห็นว่าองค์ประกอบเฉดสี (H) จะทำงานได้ไม่ค่อยมีประสิทธิภาพเมื่อองค์ประกอบค่าความอิ่มตัวของสี(S) มีค่าต่ำเนื่องจากว่าค่าความอิ่มตัวของสีเกี่ยวข้องกับแสงสีขาวที่ผสมกับเฉดสี ดังนั้นเมื่อค่าความอิ่มตัวของสีมีค่าน้อยทำให้แมปเฉดสีบริสุทธิ์ที่ได้ไม่ตรงกับบริเวณสีผิว นอกจากนี้แบบจำลองสีผิวในปริภูมิสี HSI มีข้อเสียในด้านสีผิวที่มีลักษณะโทนไปทางสีแดง ในขณะที่ Chai [8] ได้เสนอการใช้องค์ประกอบสี CbCr ในปริภูมิสี YCbCr อีกทั้ง Jones [41] พบว่าองค์ประกอบ Cr นั้นมีความเหมาะสมต่อการตรวจพบฮีโมโกลบินหรือบริเวณผิวหนังของมนุษย์มากกว่าองค์ประกอบอื่นๆ แต่อย่างไรก็ตามการใช้องค์ประกอบสี Cr ร่วมกับองค์ประกอบ Cb ตามที่ Chai เสนอนั้น Chen [10] พบว่าถ้าหากว่าการแยกส่วนที่ได้จากองค์ประกอบ Cb ไม่ได้ผลที่ดีแล้วจะทำให้การแยกส่วนสีผิวได้ผลไม่ดีทันทีแม้ว่าการแยกส่วนที่ได้จากองค์ประกอบ Cr จะผลที่ดีอย่างไรก็ตามนอกจากนี้แบบจำลองสีผิว YCbCr มีปัญหาในด้านสีผิวที่มีลักษณะสีผิวไปทางสีดำซึ่งต่างจากปริภูมิสี HSI ข้างต้น นอกจากนี้ Lui [42] ได้เสนอว่าไม่มีแบบจำลองสีผิวใดเพียงปริภูมิเดียวที่สามารถทำงานได้อย่างมีประสิทธิภาพกับทุกชนิดลำดับภาพ

ดังนั้น วิทยานิพนธ์นี้เสนอแมปสีอ้างอิงที่ใช้ในการแยกบริเวณที่สนใจบริเวณใบหน้าและบริเวณที่ไม่สนใจ โดยแมปสีอ้างอิงที่ใช้ในการนิยามสีผิวในวิทยานิพนธ์นี้นั้นเป็นการใช้องค์ประกอบเฉดสี H ของปริภูมิสี HSI และองค์ประกอบสี Cr ของปริภูมิสี YCbCr เพราะว่าการใช้องค์ประกอบสีผิวระหว่างสองปริภูมิร่วมกันทำให้อัตราส่วนการแยกส่วนบริเวณใบหน้าผิดพลาดมี

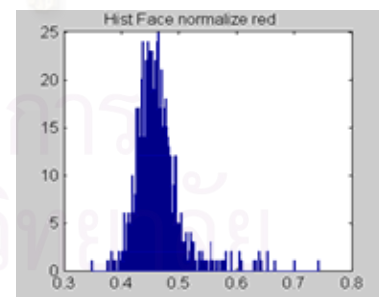
ค่าลดลงกว่าเพียงการใช้องค์ประกอบสีภายในปริภูมิสีเดียว โดยเมื่อทำการทดลอง โดยใช้แบบจำลองชนิดไม่มีตัวแปร ซึ่งได้แสดงในส่วนของผลการทดลอง

3.1.2 การกำจัดพื้นหลัง (Background Elimination)

กรรมวิธีของขั้นตอนนี้คือการขจัดบางจุดภาพที่ได้จากขั้นตอนแรกที่เป็นสัญญาณรบกวนที่มีลักษณะคล้ายสีผิว ซึ่งนิยามว่าเป็นสีผิวเทียม จากความจริงที่ว่า สีผิวของมนุษย์นั้นมีแนวโน้มที่จะมีค่าความเข้มสีของสีแดงเป็นอย่างมาก (predominance of red) แต่จะมีค่าความเข้มสีของสีน้ำเงินค่อนข้างน้อยหรือไม่มีค่าสีน้ำเงินเลยในส่วนของสีผิว นอกจากนี้ Mei-Juan Chen และ Meng-Chieh Chi พบว่าสีผิวและสีผิวปลอมนั้นอยู่ในช่วงค่าของความเข้มสีเดียวกันในองค์ประกอบสี Cr แต่ทว่าสีผิวและสีผิวเทียมนั้นจะมีช่วงค่าความแปรปรวนที่แตกต่างกันในปริภูมิสี RGB ค่าฮิสโตแกรมของโดเมนสีน้ำเงินและสีแดงแสดงให้เห็นดังรูปที่ 3.3 ผลทางสถิติได้แสดงให้เห็นว่าค่าฮิสโตแกรมของสีแดงและสีน้ำเงินที่เป็นช่วงค่าของสีผิวในปริภูมิสีนอร์มอลไลซ์แดงเขียวน้ำเงินนั้นจะมีการกระจายตัวอยู่ในช่วงแคบๆ ดังนั้นจึงสามารถนำคุณสมบัตินี้ไปใช้ขจัดสีผิวปลอมได้ โดยทำการกำหนดช่วงค่าขีดเริ่มเปลี่ยนที่เลือกจากค่าฮิสโตแกรมจากกลุ่มตัวอย่างและสร้างเงื่อนไขว่าถ้าหากว่าแต่ละจุดภาพจากขั้นตอนแรกมีค่าความเข้มสีของสีน้ำเงินอยู่ในช่วงค่าที่กำหนดและมีค่าความเข้มสีของสีแดงอยู่ในช่วงค่าที่กำหนดนี้ด้วยแล้ว หมายความว่าจุดภาพนั้นเป็นจุดภาพของสีผิวมนุษย์ ในทางตรงกันข้าม ถ้าจุดภาพมีค่าความเข้มสีของสีน้ำเงินอยู่ในช่วงค่าที่กำหนดแต่ทว่าค่าความเข้มสีของสีแดงไม่ได้อยู่ในช่วงค่าที่กำหนด จุดภาพนั้นจะถูกกำหนดให้เป็นจุดภาพของสีผิวเทียม โดยกำหนดให้ช่วงค่าของสีน้ำเงิน อยู่ระหว่าง 0.15-0.3 และช่วงค่าของสีแดงอยู่ระหว่าง 0.4-0.5



(ก)



(ข)

รูปที่ 3.3 ฮิสโตแกรมของโดเมนสีน้ำเงินและสีแดง ในปริภูมิสีนอร์มอลไลซ์แดงเขียวน้ำเงิน

(ก) ฮิสโตแกรมของโดเมนสีน้ำเงิน

(ข) ฮิสโตแกรมของโดเมนสีแดง

3.1.3 กระบวนการข้างหลัง (Postprocessing) และการติดตามวัตถุ (Object Tracking)

กระบวนการนี้เป็นกระบวนการดึงเอาวัตถุที่มีสีผิวขนาดใหญ่จากลำดับภาพออกมาโดยการใช้ตัวดำเนินการเชิงสัญญาณแบบปิดซึ่งถูกใช้งานทั่วไปในการสร้างความราบเรียบให้กับวัตถุ การเติมเต็มช่องว่างภายในวัตถุ และการขจัดสัญญาณรบกวนภายในลำดับภาพ โดยตัวดำเนินการเชิงสัญญาณแบบปิดคือการทำไดเลชันแล้วตามด้วยอีโรชัน แต่อย่างไรก็ตามกระบวนการนี้ขจัดได้เพียงสัญญาณรบกวนที่เป็นบริเวณขนาดเล็กเท่านั้นแต่ไม่สามารถขจัดสัญญาณรบกวนจากมาสก์ไปได้ทั้งหมดอย่างสมบูรณ์ ดังนั้นเพื่อที่จะขจัดสัญญาณรบกวนที่หลงเหลืออยู่จากขั้นตอนข้างต้น ซึ่งมีขนาดพื้นที่ของจุดภาพที่เป็นสัญญาณรบกวนที่ไม่ใหญ่มาก จึงทำการแบ่งภาพออกเป็นบล็อกโดยแต่ละบล็อกมีพื้นที่ 4×4 จุดภาพ แล้วพิจารณาว่า ถ้าภายในบล็อกใดมีจุดภาพมากกว่าหรือเท่ากับ 8 จุดภาพแสดงว่าบล็อกนั้นเป็นส่วนของมาสก์ แต่ถ้าบล็อกใดมีจุดภาพน้อยกว่า 8 แสดงว่าบล็อกนั้นเป็นจุดภาพที่เป็นสัญญาณรบกวนซึ่งจะทำให้บล็อกนั้นกลายเป็นบล็อกที่ว่างเปล่าไปด้วยวิธีการนี้ทำให้สามารถขจัดจุดภาพที่เป็นสัญญาณรบกวนไปได้ อย่างไรก็ตาม การใช้ข้อมูลเชิงพื้นที่ในการขจัดจุดภาพที่เป็นสัญญาณรบกวนยังคงไม่เพียงพอในกรณีที่มีวัตถุพื้นหลังปรากฏขึ้นมาอย่างทันทีทันใดทำให้มีปัญหาในการแยกส่วนเกิน ดังนั้นด้วยสมมติฐานที่ว่าลำดับภาพจำพวกวิดีโอเฟรมหรือวิดีโอคอนเฟอเรนซ์ไม่มีการเคลื่อนที่ที่เปลี่ยนแปลงตำแหน่งของใบหน้าอย่างมาก โดยทั่วไปแล้วตำแหน่งของใบหน้าอยู่ในตำแหน่งเดิมหรือใกล้เคียงตำแหน่งเดิม ดังนั้นการใช้ข้อมูลเชิงเวลามาช่วยในการขจัดจุดภาพที่เป็นสัญญาณรบกวนซึ่งเป็นจุดภาพที่เป็นสีผิวที่ปรากฏขึ้นมาอย่างทันทีทันใดและไม่ได้อยู่ในตำแหน่งที่ใกล้เคียงกับตำแหน่งของใบหน้า ด้วยการใช้การติดตามวัตถุซึ่งเป็นเสมือนหน่วยความจำสำหรับเก็บจุดภาพที่เป็นสีผิวของลำดับภาพก่อนหน้าไว้เพื่อเป็นแนวทางให้กับลำดับภาพถัดไป ดังนั้นจากข้อมูลของลำดับภาพก่อนหน้าทำให้สร้างเงื่อนไขว่าถ้าจุดภาพที่เป็นสีผิวภายในบล็อกของลำดับภาพถัดไปนั้นครอบคลุมเกินกึ่งหนึ่งของลำดับภาพก่อนหน้า บล็อกนั้นจะถูกพิจารณาว่าเป็นบริเวณของสีผิว แต่ถ้าไม่เข้าตามเงื่อนไขข้างต้นบล็อกนั้นก็จะถูกขจัดไปซึ่งด้วยวิธีการนี้ทำให้สามารถขจัดจุดภาพที่เป็นสีผิวที่ปรากฏขึ้นมาอย่างทันทีทันใดได้ ดังนั้นผลลัพธ์จากกระบวนการนี้จะประกอบไปด้วยบริเวณใบหน้าที่ปราศจากจุดภาพที่เป็นสัญญาณรบกวน.

3.2 การแยกส่วนบริเวณที่สนใจที่เป็นวัตถุที่มีความหมาย

การแยกส่วนบริเวณที่สนใจที่เป็นวัตถุ ประกอบไปด้วย การแยกส่วนเชิงพื้นที่ การแยกส่วนเชิงเวลา กรรมวิธีสเนกเพื่อให้ได้รูปร่างก่อนหน้าของบริเวณที่สนใจ กฎการตัดสินใจ กระบวนการข้างหลังและการติดตามวัตถุ

3.2.1 รูปร่างก่อนหน้าของบริเวณที่สนใจ (Shape Prior of Region-of-Interest)

การใช้บริเวณที่สนใจที่เป็นบริเวณใบหน้ามาช่วยในการแยกส่วนวัตถุที่สนใจนั้นจะนำมาใช้เพื่อกำหนดขอบเขตของวัตถุอย่างหยาบเพื่อที่จะนำไปใช้ในขั้นตอนถัดไป โดยการใช้ตำแหน่งของบริเวณใบหน้าในการกำหนดขอบเขตของวัตถุนั้นกระทำโดยการใช้ตำแหน่งของใบหน้าในการกำหนดตำแหน่งมาสก์ที่มีลักษณะรูปร่างสอดคล้องกับลำดับภาพมนุษย์ครึ่งตัวที่มีศีรษะและไหล่ (head and shoulder) โดยใช้ขอบของมาสก์นี้เป็นจุดควบคุมเริ่มต้นอย่างอัตโนมัติของสเนกในการหดตัวเข้าไปหาขอบของวัตถุเพื่อให้ได้การแยกส่วนบริเวณวัตถุที่ดีขึ้นเพื่อนำไปใช้ในขั้นตอนการแยกส่วนเชิงเวลาและการแยกส่วนเชิงพื้นที่



รูปที่ 3.4 แสดงตัวอย่างผลลัพธ์จากขั้นตอนการกำหนดขอบเขตของวัตถุจากใบหน้าและการแยกส่วนด้วยกรรมวิธีสเนก

3.2.2 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)

กรรมวิธีการแยกส่วนภาพเชิงเวลานั้นใช้ประโยชน์จากข้อมูลการเคลื่อนที่มากกว่าข้อมูลเชิงพื้นที่เพื่อกำหนดตำแหน่งและขอบของวัตถุ เนื่องจากวัตถุที่สนใจมักจะมีการเคลื่อนที่ ดังนั้นการตรวจจับการเปลี่ยนแปลงจึงเป็นโครงสร้างหลักของการแยกส่วนเชิงเวลา จุดมุ่งหมายของการแยกส่วนภาพเชิงเวลาคือเพื่อให้ได้ส่วนของพื้นที่หรือวัตถุที่ต้องการมากที่สุด เพื่อจะได้ถูกนำไปใช้โดยส่วนของการตัดสินใจได้อย่างมีประสิทธิภาพ ถ้าการแยกส่วนภาพเชิงเวลาไม่สามารถแยกส่วนของวัตถุที่ต้องการได้มากพอแล้ว กระบวนการตัดสินใจก็ให้ผลได้ไม่ดีนัก การแยกส่วนภาพเชิงเวลาใช้ประโยชน์จากภาพ 2 เฟรมที่อยู่ติดต่อกัน เนื่องจากว่า เมื่อเกิดการเคลื่อนไหวขึ้นแล้ว ก็เกิดการเปลี่ยนแปลงของความเข้มแสงด้วย การแยกส่วนภาพเชิงเวลามีกระบวนการคือ

ขั้นตอนแรกคือนำความเข้มแสงของภาพเฟรมที่ $i+1$ และภาพเฟรมที่ i มาตรวจจับการเปลี่ยนแปลงแล้วทำให้เป็นค่าสัมบูรณ์

$$FD(x,y,t) = |I(x,y,t) - I(x,y,t-1)| \quad (3.1)$$

เมื่อ $I(x,y,t)$ คือ ข้อมูลความเข้มแสงของเฟรมที่ t และ $FD(x,y,t)$ คือ ความต่างเฟรมที่ t กับเฟรมก่อนหน้า

ดังนั้นพื้นที่ส่วนที่ไม่เกิดการเคลื่อนไหวจะมีค่าเป็นศูนย์ แต่เนื่องจากอุปกรณ์ที่ใช้จับภาพอาจมีการรบกวนเนื่องจากความร้อน ความผิดพลาดที่เกิดจากการควอนไทซ์ และอื่นๆ อันอาจมีผลทำให้ค่าความเข้มแสงของส่วนที่ไม่เคลื่อนไหวไม่เป็นศูนย์ทั้งหมด วิธีแก้ไขขั้นต้นคือกำหนดค่าจุดเปลี่ยนไว้ค่าหนึ่งเพื่อใช้ประมาณว่าความแตกต่างนั้น เกิดจากการเคลื่อนไหวจริงๆ มิใช่เกิดจากความผิดพลาดหรือการรบกวนของสิ่งอื่นๆ มาส์กความต่างเฟรม (Frame Difference Mask : FDM) นิยามได้ดังนี้

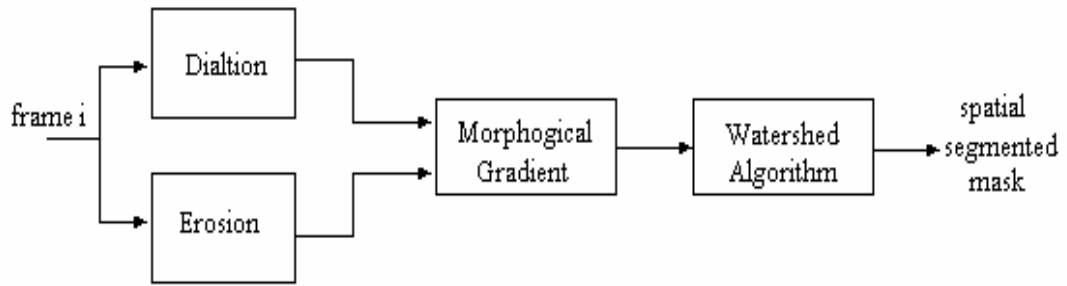
$$FDM_i(x,y,t) = \begin{cases} 1 & \text{ถ้า } FD \geq Th \\ 0 & \text{ถ้า } FD < Th \end{cases} \quad (3.2)$$

โดยที่ Th เป็นค่าขีดเริ่มเปลี่ยนที่ใช้ในการแยกจุดภาพที่มีการเคลื่อนไหวที่จริงซึ่งจำเป็นต้องกำหนดไว้ก่อนหน้า

ขั้นตอนที่สองคือหลังจากนั้นผลลัพธ์ที่ได้จะทำให้มีรูปแบบที่ง่ายโดยการใช้ตัวปฏิบัติการเชิงเส้นฐานแบบปิดและกำจัดบริเวณที่มีจำนวนจุดภาพน้อยกว่าจำนวนที่กำหนดออกไป ซึ่งอยู่บนสมมติฐานที่ว่าบริเวณที่มีขนาดเล็กไม่ควรเป็นส่วนของวัตถุและมีแนวโน้มที่จะเป็นจุดรบกวนเป็นอย่างมาก พื้นที่หลังที่ไม่ครอบคลุมในส่วนของมาส์กของการตรวจจับการเคลื่อนไหวจะถูกกำจัดออก ทำให้ได้มาส์กของทุกๆ จุดภาพที่เป็นจุดภาพของวัตถุที่เคลื่อนที่ เรียกว่า มาส์กของภาพวัตถุ

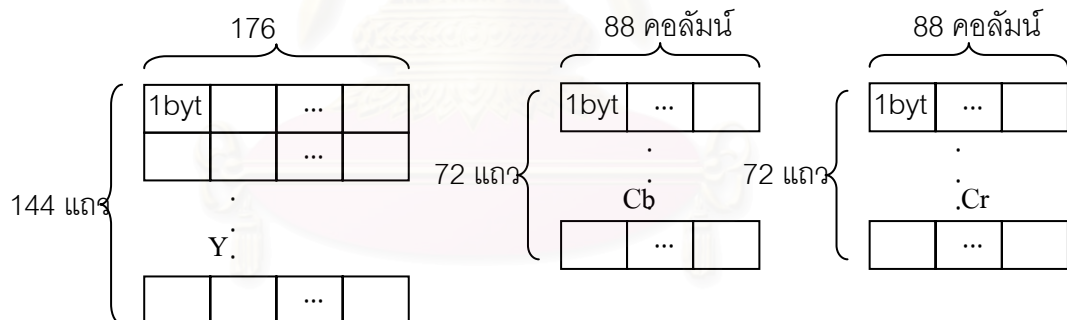
3.2.3 การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation)

ข้อได้เปรียบของการแยกส่วนเชิงพื้นที่คือให้ผลลัพธ์ที่ได้จากการแยกส่วนที่เป็นขอบของวัตถุที่แม่นยำเป็นอย่างมาก ดังนั้นการแยกส่วนภาพเชิงพื้นที่นั้น กระทำเพื่อแบ่งภาพออกเป็นส่วนย่อยๆ โดยขอบของส่วนย่อยๆ ของภาพนั้นต้องเป็นขอบของวัตถุด้วย กระบวนการที่กระทำแล้วให้ผลลัพธ์เช่นนี้คือ กรรมวิธีสันปันน้ำ โดยมีขั้นตอนก่อนการกระทำกรรมวิธีสันปันน้ำนั้นคือต้องอ่านค่าความเข้มแสงของลำดับวิดีโอเข้ามา แล้วทำกระบวนการใดเลชันกับอีโรชันเพื่อนำไปหาค่าเกรเดียนต์เชิงเส้นฐานของภาพนั้น ต่อจากนั้นจึงนำมาทำตามกรรมวิธีสันปันน้ำดังแสดงในรูปที่ 3.5



รูปที่ 3.5 โครงสร้างการแยกภาพส่วนเชิงพื้นที่

โครงสร้างของลำดับวิดีโอแบบ YCbCr 4:2:0 มีการจัดรูปแบบดังรูปที่ 3.5 โดย 1 เฟรมจะมี $176 * 144 + (176/2) * (144/2) + (176/2) * (144/2) = 38013$ ไบต์ การนำไปใช้ในการแยกส่วนภาพเชิงพื้นที่นั้นนำเฉพาะความเข้มแสงไปใช้เท่านั้น ภาพที่ได้ออกมาจากกระบวนการแยกส่วนภาพเชิงพื้นที่นั้นจะเป็นภาพขาวดำโดยแบ่งออกเป็นส่วนย่อยๆ ซึ่งแต่ละส่วนจะมีเส้นขอบเขตเป็นสีดำล้อมรอบพื้นที่ที่เป็นสีขาว รูปที่ 3.6 แสดงตัวอย่างผลลัพธ์จากขั้นตอนการแยกส่วนเชิงพื้นที่ตามขั้นตอนในรูปที่ 3.7 การประมาณค่าเกรเดียนต์ของภาพ ทำได้โดยได้เลขชั้นลบด้วยอีโรชันของภาพ



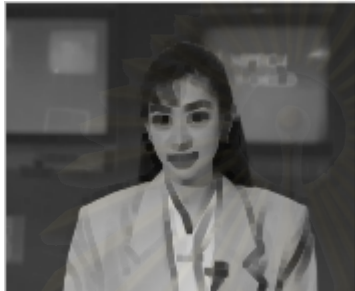
รูปที่ 3.6 โครงสร้างการจัดวางตำแหน่งไบต์



(ก)



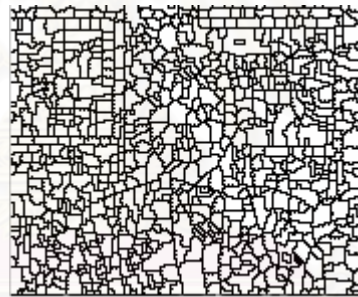
(ข)



(ค)



(ง)



(จ)

รูปที่ 3.7 แสดงตัวอย่างผลลัพธ์จากขั้นตอนการแยกส่วนเชิงพื้นที่

(ก) ลำดับภาพต้นฉบับลำดับภาพ Akiyo (ข) การปฏิบัติการไคเลชัน

(ค) การปฏิบัติการอีโรชัน

(ง) ค่าการประมาณเกรเดียนต์

(จ) กรรรมวิธีสันปันน้ำ

3.2.4 กฎการตัดสินใจ (Decision Rule)

กระบวนการนี้จะพิจารณาข้อมูลของมาสกี้ในแต่ละส่วนในกระบวนการตัดสินใจ ซึ่งนำเอาคุณสมบัติที่แตกต่างกันมาใช้ ในขั้นตอนการแยกส่วนโดยใช้ข้อมูลเชิงพื้นที่ที่จะให้จำนวนบริเวณที่แตกต่างกันเป็นจำนวนมาก แต่จะให้ข้อมูลส่วนของขอบภาพวัตถุที่ค่อนข้างแม่นยำ ขณะที่ข้อมูลเชิงเวลาจะให้ข้อมูลเพิ่มเติมสามารถบอกการเปลี่ยนแปลงและการเคลื่อนที่ของภาพวัตถุ ข้อมูล

ของการเปลี่ยนแปลงความเข้มจะให้ข้อมูลส่วนของบริเวณขอบไม่ชัดเจน แต่จะระบุบริเวณที่มีการเคลื่อนที่ของวัตถุและบริเวณที่ไม่มีการเคลื่อนที่หรือเป็นส่วนของพื้นหลังได้ ส่วนข้อมูลของการวิเคราะห์การเคลื่อนที่จะสามารถระบุตำแหน่งและทิศทางของแต่ละวัตถุได้ กระบวนการตัดสินใจนี้ จะใช้ข้อมูลเชิงพื้นที่ในการกำหนดขอบเขตของบริเวณ ซึ่งจะสร้างมาสก์ของวัตถุโดยการรวมบริเวณของแต่ละส่วนความเข้มที่ได้จากการแยกส่วนเชิงพื้นที่เข้าด้วยกันตามข้อมูลเชิงเวลา

การพิจารณาสร้างมาสก์ของภาพวัตถุ จะใช้ทั้งบริเวณที่ได้จากข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา แต่เนื่องจากการแยกส่วนเชิงพื้นที่ จะทำให้ขอบเขตมีความถูกต้องมากขึ้น จึงใช้เป็นบริเวณหลักสำหรับทำมาสก์ให้สมบูรณ์สำหรับแสดงขอบเขตของภาพวัตถุ ซึ่งมีขั้นตอนดังนี้

กรณีจุดภาพภายในบริเวณที่ได้จากการแยกส่วนเชิงพื้นที่ถูกตรวจสอบว่าเป็นฉากหน้า (foreground) ในมาสก์ของการตรวจจับการเปลี่ยนแปลง มากกว่าครึ่งหนึ่งของทั้งบริเวณพื้นที่นั้น ทุกจุดภาพของบริเวณพื้นที่นั้นจะกำหนดให้เป็นฉากหน้า ส่วนกรณีจุดภาพภายในบริเวณจากการแยกส่วนเชิงพื้นที่ถูกตรวจสอบว่าเป็นฉากหน้าซึ่งอยู่ในมาสก์ของการตรวจจับการเปลี่ยนแปลง น้อยกว่าครึ่งหนึ่งของบริเวณพื้นที่นั้น ทุกจุดภาพของบริเวณพื้นที่นั้นจะกำหนดให้เป็นพื้นหลัง

3.2.5 กระบวนการข้างหลัง (Postprocessing)

ถ้าวัตถุที่ผ่านกระบวนการแยกส่วนแล้วยังประกอบด้วยบริเวณหลายส่วนอยู่หรือยังมีบริเวณขนาดเล็กเกินไปที่ไม่น่าจะเป็นขนาดของภาพวัตถุ ซึ่งเกิดเนื่องจากความคลาดเคลื่อนของการวิเคราะห์การเคลื่อนที่ ขั้นตอนนี้จะช่วยปรับปรุงผลของการแยกส่วนภาพเพื่อแก้ปัญหาการแยกส่วนของภาพมากเกินไป โดยการรวมบริเวณที่มีลักษณะใกล้เคียงกันเข้าด้วยกัน ซึ่งจะขึ้นอยู่กับข้อกำหนด 2 ข้อ คือ ขนาด ถ้าบริเวณนั้นมีขนาดเล็กกว่าขนาดของจำนวนจุดภาพที่กำหนดไว้ บริเวณนั้นจะถูกรวมเข้ากับบริเวณข้างเคียง ถ้ามีบริเวณข้างเคียงมากกว่าหนึ่งบริเวณ บริเวณนั้นจะรวมกับบริเวณที่มีลักษณะเหมือนกันและมีขนาดใหญ่กว่าเพื่อที่จะกำจัดบริเวณที่ไม่มี ความหมาย ส่วนอีกข้อหนึ่ง คือ การรวมกัน โดยบริเวณที่มีการเคลื่อนที่เหมือนกันซึ่งอยู่ติดกันจะถูกรวมเข้าด้วยกัน ซึ่งบริเวณนี้มีแนวโน้มที่จะเป็นวัตถุที่มีการเคลื่อนที่ในทิศทางเดียวกันหรือเป็นวัตถุเดียวกัน หลังจากนั้นก็จะได้มาสก์สำหรับการแยกส่วนภาพวัตถุที่สามารถแยกส่วนภาพของภาพวัตถุได้

มาสก์ของภาพวัตถุที่ได้จะมีค่าเป็นจำนวนบวก ซึ่งแต่ละค่าจะแทนวัตถุที่แตกต่างกัน ส่วนพื้นหลังของภาพจะมีค่าเป็นศูนย์ โดยที่มาสก์ในเฟรมแรกจะมีค่าเป็นศูนย์ทั้งหมดเนื่องจากยังไม่มี การแยกส่วนของภาพเกิดขึ้น

3.2.6 การติดตามวัตถุ (Object Tracking)

หลังจากผ่านกระบวนการแยกส่วนภาพวัตถุในเฟรมหนึ่งแล้ว จะได้มาส์กของภาพวัตถุ ซึ่งจะนำมาใช้เป็นข้อมูลในการแยกส่วนภาพวัตถุในเฟรมถัดไป โดยจะนำข้อมูลของมาส์กของภาพวัตถุที่ได้เข้ามารวมใช้ในกระบวนการแยกส่วนภาพของเฟรมปัจจุบัน ซึ่งจะแทนด้วยมาส์กของการขจัดเซกเมนต์ที่ ในส่วนนี้นอกจากจะช่วยในการแยกส่วนภาพแล้วยังช่วยในการติดตามการเคลื่อนที่ของแต่ละวัตถุ ดังนั้นการติดตามวัตถุจึงมีจุดประสงค์เพื่อยังคงขอบเขตของวัตถุนั้นไว้เพื่อความรวดเร็วในการแยกส่วนภาพในเฟรมถัดไป แต่ทว่าการติดตามวัตถุจำเป็นต้องมีหน่วยความจำสำรองเก็บค่าวัตถุของภาพก่อนหน้า ดังนั้นการติดตามวัตถุจึงเป็นเสมือนหน่วยความจำหน่วยหนึ่งซึ่งมีหน้าที่เก็บมาส์กวัตถุของภาพเฟรมที่ i เพื่อนำไปใช้ในกระบวนการตัดสินใจของภาพเฟรมที่ $i+1$ ต่อไป

3.2.7 การแยกส่วนภาพเชิงพื้นที่และเวลา

การแยกส่วนภาพเชิงพื้นที่และเวลานั้นนำเสนอโดย ดัชกรณ ดันเจริญ และคณะ [4] มีโครงสร้างดังรูปที่ 3.8 ซึ่งประกอบด้วย 4 กระบวนการหลักดังนี้

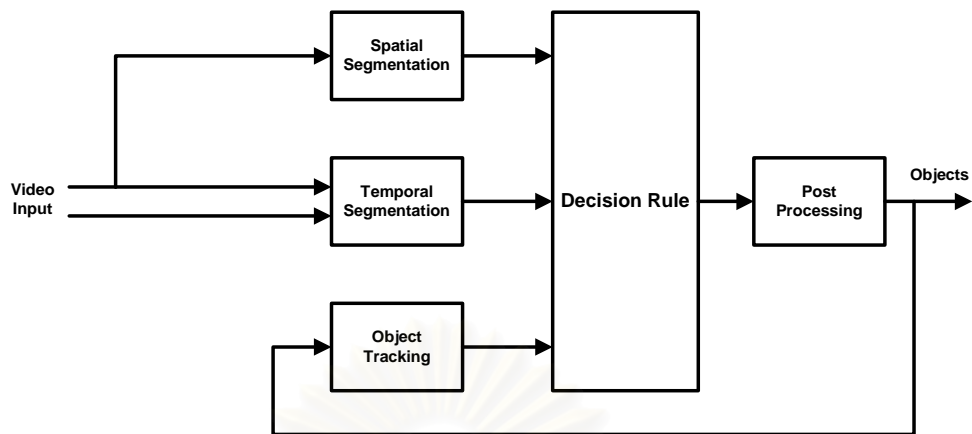
1. การแยกส่วนภาพเชิงพื้นที่ ประกอบด้วย 3 กระบวนการย่อยคือ

- การทำให้ภาพเรียบง่าย โดยใช้ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิด การประมาณเกรเดียนต์ กระบวนการย่อยนี้กระทำโดยเกรเดียนต์เชิงสัญญาณและกรรมวิธีสันปันน้ำ

2. การแยกส่วนภาพเชิงเวลา กระบวนการนี้นำภาพ 2 เฟรมมาหาค่าความแตกต่างแล้วกำหนดว่าถ้าจุดภาพใดมีค่าความแตกต่างมากกว่าค่าที่กำหนด ให้ตัดสินใจว่าจุดภาพนั้นมีความเปลี่ยนแปลง

3. กระบวนการตัดสินใจ เป็นกระบวนการรวมส่วนที่ได้จากการแยกส่วนเชิงเวลาและการแยกส่วนเชิงพื้นที่เข้าด้วยกันโดยใช้หลักการคือ ถ้าส่วนที่ได้จากการแยกส่วนเชิงเวลาครอบคลุมพื้นที่ของกระบวนการแยกส่วนภาพเชิงพื้นที่มากกว่าร้อยละ 50 ให้ตัดสินใจว่าพื้นที่นั้นเป็นวัตถุ

4. กระบวนการติดตามวัตถุ เป็นกระบวนการในตรวจจับตำแหน่งของวัตถุในเฟรมถัดไปว่าอยู่ที่ตำแหน่งใด



รูปที่ 3.8 กรรมวิธีที่นำเสนอโดย ดัชกรณันต์ ดันเจริญ และคณะ [4]

3.2.8. การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐาน (Hypothesis Testing)

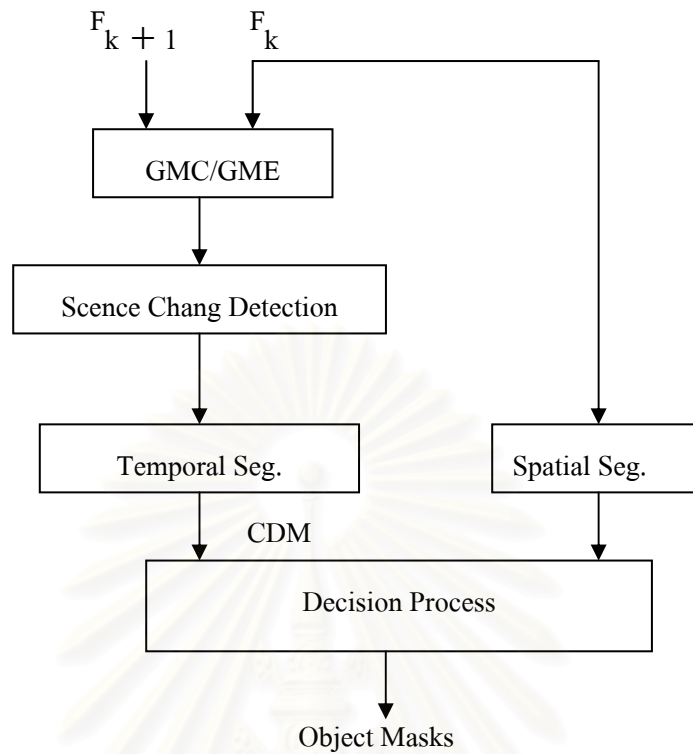
การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐานนั้นนำเสนอโดย MunchurlxKim, Jae Gark Choi, Daelee Kim, Hyung Lee [32] ซึ่งมีโครงสร้างดังรูปที่ 3.9 ประกอบด้วย 5 กระบวนการหลักดังนี้

1. การประมาณและการชดเชยการเคลื่อนที่แบบโกลบอล (Global Motion Estimation : GME และ Global Motion Compensation : GMC)

ในกระบวนการนี้ประกอบด้วยกระบวนการย่อยอีก 2 กระบวนการคือ การประมาณการเคลื่อนที่แบบค้นหาเต็ม (Full Search Motion Estimation) และการประมาณค่าเวกเตอร์การเคลื่อนที่แบบโกลบอล

2. การตรวจจับการเปลี่ยนฉาก (Scene Change Detection)

กระบวนการนี้เป็นการตรวจว่าลำดับของภาพวิดีโอที่พิจารณานั้นมีการเปลี่ยนแปลงจากเฟรมที่แล้วหรือมีการเปลี่ยนฉากเดิมเป็นฉากอื่นหรือไม่ การตรวจจับทำได้โดยวัดค่าความแตกต่างสมบูรณ์เฉลี่ย (Mean Absolute Difference) ของเฟรมสองเฟรมติดกัน ถ้ามีค่ามากกว่าค่าที่ตั้งไว้ก็ตัดสินว่ามีการเปลี่ยนฉากเกิดขึ้นแล้วไม่ต้องทำกระบวนการแยกส่วนภาพเชิงเวลา



รูปที่ 3.9 กรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee

Myoung Ho Lee, Chieteuk และ Yo-Sung Ho

3. การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)

กระบวนการนี้นำภาพ 2 เฟรมมาหาค่าความแตกต่างและทำการทดสอบสมมติฐานโดยมีการดำเนินการดังนี้

- หาค่าความแตกต่างของภาพเฟรมที่ n และ $n+3$
- กำหนดส่วนที่เป็นพื้นหลังแล้วคำนวณหาความแปรปรวน S_1^1
- แบ่งภาพที่เป็นค่าความแตกต่างออกเป็นบล็อกย่อยๆเท่ากันแล้วคำนวณค่าความแปรปรวนของบล็อกนั้นๆ S_2^1

$$\text{- ค่าวนค่าทดสอบทางสถิติ } V = \frac{S_2^1}{S_1^1}$$

- กำหนดระดับนัยสำคัญแล้วอ่านค่า V_{th} จากตารางสถิติการแจกแจงแบบ F (F -distribution)

- ถ้า $V > V_{th}$ แล้ว กำหนดให้บล็อกนั้นๆ เป็นวัตถุ ถ้าไม่ใช่ กำหนดให้บล็อกนั้นๆ เป็นพื้นหลัง

- ทำเช่นนี้จนครบทุกบล็อก

4. การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation)

กระบวนการนี้ประกอบด้วย 3 กระบวนการย่อยคือ

- การทำให้ภาพเรียบง่าย (Simplification of Image) ในกระบวนการย่อยนี้ทำให้ภาพเรียบง่ายโดยใช้ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดซึ่งเป็นตัวดำเนินการเชิงสัญญาณ

- การประมาณเกรเดียนต์ (Gradient Approximation) กระบวนการย่อยนี้กระทำโดยเกรเดียนต์เชิงสัญญาณ

- กรรมวิธีสันปันน้ำ

5. กระบวนการตัดสินใจ เป็นกระบวนการรวมส่วนที่ได้จากการแยกส่วนเชิงเวลาและการแยกส่วนเชิงพื้นที่เข้าด้วยกันโดยใช้หลักการคือ ถ้าส่วนที่ได้จากการแยกส่วนเชิงเวลาครอบครองพื้นที่ของกระบวนการแยกส่วนภาพเชิงพื้นที่มากกว่าร้อยละ 90 ให้ตัดสินใจว่าพื้นที่นั้นเป็นวัตถุ จากนั้นนำผลที่ได้ไปช่วยในกระบวนการติดตามวัตถุในเฟรมถัดไป

ในบทนี้ได้กล่าวถึงกระบวนการแยกส่วนภาพวิดีโอออกเป็นวัตถุอย่างละเอียด ซึ่งประกอบด้วย 6 กระบวนการคือ รูปร่างก่อนหน้าของบริเวณที่สนใจ การแยกส่วนภาพเชิงพื้นที่ การแยกส่วนภาพเชิงเวลา กฎการตัดสินใจ กระบวนการข้างหลัง และการติดตามวัตถุ ทั้งยังกล่าวถึงการแยกส่วนภาพโดยใช้การทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลาเพื่อที่จะใช้เป็นตัวเปรียบเทียบ ส่วนในบทที่ 4 จะแสดงถึงผลลัพธ์ที่ได้จากการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยลำดับภาพแต่ละลำดับมีการเคลื่อนไหวที่แตกต่างกันออกไป และผลการเปรียบเทียบของการแยกส่วนภาพในแต่ละวิธี

บทที่ 4

ผลการทดลอง

บทนี้จะเป็นการแสดงผลการวิจัยจากกรรมวิธีการแยกส่วนบริเวณที่สนใจที่นำเสนอ โดยแบ่งเป็นสองส่วน คือ การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า และการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุ

ชุดภาพทั้งหมดที่ใช้ในการทดลองเป็นชุดภาพที่มีลักษณะเป็นวิดีโอคอนเฟอร์เรนซ์ มีวัตถุในภาพ 1 วัตถุคือตัวคนครึ่งตัว ซึ่งชุดภาพนั้นประกอบด้วยลำดับภาพมาตรฐาน 4 ชุดภาพ ซึ่งชุดภาพแต่ละชุด มีลักษณะการเคลื่อนไหวของวัตถุและองค์ประกอบของพื้นหลังแตกต่างกัน สามารถบรรยายลักษณะการเคลื่อนไหวของวัตถุและพื้นหลังได้ดังต่อไปนี้

ลำดับภาพวิดีโอมาตรฐาน เป็นลำดับภาพที่จัดเก็บในรูปแบบ YCbCr 4:2:0

1. ลำดับภาพ Claire พื้นหลังเป็นฉากที่มีสีสม่ำเสมอ มีการเคลื่อนไหวเพียงเล็กน้อยที่บริเวณปากและศีรษะต่อเนื่องตลอดลำดับภาพ
2. ลำดับภาพ Akiyo พื้นหลังบางส่วนเป็นฉากที่มีสีที่สม่ำเสมอ และบางส่วนไม่สม่ำเสมอ มีการเคลื่อนไหวพอประมาณ
3. ลำดับภาพ Grandmother พื้นหลังบางส่วนเป็นผนังสีสม่ำเสมอ บางส่วนเป็นใบไม้ มีการเคลื่อนไหวน้อยมาก โดยเฉพาะเฟรมแรกๆ แทบไม่เคลื่อนไหวเลย
4. ลำดับภาพ Miss Am พื้นหลังเป็นฉากที่มีสีสม่ำเสมอ มีการเคลื่อนไหวเพียงเล็กน้อยที่บริเวณปากและศีรษะต่อเนื่องตลอดลำดับภาพ

ผลการทดลองจะนำเสนอ ดังนี้

1. ผลการเปรียบเทียบความแตกต่างของแต่ละปริมาณที่ใช้ในการแยกส่วนภาพที่ได้จากกระบวนการแยกส่วนด้วยองค์ประกอบสี โดยพิจารณาข้อมูลขององค์ประกอบสีตามแบบจำลองสีทั้ง 3 แบบ (HS, HCr, CbCr) โดยจะเสนอเป็นตารางแสดง อัตราความถูกต้อง และอัตราความผิดพลาดเฉลี่ย ตามแบบจำลองสีทั้ง 3 แบบเทียบกับการแยกส่วนสีผิวด้วยสายตา
2. ผลการเปรียบเทียบความแตกต่างของแต่ละกรรมวิธีที่ใช้ในการแยกส่วนภาพที่ได้จากกระบวนการแยกส่วนด้วยองค์ประกอบสี Cbcr องค์ประกอบสี HS องค์ประกอบสี HCr และองค์ประกอบสี Cr โดยจะเสนอเป็นตารางแสดง อัตราความถูกต้อง และอัตราความผิดพลาดเฉลี่ย เทียบกับการแยกส่วนสีผิวด้วยสายตา

3. ภาพที่ได้จากการแยกส่วนโดยเทคนิคที่นำเสนอ การแยกส่วนที่ได้จากการทดสอบสมมติฐานและการแยกส่วนเชิงพื้นที่และเวลา และการเปรียบเทียบอัตราผิดพลาดต่อเฟรมของลำดับภาพทดสอบ

4.1 การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า

การแยกส่วนบริเวณที่สนใจนี้ ประกอบไปด้วยกรรมวิธีต่างๆ ดังนี้ การแยกส่วนด้วยองค์ประกอบสี การขจัดพื้นหลังด้วยคุณสมบัติของสีผิวมนุษย์ การใช้กรรมวิธีเชิงสัญญาณในการขจัดจุดภาพที่เป็นสัญญาณ

4.1.1. การแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าที่ได้จากกระบวนการแยกส่วนด้วยองค์ประกอบสี

โดยพิจารณาข้อมูลขององค์ประกอบสีตามแบบจำลองสีทั้ง 3 แบบ

ตารางที่ 4.1 แสดงค่าอัตราความถูกต้องเฉลี่ยจากการใช้ข้อมูลตามแบบจำลองสีแต่ละแบบมาเป็นเวกเตอร์ลักษณะเด่น

ตารางที่ 4.1 ค่าอัตราความถูกต้องในแต่ละประมุขสีของแบบจำลองชนิดที่ไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตา

องค์ประกอบสี	อัตราความถูกต้องเฉลี่ย (ร้อยละ)			
	ลำดับภาพ Grandma	ลำดับภาพ Akiyo	ลำดับภาพ Claire	ลำดับภาพ Miss AM
HCr	82.85	89.32	92.09	90.52
HS	51.4	92.49	85.48	73.83
CbCr	85.14	92.99	92.17	93.7

ตารางที่ 4.2 ค่าอัตราความผิดพลาดในแต่ละประมุขสีของแบบจำลองชนิดที่ไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตา

องค์ประกอบสี	อัตราความผิดพลาดเฉลี่ย (ร้อยละ)			
	ลำดับภาพ Grandma	ลำดับภาพ Akiyo	ลำดับภาพ Claire	ลำดับภาพ Miss AM
HCr	30.96	36.57	26.5	41.04
HS	64.3	44.01	35.38	70.21
CbCr	34.66	70.22	26.7	41.96



(ก) ภาพต้นแบบ



(ข) องค์ประกอบสี Cb และ Cr



(ค) องค์ประกอบสี Hue และ S



(ง) องค์ประกอบสี Hue และ Cr

รูปที่ 4.1 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสีของลำดับภาพ Grandmother



(ก) ภาพต้นแบบ



(ข) องค์ประกอบสี Cb และ Cr



(ค) องค์ประกอบสี Hue และ S

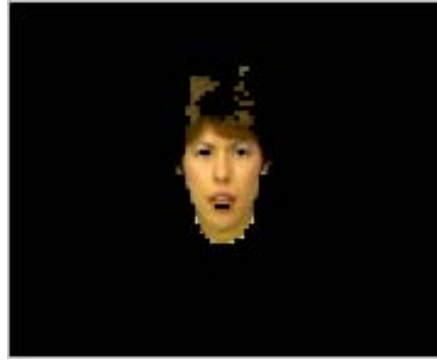


(ง) องค์ประกอบสี Hue และ Cr

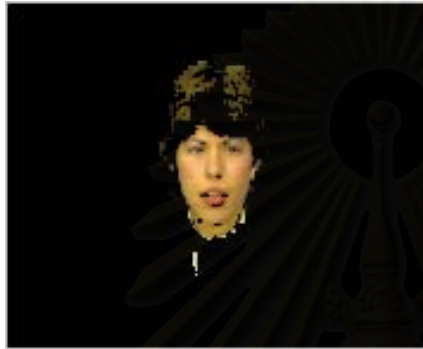
รูปที่ 4.2 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสีของลำดับภาพ Akiyo



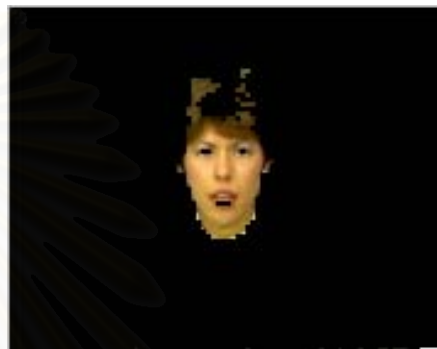
(ก) ภาพต้นแบบ



(ข) องค์ประกอบสี Cb และ Cr



(ค) องค์ประกอบสี Hue และ S

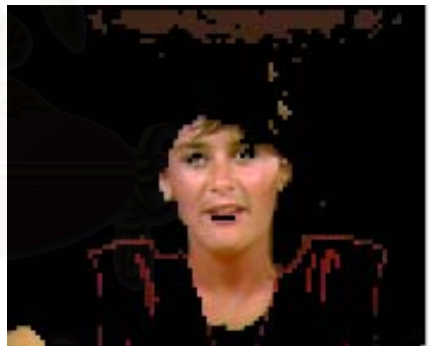


(ง) องค์ประกอบสี Hue และ Cr

รูปที่ 4.3 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสีของลำดับภาพ Claire



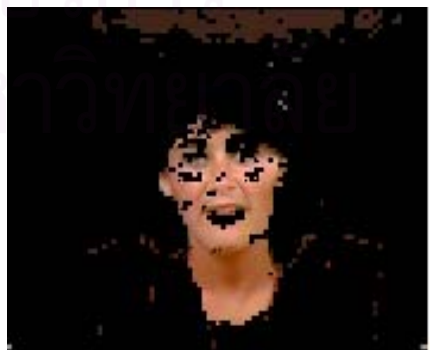
(ก) ภาพต้นแบบ



(ข) องค์ประกอบสี Cb และ Cr



(ค) องค์ประกอบสี Hue และ S



(ง) องค์ประกอบสี Hue และ Cr

รูปที่ 4.4 ผลการแยกส่วนภาพที่เปรียบเทียบของแต่ละปริภูมิสีของลำดับภาพ MissAM

4.1.2. ผลลัพธ์การแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าจากกระบวนการแยกส่วนจากกรรมวิธีที่นำเสนอและกรรมวิธีอ้างอิงต่าง ๆ

ในหัวข้อนี้เป็นการเปรียบเทียบกรรมวิธีที่นำเสนอกับกรรมวิธีอื่น ๆ โดยทำการเปรียบเทียบเชิงอัตวิสัย



รูปที่ 4.5 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่นำเสนอของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6



รูปที่ 4.6 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6



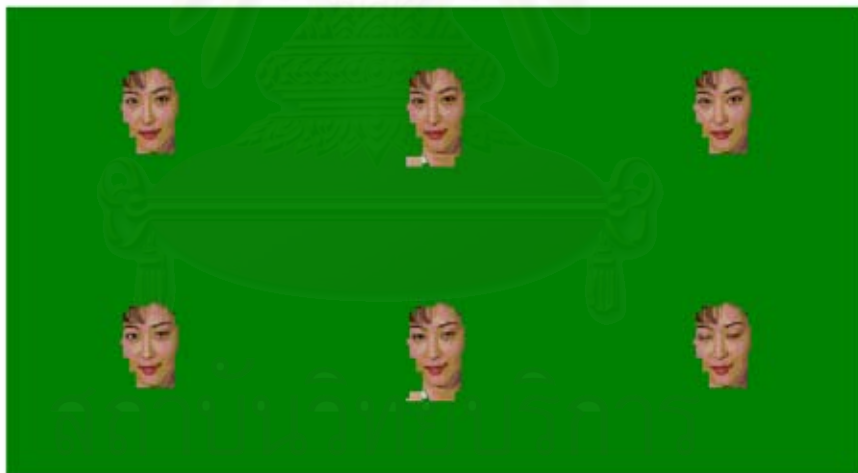
รูปที่ 4.7 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6



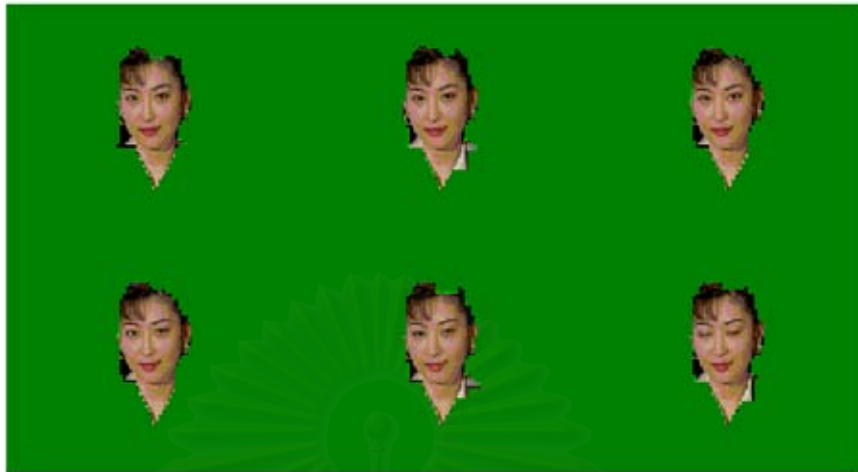
รูปที่ 4.8 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Grandmother เฟรมที่ 1 ถึง 6



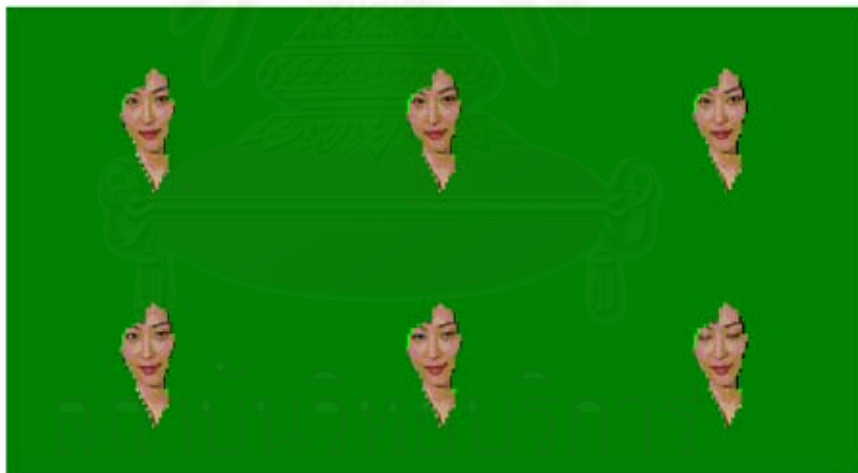
รูปที่ 4.9 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่เสนอของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6



รูปที่ 4.10 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6



รูปที่ 4.11 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6



รูปที่ 4.12 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Akiyo เฟรมที่ 1 ถึง 6



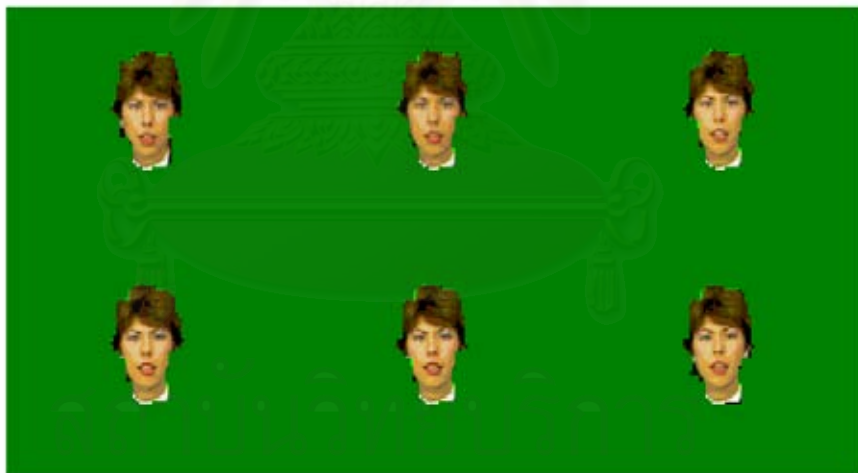
รูปที่ 4.13 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่นำเสนอของลำดับภาพ Claire เฟอร์มที่ 1 ถึง 6



รูปที่ 4.14 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Claire เฟอร์มที่ 1 ถึง 6



รูปที่ 4.15 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Claire เฟรมที่ 1 ถึง 6



รูปที่ 4.16 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Claire เฟรมที่ 1 ถึง 6



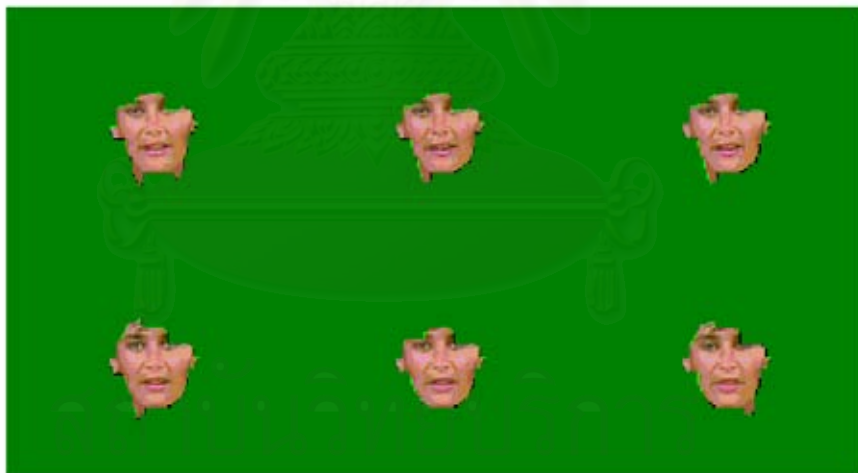
รูปที่ 4.17 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธีที่นำเสนอของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6



รูปที่ 4.18 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [8] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6



รูปที่ 4.19 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [22] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6



รูปที่ 4.20 ผลการแยกส่วนภาพบริเวณใบหน้าของกรรมวิธี [10] ของลำดับภาพ Miss Am เฟรมที่ 1 ถึง 6

4.1.3. การวัดผลของการแยกส่วนภาพบริเวณที่สนใจบริเวณใบหน้าจากกระบวนการแยกส่วนจากกรรมวิธีที่นำเสนอและกรรมวิธีอ้างอิงต่าง ๆ

ในหัวข้อนี้เป็นการเปรียบเทียบกรรมวิธีที่นำเสนอกับกรรมวิธีอื่น ๆ โดยทำการเปรียบเทียบเชิงวัตถุ โดยทำการพิจารณาที่ค่าอัตราความถูกต้องและค่าอัตราความผิดพลาดของบริเวณใบหน้าที่ยกส่วนได้ในแต่ละกรรมวิธีเทียบกับการแยกส่วนด้วยสายตา

โดยที่การวัดประสิทธิภาพของกรรมวิธีการแยกส่วนสีผิวนี้เป็นการประเมินผลเชิงปริมาณด้วยการวัดเชิงปริมาณนี้ทุกๆลำดับเฟรมที่ 5 จะถูกแยกส่วนโดยผู้ประเมินเป็นมาสก์อ้างอิงเพื่อนำเอาภาพที่แยกส่วนได้นำมาเปรียบเทียบกับมาสก์อ้างอิงนี้

ตัวแปรสำคัญที่นำมาใช้สำหรับการวัดประสิทธิภาพมีอยู่ด้วยกัน 4 ตัวแปร โดยสมมติว่า ณ ที่นี้ เราสนใจบริเวณใบหน้า จึงกำหนดให้เป็นฉากหน้า (Foreground) ดังนั้นส่วนที่ไม่ใช่ใบหน้าจะถูกกำหนดเป็นฉากหลัง (Background) ทั้งหมด

ค่า DR ย่อจาก “Detection Rate” คือ อัตราความถูกต้องในการตรวจหา มีหน่วยเป็น (%) เปอร์เซนต์ สามารถหาได้จากสมการที่ 4.1

$$DR = \frac{TP}{TP + FN} \times 100\% \quad (4.1)$$

ค่า FAR ย่อจาก “False Alarm Rate” คือ อัตราการเกิดความผิดพลาด มีหน่วยเป็น (%) เปอร์เซนต์ สามารถหาได้จากสมการที่ 4.2

$$FAR = \frac{FP}{TP + FP} \times 100\% \quad (4.2)$$

- TP ย่อจาก “True Positive” นิยามของ TP คือจุดภาพส่วนบริเวณใบหน้าที่ตรวจหาได้ถูกต้อง

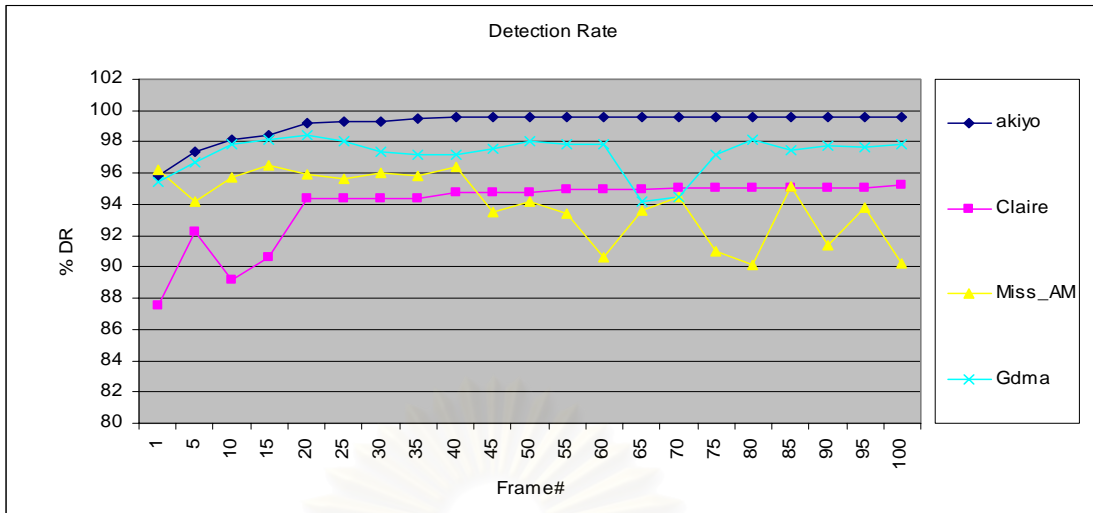
- TN ย่อจาก “True Negative” นิยามของ TN คือจุดภาพส่วนฉากหลังที่ตรวจหาได้ถูกต้อง

- FP ย่อจาก “False Positive” นิยามของ FP คือจุดภาพส่วนฉากหลังแต่ระบบตีความให้ผลเป็นค่าบริเวณใบหน้า

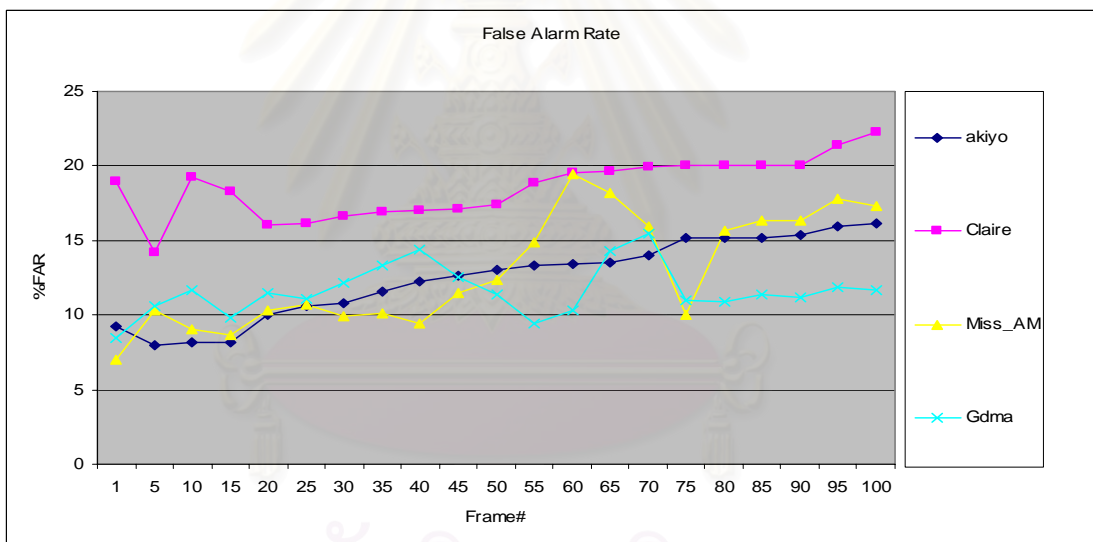
- FN ย่อจาก “False Negative” นิยามของ FN คือจุดภาพส่วนบริเวณใบหน้าแต่ระบบตีความให้ค่าเป็นฉากหลัง

$$False\ Positive\ Rate = \frac{FP}{TN + FP} \quad (4.3)$$

$$False\ Negative\ Rate = \frac{FN}{FN + TP} \quad (4.4)$$

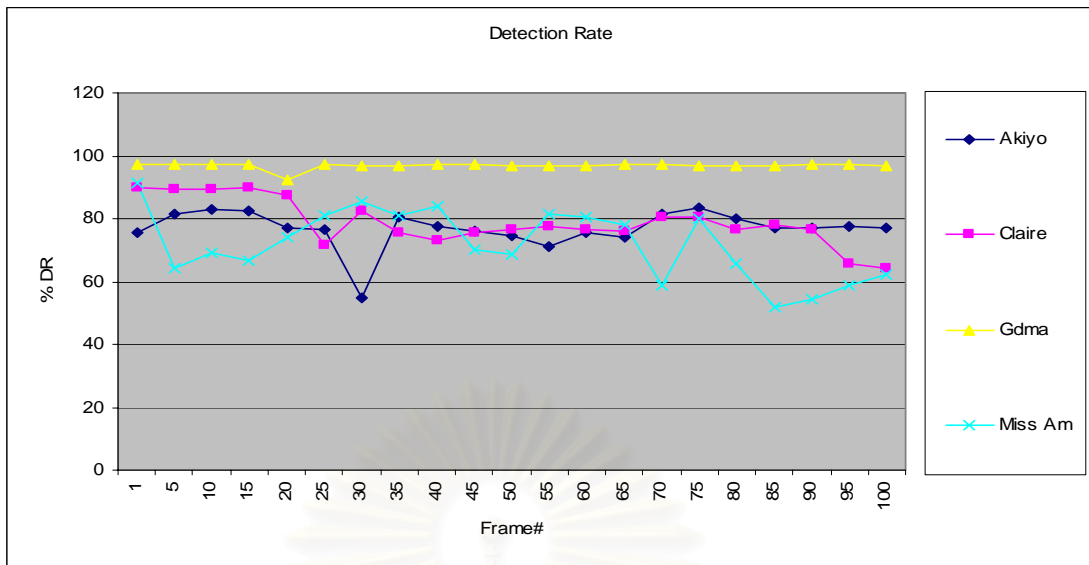


รูปที่ 4.21 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีที่นำเสนอเทียบกับ การแยกส่วนด้วยสายตา

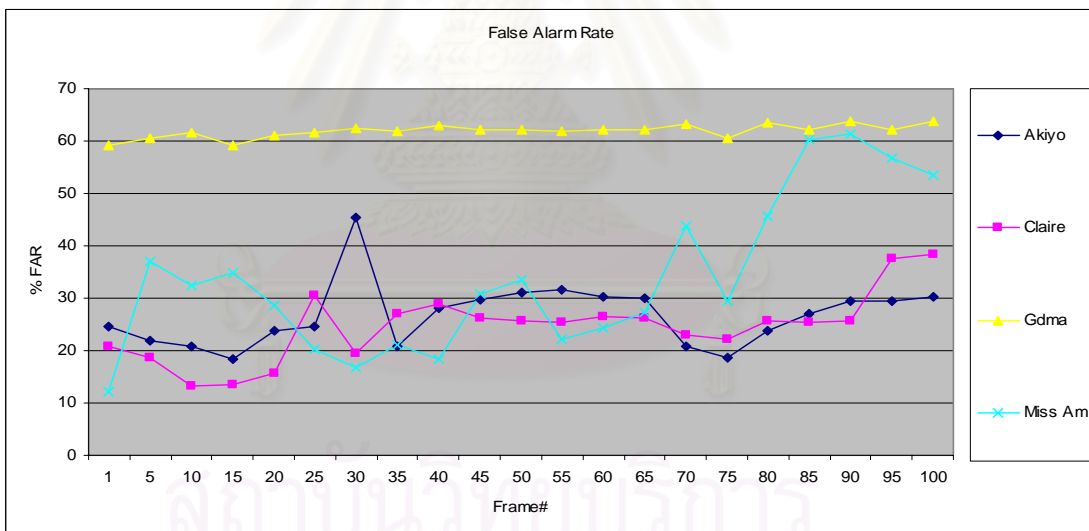


รูปที่ 4.22 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีที่นำเสนอเทียบกับ การแยกส่วนด้วยสายตา

จากรูปที่ 4.21 แสดงให้เห็นว่าอัตราความถูกต้องจะคงที่หลังจากเฟรมลำดับที่ 20 เป็นต้นไป เนื่องมาจากการใช้ความต่างเฟรมช่วยในการแยกส่วนภาพ ยกเว้นลำดับภาพ Miss Am ที่ไม่คงที่เนื่องมาจากวัตถุประสงค์การเคลื่อนที่เปลี่ยนตำแหน่งของใบหน้าไปจากตำแหน่งเดิมค่อนข้างมาก ทำให้การติดตามลำดับภาพทำงานได้ไม่ดีนัก ส่วนอัตราความผิดพลาดมีแนวโน้มสูงขึ้น เนื่องจากกระบวนการติดตามที่นำเอาการแยกส่วนเกินจากเฟรมก่อนหน้าเข้ามารวมด้วยแต่การทำเช่นนั้นก็มีข้อดี คือ ทำให้อัตราความถูกต้องเพิ่มขึ้น

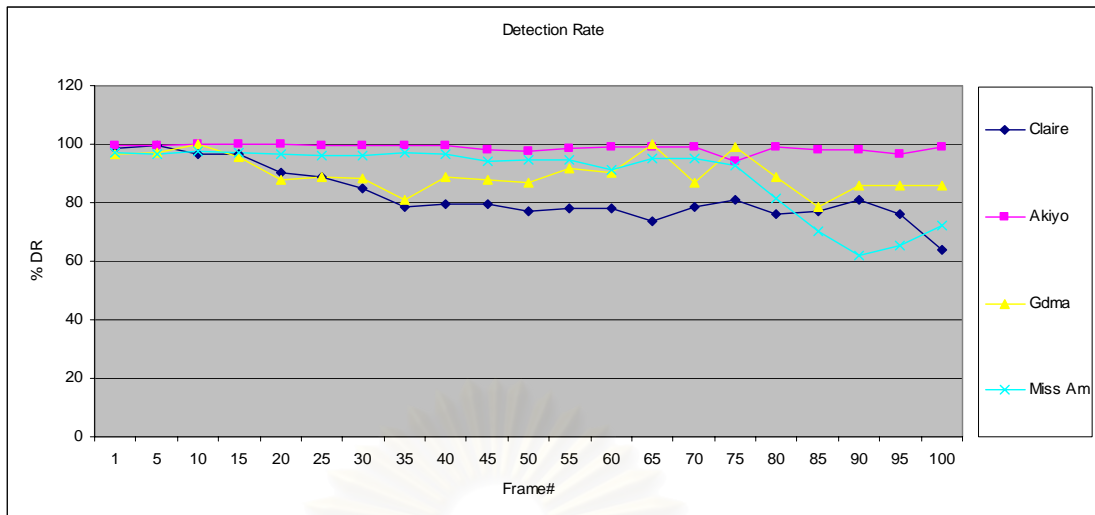


รูปที่ 4.23 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีที่ใช้องค์ประกอบสี่เหลี่ยม CbCr [8] เทียบกับการแยกส่วนด้วยสายตา

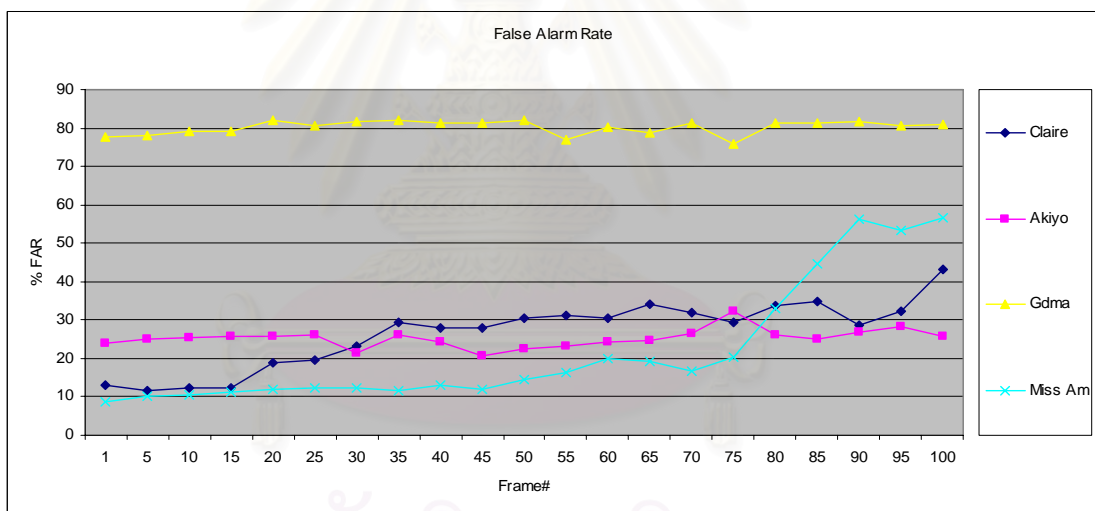


รูปที่ 4.24 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีที่ใช้องค์ประกอบสี่เหลี่ยม CbCr [8] เทียบกับการแยกส่วนด้วยสายตา

จากรูปที่ 4.23 และรูปที่ 4.24 แสดงให้เห็นว่าอัตราความถูกต้องค่อนข้างที่จะมีค่าสูง เนื่องจากการแยกส่วนมีการแยกส่วนเกินค่อนข้างมากซึ่งจะเห็นได้จากอัตราความผิดพลาดของทุกลำดับภาพที่มีค่าสูง โดยเฉพาะลำดับภาพ Grandmother เนื่องจากสีผิวของสีผมของลำดับภาพมีสีที่อยู่ในย่านเดียวกันกับย่านสีผิวซึ่งการใช้อัลกอริทึม CbCr ไม่สามารถแยกส่วนสีผิวจริงออกจากสีผิวปลอมที่เป็นสีของผมได้อย่างมีประสิทธิภาพ

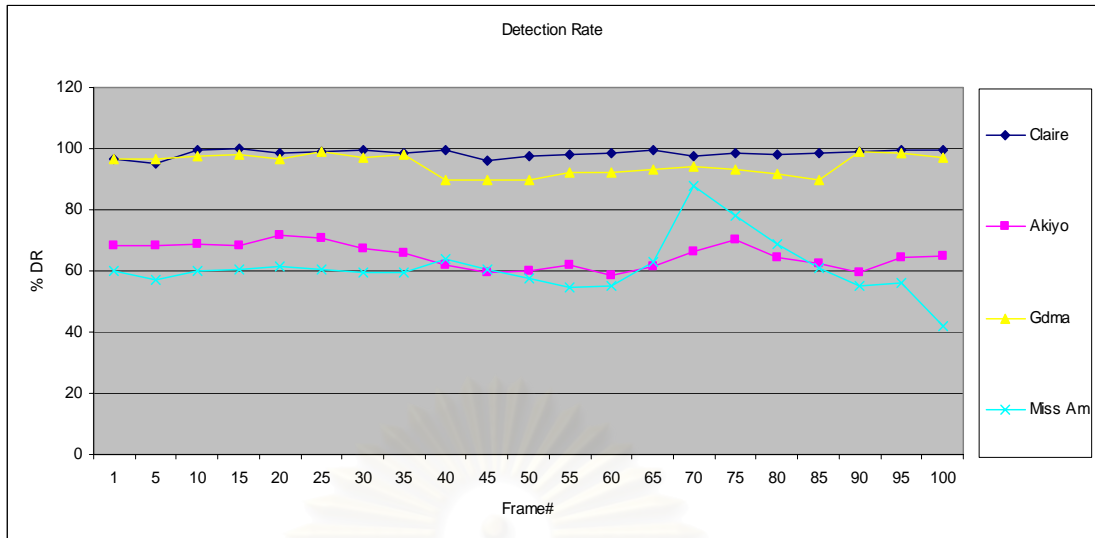


รูปที่ 4.25 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีกรรมวิธี [22] เทียบกับการแยกส่วนสี่ผิวด้วยสายตา

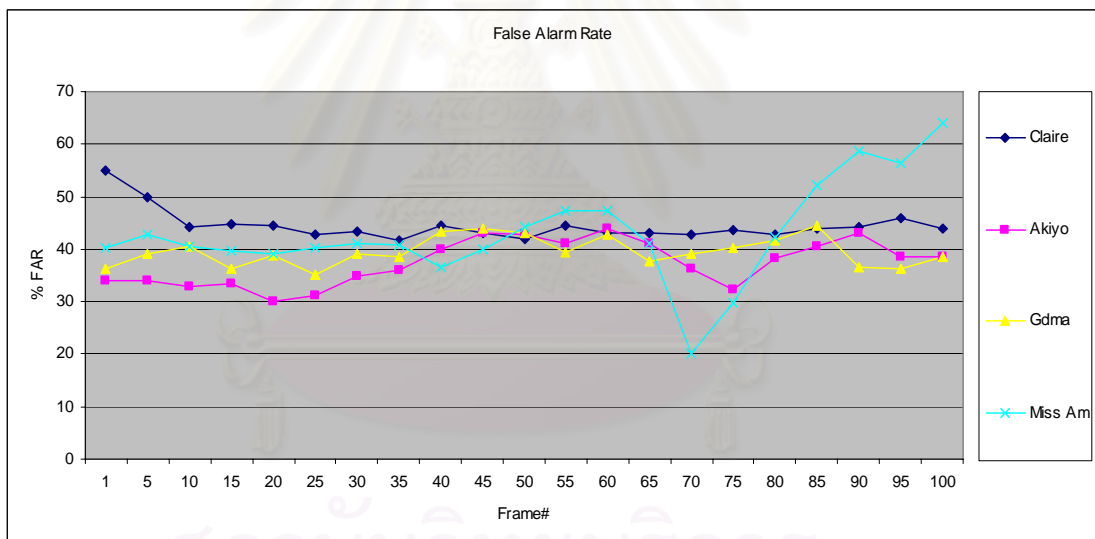


รูปที่ 4.26 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีกรรมวิธี [22] เทียบกับการแยกส่วนสี่ผิวด้วยสายตา

จากรูปที่ 4.25 และรูปที่ 4.26 แสดงให้เห็นว่าอัตราความถูกต้องมีค่าสูงแต่มีแนวโน้มลดลงในหลายลำดับภาพเนื่องจากการแยกส่วนมีการแยกส่วนขาดเพิ่มขึ้นเนื่องจากการใช้องค์ประกอบสี่ HS นั้น องค์ประกอบสี่ S นั้นมีผลกระทบกับการแยกส่วนค่อนข้างมากซึ่งเมื่อค่า S เปลี่ยนแปลงไปก็จะทำให้การแยกส่วนทำได้ไม่ดีนัก จึงทำให้อัตราความผิดพลาดของทุกลำดับภาพที่มีค่าสูง โดยเฉพาะลำดับภาพ Grandmother ที่พื้นหลังของลำดับภาพมีสี่ที่อยู่ในย่านเดียวกันกับย่านสี่ผิว ซึ่งการใช้องค์ประกอบสี่ HS ไม่สามารถแยกส่วนสี่ผิวจริงออกจากสี่ผิวปลอมที่เป็นพื้นหลังได้อย่างมีประสิทธิภาพ



รูปที่ 4.27 รูปกราฟแสดงอัตราความถูกต้องที่ได้จากการแยกส่วนของกรรมวิธีกรรมวิธี [10] เทียบกับการแยกส่วนสี่ผิวด้วยสายตา



รูปที่ 4.28 รูปกราฟแสดงอัตราความผิดพลาดที่ได้จากการแยกส่วนของกรรมวิธีกรรมวิธี [10] เทียบกับการแยกส่วนสี่ผิวด้วยสายตา

จากรูปที่ 4.27 และรูปที่ 4.28 แสดงให้เห็นว่าอัตราความถูกต้องมีแนวโน้มลดลงในลำดับ Miss am เนื่องจากการแยกส่วนมีการแยกส่วนขนาดเพิ่มขึ้นเนื่องจากการใช้องค์ประกอบสี Cr เพียงองค์ประกอบสีเดี่ยวนั้นทำให้การแยกส่วนทำได้ไม่ดีนัก จึงทำให้อัตราความผิดพลาดของทุกลำดับภาพที่มีค่าสูงอันเนื่องมาจากการแยกส่วนขนาดเป็นสาเหตุหลัก อีกทั้งการใช้ค่าความแปรปรวนในการขจัดสี่ผิวดล้อมยังทำให้การแยกส่วนขนาดมีค่าที่เพิ่มขึ้นเนื่องจากการขจัดสี่ผิวดจริงไปด้วยในบางลำดับภาพ

4.2 การแยกส่วนบริเวณที่สนใจที่เป็นวัตถุที่มีความหมาย

ในหัวข้อนี้เป็นผลการแยกส่วนวัตถุที่ได้จากกรรมวิธีที่นำเสนอ และเสนอการเปรียบเทียบเชิงอัตวิสัยกับกรรมวิธีการแยกส่วนเชิงพื้นที่และเวลาและกรรมวิธีการแยกส่วนโดยการทดสอบสมมติฐาน

4.2.1 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Claire

รูปที่ 4.29 แสดงผลจากการแยกส่วนภาพโดยกรรมวิธีที่นำเสนอ ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 1 เป็นต้นไป



รูปที่ 4.29 ผลการแยกส่วนภาพโดยกรรมวิธีที่นำเสนอ เฟรมที่ 1 ถึง 6



รูปที่ 4.30 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 6



รูปที่ 4.31 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 6

รูปที่ 4.30 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.31 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 3 ซึ่งใช้จำนวนเฟรมมากกว่าวิธีที่นำเสนอ

จากผลการทดลองจะเห็นว่า การแยกส่วนภาพด้วยกรรมวิธีที่นำเสนอสามารถแยกส่วนได้เต็มตัวโดยใช้จำนวนเฟรมที่น้อยที่สุด ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลาสามารถแยกส่วนได้เต็มตัวโดยใช้จำนวนเฟรมที่น้อยถัดมาจากกรรมวิธีที่นำเสนอ ส่วนการแยกส่วนภาพโดยการทดสอบสมมติฐานนั้นสามารถแยกได้เต็มตัวในเฟรมที่ 23

4.2.2 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Akiyo

รูปที่ 4.32 แสดงผลจากการแยกส่วนภาพโดยเทคนิคที่นำเสนอ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 1 เป็นต้นไป

รูปที่ 4.33 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัว ได้ตั้งแต่เฟรมที่ 7 เป็นต้นไป

รูปที่ 4.34 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้ครบเต็มตัวได้ตั้งแต่เฟรมที่ 1



รูปที่ 4.32 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอเฟรมที่ 1 ถึง 9



รูปที่ 4.33 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9



รูปที่ 4.34 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

4.2.3 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Grandmother

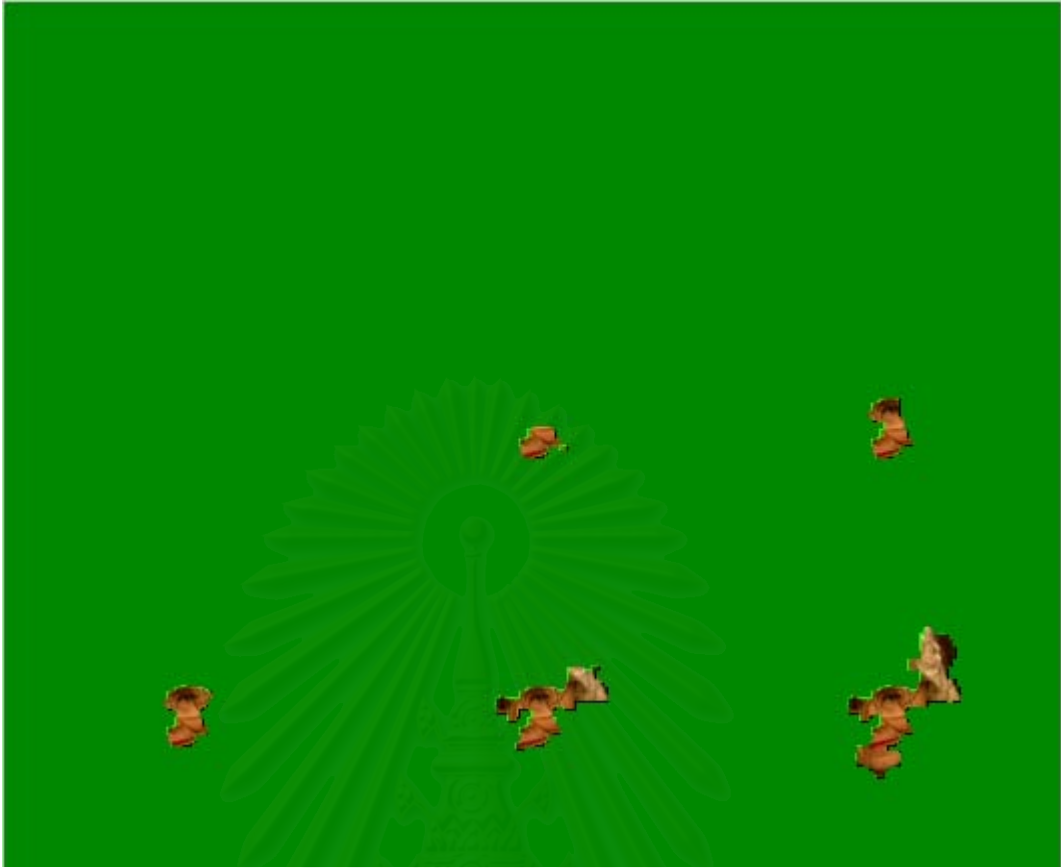
รูปที่ 4.35 แสดงผลจากการแยกส่วนภาพโดยเทคนิคที่นำเสนอ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 1 เป็นต้นไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

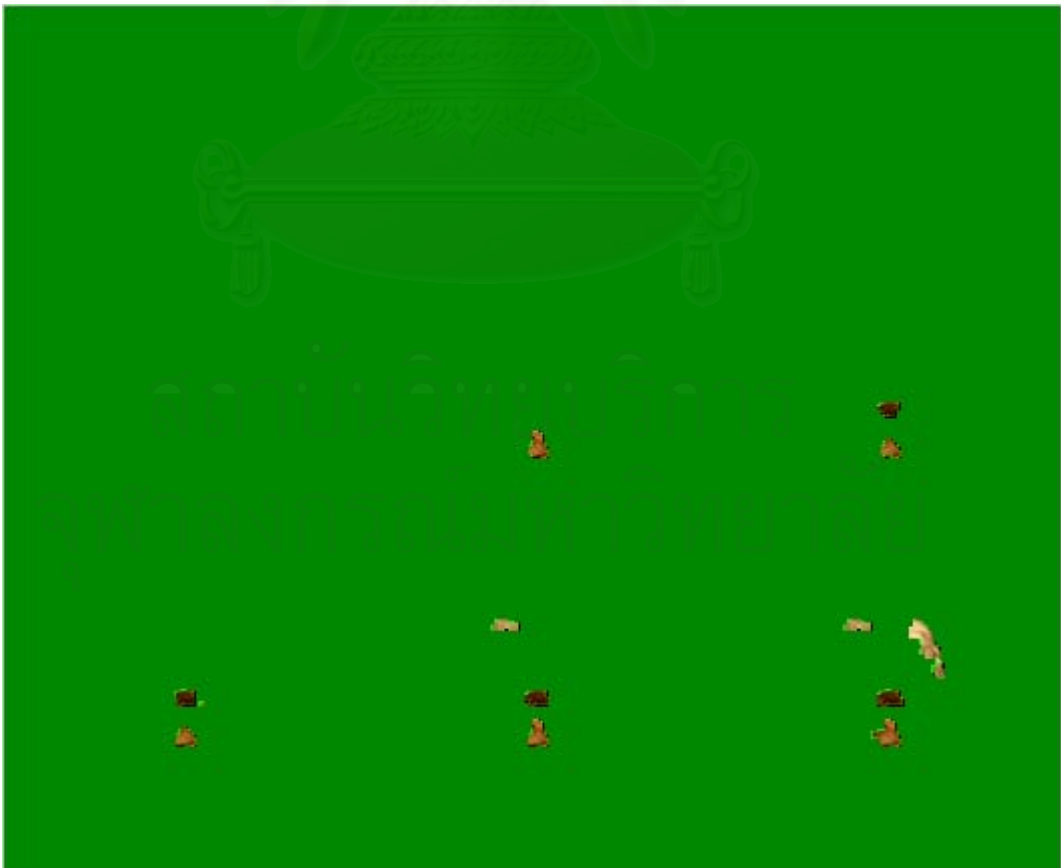


รูปที่ 4.35 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอ เฟรมที่ 1 ถึง 9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.36 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9



รูปที่ 4.37 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

รูปที่ 4.36 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว ในเฟรมแรก แต่สามารถแยกส่วนได้เต็มตัวในเฟรมที่ 35 เป็นต้นไป

รูปที่ 4.37 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว

จากผลการทดลองพบว่าการแยกส่วนภาพโดยเทคนิคที่นำเสนอสามารถแยกส่วนได้ครบเต็มตัวตั้งแต่เฟรมที่ 1 ซึ่งจากรูปที่ 4.35 จะเห็นว่ามีส่วนของเก้าอี้ติดมาด้วย เนื่องจาก เส้นผมและเก้าอี้มีสีที่คล้ายกัน ถึงแม้ว่าการเคลื่อนไหวของวัตถุมีน้อยมากโดยเฉพาะส่วนหัวไหล่และตัว แต่ที่ใช้จำนวนเฟรมเพียง 1 เฟรมในการแยกส่วนได้เต็มตัวเป็นเพราะว่ามีการกำหนดขอบเขตของวัตถุโดยใช้ตำแหน่งของใบหน้าเป็นเบื้องต้นและใช้ขอบเขตนั้นเป็นจุดเริ่มต้นของสเนกในการหดตัวเข้าหาขอบของวัตถุทำให้สามารถแยกวัตถุออกมาได้ตั้งแต่เฟรมแรก

การแยกส่วนโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลานั้น ในช่วง 39 เฟรมแรกไม่สามารถแยกส่วนได้ครบเต็มตัวเนื่องจากบริเวณหัวไหล่ยังไม่ค่อยเคลื่อนไหวมากนัก

การแยกส่วนโดยวิธีทดสอบสมมติฐานสามารถแยกส่วนออกมาได้เต็มตัวในเฟรมที่ 49 ซึ่งมีพื้นหลังติดมามากพอสมควรเพราะกระบวนการข้างหลังและกระบวนการติดตามวัตถุได้สะสมจำนวนเฟรมของวัตถุก่อนหน้านี้ไว้มากเกินไป ดังนั้นจึงเกิดการสะสมของจุดภาพที่เป็นพื้นหลังไปเรื่อย ๆ ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นไม่สามารถแยกส่วนออกมาได้ครบเต็มตัวเลย

4.2.4 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Miss America

รูปที่ 4.38 แสดงผลจากการแยกส่วนภาพโดยเทคนิคที่นำเสนอ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 1 เป็นต้นไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.38 ผลการแยกส่วนภาพโดยเทคนิคที่นำเสนอ เฟรมที่ 1 ถึง 9



รูปที่ 4.39 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9



รูปที่ 4.40 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

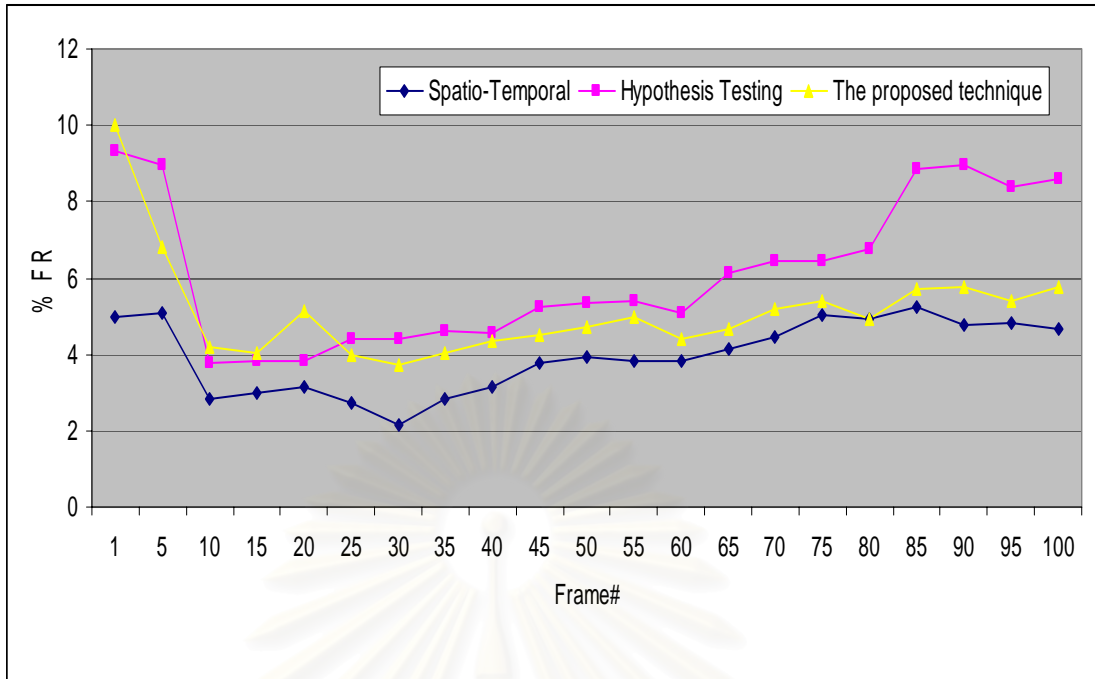
4.3 การวัดสมรรถนะของกรรมวิธีการแยกส่วนวัตถุ

4.3.1 การวัดผลการแยกส่วนภาพเชิงวัตถุ

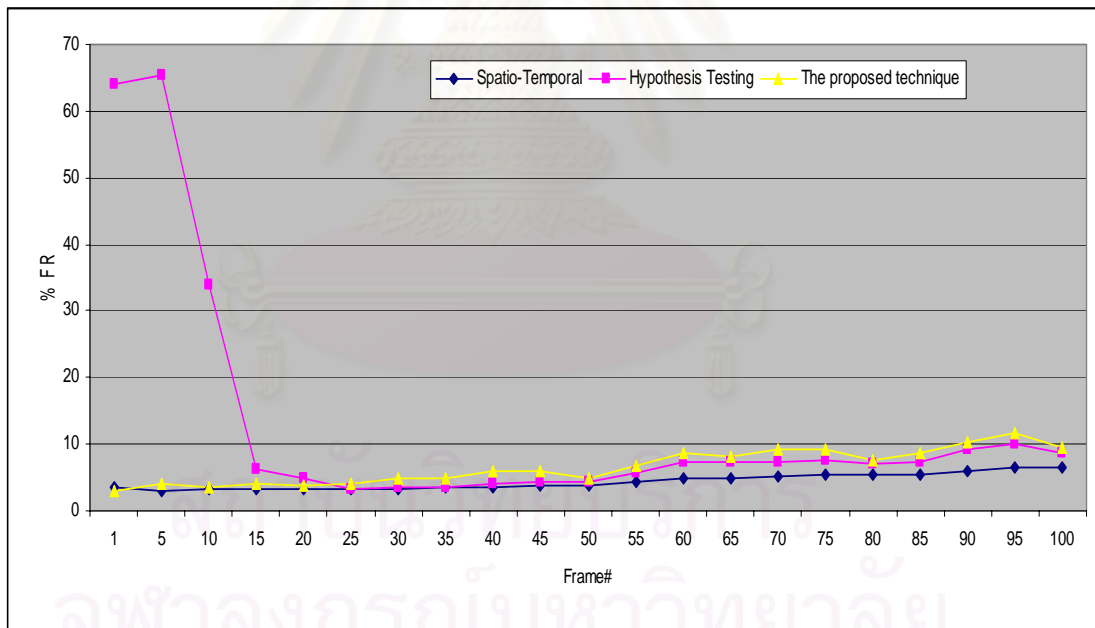
ความผิดพลาดของการแยกส่วนภาพวัตถุวัดจากจำนวนจุดภาพที่เกินและจำนวนจุดภาพที่ขาดไปของผลการแยกส่วนภาพวัตถุเมื่อเทียบกับการแยกส่วนด้วยสายตาซึ่งเป็นมาตรฐานอ้างอิงซึ่งนิยามได้ [44]

$$FR = (Now + Nbw) / N \quad (4.5)$$

โดย FR คือ อัตราความผิดพลาดต่อเฟรม Now คือ จำนวนจุดภาพที่เป็นวัตถุที่ถูกตัดสินว่าเป็นพื้นหลัง Nbw คือ จำนวนจุดภาพที่เป็นพื้นหลังแต่ถูกตัดสินว่าเป็นวัตถุ และ N คือ จำนวนจุดภาพของมาตรฐานอ้างอิง



รูปที่ 4.41 รูปกราฟแสดงค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Akiyo

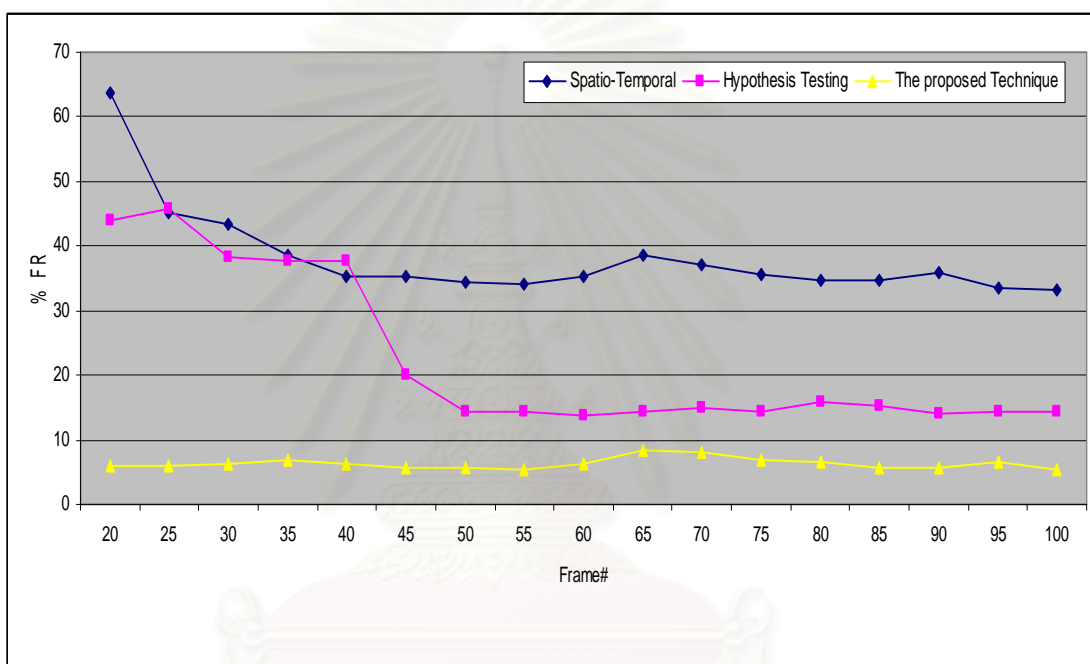


รูปที่ 4.42 รูปกราฟแสดงค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Claire

จากรูปที่ 4.41 แสดงให้เห็นว่าการแยกส่วนด้วยการทดสอบสมมติฐานมีอัตราผิดพลาดที่สูงกว่ากรรมวิธีอื่นเนื่องจากการแยกส่วนเกินที่เกิดจากการตัดสินใจผิดพลาด ส่วนกรรมวิธีที่นำเสนอมีอัตราผิดพลาดที่สูงกว่ากรรมวิธีเชิงพื้นที่และเวลาแต่น้อยกว่าการทดสอบสมมติฐานเนื่องจากลำดับ Akiyo มีการเคลื่อนที่อย่างต่อเนื่องตลอดลำดับภาพทำให้การแยกส่วนเชิงพื้นที่

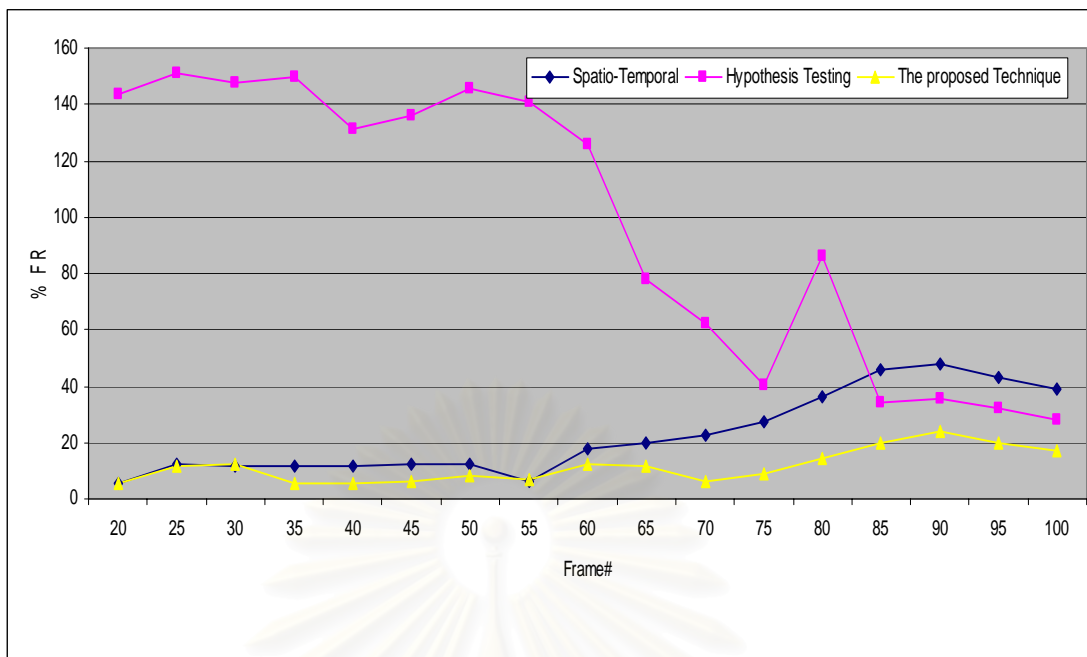
เวลาทำงานได้อย่างมีประสิทธิภาพ ส่วนกรรมวิธีที่น่าเสนอทำงานได้ดีเช่นเดียวกันเพียงแต่ว่าการใช้กรรมวิธีสเนกทำให้เกิดการแยกส่วนขาดไปเล็กน้อย

จากรูปที่ 4.42 แสดงให้เห็นว่าการแยกส่วนด้วยการทดสอบสมมติฐานมีอัตราผิดพลาดที่ค่อนข้างสูงในช่วงเฟรม 15 แรกเนื่องจากลำดับภาพไม่มีการเคลื่อนที่ตลอดทั้งตัวในช่วงเฟรมแรกทำให้กรรมวิธีนี้ไม่สามารถตัดสินใจได้ แต่หลังจากวัตถุมีการเคลื่อนที่มากพอจึงทำให้กรรมวิธีนี้สามารถทำงานได้อย่างมีประสิทธิภาพ เนื่องจากลำดับภาพมีการเคลื่อนที่ค่อนข้างคงที่ตลอดลำดับภาพทำให้อัตราผิดพลาดของแต่ละกรรมวิธีไม่แตกต่างกันมากนักในเฟรมที่ 20 เป็นต้นไป



รูปที่ 4.43 รูปกราฟแสดงค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Grandmother

จากรูปที่ 4.43 แสดงให้เห็นว่าการแยกส่วนด้วยการทดสอบเชิงพื้นที่เวลามีอัตราผิดพลาดที่ค่อนข้างสูงในทุกเฟรม เนื่องจากลำดับภาพไม่มีการเคลื่อนที่ตลอดทั้งตัวในช่วงเฟรมแรกและมีการเคลื่อนที่น้อยมาทำให้กรรมวิธีนี้ไม่สามารถแยกส่วนได้อย่างมีประสิทธิภาพ ส่วนกรรมวิธีการทดสอบสมมติฐานสามารถแยกส่วนได้ในลำดับภาพที่ 35 แต่อย่างไรก็ตามอัตราความผิดพลาดยังคงสูงกว่ากรรมวิธีที่น่าเสนอ เนื่องจากลำดับภาพมีการเคลื่อนที่ค่อนข้างน้อยคงที่ตลอดลำดับภาพดังนั้นการใช้กรรมวิธีสเนกทำให้สามารถแยกส่วนได้ในเฟรมแรกทำให้อัตราผิดพลาดของกรรมวิธีที่น่าเสนอมีค่าน้อยและค่าผิดพลาดเกิดมาจากการแยกส่วนเกินที่ไม่สามารถใช้ในการเคลื่อนที่ช่วยในการแยกส่วนได้ ดังนั้นกรรมวิธีที่น่าเสนอจึงทำงานได้อย่างมีประสิทธิภาพสูงในลำดับภาพนี้



รูปที่ 4.44 รูปกราฟแสดงค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Miss Am

จากรูปที่ 4.44 แสดงให้เห็นว่าการแยกส่วนด้วยการทดสอบเชิงพื้นที่เวลามีอัตราผิดพลาดที่สูงกว่ากรรมวิธีที่นำเสนอเนื่องจากวัตถุประสงค์การเคลื่อนที่อย่างไม่ต่อเนื่องในทุกเฟรม ทำให้เกิดการแยกส่วนที่ผิดพลาดเนื่องมาจากการแยกส่วนขาดอีกทั้งลำดับภาพไม่มีการเคลื่อนที่ตลอดทั้งตัวในช่วงเฟรมแรกทำให้การแยกส่วนบริเวณลำตัวของวัตถุมีความผิดพลาดเกิดขึ้น ส่วนกรรมวิธีการทดสอบสมมติฐานไม่สามารถแยกส่วนได้ในลำดับภาพต้นๆ เนื่องจากไม่มีการเคลื่อนที่ตลอดทั้งตัววัตถุดังที่กล่าวมาแล้วทำให้กรรมวิธีนี้ทำงานได้อย่างไม่มีประสิทธิภาพมากนัก แต่อัตราความผิดพลาดมีแนวโน้มลดลงและดีกว่ากรรมวิธีเชิงพื้นที่เวลาเมื่อกรรมวิธีก็สามารถใช้กรรมวิธีการตัดสินใจได้ในลำดับภาพที่ 80 เป็นต้นไป แต่อย่างไรก็ตามอัตราความผิดพลาดยังคงสูงกว่ากรรมวิธีที่นำเสนอ เนื่องจากการใช้กรรมวิธีสเนกทำให้กรรมวิธีนี้สามารถแยกส่วนได้ในเฟรมแรกทำให้อัตราผิดพลาดของกรรมวิธีที่นำเสนอมีค่าน้อยกว่ากรรมวิธีอื่น ดังนั้นกรรมวิธีที่นำเสนอจึงทำงานได้อย่างมีประสิทธิภาพสูงในลำดับภาพนี้เช่นเดียวกัน

4.4 การวิเคราะห์ความซับซ้อนของการแยกส่วนภาพทั้งสามวิธี

ในการแยกส่วนภาพแต่ละวิธีมีกระบวนการหลายกระบวนการที่แตกต่างกันและมีบางกระบวนการที่เหมือนกันซึ่งในการวิเคราะห์ความซับซ้อนนี้จะเป็นการชี้ให้เห็นว่าแต่ละกระบวนการและแต่ละวิธีนั้นมีความซับซ้อนเป็นอย่างไรเมื่อนำมาเปรียบเทียบกัน

จากกรรมวิธีการแยกส่วนภาพทั้ง 3 วิธีนั้นสามารถเปรียบเทียบเป็นกระบวนการได้ดังตารางที่ 4.3 ซึ่งจะเห็นว่าการแยกส่วนภาพเชิงพื้นที่และเวลาที่มีความซับซ้อนน้อยที่สุดเพราะมีกระบวนการน้อยที่สุด

การแยกส่วนโดยเทคนิคที่นำเสนอซับซ้อนกว่าการแยกส่วนภาพเชิงพื้นที่และเวลาเพราะมีการใช้กรรมวิธีสเนกเข้าช่วยในการหาขอบเขตของวัตถุที่สนใจ

การแยกส่วนโดยวิธีการทดสอบสมมติฐานนั้นซับซ้อนที่สุดเพราะจะมีการประมาณการเคลื่อนที่แบบโกลบอลและมีการค้นหาแบบเต็ม ทั้งนี้ในการแยกส่วนภาพเชิงเวลายังใช้การทดสอบสมมติฐานซึ่งมีความซับซ้อนมากกว่าการนำมาสกรีนตรวจสอบการเคลื่อนไหวมาก

ตารางที่ 4.4 แสดงจำนวนเฟรมที่ใช้การแยกส่วนได้เต็มตัวเป็นครั้งแรกโดยเปรียบเทียบกันทั้ง 3 วิธี ซึ่งจะเห็นว่า การแยกส่วนโดยเทคนิคที่นำเสนอใช้จำนวนเฟรมน้อยที่สุด

ตารางที่ 4.5 แสดงเวลาที่ใช้ในการประมวลผลในแต่ละวิธีโดยเฉลี่ยจาก 100 ลำดับภาพ โดยเปรียบเทียบกัน 3 วิธี ซึ่งประมวลผลด้วยเครื่องคอมพิวเตอร์เพนเทียม 4 ความเร็ว 2.4 GHz หน่วยความจำ 256 เมกกะไบต์ และเขียนโปรแกรมด้วย Matlab เวอร์ชัน 6.5 พบว่ากรรมวิธีที่นำเสนอพบว่าการแยกส่วนด้วยวิธีการทดสอบสมมติฐานใช้เวลาเฉลี่ยมากที่สุด ส่วนการแยกส่วนเชิงพื้นที่และเวลาใช้เวลาเฉลี่ยน้อยที่สุด ส่วนกรรมวิธีที่นำเสนอนั้นใช้เวลาเฉลี่ยอยู่ระหว่าง 2 กรรมวิธีข้างต้น อย่างไรก็ตามกรรมวิธีที่นำเสนอมถึงแม้ว่าจะใช้เวลาในการประมวลผลมากกว่ากรรมวิธีการแยกส่วนเชิงพื้นที่และเวลาแต่ก็สามารถแยกส่วนภาพวัตถุได้เต็มตัวด้วยจำนวนเฟรมที่น้อยกว่าดังตารางที่ 4.4

ตารางที่ 4.3 การเปรียบเทียบกระบวนการแยกส่วนภาพของทั้ง 3 กรรมวิธี

การแยกส่วนภาพโดยวิธี	การทดสอบสมมติฐาน	กรรมวิธีที่นำเสนอ	เชิงเวลาและพื้นที่
การประมาณการเคลื่อนที่	แบบโกลบอล และค้นหาแบบเต็ม	ไม่มี	ไม่มี
การแยกส่วนภาพเชิงพื้นที่	ตัวดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ	กรรมวิธีเสนก, ตัว ดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ	ตัวดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ
การแยกส่วนภาพเชิงเวลา	การทดสอบสมมติฐาน	มาส์กการเคลื่อนไหว	มาส์กการเคลื่อนไหว
การตัดลินใจ	ใช้เงื่อนไข	ใช้เงื่อนไข	ใช้เงื่อนไข
การติดตามวัตถุ	มี	มี	มี

ตารางที่ 4.4 เปรียบเทียบจำนวนเฟรมที่ใช้ในการแยกส่วนได้เต็มตัว

ลำดับภาพ	เฟรมที่สามารถแยกส่วนได้เต็มตัวเป็นเฟรมแรกโดยวิธี		
	กรรมวิธีที่นำเสนอ	การทดสอบ สมมติฐาน	เชิงพื้นที่ และเวลา
Claire	1	23	3
Akiyo	1	7	1
Grandmother	1	49	866
Miss Am	1	83	12

ตารางที่ 4.5 เวลาที่ใช้ในการประมวลผลในแต่ละวิธีโดยเฉลี่ยจาก 100 ลำดับภาพ

ลำดับภาพ	เวลาที่ใช้ในการประมวลผลในแต่ละเฟรม(วินาทีต่อเฟรม)		
	กรรมวิธีที่นำเสนอ	การทดสอบ สมมติฐาน	เชิงพื้นที่ และเวลา
Claire	7.25	13.35	1.82
Akiyo	6.13	13.80	1.99
Grandmother	6.30	13.34	2.00
Miss Am	6.40	14.03	2.03



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปและข้อเสนอแนะ

5.1 สรุป

วิทยานิพนธ์นี้ เสนอการแยกส่วนบริเวณที่สนใจบริเวณใบหน้าและบริเวณที่สนใจที่เป็นวัตถุ โดยแบ่งออกเป็น 2 ส่วนหลัก ดังนี้

การแยกส่วนบริเวณที่สนใจบริเวณใบหน้า

ในส่วนนี้วิทยานิพนธ์นี้สรุปความรู้พื้นฐานและกระบวนการของการแยกส่วนบริเวณที่สนใจ และได้เสนอกรรมวิธีการแยกส่วนบริเวณที่สนใจบริเวณใบหน้าที่ใช้องค์ประกอบสีในการแยกส่วนโดยเสนอกรรมวิธีการใช้องค์ประกอบสีร่วมกันระหว่าง 2 ปริภูมิสี อีกทั้งยังได้เสนอการขจัดพื้นหลังที่มีลักษณะคล้ายสีผิวโดยใช้คุณสมบัติของความจริงที่ว่าสีผิวของมนุษย์นั้นมีค่าความเข้มสีแดงมากและมีค่าความเข้มสีของสีน้ำเงินน้อย ทำให้กำหนดช่วงค่าของสีสองสีนี้เพื่อขจัดสีผิวเทียมในลำดับภาพ ทำยสุดกรรมวิธีได้ใช้การติดตามวัตถุเพื่อขจัดจุดภาพที่มีลักษณะสีผิวที่ปรากฏขึ้นมาอย่างทันทีทันใด ผลการแยกส่วนที่ได้จากกรรมวิธีที่นำเสนอให้ผลลัพธ์ที่มีอัตราความถูกต้องที่ใกล้เคียงกับกรรมวิธีอื่นๆ แต่มีอัตราความผิดพลาดที่น้อยกว่ากรรมวิธีที่ใช้ปริภูมิสีอื่นๆ

การเปรียบเทียบผลการทดลองการแยกส่วนบริเวณที่สนใจบริเวณใบหน้ากับกรรมวิธีอื่น ๆ

ผลการทดลองของกรรมวิธีการใช้องค์ประกอบสี CbCr แสดงให้เห็นว่า สามารถให้ผลการแยกส่วนในระดับดี แต่ลำดับภาพ Grandmother ไม่สามารถแยกส่วนสีผมออกไปจากบริเวณใบหน้าได้เนื่องจากช่วงค่าขององค์ประกอบสี CbCr อยู่ในช่วงค่าของสีผม และกรรมวิธีนี้ไม่มีการติดตามวัตถุทำให้เมื่อบริเวณใบหน้ามีการเคลื่อนที่ไปจากตำแหน่งเดิมไม่มากนักทำให้ต้องทำกระบวนการซ้ำเช่นเดิม ทำให้ไม่มีการนำผลการแยกส่วนเฟรมก่อนหน้ามาช่วยในการแยกส่วนทำให้บางเฟรมมีการแยกส่วนขาดเกิดขึ้นเนื่องจากเงาบริเวณลำคอที่เกิดขึ้น ทำให้อัตราความถูกต้องลดลงเนื่องจากการแยกส่วนขาดที่เกิดขึ้น

ผลการทดลองของกรรมวิธีการใช้องค์ประกอบสี HS แสดงให้เห็นว่า สามารถให้ผลการแยกส่วนได้ในระดับดีพอใช้ แต่ลำดับภาพ Grandmother ไม่สามารถแยกส่วนสีพื้นหลังออกไปได้เนื่องจากช่วงค่าขององค์ประกอบสี HS อยู่ในช่วงค่าของสีพื้นหลังเช่นเดียวกับกรรมวิธีการใช้องค์ประกอบสี CbCr เพียงแต่ว่ากรรมวิธีนี้ไม่มีกระบวนการขจัดพื้นหลังเช่นเดียวกับวิธีก่อนหน้าจึงทำให้ไม่สามารถขจัดพื้นหลังที่แยกส่วนผิดพลาดนี้ได้ ทำให้การแยกส่วนเกิดขึ้นเป็นบริเวณกว้างมากจนทำให้อัตราความผิดพลาดมีค่าสูงมาก ส่วนลำดับภาพอื่นๆ กรรมวิธีนี้มีอัตราความถูกต้องที่ใกล้เคียงกับวิธีการอื่นๆ

ผลการทดลองของกรรมวิธีการใช้องค์ประกอบสี Cr แสดงให้เห็นว่า สามารถให้ผลการแยกส่วนได้ในระดับพอใช้ เนื่องจากผลการแยกส่วนแสดงให้เห็นว่ากรรมวิธีนี้มีการแยกส่วนขาดเป็นส่วนใหญ่ทำให้อัตราความถูกต้องมีค่าน้อยกว่ากรรมวิธีอื่น เพราะการใช้องค์ประกอบสี Cr เพียงองค์ประกอบเดียวทำให้ช่วงค่าของสีผิวที่แยกส่วนได้ไม่ครอบคลุมช่วงค่าสีผิวอย่างเพียงพอ และการใช้กรรมวิธีการจัดสีผิวปลอมโดยใช้ค่าความแปรปรวนขององค์ประกอบสี RGB ก่อให้เกิดการแยกส่วนขาดที่เพิ่มขึ้น แต่อย่างไรก็ตามกรรมวิธีนี้มีความซับซ้อนน้อยที่สุดเพราะมีเพียง 2 ขั้นตอน คือ การแยกส่วนสีผิวด้วยองค์ประกอบสี Cr และการจัดพื้นหลังด้วยค่าความแปรปรวน

ผลการทดลองของกรรมวิธีที่นำเสนอที่มีการใช้องค์ประกอบสี HCr แสดงให้เห็นว่า สามารถให้ผลการแยกส่วนได้ในระดับดี เนื่องจากผลการแยกส่วนแสดงให้เห็นว่ากรรมวิธีนี้มีการแยกส่วนเกินเป็นส่วนใหญ่ทำให้อัตราความถูกต้องมีค่าที่ใกล้เคียงกับกรรมวิธี CbCr และ HS แต่มีอัตราความผิดพลาดที่น้อยกว่าเนื่องจากการตัดองค์ประกอบ S ในปริภูมิสี HSI และ Cb ในปริภูมิสี YCbCr ออก เนื่องจากองค์ประกอบสีทั้งสองนั้นก่อให้เกิดอัตราความผิดพลาด ดังนั้นกรรมวิธีนี้จึงมีค่าอัตราความผิดพลาดน้อยกว่ากรรมวิธีอื่น อีกทั้งการใช้กรรมวิธีการจัดสีผิวปลอมด้วยการใช้ช่วงค่าของสีแดงและสีน้ำเงินในการตรวจสอบว่าจุดภาพใดเป็นสีผิวจริงหรือสีผิวปลอมทำให้สามารถจัดสีผิวในลำดับภาพได้ ทำให้อัตราความถูกต้องมีค่าสูงขึ้น อีกทั้งการใช้องค์ประกอบสี Cr ร่วมกับองค์ประกอบสี H ทำให้สามารถชดเชยข้อบกพร่องของการใช้กรรมวิธีองค์ประกอบสี Cr เพียงองค์ประกอบเดียวไปได้เพราะทำให้ช่วงค่าของสีผิวครอบคลุมช่วงค่าสีผิวอย่างเพียงพอ

ดังนั้น กรรมวิธีที่นำเสนอจึงเหมาะสมกับความต้องการผลการแยกส่วนลำดับภาพที่ต้องการอัตราความผิดพลาดที่มีค่าน้อยและอัตราความถูกต้องที่อยู่ในเกณฑ์ที่เหมาะสม

การแยกส่วนบริเวณที่สนใจที่เป็นวัตถุ

ในส่วนนี้วิทยานิพนธ์นี้สรุปความรู้พื้นฐานและกระบวนการของการแยกส่วนบริเวณที่สนใจ และได้เสนอกรรมวิธีการแยกส่วนบริเวณที่เป็นภาพวัตถุสำหรับสัญญาณวิดีโอ กรรมวิธีการแยกส่วนบริเวณที่สนใจของสัญญาณวิดีโอที่นำเสนอในวิทยานิพนธ์ฉบับนี้ อยู่บนพื้นฐานของการใช้คุณลักษณะสำคัญของสัญญาณวิดีโอ ทั้งข้อมูลเชิงพื้นที่ ข้อมูลจากตำแหน่งของใบหน้าของวัตถุและข้อมูลเชิงเวลา

โดยที่ข้อมูลเชิงพื้นที่ได้มาจากข้อมูลของความเข้มแสง ซึ่งภาพในแต่ละเฟรมจะถูกแยกส่วนภาพเชิงพื้นที่เป็นบริเวณที่มีความเข้มต่อเนื่องกัน โดยการใช้ข้อมูลจากตำแหน่งของใบหน้ามากำหนดขอบเขตของวัตถุและใช้กรรมวิธีสเนกในการลดพื้นที่การแยกส่วนลง โดยให้สเนกหดตัวเข้าหาขอบของวัตถุ แล้วหลังจากนั้นจึงใช้ข้อมูลการแยกส่วนเชิงพื้นที่ที่ได้จากการปฏิบัติการเชิงสัณฐาน (morphological operation) การตรวจรู้ขอบเขตของบริเวณในภาพโดยใช้การประมาณเกรเดียนต์และการกำหนดบริเวณให้แก่จุดภาพที่มีความเข้มใกล้เคียงกันโดยกรรมวิธี

สันปันน้ำ (watershed algorithm) แล้วจึงใช้ข้อมูลการแยกส่วนเชิงเวลาเพื่อช่วยในการจัดการแยกส่วนเกินที่เกิดขึ้นจากการแยกส่วนเชิงพื้นที่ โดยที่ข้อมูลเชิงเวลานั้นได้จากการตรวจรู้การเปลี่ยนแปลงโดยการพิจารณาการเปลี่ยนแปลงด้วยค่าจุดเปลี่ยนซึ่งวิเคราะห์จากค่าการเปลี่ยนแปลงเฉลี่ยของเฟรมที่ต่อเนื่องกัน และนำเอาข้อมูลจากเฟรมก่อนหน้ามาพิจารณาร่วมกับข้อมูลที่ได้จากการแยกส่วนภาพในเฟรมปัจจุบันเพื่อช่วยในการติดตามการเคลื่อนที่ของวัตถุ หลังจากนั้นจึงนำข้อมูลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อให้สามารถแยกส่วนภาพวัตถุได้ดียิ่งขึ้น

การเปรียบเทียบผลการทดลองการแยกส่วนบริเวณที่สนใจที่เป็นวัตถุกับกรรมวิธีอื่น ๆ

ผลการทดลองของกรรมวิธีการแยกส่วนเชิงพื้นที่และเวลา แสดงให้เห็นว่า ผลลัพธ์จากการแยกส่วนพบว่ากรรมวิธีนี้ ให้ผลการแยกส่วนได้ดีกับลำดับภาพที่มีการเคลื่อนที่อย่างต่อเนื่องตลอดลำดับภาพ แต่จะให้ผลการแยกส่วนไม่ดีกับลำดับภาพที่มีการเคลื่อนที่ช้าเนื่องจากไม่สามารถตรวจจับการเคลื่อนที่ได้อย่างมีประสิทธิภาพ ทำให้อัตราความผิดพลาดมีค่าค่อนข้างสูงในลำดับภาพประเภทนี้ แต่กรรมวิธีนี้มีความซับซ้อนน้อยกว่ากรรมวิธีอื่น ๆ

ผลการทดลองของกรรมวิธีการทดสอบสมมติฐาน แสดงให้เห็นว่า ผลลัพธ์จากการแยกส่วนพบว่ากรรมวิธีนี้ จะให้ผลการแยกส่วนได้ดีกับวัตถุที่มีการเคลื่อนที่ซึ่งใช้จำนวนเฟรมในการตรวจจับการเคลื่อนที่มากกว่ากรรมวิธีการแยกส่วนเชิงพื้นที่และเวลา ทำให้สามารถแยกส่วนเต็มตัวได้ช้ากว่าแต่อัตราความผิดพลาดจะมีค่าลดลงตามลำดับเมื่อกรรมวิธีสามารถใช้กระบวนการตัดสินใจได้ แต่อย่างไรก็ตามกับลำดับภาพที่มีการเคลื่อนที่อย่างช้าต่อเนื่องตลอดลำดับภาพกรรมวิธีนี้จะให้ผลการแยกส่วนได้ไม่ดีนักเนื่องจากไม่สามารถใช้กระบวนการตัดสินใจได้เนื่องจากไม่สามารถตรวจจับการเคลื่อนที่ของวัตถุได้อย่างชัดเจน ทำให้อัตราความผิดพลาดมีค่าค่อนข้างสูง อีกทั้งกรรมวิธีนี้มีความซับซ้อนสูงกว่ากรรมวิธีอื่น ๆ

ผลการทดลองของกรรมวิธีที่นำเสนอ แสดงให้เห็นว่า ผลลัพธ์จากการแยกส่วนพบว่ากรรมวิธีนี้ จะให้ผลการแยกส่วนได้ดีกว่ากรรมวิธีอื่น ๆ ในลำดับภาพที่มีการเคลื่อนที่น้อยเนื่องจากมีการกำหนดขอบเขตของวัตถุโดยการใช้กรรมวิธีเสนก ทำให้กรรมวิธีนี้สามารถแยกส่วนลำดับภาพต่างๆ ได้ตั้งแต่ลำดับเฟรมต้นๆ ในขณะที่กรรมวิธีอื่นต้องใช้จำนวนเฟรมมากกว่าในการแยกส่วนได้เต็มตัววัตถุ ทำให้อัตราความผิดพลาดมีค่าที่น้อยกว่ากรรมวิธีอื่น ๆ ในลำดับภาพที่มีการเคลื่อนที่น้อย ส่วนในลำดับภาพที่มีการเคลื่อนที่อย่างต่อเนื่องตลอดลำดับภาพอัตราความผิดพลาดจะมีค่าที่ใกล้เคียงกับกรรมวิธีการแยกส่วนเชิงพื้นที่และเวลา โดยที่กรรมวิธีที่นำเสนอมีความซับซ้อนมากกว่ากรรมวิธีการแยกส่วนเชิงพื้นที่และเวลาแต่น้อยกว่ากรรมวิธีการทดสอบสมมติฐาน

ดังนั้น กรรมวิธีที่นำเสนอจึงเหมาะสมกับทั้งลำดับภาพที่มีการเคลื่อนที่ตลอดลำดับภาพ และลำดับภาพที่มีการเคลื่อนที่น้อย แต่กรรมวิธีนี้จะให้ผลการแยกส่วนที่ดีกว่าวิธีการอื่นๆ ในลำดับภาพที่มีการเคลื่อนที่น้อย โดยกรรมวิธีที่นำเสนอมีอัตราความผิดพลาดอยู่ในเกณฑ์ที่เหมาะสม

5.2 ข้อเสนอแนะ

ทดสอบนำกรรมวิธีข้างต้นมาใช้ร่วมกับส่วนประกอบอื่นๆ สำหรับการประยุกต์ใช้งานจริง ทางด้านการประมวลผลสัญญาณวิดีโอ

สร้างกรรมวิธีเพิ่มเติมเพื่อใช้กับลำดับภาพอื่นที่ไม่มีใบหน้าอยู่

สร้างขั้นตอนการตรวจจับอัตราเคลื่อนที่เพิ่มขึ้นเพื่อเป็นทางเลือกกว่าจะใช้หรือไม่ใช้ กรรมวิธีสเนกเพื่อลดความซับซ้อนของระบบ



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

1. Coding of Moving Pictures and Audio. ISO/IEC/ JTC1/SC29/WG11, 1998.
2. Video Coding for Low Bitrate Communication. ITU-T Recommendation. H.263 1996.
3. Changkck Kim and Jeng-Neng Hwang. Fast and Automatic Video Object Segmentation and Tracking for Content-Based Applications. IEEE Transactions on Circuits and Systems for Video Technology 12 (Feburary 2002): 122-129
4. Datchakorn Tanchaoren; S. Jitapankul; P. Kittipanya-ngam; S. Chompon; and H. Kortrakulkij. Semantic Object Segmentation for Content Based Video Coding. Proceedings Vol XIII, Image, Acoustic, Speech and Signal Processing: Part II, World Multiconference on Systemics, Cybernetics, and Informics (July 2001): 579-583.
5. Nugroho H.; Takahashi S. K.; Ooi Y. and Ozawa S. Detecting Human face from Monocular Image Sequences by Genetic Algorithms. IEEE International Conference on Acoustics Speech and Signal Processing 4 (1997): 2533-2536.
6. Hartung, J.; and Jacquin, A. Object-Oriented H.263 Compatible Video Coding Platform for Conferencing Applications. IEEE Journal on Selected Areas in Communications 16, 1 (January 1998): 42-55.
7. Jae-Beom Lee ; and Eleftheriadis, A. Spatio-temporal-assisted compatible coding for low and very low bitrate videotelephony IEEE Proceedings 1 (1996): 429-432.
8. Chai, D.; and Ngan, K. N. Face Segmentation Using Skin-Color Map in Videophone Applications. IEEE Transactions on Circuits and Systems for Video Technology 9, 4 (June 1999): 551-564.
9. Lin, C.W.; Chang, Y.-J.; and Chen, Y.C. Low-complexity face-assisted video coding. Proc. Int. Conf. Image Proc (September 2000): 207-210.

10. Mei-Juan Chen; Ming-Chieh Chi; Ching-Ting Hsu; and Jeng-Wei Chen. ROI Video coding based on H.263+ with robust skin-color detection technique. IEEE Trans. On Consumer Electronics 49 (2003): 724-730.
11. Ming-Chieh Chi; Mei-Juan Chen; and Ching-Ting Hsu. Region-of-interest video coding by fuzzy for H.263+ standard. IEEE on Circuits and Systems for Video Technology 49 (2004): 93-96.
12. Kiran Challapali; Tomas Brodsky; Yun-Ting Lin; Yong Yan; and Richard Yi Chen. Real-time Object Segmentation and Coding for Selective-Quality Video Communications. IEEE Transactions on Circuits and Systems for Video Technology. 14, 6 (June 2004): 813-824.
13. Tanchaoen, D.; Kortrakulkij, H.; Khemachai, S.; Aramvith, S.; Jitapunkul, S. Automatic face color segmentation based rate control for low bit-rate video coding. Proceedings of the 2003 International Symposium on Circuits and Systems 2 (May 2003): 384 – 387.
14. L. L. Yang; M. A. Robertson. Multiple-Face Tracking System For General Region-Of-Interest Video Coding Proceedings of the 2000 International Conference on Image Processing 1 (September 2000): 347 – 350.
15. Erik Hjelmas; and Boon Kee Low. Face Detection : A Survey. Computer Vision and Image Understanding 83 (2001): 236-274.
16. Shi Kewei; Fu Xitian; Cai Anni; and Sun Jingao. Automatic Face Segmentation in YCrCb Images. APCC/OECC Fifth Asia-Pacific Conference on Communications, and Fourth Optoelectronics and Communications Conference 2 (October 1999): 916-919
17. Chi Zhan; Wang, P. A New Method of Color Image Segmentation Based on Intensity and Hue Clustering Proceedings 15th International Conference on Pattern Recognition 3 (September 2000): 613 -616.

18. P. Sharma.; R. Reilly. Fast Marching Methods Applied to Face Location in Videophone Applications Using Color Information. IEEE International Conference on Multimedia and Expo 2 (August 2002): 141-144.
19. Kass, M.; Witkin, A.; and Terzopoulos D. Snakes: Active contour models. Int. J. comput. Vis. 1 (1987): 321 -331.
20. L. D. Cohen. On active contour models and balloons. Computer Vision and Image Processing: Image Understanding 53 (March 1991): 211-218.
21. C. Xu and J. L. Prince. Snake, Shapes and Gradient Vector Flow. IEEE Transactions on Image Processing 7 (March 1998): 359-369.
22. K. Sobottka; and I. Pitass. Face localization and facial feature extraction based on shape and color information. IEEE Int. Conf. on Image Processing 3 (1996): 483-486.
23. M. M. Fleck; D. Forsyth; and C. Bregler. Finding Naked People. Proceedings of 4th European Conference on Computer Vision Springer-Verlag 2 (April 1996): 592-602.
24. J. Fritsch; S. Lang; M. Kleinhagenbrock; G. A. Fink; and G. Sagerer. Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection IEEE Int. Workshop on Robot and Human Interactive Communication (September 2002): 337-343.
25. Douglas Chai; and Abdesselam Bouzerdoum. Coding Videophone Sequences at Better Perceptual Quality By Using Face Localization and Redistribution. IEEE International Symposium on Intelligent Signal Processing and Communication Systems 1 (November 2000): 22-26.
26. Huynh-Thu, Q.; Meguro, M.; Kaneko, M. Skin-Color Extraction in Images with Complex Background and Varying Illumination. IEEE Workshop on Applications Computer Vision 6 (December 2002): 280-285.

27. Chen Sipei; Li Zongjie; and Zhang Guilin. A renovated Algorithm for Extracting Moving Target from Background in Real Time Video Tracking System. International Conference on Pattern Recognition (November 1988): 646-648.
28. Fabrice Heitz; and Patrick Bouthemy. Motion Estimation and Segmentation using A Global Bayesian Approach. International Conference on Acoustics, Speech, and Signal Processing (April 1990): 2305-2308.
29. Chiou-Shann Fuh; and Petros Maragos. Affine Models for Image Matching and Motion Detection. International Conference on Acoustics, Speech, and Signal Processing (April 1991): 2409-2412.
30. Raynard O. Hinds; and Thrasyvoulos N. Pappas. An Adaptive Clustering Algorithm for Segmentation of Video Sequences. International Conference on Acoustics, Speech, and Signal Processing 4 (May 1995): 2427-2430.
31. Jean-Marc Odobez and Patrick Bouthemy. MRF-Based Motion Segmentation Exploiting A 2D Motion Model Robust Estimation. International Conference on Image Processing 3 (October 1995): 628-631.
32. Munchurl Kim; J. G. Chol; D. Kim; H. Lee; M. H. Lee; C. Ahn; and Y. Ho. A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information. IEEE Trans. Circuits Syst. Video Technol 9 (December 1999): 1216-1226.
33. Candemir Toklu; A. Murat Tekalp; and A. Tanju Erden. Semi-Automatic Video Object Segmentation in the Presence of Occlusion. Transaction on Circuits and System for Video Technology 10 No.4 (June 2000): 624-629.
34. Yaakov Tsaig; and A. Averbuch. Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach. IEEE Transaction on Circuits and System for Video Technology 12 (July 2002): 597-612.

35. Hui zhu; and Zaiming Li. A video segmentation algorithm based on spatial-temporal information. International Conference on Communications, Circuits and Systems and West Sino Expositions 1 (June-July 2002): 566 -569.
36. Jong Bae Kim; and Hang Joon Kim. A wavelet-based watershed image segmentation for VOP generation. The 16th International Conference on Pattern Recognition (August 2002): 505 -508.
37. Hotter, M.; and Thoma, R. Image Segmentation Based on Object Oriented Mapping Parameter Estimation. IEEE Transactions on Signal Processing 15 (1998): 315-334.
38. Aach,T.; Kaup, A.; and Mester, R. Statistical Model Based Change Detection in Moving Video. IEEE Transactions on Signal Processing 31 (March 1993): 165-180.
39. Neri, A.; Colonnese, S.; Russo, G.; and Ralone, P. Automatic Moving Object and Background Separation. IEEE Transactions on Signal Processing 66 (1998): 219-232.
40. Daehee Kim; and Yo-Sung Ho. Shape Training for Video Object Segmentation. Int'l Workshop on Very Low Bitrate Video (VLBV) Coding (October 2001): 186-190.
41. M. J. Jones; and J. M. Rehg Statistical Color Models with Application to Skin Detection. Proc IEEE Conf. on Computer Vision and Pattern Recognition 1 (1999): 274-280.
42. J. Liu; and Y. H. Yang Multiresolution Color Image Segmentation. IEEE Transaction on Pattern Analysis and Machine Intelligence 16 (1994): 689-700.
43. Jongmoo Choi; Sanghoon Lee; and Juneho Yi A Real-Time Face Recognition Systems Using Multiple Mean Faces And Dual Mode Fisherfaces. IEEE Symposium on Industrial Electronics 3 (2001): 1686-1689.

44. Yi Liu; and Yuan F. Zheng Video Object Segmentation and Tracking Using w-Learning Classification. IEEE Transaction on Circuits and Systems For Video Technology 15 (July 2005): 885-899.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทความทางวิชาการที่ได้รับการเผยแพร่

1. Teerayoot Sawangsri, Vorapoj Patanavijit and Somchai Jitapunkul. "Face Segmentation Based on Hue-Cr Component And Morphological Technique," IEEE International Symposium on Circuit and Systems (ISCAS 2005), May 2005, Kobe, Japan
2. Teerayoot Sawangsri, Vorapoj Patanavijit and Somchai Jitapunkul. "Face Segmentation Using Novel Skin-Color Map And Morphological Technique," Proceedings of the International Conference on Signal Processing (ICSP 2004), November 2004, Istanbul, Turkey.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Face Segmentation Based on Hue-Cr Components And Morphological Technique

Teerayoot Sawangsri, Vorapoj Patanavijit, and Somchai Jitapunkul

Department of Electrical Engineering
Chulalongkorn University
Bangkok, Thailand

teerayoot.s@student.chula.ac.th, vorapoj.p@student.chula.ac.th, and somchai.j@chula.ac.th

Abstract—The paper proposes a novel algorithm to dynamically define the Region of Interest (ROI) videophone application. The algorithm uses the color information Hue and Cr to find the skin-color pixels and also use range of threshold obtained from red and blue components in normalized RGB color space to remove nonskin-color pixels because the human skin tends to have a predominance of red and nonpredominance of blue. Post-processing is used to remove such noises by a morphological operator. Moreover, the algorithm performs temporal filtering to remove skin-color pixels that immediately appear from frame to frame by using object tracking process to perform as memory for collecting skin-color objects obtained from previous frame to guide the next frame. The experimental results confirm the effectiveness of the proposed algorithm.

I. INTRODUCTION

The demand for applications of the digital video communications, such as videophone and videoconferencing, has increased considerably then coding for very low rate channels requires novel compression schemes. The content-based coding offers a more flexible and intelligent approach than traditional methods embodied by H.263/H.263+ and MPEG-4rstandard.

The developed video coder improves the quality of perceptually important image regions, relative to traditional coding standard, at the expense of perceptually insignificant portions of the image.

Faces are an important component of image sequences (i.e. the head and shoulders videophone and videoconferencing) and faces have some distinct features from background therefore a skin segmentation plays an important role in recent color-based approaches to human face segmentation [5] and [6].

In these approaches, regions of the input image that have skin colors are segmented to give an initial estimate of the face locations. These segmented skin regions will be further processed in the later stages of a face segmentation that utilizes several color spaces to label pixels as skin including normalized RGB [1], HSV [5] and YCbCr [2] and [6].

This paper proposes a skin segmentation algorithm for color image sequences. The proposed algorithm uses a combined human skin color model to classify skin-colors and

nonskin-colors. The model is based on the nonparametric histogram color model [2], [3], [5] and [6]. Compared to many existing skin segmentation approaches, the proposed algorithm combine two color spaces for skin segmentation and use property of skin color to set threshold for removing some noises that similar to skin-color.

This paper is organized as follows: Section II. explains the human skin color model. Section III. presents the skin segmentation algorithm. Section IV. discussed experimental results and Section V. concludes this paper.

II. HUMAN SKIN COLOR MODEL

A human skin color model is used to decide either a pixel is skin color or nonskin-color. This model is characterized by a classification algorithm and a color space used to represent pixel color. Color spaces used in skin color segmentation include YCbCr, HSV and normalized RGB.

The skin color model used in this paper is based on the nonparametric histogram color model [2], [5] and [6]. Jones and Rehg [8] compare the performance of histogram and parametric method, mixture models, for skin detection, they find histogram models to be superior in accuracy and computational cost.

There are many color models used for the modeling of human skin-color. Liu [4] has contended that there does not exist a single color models that work best for all kinds of image sequences therefore it is very important to choose the suitable color space to be combined for the modeling of human skin-color. Sobottka and Pitas [5] have operated for the HSV color space that is compatible with human color perception. They have considered hue and saturation as discriminating color information that describes the human skin color. This information can be defined as a priori and used subsequently as reference for any skin color. They denote R_S and R_H as the respective ranges of Hue and Saturation values that correspond to skin-color and they have chosen the parameters as follows : $R_S = [0.23 \ 0.68]$ and $R_H = [0^\circ \ 50^\circ]$. Consider an input image sequence of $M \times N$ pixel. Hence, we can obtain a reference map by following criterion.

$$\text{Map}_{\text{HS}(x,y)} = \begin{cases} 255, & \text{Hue} \in R_H \cap S \in R_S \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $x = 1, 2, \dots, M$ and $y = 1, 2, \dots, N$.

The paper shows the reference maps in Fig. 1b-1d that match all skin-color regions besides some noise pixels that are similar to skin-color however Hue (H) is not reliable for the discrimination task when the saturation is low [9]. Because saturation defines the relative purity or the amount of white light mixed with a hue that is a color attribute as a pure color, skin-color pixels of Saturation component are wider distribution than hue component. Hence, we just adopt the color information of the hue component for the modeling of human skin-color for decrement effect from Saturation component.

Chai and Ngan [6] have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e., Cr and Cb) values narrowly and consistently distributed in the YCbCr color space. They denote R_{Cr} and R_{Cb} as the respective ranges of Cr and Cb values that correspond to skin-color and they set $R_{Cr} = [133 \ 173]$ and $R_{Cb} = [77 \ 127]$.

$$\text{Map}_{\text{Cb}\&\text{Cr}(x,y)} = \begin{cases} 255, & \text{Cb} \in R_{Cb} \cap \text{Cr} \in R_{Cr} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $x = 1, 2, \dots, M/2$ and $y = 1, 2, \dots, N/2$.

The paper shows the reference maps in Fig. 2b-d. Notice that Fig. 2b match all skin-color regions but it is very over segmentation because many background can not be eliminated by Cb value. Fig. 2c and Fig. 2d match all skin-color regions besides some noise pixels that are similar to skin-color.

Moreover, Mei-Juan Chen and Ming-Chieh Chi [7] found an interesting problem, though Chai and Ngans' method is well but if the skin color region of Map_{Cb} is poor even though Map_{Cr} is still strong to present the region of skin-color, the skin-color region of $\text{Map}_{Cb\&Cr}$ will be poor. Moreover, the distribution of skin areas is consistent across different races in the Cb and Cr color spaces [4]. Then, we can only adopt the color information of Cr component for the modeling of human skin-color.

Because there does not exist a single color models that work best for all kinds of image sequences, this paper proposes a combined color spaces that are suitable for skin pixel classification under different conditions, i.e. different races and varying illuminations. We use color information of Hue and Cr to define skin-color and estimate a reference map in YCbCr and HSV color space. Moreover, R_{Cr} and R_H is denoted as the respective ranges of Cr and Hue values that correspond to skin-color and $R_{Cr} = [133 \ 173]$ [6] and $R_H = [0^\circ \ 50^\circ]$ [5].

$$\text{Map}_{\text{Cr}\&\text{Hue}(x,y)} = \begin{cases} 255, & \text{Cr} \in R_{Cr} \cap \text{Hue} \in R_H \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

For Hue color component, where $x = 1, 2, \dots, M$ and $y = 1, 2, \dots, N$. For Cr color component, $x = 1, 2, \dots, M/2$ and $y = 1, 2, \dots, N/2$.

The reference maps result is shown in Fig. 3b. Notice that Fig. 3b matches all skin-color regions and have a little some noise pixels that are similar to skin-color.

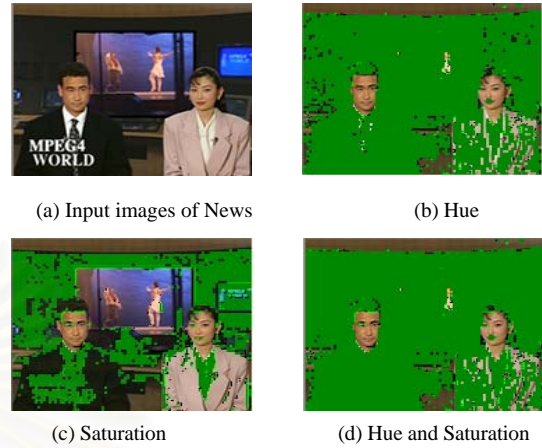


Figure 1. Result of different color components reference maps in YCbCr color space.

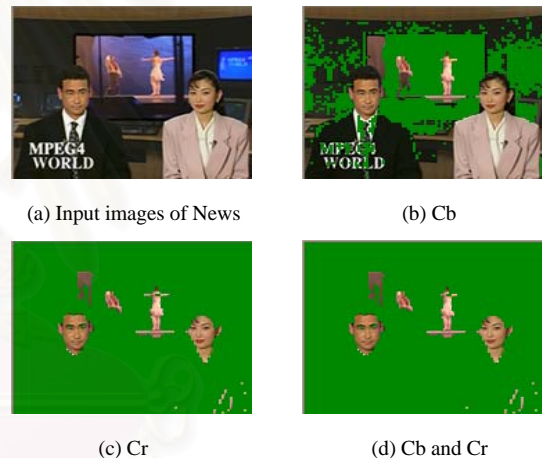


Figure 2. Result of different color components reference maps in YCbCr color space.



Figure 3. Result of proposed color components reference maps in YCbCr and HSV color space.

III. SKIN-COLOR SEGMENTATION

As shown in Fig. 4, the proposed method contains three stages

A. Color Segmentation

The first stage of the color segmentation algorithm involves the use of color information in a low-level region segmentation process to define region of interest. The aim is to classify pixels of the input image sequence into skin color and nonskin-color. To do so, the paper have devised a skin-color reference map thus we can easily discriminate the region of interest and noninterest via the proposed reference map that this paper use color information of Hue and Cr component to define skin-color.

B. Background Elimination

This section presents an algorithm to remove some noise pixels that are similar to skin-color, pseudo-skin-color, from first state. It is a fact that human skin tend to have a predominance of red and nonpredominance of blue. Furthermore, Mei-Juan Chen and Ming-Chieh Chi [7] found that skin-colors and pseudo-skin-colors locate at the same range, $R_{Cr} = [133 \ 173]$, in Cr component but these two colors are different in RGB color space. The blue and red histogram of skin-color pixels of Fig. 2a is shown in Fig. 5a and Fig. 5b, respectively. The statistical results show that the histograms of red and blue for skin-color pixels in a normalized RGB color space are narrowly distributed because the normalization was employed to minimize the dependence on the luminance values then the paper can utilize the characteristic to remove pseudo-skin-color. And we can set R_{blue} and R_{red} as a threshold selected empirically from the histogram of samples and check that if each pixel from first state has the value of Blue in range of R_{blue} and has the value of Red in range of R_{red} that is real skin-color pixel. On the other hand, if the each pixel has the value of Blue in range of R_{blue} but has not the value of Red in range of R_{red} then this pixel is classified to be pseudo-skin-color pixel and we set $R_{blue} = [0.15 \ 0.3]$ and $R_{red} = [0.4 \ 0.5]$.

C. Post-processing and Object Tracking

Next, the process extracts the large skin-color objects from the image by using morphologic closing operations that are commonly used to smooth, fill in, and remove objects in an image sequence. Morphologic closing is equivalent to a dilation followed by an erosion. Nevertheless, these operations can only reduce some noise areas but cannot completely remove noise from the mask. In order to remove the remaining noise, the image is classified into 4×4 pixels block and these blocks are then categorized into skin-color and nonskin-color blocks. If there are more than 8 pixels in the block, it will be assigned to full pixel block, otherwise it will be assigned as empty block then each block that has the same value will be considered as a pixel. Finally, the algorithm performs temporal filtering to remove skin-color pixels that immediately appear from frame to frame by using the object tracking process that performs as memory for collecting skin-color object obtained from previous frame to guide the next frame. If the skin-color in next frame cover the skin-color in previous frame more than 50% then skin-color regions will be not removed, otherwise some regions that immediately appear and cover less than 50% of previous

frame, will be removed. As a result, the output mask of this step should contain the face region with minimal or no noise and the next section will show the experimental results.

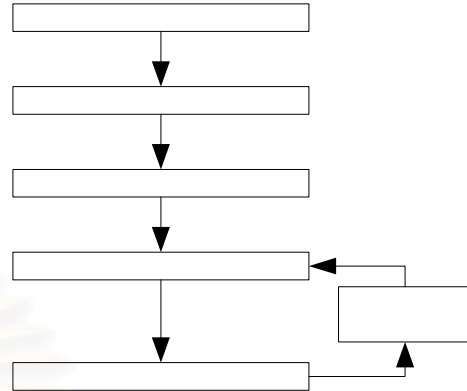
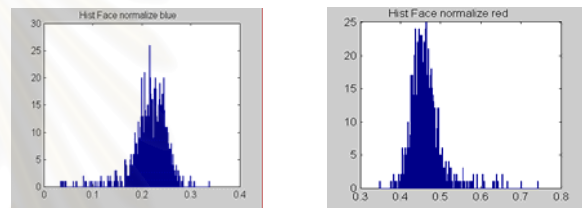


Figure 4. Block diagram of videophone skin-color segmentation algorithm



(a) Histogram of blue domain

(b) Histogram of red domain

Figure 5. The histograms of Blue and Red domain of Fig. 2a in normalized RGB color space.

IV. EXPERIMENTAL RESULTS

The experimental results are performed using three videophone sequences: News, Grandma and Carphone. These sequences are video standard sequences in QCIF format, which can be categorized by moving characteristic. First, News sequence has less movement because only their heads and shoulders move. Second, Grandma sequence has very slowly movement along the sequence. Finally, Carphone sequence has movement very much. At the first stage of the algorithm color segmentation, the reference maps result produced by the proposed criterion is shown in Fig. 6d, Fig. 7d and Fig. 8d that are compared with the Cb and Cr method [6] in Fig. 6b, Fig. 7b and Fig. 8b and Hue and Saturation method [5] in Fig. 6c, Fig. 7c and Fig. 8c.

The performance of the skin-color segmentation methodology is evaluated quantitatively. To quantitatively measure the accuracy of the proposed methodology, each frame was manually segmented into skin and nonskin pixels. The manually segmented images serve as a reference to which the automatically segmented images are compared. The detection rate and false alarm rate are expressed as a percentage, which represent in equation (4) and (5) respectively.

$$DR = \frac{TP}{TP + FN} \quad (4)$$

$$FAR = \frac{FP}{TP + FP} \quad (5)$$

where :

TP, True Positive, is defined as a group of pixels that are foreground and the system distinguish that a group of pixels are foreground.

FP, False Positive, is defined as a group of pixels that are background but the system distinguish that a group of pixels are foreground.

FN, False Negative, is defined as a group of pixels that are foreground but the system distinguish that a group of pixels are background. The detection rates and false alarm rates for Fig. 6 – Fig. 8 video sequences are shown in Table I and Table II respectively. These results show that the proposed segmentation algorithm could provide the face segmentation more efficient and it also has less affect from the fast or slow moving object in video sequences.

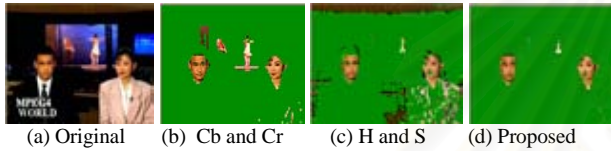


Figure 6. Comparison of different skin-color segmentation method of News sequence

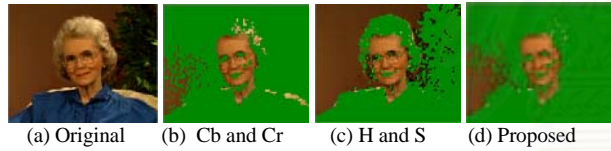


Figure 7. Comparison of different skin-color segmentation method of Grandma sequence

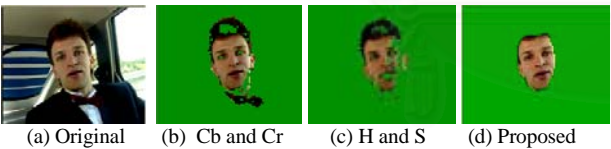


Figure 8. Comparison of different skin-color segmentation method of Carphone sequence

TABLE I
DETECTION RATES.

Sequence	Detection Rate		
	CbCr	HS	Proposed CrHue
Car phone	87.56 %	70.48 %	72.09%
Grandma	70.11 %	76.79 %	71.63 %
News	90.17 %	90.97 %	93.79 %

TABLE II
FALSE ALARM RATES.

Sequence	False Alarm Rate		
	CbCr	HS	Proposed CrHue
Car phone	27.90 %	29.51%	12.43%
Grandma	29.88 %	43.20 %	20.36%
News	13.20 %	19.02%	6.82%

V. CONCLUSION

This paper proposed a novel algorithm for skin segmentation of color image sequence. The proposed algorithm uses a combined human skin color model based on the nonparametric color model to classify skin-colors and nonskin-colors. We use property of skin color to set threshold for removing some noises that similar to skin-color because the human skin tend to have a predominance of red and nonpredominance of blue and we use morphologic closing operations to smooth, fill in, and remove objects in an image sequence.

Finally, the object tracking process performs as memory for collecting skin-color objects obtained from previous frame to guide the next frame in order to remove skin-color pixels that immediately appear from frame to frame. The experimental results show the satisfying subjective test results.

ACKNOWLEDGMENT

The authors would like to express the grateful thanks to grant from government research and development in cooperative project between Department of Electrical Engineering and Private Sector for Research for supporting this work and development under Chulalongkorn University.

REFERENCES

- [1] J.L. Crowley and J.M. Berard, "Multi-Modal Tracking of Faces for Video Communications," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 640 – 645, 1997.
- [2] D. Tancharoen, S. Jitapankul, P. Kittipanya-ngam, S. Chompon, and H. Kortrakulkij, "Semantic Object Segmentation for Content Based Video Coding," Proceedings Vol. XIII, Image, Acoustic, Speech and Signal Processing: Part II, World Multiconference on Systemics, Cybernetics, and Informics, pp. 579-583, July 2001.
- [3] L. Sigal., S. Sclaroff and V. Athitsos, "Skin Color-Based Video Segmentation under Time-Varying Illumination," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, pp. 862-877, July. 2004.
- [4] J. Liu and Y.-H. Yang, "Multiresolution Color Image Segmentation," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.16, No.7, pp. 689-700, 1994.
- [5] K. Sobottka and I. Pitass, "Face localization and facial feature extraction based on shape and color information," IEEE Int. Conf. on Image Processing, Vol. III, pp. 483-486, 1996.
- [6] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," IEEE Trans. on Circuits and System for video Technology, Vol.9, No.4, pp. 551-564, 1999.
- [7] Mei-Juan, Ming-Chieh Chi, Ching-Ting Hsu and Jeng-Wei Chen, "ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique," IEEE Trans. on Consumer electronics, Vol.49, No.3, pp. 724-730, 2003.
- [8] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 1, pp. 274-280, 1999.
- [9] Jongmoo Choi, Sanghoon Lee, Chilgee Lee and Juneho Yi, "A Real-Time Face Recognition Systems Using Multiple Mean Faces And Dual Mode Fisherfaces," IEEE Symposium on Industrial Electronics, Vol.3, pp. 1686-1689, 2001.

Face Segmentation Using Novel Skin-Color Map And Morphological Technique

Teerayoot Sawangsri, Vorapoj Patanavijit, and Somchai Jitapunkul

Abstract—The paper proposes a novel algorithm to dynamically define the Region of Interest (ROI) videophone application. The algorithm uses the color information Hue and Cr to find the skin-color pixels and also use range of threshold obtained from red and blue components in normalized RGB color space to remove nonskin-color pixels because the human skin tends to have a predominance of red and non-predominance of blue. Post-processing is used to remove such noises by a morphological operator. Moreover, the algorithm performs temporal filtering to remove skin-color pixels that immediately appear from frame to frame by using object tracking process to perform as memory for collecting skin-color objects obtained from previous frame to guide the next frame. The experimental results confirm the effectiveness of the proposed algorithm.

Keywords—face segmentation, human skin segmentation, video object segmentation, Region-of-Interest (ROI) video coding.

I. INTRODUCTION

THE demand for applications of the digital video communications, such as videophone and videoconferencing, has increased considerably then coding for very low rate channels requires novel compression schemes. The content-based coding offers a more flexible and intelligent approach than traditional methods embodied by H.263/H.263+ and MPEG-4rstandard.

The developed video coder improves the quality of perceptually important image regions, relative to traditional coding standard, at the expense of perceptually insignificant portions of the image.

Faces are an important component of image sequences (i.e. the head and shoulders videophone and videoconferencing) and faces have some distinct features from background therefore a skin segmentation plays an important role in recent color-based approaches to human face segmentation [5] and [6].

In these approaches, regions of the input image that have skin colors are segmented to give an initial estimate of the face locations. These segmented skin regions will be further

processed in the later stages of a face segmentation that utilizes several color spaces to label pixels as skin including normalized RGB [1], HSV [5] and YCbCr [2] and [6].

This paper proposes a skin segmentation algorithm for color image sequences. The proposed algorithm uses a combined human skin color model to classify skin-colors and nonskin-colors. The model is based on the nonparametric histogram color model [2], [3], [5] and [6]. Compared to many existing skin segmentation approaches, the proposed algorithm combine two color spaces for skin segmentation and use property of skin color to set threshold for removing some noises that similar to skin-color.

This paper is organized as follows: Section II. explains the human skin color model analysis. Section III. presents the skin segmentation algorithm. Section IV. discussed experimental results and Section V. concludes this paper.

II. HUMAN SKIN-COLOR MODEL ANALYSIS

A human skin color model is used to decide either a pixel is skin color or nonskin-color. This model is characterized by a classification algorithm and color space used to represent pixel color. Color spaces used in skin color segmentation include YCbCr, HSV and normalized RGB.

The skin color model used in this paper is based on the nonparametric histogram color model [2], [5] and [6]. Jones and Rehg [8] compare the performance of histogram and parametric method, mixture models, for skin detection, they find histogram models to be superior in accuracy and computational cost.

There are many color models used for the modeling of human skin-color. Liu [4] has contended that there does not exist a single color models that work best for all kinds of image sequences therefore it is very important to choose the suitable color space to be combined for the modeling of human skin-color. Sobottka and Pitas [5] have operated for the HSV color space that is compatible with human color perception. They have considered hue and saturation as discriminating color information that describes the human skin color. This information can be defined as a priori and used subsequently as reference for any skin color. They denote R_S and R_H as the respective ranges of Hue and Saturation values that correspond to skin-color and they have chosen the parameters as follows : $R_S = [0.23 \ 0.68]$ and $R_H = [0^\circ \ 50^\circ]$.

Manuscript received November 5, 2004. This work was supported in part by the cooperative project between Department of Electrical Engineering, Chulalongkorn University and Private Sector.

T. Sawangsri is with Chulalongkorn University, Thailand. (e-mail: teerayoot.s@student.chula.ac.th).

V. Patanavijitr is with the Electrical Engineering Department, Asumption University, Thailand. (e-mail: vorapoj.p@student.chula.ac.th).

S. Jitapunkul is with the Electrical Engineering Department, Chulalongkorn University, Thailand.

Consider an input image sequence of $M \times N$ pixel. Hence, we can obtain a reference map by following criterion.

Please check with your editor on whether to submit your manuscript by hard copy or electronically for review. If hard copy, submit photocopies such that only one column appears per page.

$$\text{Map}_{\text{HS}(x,y)} = \begin{cases} 255, & \text{Hue} \in R_H \cap S \in R_S \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $x = 1, 2, \dots, M$ and $y = 1, 2, \dots, N$.

The paper shows the reference maps in Fig. 1b-1d that match all skin-color regions besides some noise pixels that are similar to skin-color however Hue (H) is not reliable for the discrimination task when the saturation is low [9]. Because saturation defines the relative purity or the amount of white light mixed with a hue that is a color attribute as a pure color, skin-color pixels of Saturation component are wider distribution than hue component. Hence, we just adopt the color information of the hue component for the modeling of human skin-color for decrement effect from Saturation component.

Chai and Ngan [6] have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e., Cr and Cb) values narrowly and consistently distributed in the YCbCr color space. They denote R_{Cr} and R_{Cb} as the respective ranges of Cr and Cb values that correspond to skin-color and they set $R_{Cr} = [133 \ 173]$ and $R_{Cb} = [77 \ 127]$.

$$\text{Map}_{\text{Cb}\&\text{Cr}(x,y)} = \begin{cases} 255, & \text{Cb} \in R_{Cb} \cap \text{Cr} \in R_{Cr} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $x = 1, 2, \dots, M/2$ and $y = 1, 2, \dots, N/2$.

The paper shows the reference maps in Fig. 2b-d. Notice that Fig. 2b match all skin-color regions but it is very over segmentation because many background can not be eliminated by Cb value. Fig. 2c and Fig. 2d match all skin-color regions besides some noise pixels that are similar to skin-color.

Moreover, Mei-Juan Chen and Ming-Chieh Chi [7] found an interesting problem, though Chai and Ngans' method is well but if the skin color region of Map_{Cb} is poor even though Map_{Cr} is still strong to present the region of skin-color, the skin-color region of $\text{Map}_{Cb\&Cr}$ will be poor. Moreover, the distribution of skin areas is consistent across different races in the Cb and Cr color spaces [4]. Then, we can only adopt the color information of Cr component for the modeling of human skin-color.

Because there does not exist a single color models that work best for all kinds of image sequences, this paper proposes a combined color spaces that are suitable for skin pixel classification under different conditions, i.e. different races and varying illuminations. We use color information of Hue and Cr to define skin-color and estimate a reference map in YCbCr and HSV color space. Moreover, R_{Cr} and R_H is denoted as the respective ranges of Cr and Hue values that correspond to skin-color and $R_{Cr} = [133 \ 173]$ [6] and $R_H = [0^\circ \ 50^\circ]$ [5].

$$\text{Map}_{\text{Cr}\&\text{Hue}(x,y)} = \begin{cases} 255, & \text{Cr} \in R_{Cr} \cap \text{Hue} \in R_H \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

For Hue color component, where $x = 1, 2, \dots, M$ and $y = 1, 2, \dots, N$. For Cr color component, $x = 1, 2, \dots, M/2$ and $y = 1, 2, \dots, N/2$.

The reference maps result is shown in Fig.1b. Notice that Fig. 3b matches all skin-color regions and have a little some noise pixels that are similar to skin-color.

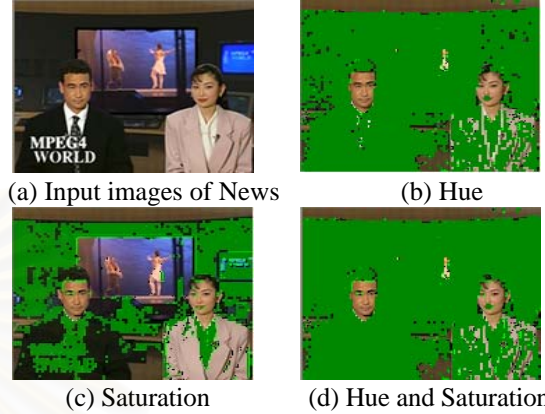


Figure 1. Result of different color components reference maps in YCbCr color space.

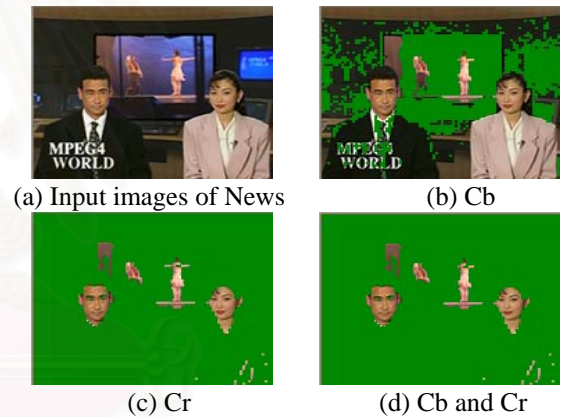


Figure 2. Result of different color components reference maps in YCbCr color space.



Figure 3. Result of proposed color components reference maps in YCbCr and HSV color space.

III. SKIN-COLOR SEGMENTATION

As shown in Fig. 4, the proposed method contains three stages

A. Color Segmentation

The first stage of the color segmentation algorithm involves the use of color information in a low-level region

segmentation process to define region of interest. The aim is to classify pixels of the input image sequence into skin color and nonskin-color. To do so, the paper have devised a skin-color reference map thus we can easily discriminate the region of interest and noninterest via the proposed reference map that this paper use color information of Hue and Cr component to define skin-color.

B. Background Elimination

This section presents an algorithm to remove some noise pixels that are similar to skin-color, pseudo-skin-color, from first state. It is a fact that human skin tend to have a predominance of red and non-predominance of blue. Furthermore, Mei-Juan Chen and Ming-Chieh Chi [7] found that skin-colors and pseudo-skin-colors locate at the same range, $R_{Cr} = [133\ 173]$, in Cr component but these two colors are different in RGB color space. The blue and red histogram of skin-color pixels of Fig. 2a is shown in Fig. 5a and Fig. 5b, respectively. The statistical results show that the histograms of red and blue for skin-color pixels in a normalized RGB color space are narrowly distributed because it was employed to minimize the dependence on the luminance values then the paper can utilize the characteristic to remove pseudo-skin-color. And we can set R_{blue} and R_{red} as a threshold selected empirically from the histogram of samples and check that if each pixel from first state has the value of Blue in range of R_{blue} and has the value of Red in range of R_{red} that is real skin-color pixel. On the other hand, if the each pixel has the value of Blue in range of R_{blue} but has not the value of Red in range of R_{red} then this pixel is classified to be pseudo-skin-color pixel and we set $R_{blue} = [0.15\ 0.3]$ and $R_{red} = [0.4\ 0.5]$.

C. Post-processing And Object Tracking

Next, the process extracts the large skin-color objects from the image by using morphologic closing operations that are commonly used to smooth, fill in, and remove objects in an image sequence. Morphologic closing is equivalent to a dilation followed by an erosion. Nevertheless, these operations can only reduce some noise areas but cannot completely remove noise from the mask. In order to remove the remaining noise, the image is classified into 4x4 pixels block and these blocks are then categorized into skin-color and nonskin-color blocks. If there are more than 8 pixels in the block, it will be assigned to full pixel block, otherwise it will be assigned as empty block then each block that has the same value will be considered as a pixel. Finally, the algorithm performs temporal filtering to remove skin-color pixels that immediately appear from frame to frame by using the object tracking process that performs as memory for collecting skin-color object obtained from previous frame to guide the next frame. If the skin-color in next frame cover the skin-color in previous frame more than 50% then skin-color regions will be not removed, otherwise some regions that immediately appear and cover less than 50% of previous frame, will be removed. As a result, the output mask of this step should contain the face region with minimal or no noise and the next section will show the experimental results.

IV. EXPERIMENTAL RESULT

The experimental results are performed using three videophone sequences: News, Grandma and Carphone. These sequences are video standard sequences in QCIF format, which can be categorized by moving characteristic. First, News sequence has less movement because only their heads and shoulders move. Second, Grandma sequence has very slowly movement along the sequence. Finally, Carphone sequence has movement very much. At the first stage of the algorithm color segmentation, the reference maps result produced by the proposed criterion is shown in Fig. 6d, Fig. 7d and Fig. 8d that are compared with the Cb and Cr method [6] in Fig. 6b, Fig. 7b and Fig. 8b and Hue and Saturation method [5] in Fig. 6c, Fig. 7c and Fig. 8c.

These results show that the proposed segmentation algorithm could provide the face segmentation more efficient and it also has less affect from the fast or slow moving object in video sequences.

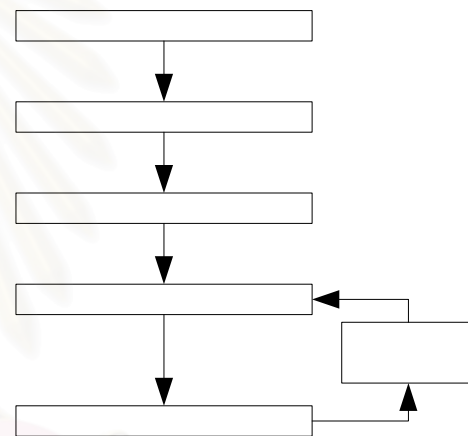
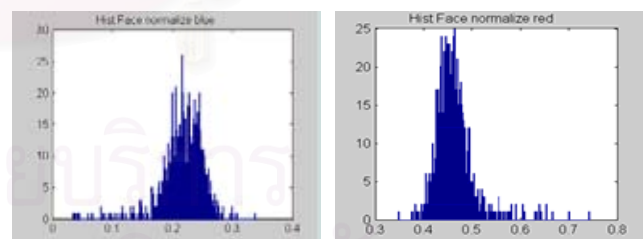


Figure 4. Block diagram of videophone skin-color segmentation algorithm



(a) Histogram of blue domain

(b) Histogram of red domain

Figure 5. The histograms of Blue and Red domain of Fig. 2a in normalized RGB color space.

V. CONCLUSION

This paper proposed a novel algorithm for skin segmentation of color image sequences. The proposed algorithm uses a combined human skin color model based on the nonparametric color model to classify skin-colors and nonskin-colors. We use property of skin color to set threshold for removing some noises that similar to skin-color because the human skin tend to have a predominance of red

and non-predominance of blue and we use morphologic closing operations to smooth, fill in, and remove objects in an image sequence.

Finally, the object tracking process performs as memory for collecting skin-color objects obtained from previous frame to guide the next frame in order to remove skin-color pixels that immediately appear from frame to frame. The experimental results show the satisfying subjective test results.

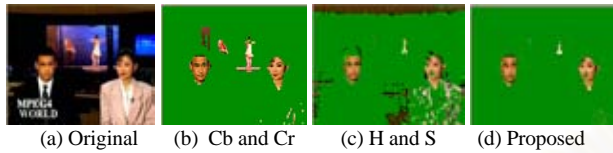


Figure 6. Comparison of different skin-color segmentation method of News sequence

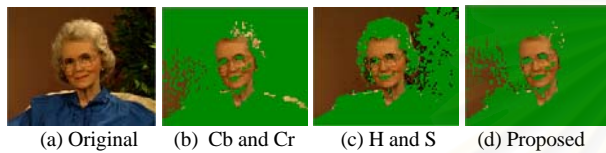


Figure 7. Comparison of different skin-color segmentation method of Grandma sequence

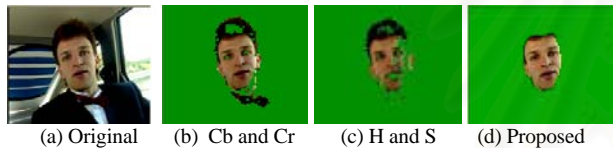


Figure 8. Comparison of different skin-color segmentation method of Carphone sequence

ACKNOWLEDGMENT

The authors would like to express the grateful thanks to grant from government research and development in cooperative project between Department of Electrical Engineering and Private Sector for Research for supporting this work and development under Chulalongkorn University.

REFERENCES

- [1] J. L. Crowley and J.M. Berard, "Multi-Modal Tracking of Faces for Video Communications," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 640 – 645, 1997.
- [2] D. Tancharoen, S. Jitapankul, P. Kittipanya-ngam, S. Chompon, and H. Kortrakulkij, "Semantic Object Segmentation for Content Based Video Coding," *Proceedings Vol. XIII, Image, Acoustic, Speech and Signal Processing: Part II, World Multiconference on Systemics, Cybernetics, and Informics*, pp. 579-583, July 2001.
- [3] L. Sigal., S. Sclaroff and V. Athitsos, "Skin Color-Based Video Segmentation under Time-Varying Illumination," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 862-877, July 2004.
- [4] J. Liu and Y.-H. Yang, "Multiresolution Color Image Segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.16, No.7, pp. 689-700, 1994.
- [5] K. Sobottka and I. Pitass, "Face localization and facial feature extraction based on shape and color information," *IEEE Int. Conf. on Image Processing*, Vol. III, pp. 483-486, 1996.
- [6] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *IEEE Trans. on Circuits and System for video Technology*, Vol.9, No.4, pp. 551-564, 1999.
- [7] Mei-Juan, Ming-Chieh Chi, Ching-Ting Hsu and Jeng-Wei Chen, "ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique," *IEEE Trans. on Consumer electronics*, Vol.49, No.3, pp. 724-730, 2003.
- [8] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 274-280, 1999.
- [9] Jongmoo Choi, Sanghoon Lee, Chilgee Lee and Juneho Yi, "A Real-Time Face Recognition Systems Using Multiple Mean Faces And Dual Mode Fisherfaces," *IEEE Symposium on Industrial Electronics*, Vol.3, pp. 1686-1689, 2001.

ประวัติผู้เขียนวิทยานิพนธ์

นายธีรยุทธ สว่างศรี เข้ารับการศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒในปีการศึกษา 2540 สำเร็จการศึกษาปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ในปีการศึกษา 2544 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ในสังกัดห้องปฏิบัติการกรรมวิธีสัตยญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า ที่จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2544



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย