

การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติโดยใช้เทคนิคการตัดสลับใจแบบเบย์



นายศักดิ์ เชมชัย

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

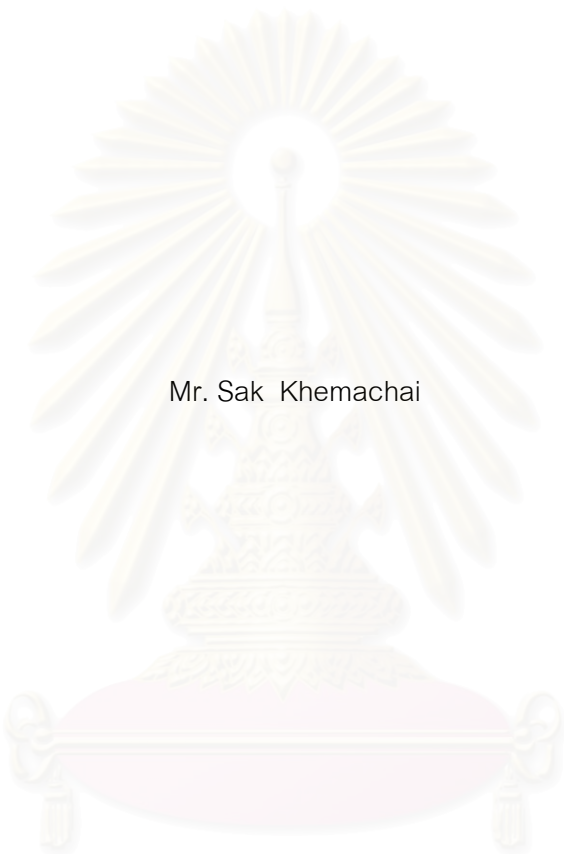
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2546

ISBN 974-17-3956-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AUTOMATIC SEGMENTATION OF VIDEO OBJECTS USING BAYESIAN DECISION TECHNIQUE



Mr. Sak Khemachai

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

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

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2003

ISBN 974-17-3956-7

ศักดิ์ เขมชัย : การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติโดยใช้เทคนิคการตัดสินใจแบบเบย์ (AUTOMATIC SEGMENTATION OF VIDEO OBJECTS USING BAYESIAN DECISION TECHNIQUE) อ. ที่ปรึกษา : รศ. ดร.สมชาย จิตะพันธ์กุล , 99 หน้า. ISBN 974-17-3956-7.

วิทยานิพนธ์นี้นำเสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุตามข้อกำหนดของ MPEG-4 การแยกส่วนภาพวัตถุนี้ประกอบด้วยกระบวนการ 5 กระบวนการคือ กระบวนการแยกส่วนภาพเชิงพื้นที่ กระบวนการแยกส่วนภาพเชิงเวลา กระบวนการตัดสินใจ กระบวนการข้างหลัง และกระบวนการติดตามวัตถุ การแยกส่วนภาพวิดีโอที่ถูกพิจารณาในวิทยานิพนธ์นี้มีการนำทฤษฎีการตัดสินใจแบบเบย์มาตัดสินใจว่าจุดภาพใดคือจุดภาพที่เป็นวัตถุ โดยที่เวกเตอร์ลักษณะเด่นที่ใช้คือองค์ประกอบสีของ Y และ Cb ร่วมกัน

กระบวนการแยกส่วนภาพเชิงพื้นที่นั้นใช้ตัวดำเนินการเชิงสัญญาณในการดำเนินการ การแยกส่วนภาพเชิงเวลานั้น ใช้ความแตกต่างกันของความเข้มแสงในการตรวจจับการเคลื่อนไหวและมีการสแกนเพื่อให้ได้วัตถุอย่างคร่าวๆ กระบวนการข้างหลังจะช่วยกรองสิ่งรบกวนที่ได้จากกระบวนการตัดสินใจและกระบวนการตรวจจับวัตถุนั้น จะเป็นกระบวนการช่วยค้นหาว่าวัตถุอยู่ตำแหน่งใดในภาพถัดไป จากผลการทดลองพบว่าการแยกส่วนภาพโดยวิธีการตัดสินใจแบบเบย์ให้ค่าอัตราผิดพลาดน้อยที่สุดเมื่อเทียบกับการแยกส่วนภาพโดยวิธีการทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา โดยใช้เวลาในการประมวลผลมากกว่าการแยกส่วนภาพเชิงพื้นที่และเวลา 3 เท่า แต่น้อยกว่าการแยกส่วนภาพโดยวิธีทดสอบสมมติฐานอยู่ครึ่งหนึ่ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่อนิสิต.....
สาขาวิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2546.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4470601121 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: VIDEO CODING OBJECT SEGMENTATION / BAYESIAN DECISION /
HYPOTHESIS TESTING / MORPHOLOGICAL OPERATOR

SAK KHEMACHAI : AUTOMATIC SEGMENTATION OF VIDEO OBJECTS USING
BAYESIAN DECISION TECHNIQUE. THESIS ADVISOR: ASSOC. PROF. SOMCHAI
JITAPUNKUL, Dr.Ing., 99 pp. ISBN 974-17-3956-7.

This thesis proposes a video object segmentation algorithm according to the MPEG-4 requirements which provide the object based video coding standard. This algorithm consists of five processes: spatial segmentation process, temporal segmentation process, decision process, postprocessing, and object tracking process. The Bayesian decision is exploited in order to segment video object in the decision process. The feature vector used in the decision process consists of Y and Cb color components in combination.

Spatial segmentation process segments image into many region by morphological operators. In temporal segmentation process, change detection of intensity and scanning process are used to segment roughly object. Postprocessing is the final process to segment fine object. Object tracking guide the next frame to locate the object. The experimental results show that this proposed technique has the least fault rate when compare with Hypothesis testing segmentation technique and spatio-temporal segmentation technique.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department.....Electrical Engineering..... Student's Signature.....
Field of study.....Electrical Engineering..... Advisor's signature.....
Academic year.....2003..... Co-advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลืออย่างดียิ่งของท่านอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล ซึ่งได้ให้ความกรุณาแนะนำและเสนอแนะข้อคิดเห็นต่างๆ ในการทำวิจัยมาตลอด ผู้วิจัยจึงขอกราบขอบพระคุณ

ขอขอบคุณ โครงการวิจัยร่วมระหว่างภาครัฐบาลและภาคเอกชนที่ให้ความสนับสนุนในส่วนของคุณและอุปกรณ์สำหรับช่วยทำงานวิจัย ซึ่งทำให้งานวิจัยดำเนินไปได้ด้วยความราบรื่น

กราบขอบพระคุณบิดามารดา ผู้ซึ่งทำให้ผู้วิจัยได้รับแต่สิ่งดีๆ เสมอมา ขอบพระคุณสำหรับความรักความห่วงใยที่ไม่เคยลดน้อยลงเลย ขอบคุณคุณอรอุมา พลัฒ์พลี ที่คอยเป็นกำลังใจและช่วยเหลือในหลายๆ ส่วนตลอดเวลา

ขอบคุณเพื่อนๆ พี่ๆ น้องๆ ทุกคน ที่ผู้วิจัยได้รู้จักไม่ว่าจะเป็นภายในห้องปฏิบัติการ การกรรมวิธีสัญญาณดิจิทัลหรือที่ได้ก็ตาม สำหรับความช่วยเหลือและคำทักเตือนอันมีค่าทำให้วิทยานิพนธ์เสร็จสิ้นได้ในที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ฌ
สารบัญภาพ	ญ
บัญชีคำศัพท์.....	ฎ
บทที่	
1. บทนำ.....	1
1.1 แนวทางของงานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุ.....	1
1.2 แนวทางของวิทยานิพนธ์.....	12
1.3 เป้าหมายและขอบเขตวิทยานิพนธ์	12
1.4 ขั้นตอนและวิธีการดำเนินงาน.....	12
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	12
1.6เค้าโครงวิทยานิพนธ์.....	13
2. ทฤษฎีพื้นฐาน	14
2.1 พื้นฐานของสี	14
2.1.1 แบบจำลองสี RGB (RGB color model)	14
2.1.2 แบบจำลองสี YUV (YUV color model).....	16
2.1.3 แบบจำลองสี YIQ (YIQ color model).....	16
2.1.4 แบบจำลองสี YCbCr (YCbCr color model).....	17
2.1.5 แบบจำลองสี HSI (HSI color model).....	20
2.2 ตัวดำเนินการเชิงสัณฐาน (Morphological Operator)	27
2.2.1 ตัวดำเนินการเชิงสัณฐานบนภาพไบนารี	27
2.2.2 ตัวดำเนินการเชิงสัณฐานต่อภาพสีเทา.....	31
2.2.3 เกรเดียนต์เชิงสัณฐาน (Morphological Gradient).....	34
2.2.4 อัลกอริทึมสันปันน้ำ (Watershed Algorithm)	34
2.2.4.1 ขั้นตอนการสร้างเขื่อน.....	35

บทที่	หน้า
2.3 ทฤษฎีการตัดสินใจของเบย์ (Bayesian Decision Theory)	36
3. การแยกส่วนภาพวิดีโอโดยใช้เทคนิคการตัดสินใจแบบเบย์.....	38
3.1 การแยกส่วนภาพทางพื้นที่ (Spatial Segmentation).....	38
3.2 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)	39
3.3 กระบวนการตัดสินใจ (Decision Process).....	42
3.4 กระบวนการข้างหลัง (Post Processing)	45
3.5 การติดตามวัตถุ (Object Tracking)	46
3.6 การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐาน (Hypothesis Testing)	46
3.7 การแยกส่วนภาพเชิงพื้นที่และเวลา.....	49
4. ผลการทดลอง.....	51
4.1 ผลการแยกส่วนภาพที่ได้จากกระบวนการข้างหลังโดยพิจารณาข้อมูลขององค์ประกอบสี่ตามแบบจำลองสี่ทั้ง 3 แบบ.....	52
4.2 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Claire.....	54
4.3 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Mother&Daughter.....	58
4.4 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Grandmother.....	63
4.5 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Suzie.....	69
4.6 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak7.....	74
4.7 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak35.....	79
4.8 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak39.....	85
4.9 การวิเคราะห์ความซับซ้อนของการแยกส่วนภาพทั้งสามวิธี.....	90
5. สรุปและข้อเสนอแนะ.....	93
5.1 สรุป.....	93
5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต.....	94
รายการอ้างอิง.....	95
ประวัติผู้เขียนวิทยานิพนธ์.....	99

สารบัญตาราง

ตารางที่	หน้า
2.1 ลักษณะของการเข้ารหัสสัญญาณวิดีโอระบบ 525 เส้น และระบบ 625 เส้น.....	19
4.1 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ 100 ภาพ.....	53
4.2 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Claire 45 ภาพ.....	57
4.3 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Mother&Daughter 90 ภาพ.....	63
4.4 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Grandmother 200 ภาพ.....	68
4.5 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Suzie 90 ภาพ.....	74
4.6 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak7 200 ภาพ.....	79
4.7 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak35 200 ภาพ.....	84
4.8 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak39 190 ภาพ.....	90
4.9 การเปรียบเทียบกระบวนการแยกส่วนภาพของทั้ง 3 กรรมวิธี.....	91
4.10 เปรียบเทียบจำนวนเฟรมที่ใช้ในการแยกส่วนได้เต็มตัว.....	91
4.11 เวลาที่ใช้ในการประมวลผลในแต่ละวิธีโดยเฉลี่ยจาก 100 เฟรม.....	92

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพประกอบ	หน้า
รูปที่ 1.1 กรรมวิธีที่เสนอโดย Fabrice Mosheni, Frederic Dufaux และ Murat Kunt	5
รูปที่ 1.2 กรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee Myoung Ho Lee, Chieteuk และ Yo-Sung Ho	7
รูปที่ 1.3 กรรมวิธีที่เสนอโดย Aljoscha Smolic, Thomas Sikora และ Jens – Rainer Ohm.....	8
รูปที่ 1.4 กรรมวิธีที่เสนอโดย Yaakov Tsaig และ Amir Averbuch.....	10
รูปที่ 2.1 พิกัดแบบจำลองสี RGB.....	15
รูปที่ 2.2 ลักษณะค่าแกมมาที่ใช้ในการแสดงผล.....	15
รูปที่ 2.3 ความแตกต่างของพิกัด UV และพิกัด IQ	17
รูปที่ 2.4 ตำแหน่งของจุดตัวอย่างหรือจุดภาพของ YCbCr.....	18
รูปที่ 2.5 พิกัดแบบจำลองสี HSI	21
รูปที่ 2.6 แสดงตำแหน่งที่ใช้ในการแปลงพิกัดจากแบบจำลองสี RGB เป็น HSI.....	23
รูปที่ 2.7 การฉายจุด P ลงบนแกนสีแดง	25
รูปที่ 2.8 ตัวอย่างการเลื่อนตำแหน่ง การสะท้อน คอมพลิเมนต์ และความแตกต่างกันระหว่างเซต	28
รูปที่ 2.9 การทำไดเลชันบนภาพไบนารี.....	29
รูปที่ 2.10 การทำอีโรชันบนภาพไบนารี.....	30
รูปที่ 2.11 การกระทำตัวดำเนินการแบบเปิดและแบบปิด.....	32
รูปที่ 2.12 การทำไดเลชันและอีโรชันแสดงในรูปแบบ 1 มิติ	33
รูปที่ 2.13 อัลกอริทึมสันปันน้ำ 1 มิติ	35
รูปที่ 2.14 การสร้างเขื่อน.....	36
รูปที่ 3.1 โครงสร้างของแนวคิดวิทยานิพนธ์.....	38
รูปที่ 3.2 โครงสร้างการแยกภาพส่วนทางพื้นที่	38
รูปที่ 3.3 โครงสร้างการจัดวางตำแหน่งไปต์.....	39
รูปที่ 3.4 แสดงตัวอย่างผลลัพธ์จากอัลกอริทึมสันปันน้ำ.....	39
รูปที่ 3.5 โครงสร้างการแยกส่วนเชิงเวลา	40
รูปที่ 3.6 แสดงวิธีการสแกน.....	41
รูปที่ 3.7 ตัวอย่างมาสก์ที่ได้จากการแยกส่วนภาพเชิงเวลา.....	42
รูปที่ 3.8 ตัวอย่างมาสก์ที่ได้จากกระบวนการติดตามวัตถุ.....	42

ภาพประกอบ	หน้า
รูปที่ 3.9 ตัวอย่างมาส์กที่ได้จากการกระทำ OR.....	43
รูปที่ 3.10 ข้อมูลขององค์ประกอบสีที่ใช้เป็นเวกเตอร์ลักษณะเด่น.....	43
รูปที่ 3.11 ตัวอย่างมาส์กที่ได้การตัดสีในใจในรูปแบบของภาพสี.....	44
รูปที่ 3.12 โครงสร้างของกระบวนการข้างหลัง.....	45
รูปที่ 3.13 ตัวอย่างมาส์กที่ได้จากการกระบวนการข้างหลัง (มาส์กของวัตถุ).....	46
รูปที่ 3.14 กรรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee Myoung Ho Lee, Chieteuk และ Yo-Sung Ho	47
รูปที่ 3.15 กรรรมวิธีที่นำเสนอโดย ดัชกรณม์ ต้นเจริญ และคณะ.....	50
รูปที่ 4.1 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 1 ถึง 6.....	54
รูปที่ 4.2 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 6.....	54
รูปที่ 4.3 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 6.....	55
รูปที่ 4.4 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Claire	56
รูปที่ 4.5 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสีว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Claire โดยเปรียบเทียบกันทั้งสี่วิธี.....	57
รูปที่ 4.6 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสีในใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา	58
รูปที่ 4.7 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 1 ถึง 9.....	59
รูปที่ 4.8 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9.....	59
รูปที่ 4.9 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 6.....	60
รูปที่ 4.10 ผลการแยกส่วนวัตถุได้ครบเต็มตัว	60
รูปที่ 4.11 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Mother&Daughter.....	61
รูปที่ 4.12 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสีว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Mother&Daughter โดยเปรียบเทียบกันทั้งสี่วิธี.....	62
รูปที่ 4.13 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสีในใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	63
รูปที่ 4.14 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 31 ถึง 39.....	64
รูปที่ 4.15 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 31 ถึง 39.....	64
รูปที่ 4.16 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 31 ถึง 39	65

ภาพประกอบ	หน้า
รูปที่ 4.17 ผลการแยกส่วนวัตถุได้ครบเต็มตัว.....	66
รูปที่ 4.18 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Grandmother.....	66
รูปที่ 4.19 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Grandmother โดยเปรียบเทียบกันทั้งสี่วิธี.....	67
รูปที่ 4.20 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	68
รูปที่ 4.21 ผลการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ เฟรมที่ 2 ถึง 10.....	69
รูปที่ 4.22 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 2 ถึง 10.....	70
รูปที่ 4.23 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 2 ถึง 10.....	70
รูปที่ 4.24 ผลการแยกส่วนวัตถุได้ครบเต็มตัว.....	71
รูปที่ 4.25 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Suzie.....	72
รูปที่ 4.26 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Suzie โดยเปรียบเทียบกันทั้งสี่วิธี.....	73
รูปที่ 4.27 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	73
รูปที่ 4.28 ผลการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ เฟรมที่ 3 ถึง 11.....	74
รูปที่ 4.29 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 3 ถึง 11.....	75
รูปที่ 4.30 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 3 ถึง 11.....	76
รูปที่ 4.31 ผลการแยกส่วนวัตถุได้ครบเต็มตัว.....	76
รูปที่ 4.32 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak7.....	77
รูปที่ 4.33 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Sak7 โดยเปรียบเทียบกันทั้งสี่วิธี.....	78
รูปที่ 4.34 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	79
รูปที่ 4.35 ผลการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ เฟรมที่ 1 ถึง 9.....	80
รูปที่ 4.36 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9.....	80
รูปที่ 4.37 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9.....	81
รูปที่ 4.38 ผลการแยกส่วนวัตถุได้ครบเต็มตัว.....	82

ภาพประกอบ	หน้า
รูปที่ 4.39 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak35.....	82
รูปที่ 4.40 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Sak35 โดยเปรียบเทียบกันทั้งสี่วิธี.....	83
รูปที่ 4.41 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	84
รูปที่ 4.42 ผลการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ เฟรมที่ 1 ถึง 9.....	85
รูปที่ 4.43 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9.....	86
รูปที่ 4.44 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9.....	86
รูปที่ 4.45 ผลการแยกส่วนวัตถุได้ครบเต็มตัว.....	87
รูปที่ 4.46 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak39.....	88
รูปที่ 4.47 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ของลำดับภาพวิดีโอ Sak39 โดยเปรียบเทียบกันทั้งสี่วิธี.....	89
รูปที่ 4.48 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพ โดยวิธีพิจารณาทางสายตา.....	89

บัญชีคำศัพท์

Affine Model	แบบจำลองแอฟฟายน์
Alternative Hypothesis	สมมติฐานอื่น
Background Music	เสียงดนตรีพื้นหลัง
Bayesian Decision	การตัดสินใจแบบเบย์
Bitstream	สายข้อมูลบิต
Cathode-Ray	รังสีแคโทด
Closing Operator	ตัวดำเนินการแบบปิด
Color-Difference	ความแตกต่างกันของสี
Color Model	แบบจำลองสี
Conditional Risk	ค่าความเสี่ยงเงื่อนไข
Decision Process	กระบวนการตัดสินใจ
Dilation	ไดเลชัน
Dynamic Sprite	สไปรท์พลวัต
Erosion	อีโรชัน
Fault Rate	อัตราความผิดพลาด
F-Distribution	การแจกแจงแบบ F
Feature Vector	เวกเตอร์ลักษณะเด่น
Gamma correction	การแก้ไขแกมมา
Gibbs Formula	สูตรของกิบส์
Global Operation	ตัวดำเนินการแบบโกลบอล
HSI	Hue Saturation Intensity
Hough Transform	ฮูทรานสฟอร์ม
Hypothesis Test	การทดสอบสมมติฐาน
Joint likelihood function	ฟังก์ชันความน่าจะเป็นไปได้ร่วม
Least Square	กำลังสองน้อยที่สุด
Markov Random Field	มาร์คอฟแรนดอมฟิลด์

Maximize	แมกซิไมซ์
Minimize	มินิไมซ์
Morphological Gradient	เกรเดียนต์เชิงสัญฐาน
Morphological Operator	ตัวดำเนินการเชิงสัญฐาน
Motion Compensation	การชดเชยการเคลื่อนที่
Motion Estimation	การประมาณการเคลื่อนที่
Motion Vector	เวกเตอร์การเคลื่อนที่
Moving Object	วัตถุเคลื่อนไหว
Null Hypothesis	สมมติฐานหลัก
Object Tracking	การติดตามวัตถุ
Opening Operator	ตัวดำเนินการแบบเปิด
Preprocessing	กระบวนการก่อน
Postprocessing	กระบวนการข้างหลัง
Posteriori Probability	ความน่าจะเป็นภายหลัง
Prior Probability	ความน่าจะเป็นก่อน
Red Green Blue	RGB
Robust Motion Estimation	การประมาณการเคลื่อนที่แบบมั่งคงทนทาน
Spatial Segmentation	การแยกส่วนภาพทางพื้นที่
Sprite	สไปรท์
Static Sprite	สไปรท์สถิตย
Subjective	อัตวิสัย
Temporal Segmentatin	การแยกส่วนภาพเชิงเวลา
Video Browsing	วิดีโอเบราว์เซอร์
Video Database	ฐานข้อมูลวิดีโอ
Watershed algorithm	อัลกอริทึมสันปันน้ำ

บทที่ 1

บทนำ

มาตรฐานการเข้ารหัสสัญญาณวิดีโอ MPEG-4 นั้น มีรูปแบบใหม่ที่เพิ่มมาจาก MPEG-2 คือ มีส่วนของการเข้ารหัสแบบวัตถุ นั้นหมายความว่า สามารถที่จะเข้ารหัสภาพหรือเสียงโดยตัวของแต่ละวัตถุเอง เช่น เข้ารหัสวัตถุเคลื่อนไหว (moving object) แยกจากภาพพื้นหลัง (background) หรือเสียงของวัตถุนั้นๆแยกออกจากเสียงดนตรีพื้นหลัง (background music) การเข้ารหัสแบบวัตถุนี้ สามารถทำให้เกิดการใช้งานใหม่ๆมากมาย เช่น ง่ายที่จะเข้าถึงสายข้อมูลบิต (bitstream) การโต้ตอบกับวัตถุ สามารถเปลี่ยนฉากหลังได้ และสามารถนำวัตถุอื่นมาประกอบกันซึ่งเหมาะกับการประยุกต์ทางมัลติมีเดีย ทั้งนี้มาตรฐานใหม่ เช่น MPEG-7 ซึ่งมีการประยุกต์ทางการจัดการข้อมูลมัลติมีเดีย เช่น วิดีโอเบราว์เซอร์ (video browsing) ฐานข้อมูลวิดีโอ (video database) เป็นต้น ก็ยังจำเป็นต้องใช้ข้อมูลจากการแยกส่วน ดังนั้นการแยกส่วนภาพวิดีโอออกเป็นวัตถุจึงมีความสำคัญและจำเป็น

ในบทนี้จะอธิบายแนวทางของงานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุ แนวทางของวิทยานิพนธ์ เป้าหมายและขอบเขตของวิทยานิพนธ์ ขั้นตอนและวิธีดำเนินงานวิจัย ประโยชน์ที่คาดว่าจะได้รับ และเค้าโครงวิทยานิพนธ์

1.1 แนวทางของงานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุ

งานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุนั้น เริ่มขึ้นประมาณช่วงปี ค.ศ. 1980 เป็นต้นมา โดยช่วงเริ่มแรกนั้น เป็นการนำเทคนิคที่ใช้กับภาพนิ่งมาประยุกต์ใช้กับภาพวิดีโอในแต่ละเฟรม การแยกส่วนภาพวิดีโอออกเป็นวัตถุที่มีความหมายนั้น เริ่มวิจัยกันอย่างแพร่หลายตั้งแต่ปี ค.ศ.1994 เป็นต้นมา เนื่องจากว่าเป็นปีที่มีการร่างมาตรฐานการเข้ารหัสวิดีโอ MPEG-4 ทั้งนี้เพราะว่ามาตรฐาน MPEG-4 เป็นมาตรฐานที่ไม่เพียงแต่กล่าวถึงการบีบอัดข้อมูลเท่านั้น ยังครอบคลุมถึงการเข้ารหัสแบบวัตถุด้วย เพราะฉะนั้นจึงเป็นจุดที่ทำให้งานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุที่มีความหมายแพร่หลายมากขึ้น และยังคงมีความสำคัญจนถึงปัจจุบัน เพราะมาตรฐาน MPEG-7 ที่เริ่มร่างมาตั้งแต่ปี ค.ศ.1997 ได้กล่าวถึงการจัดการข้อมูลมัลติมีเดียเชิงวัตถุอย่างชัดเจน และจุดสำคัญอีกประการคือ ผลลัพธ์จากการแยกส่วนภาพโดยกรรมวิธีต่าง ๆ นั้น แม้ว่าจะได้ผลดีก็ตามแต่ยังคงมีข้อจำกัดอยู่มากเมื่อต้องการจะนำไปใช้งานจริง เพราะฉะนั้น การแยกส่วนภาพวิดีโอออกเป็นวัตถุจึงยังคงเป็นประเด็นที่สำคัญและสามารถทำการศึกษาต่อไปได้อีกมาก

งานวิจัยด้านการแยกส่วนวิดีโอสามารถแบ่งตามแนวทางอย่างคร่าวๆ ดังนี้

- การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติ

การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติ เป็นการนำภาพวิดีโอมาทำการแยกส่วน โดยที่ไม่ต้องใช้ผู้ใช้ช่วยกำหนดข้อมูลบางอย่าง โดยกรรมวิธีของแต่ละบทความนั้น อาจจะทำได้ทั้งเฟรม สองเฟรมติดกัน หรือทำทุกเฟรมซ้ำหลายรอบก็ได้

- การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบกึ่งอัตโนมัติ

การแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบกึ่งอัตโนมัติ เป็นการนำภาพวิดีโอมาทำการแยกส่วน โดยที่ต้องใช้ผู้ใช้ช่วยกำหนดข้อมูลพื้นฐานบางอย่างเช่น สี ตำแหน่ง หรือรูปร่างอย่างคร่าวๆ ของวัตถุ เป็นต้น

ทั้งสองแนวทางนี้ยังสามารถพัฒนาต่อไปในเชิงของเวลาจริง (real time) ได้อีกด้วย ในส่วนถัดไป จะกล่าวถึงงานวิจัยที่ผ่านมาและกรรมวิธีของแต่ละงานที่น่าเสนออย่างคร่าวๆ

ในปี ค.ศ. 1988 Chen Sipei, Li Zongjie และ Zhang Guilin ได้เสนอกรรมวิธีการสกัดวัตถุเคลื่อนที่ออกจากพื้นหลังในระบบติดตามวิดีโอแบบเวลาจริง [1] 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

ในปี ค.ศ. 1990 Fabric HEITZ และ Patrick BOUTHMY เสนอการประมาณเคลื่อนที่และการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้วิธีโกลบอลเบย์เซียน [2] ซึ่งกรรมวิธีดังกล่าวนี้เป็นการหาขอบเขตของวัตถุที่เคลื่อนไหวภายในเฟรม โดยใช้ข้อมูลจากการตรวจจับขอบของภาพนิ่ง และข้อมูลจากการประมาณการเคลื่อนที่ ซึ่งการหาข้อมูลเหล่านั้นจะกระทำผ่านทางแบบจำลองเชิงสถิติมาร์คอฟแรนดอมฟิลด์ (Markov Random Field)

กรรมวิธีนี้ให้ผลลัพธ์และความเที่ยงตรงสูงเมื่อเป็นลำดับภาพที่สังเคราะห์ขึ้นเองโดยเทียบกับวิธีหาขอบโดยการไหลของแสง (Optical Flow) วิธีอื่นๆ แต่เมื่อนำมาใช้กับลำดับภาพจริงยังคงให้ผลที่ดีเฉพาะตรงบริเวณที่เคลื่อนไหวเท่านั้น

ในปี ค.ศ. 1991 Chiou-Shann Fuh และ Petros Maragos เสนอการตรวจจับการเคลื่อนที่ของภาพโดยใช้ แบบจำลองแอฟฟายน์ (Affine Models) [3] ซึ่งเป็นแบบจำลองการ

เคลื่อนที่ของวัตถุ (ซึ่งจะนำมาใช้กับทั้งภาพก็ได้) 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) เป็นตำแหน่งของจุดภาพในเฟรมอ้างอิง A_1 ถึง A_6 เป็นค่าพารามิเตอร์ที่จะทำให้เกิดการเคลื่อนที่ตามแบบจำลองแอฟฟายน์ การหาเวกเตอร์การเคลื่อนที่ของแบบจำลองแอฟฟายน์ทำได้โดยทำให้ฟังก์ชันความผิดพลาดมีค่าน้อยที่สุดคือ

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

โดย I_{i+1} คือความเข้มแสงของภาพเฟรมที่ i และ I_i คือความเข้มแสงของภาพเฟรมที่ $i+1$

จากกรรมวิธีที่นำเสนอได้ทำการทดลองกับภาพที่สังเคราะห์การเคลื่อนไหวขึ้น ซึ่งผลลัพธ์ที่ได้มีความราบเรียบกว่าการประมาณด้วยการไหลของแสงโดยวิธีอื่นๆ และยังสามารถนำไปประยุกต์ใช้กับภาพอินฟราเรดแสดงการเคลื่อนไหวของเมฆด้วย ส่วนข้อด้อยของกรรมวิธีนี้คือ ต้องการการคำนวณสูงมาก และยังคงต้องสอดคล้องกับข้อสมมติที่ว่า ภาพ (หรือวัตถุ) ต้องหมุนไปในทิศทางเดียวกัน

ในปี ค.ศ. 1992 Samir Shaltaf และ Nuder M. Namazi เสนอการใช้ EM (Expectation Maximization) ประมาณการเคลื่อนที่ของวัตถุซึ่งมีการเคลื่อนไหวตามจำลองแอฟฟายน์ [4] การประมาณโดยวิธี EM นั้นมีความซับซ้อนสูงและต้องการการคำนวณมาก ซึ่งต้องการการคำนวณไม่ต่ำกว่า 250 รอบการคำนวณ ผลลัพธ์จากกรรมวิธีนี้มีการผิดพลาดจากค่าจริงไม่เกินร้อยละ 5 ผลการทดลองที่นำเสนอเป็นการใช้กับภาพจริงที่สังเคราะห์การเคลื่อนที่เอง

ในปี ค.ศ. 1994 John Y.A. Wang และ Edward H. Adelson เสนอการแทนภาพเคลื่อนไหวด้วยการแยกออกเป็นชั้นๆ [5] กรรมวิธีนี้เป็นการหาเวกเตอร์การเคลื่อนที่ โดยสมมติว่าการเคลื่อนที่นั้นเป็นไปตามแบบจำลองแอฟฟายน์ แล้วจัดกลุ่มการเคลื่อนที่ที่มีลักษณะเหมือนกันไว้เป็นกลุ่มเดียวกัน ซึ่งแต่ละกลุ่มจะแยกเป็นชั้นๆ

การหาเวกเตอร์การเคลื่อนที่ระดับจุดภาพ โดยสมมติการเคลื่อนที่ (หรือการไหลของแสง) ของภาพด้วยสมการ

$$I_t(x - V_x(x, y), y - V_y(x, y)) = I_{t+1}(x, y) \quad (1.4)$$

แล้วหาคำตอบโดยวิธีกำลังสองน้อยที่สุด (Least Square) ดังสมการ

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum -I_x I_t \\ \sum -I_y I_t \end{bmatrix} \quad (1.5)$$

โดย I_x, I_y, I_t คืออนุพันธ์ย่อยของความเข้มแสงของภาพที่ตำแหน่ง (x, y) และเวลา t ตามลำดับ กรรมวิธีการหาเวกเตอร์การเคลื่อนที่ทำตามวิธีของ Shizawa และ Mare [5] จากนั้นก็แยกเวกเตอร์การเคลื่อนที่ออกเป็นกลุ่มๆ ตามแบบจำลองแอฟฟายน์โดยใช้วิธีทดสอบสมมติฐาน (Hypothesis testing) ในการทดสอบความเชื่อมั่นนั้นเมื่อได้กลุ่มของเวกเตอร์การเคลื่อนที่แล้วก็นำมากำหนดว่ากลุ่มไหนเป็นชั้นใด ผลการทดลองแสดงให้เห็นว่าสามารถแยกได้เป็นชั้นๆ แต่ในลำดับภาพ MPEG Calendar นั้น วัตถุที่เป็นปฏิทินยังไม่สามารถแยกออกจากพื้นหลังได้ และไม่แน่ใจว่าสามารถใช้แยกกับลำดับวิดีโอชนิดที่เป็นวิดีโอคอนเฟอร์เรนซ์หรือไม่ อย่างไรก็ตามสามารถนำผลการทดลองไปประยุกต์กับงานอื่นๆ ได้อีก เช่นการบีบอัดข้อมูลและการแยกวัตถุอย่างมีความหมาย แต่ก็ต้องแลกกับความซับซ้อนในการคำนวณที่สูงมาก

ในปี ค.ศ. 1995 Raynard O. Hinds และ Thrasyvoulos N. Pappas เสนอกรรมวิธีการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยสามารถปรับเปลี่ยนการแบ่งกลุ่มได้ [6] แนวคิดของวิธีนี้คือนำกรรมวิธีที่ใช้กับภาพใบนารี 2 มิติ ไปใช้กับภาพวิดีโอทุกภาพ โดยกรรมวิธีที่ใช้มีทั้งเชิงพื้นที่และเชิงเวลา

ผลลัพธ์ของกรรมวิธีที่เสนอได้ออกมาเป็นภาพใบนารีสเก็ตซ์ ในการกระทำแต่ละครั้งจะต้องใช้ภาพทั้งหมดทุกเฟรม เพราะฉะนั้นไม่สามารถเป็นเวลาจริงได้และต้องการการคำนวณสูงมากแต่กรรมวิธีที่เสนอเป็นแบบอัตโนมัติ ซึ่งไม่ใช่คนเกี่ยวข้องเลย

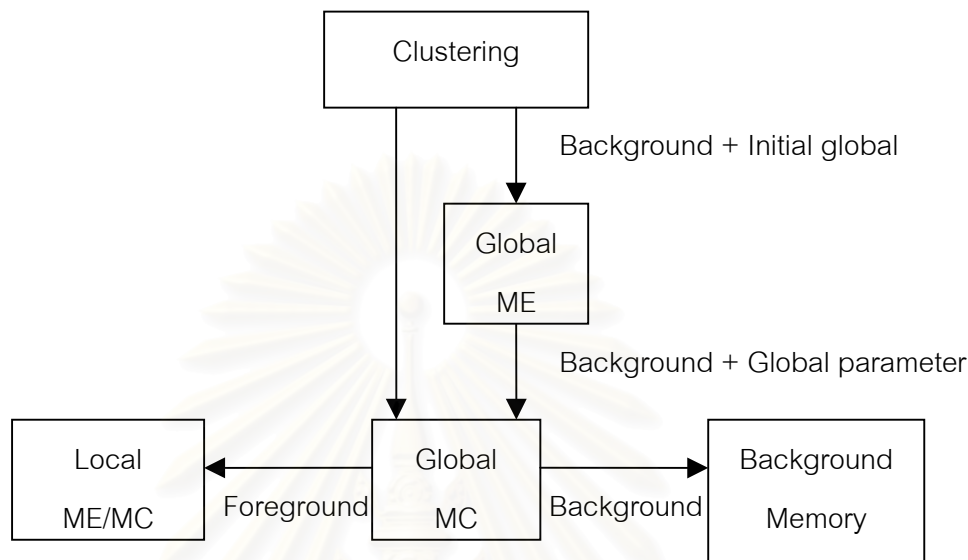
Fabrice Mosheni, Frederic Dufaux และ Murat Kunt เสนอการประมาณการเคลื่อนที่ของพื้นหน้าและพื้นหลังแยกกัน โดยทำการประมาณการเคลื่อนที่ทั้งแบบโกลบอล (Global) และแบบท้องถิ่น (Local) แนวคิดของกรรมวิธีสามารถแสดงให้เห็นได้ดังรูปที่ 1.1 [7]

เริ่มแรกนำภาพมาแยกออกเป็น 2 ส่วน คือ ส่วนพื้นหน้าและส่วนพื้นหลัง โดยวิธี k-mean จากนั้นนำพื้นหลังมาประมาณการเคลื่อนที่แบบโกลบอล (Global) และนำมาชดเชยการเคลื่อนที่กับพื้นหน้า แล้วเก็บพื้นหลังไว้เป็นตัวอย่างในเฟรมถัดไป ส่วนพื้นหน้านำไปหาการประมาณการเคลื่อนที่และการชดเชยการเคลื่อนที่แบบเข้าคู่บล็อก การประมาณการเคลื่อนที่แบบโกลบอล ทำโดยใช้โครงสร้างพีระมิด เกาส์

ผลลัพธ์ที่ได้จากการทำกรรมวิธีนี้ วัดสมรรถนะโดยเทียบค่าความผิดพลาดกำลังสองเฉลี่ย (Mean Square Error) กับการประมาณการเคลื่อนที่แบบคลาสสิกอื่นๆ กรรมวิธีที่นำเสนอให้ค่าอัตราบิตที่ต่ำกว่าในทุกๆ เฟรมของลำดับภาพที่ใช้ทดสอบ

Jean-Mare Odobez และ Patrick Bouthemy เสนอการใช้การเคลื่อนที่ช่วยในการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยสมมติว่าพื้นหลังเป็นไปตามแบบจำลองมาร์คอฟแรนดอมฟิลด์ ซึ่งแบบจำลองของการประมาณการเคลื่อนที่นั้นจะใช้แบบจำลอง 2 มิติ และประมาณการเคลื่อนที่แบบทนทาน (Robust estimation) [8] ผลลัพธ์ที่ได้จากการกระทำโดยแนวความคิดนี้ สามารถ

แยกส่วนตัวคนหรือวัตถุเคลื่อนที่อื่นๆ ออกมาจากพื้นหลังได้ดีพอสมควร แต่การคำนวณค่อนข้างซับซ้อนยุ่งยาก และมีค่าตัวแปรบางค่าต้องการจากการทดลอง ซึ่งไม่ตรงกันตามแต่ละชนิดของวิดีโอ



รูปที่ 1.1 กรรมวิธีที่เสนอโดย Fabrice Mosheni, Frederic Dufaux และ Murat Kunt [7]

Alatan A. Aydin และ Levent Onural เสนอการประมาณการเคลื่อนที่ 2 มิติ และการประมาณโครงสร้างโดยพิจารณาวัตถุเป็นพื้นฐาน [9] แนวคิดคือ หาค่าพารามิเตอร์ของการเคลื่อนที่แบบ 2 มิติ ซึ่งการหาค่าพารามิเตอร์นั้นใช้สูตรกิบส์ (Gibbs formula) เมื่อหาค่าพารามิเตอร์ได้แล้ว จะได้วัตถุออกมา แล้วนำวัตถุที่ได้นั้นมาทำการประมาณการเคลื่อนที่แบบ 3 มิติ ซึ่งจะได้ค่าพารามิเตอร์ของการเคลื่อนที่แบบ 3 มิติออกมา แล้วนำไปหาความลึกของวัตถุนั้นอีกครั้ง กรรมวิธีนี้เมื่อใช้กับภาพสังเคราะห์โดยคอมพิวเตอร์แล้วสามารถหาค่าพารามิเตอร์ทั้ง 2 มิติ และ 3 มิติ ได้ถูกต้อง แต่เมื่อใช้กับภาพจริงกลับหาค่าพารามิเตอร์ได้ถูกเป็นบางส่วน ซึ่งไม่สามารถระบุวัตถุทั้งตัวออกมาได้

ในปี ค.ศ. 1997 Michael M.Chang, Murat Tekalp และ M. Ibrahim Sezan เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุพร้อมกับการประมาณการเคลื่อนที่ [10] แนวคิดของกรรมวิธีนี้คือนำข้อมูลของการประมาณเคลื่อนที่ (เวกเตอร์การเคลื่อนที่) มาจัดกลุ่มเพื่อระบุว่ากลุ่มของเวกเตอร์ที่เหมือนกันกลุ่มใดคือวัตถุใด กลุ่มใดคือพื้นหลัง และอยู่บริเวณไหน

ผลลัพธ์จากการทดลองได้ผลดีกับลำดับวิดีโอมาตรฐาน แต่ภาพที่สังเคราะห์การเคลื่อนที่เองยังคงได้ผลไม่ดีนัก ข้อจำกัดของกรรมวิธีนี้อยู่บางข้อคือ ต้องกำหนดจำนวนวัตถุตั้งแต่แรก และมีความซับซ้อนสูงมาก

Fernando Pereira ได้กล่าวถึงข้อดีของการเข้ารหัสวิดีโอแบบพื้นฐานคอนเทนท์ ซึ่งชี้ให้เห็นว่าจะเกิดความสะดวกรวดเร็วกว่าผู้ให้บริการในโครงข่ายและจัดการข้อมูลได้สะดวกกว่าแบบเดิม ทั้งยังสามารถลดแบนด์วิดท์ได้ด้วย [11]

Georgi D. Borshukov, Gozde Bozdagi, Yucel Altunbasak และ A. Murat Tekalp เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ [12] โดยแนวคิดคล้ายกับ Michael M. Chang และคณะ [10] แต่กรรมวิธีมีความซับซ้อนน้อยกว่า

ในปี ค.ศ. 1998 Thomas Meier และ King N. Ngan เสนอการแยกส่วนวิดีโอแบบอัตโนมัติสำหรับระนาบวิดีโอ (Video object plan) [13] แนวคิดของกรรมวิธีนี้คือใช้ Hausdorff distance ในการติดตามวัตถุ

ผลลัพธ์ที่ได้จากกรรมวิธีนี้ค่อนข้างเฉพาะมากกับลำดับวิดีโอที่นำมาทดสอบ ถ้านำมาใช้กับลำดับวิดีโออื่นๆ ทั่วไปจะได้ผลไม่ดี แต่มีข้อดีคือความซับซ้อนในการคำนวณน้อย

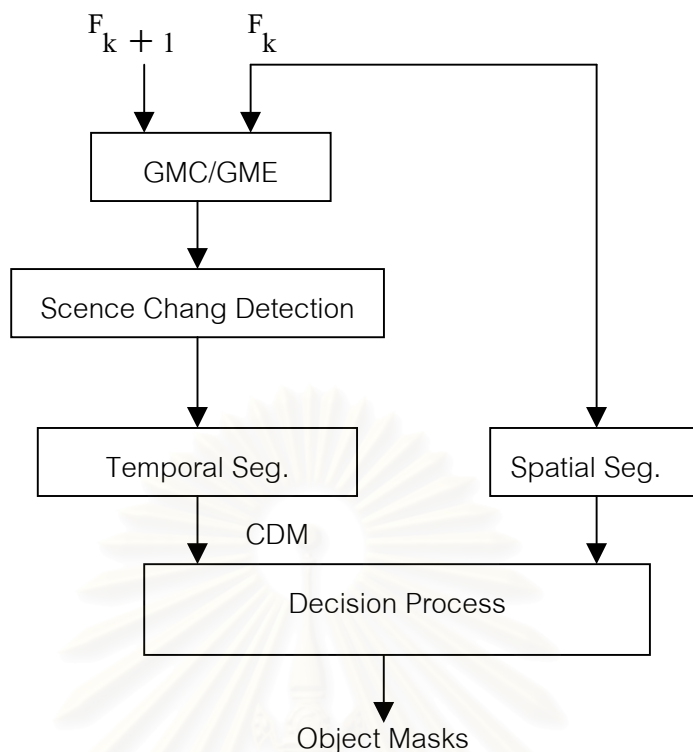
Chee Sun Won เสนอการแยกส่วนภาพเชิงพื้นที่ โดยการหา Maximum a posteriori (MAP) แบบพื้นฐานบล็อก [14] แนวคิดของกรรมวิธีนี้คือ แบ่งภาพนิ่งออกเป็นพื้นที่ย่อย โดยใช้บล็อกเป็นพื้นฐานในการแบ่งกลุ่ม แล้วทำการ optimal block label realization ผลลัพธ์ที่ออกมาแสดงการแยกส่วนทางพื้นที่ได้ โดยแบ่งตามระดับที่เรากำหนดเองได้

ในปี ค.ศ. 1999 Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee Myoung Ho Lee, Chieteuk และ Yo-Sung Ho เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุแบบอัตโนมัติโดยพิจารณาข้อมูลเชิงพื้นที่และเชิงเวลาร่วมกัน [15]

โครงสร้างของกรรมวิธีนี้แสดงดังรูปที่ 1.2 นำภาพเฟรมที่ $k+1$ และเฟรมที่ k มาประมาณการเคลื่อนที่แบบโกลบอล และขจัดขบวนการเคลื่อนที่แบบโกลบอล จากนั้นนำไปตรวจจับหาการเปลี่ยนฉาก แล้วผ่านกระบวนการแยกส่วนภาพเชิงเวลา เพื่อนำมาตัดสินใจว่าพื้นที่ใดเป็นวัตถุ พื้นที่ใดเป็นพื้นหลัง โดยพิจารณาเทียบกับภาพที่ได้จากการแยกส่วนภาพทางพื้นที่

การประมาณการเคลื่อนที่แบบโกลบอลทำได้โดยประมาณการเคลื่อนที่ที่ใช้แบบจำลองแอฟฟายน์ ซึ่งหาเวกเตอร์การเคลื่อนที่โดยใช้วิธีของกำลังสองน้อยที่สุด การแยกส่วนทางเวลาทำได้โดยใช้การทดสอบสมมติฐาน ซึ่งจำเป็นต้องรู้ข้อมูลของพื้นหลังก่อนด้วย การแยกส่วนภาพเชิงพื้นที่ใช้กรรมวิธีสันปันน้ำ (Watershed Algorithm) การตัดสินใจก็พิจารณาว่าภาพที่ได้จากการแยกส่วนเชิงพื้นที่ พื้นที่ใดอินเตอร์เซกกับพื้นที่ที่ได้จากการแยกส่วนเชิงเวลา ถ้าครอบคลุมมากกว่าร้อยละ 90 ก็ให้ตัดสินใจว่าเป็นวัตถุ

ผลลัพธ์จากการทดลองสามารถแยกส่วนได้ แต่ที่นำมาเสนอในบทความคือภาพที่เป็นวิดีโอคอนเฟอร์เรนซ์เพียง 2 ลำดับภาพวิดีโอมาตรฐาน



รูปที่ 1.2 กรรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee
Myoung Ho Lee, Chietek และ Yo-Sung Ho [15]

Aljoscha Smolic, Thomas Sikora และ Jens-Rainer Ohm เสนอการประมาณการเคลื่อนที่แบบ Long-Term การประยุกต์ใช้สำหรับการเข้ารหัสแบบสไปรท์ (Sprite coding) การอธิบายคอนเทนต์และการแยกส่วนภาพวิดีโอออกเป็นวัตถุ [16]

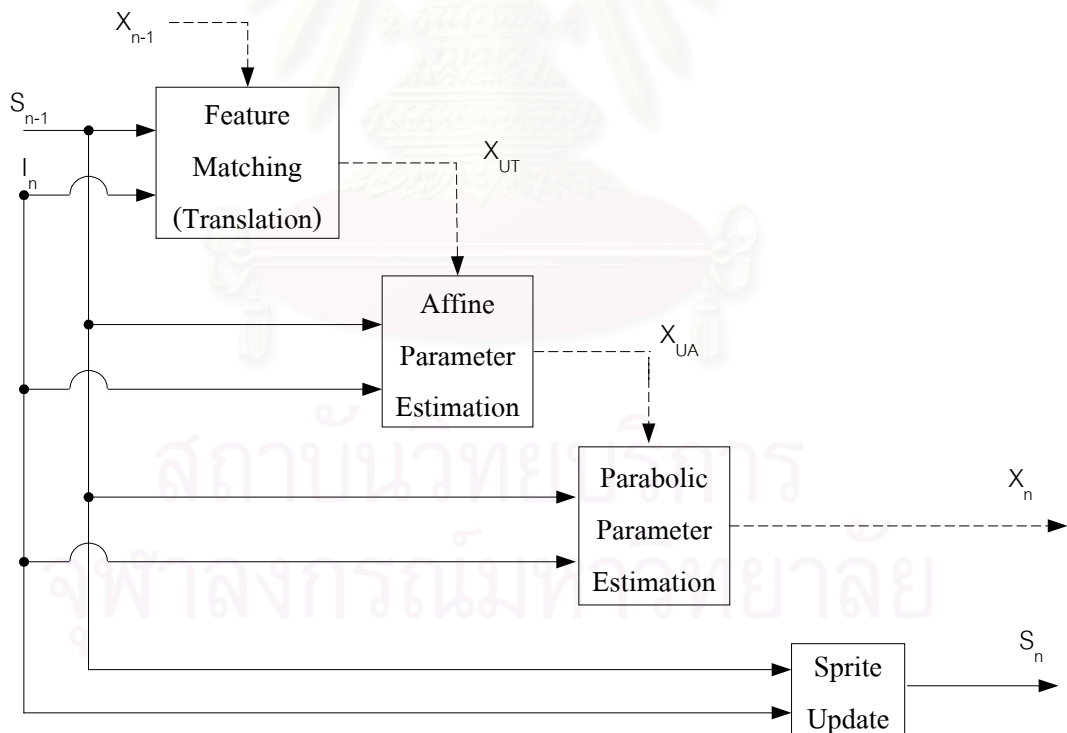
สไปรท์ (Sprite) คือการนำพื้นหลังของภาพวิดีโอทุกเฟรมมาทำเป็นภาพมุกกว้าง การที่จะกระทำเช่นนี้ได้พื้นหลังต้องมีการเคลื่อนที่ในมุกกว้าง คือมีการเคลื่อนที่ของกล้อง ซึ่งการนำภาพพื้นหลังมาทำเป็นพื้นหลังแบบมุกกว้างเป็นความต้องการของทั้ง MPEG-4 Version 1 และ Version 2 ซึ่ง Version 1 นั้นจะเป็น สไปรท์สถิตย์ (Static Sprite) ซึ่งเป็นการกระทำแบบออฟไลน์ (Off-Line) ไม่เหมาะสมกับการกระทำแบบเวลาจริง และเป็น Long-Term parameter ส่วน Version 2 จะเป็นสไปรท์พลวัต (Dynamic Sprite) ซึ่งเป็นออนไลน์ (On-Line) การขจัดเซกการเคลื่อนที่เป็นแบบโกลบอลและเป็น Short-Term parameter ไม่ว่าจะเป็สไปรท์สถิตย์หรือสไปรท์พลวัต การประมาณการเคลื่อนที่แบบโกลบอลก็ยังเป็นสิ่งจำเป็นต่อทั้ง 2 แบบ รูปที่ 1.3 แสดงการหาสไปรท์ที่กรรรมวิธีนี้เสนอ

ขั้นตอนเริ่มแรก นำภาพเฟรมที่ n , X_{n-1} และ Sprite ของเฟรมที่ $n-1$ มาหาเวกเตอร์การเคลื่อนที่ X_{UT} โดยการหาจะทำโดยเลือกจุดที่คิดว่าเป็นจุดสำคัญมา N จุดจากนั้นหาโกลบอลทรานสเลชันด้วยวิธี Roburst M-estimator ได้ X_{UT} ออกมา

ขั้นที่สอง นำ X_{UT} ที่ได้มาประมาณการเคลื่อนที่แบบโกลบอล โดยใช้แบบจำลองแอฟฟายน์ได้ X_{UA}

ขั้นที่สาม นำ X_{UA} ที่ได้มาทำการประมาณการเคลื่อนที่แบบโกลบอล โดยใช้แบบจำลองพาราโบลาได้ X_n แบบจำลองพาราโบลาเป็นแบบจำลองการเคลื่อนที่ของภาพแบบ 9 พารามิเตอร์ ต่างจากแบบจำลองแอฟฟายน์ ที่มีเพียง 6 พารามิเตอร์

ขั้นตอนที่สี่ คือนำ X_n ที่ได้มาทำการประมาณภาพสไปรท์เฟรมที่ n ผลลัพธ์จากกรรมวิธีนี้ดีมากสามารถแยกวัตถุออกจากพื้นหลังที่เคลื่อนที่ได้อย่างสมบูรณ์ แต่จากกรรมวิธีที่แสดงให้เห็นจะพบว่ามีความซับซ้อนในการคำนวณสูงมาก อย่างไรก็ตามกรรมวิธีจะให้ผลลัพธ์ที่ดีเมื่อสอดคล้องกับข้อสมมติ คือ แบบจำลองการเคลื่อนที่สอดคล้องกับการเคลื่อนไหวที่ปรากฏจริงในวิดีโอ ขนาดของวัตถุไม่ใหญ่มากเกินไป และวัตถุไม่ได้อยู่ที่ขอบของภาพ



รูปที่ 1.3 กรรมวิธีที่เสนอโดย Aljoscha Smolic, Thomas Sikora และ Jens-Rainer Ohm [16]

ในปี ค.ศ. 2000 Candemir Toklu, A. Murat Tekalp และ A. Tanju Erden เสนอการแยกส่วนวิดีโอแบบกึ่งอัตโนมัติ (Semi - Automatic) ในกรณีที่เกิดการบดบังกันของวัตถุ [17] แนวคิดคือ ใช้โครงข่ายเมช (Mesh) สี และตำแหน่งเข้าช่วยในการแยกส่วน ซึ่งผู้ใช้ต้องช่วยแยกส่วนวัตถุทุกๆ 10 เฟรม

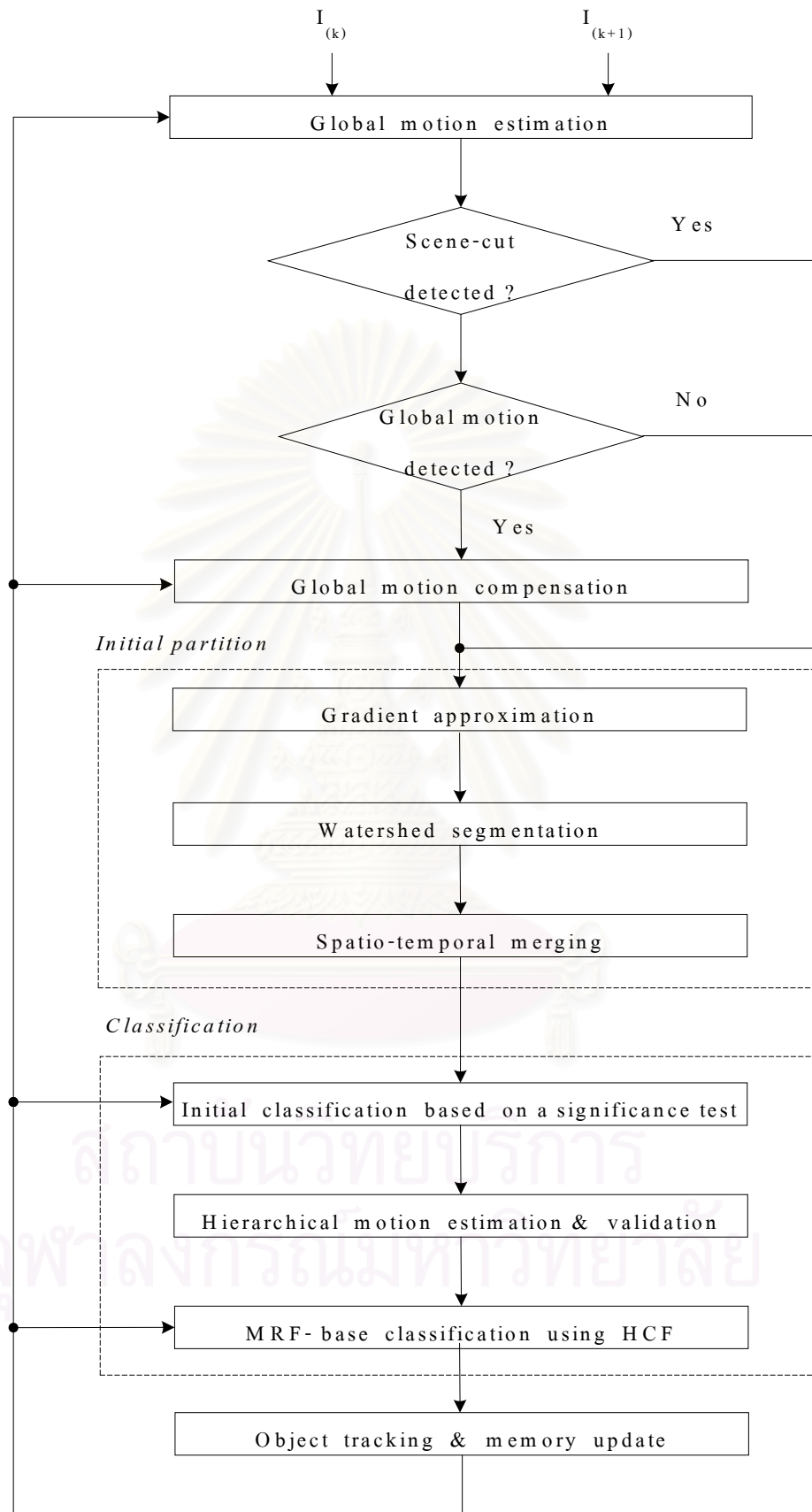
Salkmann Ji และ Hyun Wook Park เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุที่กำลังเคลื่อนไหว โดยพิจารณาบนโดเมน DCT (Discrete Cosine Transform) [18] แนวคิดคือใช้ข้อมูลการเคลื่อนไหวแบบเข้าคู่บล็อกที่มีอยู่แล้วผสมกับข้อมูลที่ได้จากโดเมน DCT นำมาเข้าการแยกส่วน

Hieu T. Nguyen, Marcel Worring และ Anuj Dev เสนอการตรวจจับวัตถุเคลื่อนไหวโดยใช้ข้อมูลการเคลื่อนที่ [19] ซึ่งแนวคิดคล้ายกับ Michael M. Chang และคณะ [10] ดังที่ได้กล่าวไป

ในปี ค.ศ. 2001 Mark Everingham และ Barry Thomas เสนอการแยกส่วนวัตถุที่มีรูปร่างไม่แข็งเกร็ง โดยใช้แบบจำลองของฮิสโทแกรม (histogram) [20] แนวคิดคือให้ผู้ใช้กำหนดกรอบวัตถุอย่างคร่าวๆ โดยวงรี แล้วนำมาหาค่าการกระจายความเข้มแสงของวัตถุนั้น แล้วจึงนำไปใช้กับภาพอื่นๆ ซึ่งภาพทดสอบเป็นภาพของสัตว์ในป่าเป็นส่วนใหญ่

Janez Zaletelj และ Jurij F.Tasic เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้ตัวตรวจจับขอบ [21] แนวคิดคือให้ผู้ใช้ลากเส้นขอบรูปร่างวัตถุก่อน จากนั้นจะใช้ขอบนี้ในการตรวจจับวัตถุในเฟรมต่อไป

ในปี 2002 Yaakov Tsaig และ Amir Averbuch เสนอการแยกส่วนวัตถุเคลื่อนที่แบบอัตโนมัติ โดยใช้เทคนิคการกำหนดชื่อพื้นที่ [22] กรรมวิธีสามารถแสดงได้ดังรูปที่ 1.4 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 จากนั้นตรวจสอบว่ามี การเคลื่อนไหวของกล้องหรือเปล่า ถ้ามีให้ทำการชดเชยแบบโกลบอล จากนั้นนำมาหาเกรเดียนต์ แล้วรวมพื้นที่ย่อยโดยใช้ข้อมูลเชิงพื้นที่และเชิงเวลามาพิจารณาร่วมกัน เพื่อแยกส่วนอย่างคร่าวๆ แล้วนำไปแบ่งส่วน โดยวิธีทางความน่าจะเป็นและสถิติ เริ่มด้วยแบ่งส่วนโดยใช้วิธีทดสอบทางสถิติ แบ่งว่าพื้นที่ใดคือพื้นหน้า พื้นที่ใดคือพื้นหลัง จากนั้นก็ใช้แบบจำลองมาร์คอฟแรนดอมฟิลด์ ในการแบ่งส่วนอย่างละเอียด แล้วทำการมินิไมซ์ (Minimize) ด้วยวิธี HCF (Highest Confident First)



รูปที่ 1.4 กรรณวิธีที่เสนอโดย Yaakov Tsaig และ Amir Averbuch [22]

ผลลัพธ์ที่ได้ใช้ได้ดีกับลำดับวิดีโอมาตรฐานถึง 5 ลำดับวิดีโอ ซึ่งการแยกส่วนเป็นไปอย่างอัตโนมัติ แต่จากการวิเคราะห์ห้ดูแล้ว กรรมวิธีจะให้ผลลัพธ์ที่ดีสำหรับลำดับภาพที่มีการเคลื่อนไหวปานกลางถึงมาก ลำดับภาพที่มีการเคลื่อนไหวน้อยยังไม่นำมาแสดงผล ทั้งนี้กรรมวิธีที่เสนอยังมีความซับซ้อนในการคำนวณสูงมากเนื่องจากการประมาณการเคลื่อนที่ และยังคงต้องการ มิโนมิซ (Minimize) ฟังก์ชันความน่าจะเป็นอีกด้วย

Giusto D.D., F. Massidda และ C. Perra เสนอกรรมวิธีการแยกส่วนภาพวิดีโอออกเป็นวัตถุอย่างรวดเร็ว [23] โดยแนวคิดคือ หาความแตกต่างของภาพแล้วนำไปตรวจจับขอบ ซึ่งก็มีความรวดเร็วในการประมวลผล แต่ผลลัพธ์ที่เสนอก่อนข้างจะเป็นลำดับที่มีความเฉพาะสูงมาก

Ebroul Izquierdo, Jianhui Xia และ Roland Mech ได้อธิบายถึงซอฟต์แวร์ AM 5.0 ของกลุ่ม COST 211 ว่าพัฒนาจนถึง AM 5.3 แล้ว มีความเปลี่ยนแปลงคือสามารถใช้เป็นกราฟิกแบบยูสเซอร์อินเตอร์เฟซ (User Interface) แล้วมีสมบัติต่างๆ เพิ่มขึ้นจากเดิมด้วย [24]

Hui zhu และ Zaiming Li เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยใช้ข้อมูลเชิงพื้นที่และเวลาพิจารณาร่วมกัน [25] แนวคิดก็ไม่ต่างจากหลายบทความที่กล่าวมา ที่แตกต่างคือในส่วนของประมาณการเคลื่อนที่แบบโกลบอลนั้นใช้การค้นหาแบบเต็ม (full search) และหารามิเตอร์ของแบบจำลองแอฟฟายน์ ด้วยวิธีกำลังสองน้อยที่สุด

Jong Bae Kim และ Hang Joon Kim เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ โดยใช้การแปลงเวฟเลทร่วมกับกรรมวิธีสันปันน้ำ [26] แนวคิดคือ ทำการแปลงเวฟเลทกับภาพก่อนแล้วนำส่วนความละเอียดต่ำ (Low Resolution) ของชั้นที่ 2 มาทำด้วยอัลกอริทึมสันปันน้ำ (Watershed Algorithm) เพื่อแยกส่วนภาพ กรรมวิธีมีความรวดเร็วในการประมวลผล แต่ผลลัพธ์ยังคงจำกัดอยู่เฉพาะบางลำดับภาพวิดีโอ

1.2 แนวทางของวิทยานิพนธ์

ในวิทยานิพนธ์นี้ เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้เทคนิคการตัดสินใจแบบเบย์ ซึ่งกรรมวิธีจะเป็นแบบอัตโนมัติ โดยจะนำไปประยุกต์ใช้กับวิดีโอที่มีลักษณะเป็นวิดีโอเฟรมหรือวิดีโอคอนเฟอเรนซ์ ที่ประกอบด้วยวัตถุ 2 วัตถุคือ บุคคลที่อยู่หน้าจอและพื้นหลัง โดยลักษณะของพื้นหลังต้องอยู่นิ่งและกล้องก็ต้องอยู่นิ่งด้วย ซึ่งมีกระบวนการหลักๆ ดังนี้คือ การแยกส่วนภาพเชิงพื้นที่ (Spatial segmentation) การแยกส่วนภาพเชิงเวลา (Temporal segmentation) การติดตามวัตถุ (Object tracking) และกระบวนการตัดสินใจ (Decision process)

1.3 เป้าหมายและขอบเขตวิทยานิพนธ์

1. สามารถแยกส่วนภาพตัวคนกับพื้นหลังที่อยู่นิ่งออกจากกันได้
2. การแยกส่วนภาพเป็นแบบอัตโนมัติ
3. สามารถแยกวัตถุได้อย่างน้อย 3 วิดีโอที่มีลักษณะการเคลื่อนไหวของวัตถุที่ต่างกัน
4. สามารถแยกวัตถุได้เต็มตัวโดยใช้จำนวนเฟรมน้อยกว่าและมีอัตราผิดพลาดเฉลี่ยต่ำกว่า เมื่อเปรียบเทียบกับวิธีการแยกส่วนโดยวิธีทดสอบสมมติฐาน [15] และการแยกส่วนภาพเชิงพื้นที่และเวลา [27]

1.4 ขั้นตอนและวิธีการดำเนินงาน

1. ศึกษาบทความที่เกี่ยวข้องกับวิทยานิพนธ์
 - 1.1) ศึกษาบทความที่เกี่ยวข้องกับเทคนิคการแยกส่วนภาพทั้งแบบอัตโนมัติและกึ่งอัตโนมัติ
 - 1.2) ศึกษาทฤษฎีการตัดสินใจแบบเบย์
2. ทดลองเทคนิคต่างๆ ที่ศึกษามาเพื่อชี้ถึงข้อดีและข้อเสีย
3. เสนอเทคนิคการแยกส่วนภาพที่คาดว่าจะให้ผลลัพธ์ที่ดีและทดสอบเทคนิคนั้น
4. วิเคราะห์ผลของเทคนิคที่นำเสนอถึงข้อดี ข้อเสีย และข้อที่ควรปรับปรุง
5. รวบรวมและสรุปผลการวิจัยเพื่อเขียนวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงกระบวนการและเทคนิคที่เหมาะสมสำหรับการแยกส่วนภาพวิดีโอออกเป็นวัตถุ
2. ทำให้สามารถวางแผนแนวทางที่เหมาะสมว่า การแยกส่วนภาพควรดำเนินไปในแนวทางใด
3. สามารถนำเทคนิคที่เสนอไปประยุกต์ลงบนอุปกรณ์สื่อสารที่สนับสนุนด้านมัลติมีเดียต่อไปในอนาคตได้

1.6 คำโครงวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้แบ่งรายละเอียดออกเป็น 5 บท ดังต่อไปนี้

บทที่ 1 บทนำ กล่าวถึงแนวทางของงานวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุ แนวทางของวิทยานิพนธ์ เป้าหมายและขอบเขตของวิทยานิพนธ์ ขั้นตอนและวิธีดำเนินงานวิจัย ประโยชน์ที่คาดว่าจะได้รับและคำโครงวิทยานิพนธ์

บทที่ 2 ความรู้พื้นฐาน กล่าวถึงพื้นฐานของสี แบบจำลองสี การแปลงจากพิกัดแบบจำลองสี RGB เป็นพิกัดแบบจำลองสี HSI ตัวดำเนินการเชิงสัณฐาน (Morphological Operator) ทฤษฎีการตัดสินใจแบบเบย์ (Bayesian Decision Theory)

บทที่ 3 การแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้เทคนิคการตัดสินใจแบบเบย์ กล่าวถึงการแยกส่วนภาพทางพื้นที่ การแยกส่วนภาพเชิงเวลา กระบวนการตัดสินใจ และการติดตามวัตถุ

บทที่ 4 ผลการวิจัย กล่าวถึง ผลการทดลองเพื่อทดสอบแนวความคิด ผลของเทคนิคที่นำเสนอในเชิงอัตวิสัย (subjective) และแสดงเป็นแผนภูมิเส้นโดยเปรียบเทียบกับวิธีการทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา

บทที่ 5 บทสรุป กล่าวถึง สรุปผลการวิจัย ข้อเสนอแนะ

บทที่ 2

ทฤษฎีพื้นฐาน

2.1 พื้นฐานของสี

แบบจำลองสี (color model)

จุดประสงค์ของการมีแบบจำลองสี คือใช้ในการอ้างอิงและเป็นมาตรฐานเดียวกัน โดยในวิทยานิพนธ์นี้จะกล่าวถึง 5 แบบจำลองสี คือ แบบจำลองสี RGB แบบจำลองสี YIQ แบบจำลองสี YUV แบบจำลองสี YCbCr [28] และแบบจำลองสี HSI [29]

2.1.1 แบบจำลองสี RGB (RGB color model)

ระบบสี RGB (แดง น้ำเงิน เขียว) เป็นระบบพื้นฐานของสีที่ใช้ในคอมพิวเตอร์กราฟิก และหน่วยเก็บความจำของภาพ เพราะว่าหลอดภาพของจอคอมพิวเตอร์นั้นใช้สารเรืองแสงสีแดง สีเขียว และสีน้ำเงิน ในการผลิตสีอื่นๆออกมา นั่นคือสีแดง สีเขียว และสีน้ำเงิน เป็นองค์ประกอบของสีทุกๆสี ซึ่งในการประกอบกันเป็นสีต่างๆนั้นเกิดจากการรวมกันของสามองค์ประกอบดังแสดงในรูปที่ 2.1

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

การแก้ไขแกมมา (Gamma correction)

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

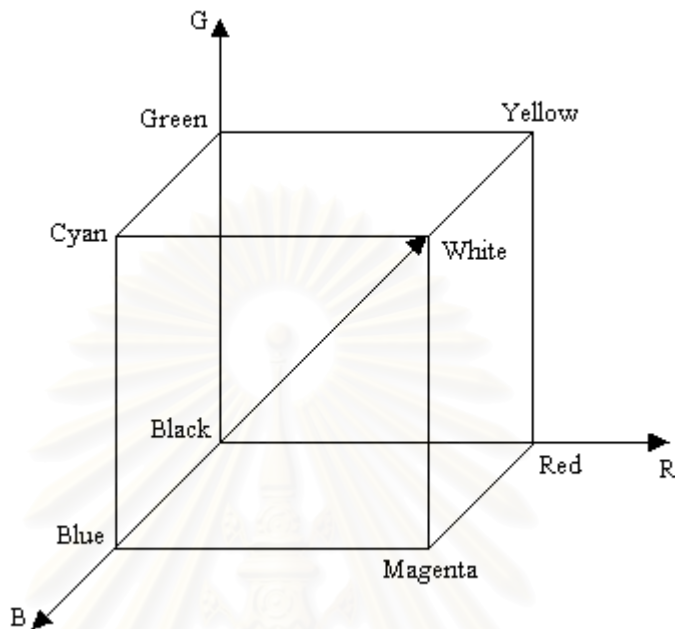
$$I_{\text{แสดงหน้าจอ}} \propto V^\gamma \quad (2.1)$$

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

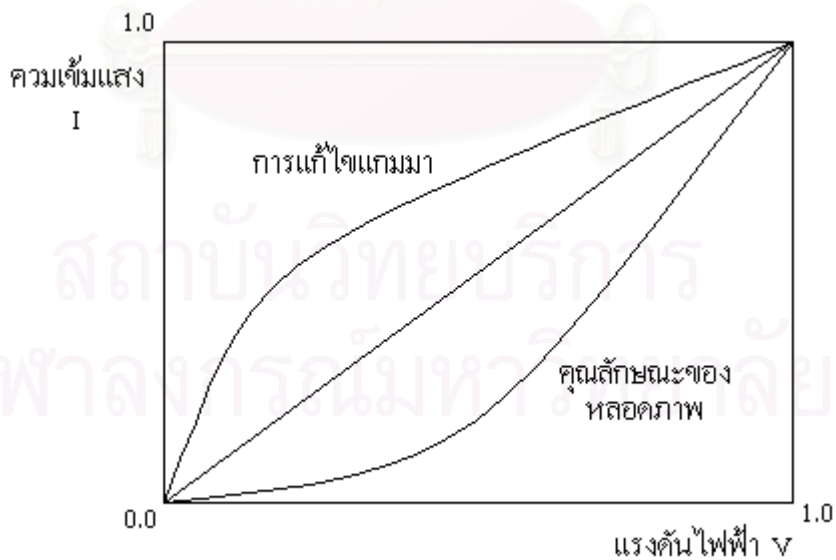
V คือ แรงดันทางไฟฟ้าของสัญญาณภาพที่ได้รับ

γ คือ ค่าแกมมาที่ใช้ในการแสดงผล มีค่าอยู่ในช่วง 2.0 ถึง 3.0 ซึ่งขึ้นกับชนิดของหลอดภาพ โดยทั่วไปจะมีค่า 2.2 สำหรับระบบ NTSC ดังแสดงในรูปที่ 2.2 และมีค่า 2.8 สำหรับระบบ

PAL/SECAM ซึ่งค่าสัญญาณแรงดันอินพุทจะถูกทำการนอร์มอลไลซ์ (nomalize) ให้อยู่ในช่วง 0 ถึง 1



รูปที่ 2.1 พิกัดแบบจำลองสี RGB [28]



รูปที่ 2.2 ลักษณะค่าแกมมาที่ใช้ในการแสดงผล [28]

การชดเชยสำหรับกระบวนการที่ไม่เป็นเชิงเส้นจะทำที่ด้านส่ง ซึ่งข้อมูลเชิงเส้น RGB นั้น จะเรียกว่า การแก้ไขแกมมา (gamma-corrected) การแก้ไขทำได้โดยนำสัญญาณจริงมายกกำลัง 0.45 [28] เช่น $V^{0.45}$ เพราะฉะนั้น เมื่อส่งไปและแสดงผล จะทำให้ค่า $I_{\text{แสดงหน้าจอ}}$ มีค่าเท่ากับ $V^{0.45 \cdot Y}$ ถ้าให้ค่าแกมมาเป็น 2.2 จะได้ $I_{\text{แสดงหน้าจอ}} = V$ เพราะฉะนั้นภาพที่ได้ก็จะเป็นเชิงเส้นต่อการรับชม ประโยชน์ของการแก้ไขแกมมาที่ด้านส่งคือทำให้ต้นทุนของเครื่องรับโทรทัศน์ลดลงและส่วนของ ภาพที่มีดีจะถูกขยายขึ้น

2.1.2 แบบจำลองสี YUV (YUV color model)

แบบจำลองสี YUV คือรูปแบบพื้นฐานที่ใช้โดยมาตรฐานของโทรทัศน์ระบบ NTSC, PAL และ SECAM ซึ่ง Y คือองค์ประกอบระดับความเข้มแสง U และ V คือองค์ประกอบของสี สัญญาณ YUV หาได้จากสัญญาณ RGB ที่แก้ไขแกมมาแล้วดังนี้ [28]

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (2.2)$$

ซึ่งสามารถแสดง U และ V ได้อีกรูปแบบคือ

$$U = 0.492 (B' - Y) \quad (2.3)$$

$$V = 0.877 (R' - Y) \quad (2.4)$$

โดย R', G', B' แสดงถึงสัญญาณ RGB ที่ได้รับการแก้ไขแกมมาแล้ว

ในระบบโทรทัศน์สีเช่น HDTV นั้น สัญญาณ YUV จะหาได้โดยใช้ **เมตริกซ์** การแปลง RGB เป็น YUV ซึ่งโดยทั่วไปแล้ว สัญญาณความแตกต่างกันของสี (U และ V) จะถูกชักตัวอย่างเป็นครึ่งหนึ่งของ Y ในมิติทางขนาดของภาพ เพราะระบบการรับรู้ของมนุษย์ต่อความเข้มแสงจะมีมากกว่าความเข้มสี ซึ่งเป็นประโยชน์อย่างยิ่งในการที่จะช่วยลดอัตราบิตของเทคนิคการเข้ารหัสสัญญาณวิดีโอ

2.1.3 แบบจำลองสี YIQ (YIQ color model)

แบบจำลองสี YIQ สืบทอดมาจาก YUV และเป็นส่วนเสริมของระบบโทรทัศน์แบบ NTSC ซึ่ง I นั้นหมายถึง เฟสตรงกัน (in-phase) Q หมายถึง เฟสต่างกัน 90 องศา (quadrature-phase) [28] โดยมีสมการการแปลงคือ

$$Y = 0.299R' + 0.587G' + 0.114B' \quad (2.5)$$

$$I = 0.596R' - 0.275G' - 0.321B' \quad (2.6)$$

$$= 0.736(R' - Y) - 0.268(B' - Y) \quad (2.7)$$

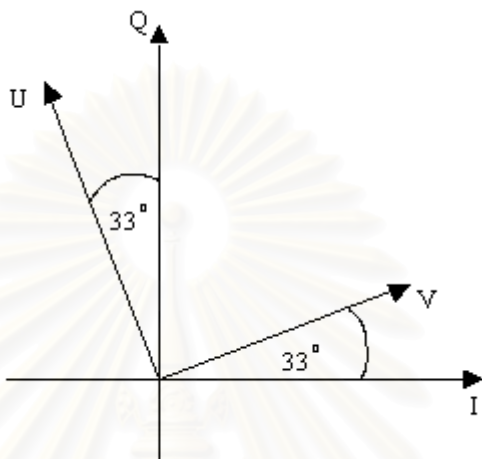
$$Q = 0.212R' - 0.523G' + 0.311B' \quad (2.8)$$

$$= 0.478(R' - Y) + 0.413(B' - Y) \quad (2.9)$$

ซึ่งสามารถเขียนได้อีกรูปแบบคือ

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos 33^\circ & \sin 33^\circ \\ -\sin 33^\circ & \cos 33^\circ \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \quad (2.10)$$

จากรูปที่ 2.3 จะเห็นได้ว่า พิกัด UV จะต่างจากพิกัด IQ อยู่ 33 องศา

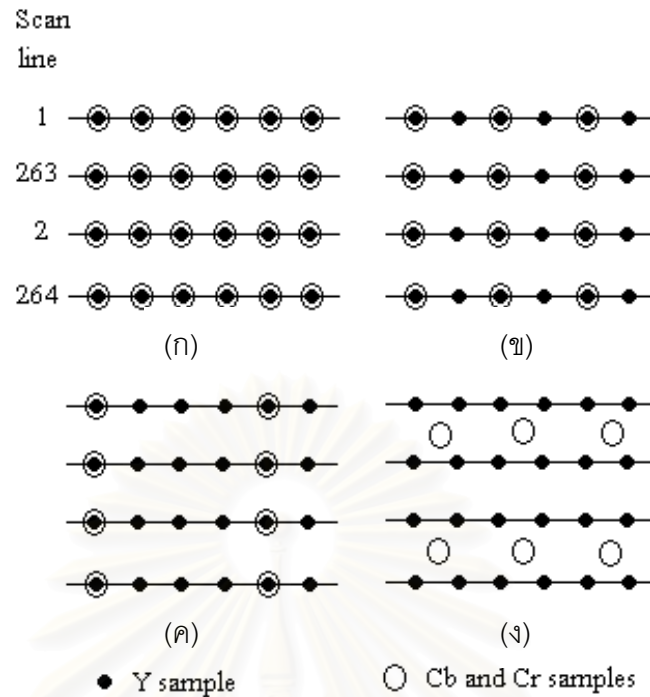


รูปที่ 2.3 ความแตกต่างของพิกัด UV และพิกัด IQ

2.1.4 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 ซึ่งเป็นความถี่ของคลื่นพาห่อย่อยสำหรับระบบ NTSC ตารางที่ 2.1 แสดงลักษณะการเข้ารหัสสัญญาณวิดีโอระบบ 525 เส้นและระบบ 625 เส้น

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 โดยจะเปลี่ยนเพียงแค่ตำแหน่งของ 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 2.4 ตำแหน่งของจุดตัวอย่างหรือจุดภาพของ YCbCr

(ก) 4:4:4 (ข) 4:2:2 (ค) 4:1:1 (ง) 4:2:0 [28]

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

สัญญาณ YCbCr นั้นหาได้จากการแปลงดังนี้

$$Y = 0.299R' + 0.587G' + 0.114B' \quad (2.11)$$

$$Cb = -0.169R' - 0.331G' + 0.500B' \quad (2.12)$$

$$Cr = 0.500R' - 0.419G' - 0.081B' \quad (2.13)$$

สัญญาณความแตกต่างกันของสีเป็นไปตามสมการ

$$(B - Y) = -0.299R' - 0.587G' + 0.886B' \quad (2.14)$$

$$(R - Y) = 0.701R' - 0.587G' - 0.114B' \quad (2.15)$$

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

ตารางที่ 2.1 ลักษณะของการเข้ารหัสสัญญาณวิดีโอระบบ 525 เส้น และระบบ 625 เส้น [28]

	ระบบ 525 เส้น 60รอบต่อวินาที	ระบบ 625 เส้น 50 รอบต่อวินาที
การเข้ารหัสสัญญาณ	YCbCr	
จำนวนจุดภาพต่อเส้น :		
● ความเข้มแสง (Y)	858	846
● ความแตกต่างกันของสี(Cb,Cr)	429	432
โครงสร้างการชักตัวอย่าง	ออร์โทโกนอล เส้น ฟิลด์ และ เฟรม Cb และ Crถูกชักตัวอย่าง ในตำแหน่งเลขคี่ (1, 3, 5, ...) ของ Y ทุกเส้น	
การชักตัวอย่างความถี่ :		
● ความเข้มแสง	13.5 เมกะเฮิร์ตซ์	
● ความแตกต่างกันของสี	6.75 เมกะเฮิร์ตซ์	
รูปแบบการเข้ารหัส :	การควอนไทซ์แบบยูนิฟอร์ม PCM	
จำนวนของจุดภาพต่อ	8 บิตต่อจุดภาพ	
เส้นแอกที่พิดิจิทัล :		
● ความเข้มแสง	720	
● ความแตกต่างกันของสี	360	
ระดับการควอนไทซ์ :		
● สเกล	0-255	
● ความเข้มแสง	220 ระดับ (16 (สีดำ) ถึง 235 (สีขาว))	
● ความแตกต่างกันของสี	225 ระดับ (16 ถึง 240 ค่า 128 จะตรงกับ 0 ในสเกล 0 ถึง 255)	

จุฬาลงกรณ์มหาวิทยาลัย

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

$$\bar{Y} = 219Y + 16 \quad (2.16)$$

$$\bar{Cb} = 224[0.564(B-Y)] + 128 = 126(B-Y) + 128 \quad (2.17)$$

$$\bar{Cr} = 224[0.713(B-Y)] + 128 = 160(B-Y) + 128 \quad (2.18)$$

ถ้าค่าที่ออกมาเป็นทศนิยมแล้วจะปัดเป็นค่าจำนวนเต็มทีใกล้เคียงที่สุด สังเกตว่าถ้าค่า Cb เป็น 0 จะแปลงเป็นจำนวนเต็มคือ 128 หรือสามารถเขียนการแปลงได้อีกแบบคือ

$$\bar{Y} = 219 Y + 16 \quad (2.19)$$

$$\bar{Cb} = 224Cb + 128 \quad (2.20)$$

$$\bar{Cr} = 224Cr + 128 \quad (2.21)$$

2.1.5 แบบจำลองสี HSI (HSI color model)

HSI (Hue Saturation Intensity) ประกอบด้วย 3 องค์ประกอบเช่นเดียวกับแบบจำลองสีแบบอื่น โดย Hue คือ คุณสมบัติของสีที่อธิบายถึงความบริสุทธิ์ของสี (สีเหลือง สีส้ม หรือสีแดง) ขณะที่ความอิ่มตัวของสี (Saturation) เป็นการวัดองศาความบริสุทธิ์ของสีที่ถูกเจือโดยแสงสีขาว เช่นถ้ามีแสงสีขาวอยู่มากจะมีค่าความอิ่มตัวน้อย แต่ถ้ามีแสงสีขาวเจืออยู่น้อยแสดงว่ามีความอิ่มตัวมาก แบบจำลองสี HSI นั้นเป็นที่ใช้กันอย่างแพร่หลายเนื่องจากความสอดคล้องกับกลไกในการรับรู้ของดวงตามนุษย์ที่มีเส้นประสาท 2 ชุดคือ rod และ cone ที่มีการตอบสนองที่ดีต่อความเข้มแสงและสีตามลำดับ กล่าวคือ ระบบ HSI มี I ซึ่งเป็นความเข้มแสงเป็นองค์ประกอบหนึ่งในแบบจำลอง องค์ประกอบ Hue และความอิ่มตัวนั้นมีความสัมพันธ์กับการรับรู้สีของมนุษย์ ทำให้แบบจำลอง HSI เป็นเครื่องมือที่สำคัญในการพัฒนาอัลกอริทึมทางการประมวลผลภาพ ซึ่งการพัฒนานั้นขึ้นกับสมบัติการรับรู้สีของมนุษย์เป็นสำคัญ

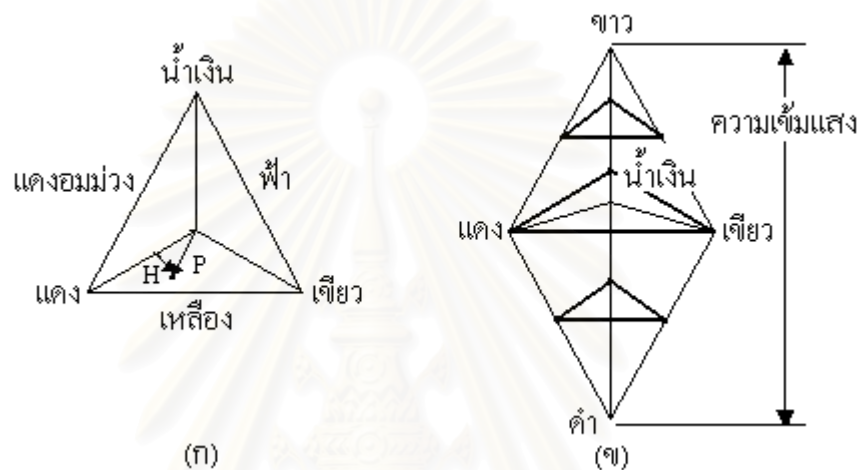
การแปลงจาก RGB เป็น HSI [29]

แบบจำลองสี RGB สามารถแสดงออกมาได้เป็นรูปสี่เหลี่ยมลูกบาศก์ดังรูปที่ 2.1 แต่แบบจำลอง HSI แสดงได้ดังรูปที่ 2.5(ก) โดยที่ความเข้มแสงมีค่าเปลี่ยนแปลงตามแนวแกนที่เชื่อมต่อระหว่างปลายยอดของพีระมิดทั้ง 2 จุด ดังรูปที่ 2.5(ข)

จากรูปที่ 2.5(ก) ที่จุดสี P มีค่า H คือ ค่าที่เป็นมุมนับจากแกนสีแดง ดังนั้นเมื่อ $H = 0$ องศา สีนั้นจะเป็นสีแดง ถ้าเป็นสีเหลือง H จะมีค่า 60 องศา เป็นต้น ความอิ่มตัว S ของจุดสี P คือ องศาของจุดสีนั้นได้ถูกเจือจางแสงสีขาว และความอิ่มตัวเป็นสัดส่วนโดยตรงกับระยะทางจากจุด

P ไปยังจุดเซนทรอยด์ (centroid) ของสามเหลี่ยม ถ้าจุด P ยิ่งไกลจากจุดเซนทรอยด์ของสามเหลี่ยมมากเท่าไรแล้ว ค่าความเข้มตัวของสีก็จะยิ่งมากกว่าเท่านั้น

การวัดระดับความเข้มแสงของแบบจำลอง HSI นั้นวัดได้โดยเทียบจากเส้นตั้งฉาก ซึ่งเส้นตั้งฉากนั้นจะลากผ่านจุดเซนทรอยด์ของสามเหลี่ยม ถ้าเส้นตั้งฉากไปทางปลายล่างของพีระมิดจะมีความเข้มแสงที่ลดลงไปทางสีดำ ส่วนเส้นตั้งฉากไปทางปลายบนของพีระมิดจะมีความเข้มแสงที่เป็นสีขาว



รูปที่ 2.5 พิกัดแบบจำลองสี HSI [29]

การนิยามสีในแบบจำลองสี HSI นั้นจะนิยามบนค่าสีแดง สีเขียว และสีน้ำเงิน ซึ่งถูกนอร์มอลไลซ์แล้ว ค่าสีแดง สีเขียว และสีน้ำเงิน ที่ถูกนอร์มอลไลซ์แสดงตามสมการต่อไปนี้

$$r = \frac{R}{R + G + B} \quad (2.22)$$

$$g = \frac{G}{R + G + B} \quad (2.23)$$

$$b = \frac{B}{R + G + B} \quad (2.24)$$

โดยค่า R, G และ B คือ ค่าของสีแดง สีเขียว และสีน้ำเงิน ที่ได้ทำการนอร์มอลไลซ์ให้อยู่ในช่วง 0 ถึง 1 และค่า r, g และ b นี้ก็อยู่ในช่วง 0 ถึง 1 เช่นกันและยังสอดคล้องกับเงื่อนไข

$$r + g + b = 1 \quad (2.25)$$

โดยค่าความเข้มแสงของแบบจำลองสี HSI หาได้จากการแปลงจาก RGB ได้ดังนี้

$$I = \frac{1}{3}(R+G+B) \quad (2.26)$$

ซึ่งจะมีค่าอยู่ในช่วง 0 ถึง 1

ค่า H และ S นั้นหาได้จากการสร้างสามเหลี่ยมดังในรูปที่ 2.6 ซึ่งในการอธิบายที่มานั้นจำเป็นที่จะต้องกำหนดเงื่อนไขและสัญลักษณ์ดังต่อไปนี้

1) จุด W มีพิกัด $(1/3, 1/3, 1/3)$

2) จุดสี P ใดๆมีพิกัด (r, g, b)

3) w คือ เวกเตอร์ที่ชี้จากจุดกำเนิดไปยังจุด W

p_R คือ เวกเตอร์ที่ชี้จากจุดกำเนิดไปยังจุด P_R

p คือ เวกเตอร์ที่ชี้จากจุดกำเนิดไปยังจุด P

4) เส้น PQ , โดย $i = R, G, B$ สร้างขึ้นให้ผ่านจุด W

5) กำหนดให้ $r_0 = R/l$, $g_0 = G/l$ และ $b_0 = B/l$ โดย l คือค่าความเข้มสีที่หาได้จากสมการที่ (2.26) จากรูปที่ 2.6(ก) จะเห็นได้ว่าบนเส้นตรง P_RQ_R นั้นจะมีค่า $g_0 = b_0$ เช่นเดียวกับเส้นตรง P_BQ_B นั้นมี $g_0 = r_0$ และ เส้นตรง P_GQ_G นั้นมี $b_0 = r_0$

6) จุดทุกจุดในสามเหลี่ยม $P_RQ_RP_G$ จะมีค่า $g_0 \geq b_0$ จุดทุกจุดในสามเหลี่ยม $P_RQ_RP_B$ จะมีค่า $b_0 \geq g_0$ จุดทุกจุดในสามเหลี่ยม $P_RQ_BP_B$ จะมีค่า $r_0 \geq g_0$ เช่น P_RQ_R เป็นเส้นที่แบ่งพื้นที่สามเหลี่ยม $P_RP_BP_G$ ออกเป็นสองส่วน โดยแยกตามเงื่อนไขคือ สามเหลี่ยมบนมีค่า $g_0 > b_0$ และสามเหลี่ยมล่างมีค่า $g_0 < b_0$ P_GQ_G เป็นเส้นที่แบ่งพื้นที่สามเหลี่ยม $P_RP_BP_G$ ออกเป็นสองส่วน โดยแยกตามเงื่อนไขคือ สามเหลี่ยมบนมีค่า $b_0 > r_0$ และ สามเหลี่ยมล่างมีค่า $b_0 < r_0$ เช่น P_BP_B เป็นเส้นที่แบ่งพื้นที่สามเหลี่ยม $P_RP_BP_G$ ออกเป็นสองส่วน โดยแยกตามเงื่อนไขคือ สามเหลี่ยมซ้ายมีค่า $r_0 > g_0$ และ สามเหลี่ยมขวามีค่า $r_0 < g_0$

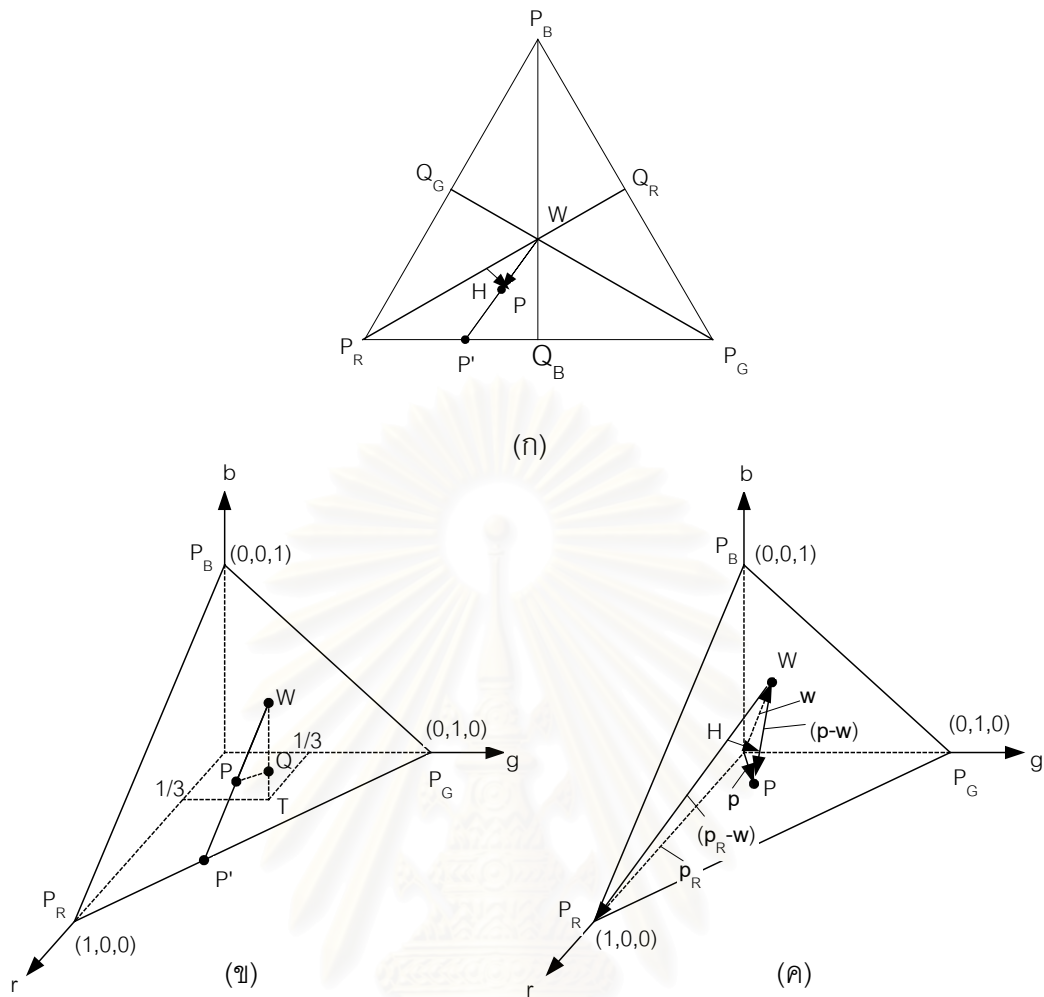
7) $|WQ_i| / |P_iQ_i| = 1/3$ และ $|WP_i| / |P_iQ_i| = 2/3$ โดย $| \cdot |$ คือค่าขนาดของระยะ

8) กำหนดให้ สามเหลี่ยม WP_RP_G เรียกว่า ส่วน RG , สามเหลี่ยม WP_GP_B เรียกว่า ส่วน GB , สามเหลี่ยม WP_BP_R เรียกว่า ส่วน BR

จากรูปที่ 2.6 ค่า Hue ของจุดสีใดๆ คือ ค่ามุมระหว่างส่วนของเส้นตรง WP_R และ WP สามารถเขียนให้อยู่ในรูปเวกเตอร์ดังรูปที่ 2.6(ค) คือมุมระหว่าง $(p_r - w)$ กับ $(p - w)$ เช่น สีแดงมีค่า $H = 0$ องศา ถ้าเป็นสีเขียวมีค่า $H = 120$ องศา เป็นต้น ค่า H ยังสามารถเขียนให้อยู่ในรูปผลการคูณกันภายในได้ดังนี้

$$(p-w) \cdot (p_r - w) = \|p_r - w\| \|p - w\| \cos H \quad (2.27)$$

โดย $\| \cdot \|$ คือ ตัวกระทำนอร์มของเวกเตอร์



รูปที่ 2.6 แสดงตำแหน่งที่ใช้ในการแปลงพิกัดจากแบบจำลองสี RGB เป็น HSI [29]

จากเงื่อนไข (1) และ (2) จะได้

$$\|p-w\| = \left[\left(r - \frac{1}{3}\right)^2 + \left(g - \frac{1}{3}\right)^2 + \left(b - \frac{1}{3}\right)^2 \right]^{1/2} \tag{2.28}$$

แทนค่า r, g และ b จาก สมการที่ (2.22) ถึงสมการที่ (2.26) ลงสมการที่ (2.28)

$$\|p-w\| = \left[\frac{9(R^2 + G^2 + B^2) - 3(R + G + B)}{9(R + G + B)^2} \right] \tag{2.29}$$

p_R มีพิกัด (1,0,0) และ w มีพิกัด (1/3,1/3,1/3) เพราะฉะนั้น

$$\|p_R-w\| = \left(\frac{2}{3}\right)^{1/2} \tag{2.30}$$

จากนิยามผลคูณภายในของ 2 เวกเตอร์ $a \cdot b = a^T b = a_1 b_1 + a_2 b_2 + a_3 b_3$ แล้ว

$$(p-w) \cdot (p_R - w) = \frac{2}{3} \left(r - \frac{1}{3}\right) - \frac{1}{3} \left(g - \frac{1}{3}\right) + \frac{1}{3} \left(b - \frac{1}{3}\right) \quad (2.31)$$

$$= \frac{2R - G - B}{3(R + G + B)} \quad (2.32)$$

จากสมการที่ (2.27)

$$H = \arccos \left[\frac{(p-w) \cdot (p_R - w)}{\|p-w\| \|p_R - w\|} \right]^{-1} \quad (2.33)$$

$$H = \frac{1}{2} \frac{[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{1/2}} \quad (2.34)$$

ค่า H นี้อยู่ในช่วง 0 ถึง 180 องศา เท่านั้น ถ้า $b_0 > g_0$ แล้ว H จะมากกว่า 180 องศา ดังนั้น จึงใช้ $H = 360 - H_{\text{น้อยกว่า } 180}$ เพื่อหาค่าออกมา (การหาผลคูณภายในของเวกเตอร์นั้น มุมระหว่างเวกเตอร์จะอยู่ในช่วง 0 ถึง 180 องศา ค่า H ที่หามาได้ก็จะอยู่ในช่วง 0 ถึง 180 องศาเช่นกัน แต่ถ้า $b_0 < g_0$ แล้ว ค่ามุมระหว่างเวกเตอร์ $(p_R - w)$ และ $(p-w)$ ก็อยู่ในช่วง 0 ถึง 180 แต่ค่า H เป็นค่าที่นับจากแกนสีแดง เพราะฉะนั้นเมื่อ P อยู่ในสามเหลี่ยม $P_R Q_R P_B$ (ดังนั้นค่า H จึงต้องลบออกจาก 360)

S คือค่าความอิมิตัวของสี จากรูปที่ 2.6(ข) จุด T คือจุดของ W ที่ฉายลงบนระนาบ rg นิยามของ S คือ

$$S = \frac{|WP|}{|WP'|} = \frac{|WQ|}{|WT|} = \frac{|WT| - |QT|}{|WT|} \quad (2.35)$$

เมื่อ $|WT| = 1/3$ และ $|QT| = b$ ดังนั้นในส่วนของ RG

$$S = 3 \left(\frac{1}{3} - b \right) \quad (2.36)$$

$$= 1 - 3b \quad (2.37)$$

$$= 1 - b_0 \quad (2.38)$$

จากเงื่อนไข (5) พบว่า b_0 คือค่า $\min(r_0, g_0, b_0)$ ภายในส่วนของ RG ดังนั้น

$$S = 1 - \min(r_0, b_0, g_0) \quad (2.39)$$

$$= 1 - \frac{3}{R + B + G} [\min(R, G, B)] \quad (2.40)$$

ดังนั้นสามารถสรุปสมการการแปลงได้ดังนี้

$$I = \frac{1}{3} (R + G + B) \quad (2.41)$$

$$S = 1 - \frac{3}{R+B+G} [\min(R, G, B)] \quad (2.42)$$

$$H = \frac{\frac{1}{2} [(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{1/2}} \quad (2.43)$$

การแปลงจาก HSI เป็น RGB

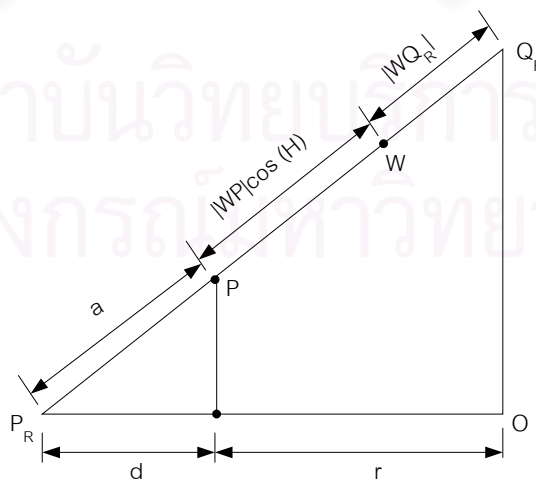
ค่า H จะอยู่ในช่วง 0 ถึง 1 คือ 0 ถึง 360 องศา ซึ่งสามารถแยกเป็นส่วนๆได้ตามองศาคือ ส่วน RG ส่วน GB และ ส่วน BR

ส่วน RG ($0 < H \leq 120$) จากสมการที่ (2.36)

$$b = \frac{1}{3}(1 - s) \quad (2.44)$$

ค่า r คือค่าของจุด P ฉายลงบนแกนสีแดง พิจารณาสามเหลี่ยม $P_R O Q_R$ ดังรูปที่ 2.7 โดย O คือจุดกำเนิดในระบบพิกัด rgb เส้นประในรูปแสดงการฉายจุด P ลงบนแกนสีแดงซึ่งจุดที่ฉายลงมามีค่า r จากจุด P ถึงจุด W มีระยะ $|WP| \cos H$ (ดูรูปที่ 2.6 ประกอบ) ด้วยสมบัติของสามเหลี่ยมคล้ายจะได้

$$\frac{|P_R Q_R|}{|P_R O|} = \frac{a}{d} \quad (2.45)$$



รูปที่ 2.7 การฉายจุด P ลงบนแกนสีแดง [29]

จาก $|P_R O| = 1$, $d = 1 - r$ และ $a = |P_R Q_R| - (|WP| \cos H + |WQ_R|)$ แทนลงในสมการที่ (2.45) ได้

$$r = \frac{|WQ_R|}{|P_R Q_R|} + \frac{|WP|}{|P_R Q_R|} \cos H \quad (2.46)$$

$$= \frac{1}{3} + \frac{|WP|}{|P_R Q_R|} \cos H \quad (2.47)$$

จากรูปที่ 2.6(ค) $|P_R Q_R| = 3|WQ_R|$ จากสมการที่ (2.35) $|WP| = S|WP'|$ มุม $P_R WQ_B = 60$ องศา เพราะฉะนั้น $|WQ_B| = |WP'| \cos(60 - H) = |WQ_R|$ เพราะฉะนั้น

$$r = \frac{1}{3} + \frac{S |WQ_R| \cos H}{|P_R Q_R| \cos(60 - H)} \quad (2.48)$$

$$r = \frac{1}{3} \left[1 + \frac{S \cos H}{\cos(60 - H)} \right] \quad (2.49)$$

$$g = 1 - (r + b) \quad (2.50)$$

ดังนั้นสรุปเป็นสมการการแปลงได้ดังนี้

ส่วน RG สำหรับ $0 < H \leq 120$

$$b = \frac{1}{3} (1 - s) \quad (2.51)$$

$$r = \frac{1}{3} \left[1 + \frac{S \cos H}{\cos(60 - H)} \right] \quad (2.52)$$

$$g = 1 - (r + b) \quad (2.53)$$

ส่วน GB $120 < H \leq 240$

$$H = H - 120 \quad (2.54)$$

$$r = \frac{1}{3} (1 - s) \quad (2.55)$$

$$g = \frac{1}{3} \left[1 + \frac{S \cos H}{\cos(60 - H)} \right] \quad (2.56)$$

$$b = 1 - (r + g) \quad (2.57)$$

ส่วน BR $240 < H \leq 360$

$$H = H - 240^\circ \quad (2.58)$$

$$g = \frac{1}{3} (1 - s) \quad (2.59)$$

$$b = \frac{1}{3} \left[1 + \frac{S \cos H}{\cos(60 - H)} \right] \quad (2.60)$$

$$r = 1 - (g + b) \quad (2.61)$$

โดย $R = 3lr$, $G = 3lg$ และ $B = 3lb$

2.2 ตัวดำเนินการเชิงสัณฐาน (Morphological Operator) [29]

ตัวดำเนินการเชิงสัณฐานที่จะกล่าวถึงในวิทยานิพนธ์นี้ประกอบด้วย ไดเลชัน (Dilation) อีโรชัน (Erosion) ตัวดำเนินการแบบเปิด (Opening Operator) ตัวดำเนินการแบบปิด (Closing Operator) และกรรมวิธีสันปันน้ำ (Watershed Algorithm)

2.2.1 ตัวดำเนินการเชิงสัณฐานบนภาพไบนารี

ไดเลชันและอีโรชัน เป็นตัวดำเนินการเชิงสัณฐาน ซึ่งเป็นพื้นฐานของตัวดำเนินการเชิงสัณฐานอื่นๆต่อไป

ให้ A และ B เป็นเซตใน Z^2 โดย Z คือเซตของจำนวนเต็มบวก ที่มีองค์ประกอบ $a = (a_1, a_2)$ และ $b = (b_1, b_2)$ ตามลำดับ การเลื่อนตำแหน่งของ A โดย $x = (x_1, x_2)$ คือ

$$(A)_x = \{c \mid c = a + x, \text{ เมื่อ } a \in A\} \quad (2.62)$$

การสะท้อนของ B คือ

$$\bar{B} = \{x \mid x = -b, \text{ เมื่อ } b \in B\} \quad (2.63)$$

คอมพลีเมนต์ (Complement) ของเซต A คือ

$$A^c = \{x \mid x \notin A\} \quad (2.64)$$

ความแตกต่างกันระหว่างเซต A และ B คือ

$$A - B = \{x \mid x \in A, x \notin B\} \quad (2.65)$$

$$= A \cap B^c \quad (2.66)$$

รูปที่ 2.8 แสดงภาพของการเลื่อนตำแหน่ง การสะท้อน คอมพลีเมนต์ และความแตกต่างกันระหว่างเซต

ไดเลชันของ A โดย B คือ

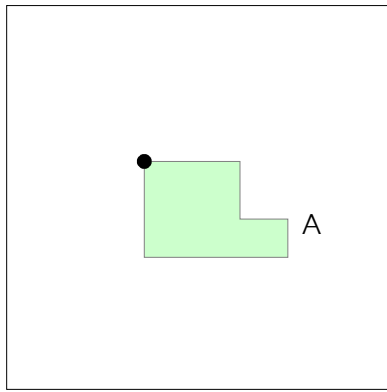
$$A \oplus B = \{x \mid (\bar{B})_x \cap A \neq \phi\} \quad (2.67)$$

โดย A และ B เป็นเซตใน Z^2 , ϕ เป็นเซตว่าง

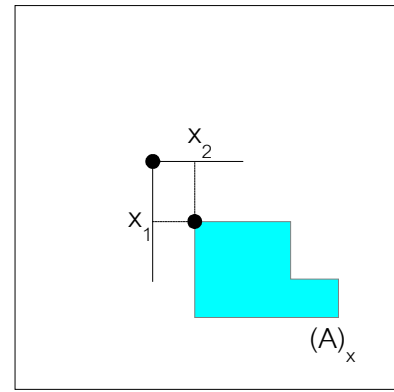
กระบวนการไดเลชันเป็นการสะท้อนเซต B แล้วเลื่อนตำแหน่งไป x ไดเลชันของ A โดย B คือ การอินเตอร์เซกระหว่าง A กับ B ซึ่งสามารถเขียนได้อีกแบบคือ

$$A \oplus B = \{x \mid [(\bar{B})_x \cap A] \subseteq A\} \quad (2.68)$$

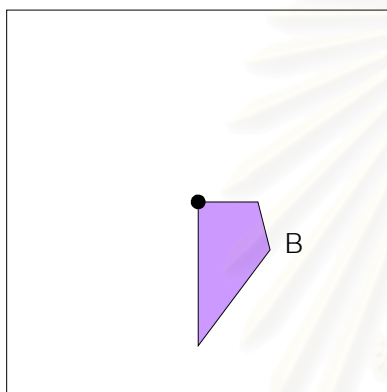
เซต B เรียกว่า โครงสร้างย่อยในการไดเลชัน รูปที่ 2.9 แสดงการทำไดเลชัน



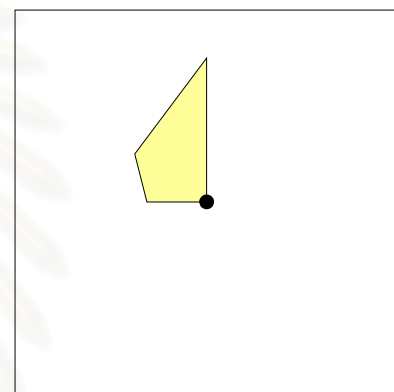
(ก)



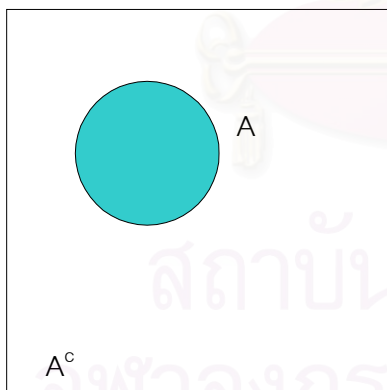
(ข)



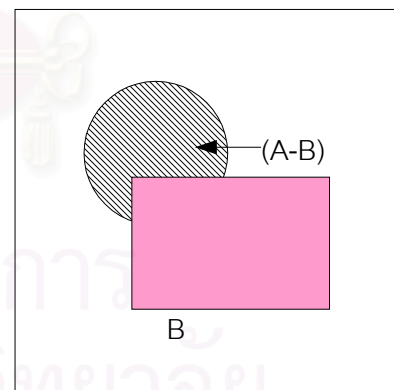
(ค)



(ง)

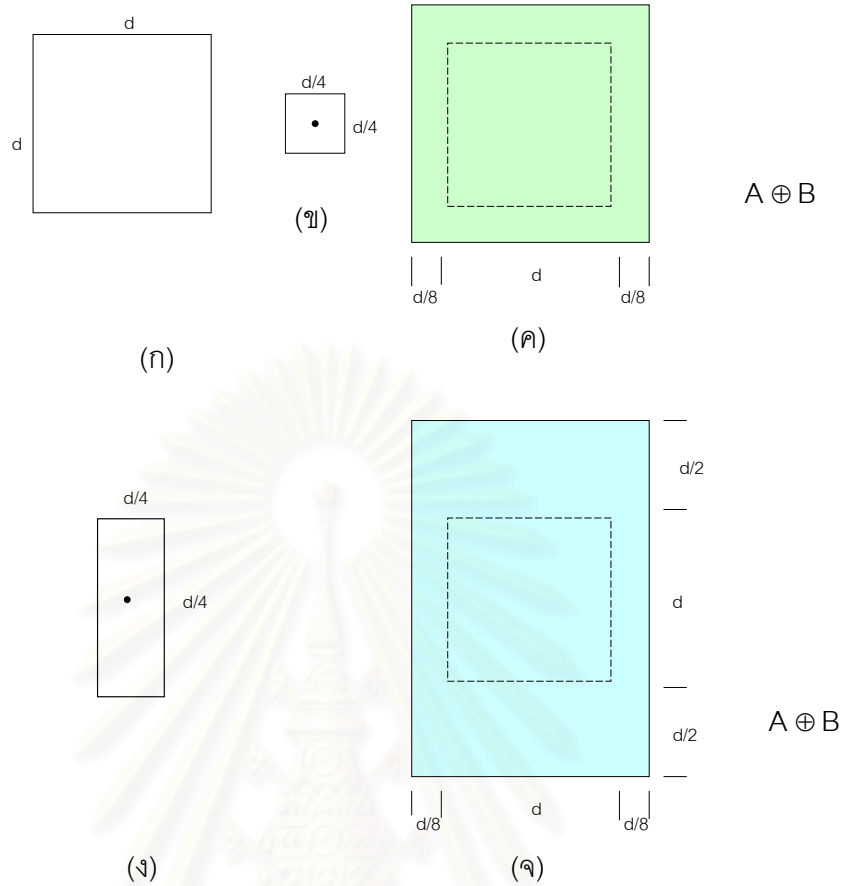


(จ)



(ฉ)

รูปที่ 2.8 ตัวอย่างการเลื่อนตำแหน่ง การสะท้อน คอมพลีเมนต์ และความแตกต่างกันระหว่างเซต
 (ก) ภาพต้นแบบ A (ข) การเลื่อนตำแหน่งของ A (ค) ภาพต้นแบบ B (ง) การสะท้อนของ B
 (ฉ) คอมพลีเมนต์ของ A (จ) ความแตกต่างกันระหว่างเซต A และ B [29]



รูปที่ 2.9 การทำไดเลชันบนภาพไบนารี (ก) เซต A (ข) เซต B แบบที่ 1 (ค) ผลลัพธ์จากการทำไดเลชันต่อเซต A โดยเซต B แบบที่ 1 (ง) เซต B แบบที่ 2 (จ) ผลลัพธ์จากการทำไดเลชันต่อเซต A โดยเซต B แบบที่ 2 [29]

อีโรชันของ A โดย B คือ

$$A \ominus B = \{x \mid (B)_x \subseteq A\} \tag{2.69}$$

โดย A และ B เป็นเซตใน Z^2 อีโรชันของ A โดย B คือ การสะท้อนเซต B แล้วเลื่อนตำแหน่งไป x บนเซต A ซึ่งสมาชิกในเซตนั้นต้องเป็นสับเซตของ A รูปที่ 2.10 แสดงการทำอีโรชันไดเลชันและอีโรชันเป็นคู่การกระทำซึ่งกันและกันโดยพิจารณาสมการต่อไปนี้

$$(A \ominus B)^c = \{x \mid (B)_x \subseteq A\}^c \tag{2.70}$$

$$= \{x \mid (B)_x \cap A^c = \phi\}^c \tag{2.71}$$

$$= \{x \mid (B)_x \cap A^c \neq \phi\} \tag{2.72}$$

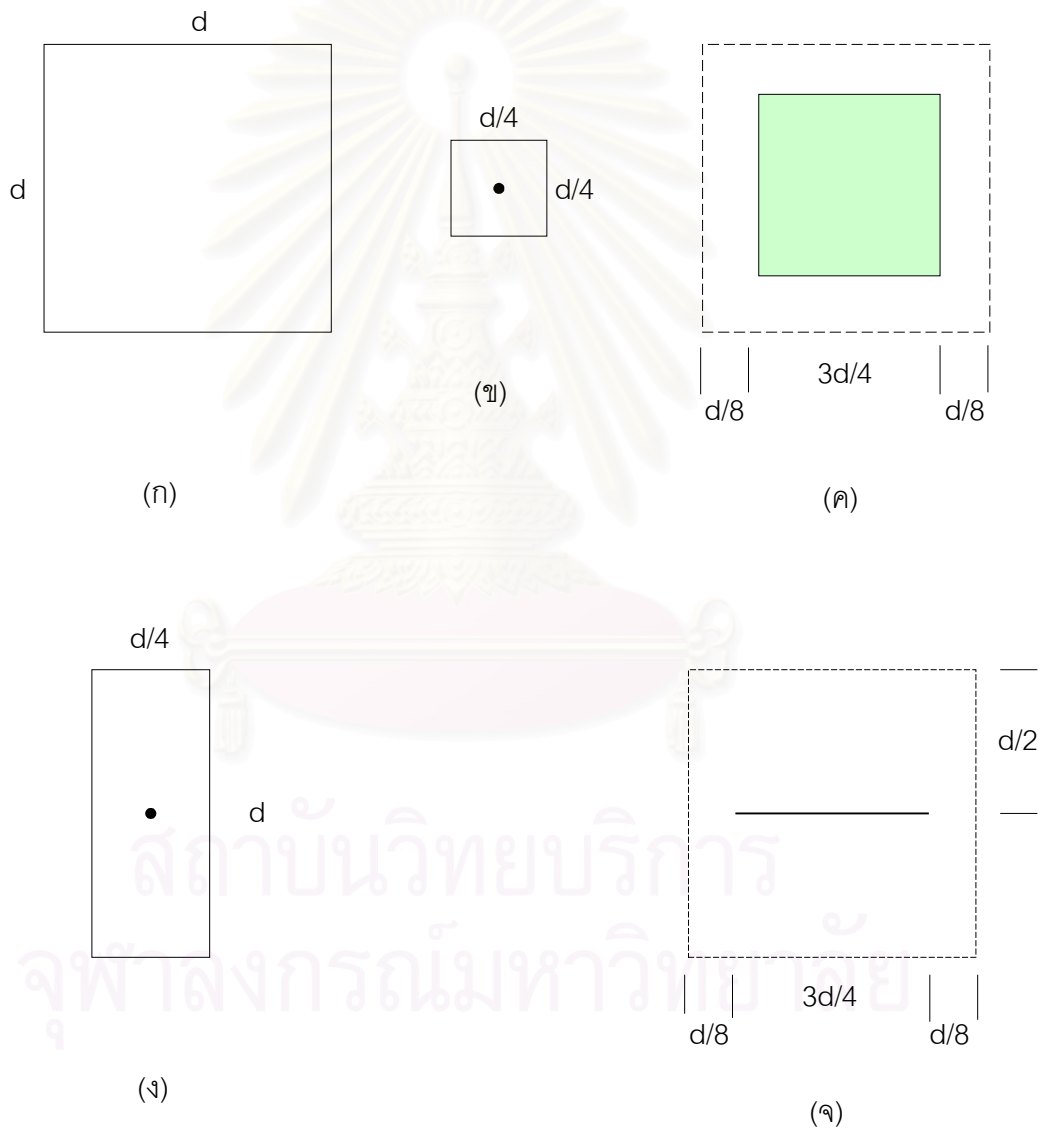
$$= A^c \oplus \overline{B} \tag{2.73}$$

ตัวดำเนินการแบบเปิด (Opening Operator) และ ตัวดำเนินการแบบปิด (Closing Operator)

การกระทำแบบเปิดต่อเซต A โดยเซตโครงสร้างย่อย B คือ

$$A \circ B = (A \ominus B) \oplus B \quad (2.74)$$

ถ้าพิจารณาจากรูปสมการจะพบว่าเป็นการกระทำอีโรชั่นต่อ A โดย B แล้วนำผลที่ได้มากระทำไดเลชันโดย B อีกครั้ง



รูปที่ 2.10 การทำอีโรชั่นบนภาพไบนารี (ก) เซต A (ข) เซต B แบบที่ 1 (ค) ผลลัพธ์จากการทำอีโรชั่นต่อเซต A โดยเซต B แบบที่ 1 (ง) เซต B แบบที่ 2 (จ) ผลลัพธ์จากการทำอีโรชั่นต่อเซต A โดยเซต B แบบที่ 2 [29]

การกระทำแบบปิดต่อเซต A โดยเซตโครงสร้างย่อย B คือ

$$A \cdot B = (A \oplus B) \ominus B \quad (2.75)$$

ถ้าพิจารณาจากรูปสมการจะพบว่าเป็นการกระทำได้เลขชั้นต่อ A โดย B แล้วนำผลที่ได้มากระทำอี
โรชันโดย B อีกครั้ง รูปที่ 2.11 แสดงผลตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดคุณสมบัติ
สมบัติความเป็นคู่ของตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดพิจารณาได้จากสมการดัง
ต่อไปนี้

$$(A \cdot B)^c = [(A \oplus B) \ominus B]^c \quad (2.76)$$

$$= (A \oplus B)^c \oplus \overline{B} \quad (2.77)$$

$$= (A^c \ominus \overline{B}) \oplus \overline{B} \quad (2.78)$$

$$= A^c \circ \overline{B} \quad (2.79)$$

2.2.2 ตัวดำเนินการเชิงสัจฐานต่อภาพระดับสี่เทา

โดเลขชั้นในภาพระดับสี่เทานิยามได้ดังนี้

$$f \oplus b = \max\{f(s-x, t-y) + b(x,y) \mid (s-x), (t-y) \in D_f; (x,y) \in D_b\} \quad (2.80)$$

โดย D_f และ D_b คืออาณาจักรของ f และ b ตามลำดับ b คือโครงสร้างย่อย เงื่อนไขที่ $(s-x)$ และ $(t-y)$
ต้องอยู่ในอาณาจักร f ตรงกับเงื่อนไขของโดเลขชั้นในภาพไบนารีที่เซตสองเซตจะต้องมีส่วนที่
ซ้อนทับกันอย่างน้อย 1 องค์ประกอบย่อย ในการอธิบายจะลดรูปให้เหลือเพียง 1 มิติ ดังนั้นสม
การที่ (2.64) จะลดรูปเหลือดังนี้

$$(f \oplus b) = \max\{f(s-x) + b(x) \mid (s-x) \in D_f; (x) \in D_b\} \quad (2.81)$$

จากรูปที่ 2.15 แสดงให้เห็นการทำโดเลขชั้น โดยจะเหมือนการทำคอนโวลูชัน คือ เลื่อน $b(x)$ ไปบน f
(s) ผลลัพธ์ที่ได้ แสดงดังรูปที่ 2.12(ง)

อีโรชันในภาพระดับสี่เทานิยามดังนี้

$$(f \ominus b)(s,t) = \min\{f(s+x, t+y) - b(x, y) \mid (s+x), (t+y) \in D_f; (x,y) \in D_b\} \quad (2.82)$$

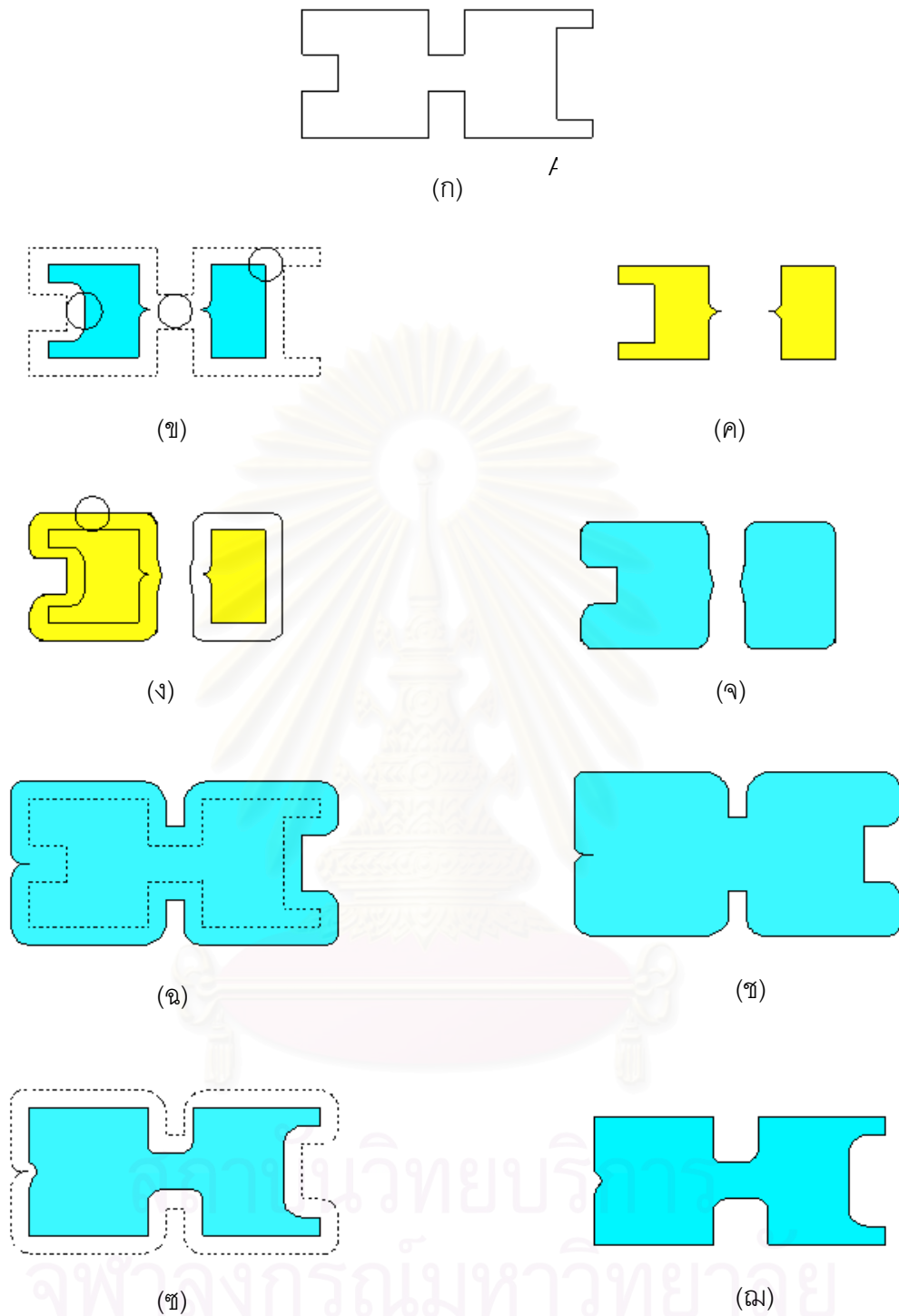
โดย D_f และ D_b เป็นอาณาจักร f และ b ตามลำดับ เงื่อนไขที่ $(s+x)$ และ $(t+y)$ ต้องอยู่ใน
อาณาจักร f เหมือนกับเงื่อนไขของอีโรชันในภาพไบนารีขณะที่โครงสร้างย่อยต้องอยู่ในเซตที่ถูก
อีโรชัน ในการอธิบายจะใช้รูป 1 มิติดังสมการ

$$(f \ominus b) = \max\{f(s+x) - b(x) \mid (s+x) \in D_f; (x) \in D_b\} \quad (2.83)$$

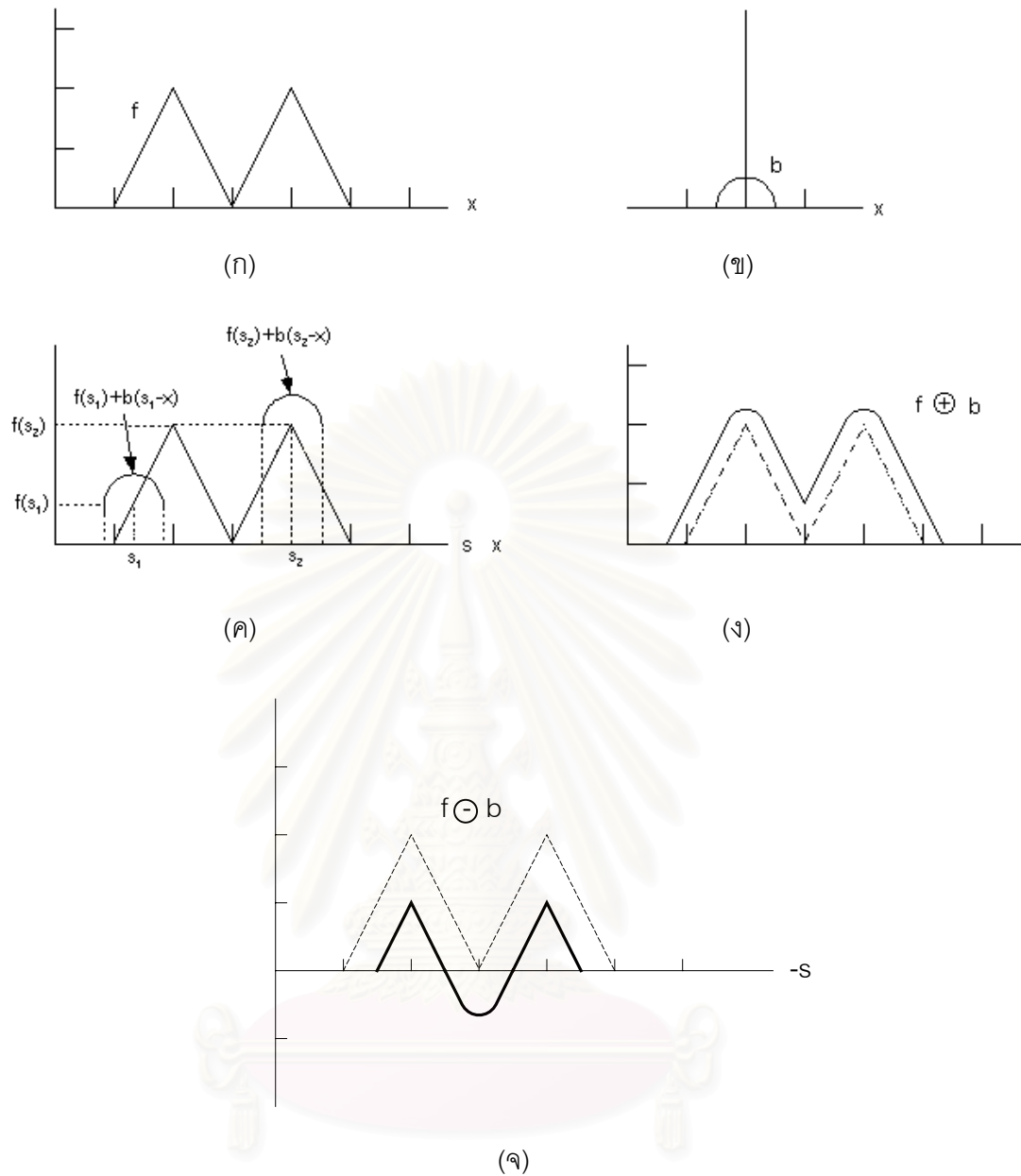
รูปที่ 2.12(จ) แสดงผลลัพธ์ของการทำอีโรชันใน 1 มิติ โดเลขชั้นและอีโรชันมีสมบัติความเป็นคู่ดังนี้

$$(f \ominus b)^c(x, y) = \overline{(f^c \oplus b^c)}(x, y) \quad (2.84)$$

โดย $f^c = -f(x, y)$ และ $b^c = b(-x, -y)$



รูปที่ 2.11 การกระทำตัวดำเนินการแบบเปิดและแบบปิด (ก) ภาพต้นแบบ A (ข) การทำอิโรชันต่อภาพต้นแบบ A ด้วยรูปวงกลม (ค) ผลลัพธ์การทำอิโรชัน (ง) การทำไดเลชันด้วยรูปวงกลม (จ) ผลลัพธ์จากการกระทำตัวดำเนินการแบบเปิด (ฉ) การทำไดเลชันต่อภาพต้นแบบ A ด้วยรูปวงกลม (ช) ผลลัพธ์การทำไดเลชัน (ฌ) การทำอิโรชันด้วยรูปวงกลม (ฎ) ผลลัพธ์จากการกระทำตัวดำเนินการแบบปิด [29]



รูปที่ 2.12 การทำไดเลชันและอีโรชันแสดงในรูปแบบ 1 มิติ (ก) ฟังก์ชันต้นแบบ f (ข) โครงสร้างย่อย b (ค) การทำไดเลชันต่อฟังก์ชันต้นแบบ f ด้วยโครงสร้างย่อย b (ง) ผลลัพธ์จากการทำไดเลชัน (จ) ผลลัพธ์การทำอีโรชันต่อฟังก์ชันต้นแบบ f ด้วยโครงสร้างย่อย b [29]

ตัวดำเนินการแบบเปิดมีนิยามดังนี้

$$f \circ b = (f \ominus b) \oplus b \quad (2.85)$$

ตัวดำเนินการแบบปิดมีนิยามดังนี้

$$f \bullet b = (f \oplus b) \ominus b \quad (2.86)$$

ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดมีสมบัติปิดดังนี้

$$(f \bullet b)^c = f^c \circ \bar{b} \quad (2.87)$$

2.2.3 เกรเดียนต์เชิงสัณฐาน (Morphological Gradient)

ในกระบวนการตรวจจับขอบมีตัวดำเนินการที่สำคัญคือเกรเดียนต์เชิงสัณฐาน ซึ่งมีนิยามดังนี้

$$G(f) = (f \oplus b) - (f \ominus b) \quad (2.88)$$

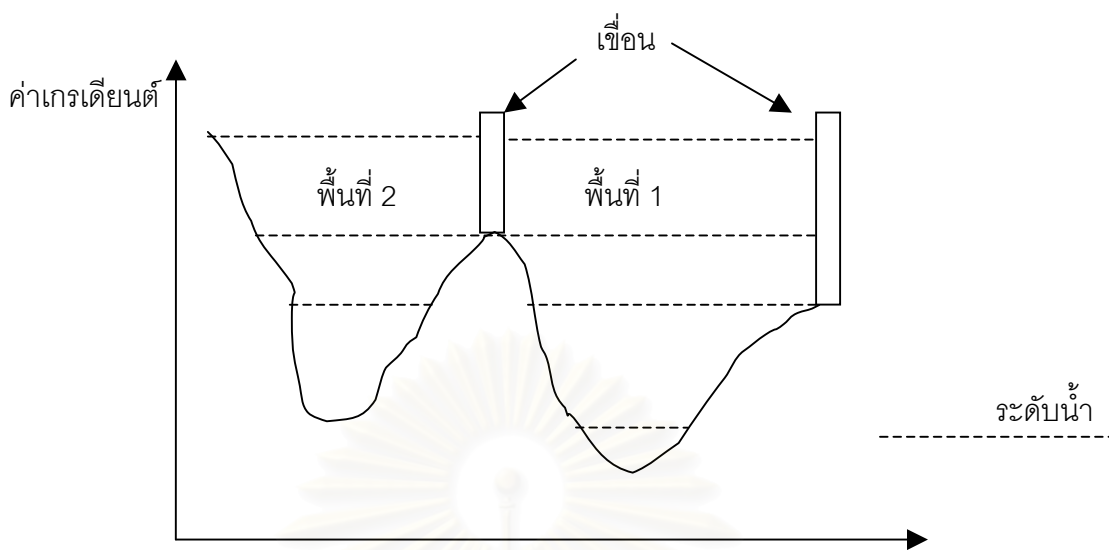
เกรเดียนต์เชิงสัณฐาน คือการตรวจจับขอบของภาพวิธีหนึ่ง ภาพที่ได้จะแสดงถึงขอบของวัตถุรวมถึงการเปลี่ยนแปลงทางสีภายในภาพอย่างชัดเจน โดยทั่วไปแล้วจะใช้ค่าเกรเดียนต์ขององค์ประกอบความเข้มแสงของภาพ บางครั้งควรนำองค์ประกอบของความเข้มสีมาช่วยในการคิดด้วย เพราะว่าในบางกรณี องค์ประกอบของความเข้มสีนั้นให้ผลที่ดีกว่า แต่การนำมาใช้ จะพิจารณาจากตัวถ่วงน้ำหนักคูณกับค่าเกรเดียนต์ในแต่ละองค์ประกอบแล้วตัดสินใจว่าค่าใดให้ค่าเกรเดียนต์มากที่สุดก็ใช้ค่านั้น [15]

$$\max \{ \omega_1 G_r, \omega_2 G_{Cb}, \omega_3 G_{Cr} \} \quad (2.89)$$

โดย $\omega_1, \omega_2, \omega_3$ คือตัวถ่วงน้ำหนักขององค์ประกอบความเข้มสีและความเข้มแสงตามลำดับ ตัวดำเนินการเกรเดียนต์นั้นทำให้เกิดขอบในภาพ แต่ยังไม่สามารถแยกภาพออกเป็นบริเวณหลายๆส่วนได้ หรือไม่สามารแบ่งได้ว่าที่ใดในภาพคือบริเวณที่มีค่าเกรเดียนต์ต่ำสุด การแยกภาพออกเป็นหลายๆบริเวณจะใช้อัลกอริทึมสันปันน้ำ (Watershed Algorithm)

2.2.4 อัลกอริทึมสันปันน้ำ

อัลกอริทึมสันปันน้ำ เป็นการดำเนินการเชิงสัณฐานชนิดหนึ่งซึ่งสามารถระบุค่าเกรเดียนต์ต่ำที่สุดออกมาเป็นพื้นที่ที่สามารถเห็นได้ รูปที่ 2.13 เป็นรูปที่แสดงพื้นฐานการทำงานของอัลกอริทึมสันปันน้ำ โดยแสดงอยู่ในรูป 1 มิติ แกนตั้งแสดงค่าเกรเดียนต์ของภาพ จะเห็นว่า มีลักษณะเป็นความชัน และมีจุดต่ำสุด วิธีการค้นหาว่า จุดต่ำสุดอยู่ที่ใดคือ เริ่มแรกใส่น้ำเข้าไปในแอ่งที่ระดับหนึ่งแล้วค่อยๆเพิ่มขึ้นไปเรื่อยๆ ถ้าระดับน้ำของสองแอ่งใดๆเชื่อมติดกันแล้ว จะมีการสร้างเขื่อนขึ้นตรงนั้น ความสูงของเขื่อนจะต้องมีค่ามากกว่าหรืออย่างน้อยเท่ากับค่าเกรเดียนต์ที่มากที่สุด นั่นหมายความว่า ตัวเขื่อนจะเป็นตัวแบ่งว่าที่ตรงนั้นมีจุดต่ำสุดของค่าเกรเดียนต์ภายในบริเวณนั้นๆ นั่นคือ ถ้านำภาพเกรเดียนต์มาทำกระบวนการตามอัลกอริทึมสันปันน้ำแล้วจะได้ภาพที่แบ่งเขตออกเป็นจุดที่เกรเดียนต์ต่ำสุดได้ ซึ่งหมายถึง สามารถแยกขอบของวัตถุภายในภาพออกมาได้อย่างชัดเจนด้วย



รูปที่ 2.13 อัลกอริทึมสันปันน้ำ 1 มิติ

2.2.4.1 ขั้นตอนการสร้างเขื่อน

พิจารณารูปที่ 2.14(ก) เป็นภาพของแอ่งน้ำสองแอ่งที่ได้จากการทำเกรดเดียนต์ ให้ M_1 และ M_2 เป็นเซตของจุดตำแหน่งของ 2 พื้นที่นั้น และให้ $C_{n-1}[M_1]$ และ $C_{n-1}[M_2]$ เป็นเซตที่บอกว่า 2 พื้นที่นั้นอยู่ในระดับ $n-1$ รูปที่ 2.14(ข) เป็นระดับ n กำหนดสัญลักษณ์คือ q จะเห็นว่า ทั้ง 2 แอ่งเกิดการรวมกัน ถ้ากำหนดการยูเนียนกันของ M_1 และ M_2 คือ $C[n-1]$ แล้ว การที่จะแยกพื้นที่ 2 พื้นที่ใน ระดับ n นั้น สามารถทำได้โดยใช้การกระทำและ (AND) หรือเขียนได้ว่า $q \cap C[n-1]$

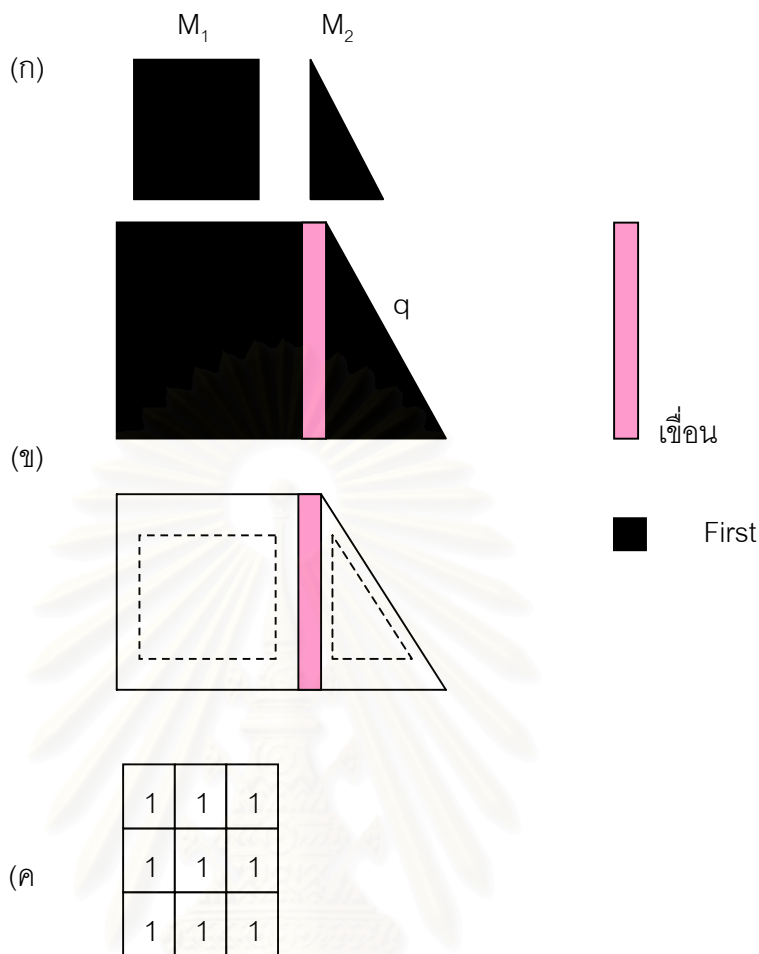
เมื่อเกิดการรวมกันของ 2 แอ่งเช่นนี้จะต้องมีการสร้างเขื่อน วิธีการสร้างคือ ใช้มาสก์ขนาด 3×3 ดังรูปที่ 2.14(ค) ทำได้เลขชั้นกับ 2 แอ่งนั้นผลลัพธ์ที่ได้จะทำให้ 2 แอ่งขยายตัวเองออกโดยการขยายต้องสอดคล้องกับเงื่อนไขคือ

1) ทำได้เลขชั้นแล้วผลลัพธ์ที่ออกมาต้องอยู่ในพื้นที่ q

2) ผลลัพธ์ที่ออกมาจากการทำได้เลขชั้นจะต้องไม่ซ้อนทับกัน

จากเงื่อนไข 2 ข้อนี้ จะเกิดจุดที่ไม่มีการทำได้เลขชั้นและอยู่ภายในพื้นที่ q ซึ่งจุดนั้นก็คือจุดที่กำหนดเป็นเขื่อนนั่นเอง ในรูปที่ 2.14(ข) จะเห็นว่าต้องทำการได้เลขชั้นถึง 2 ครั้งจึงจะได้จุดที่เป็นเขื่อน ความสูงของเขื่อนก็ต้องกำหนดให้มีค่ามากกว่าค่าเกรดเดียนต์สูงสุด เมื่อทำเช่นนี้จนครบทุกระดับความสูงของเกรดเดียนต์แล้วจะได้เขื่อนที่แยกภาพออกเป็นส่วนๆออกมาได้

จุดด้อยของอัลกอริทึมนี้คือส่วนของภาพที่แยกออกมาได้จะมีมากเกินไป ดังนั้นจะต้องมีกระบวนการบางอย่างที่ใช้สำหรับรวมพื้นที่บางส่วนเข้าด้วยกัน พื้นที่ที่แยกออกมานั้นมีหลายส่วนที่ยังชี้ชัดไม่ได้ว่าเป็นส่วนพื้นหลัง (background) หรือ ส่วนพื้นหน้า (foreground)



รูปที่ 2.14 การสร้างเช็อน (ก) แอ่ง 2 แอ่ง ในระดับ $n - 1$ (ข) การรวมของ 2 แอ่งในระดับ n
 (ค) มาตรฐานที่ใช้ทำไดเลชัน [29]

2.3 ทฤษฎีการตัดสินใจของเบย์ (Bayesian Decision Theory)

ทฤษฎีการตัดสินใจของเบย์ นั้นจะตัดสินใจโดยเลือกกระทำเมื่อเหตุการณ์ที่กระทำนั้นมีค่าความเสี่ยงน้อยที่สุด [30]

ให้ x คือ เวกเตอร์ลักษณะเด่น (feature vector)

$\{w_1, \dots, w_c\}$ เป็นเซตของคลาส w_j โดย w_j คือคลาสของสิ่งที่สนใจ

$\{\alpha_1, \dots, \alpha_c\}$ เป็นเซตของการกระทำที่เป็นไปได้ โดย α_i คือการกระทำที่ i (ในที่นี้หมายถึงเลือกกว่าเป็นคลาส i)

$\lambda(\alpha_i|w_j)$ คือ ฟังก์ชันการสูญเสีย ซึ่งเป็นค่าความสูญเสียเมื่อเลือกการกระทำ α_i โดยคลาสนั้นคือ w_j

$p(x|w_j)$ คือ ฟังก์ชันความหนาแน่นความน่าจะเป็นของ x เมื่อ x เป็นของคลาส w_j

$P(w_j)$ คือ ความน่าจะเป็นก่อน (prior probability) เมื่อเป็นคลาส w_j แล้ว ความน่าจะเป็นภายหลัง (posterior probability) $P(w_j | \mathbf{x})$ สามารถหาได้จาก $p(\mathbf{x} | w_j)$ โดยสูตรของเบย์ดังนี้

$$P(w_j | \mathbf{x}) = \frac{p(\mathbf{x} | w_j)P(w_j)}{p(\mathbf{x})} \quad (2.90)$$

โดย

$$p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x} | w_j)P(w_j) \quad (2.91)$$

ค่าความเสี่ยงเงื่อนไข (conditional risk) เมื่อเลือก \mathbf{x} เป็น α_i คำนวณได้จาก

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i | w_j) P(w_j | \mathbf{x}) \quad (2.92)$$

ดังนั้น การตัดสินใจแบบเบย์คือการหาค่าความเสี่ยงเงื่อนไขให้น้อยที่สุด

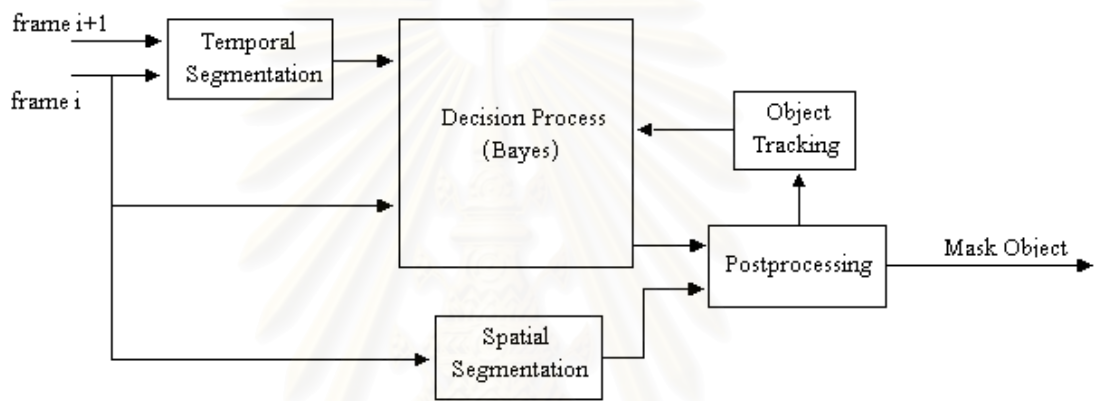
ในบทนี้กล่าวถึง พื้นฐานของสี่ แบบจำลองสี่แบบต่างๆ ตัวดำเนินการเชิงเส้นฐาน และ ทฤษฎีการตัดสินใจของเบย์ ซึ่งทฤษฎีเหล่านี้จะนำมาเป็นพื้นฐานในการพัฒนาแนวทางอัลกอริทึมที่จะนำเสนอในบทที่ 3 ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

การแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้เทคนิคการตัดสินใจแบบเบย์

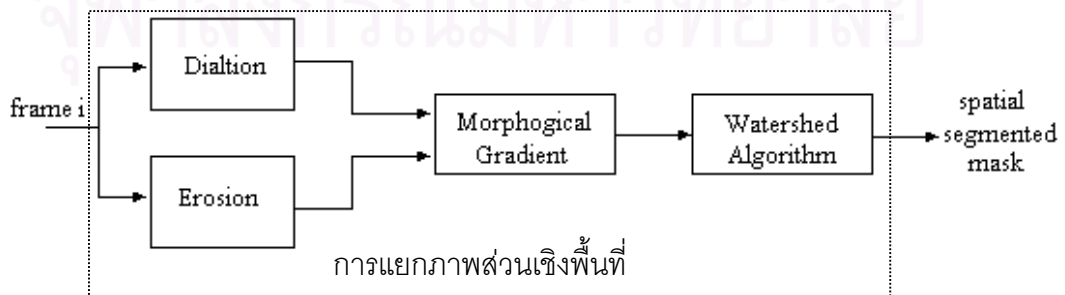
การวิจัยด้านการแยกส่วนภาพวิดีโอออกเป็นวัตถุมีทั้งแบบอัตโนมัติและกึ่งอัตโนมัติ โดยใช้เทคนิคต่าง ๆ กันดังได้นำเสนอในบทที่ 1 นั้น มีหลากหลายวิธี ในวิทยานิพนธ์นี้ จะเสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้เทคนิคการตัดสินใจแบบเบย์ รูปที่ 3.1 เป็นโครงสร้างแนวคิดของวิทยานิพนธ์ [31] ทางด้านอินพุตเป็นภาพ 2 เฟรมที่ติดต่อกัน โดยผลลัพธ์ที่ได้คือมาสก์ของวัตถุ ในขอบเขตของวิทยานิพนธ์นี้ วัตถุคือตัวคนที่เคลื่อนไหวภายในภาพโดยลักษณะของวิดีโอจะเป็นวิดีโอคอนเฟอเรนซ์ แต่ละส่วนในรูปที่ 3.1 มีกรรมวิธีดังนี้



รูปที่ 3.1 โครงสร้างของแนวคิดวิทยานิพนธ์

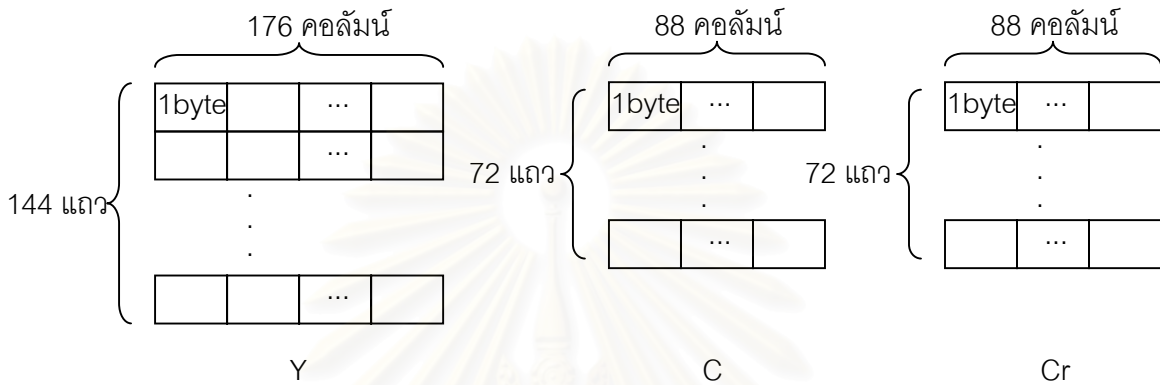
3.1 การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation)

การแยกส่วนภาพเชิงพื้นที่นั้น กระทำเพื่อแบ่งภาพออกเป็นส่วนย่อยๆ โดยขอบของส่วนย่อยๆ ของภาพนั้นต้องเป็นขอบของวัตถุด้วย กระบวนการที่กระทำแล้วให้ผลลัพธ์เช่นนี้คือ อัลกอริทึมสันปันน้ำ โดยก่อนจะมีการกระทำอัลกอริทึมสันปันน้ำนั้น จะต้องอ่านค่าความเข้มแสงของลำดับวิดีโอเข้ามา แล้วทำการประมวลผลได้เลขชี้กับอีโรชัน เพื่อนำไปหาค่าเกรเดียนต์เชิงสัญญาณของภาพนั้น ต่อจากนั้นจึงนำมาทำตามอัลกอริทึมสันปันน้ำดังแสดงในรูปที่ 3.2

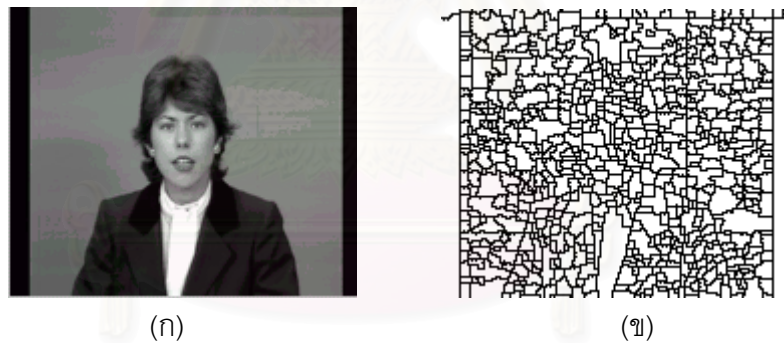


รูปที่ 3.2 โครงสร้างการแยกภาพส่วนเชิงพื้นที่

โครงสร้างของลำดับวิดีโอแบบ YCbCr 4:2:0 มีการจัดรูปแบบดังรูปที่ 3.3 โดย 1 เฟรมจะมี $176 * 144 + (176/2) * (144/2) + (176/2) * (144/2) = 38013$ ไบต์ การนำไปใช้ในการแยกส่วนภาพเชิงพื้นที่นั้นจะนำเฉพาะความเข้มแสงไปใช้เท่านั้น ภาพที่ได้ออกมาจากระบบการแยกส่วนภาพเชิงพื้นที่นั้นจะเป็นภาพขาวดำโดยแบ่งออกเป็นส่วยย่อยๆ ซึ่งแต่ละส่วนจะมีเส้นขอบเขตเป็นสีดำล้อมรอบพื้นที่ที่เป็นสีขาว รูปที่ 3.4 แสดงตัวอย่างผลลัพธ์จากอัลกอริทึมสันปันน้ำ



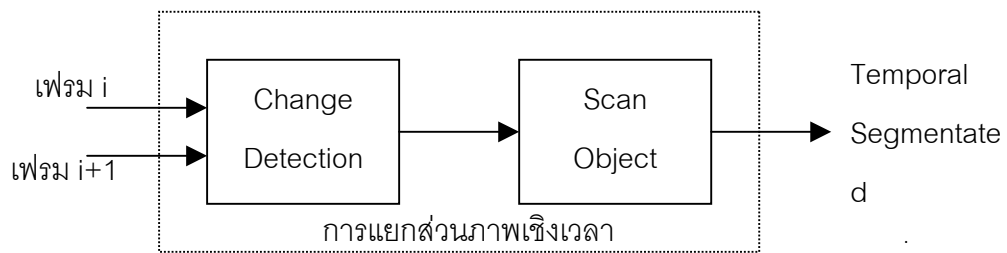
รูปที่ 3.3 โครงสร้างการจัดวางตำแหน่งไบต์



รูปที่ 3.4 แสดงตัวอย่างผลลัพธ์จากอัลกอริทึมสันปันน้ำ (ก) รูปต้นฉบับ (ข) ผลลัพธ์จากอัลกอริทึมสันปันน้ำ

3.2 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)

เป้าหมายของการแยกส่วนภาพเชิงเวลาคือเพื่อให้ได้ส่วนของพื้นที่หรือวัตถุที่ต้องการมากที่สุด เพื่อจะได้ถูกนำไปใช้โดยส่วนของการตัดสินใจได้อย่างมีประสิทธิภาพ ถ้าการแยกส่วนภาพเชิงเวลายังไม่สามารถแยกส่วนของวัตถุที่ต้องการได้มากพอแล้ว กระบวนการตัดสินใจก็ให้ผลที่ดีได้ยาก การแยกส่วนภาพเชิงเวลาใช้ประโยชน์จากภาพ 2 เฟรมที่อยู่ติดต่อกัน เนื่องจากว่า ถ้าเกิดการเคลื่อนไหวขึ้นแล้ว ก็จะมีการเปลี่ยนแปลงของความเข้มแสงด้วย การแยกส่วนภาพเชิงเวลาสามารถเขียนเป็นกระบวนการได้ดังรูปที่ 3.5



รูปที่ 3.5 โครงสร้างการแยกส่วนเชิงเวลา

เริ่มต้นจะนำความเข้มแสงของภาพเฟรมที่ $i+1$ และภาพเฟรมที่ i มาตรวจจับการเปลี่ยนแปลงแล้วทำเป็นค่าสัมบูรณ์ เพราะฉะนั้นพื้นที่ส่วนที่ไม่เกิดการเคลื่อนไหวก็จะมีค่าเป็นศูนย์ แต่เนื่องจากอุปกรณ์ที่ใช้จับภาพอาจมีการรบกวนเนื่องจากความร้อน ความผิดพลาดที่เกิดจากการควอนไทซ์ และอื่นๆ [15] อันอาจมีผลทำให้ค่าความเข้มแสงของส่วนที่ไม่เคลื่อนไหวไม่เป็นศูนย์ทั้งหมด วิธีแก้ไขขั้นต้นคือกำหนดค่าจุดเปลี่ยนไว้ค่าหนึ่งเพื่อใช้ประมาณว่าความแตกต่างนั้น เกิดจากการเคลื่อนไหวจริงๆ มิใช่เกิดจากความผิดพลาดหรือการรบกวนของสิ่งอื่นๆ มาสกัดตรวจจับการเคลื่อนไหว (Change Detection Mask : CDM) นิยามได้ดังนี้

$$CDM_i(x,y) = \begin{cases} 1 & \text{ถ้า } d_i \geq T_i \\ 0 & \text{ถ้า } d_i < T_i \end{cases} \quad (3.1)$$

โดยที่ $d_i = |p_i(x,y) - p_{i-1}(x,y)|$

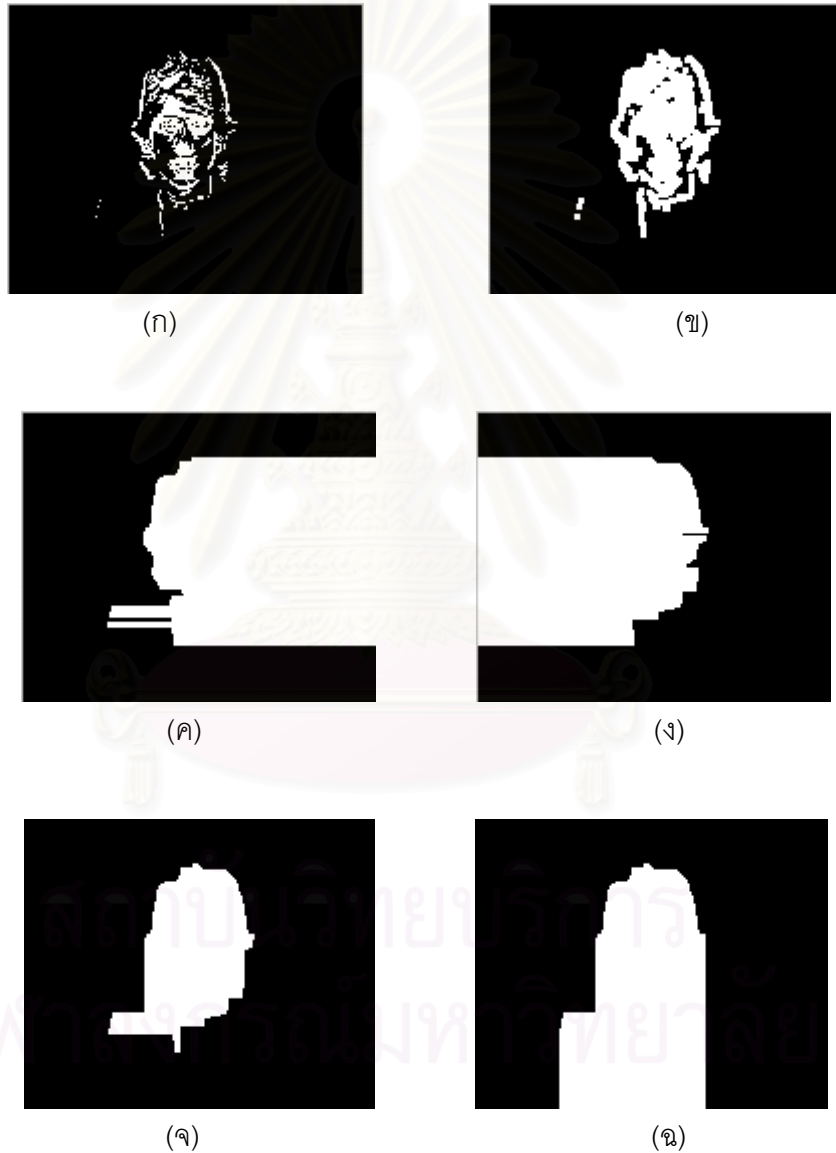
$p_i(x,y)$ คือ ค่าระดับสีเทาที่ตำแหน่ง (x,y) ในเฟรมที่ i

T_i เป็นค่าเทรชโฮลด์ ซึ่งมีค่าแตกต่างกันไปตามระบบรับภาพแต่ละระบบ

ผลลัพธ์ที่ได้จะถูกนำมาทำไคเลชันก่อนเพื่อขยายจุดการเปลี่ยนแปลงให้ใหญ่ขึ้น แล้วสแกนหาวัตถุ โดยสมมติว่าจุดของการเปลี่ยนแปลงที่ตรวจจับได้ เป็นขอบของวัตถุและวัตถุคือสิ่งที่อยู่ภายในจุดขอบเหล่านั้นดังรูปที่ 3.6

การสแกนครั้งแรกจะสแกนจากซ้ายไปขวาและบนลงล่าง เมื่อพบจุดการเปลี่ยนแปลงในแถวใดแถวหนึ่งแล้ว ให้ส่วนที่เหลือในแถวนั้นเป็นสีขาวนับตั้งแต่จุดนั้นไปทางขวาดังในรูปที่ 3.6(ค) การสแกนครั้งที่สองจะสแกนจากขวามาซ้ายและบนลงล่าง เมื่อพบจุดการเปลี่ยนแปลงในแถวแล้ว ให้ส่วนที่เหลือในแถวนั้นเป็นสีขาวนับจากจุดนั้นไปทางซ้ายดังรูปที่ 3.6(ง) นำผลลัพธ์ของทั้งสองครั้งมากระทำ AND ต่อกัน ได้ผลลัพธ์ดังรูปที่ 3.6(จ) แล้วจึงสแกนครั้งสุดท้ายจะสแกนจากบนลง

ล่างและซ้ายไปขวาเมื่อพบจุดการเปลี่ยนแปลงในคอลัมน์นั้นแล้วให้ส่วนที่เหลือในคอลัมน์เป็นสีขาวนับจากจุดนั้นไปข้างล่างผลลัพธ์ดังรูปที่ 3.6(ข) เพื่อนำไปใช้เป็นมาสก์ให้กับกระบวนการตัดสินใจต่อไป เป้าหมายของการสแกนคือต้องการตรวจจับส่วนของวัตถุให้ได้มากที่สุดโดยใช้ข้อมูลจากการตรวจจับการเคลื่อนไหว โดยทั่วไปลักษณะของวิดีโอคอนเฟอร์เรนซ์คือ มีตัวคนปรากฏอยู่ครึ่งตัวและส่วนที่เคลื่อนไหวโดยมากคือส่วนของใบหน้า ดังนั้นการแนวคิดของการสแกนจะใช้ประโยชน์จากความจริงข้อดังกล่าว



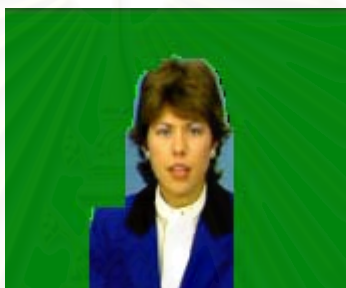
รูปที่ 3.6 แสดงวิธีการสแกน (ก) CDM (ข) CDM ที่ได้รับการทำไคเลชัน (ค) สแกนครั้งแรก (ง) สแกนครั้งที่สอง (จ) การกระทำ AND (ฉ) สแกนครั้งที่สาม

3.3 กระบวนการตัดสินใจ

กระบวนการตัดสินใจจะใช้การตัดสินใจแบบเบย์ (ดังกล่าวแล้วในบทที่ 2) สามารถแบ่งได้เป็นสองกรณีคือกรณีที่มีมาส์กจากกระบวนการติดตามวัตถุและกรณีที่ไม่มีมาส์กจากกระบวนการติดตามวัตถุ

ก) กรณีที่ไม่มีมาส์กจากกระบวนการติดตามวัตถุ

กรณีนี้จะเกิดขึ้นเป็นการตัดสินใจในเฟรมแรกเท่านั้น กระบวนการคือนำมาส์กที่ได้จากการแยกส่วนภาพเชิงเวลามาพิจารณาเป็น 2 คลาส คือคลาสของวัตถุ และคลาสของพื้นหลัง แสดงดังรูปที่ 3.7



รูปที่ 3.7 ตัวอย่างมาส์กที่ได้จากการแยกส่วนภาพเชิงเวลา

ข) กรณีที่มีมาส์กจากกระบวนการติดตามวัตถุ

จะกระทำโดยนำมาส์กที่ได้จากการแยกส่วนเชิงเวลาและมาส์กที่ได้จากกระบวนการติดตามวัตถุมากระทำ OR กัน แล้วพิจารณาออกเป็นสองคลาส คือ คลาสของวัตถุและคลาสของพื้นหลัง มาส์กที่ได้จากกระบวนการติดตามวัตถุแสดงดังรูปที่ 3.8 มาส์กที่ได้จากการกระทำ OR แสดงดังรูปที่ 3.9

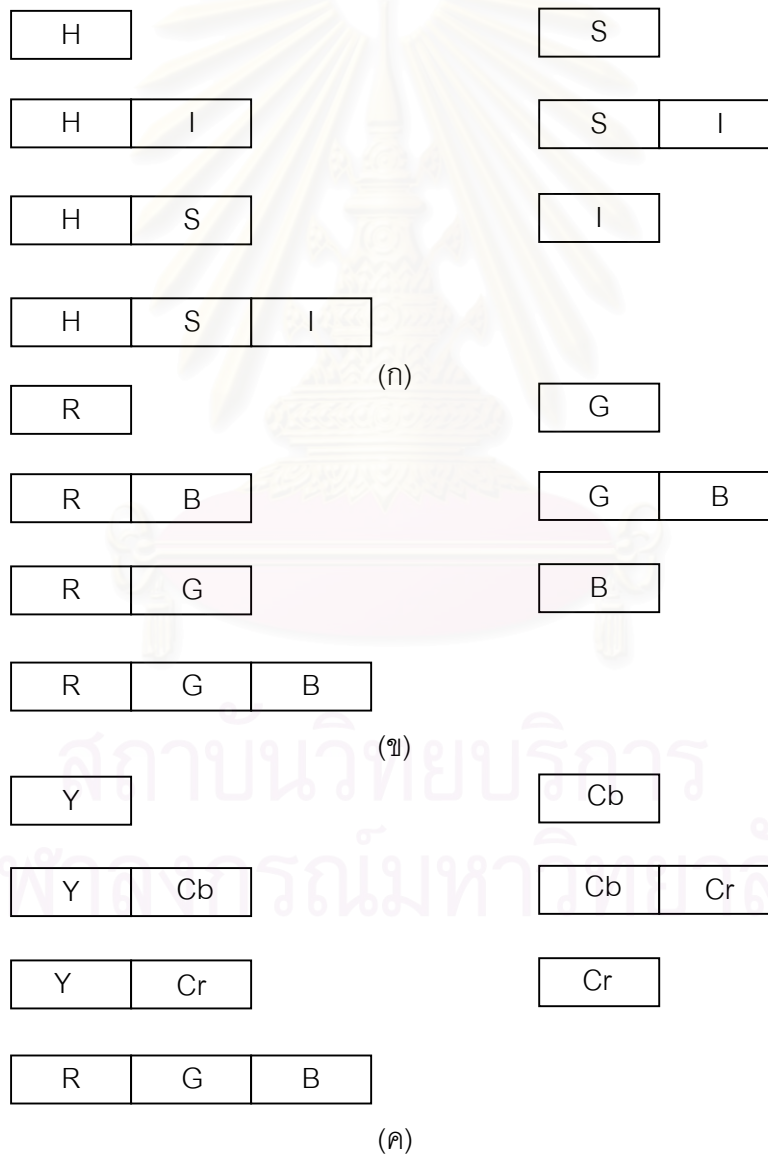


รูปที่ 3.8 ตัวอย่างมาส์กที่ได้จากกระบวนการติดตามวัตถุ



รูปที่ 3.9 ตัวอย่างมาสก์ที่ได้จากการกระทำ OR

ในการกระบวนการตัดสีใจจะใช้เวกเตอร์ลักษณะเด่น x ซึ่งเป็นข้อมูลขององค์ประกอบสี ตั้งแต่ 1 ถึง 3 องค์ประกอบ ในวิทยานิพนธ์นี้จะทดลองตามแบบจำลองสีทั้ง 3 แบบ โดยแสดงดังรูปที่ 3.10



รูปที่ 3.10 ข้อมูลขององค์ประกอบสีที่ใช้เป็นเวกเตอร์ลักษณะเด่น (ก) องค์ประกอบสี HSI

(ข) องค์ประกอบสี RGB (ค) องค์ประกอบสี YCbCr

กระบวนการตัดสินใจมี 2 ขั้นตอน ขั้นตอนแรกทำการหาความน่าจะเป็น (likelihood) ขั้นตอนที่สองตัดสินใจว่าจุดภาพนั้นเป็นวัตถุหรือพื้นหลัง จาก $R(\alpha_i | \mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i | w_j) P(w_j | \mathbf{x})$ (สมการที่ (2.94)) โดย c คือจำนวนของคลาส ในที่นี้มี 2 คลาส คือคลาสวัตถุและคลาสพื้นหลัง กำหนดให้ α_1 คือการเลือกให้เป็นวัตถุ α_2 คือการเลือกให้เป็นพื้นหลัง w_1 คือ คลาสของวัตถุ w_2 คือ คลาสของพื้นหลัง \mathbf{x} คือ ข้อมูล 24 บิต ของ YCb และ Cr ของภาพเฟรมที่ i $\lambda(\alpha_i | w_j)$ คือ ค่าความสูญเสียเมื่อเลือกว่าเป็นคลาส w_j โดยเงื่อนไขคือ w_j (ย่อเป็น λ_{ij})

$$R(\alpha_1 | \mathbf{x}) = \lambda_{11} P(w_1 | \mathbf{x}) + \lambda_{12} P(w_2 | \mathbf{x}) \quad (3.2)$$

$$R(\alpha_2 | \mathbf{x}) = \lambda_{21} P(w_1 | \mathbf{x}) + \lambda_{22} P(w_2 | \mathbf{x}) \quad (3.3)$$

โดยการตัดสินใจของเบย์จะตัดสินใจว่าเป็น w_1 เมื่อ

$$R(\alpha_1 | \mathbf{x}) < R(\alpha_2 | \mathbf{x}) \quad (3.4)$$

แทน สมการที่ (3.2) และ สมการที่ (3.3) ลง สมการที่ (3.4) ได้

$$(\lambda_{21} - \lambda_{11}) P(w_1 | \mathbf{x}) > (\lambda_{12} - \lambda_{22}) P(w_2 | \mathbf{x}) \quad (3.5)$$

$$(\lambda_{21} - \lambda_{11}) p(\mathbf{x} | w_1) P(w_1) > (\lambda_{12} - \lambda_{22}) p(\mathbf{x} | w_2) P(w_2) \quad (3.6)$$

$$\frac{p(\mathbf{x} | w_1)}{p(\mathbf{x} | w_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P(w_1)} \quad (3.7)$$

โดยให้ $\lambda_{11} = \lambda_{22} = 0$ เพราะว่าเป็นกรณีที่เกิดการสูญเสียน้อยที่สุด

$\lambda_{21} = \lambda_{12} = 1$ เนื่องจากต้องการให้เกิดความสูญเสียที่เท่าๆกัน

$$\frac{p(\mathbf{x} | w_1)}{p(\mathbf{x} | w_2)} > \frac{P(w_2)}{P(w_1)} \quad (3.8)$$

เพราะฉะนั้นการตัดสินใจว่า จุดภาพใดเป็นวัตถุหรือพื้นหลัง จะดูจาก \mathbf{x} ที่จุดภาพนั้นว่าสอดคล้องกับสมการ (3.8) หรือไม่ ถ้าสอดคล้องจะถือว่าเป็นวัตถุ ถ้าไม่สอดคล้องจะถือว่าเป็นพื้นหลัง ผลลัพธ์ที่ได้จากกระบวนการตัดสินใจจะเป็นมาสก์ขาวดำ ซึ่งจะนำไปใช้โดยกระบวนการข้างหลังต่อไป รูปที่ 3.11 แสดงมาสก์ที่ได้ในรูปแบบของภาพสี



รูปที่ 3.11 ตัวอย่างมาสก์ที่ได้จากการตัดสินใจในรูปแบบของภาพสี

ค่าความน่าจะเป็นก่อน $P(w_1)$ หาได้จาก จำนวนจุดภาพที่เป็นการเคลื่อนไหวต่อจำนวนจุดภาพทั้งหมด เช่นจากรูปที่ 3.6 (ข) จะเห็นได้ว่า จำนวนจุดภาพที่เป็นการเคลื่อนไหวคือส่วนที่เป็นสีขาว จำนวนจุดภาพทั้งหมดคือพื้นที่ของขนาดของภาพ ส่วนค่าความน่าจะเป็นก่อน $P(w_2)$ นั้นก็คือจุดภาพที่ไม่ใช่จุดภาพที่เคลื่อนไหวต่อจำนวนจุดภาพทั้งหมด จากรูปที่ 3.6 (ข) จุดภาพที่ไม่ใช่จุดภาพที่เคลื่อนไหวก็คือส่วนที่เป็นสีดำ สามารถแสดงได้ดังสมการ

$$P(w_1) = n_{\text{เคลื่อนไหว}} / n_{\text{ทั้งหมด}} \quad (3.9)$$

$$P(w_2) = n_{\text{ไม่เคลื่อนไหว}} / n_{\text{ทั้งหมด}} \quad (3.10)$$

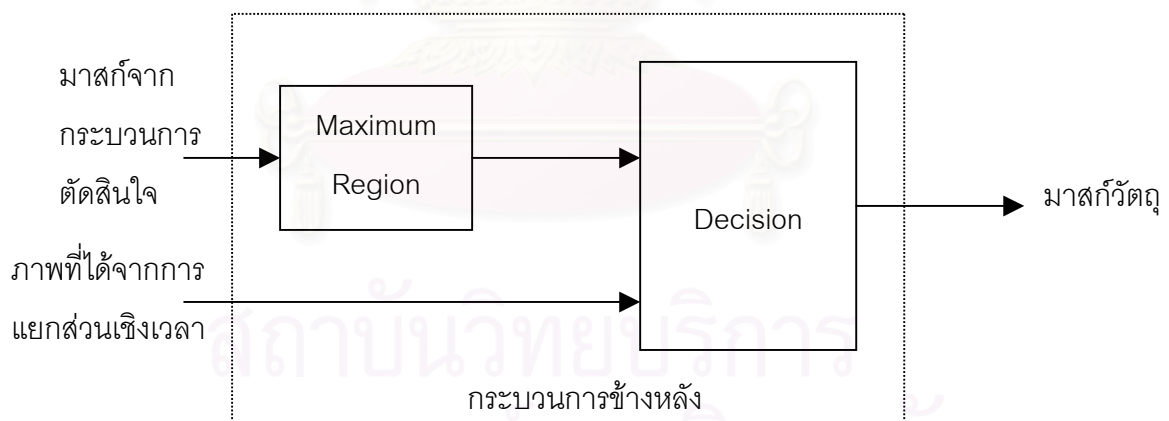
โดย $n_{\text{เคลื่อนไหว}}$ คือจำนวนจุดภาพที่กำหนดว่าเป็นจุดภาพที่เคลื่อนไหว หาได้โดยการสแกนครั้งที่สามในกระบวนการแยกส่วนภาพเชิงเวลา

$n_{\text{ไม่เคลื่อนไหว}}$ คือจำนวนจุดภาพที่กำหนดว่าเป็นจุดภาพที่ไม่เคลื่อนไหว หาได้โดยการสแกนครั้งที่สามในกระบวนการแยกส่วนภาพเชิงเวลา

$n_{\text{ทั้งหมด}}$ คือ จำนวนจุดภาพทั้งหมดของภาพนั้น

3.4 กระบวนการข้างหลัง

마스크ที่ได้จากกระบวนการตัดสินใจ ค่อนข้างมีสิ่งรบกวนมาก ฉะนั้นจะต้องกำจัดออกก่อนแล้วจึงนำไปรวมกับภาพที่ได้จากกระบวนการแยกส่วนภาพเชิงพื้นที่ กรรมวิธีแสดงดังรูปที่ 3.12



รูปที่ 3.12 โครงสร้างของกระบวนการข้างหลัง

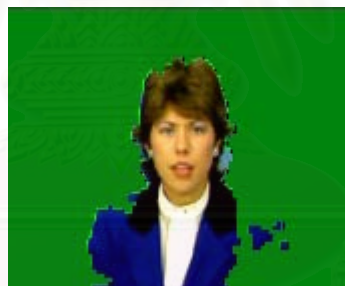
พื้นที่มากที่สุด (Maximum Region) เป็นกระบวนการหาพื้นที่ที่มากที่สุดโดยอินพุตคือ 마스크ที่ได้จากกระบวนการตัดสินใจ ดังได้กล่าวแล้วว่ากระบวนการตัดสินใจจะให้มาสก์ที่มีวัตถุและสิ่งรบกวนที่เป็นจุดเล็กจุดน้อยติดมาด้วยเป็นจำนวนมาก เพราะฉะนั้นจำเป็นต้องกำจัดจุดเล็กจุดน้อยนั้นออก โดยทำการหาจุดที่มีพื้นที่มากที่สุด แล้วกำหนดให้พื้นที่นั้นเป็นพื้นที่ของวัตถุ ส่วนจุดเล็กจุดน้อยนั้น กำหนดให้เป็นพื้นที่ของพื้นหลัง เมื่อเป็นเช่นนี้แล้ว จะได้มาสก์ที่มีพื้นที่เพียงพื้น

ที่เดียว ซึ่งอนุมานได้ว่าเป็นวัตถุที่สนใจ ในการตัดสินใจ (decision) นั้น จะเป็นการตัดสินใจว่าพื้นที่ที่ได้จากกระบวนการแยกส่วนเชิงพื้นที่ว่า ส่วนใดเป็นวัตถุ ส่วนใดเป็นพื้นหลัง โดยใช้เงื่อนไขว่า ถ้าพื้นที่ย่อยในภาพที่ได้จากการแยกส่วนเชิงพื้นที่ พื้นที่ใดถูกรอบครองโดยพื้นที่ที่ได้จากกระบวนการพื้นที่มากที่สุด เกินครั้งหนึ่งแล้ว จะกำหนดให้พื้นที่ย่อยนั้นเป็นวัตถุ สามารถแสดงเป็นสมการได้ดังนี้

$$P = N(R_t^{\text{proj}}, R_s) / N(R_s) \quad (3.11)$$

โดย R_s คือ พื้นที่ย่อยในภาพที่ได้จากการแยกส่วนเชิงพื้นที่ R_t^{proj} คือ พื้นที่ย่อยที่ได้จากการแยกส่วนเชิงเวลาเมื่อฉายลงบนภาพที่ได้จากการแยกส่วนเชิงพื้นที่ $N(R_s)$ คือจำนวนจุดภาพในพื้นที่ย่อย R_s

สาเหตุที่ใช้การแยกส่วนเชิงพื้นที่มาใช้ เพราะว่าพื้นที่ย่อยของภาพที่ได้จากการแยกส่วนเชิงเวลานั้น มีส่วนขอบของวัตถุอยู่ (เพราะเป็นการหาเกรเดียนต์) ดังนั้นจึงมั่นใจได้ระดับหนึ่งว่าพื้นที่ที่กำหนดเป็นวัตถุจะตรงกับขอบของวัตถุด้วย ผลลัพธ์ที่ได้จากการกระบวนการข้างหลังจะเป็นมาสก์ของวัตถุในภาพเฟรมที่ i และผลที่ได้จะถูกเก็บไว้ใช้เป็นตัวติดตามโดยกระบวนการติดตามวัตถุต่อไป ผลลัพธ์ที่ได้แสดงในรูปที่ 3.13



รูปที่ 3.13 ตัวอย่างมาสก์ที่ได้จากการกระบวนการข้างหลัง (มาสก์ของวัตถุ)

3.5 การติดตามวัตถุ (Object Tracing)

การติดตามวัตถุเป็นเหมือนหน่วยความจำหน่วยหนึ่งซึ่งมีหน้าที่เก็บมาสก์วัตถุของภาพเฟรมที่ i เพื่อนำไปใช้ในกระบวนการตัดสินใจของภาพเฟรมที่ $i+1$ ต่อไป

3.6 การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐาน (Hypothesis Testing)

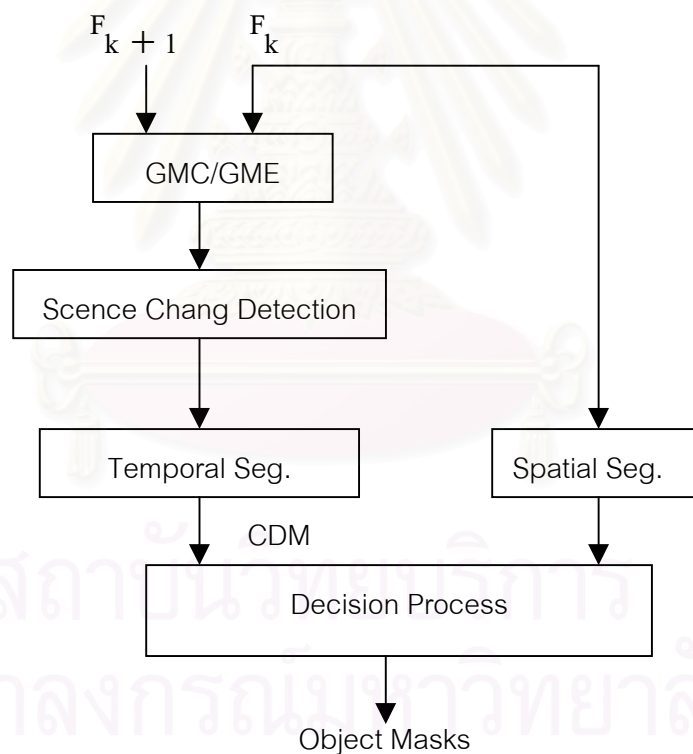
การแยกส่วนภาพโดยวิธีการทดสอบสมมติฐานนั้นนำเสนอโดย MunchurixKim, Jae Gark Choi, Daelee Kim, Hyung Lee [15] ซึ่งมีโครงสร้างดังรูปที่ 3.14 ประกอบด้วย 5 กระบวนการหลักดังนี้

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 การตรวจจับทำได้โดยวัดค่าความแตกต่างสัมบูรณ์เฉลี่ย (Mean Absolute Difference) ของเฟรมสองเฟรมติดกัน ถ้ามีค่ามากกว่าค่าที่ตั้งไว้ก็ตัดสินว่ามีการเปลี่ยนฉากเกิดขึ้นแล้วไม่ต้องทำกระบวนการแยกส่วนภาพเชิงเวลา



รูปที่ 3.14 กรรมวิธีที่เสนอโดย Munchurl Kim, Jae Gark Choi, Daelee Kim, Hyung Lee

Myoung Ho Lee, Chieteuk และ Yo-Sung Ho [15]

3. การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)

กระบวนการนี้นำภาพ 2 เฟรมมาหาค่าความแตกต่างและทำการทดสอบสมมติฐานโดยมีการดำเนินการดังนี้

- หาค่าความแตกต่างของภาพเฟรมที่ n และ $n+3$
- กำหนดส่วนที่เป็นพื้นหลังแล้วคำนวณหาความแปรปรวน S_1^1
- แบ่งภาพที่เป็นค่าความแตกต่างออกเป็นบล็อกย่อยๆ เท่ากันแล้วคำนวณค่าความแปรปรวนของบล็อกนั้นๆ S_2^1
- คำนวณค่าทดสอบทางสถิติ $V = \frac{S_2^1}{S_1^1}$
- 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

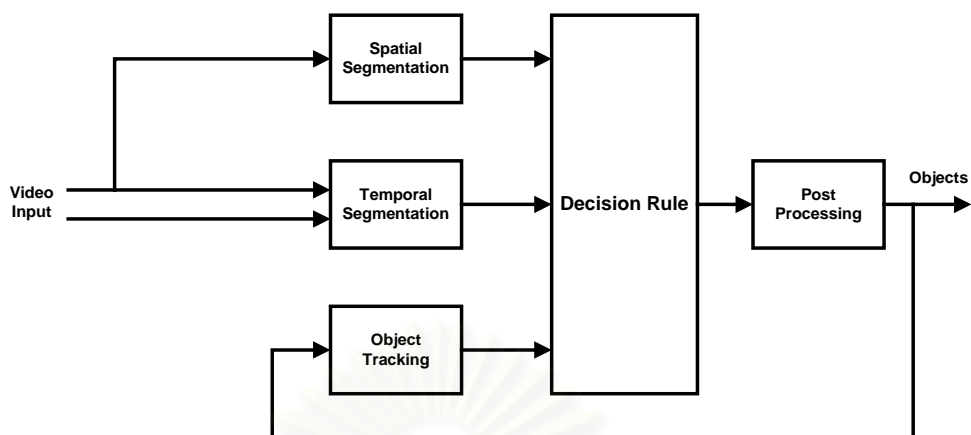
5. กระบวนการตัดสินใจ เป็นกระบวนการรวมส่วนที่ได้จากการแยกส่วนเชิงเวลาและการแยกส่วนเชิงพื้นที่เข้าด้วยกันโดยใช้หลักการคือ ถ้าส่วนที่ได้จากการแยกส่วนเชิงเวลาครอบครอง

พื้นที่ของกระบวนการแยกส่วนภาพเชิงพื้นที่ที่มากกว่าร้อยละ 90 ให้ตัดสินใจว่าพื้นที่นั้นเป็นวัตถุ จากนั้นนำผลที่ได้ไปช่วยในกระบวนการติดตามวัตถุในเฟรมถัดไป

3.7 การแยกส่วนภาพเชิงพื้นที่และเวลา

การแยกส่วนภาพเชิงพื้นที่และเวลานั้นนำเสนอโดย ดัชกรรณ ดันเจริญ และคณะ [27] มีโครงสร้างดังรูปที่ 3.15 ซึ่งประกอบด้วย 4 กระบวนการหลักดังนี้

1. การแยกส่วนภาพเชิงพื้นที่ ประกอบด้วย 3 กระบวนการย่อยคือ
 - การทำให้ภาพเรียบง่าย ในกระบวนการย่อยนี้ทำให้ภาพเรียบง่ายโดยใช้ตัวดำเนินการแบบเปิดและตัวดำเนินการแบบปิดซึ่งเป็นตัวดำเนินการเชิงสัญญาณ
 - การประมาณเกรเดียนต์ กระบวนการย่อยนี้กระทำโดยเกรเดียนต์เชิงสัญญาณ
 - กรรมวิธีสันปันน้ำ
2. การแยกส่วนภาพเชิงเวลา กระบวนการนี้นำภาพ 2 เฟรมมาหาค่าความแตกต่างแล้วกำหนดว่าถ้าจุดภาพใดมีค่าความแตกต่างมากกว่าค่าที่กำหนด ให้ตัดสินใจว่าจุดภาพนั้นมีความเปลี่ยนแปลง
3. กระบวนการตัดสินใจ เป็นกระบวนการรวมส่วนที่ได้จากการแยกส่วนเชิงเวลาและการแยกส่วนเชิงพื้นที่เข้าด้วยกันโดยใช้หลักการคือ ถ้าส่วนที่ได้จากการแยกส่วนเชิงเวลาครอบครองพื้นที่ของกระบวนการแยกส่วนภาพเชิงพื้นที่ที่มากกว่าร้อยละ 50 ให้ตัดสินใจว่าพื้นที่นั้นเป็นวัตถุ
4. กระบวนการติดตามวัตถุ เป็นกระบวนการในตรวจจับตำแหน่งของวัตถุในเฟรมถัดไปว่าอยู่ที่ตำแหน่งใด



รูปที่ 3.15 กรรมวิธีที่นำเสนอโดย ดัชกรณ์ ต้นเจริญ และคณะ [27]

ในบทนี้ได้กล่าวถึงกระบวนการแยกส่วนภาพวิดีโอออกเป็นวัตถุอย่างละเอียด ซึ่งประกอบด้วย 5 กระบวนการคือ การแยกส่วนภาพเชิงพื้นที่ การแยกส่วนภาพเชิงเวลา กระบวนการตัดสินใจ กระบวนการข้างหลัง และการติดตามวัตถุ ทั้งยังกล่าวถึงการแยกส่วนภาพโดยใช้การทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลาเพื่อที่จะใช้เป็นตัวเปรียบเทียบ ส่วนในบทที่ 4 จะแสดงถึงผลลัพธ์ที่ได้จากการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยลำดับภาพแต่ละลำดับมีการเคลื่อนไหวที่แตกต่างกันออกไป และผลการเปรียบเทียบของการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์และการทดสอบสมมติฐานด้วย

บทที่ 4

ผลการทดลอง

ในบทนี้จะเป็นการแสดงผลการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยใช้เทคนิคการตัดสินใจแบบเบย์ ที่ได้กล่าวแล้วในบทที่ 3 การประเมินผลจะนำเสนอในรูปแบบของอัตวิสัย (subjective) ตารางและแผนภูมิ

ชุดภาพทั้งหมดที่ใช้ในการทดลองเป็นชุดภาพที่มีลักษณะเป็นวิดีโอคอนเฟอร์เรนซ์ มีวัตถุในภาพ 1 วัตถุคือตัวคนครึ่งตัว ซึ่งชุดภาพนั้นประกอบด้วยลำดับภาพมาตรฐาน 4 ชุดภาพและลำดับภาพที่จัดทำขึ้นในห้องปฏิบัติการ 3 ชุดภาพ ซึ่งชุดภาพแต่ละชุด มีลักษณะการเคลื่อนไหวของวัตถุและองค์ประกอบของพื้นหลังแตกต่างกัน สามารถบรรยายลักษณะการเคลื่อนไหวของวัตถุและพื้นหลังได้ดังต่อไปนี้

ลำดับภาพวิดีโอมาตรฐาน เป็นลำดับภาพที่จัดเก็บในรูปแบบ YCbCr 4:2:0 ปรับค่าเทรสโพลในกระบวนการแยกส่วนภาพเชิงเวลาไว้ที่ 0.02

1. ลำดับภาพ Claire พื้นหลังเป็นฉากที่มีสีสม่ำเสมอ มีการเคลื่อนไหวเพียงเล็กน้อยที่บริเวณปากและศีรษะต่อเนื่องตลอดลำดับภาพ
2. ลำดับภาพ Mother&Daughter พื้นหลังบางส่วนเป็นฉากที่มีสีสม่ำเสมอ และบางส่วนไม่สม่ำเสมอเนื่องจากเป็นส่วนของรูปภาพติดผนัง มีการเคลื่อนไหวที่บริเวณปาก ศีรษะ และลำตัวของแม่ทั้งยังมีการหันศีรษะของลูกด้วย
3. ลำดับภาพ Grandmother พื้นหลังบางส่วนเป็นผนังสีสม่ำเสมอ บางส่วนเป็นใบไม้ มีการเคลื่อนไหวน้อยมาก โดยเฉพาะเฟรมแรกๆ แทบไม่เคลื่อนไหวเลย
4. ลำดับภาพ Suzie มีพื้นหลังปรากฏอยู่น้อย เนื่องจากวัตถุมีขนาดใหญ่และมีการเคลื่อนไหวโดยการสะบัดผมด้วย

ลำดับภาพวิดีโอที่จัดทำขึ้นในห้องปฏิบัติการ เป็นลำดับภาพที่จัดเก็บในรูปแบบ RGB 24 บิตสี ไม่มีการบีบอัด ปรับค่าเทรสโพลในกระบวนการแยกส่วนภาพเชิงเวลาที่ 0.075

1. ลำดับภาพ Sak7 เป็นลำดับภาพที่วัตถุมีการเคลื่อนไหวตลอดเวลา มีการหันศีรษะ ตัวคนยืนหันหน้าเข้าหากล้อง มีฉากหลังเป็นเพดาน
2. ลำดับภาพ Sak35 เป็นลำดับภาพที่วัตถุมีการเคลื่อนไหวตลอดเวลา มีการเอียงศีรษะและขยับไหล่ ตัวคนนั่งหันหน้าเข้าหากล้อง ตำแหน่งของตัวคนอยู่ทางซ้ายของภาพ มีฉากหลังเป็นเพดานและห้องปฏิบัติการซึ่งมีรายละเอียดและความซับซ้อนภายในภาพสูง

3. ลำดับภาพ Sak39 เป็นลำดับภาพที่วัตถุมีการเคลื่อนไหวทางศีรษะและปากตลอดเวลา บริเวณหัวไหล่ไม่ค่อยเคลื่อนไหวมากนัก ตัวคนนั่งหันหน้าเข้าหากล้อง โดยตัววัตถุอยู่ตรงกลางของจอภาพ มีฉากหลังเป็นเพดานและห้องปฏิบัติการ

ผลการทดลองจะนำเสนอ ดังนี้

1. ผลการแยกส่วนภาพที่ได้จากกระบวนการข้างหลัง โดยพิจารณาข้อมูลขององค์ประกอบสีตามแบบจำลองสีทั้ง 3 แบบ (HSI, RGB, YCbCr) ซึ่งจะใช้ Mother&Daughter และ Sak7 เป็นตัวทดสอบ โดยจะเสนอเป็นตารางแสดง อัตราความผิดพลาดเฉลี่ย (Average Fault Rate) ตามแบบจำลองสีทั้ง 3 แบบ

2. แผนภูมิภาพแสดง อัตราความผิดพลาดต่อเฟรม ซึ่งอัตราความผิดพลาดนิยามได้ดังนี้

$$FR = (N_{OW} + N_{BW})/N \quad (4.1)$$

โดย FR คืออัตราความผิดพลาดต่อเฟรม N_{OW} คือ จำนวนจุดภาพที่เป็นวัตถุวิดีโอ แต่ถูกตัดสินว่าเป็นพื้นหลัง N_{BW} คือ จำนวนจุดภาพที่เป็นพื้นหลัง แต่ถูกตัดสินว่าเป็นวัตถุ N คือ จำนวนจุดภาพที่เป็นวัตถุ

ภาพอ้างอิงที่เป็นตัวเทียบคือภาพที่ได้จากการแยกส่วนภาพโดยพิจารณาทางสายตา เพราะฉะนั้นภาพที่ได้จากการแยกส่วนโดยพิจารณาทางสายตาจะมีค่าความผิดพลาดเป็นศูนย์

3. แผนภูมิภาพแสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมด ที่ได้จากเทคนิคการตัดสินใจแบบเบย์เทียบกับค่าที่ได้จากการแยกส่วนภาพโดยพิจารณาทางสายตา การแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยวิธีการทดสอบสมมติฐาน และการแยกส่วนเชิงพื้นที่และเวลา

4. ภาพที่ได้จากการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยพิจารณาทางสายตา การแยกส่วนที่ได้จากการทดสอบสมมติฐานและการแยกส่วนเชิงพื้นที่และเวลา โดยจะแสดงภาพในเฟรมที่มีค่าอัตราความผิดพลาดมากที่สุด ทั้งยังแสดงให้เห็นส่วนที่เกิดจากการตัดสินใจผิดพลาดด้วย

4.1 ผลการแยกส่วนภาพที่ได้จากกระบวนการข้างหลัง โดยพิจารณาข้อมูลขององค์ประกอบสีตามแบบจำลองสีทั้ง 3 แบบ

ตารางที่ 4.1 แสดงค่าอัตราผิดพลาดเฉลี่ยจากการใช้ข้อมูลตามแบบจำลองสีแต่ละแบบมาเป็นเวกเตอร์ลักษณะเด่น ซึ่งพบว่าในลำดับภาพ Mother&Daughter เมื่อนำองค์ประกอบของ Y และ Cb มาเป็นเวกเตอร์ลักษณะเด่นร่วมกันแล้วจะให้ค่าอัตราผิดพลาดเฉลี่ยน้อยที่สุด จากผลการทดลอง เมื่อใช้องค์ประกอบสี GB เป็นเวกเตอร์ลักษณะเด่นก็สามารถแยกส่วนได้เช่นกัน โดยมีค่าอัตราผิดพลาดเฉลี่ยเป็นร้อยละ 8.04 สาเหตุที่เป็นเช่นนี้เป็นเพราะว่าการแปลงองค์ประกอบสี

จากแบบจำลองสี YCbCr เป็นแบบจำลองสี RGB นั้นเป็นการแปลงโดยใช้เมตริกซ์การแปลงในบทที่ 2 ซึ่งเป็นการแปลงแบบเชิงเส้น ซึ่งสรุปได้ว่าองค์ประกอบสี YCb มีความสัมพันธ์เป็นเชิงเส้นกับองค์ประกอบสี GB

ตารางที่ 4.1 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ 100 ภาพ

องค์ประกอบสี	อัตราผิดพลาดเฉลี่ย (ร้อยละ)	
	ลำดับภาพ Mother&Daughter	ลำดับภาพ Sak7
Y	61.96	7.87
Cb	33.88	95.20
Cr	24.68	82.26
Y Cb	7.28	4.65
Y Cr	13.29	4.51
Cb Cr	12.16	78.58
Y Cb Cr	34.54	6.20
R	62.68	18.17
G	68.42	5.88
B	52.98	4.57
R G	12.25	5.60
R B	11.41	4.65
G B	8.04	4.73
R G B	13.07	6.20
H	21.48	82.17
S	38.13	65.20
I	62.68	8.58
H S	9.77	5.57
H I	13.08	4.91
S I	12.27	4.77
H S I	13.09	6.20

ในลำดับภาพ Sak7 พบว่า มืองค์ประกอบสีที่สามารถนำมาเป็นเวกเตอร์ลักษณะเด่นได้หลายองค์ประกอบ เช่น YCb YCr B RB GB I HI และ SI ซึ่งจากผลการทดลองนั้น YCb มีค่าอัตราผิดพลาดไม่เกินร้อยละ 10 จากผลการทดลองจะเห็นว่ามีเพียงองค์ประกอบสี YCb ที่สามารถแยกส่วนออกมาได้ทั้ง 2 ลำดับภาพ

4.2 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Claire

รูปที่ 4.1 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 4 เป็นต้นไป



รูปที่ 4.1 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 1 ถึง 6



รูปที่ 4.2 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 6



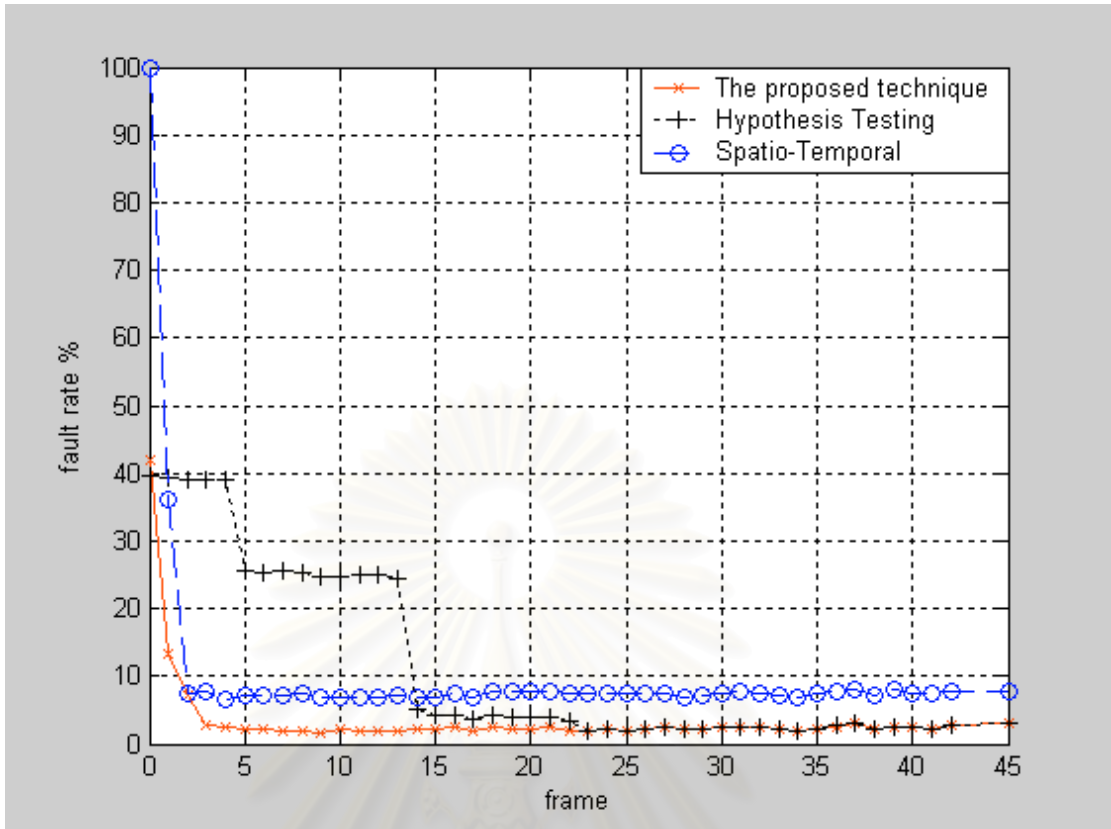
รูปที่ 4.3 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 6

รูปที่ 4.2 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.3 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 6 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 3

จากผลการทดลองจะเห็นว่า การแยกส่วนภาพเชิงพื้นที่และเวลาสามารถแยกส่วนได้เต็มตัวโดยใช้จำนวนเฟรมที่น้อยที่สุด การแยกส่วนภาพโดยการทดสอบสมมติฐานนั้นสามารถแยกได้เต็มตัวในเฟรมที่ 23

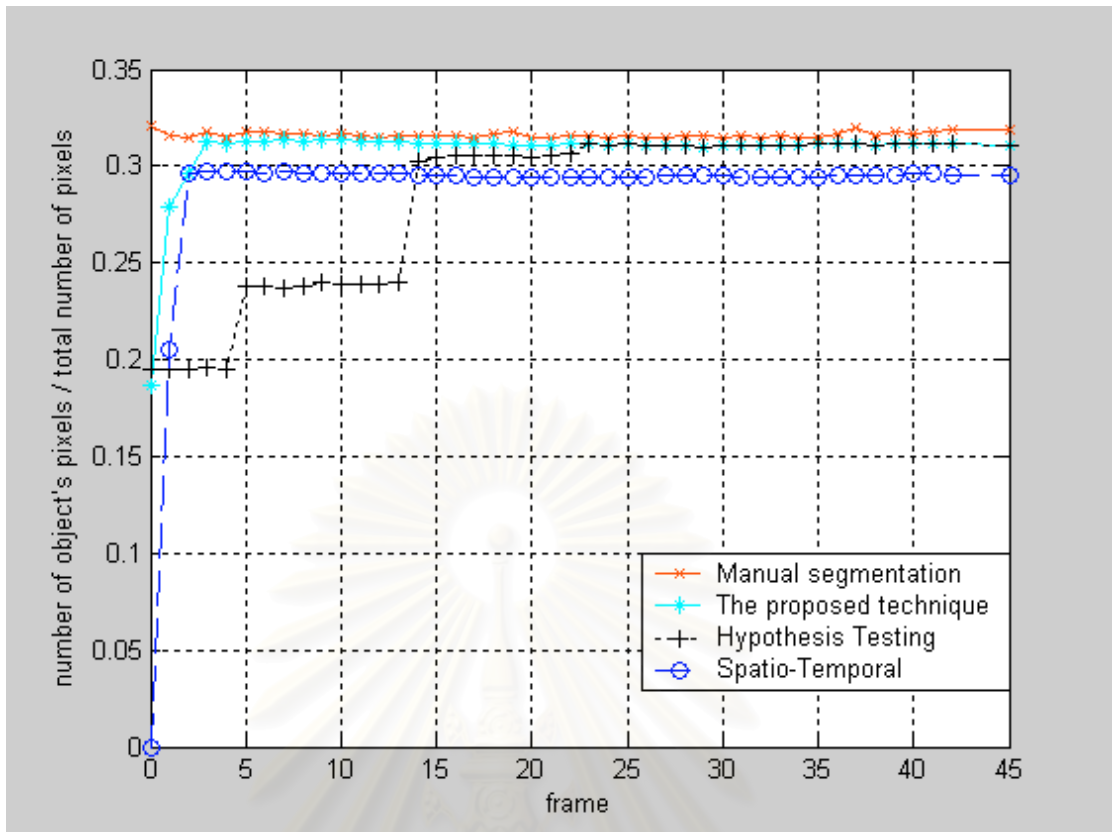
รูปที่ 4.4 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Claire ใน 45 เฟรมแรก ซึ่งจะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด ส่วนการแยกส่วนภาพโดยการทดสอบสมมติฐานนั้นจะมีค่าความผิดพลาดต่อเฟรมใกล้เคียงกับกรรมวิธีที่นำเสนอก็ต่อเมื่อหลังจากแยกส่วนได้เต็มตัวแล้ว (หลังจากเฟรมที่ 23) การแยกส่วนภาพเชิงพื้นที่และเวลานั้นมีค่าอัตราความผิดพลาดที่ต่ำกว่าร้อยละ 10 ตั้งแต่แยกส่วนได้เต็มตัว (หลังจากเฟรมที่ 3) การที่ค่าความผิดพลาดของการแยกส่วนภาพเชิงพื้นที่มีค่าร้อยละ 100 ในเฟรมแรกเพราะว่า ในเฟรมแรกนั้นวิธีนี้ยังไม่สามารถแยกส่วนใดๆออกมาได้เลย



รูปที่ 4.4 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Claire

รูปที่ 4.5 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Claire โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา

การแยกส่วนภาพโดยวิธีพิจารณาทางสายตานั้นสามารถอนุมานได้ว่าเป็นวิธีที่ให้ผลการแยกส่วนภาพโดยสมบูรณ์ที่สุด ซึ่งจากรูปที่ 4.5 นั้นจะเห็นว่า การแยกส่วนภาพโดยวิธีที่น่าเสนอนั้นหลังจากแยกส่วนได้เต็มตัวแล้ว มีจำนวนจุดภาพที่ตัดสินว่าเป็นวัตถุได้ใกล้เคียงกับวิธีพิจารณาทางสายตามากที่สุด โดยการแยกส่วนภาพด้วยวิธีทดสอบสมมติฐานนั้นจะให้ผลที่ใกล้เคียงหลังจากแยกส่วนได้เต็มตัวแล้ว ส่วนการแยกส่วนภาพเชิงเวลานั้น ก็ให้ผลที่ใกล้เคียงกับการแยกส่วนโดยวิธีพิจารณาทางสายตาเช่นกัน



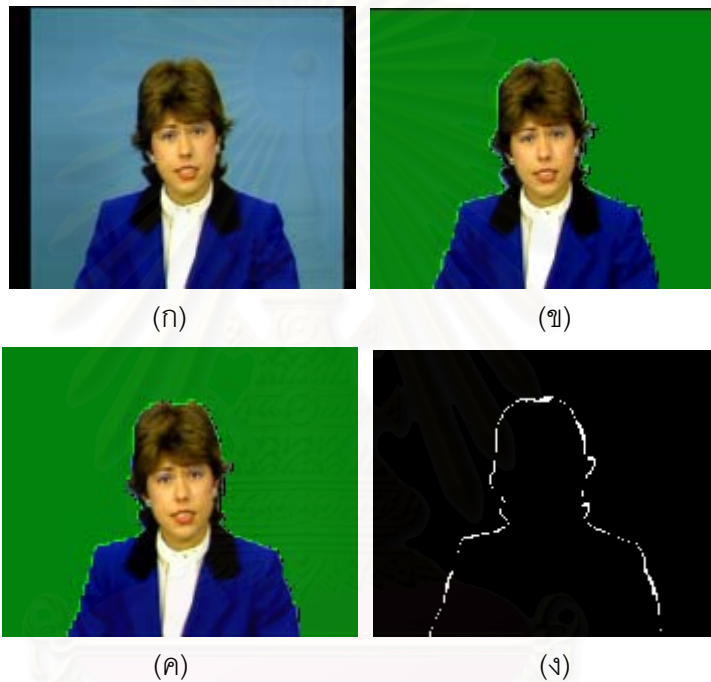
รูปที่ 4.5 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Claire โดยเปรียบเทียบกันทั้งสี่วิธี

ตารางที่ 4.2 แสดงค่าอัตราผิดพลาดเฉลี่ย 45 เฟรม โดยจะเห็นว่าการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์มีค่าต่ำกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานและวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา

4.2 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Claire 45 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสินใจแบบเบย์	3.54
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	11.57
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	10.08

รูปที่ 4.6(ก) แสดงภาพเฟรมที่ 37 ซึ่งเป็นภาพที่วิธีการแยกส่วนโดยกรรมวิธีที่นำเสนอมีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 3.02 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคที่นำเสนอมีค่า 0.31 และ ค่าที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.32 รูปที่ 4.6(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาในเฟรมที่ 37 รูปที่ 4.6(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 37 รูปที่ 4.6(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)



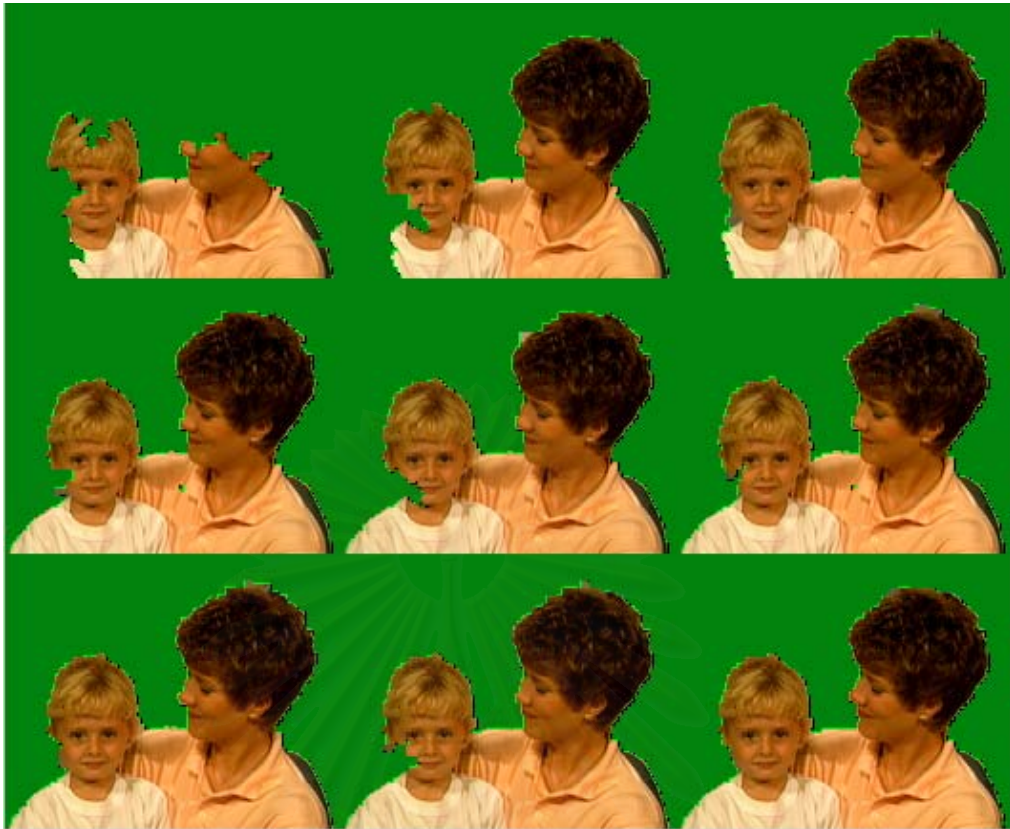
รูปที่ 4.6 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 37 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)

4.3 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Mother&Daughter

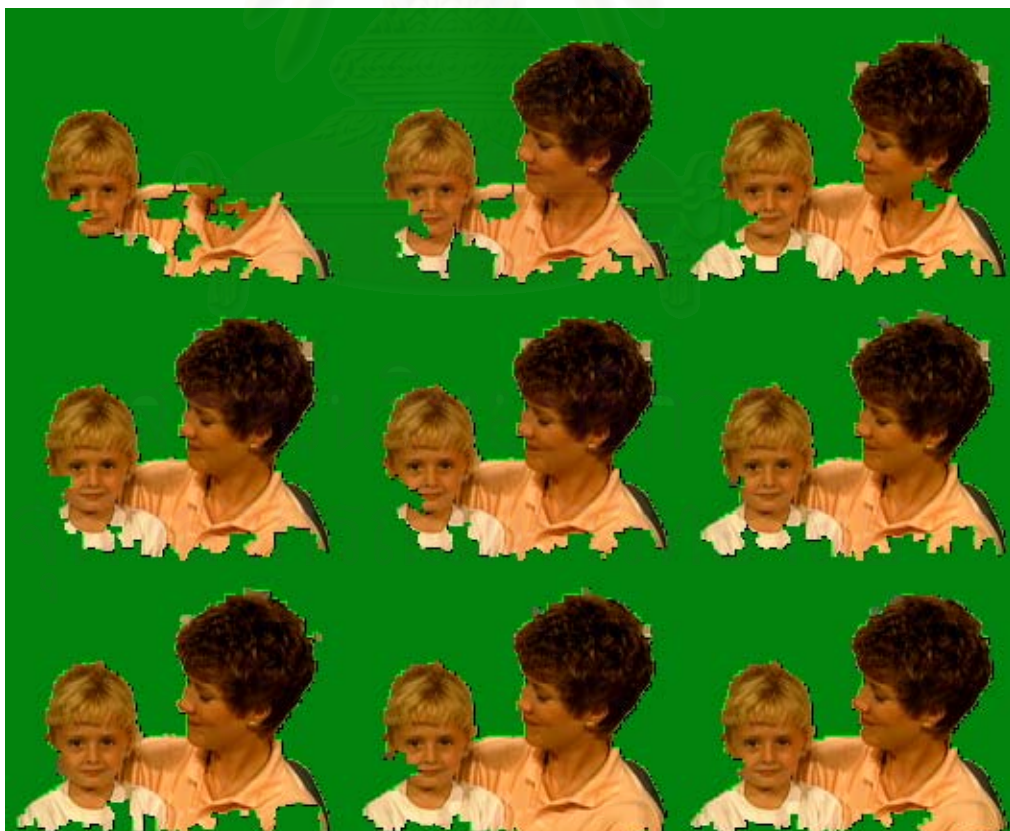
รูปที่ 4.7 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 7 เป็นต้นไป

รูปที่ 4.8 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.9 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว



รูปที่ 4.7 ผลการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์ เฟรมที่ 1 ถึง 9



รูปที่ 4.8 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9



รูปที่ 4.9 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

จากผลการทดลองจะเห็นว่า การแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์สามารถแยกวัตถุได้เต็มตัวตั้งแต่เฟรมที่ 7 ในขณะที่การแยกส่วนโดยวิธีทดสอบสมมติฐานนั้นสามารถแยกส่วนได้ครบในเฟรมที่ 89 ส่วนวิธีการแยกส่วนเชิงพื้นที่และเวลานั้นสามารถแยกได้ครบในเฟรมที่ 19 ดังแสดงในรูปที่ 4.10



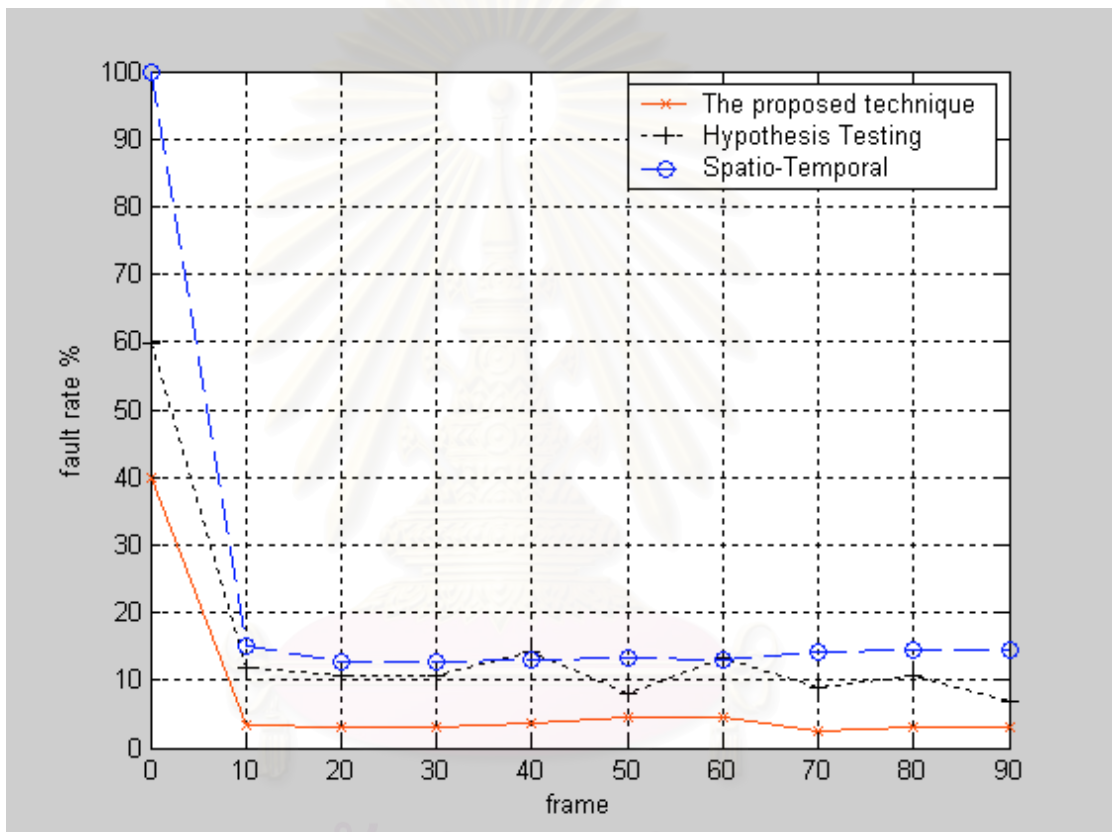
(ก)

(ข)

รูปที่ 4.10 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 89
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 19

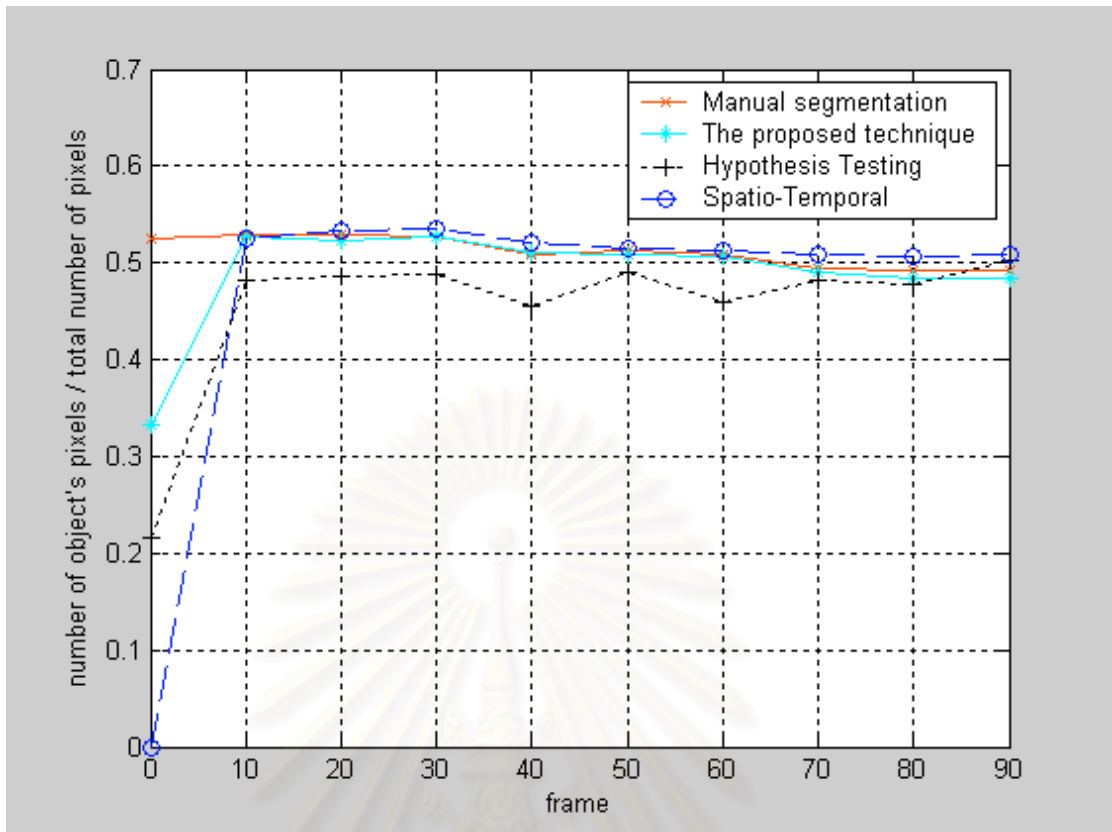
รูปที่ 4.11 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Mother&Daughter ใน 90 เฟรมแรก ซึ่งจะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด ส่วนการแยกส่วนภาพโดยการทดสอบสมมติฐานและ การแยกส่วนภาพเชิงพื้นที่และเวลานั้น จะมีค่าความผิดพลาดต่อเฟรมใกล้เคียงกัน

แม้ว่าในช่วงเฟรมแรกๆนั้น ทั้งสองวิธีนี้จะยังไม่สามารถแยกวัตถุได้ครบก็ตาม แต่ค่าอัตราผิดพลาดยังมีค่าต่ำ เนื่องจากว่าในช่วงเฟรมแรกๆแม้จะแยกยังไม่ได้ครบเต็มตัวแต่ก็ได้พื้นที่ของวัตถุโดยส่วนใหญ่แล้ว (ดังรูปที่ 4.8 และ รูปที่ 4.9)



รูปที่ 4.11 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Mother&Daughter

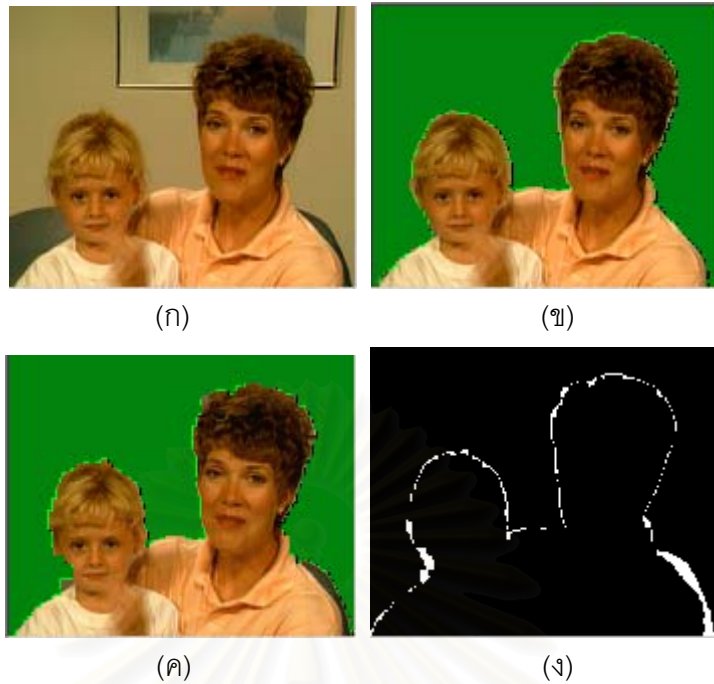
รูปที่ 4.12 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Mother&Daughter โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐาน และการแยกส่วนภาพเชิงพื้นที่และเวลา



รูปที่ 4.12 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Mother&Daughter โดยเปรียบเทียบกันทั้งสี่วิธี

จากรูปที่ 4.12 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดมีค่าใกล้เคียงกันทั้งสามวิธีการแยกส่วนภาพ ซึ่งจากรูปจะเห็นได้ชัดว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานนั้นมีแนวโน้มการลู่เข้าค่าที่ได้จากการแยกส่วนโดยวิธีพิจารณาทางสายตาเข้าที่เร็วที่สุด เนื่องจากว่าในลำดับภาพ Mother&Daughter นั้น การแยกส่วนโดยวิธีทดสอบสมมติฐานสามารถแยกส่วนได้ครบเข้าที่เร็วที่สุด (ใช้จำนวนเฟรมในการแยกส่วนมากที่สุด)

รูปที่ 4.13(ก) แสดงภาพเฟรมที่ 60 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้ครบเต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 4.65 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.507 และค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.509 รูปที่ 4.13(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาในเฟรมที่ 60 รูปที่ 4.13(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 60 รูปที่ 4.13(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว) ส่วนที่เกิดจากการตัดสินใจผิดพลาดนั้นเป็นส่วนของเก้าอี้



รูปที่ 4.13 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสีใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 60 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสีใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสีใจผิดพลาด (สีขาว)

ตารางที่ 4.3 แสดงค่าอัตราผิดพลาดเฉลี่ย 90 เฟรม โดยจะเห็นว่าการแยกส่วนโดยใช้เทคนิคการตัดสีใจแบบเบย์มีค่าต่ำกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานและวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา

ตารางที่ 4.3 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Mother&Daughter 90 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสีใจแบบเบย์	7.08
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	15.52
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	22.29

4.4 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Grandmother

รูปที่ 4.14 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสีใจแบบเบย์ ตั้งแต่เฟรมที่ 31 ถึง 39 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 35 เป็นต้นไป



รูปที่ 4.14 ผลการแยกส่วนภาพโดยเทคนิคการตัดลึนใจแบบเบย์ เพรมที่ 31 ถึง 39



รูปที่ 4.15 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เพรมที่ 31 ถึง 39

รูปที่ 4.15 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 31 ถึง 39 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.16 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 31 ถึง 39 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว



รูปที่ 4.16 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 31 ถึง 39

จากผลการทดลองพบว่าการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์สามารถแยกส่วนได้ครบเต็มตัวตั้งแต่เฟรมที่ 35 ซึ่งจากรูปที่ 4.14 จะเห็นว่ามีส่วนของเก้าอี้ติดมาด้วย เนื่องจากเส้นผมและเก้าอี้มีสีที่คล้ายกัน การที่ใช้จำนวนเฟรมถึง 35 เฟรมเพื่อที่จะใช้แยกส่วนได้เต็มตัวเป็นเพราะว่าการเคลื่อนไหวของวัตถุมีน้อยมากโดยเฉพาะส่วนหัวไหล่และตัว

การแยกส่วนโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลานั้น ในช่วง 39 เฟรมแรกไม่สามารถแยกส่วนได้ครบเต็มตัวเนื่องจากบริเวณหัวไหล่ยังไม่ค่อยเคลื่อนไหวมากนัก

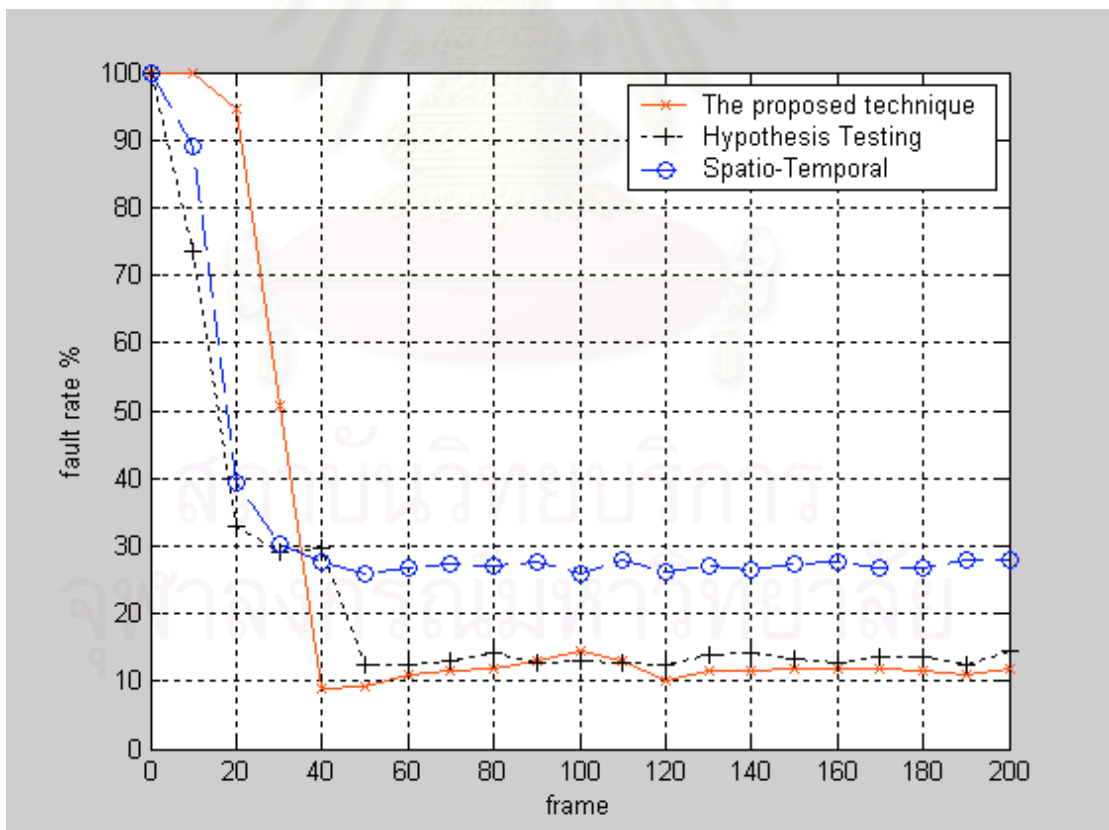
การแยกส่วนโดยวิธีทดสอบสมมติฐานสามารถแยกส่วนออกมาได้เต็มตัวในเฟรมที่ 49 ซึ่งมีพื้นหลังติดมามากพอสมควรเพราะกระบวนการข้างหลังและกระบวนการติดตามวัตถุได้สะสม

จำนวนเฟรมของวัตถุก่อนหน้าไว้มากเกินไป ดังนั้นจึงเกิดการสะสมของจุดภาพที่เป็นพื้นหลังไปเรื่อยๆ ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นไม่สามารถแยกส่วนออกมาได้ครบเต็มตัวเลย

รูปที่ 4.17 แสดงผลการแยกส่วนโดยวิธีทดสอบสมมติฐานในเฟรมที่ 49 และผลการแยกส่วนภาพเชิงพื้นที่และเวลาในเฟรม 866 ซึ่งเป็นเฟรมสุดท้าย



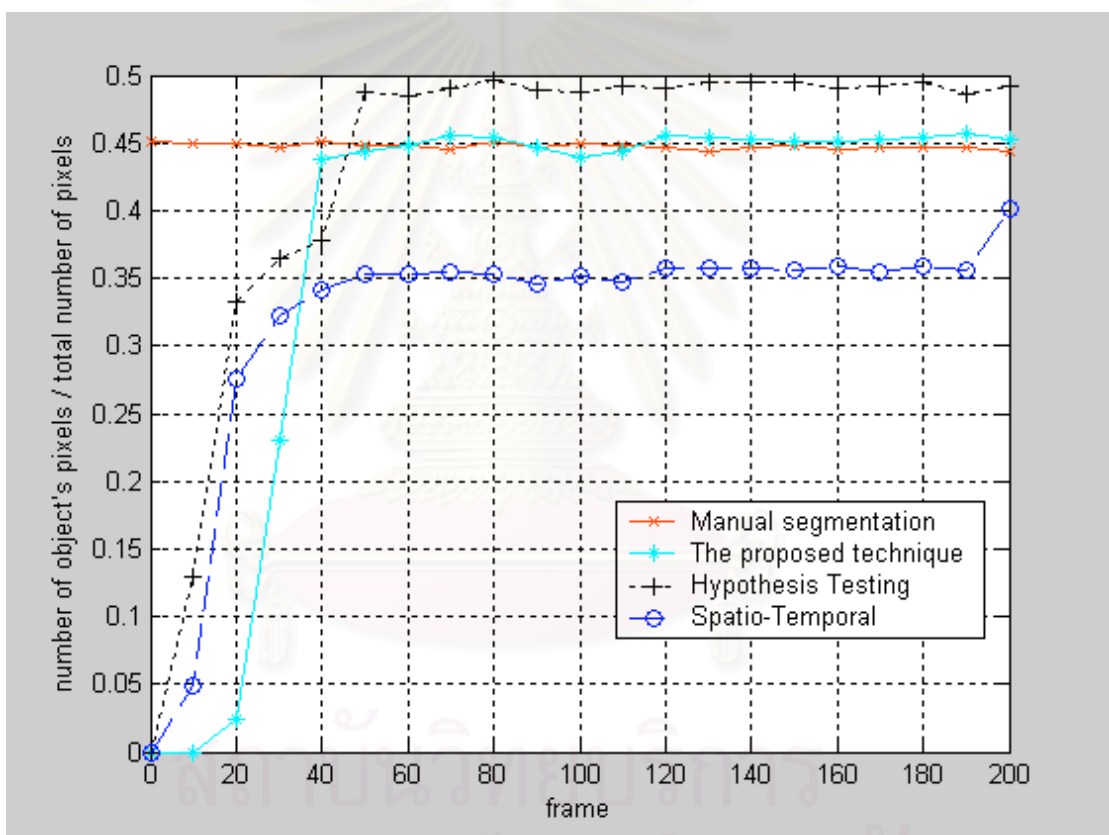
รูปที่ 4.17 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 49
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 866



รูปที่ 4.18 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Grandmother

รูปที่ 4.18 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Grandmother ในช่วง 200 เฟรม ซึ่งจะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยใกล้เคียงกับการแยกส่วนภาพโดยวิธีทดสอบสมมติฐาน ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลามีค่าความผิดพลาดต่อเฟรมสูงที่สุด การที่การแยกส่วนภาพเชิงพื้นที่และเวลามีค่าอัตราผิดพลาดมากเนื่องจากไม่สามารถแยกส่วนได้เต็มตัว

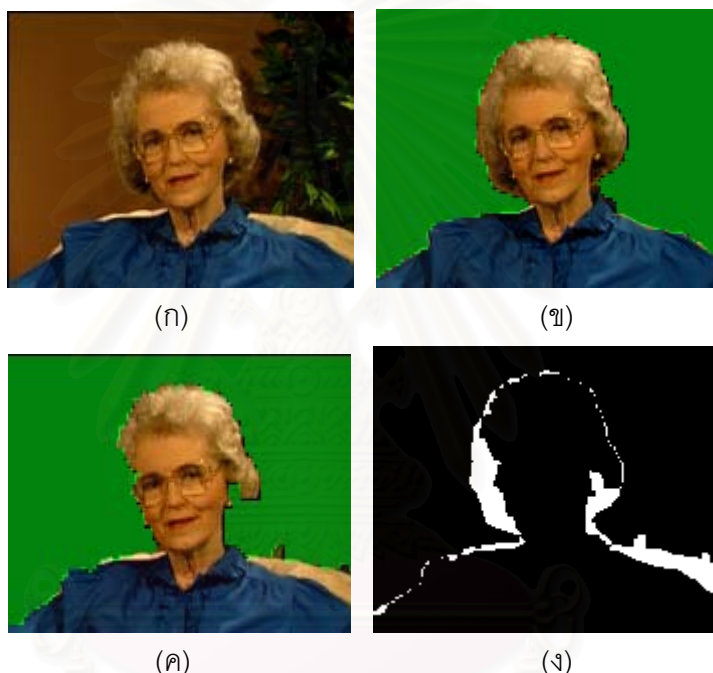
รูปที่ 4.19 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Grandmother โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา



รูปที่ 4.19 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Grandmother โดยเปรียบเทียบกันทั้งสี่วิธี

จากรูปที่ 4.19 การแยกส่วนโดยวิธีทดสอบสมมติฐานมีจำนวนจุดภาพที่ถูกตัดสินว่าเป็นวัตถุมากขึ้นไปเพราะเกิดการตัดสินใจพื้นที่หลังรวมเข้าไปด้วย ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นมีจำนวนจุดภาพที่ถูกตัดสินว่าเป็นวัตถุน้อยเนื่องจากไม่สามารถแยกส่วนได้ครบเต็มตัว

รูปที่ 4.20(ก) แสดงภาพเฟรมที่ 100 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 14.34 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.45 และค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.44 รูปที่ 4.20(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาในเฟรมที่ 100 รูปที่ 4.20(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 100 รูปที่ 4.20(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)



รูปที่ 4.20 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 100 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)

ตารางที่ 4.4 แสดงค่าอัตราผิดพลาดเฉลี่ย 20 เฟรม โดยจะเห็นว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานมีค่าต่ำกว่าการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์และวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา

สาเหตุที่ค่าอัตราความผิดพลาดโดยเฉลี่ยของการแยกส่วนโดยวิธีทดสอบสมมติฐานมีค่าน้อยกว่าการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์เนื่องจากว่าในช่วง 40 เฟรมแรกนั้นค่าอัตราความผิดพลาดโดยเฉลี่ยของการแยกส่วนโดยวิธีทดสอบสมมติฐานมีค่าน้อยกว่าการแยก

ส่วนโดยใช้เทคนิคการตัดสใจแบบเบย์ เพราะช่วงแรกของเทคนิคการตัดสใจแบบเบย์ไม่สามารถตรวจจับการเคลื่อนไหวได้มากนักต่างจากการแยกส่วนโดยวิธีทดสอบสมมติฐานที่ตรวจจับการเคลื่อนไหวได้มากกว่า จึงทำให้ปรากฏเห็นวัตถุได้ก่อน ดังนั้นมีผลทำให้ค่าอัตราความมืค่าต่ำกว่า

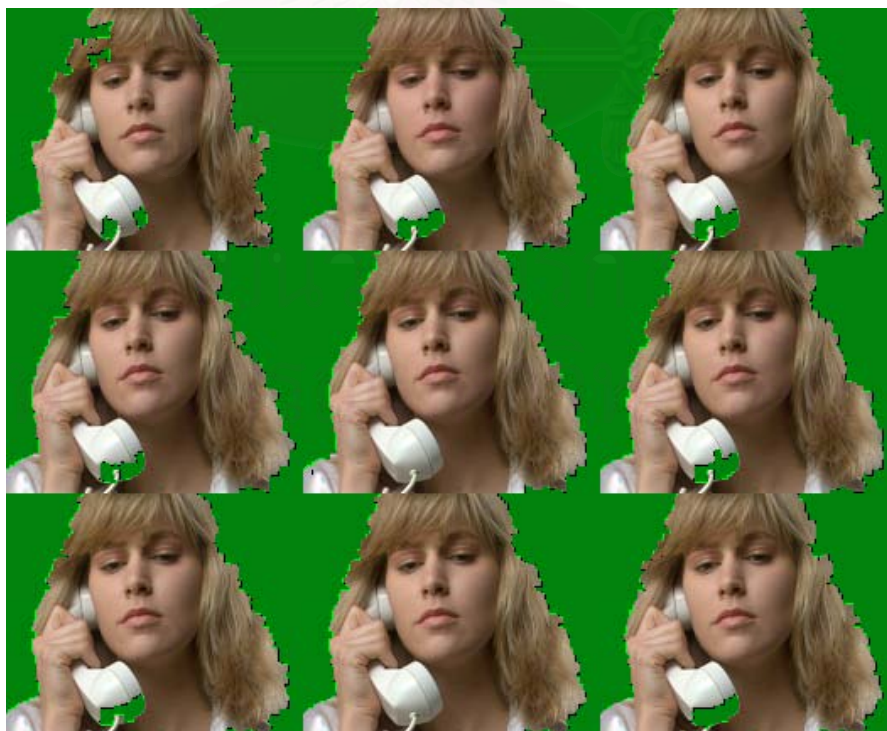
ตารางที่ 4.4 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Grandmother 200 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสใจแบบเบย์	25.80
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	22.70
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	34.20

4.5 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Suzie

รูปที่ 4.21 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสใจแบบเบย์ ตั้งแต่เฟรมที่ 2 ถึง 10 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 6 เป็นต้นไป

รูปที่ 4.22 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 2 ถึง 10 ซึ่งจะเห็นว่ายังไม่สามารถแยกส่วนได้เต็มตัว



รูปที่ 4.21 ผลการแยกส่วนภาพโดยเทคนิคการตัดสใจแบบเบย์ เฟรมที่ 2 ถึง 10



รูปที่ 4.22 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 2 ถึง 10



รูปที่ 4.23 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 2 ถึง 10

รูปที่ 4.23 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 2 ถึง 10 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว

จากผลการทดลองพบว่าการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์สามารถแยกส่วนได้ครบเต็มตัวตั้งแต่เฟรมที่ 6 ซึ่งจากรูปที่ 4.21 จะเห็นว่าบางส่วนของโทรศัพท์ด้านรับเสียงพูด ถูกตัดสินใจว่าเป็นพื้นหลังเนื่องจากเป็นส่วนที่มีสีคล้ายคลึงกับพื้นหลัง

การแยกส่วนโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลานั้น ในช่วง 10 เฟรมแรกไม่สามารถแยกส่วนได้ครบเต็มตัวเนื่องจากบริเวณใบหน้าและคอยังไม่ค่อยเคลื่อนไหวมากนัก

รูปที่ 4.24 แสดงผลการแยกส่วนโดยวิธีทดสอบสมมติฐานในเฟรมที่ 45 และผลการแยกส่วนภาพเชิงพื้นที่และเวลาในเฟรม 14 ซึ่งเป็นเฟรมสุดท้าย

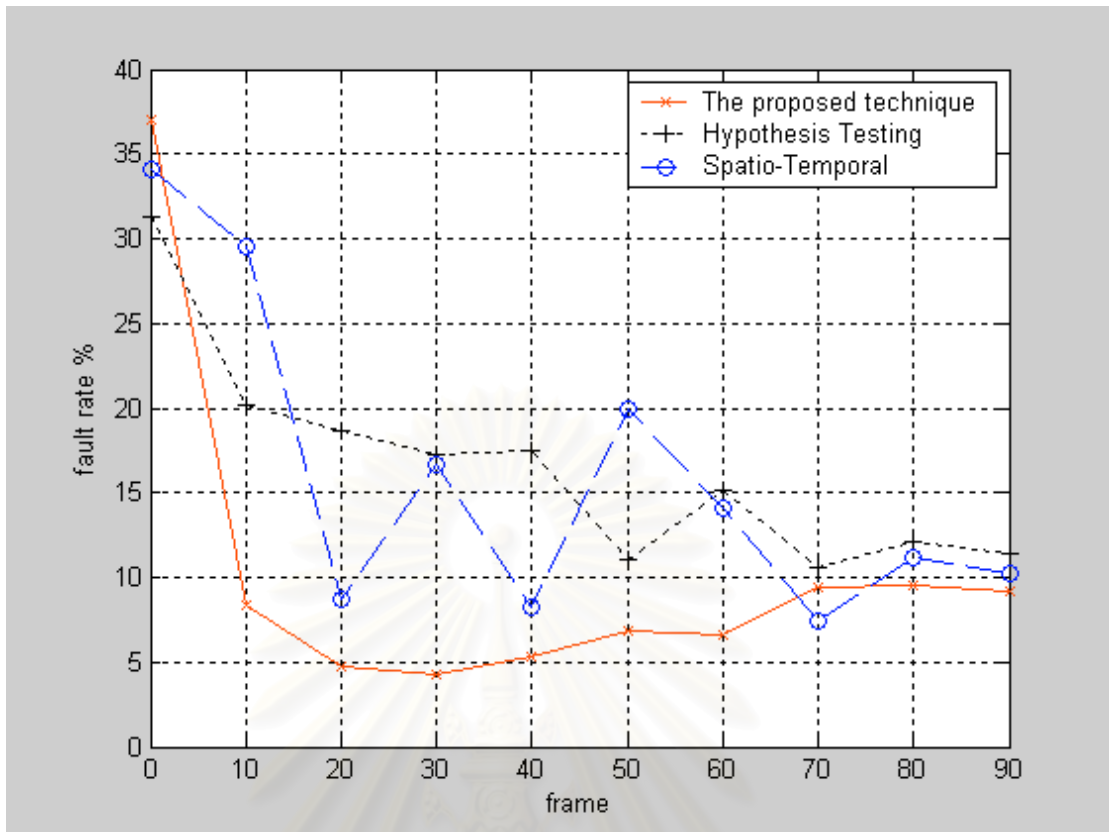


(ก)

(ข)

รูปที่ 4.24 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 45
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 14

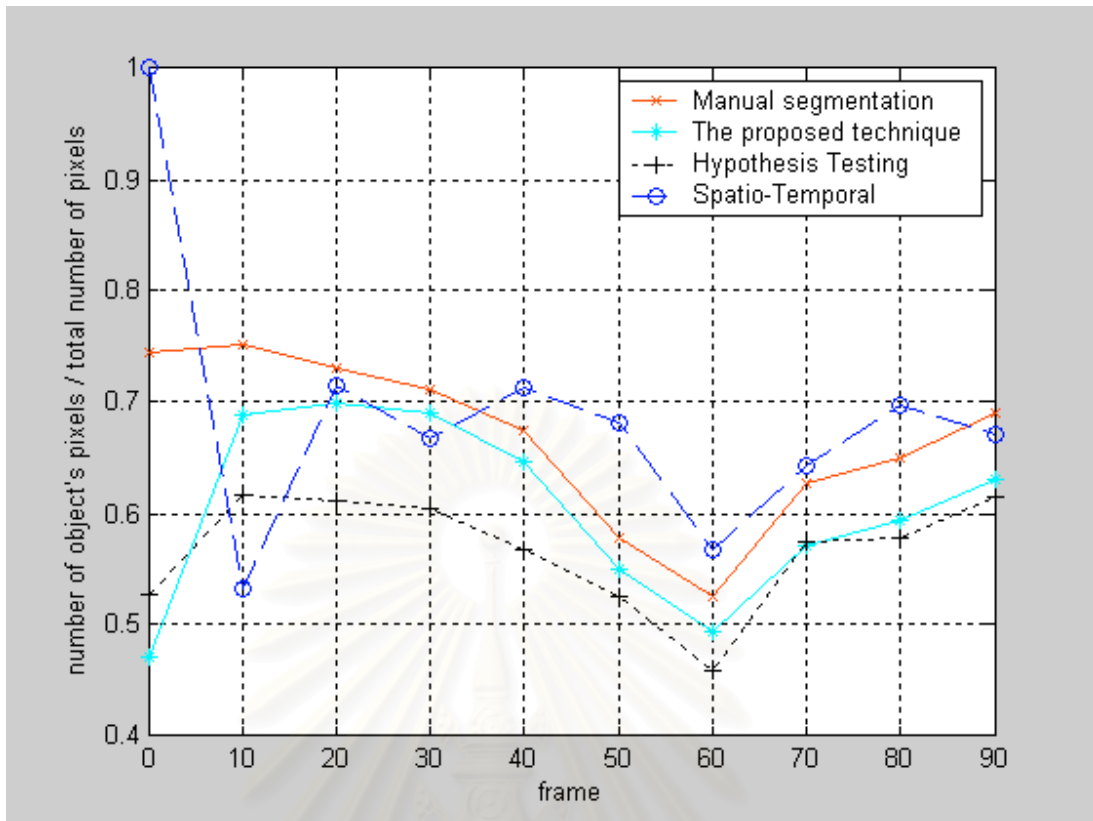
รูปที่ 4.25 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Suzie ในช่วง 90 เฟรม ซึ่งจะเห็นว่าในช่วง 70 เฟรมแรกกรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด แต่หลังจาก เฟรมที่ 70 ไปแล้วความผิดพลาดมีค่าใกล้เคียงกัน การที่ในช่วง 70 เฟรมแรกการแยกส่วนภาพเชิงพื้นที่และเวลา และการแยกส่วนโดยวิธีทดสอบสมมติฐานมีค่าอัตราผิดพลาดมากนั้นเนื่องจากไม่สามารถแยกส่วนได้เต็มตัว



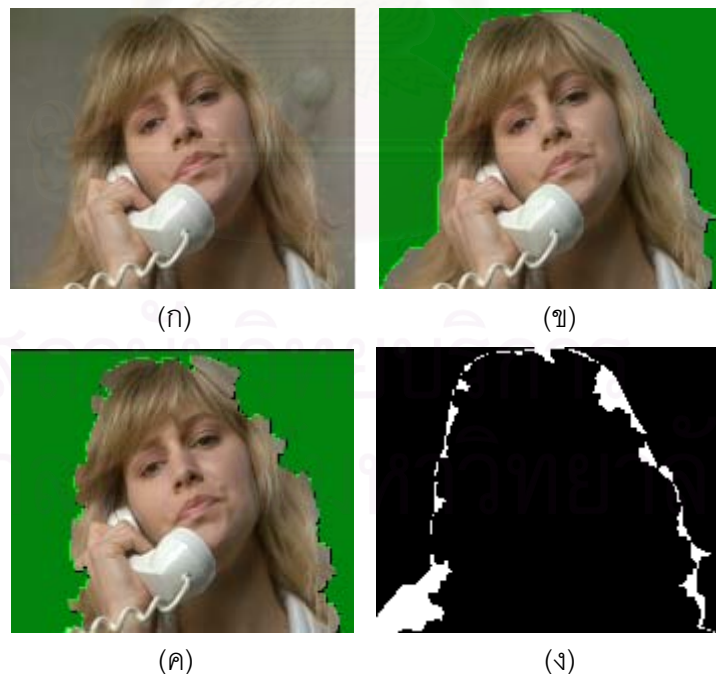
รูปที่ 4.25 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Suzie

รูปที่ 4.26 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Suzie โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา ซึ่งจะพบว่าค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้มีค่าใกล้เคียงทั้ง 4 วิธี (หลังจากเฟรมที่ 70 เป็นต้นไป) เพราะว่าทั้ง 4 วิธีสามารถแยกส่วนวัตถุส่วนใหญ่ออกมาได้

รูปที่ 4.27(ก) แสดงภาพเฟรมที่ 80 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 9.53 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.60 และค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.65 รูปที่ 4.27(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาในเฟรมที่ 80 รูปที่ 4.27(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 80 รูปที่ 4.27(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)



รูปที่ 4.26 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Suzie โดยเปรียบเทียบกันทั้งสี่วิธี



รูปที่ 4.27 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 80 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)

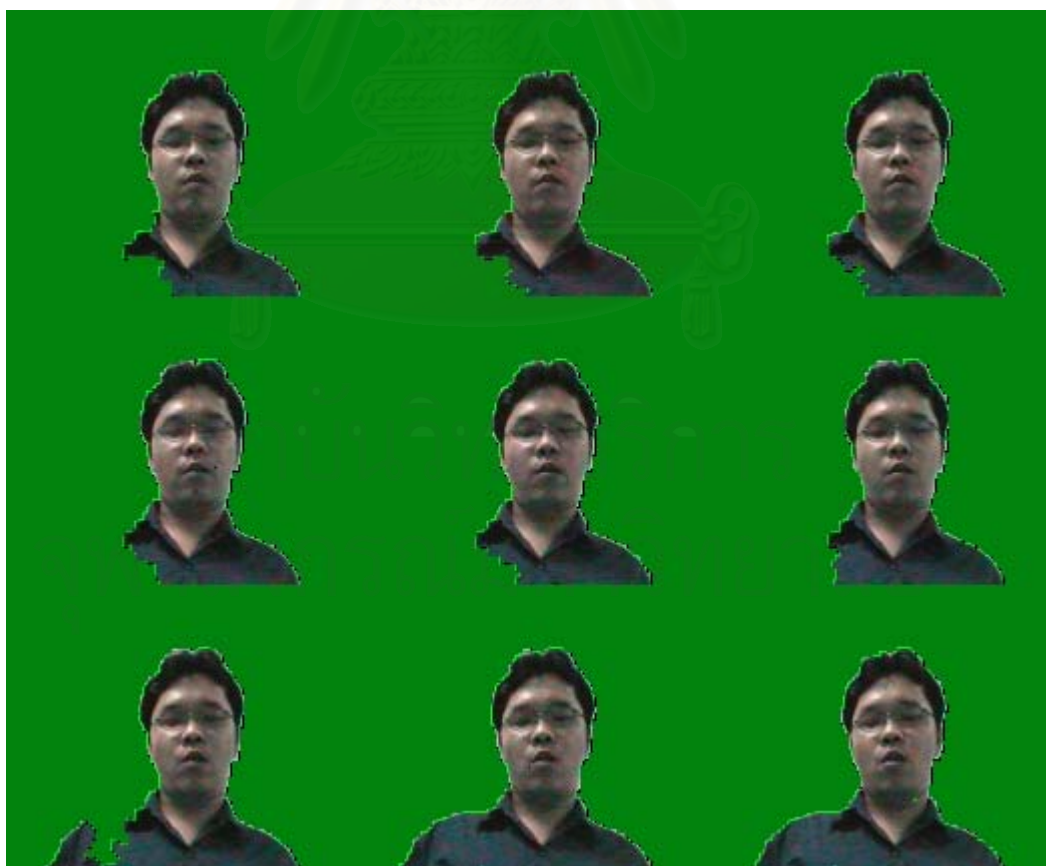
ตารางที่ 4.5 แสดงค่าอัตราผิดพลาดเฉลี่ย 90 เฟรม โดยจะเห็นว่าการแยกส่วนโดยใช้เทคนิคการตัดสีในใจแบบเบย์มีค่าต่ำกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานและวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา

ตารางที่ 4.5 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Suzie 90 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสีในใจแบบเบย์	10.13
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	16.50
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	16.02

4.6 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak7

รูปที่ 4.28 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ ตั้งแต่เฟรมที่ 3 ถึง 11 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 10 เป็นต้นไป



รูปที่ 4.28 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 3 ถึง 11

รูปที่ 4.29 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 3 ถึง 11 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.30 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 3 ถึง 11 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว

จากผลการทดลองจะเห็นว่าการแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์สามารถแยกวัตถุได้เต็มตัวตั้งแต่เฟรมที่ 10 ในขณะที่การแยกส่วนโดยวิธีทดสอบสมมติฐานนั้นสามารถแยกส่วนได้ครบในเฟรมที่ 21 แต่จะพบว่ามีส่วนของพื้นหลังถูกตัดสินใจว่าเป็นพื้นหน้าด้วยเช่น บริเวณหัวไหล่และศีรษะ ส่วนวิธีการแยกส่วนเชิงพื้นที่และเวลานั้นสามารถแยกได้ครบในเฟรมที่ 329 ซึ่งมีส่วนของพื้นหลังติดมาก่อนข้างมาก ดังแสดงในรูปที่ 4.31



รูปที่ 4.29 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 3 ถึง 11



รูปที่ 4.30 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 3 ถึง 11



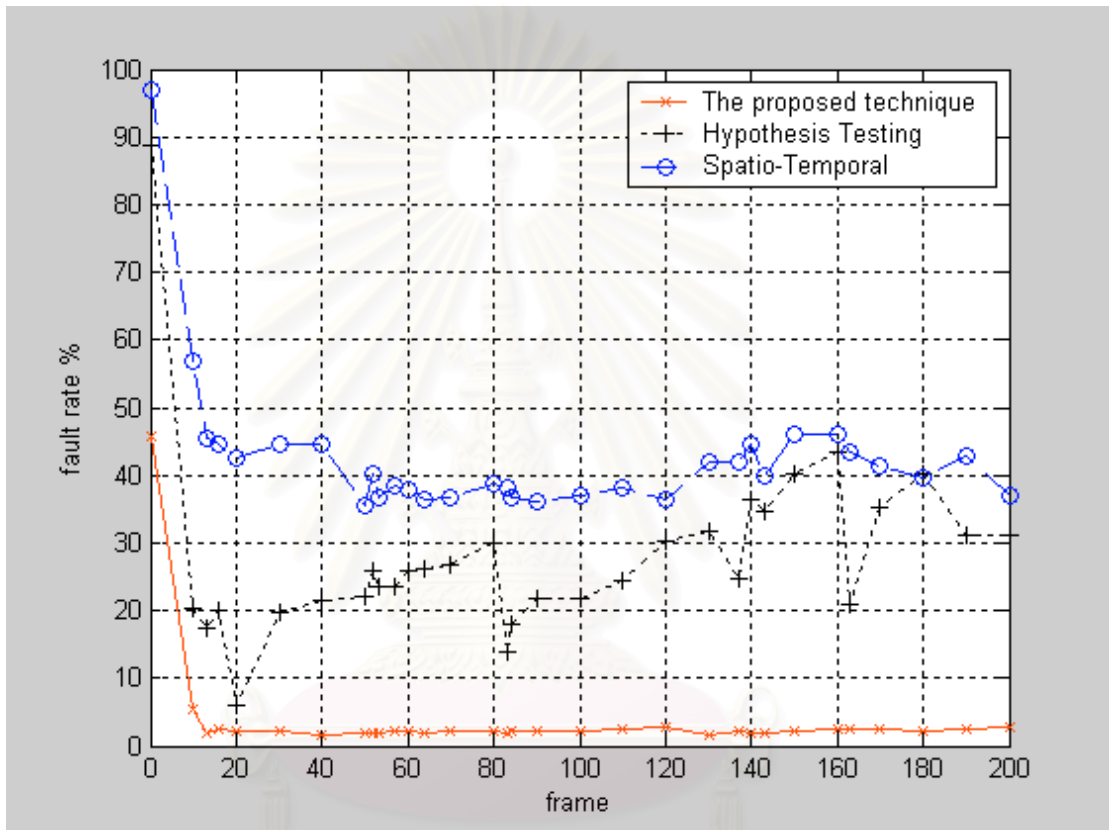
(ก)

(ข)

รูปที่ 4.31 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 21
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 329

รูปที่ 4.32 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak7 ใน 200 เฟรมแรก ซึ่ง จะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด ส่วนการแยกส่วนภาพโดยการ ทดสอบสมมติฐานมีค่าความผิดพลาดต่อเฟรมสูงกว่า และ การแยกส่วนภาพเชิงพื้นที่และเวลามี ค่าความผิดพลาดต่อเฟรมสูงที่สุด

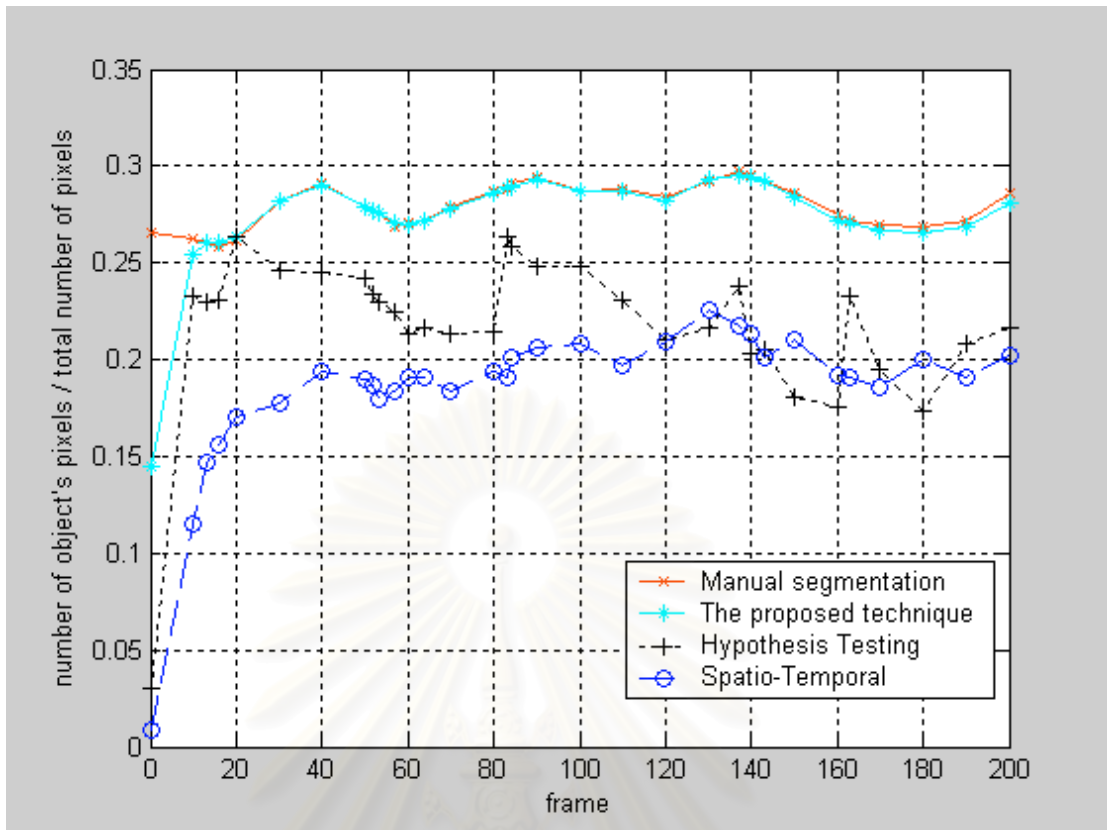
การที่อัตราความผิดพลาดต่อเฟรมของการแยกส่วนภาพโดยการทดสอบสมมติฐาน และการแยกส่วนภาพเชิงพื้นที่และเวลามีค่าความผิดพลาดต่อเฟรมสูง เนื่องจากว่ามี การตัดสินจุดภาพที่เป็นพื้นหลังให้เป็นวัตถุ ซึ่งมีจำนวนจุดที่การตัดสินผิดมากเกินไป สาเหตุเกิดจากระบบ การข้างหลังและกระบวนการติดตามวัตถุที่สะสมจำนวนเฟรมของวัตถุไว้มากเกินไป ดังนั้นเมื่อเกิด การตัดสินผิดในเฟรมแรกๆ ทำให้เกิดการสะสมของจุดที่ตัดสินผิดไปเรื่อยๆ



รูปที่ 4.32 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak7

รูปที่ 4.33 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak7 โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา

จากรูปที่ 4.33 การแยกส่วนโดยวิธีทดสอบสมมติฐานนั้นมีแนวโน้มการลู่เข้าค่าที่ได้จากการแยกส่วนโดยวิธีพิจารณาทางสายตาในช่วง 20 เฟรมแรก แต่หลังจากนั้นก็ไม่ได้ลู่เข้าซึ่งแสดงให้เห็นว่าไม่สามารถตรวจจับวัตถุได้ดีพอแม้จะแยกส่วนของตัวออกมาได้ครบในเฟรมที่ 20 แล้วก็ตาม

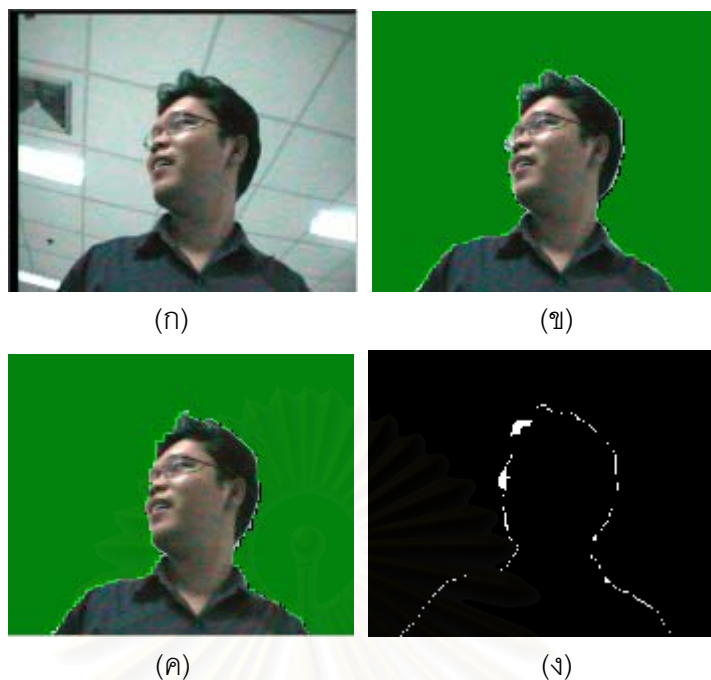


รูปที่ 4.33 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak7 โดยเปรียบเทียบกันทั้งสี่วิธี

การแยกส่วนภาพเชิงพื้นที่และเวลานั้นในช่วง 200 เฟรมแรกยังไม่ลู่เข้า เนื่องจากว่ายังไม่สามารถแยกส่วนได้ครบเต็มตัว

รูปที่ 4.34(ก) แสดงภาพเฟรมที่ 200 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัวโดยเทคนิคการตัดสินใจแบบเบย์ ซึ่งอัตราความผิดพลาดมีค่าร้อยละ 2.76 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.280 และ ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.286 รูปที่ 4.34(ข) แสดงผลการแยกส่วนภาพวิธีพิจารณาทางสายตาในเฟรมที่ 200 รูปที่ 4.34(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 200 รูปที่ 4.34(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)

ตารางที่ 4.6 แสดงค่าอัตราผิดพลาดเฉลี่ย 200 เฟรม โดยจะเห็นว่าการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์มีค่าต่ำกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานและวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา



รูปที่ 4.34 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสีใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 200 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสีใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสีใจผิดพลาด (สีขาว)

ตารางที่ 4.6 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak7 200 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสีใจแบบเบย์	3.63
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	28.03
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	42.61

4.7 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak35

รูปที่ 4.35 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสีใจแบบเบย์ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 6 เป็นต้นไป

รูปที่ 4.36 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว



รูปที่ 4.35 ผลการแยกส่วนภาพโดยเทคนิคการตัดสนใจแบบเบย์ เฟรมที่ 1 ถึง 9



รูปที่ 4.36 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9

รูปที่ 4.37 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว



รูปที่ 4.37 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

การแยกส่วนโดยวิธีทดสอบสมมติฐานนั้น ในช่วง 9 เฟรมแรกไม่สามารถแยกส่วนได้ครบเต็มตัวแต่สามารถแยกส่วนวัตถุออกมาได้เป็นส่วนใหญ่แล้ว ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลายังไม่สามารถแยกบริเวณตัวออกมาได้เลย เป็นเพราะในกระบวนการแยกส่วนภาพเชิงเวลายังไม่สามารถตรวจจับการเคลื่อนไหวของหัวไหล่และลำตัวได้

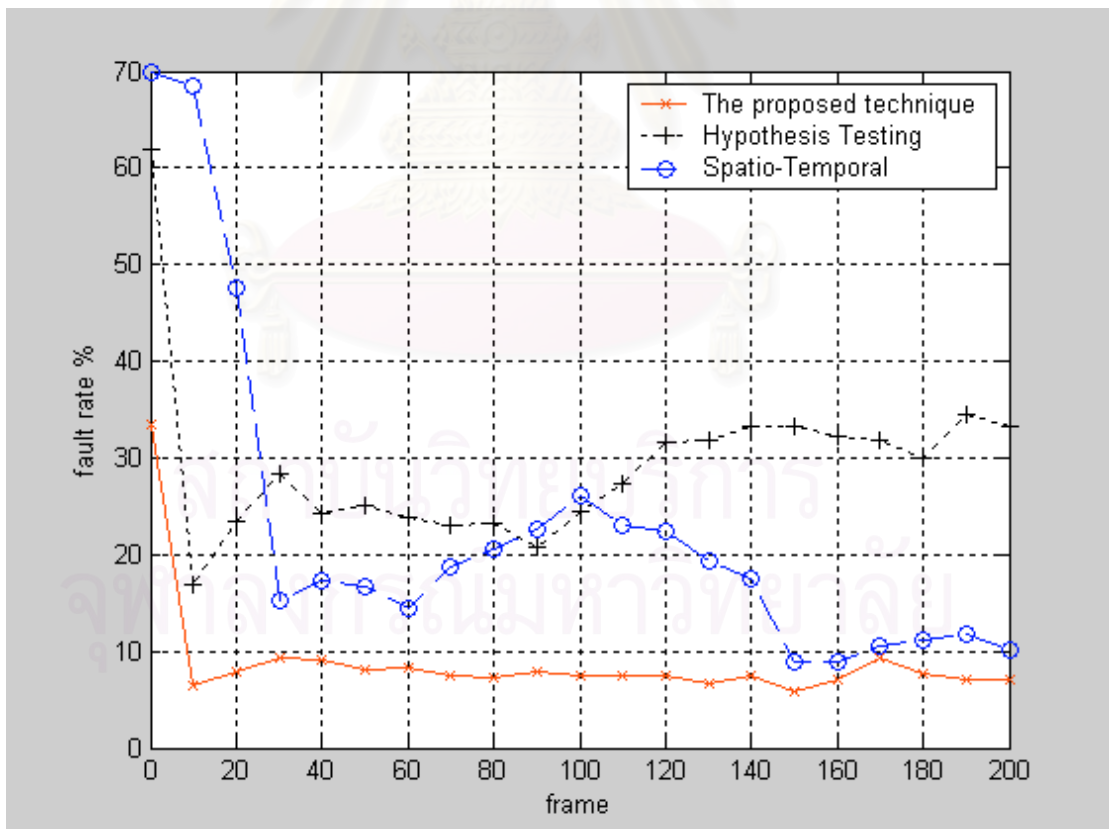
การแยกส่วนโดยวิธีทดสอบสมมติฐานสามารถแยกส่วนออกมาได้เต็มตัวในเฟรมที่ 25 ซึ่งมีพื้นหลังติดมากพอสมควรเป็นเพราะเกิดจากกระบวนการข้างหลังและกระบวนการติดตามวัตถุที่สะสมจำนวนเฟรมของวัตถุก่อนหน้ามากเกินไป ดังนั้นจึงเกิดการสะสมของจุดภาพที่เป็นพื้นหลังไปเรื่อยๆ

รูปที่ 4.38 แสดงผลการแยกส่วนโดยวิธีทดสอบสมมติฐานในเฟรมที่ 25 และผลการแยกส่วนภาพเชิงพื้นที่และเวลาในเฟรม 135



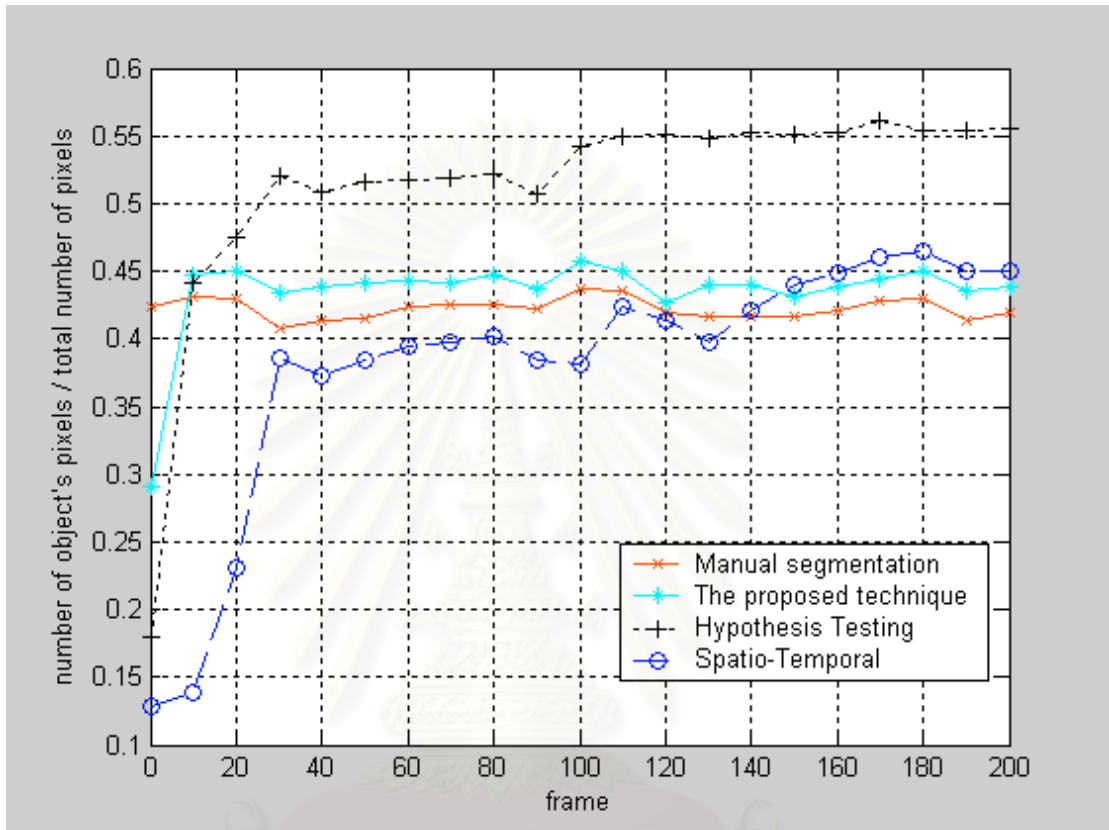
รูปที่ 4.38 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 25
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 135

รูปที่ 4.39 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak35 ในช่วง 200 เฟรม ซึ่งจะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด การแยกส่วนภาพเชิงพื้นที่ละเวลาจะมีอัตราผิดพลาดใกล้เคียงกับกรรมวิธีที่นำเสนอ ส่วนการแยกส่วนภาพโดยวิธีทดสอบสมมติฐานนั้นมีค่าความผิดพลาดสูงสุดเพราะเกิดการสะสมความผิดพลาดไว้ในกระบวนการติดตามวัตถุและกระบวนการข้างหลัง



รูปที่ 4.39 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak35

รูปที่ 4.40 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak35 โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา

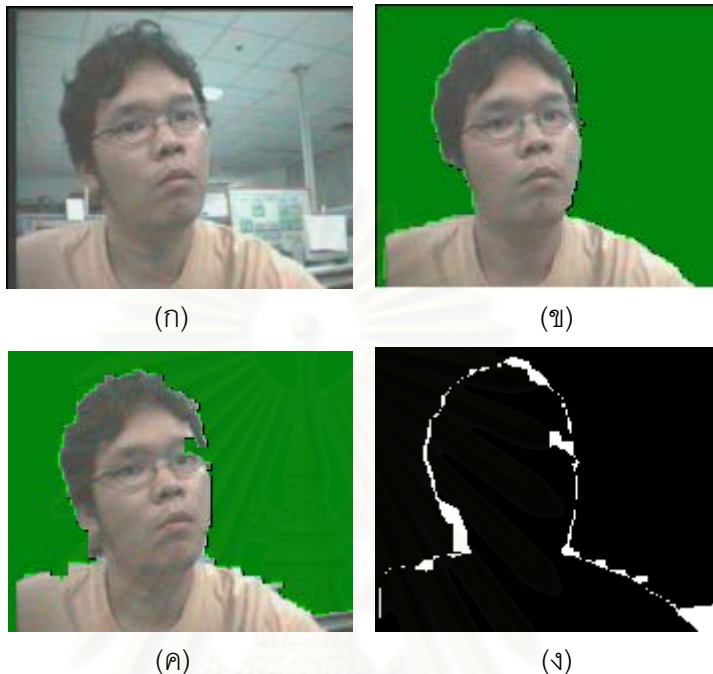


รูปที่ 4.40 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak35 โดยเปรียบเทียบกันทั้งสี่วิธี

จากรูปที่ 4.40 จะพบว่าค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้มีเพียงการแยกส่วนภาพโดยวิธีทดสอบสมมติฐานเท่านั้นที่มีค่าสูงที่สุด ส่วนอีก 3 วิธีนั้นมีค่าใกล้เคียงกัน

รูปที่ 4.41(ก) แสดงภาพเฟรมที่ 170 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 9.34 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.45 และค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตามีค่า 0.43 รูปที่ 4.41(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณา

ทางสายตาในเฟรมที่ 170 รูปที่ 4.41(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสนใจแบบเบย์
ในเฟรมที่180 รูปที่ 4.41(ง) แสดงส่วนที่เกิดจากการตัดสนใจผิดพลาด (สีขาว)



รูปที่ 4.41 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสนใจแบบเบย์และการแยกส่วนภาพโดย
วิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 170 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค)
การแยกส่วนโดยใช้เทคนิคการตัดสนใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสนใจผิดพลาด (สีขาว)

ตารางที่ 4.7 แสดงค่าอัตราผิดพลาดเฉลี่ย 200 เฟรม โดยจะเห็นว่าการแยกส่วนโดยวิธี
ทดสอบสมมติฐานมีค่าต่ำกว่าการแยกส่วนโดยใช้เทคนิคการตัดสนใจแบบเบย์และวิธีการแยกส่วน
ภาพเชิงพื้นที่และเวลา

ตารางที่ 4.7 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak35 200 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสนใจแบบเบย์	8.82
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	29.26
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	22.93

4.8 ผลการแยกส่วนภาพของลำดับภาพวิดีโอ Sak39

รูปที่ 4.42 แสดงผลจากการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าสามารถแยกส่วนได้เต็มตัวตั้งแต่เฟรมที่ 9 เป็นต้นไป

รูปที่ 4.43 แสดงผลจากการแยกส่วนภาพโดยการทดสอบสมมติฐาน ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้เต็มตัว

รูปที่ 4.44 แสดงผลจากการแยกส่วนภาพเชิงพื้นที่และเวลา ตั้งแต่เฟรมที่ 1 ถึง 9 ซึ่งจะเห็นว่าไม่สามารถแยกส่วนได้ครบเต็มตัว

จากผลการทดลองพบว่า การแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์สามารถแยกส่วนได้ครบเต็มตัวตั้งแต่เฟรมที่ 6 ซึ่งจากรูปที่ 4.42 จะเห็นว่ามีส่วนของวัตถุและบางส่วนของพื้นหลังถูกตัดสีผิดอยู่บ้าง ทั้งนี้เนื่องมาจากในลำดับภาพวิดีโอ Sak39 นี้ มีรายละเอียดของภาพ โดยเฉพาะส่วนที่เป็นพื้นหลังมาก และสีของพื้นหลังบางส่วนก็คล้ายกับพื้นหน้า



รูปที่ 4.42 ผลการแยกส่วนภาพโดยเทคนิคการตัดสีในใจแบบเบย์ เฟรมที่ 1 ถึง 9



รูปที่ 4.43 ผลการแยกส่วนภาพโดยการทดสอบสมมติฐาน เฟรมที่ 1 ถึง 9



รูปที่ 4.44 ผลจากการแยกภาพส่วนเชิงพื้นที่และเวลา เฟรมที่ 1 ถึง 9

การแยกส่วนโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลานั้น ในช่วง 9 เฟรมแรกไม่สามารถแยกส่วนได้ครบเต็มตัวเนื่องจากบริเวณหัวไหล่ยังไม่ค่อยเคลื่อนไหวมากนัก แต่การแยกส่วนโดยวิธีทดสอบสมมติฐานยังสามารถแยกส่วนออกมาได้เต็มตัวในเฟรมที่ 89 ซึ่งมีพื้นหลังติดมามากพอสมควรเป็นเพราะเกิดจากระบวนการข้างหลังและกระบวนการติดตามวัตถุที่สะสมจำนวนเฟรมของวัตถุก่อนหน้านี้ไว้มากเกินไป ดังนั้นจึงเกิดการสะสมของจุดภาพที่เป็นพื้นหลังไปเรื่อยๆ ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นไม่สามารถแยกส่วนออกมาได้ครบเต็มตัวเลย

รูปที่ 4.45 แสดงผลการแยกส่วนโดยวิธีทดสอบสมมติฐานในเฟรมที่ 89 และผลการแยกส่วนภาพเชิงพื้นที่และเวลาในเฟรม 195 ซึ่งเป็นเฟรมสุดท้าย



(ก)

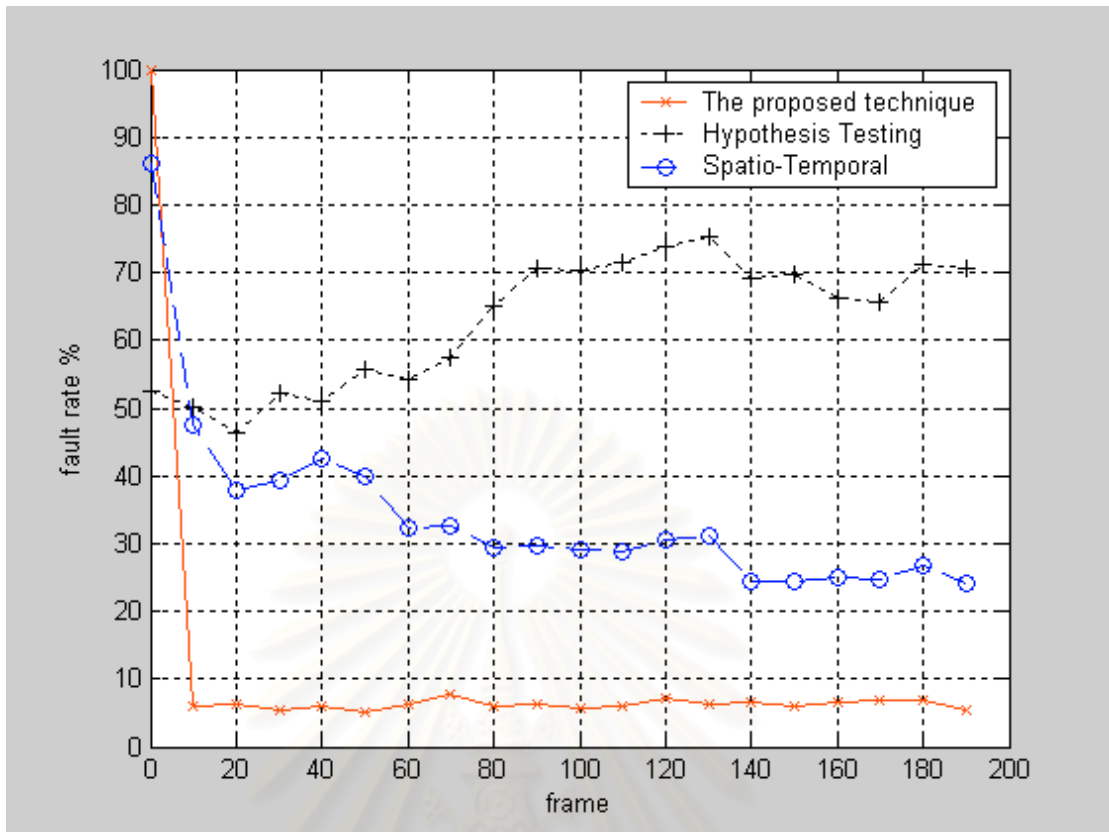
(ข)

รูปที่ 4.45 ผลการแยกส่วนวัตถุได้ครบเต็มตัว โดย (ก) วิธีทดสอบสมมติฐาน เฟรมที่ 89
(ข) วิธีเชิงพื้นที่และเวลา เฟรมที่ 195

รูปที่ 4.46 แสดงอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak39 ในช่วง 190 เฟรม ซึ่งจะเห็นว่ากรรมวิธีที่นำเสนอมีค่าอัตราผิดพลาดต่อเฟรมน้อยที่สุด ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลามีค่าความผิดพลาดต่อเฟรมสูงกว่า และ การแยกส่วนภาพโดยการทดสอบสมมติฐานมีค่าความผิดพลาดต่อเฟรมสูงที่สุด

การที่การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานมีค่าอัตราผิดพลาดมากขึ้นเนื่องจากว่าสะสมความผิดพลาดไว้ ส่วนอัตราผิดพลาดของการแยกส่วนภาพเชิงพื้นที่และเวลามีค่าน้อยกว่านั้นเพราะว่ายังไม่สามารถแยกส่วนออกมาได้ครบเต็มตัว

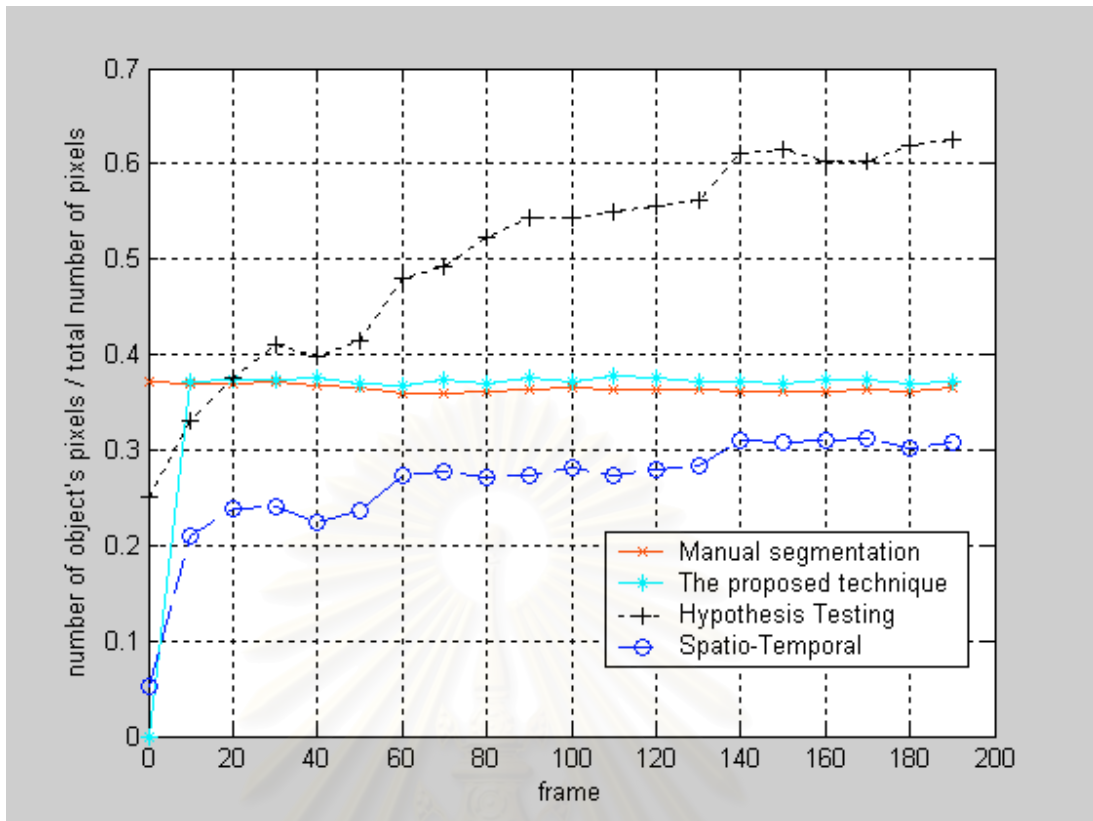
รูปที่ 4.47 แสดงค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak39 โดยเปรียบเทียบการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา การแยกส่วนภาพด้วยวิธีการตัดสินใจแบบเบย์ การแยกส่วนภาพโดยวิธีทดสอบสมมติฐานและการแยกส่วนภาพเชิงพื้นที่และเวลา



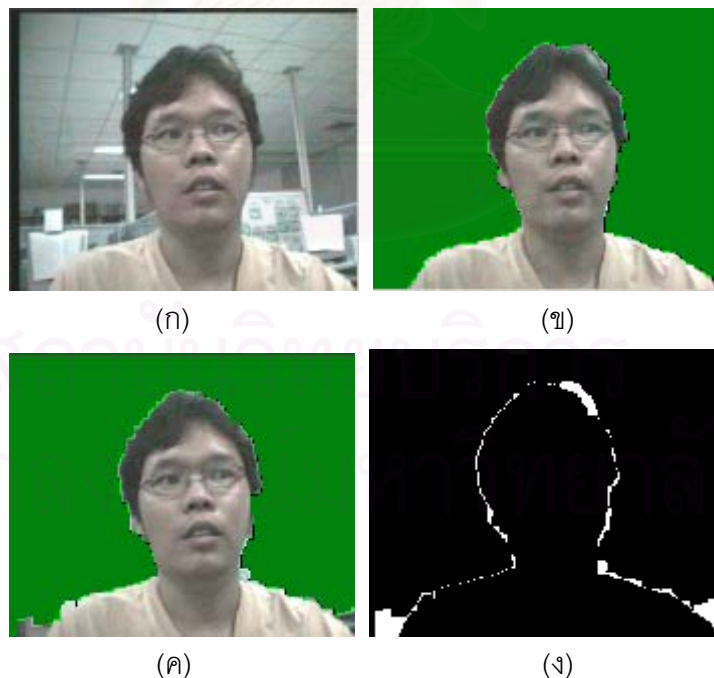
รูปที่ 4.46 ค่าอัตราผิดพลาดต่อเฟรมของลำดับภาพวิดีโอ Sak39

จากรูปที่ 4.47 การแยกส่วนโดยวิธีทดสอบสมมติฐานนั้นมีแนวโน้มไม่ถูเข้าค่าที่ได้จากการแยกส่วนโดยวิธีพิจารณาทางสายตา เนื่องจากว่าเกิดการสะสมจำนวนจุดที่ถูกตัดสินให้เป็นวัตถุมากขึ้นเรื่อยๆ ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นไม่ถูเข้า เนื่องจากว่ายังไม่สามารถแยกส่วนได้ครบเต็มตัว

รูปที่ 4.48(ก) แสดงภาพเฟรมที่ 70 ซึ่งเป็นภาพที่มีอัตราความผิดพลาดสูงที่สุดหลังจากแยกวัตถุได้เต็มตัว โดยอัตราความผิดพลาดมีค่าร้อยละ 7.74 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากเทคนิคการตัดสินใจแบบเบย์มีค่า 0.37 และค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดที่ได้จากการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาตามีค่า 0.36 รูปที่ 4.48(ข) แสดงผลการแยกส่วนภาพโดยวิธีพิจารณาทางสายตาในเฟรมที่ 70 รูปที่ 4.48(ค) แสดงผลการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ในเฟรมที่ 70 รูปที่ 4.48(ง) แสดงส่วนที่เกิดจากการตัดสินใจผิดพลาด



รูปที่ 4.47 ค่าอัตราส่วนของจุดภาพที่ถูกตัดสินว่าเป็นวัตถุต่อจำนวนจุดภาพทั้งหมดของลำดับภาพวิดีโอ Sak39 โดยเปรียบเทียบกันทั้งสี่วิธี



รูปที่ 4.48 เปรียบเทียบผลการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์และการแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ก) ภาพเฟรมที่ 70 (ข) การแยกส่วนภาพโดยวิธีพิจารณาทางสายตา (ค) การแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์ (ง) ส่วนที่เกิดจากการตัดสินใจผิดพลาด (สีขาว)

ตารางที่ 4.8 แสดงค่าอัตราผิดพลาดเฉลี่ย 190 เฟรม โดยจะเห็นว่าการแยกส่วนโดยใช้เทคนิคการตัดสินใจแบบเบย์มีค่าต่ำกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานและวิธีการแยกส่วนภาพเชิงพื้นที่และเวลา

ตารางที่ 4.8 ค่าอัตราความผิดพลาดโดยเฉลี่ยจากภาพวิดีโอ Sak39 190 ภาพ

	อัตราผิดพลาดเฉลี่ย (ร้อยละ)
การแยกส่วนโดยวิธีการตัดสินใจแบบเบย์	10.89
การแยกส่วนโดยวิธีทดสอบสมมติฐาน	62.96
การแยกส่วนโดยวิธีเชิงพื้นที่และเวลา	34.31

4.9 การวิเคราะห์ความซับซ้อนของการแยกส่วนภาพทั้งสามวิธี

ในการแยกส่วนภาพแต่ละวิธีมีกระบวนการหลายกระบวนการที่แตกต่างกันและมีบางกระบวนการที่เหมือนกันซึ่งในการวิเคราะห์ความซับซ้อนนี้จะเป็นการชี้ให้เห็นว่าแต่ละกระบวนการและแต่ละวิธีนั้นมีความซับซ้อนเป็นอย่างไรเมื่อนำมาเปรียบเทียบกัน

จากกรรมวิธีการแยกส่วนภาพทั้ง 3 วิธีนั้นสามารถเปรียบเทียบเป็นกระบวนการได้ดังตารางที่ 4.9 ซึ่งจะเห็นว่าการแยกส่วนภาพเชิงพื้นที่และเวลามีความซับซ้อนน้อยที่สุดเพราะมีกระบวนการน้อยที่สุด

การแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์ซับซ้อนกว่าการแยกส่วนภาพเชิงพื้นที่และเวลาในส่วนการตัดสินใจเพราะมีการใช้กฎของเบย์เข้ามาช่วยในส่วนการตัดสินใจ

การแยกส่วนโดยวิธีการทดสอบสมมติฐานนั้นซับซ้อนที่สุดเพราะว่ามีการประมาณการเคลื่อนที่แบบโกลบอลและมีการค้นหาแบบเต็ม ทั้งนี้ในการแยกส่วนภาพเชิงเวลายังใช้การทดสอบสมมติฐานซึ่งมีความซับซ้อนมากกว่าการใส่มาสก์ตรวจจับการเคลื่อนไหวมาก

ตารางที่ 4.10 แสดงจำนวนเฟรมที่ใช้การแยกส่วนได้เต็มตัวเป็นครั้งแรกโดยเปรียบเทียบกันทั้ง 3 วิธี ซึ่งจะเห็นว่า การแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์ใช้จำนวนเฟรมน้อยที่สุดยกเว้นลำดับภาพ Claire เท่านั้นที่การแยกส่วนภาพเชิงพื้นที่และเวลาใช้จำนวนเฟรมน้อยที่สุดคือ ใช้เพียง 3 เฟรม แต่อย่างไรก็ตาม การแยกส่วนภาพโดยเทคนิคการตัดสินใจแบบเบย์นั้นใช้เพียง 6 เฟรมเท่านั้นซึ่งไม่ต่างกันมากนัก

ตารางที่ 4.9 การเปรียบเทียบกระบวนการแยกส่วนภาพของทั้ง 3 กรรมวิธี

การแยกส่วนภาพโดยวิธี	การทดสอบสมมติฐาน	การตัดสินใจแบบเบย์	เชิงเวลาและพื้นที่
การประมาณการเคลื่อนที่	แบบโกลบอล และค้นหาแบบเต็ม	ไม่มี	ไม่มี
การแยกส่วนภาพเชิงพื้นที่	ตัวดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ	ตัวดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ	ตัวดำเนินการ เชิงสัญญาณ และกรรมวิธีสันปันน้ำ
การแยกส่วนภาพเชิงเวลา	การทดสอบสมมติฐาน	มาส์กการเคลื่อนไหว	มาส์กการเคลื่อนไหว
การตัดสินใจ	ใช้เงื่อนไข	ใช้กฎของเบย์ และเงื่อนไข	ใช้เงื่อนไข
การติดตามวัตถุ	มี	มี	มี

ตารางที่ 4.10 เปรียบเทียบจำนวนเฟรมที่ใช้ในการแยกส่วนได้เต็มตัว

ลำดับภาพ	เฟรมที่สามารถแยกส่วนได้เต็มตัวเป็นเฟรมแรกโดยวิธี		
	การตัดสินใจ แบบเบย์	การทดสอบ สมมติฐาน	เชิงพื้นที่ และเวลา
Claire	6	23	3
Mother&Daughter	7	89	19
Grandmother	35	49	866
Suzie	6	45	14
Sak7	10	21	319
Sak35	6	25	135
Sak39	9	89	195

ตารางที่ 4.11 แสดงเวลาที่ใช้ในการแยกส่วนภาพโดยเปรียบเทียบกัน 3 วิธี ซึ่งประมวลผลด้วยเครื่องคอมพิวเตอร์เพนเทียม 4 ความเร็ว 2 GHz หน่วยความจำ 512 เมกกะไบท์ และเขียนโปรแกรมด้วย Matlab เวอร์ชัน 6.1 ผลการทดลองพบว่า การแยกส่วนเชิงพื้นที่และเวลานั้นใช้เวลาเฉลี่ยน้อยที่สุดคือประมาณ 2 วินาทีต่อเฟรม ส่วนการแยกส่วนภาพวิดีโอด้วยวิธีการทดสอบสมมติฐานใช้เวลาเฉลี่ยมากที่สุดคือประมาณ 14 วินาทีต่อเฟรม เนื่องจากว่ามีความซับซ้อนสูงที่สุดเมื่อเทียบกับอีก 2 วิธี การแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยการตัดสินใจแบบเบย์จะใช้เวลาเฉลี่ยประมาณ 6 วินาที ซึ่งจะเห็นว่าเวลาที่ใช้ในการแยกส่วนภาพวิดีโอออกเป็นวัตถุโดยการตัดสินใจแบบเบย์นั้นจะใช้เวลามากกว่าการแยกส่วนเชิงพื้นที่และเวลาประมาณ 3 เท่า ส่วนการแยกส่วนภาพวิดีโอออกเป็นวัตถุด้วยวิธีการทดสอบสมมติฐานจะใช้เวลามากกว่าการแยกส่วนเชิงพื้นที่และเวลาประมาณ 6 เท่า ซึ่งการแยกส่วนภาพวิดีโอออกเป็นวัตถุด้วยวิธีการทดสอบสมมติฐานจะใช้เวลามากกว่าการแยกส่วนโดยการตัดสินใจแบบเบย์ประมาณ 2 เท่า

จากข้อมูลในตารางที่ 4.11 ทำให้สรุปได้ว่า การแยกส่วนภาพโดยการทดสอบสมมติฐานนั้นมีความซับซ้อนมากที่สุด ส่วนการแยกส่วนภาพเชิงพื้นที่และเวลานั้นมีความซับซ้อนน้อยที่สุดเมื่อเปรียบเทียบกันทั้ง 3 วิธี ซึ่งสอดคล้องกับผลจากตารางที่ 4.9

ตารางที่ 4.11 เวลาที่ใช้ในการประมวลผลในแต่ละวิธีโดยเฉลี่ยจาก 100 เฟรม

ลำดับภาพ	เวลาที่ใช้ในการประมวลผลในแต่ละเฟรม (วินาที/เฟรม) (เฉลี่ยจาก 100 เฟรม)		
	การทดสอบสมมติฐาน	การตัดสินใจแบบเบย์	เชิงพื้นที่และเวลา
Claire	13.35	5.88	1.82
Mother&Daughter	13.95	6.25	2.03
Grandmother	13.34	5.73	2.00
Suzie	14.03	6.40	2.03
Sak7	13.80	6.13	1.99
Sak35	13.75	6.11	1.93
Sak39	14.39	6.12	2.03

บทที่ 5

สรุปและข้อเสนอแนะ

5.1 สรุป

วิทยานิพนธ์นี้ เสนอการแยกส่วนภาพวิดีโอออกเป็นวัตถุ จุดหลักของกรรมวิธีที่นำเสนอคือการนำทฤษฎีการตัดสินใจแบบเบย์มาใช้ตัดสินว่าจุดใดในภาพเป็นวัตถุหรือพื้นหลัง โดยกรรมวิธีที่นำเสนอเป็นแบบอัตโนมัติ ลักษณะของวิดีโอที่สอดคล้องกับกรรมวิธีที่นำเสนอนั้นจะต้องมี 2 วัตถุ คือ บุคคลเคลื่อนไหว และ พื้นหลังที่อยู่นิ่ง กรรมวิธีที่นำเสนอนั้นประกอบด้วย 5 กระบวนการคือ การแยกส่วนภาพทางพื้นที่ การแยกส่วนภาพทางเวลา กระบวนการตัดสินใจ กระบวนการซ้ำหลัง และการติดตามวัตถุ

ผลการทดลองแสดงให้เห็นว่า กรรมวิธีที่นำเสนอนั้นใช้จำนวนเฟรมในการแยกส่วนได้เต็มตัวน้อยกว่า เมื่อเทียบกับการทดสอบสมมติฐาน และการแยกส่วนเชิงพื้นที่และเวลา ทั้งยังมีค่าอัตราความผิดพลาดเฉลี่ยโดยส่วนใหญ่น้อยที่สุดอีกด้วย เมื่อเปรียบเทียบความซับซ้อนแล้วพบว่าการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์มีความซับซ้อนปานกลาง โดยมีความซับซ้อนน้อยกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานแต่มีความซับซ้อนมากกว่าการแยกส่วนเชิงพื้นที่และเวลา โดยเวลาที่ใช้ในการแยกส่วนโดยเทคนิคการตัดสินใจแบบเบย์มีค่าเฉลี่ยประมาณ 6 วินาทีต่อเฟรม ซึ่งมากกว่าการแยกส่วนเชิงพื้นที่และเวลาประมาณ 2 เท่า แต่น้อยกว่าการแยกส่วนโดยวิธีทดสอบสมมติฐานอยู่ประมาณครึ่งหนึ่ง ปัจจัยที่มีผลต่อกรรมวิธีเป็นอย่างยิ่งคือ องค์ประกอบของสี คุณภาพของระบบรับภาพวิดีโอ (กล้องที่เลือกใช้) การเคลื่อนไหวของวัตถุ และความแตกต่างกันของสีระหว่างวัตถุกับพื้นหลัง จากการทดลองยังพบว่าองค์ประกอบที่ให้ผลการแยกส่วนออกมาดีที่สุดคือ YCb ในส่วนของกล้องวิดีโอ นั้น ถ้าระบบรับภาพมีความไวต่อการเปลี่ยนแปลงของความเข้มแสงและความเข้มสีมาก (หรือปรับโฟกัสอัตโนมัติได้) แล้วจะส่งผลกระทบต่อผลลัพธ์ที่ได้จากการแยกส่วนภาพทางเวลาอย่างยิ่ง เช่น อาจจะถูกกำหนดว่าเป็นส่วนเคลื่อนไหวทั้งที่จริงแล้วเป็นส่วนที่อยู่นิ่ง ปัจจัยนี้มีผลมากเมื่อลำดับวิดีโอเป็นการผลิตขึ้นจากกล้องวิดีโอเว็บ เพราะกล้องวิดีโอเว็บมีระบบรับภาพที่มีคุณภาพต่ำและไวต่อการเปลี่ยนแปลงความเข้มแสงมาก ทั้งยังปรับโฟกัสได้เองอัตโนมัติอีกด้วย การเคลื่อนไหวของวัตถุก็เช่นกัน ถ้าเคลื่อนไหวน้อยจนเกิดความแตกต่างไม่เกินค่าเทรชโฮลหรือไม่เคลื่อนไหวแล้วทำให้ตรวจจับความเคลื่อนไหวไม่ได้หรือได้ก็ไม่สามารถตรวจจับวัตถุได้เต็มตัว ปัจจัยในเรื่องความแตกต่างกันของสีระหว่างวัตถุกับพื้นหลังนั้นสำคัญอย่างยิ่ง เพราะแม้จะแยกส่วนภาพทางเวลาได้ดีเพียงไรก็ไม่สามารถแยกว่าเป็นวัตถุหรือพื้นหลังได้ ถ้าสีของวัตถุและพื้นหลังเป็นสีเดียวกันหรือใกล้เคียงกัน ซึ่งปัญหานี้เกิดจากลักษณะเด่นที่นำมาใช้มีสหสัมพันธ์ต่อกันมาก ดังนั้นจึงควรรหาลักษณะเด่นอื่นๆ มาช่วยและควรมีสหสัมพันธ์ต่อกันน้อย

5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

แนวทางการวิจัยที่ควรได้รับการพัฒนาต่อ

1. ควรนำเวกเตอร์การเคลื่อนที่มาร่วมเป็นส่วนหนึ่งในการแยกส่วน
2. ควรค้นหาลักษณะพิเศษอื่นๆ ที่ใช้เป็นลักษณะเด่นนอกเหนือจาก ความเข้มแสง ความเข้มสี เช่น อาจจะเป็นข้อมูลของรูปร่างวัตถุ ข้อมูลของไบหน้า เพื่อช่วยในการแยกส่วน
3. ควรประยุกต์กรรมวิธีต่อไปเพื่อใช้กับลักษณะวิดีโอแบบอื่นๆ ที่นอกเหนือจากวิดีโอคอนเฟอร์เรนซ์ เช่น ระบบตรวจจับการเคลื่อนไหว เป็นต้น



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

1. 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 14-17 Nov. 1988 : 646-648.
2. Fabrice HEITZ and Patrick BOUTHEMY. Motion Estimation and Segmentation using A Global Bayesian Approach. International Conference on Acoustics, Speech, and Signal Processing 3-6 April 1990 : 2305-2308.
3. Chiou-Shann Fuh and Petros Maragos. Affine Models for Image Matching and Motion Detection. International Conference on Acoustics, Speech, and Signal Processing 14-17 April 1991 : 2409-2412.
4. Samir Shaltar and Nader M. Namazi. Application of The EM Technique to Estimation of Affine Modeled Image Motion. Proceedings of the 35th Midwest Symposium on Circuits and Systems 1992 : 9-12.
5. John Y. A. Wang and Edward H. Adelson. Representing Moving Images with Layers. IEEE Tran. Image Processing Vol.3 May 1994 : 625-638.
6. Raynard O. Hinds and Thrasyvoulos N. Pappas. An Adaptive Clustering Algorithm for Segmentation of Video Sequences. International Conference on Acoustics, Speech, and Signal Processing Vol.4 9-12 May 1995 : 2427-2430.
7. Fabrice Moscheni, Frederic Dufaux and Murat Kunt. A New Two-Stage Global/Local Motion Estimation Based on A Background/Foreground Segmentation. International Conference on Acoustics, Speech, and Signal Processing Vol.4 9-12 May 1995 : 2261-2264.
8. Jean-Mare Odobez and Patrick Bouthemy. MRF-Based Motion Segmentation Exploiting A 2D Motion Model Robust Estimation. International Conference on Image Processing Vol.3 23-26 Oct. 1995 : 628-631.

9. Alatan A. Aydin and Levent Onural. Object Based 3-D Motion and Structure Estimation. International Conference on Image Processing Vol.1 23-26 Oct. 1995 : 390-393.
10. Michael M. Chang, A. Murat Tekalp and M.Ibrahim Sezan. Simultaneous Motion Estimation and Segmentation. IEEE Transactions on Image Processing Vol.6 Issue. 9 Sept. 1997 : 1326-1333.
11. Fernando Pereira. Very Low Bitrate Video Coding: The Content-Based Approach. VLVB'97 July 1997.
12. Georgi D. Borshukov, Gozde Bozdagi, Yucel Altunbasak and A. Murat Tekalp. Motion segmentation by multistage affine classification. IEEE Tran. Image Processing Vol.6 Issue.11 , Nov. 1997 : 1591-1594.
13. Thomas Meier and K. N. Ngan. Automatic Segmentation of Moving Objects for Video Object Plane Generation. IEEE Trans. Circuits Syst. Video Technol. vol.8 Sep. 1998 : 525-528.
14. Chee Sun Won. A Block-Based MAP Segmentation for Image Compressions. IEEE Trans. Circuits Syst. Video Technol. vol.8 Sep. 1998 : 592-601.
15. 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. vol.9 Dec. 1999 : 1216-1226.
16. Aljoscha Smolic, Thomas Sikora and Jens – Rainer Ohm. Long-term global motion estimation and its application for sprite coding, content description, and segmentation. IEEE Trans. Circuits Syst. Video Technol. Vol.9 Issue.8 , Dec. 1999 : 1227-1242.
17. Candemir Toklu, A. Murat Tekalp and A. Tanju Erden. Semi-Automatic Video Object Segmentation in the Presence of Occlusion. Trans. Circuits Syst. Video Technol. No.4 Vol.10 June 2000 : 624-629.

18. Salkmann Ji and Hyun Wook Park. Moving Object Segmentation in DCT-Based Compressed Video. Electronics Letters No.21 Vol.36 Oct 2000 : 1769.
19. Hieu T. Nguyen, Marcel Worring and Anuj Dev. Detection of Moving Objects in Video using a Robust Motion Similarity Measure. IEEE Tran. Image Processing Vol. 9 Jan 2000 : 137-141.
20. Mark Everingham and Barry Thomas. Supervised Segmentation and Tracking of Nonrigid Objects using A Mixture of Histograms Model. International Conference on Image Processing Vol.1 7-10 Oct. 2001 : 62-65.
21. Janez Zaletelj and Jurij F. Tasic. Video Object Segmentation Based on Edge Tracking. International Conference on Image Processing Vol.2 7-10 Oct. 2001 : 813-816.
22. Yaakov Tsaig and A. Averbuch. Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach. IEEE Trans. Circuits Syst. Video Technol. vol.12 July 2002 : 597-612. Aug. 1992 : 1324-1327.
23. Giusto D.D., F. Massidda and C. Perra. A fast algorithm for video segmentation and object tracking. 14th International Conference on Digital Signal Processing Vol.2 , 1-3 July 2002 : 697-700.
24. Ebroul Izquierdo, Jianhui Xia and Roland Mech. A generic video analysis and segmentation system. ICASSP '02 Vol.4 13-17 May 2002 : IV-3592 -IV-3595.
25. 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 Vol.1 29 June-1 July 2002 : 566 -569.
26. Jong Bae Kim and Hang Joon Kim. A wavelet-based watershed image segmentation for VOP generation. The 16th International Conference on Pattern Recognition 11-15 Aug. 2002 : 505 -508.

27. Datchakorn 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 July 2001 : 579-583.
28. K. R. Rao and J. J. Hwang, Techniques and Standards for Image, Video, and Audio Coding New jersey: Prentice Hall, 1996.
29. Rafael C. Gonzales and Richard E. Woods, Digital Image Processing Addison-Wesley, 1992.
30. Richard O. Duda, Peter E. Hart, David G.Stork, Pattern Classification 2nd Edition, John Wiley, 2002.
31. S. Khemachai, D. Tancharoen, S. Chomphan and S. Jitapankul, Content Based Object Segmentation using Bayesian Decision Technique. The 3rd International Symposium on Communication and Information Technologies Vol. 2 3-5 Sep. 2003.

ประวัติผู้เขียนวิทยานิพนธ์

นายศักดิ์ เขมชัย เข้ารับการศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2540 สำเร็จการศึกษาระดับปริญญาตรีวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ในปีการศึกษา 2543 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ในสังกัดห้องปฏิบัติการกรรมวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า ที่จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2544



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย