

การแยกส่วนภาพวัตถุสำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของคอนเทนต์
ตามข้อกำหนดของ MPEG-4



นายด้ชกรณ์ ตันเจริญ

สถาบันวิทยบริการ
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย


ปีการศึกษา 2543

ISBN 974-13-0049-2

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

OBJECT SEGMENTATION FOR CONTENT BASED VIDEO CODING
ACCORDING TO MPEG-4

Mr.Datchakorn Tancharoen



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 2000

ISBN 974-13-0049-2

ดัชกรณั ดันเจริญู : การแยกส่วนภาพวัตถุสำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของคอนเทนต์ตามข้อกำหนดของ MPEG-4 (OBJECT SEGMENTATION FOR CONTENT BASED VIDEO CODING ACCORDING TO MPEG-4) อ.ที่ปรึกษา : รศ.ดร.สมชาย จิตะพันธ์กุล, อาจารย์ที่ปรึกษาร่วม : ดร.เสถียร เตรียมล้ำเลิศ, 134 หน้า. ISBN 974-13-0049-2.

วิทยานิพนธ์นี้นำเสนอกรรมวิธีการแยกส่วนภาพวัตถุของสัญญาณวิดีโอตามข้อกำหนดของ MPEG-4 (Moving Picture Expert Group - 4) ซึ่งได้กำหนดมาตรฐานการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ กรรมวิธีการแยกส่วนภาพวัตถุนี้อยู่บนพื้นฐานของการรวมกันของคุณลักษณะหลายอย่าง ประกอบด้วยข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา การแยกส่วนภาพเชิงพื้นที่ประกอบด้วยขั้นตอนการสร้างความเรียบง่ายให้กับภาพ การตรวจรู้ขอบเขตของบริเวณ และการกำหนดบริเวณให้แก่จุดภาพโดยกรรมวิธีสันปันน้ำ ตามลำดับ วิทยานิพนธ์ฉบับนี้ได้นำเอาการปฏิบัติการเชิงสัจฐานแบบเปิด-ปิดมาใช้ในการสร้างความเรียบง่ายให้แก่ภาพ และการประมาณเกรเดียนต์แบบหลายชั้นมาใช้ในการตรวจรู้ขอบเขตของบริเวณ กรรมวิธีที่เสนอนี้สามารถลดจำนวนบริเวณและจำนวนจุดภาพของขอบเขตที่ไม่ระบุบริเวณ เพื่อลดการใช้ทรัพยากรของการคำนวณในขั้นตอนกรรมวิธีสันปันน้ำ

การแยกส่วนเชิงเวลาจะใช้การตรวจรู้การเปลี่ยนแปลงและตัวปฏิบัติการเชิงสัจฐานในการสร้างความเรียบง่ายให้แก่มาส์กของภาพวัตถุที่เปลี่ยนแปลง และใช้กฎการตัดสินใจในการกำหนดบริเวณร่วมกับการใช้หน่วยความจำเก็บผลของภาพวัตถุในเฟรมก่อนหน้าเพื่อกำหนดขอบเขตของภาพวัตถุที่ชัดเจน และช่วยในการติดตามการเคลื่อนที่ของวัตถุอย่างแม่นยำ วิธีนี้จะช่วยให้สามารถตรวจรู้วัตถุที่เคลื่อนที่ได้อย่างต่อเนื่อง

ผลจากการนำกรรมวิธีการแยกส่วนภาพวัตถุที่เสนอนี้มาประมวลผลกับลำดับภาพการประชุมทางวิดีโอ พบว่าโดยคุณภาพเชิงอัตวิสัย สามารถแยกส่วนภาพวัตถุเคลื่อนที่ได้ภาพวัตถุที่มีความหมายและมีคุณภาพที่ดี เมื่อเปรียบเทียบกับผลจาก COST 211 (Cooperation Europeenne dans le recherche Scientifique et Technique 211) และโดยคุณภาพเชิงวัตถุจะมีความผิดพลาดของผลการแยกส่วนเมื่อเทียบกับมาส์กของภาพวัตถุอ้างอิงไม่เกินร้อยละ 15 ของขนาดภาพวัตถุ ขณะที่การตรวจรู้บริเวณของภาพวัตถุเคลื่อนที่ขาดไปไม่เกินร้อยละ 5 ของขนาดภาพวัตถุ ดังนั้นกรรมวิธีนี้จึงให้ภาพวัตถุที่มีความหมายและมีฟังก์ชันบนพื้นฐานของคอนเทนต์ตามข้อกำหนดของ MPEG-4

ภาควิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต.....
สาขาวิชา	วิศวกรรมไฟฟ้า	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา	2543	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4170305121 : MAJOR ELECTRICAL ENGINEERING

KEY WORD : SEGMENTATION / CONTENT BASED / VIDEO CODING / MORPHOLOGICAL / MPEG-4

DATCHAKORN TANCHAROEN : OBJECT SEGMENTATION FOR CONTENT BASED VIDEO CODING ACCORDING TO MPEG-4. THESIS ADVISOR : ASSOC. PROF. SOMCHAI JITAPUNKUL, Dr. Ing., THESIS CO-ADVISOR : SATIEN TRIAMLUMLERD, Ph.D., 134 pp. ISBN 974-13-0049-2.

This thesis proposes a video object segmentation algorithm according to the MPEG-4 which provides the object based video coding standard. This algorithm is based on the combination of multiple features including spatial and temporal informations. The spatial segmentation includes image simplification, boundary detection and region assignment by the watershed algorithm. The morphological opening and closing are used in image simplification process, subsequently, the multiscale gradient approximation is used in boundary detection. The proposed algorithm can reduce the number of regions and uncertain pixels in order to reduce the computational load in the watershed process.

By the way, the temporal segmentation using change detection and morphological operator can create the change mask. In addition, the decision process and the memory of the previous segmentation result are used to determine object boundary and detect object tracking.

Exploiting the proposed algorithm to segment the moving object in the video conferencing sequence, the subjective test showed meaningful and good quality video objects compared with those of COST211 results. Moreover, the objective test showed an error of segmentation result when compared with reference mask, which is less than 15 percent of the object size, while the miss of moving object is less than 5 percent of the object size. Therefore, this algorithm can provide the semantic meaningful video objects and preserve the content based functionalities according to MPEG-4.

Department	Electrical Engineering	Student's signature.....
Field of Study	Electrical Engineering	Advisor's signature.....
Academic year	2000	Co-advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยคำแนะนำและความช่วยเหลืออย่างดียิ่งของ อาจารย์ที่ปรึกษาวิทยานิพนธ์ คือ รศ. ดร.สมชาย จิตะพันธ์กุล พร้อมทั้งแรงกระตุ้นและคำแนะนำของ อาจารย์ที่ปรึกษาร่วม คือ ดร.เสถียร เตรียมล้ำเลิศ นักวิจัยสถาบันวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

ขอขอบคุณ ศาสตราจารย์ ดร.ประสิทธิ์ ประพัฒน์มงคล และอาจารย์สุวิทย์ นาคพิระยุทธ อาจารย์ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย ที่แนะนำแนวทางดำเนินงานและให้ข้อคิดต่าง ๆ สำหรับงานวิจัย

นอกจากนี้ ผู้วิจัยต้องขอขอบคุณ ดร.วุฒิพงศ์ อารีกุล อาจารย์ภาควิชาวิศวกรรมไฟฟ้า มหาวิทยาลัยเกษตรศาสตร์ Dr.K. R. Rao อาจารย์ห้องปฏิบัติการวิจัย Digital Image Processing ของ Texas University at Arlington สหรัฐอเมริกา ที่สุพัฒนา รุ่นพี่ปริญญาเอกของมหาวิทยาลัย ไต่เกี้ยว และพีทฤษดา นักวิจัยของมหาวิทยาลัยแคลิฟอร์เนีย ที่ได้ให้แหล่งข้อมูลสำหรับการทำวิจัย ขอขอบคุณ สวทช. ที่ได้สนับสนุนทุนการศึกษาให้ผู้วิจัยตามโครงการ TGIST (Thai Graduated Institute of Science and Technology) และขอขอบคุณเพื่อน ๆ พี่น้องทุก ๆ คนที่ให้คำแนะนำและ ความช่วยเหลือตลอดระยะเวลาที่ศึกษาอยู่

สุดท้ายผู้วิจัยขอขอบคุณบิดามารดาและครอบครัวที่ให้กำลังใจและการสนับสนุนแก่ผู้วิจัย เสมอมาจนสำเร็จการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ (ต่อ)

บทที่	หน้า
1.4 แนวทางของวิทยานิพนธ์.....	15
1.5 วัตถุประสงค์ของวิทยานิพนธ์.....	15
1.6 ขอบเขตของวิทยานิพนธ์.....	16
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	16
1.8 ขั้นตอนและวิธีการดำเนินการ.....	16
1.9 ภาพรวมของวิทยานิพนธ์.....	17
2. ทฤษฎีที่เกี่ยวข้อง.....	18
2.1 ลักษณะของสัญญาณวิดีโอ (Video Signal).....	18
2.2 การเข้ารหัสสัญญาณวิดีโอ (Video Coding).....	20
2.2.1 การเข้ารหัสภายในเฟรม (Intra-Frame Coding).....	21
2.2.2 การเข้ารหัสระหว่างเฟรม (Inter-Frame Coding).....	22
2.3 มาตรฐาน MPEG-4.....	23
2.4 การแยกส่วนสำหรับการสร้างระนาบวัตถุของสัญญาณวิดีโอ (Video Object Plane : VOP) ตามมาตรฐาน MPEG-4.....	28
2.5 ความรู้พื้นฐานเกี่ยวกับการประมาณการเคลื่อนที่ (Motion Estimation).....	32
2.5.1 วิธีการไหลของแสง (Optical Flow Method).....	32
2.5.2 วิธีเพิลรีเคอร์ซีฟ (Pel-Recursive Method).....	33
2.5.3 กรรรมวิธีการเข้าคู่บล็อก (Block Matching Algorithm: BMA).....	34
2.5.3.1 การประมาณการเคลื่อนที่แบบบล็อก (Block Motion Estimation).....	34
2.5.4 วิธีพาราเมตริกซ์ (Parametric Methods).....	35
2.6 งานวิจัยที่เกี่ยวข้องทางการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ.....	37
2.6.1 การพัฒนางานวิจัยทางการเข้ารหัสรูปทรง (Shape Coding).....	37
2.6.2 การวิเคราะห์การเคลื่อนที่และการแยกส่วนภาพวัตถุ.....	37
2.6.3 งานวิจัยทางการแยกส่วนภาพวัตถุ (Object Segmentation).....	38
2.6.3.1 ระบบแบบผู้ใช้ช่วยกำหนดขอบเขตเบื้องต้น (User Assisted System).....	39

สารบัญ (ต่อ)

บทที่	หน้า
2.6.3.2 ระบบการแยกส่วนภาพวัตถุแบบอัตโนมัติ (Automatic Segmentation).....	39
2.6.4 การแยกส่วนการเคลื่อนที่โดยการเติบโตของบริเวณ (Region Growing).....	39
2.6.5 การแยกส่วนเชิงพื้นที่-เวลา (Spatio-Temporal Segmentation).....	40
2.7 การตรวจจับขอบ (Edge Detection).....	41
2.7.1 การตรวจจับขอบด้วยค่าเกรเดียนต์.....	42
2.7.2 การตรวจจับขอบแบบแคนนี่ (Canny).....	42
3. กรรมวิธีที่น่าสนใจ.....	44
3.1 ความสำคัญของการแยกส่วนภาพวัตถุในการเข้ารหัสสัญญาณวิดีโอ บนพื้นฐานของหน่วยวัตถุ.....	44
3.2 แนวความคิด.....	45
3.3 กรรมวิธีของการแยกส่วนภาพวัตถุสำหรับสัญญาณวิดีโอ.....	46
3.3.1 กระบวนการขั้นต้น (Preprocessing).....	46
3.3.2 การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation).....	47
3.3.2.1 การสร้างภาพให้มีรูปแบบที่เรียบง่าย (Image Simplification).....	47
3.3.2.2 การประมาณเกรเดียนต์ (Gradient Approximation).....	50
3.3.3 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation).....	52
3.3.3.1 การวิเคราะห์การเคลื่อนที่ของลำดับภาพ (Motion Analysis).....	52
3.3.3.2 การตรวจจับการเปลี่ยนแปลง (Change Detection).....	53
3.3.4 การนำทางการเคลื่อนที่ของวัตถุ (Object Tracking).....	55
3.3.5 กฎการตัดสินใจ (Decision Rule).....	55
3.3.6 กระบวนการภายหลัง (Postprocessing).....	56
3.4 การประเมินประสิทธิภาพของกรรมวิธีการแยกส่วนภาพวัตถุ.....	57
3.4.1 ขอบเขตงานของ EUROPEAN COST 211.....	57
3.4.1.1 โครงการทำงานของ COST โดยทั่วไป.....	57

สารบัญ (ต่อ)

บทที่	หน้า
3.4.1.2 ความเป็นมาของ COST 211.....	57
3.4.2 การเปรียบเทียบผลการแยกส่วนภาพวัตถุกับแบบจำลองการวิเคราะห์ของ COST 211.....	58
3.4.2.1 ข้อกำหนดของกรรมวิธีการแยกส่วนภาพวัตถุของ COST 211.....	59
3.4.2.2 เกณฑ์ในการประเมินผลการแยกส่วนภาพ.....	62
4. ผลการวิจัย.....	63
4.1 การแยกส่วนเชิงพื้นที่.....	63
4.1.1 การทำให้ภาพมีความเรียบง่าย (Image Simplification).....	63
4.1.2 การตรวจรู้ขอบด้วยวิธีการประมาณเกรเดียนต์.....	66
4.1.2.1 การประมาณเกรเดียนต์ (Gradient Approximation).....	66
4.1.2.2 การประมาณเกรเดียนต์แบบหลายชั้น (Multiscale Gradient).....	67
4.1.3 การใช้กรรมวิธีสีน้ำ (Watershed Algorithm).....	69
4.2 การแยกส่วนเชิงเวลา (Temporal Segmentation).....	71
4.2.1 การวิเคราะห์การเปลี่ยนแปลงความเข้มแสงระหว่างเฟรม.....	71
4.2.2 การตรวจรู้การเปลี่ยนแปลง.....	75
4.2.2.1 การปรับตัวของค่าจุดเปลี่ยนโดยอัตโนมัติ.....	75
4.2.2.2 การตรวจรู้การเปลี่ยนแปลงของแต่ละจุดภาพ.....	76
4.2.2.3 การวิเคราะห์การเปลี่ยนแปลงเป็นบล็อกขนาด 4x4 จุดภาพ.....	77
4.2.2.4 การวิเคราะห์การเปลี่ยนแปลงเป็นบล็อกขนาด 8x8 จุดภาพ.....	78
4.2.3 การกรองสัญญาณรบกวน และการเติมเต็มภาพวัตถุ.....	78
4.2.4 การใช้หน่วยความจำช่วยในการตรวจรู้.....	79
4.3 การวิเคราะห์ลำดับภาพ.....	81
4.3.1 ลำดับภาพ Akiyo.....	81

สารบัญ (ต่อ)

บทที่	หน้า
4.3.2 ลำดับภาพ Claire.....	82
4.3.3 ลำดับภาพ Erik.....	82
4.3.4 ลำดับภาพ Hall Monitor.....	83
4.3.5 ลำดับภาพ Mother & Daughter.....	83
4.3.6 ลำดับภาพ Suzie.....	84
4.3.7 ลำดับภาพ Table Tennis.....	84
4.3.8 ลำดับภาพ Trevor.....	85
4.4 การแยกส่วนภาพวัตถุโดยการวิเคราะห์เชิงพื้นที่-เวลา.....	85
4.5 การวัดสมรรถนะของกรรมวิธีการแยกส่วนภาพวัตถุ.....	90
4.5.1 การวัดผลการแยกส่วนภาพเชิงอัตวิสัย.....	90
4.5.3 การวัดผลการแยกส่วนภาพเชิงวัตถุ.....	95
4.5.2 ความต่อเนื่องของผลภาพวัตถุ.....	101
5. สรุปและข้อเสนอแนะ.....	104
5.1 สรุป.....	104
5.2 ข้อเสนอแนะ.....	105
รายการอ้างอิง.....	106
ภาคผนวก.....	111
ภาคผนวก ก.....	112
ภาคผนวก ข.....	120
ประวัติผู้เขียน.....	134

สารบัญตาราง

	หน้า
ตารางที่ 1.1 มาตรฐานการเข้ารหัสสัญญาณวิดีโอของ ITU-T.....	4
ตารางที่ 1.2 มาตรฐานการเข้ารหัสสัญญาณวิดีโอของ MPEG.....	5
ตารางที่ 2.1 แสดงรูปแบบของสัญญาณวิดีโอแบบ SIF และ CIF.....	20
ตารางที่ 3.1 แสดงชื่อและลักษณะสำคัญของลำดับภาพที่ใช้ทดสอบกรรมวิธีการแยกส่วนภาพ.....	59
ตารางที่ 4.1 แสดงค่าเฉลี่ยของค่าเบี่ยงเบนมาตรฐานของภาพที่ถูกสร้างให้มีความเรียบง่าย ของลำดับภาพ Akiyo ใน 10 เฟรมด้วยวิธีต่าง ๆ.....	65
ตารางที่ 4.2 แสดงจำนวนบริเวณเฉลี่ยจากการตรวจจู่ขอบด้วยการประมาณเกรเดียนต์.....	68
ตารางที่ 4.3 แสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณเฉลี่ยใน 10 เฟรม ระหว่างการประมาณเกรเดียนต์ แบบขั้นเดียว และแบบหลายขั้น.....	68
ตารางที่ 4.4 ผลของการแยกส่วนภาพเชิงพื้นที่ของภาพ Claire โดยกรรมวิธีสันปันน้ำ.....	71
ตารางที่ 4.5 แสดงค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211เทียบกับ มาตรฐานอ้างอิงของลำดับภาพ Erik ใน 50 เฟรม.....	96
ตารางที่ 4.6 แสดงร้อยละของค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST 211 เทียบกับมาตรฐานอ้างอิงของลำดับภาพ Erik ใน 50 เฟรม.....	98
ตารางที่ 4.7 แสดงค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211เทียบกับ มาตรฐานอ้างอิงของลำดับภาพ Hall Monitor ใน 50 เฟรม.....	99
ตารางที่ 4.8 แสดงร้อยละของค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211 เทียบกับมาตรฐานอ้างอิงของลำดับภาพ Hall Monitor ใน 50 เฟรม.....	100

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

	หน้า
รูปที่ 1.1	การพัฒนาของมาตรฐานการเข้ารหัสสัญญาณวิดีโอ..... 3
รูปที่ 1.2	คุณภาพและอัตราบิตตามข้อกำหนดของมาตรฐานการเข้ารหัสสัญญาณวิดีโอ..... 4
รูปที่ 2.1	โครงสร้างของสัญญาณภาพในรูปแบบอัตราสุ่ม 4:2:0..... 19
รูปที่ 2.2	ตัวอย่างการแปลงโดยการสุ่มลง (down sampling)..... 19
รูปที่ 2.3	บล็อกไดอะแกรมของการเข้ารหัสสัญญาณวิดีโอแบบที่มีการบีบอัด..... 21
รูปที่ 2.4	ภาพแสดงโครงสร้างพื้นฐานตัวเข้ารหัสสัญญาณวิดีโอของ MPEG-4..... 24
รูปที่ 2.5	ภาพแสดงโครงสร้างพื้นฐานตัวถอดรหัสสัญญาณวิดีโอของ MPEG-4..... 24
รูปที่ 2.6	รูปแบบในการประมาณการเคลื่อนที่แบบการจับคู่บล็อก..... 25
รูปที่ 2.7	การเข้ารหัสรูปทรงโดย Content-based Arithmetic Encoding (CAE)..... 26
รูปที่ 2.8	ส่วนต่างๆ ที่เข้ารหัสบนพื้นฐานของหน่วยวัตถุ..... 27
รูปที่ 2.9	กรรมวิธี Shape Adaptive DCT (SA-DCT)..... 27
รูปที่ 2.10	รูปแบบการเข้ารหัสโดยกรรมวิธี Zerotree..... 28
รูปที่ 2.11	โครงสร้างของการรวมข้อมูลสำหรับการแยกส่วนภาพวัตถุ..... 29
รูปที่ 2.12	โครงสร้างการแยกส่วนภาพวัตถุโดยกรรมวิธีการเติบโตของบริเวณ (Region Growing)..... 40
รูปที่ 2.13	แสดงถึงระดับความเข้มของขอบ และอนุพันธ์อันดับ 1 ของระดับความเข้ม..... 41
รูปที่ 2.14	บริเวณภาพขนาด 3x3 จุดภาพ และมาสก์ขนาด 3x3 สำหรับใช้หาค่า Gx และ Gy..... 42
รูปที่ 3.1	โครงสร้างของการแยกส่วนภาพวัตถุและการเข้ารหัส MPEG-4 สำหรับการเข้ารหัส สัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ (Object Based Video Coding)..... 44
รูปที่ 3.2	โครงสร้างของกรรมวิธีการแยกส่วนภาพวัตถุโดยการรวมกันของหลายคุณลักษณะ..... 46
รูปที่ 3.3	โครงสร้างแบบจำลองการแยกส่วนภาพวัตถุของ COST 211 (COST AM)..... 58
รูปที่ 3.4	รูปแบบของแฟ้มข้อมูลลำดับภาพ (file format of image sequence)..... 60
รูปที่ 3.5	ลำดับภาพทดสอบรูปแบบ CIF..... 60
รูปที่ 3.6	ลำดับภาพทดสอบรูปแบบ QCIF..... 61
รูปที่ 3.7	การแปลงขนาดของภาพ..... 62

สารบัญภาพ (ต่อ)

หน้า

รูปที่ 4.1	การปฏิบัติการเชิงเรขาคณิตของภาพ Akiyo.....	64
	(ก) การปฏิบัติการแบบเปิด	
	(ข) การปฏิบัติการแบบปิด	
	(ค) การปฏิบัติการแบบเปิด-ปิด	
	(ง) การปฏิบัติการแบบปิด-เปิด	
	(จ) การปฏิบัติการแบบเปิด-ปิดด้วยการสร้างใหม่บางส่วน	
รูปที่ 4.2	แสดงค่าเบี่ยงเบนมาตรฐานของภาพที่ถูกสร้างให้มีความเรียบง่ายในลำดับภาพ Akiyo เฟรมที่ 1 ถึง 10.....	65
รูปที่ 4.3	ภาพแสดงขั้นตอนของการประมาณเกรเดียนต์.....	66
	(ก) การปฏิบัติการไคเลชัน	
	(ข) การปฏิบัติการอีโรชัน	
	(ค) ค่าการประมาณเกรเดียนต์	
	(ง) ค่าเกรเดียนต์ที่มีค่าสูงกว่าค่าจุดเปลี่ยน	
รูปที่ 4.4	แสดงการประมาณเกรเดียนต์แบบหลายชั้นของภาพ Claire.....	67
รูปที่ 4.5	แสดงจำนวนบริเวณเฉลี่ยจากการตรวจรู้ขอบในเฟรมที่ 1 ถึง 10 ของการประมาณเกรเดียนต์แบบชั้นเดียวและแบบหลายชั้น.....	67
รูปที่ 4.6	แสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณในเฟรมที่ 1 ถึง 10 ระหว่างการประมาณเกรเดียนต์แบบชั้นเดียว และแบบหลายชั้น.....	68
รูปที่ 4.7	ผลการแยกส่วนภาพเชิงพื้นที่โดยใช้กรรมวิธีสันปันน้ำของลำดับภาพ Claire (ก) เฟรมที่ 1 (ข) เฟรมที่ 10.....	69
รูปที่ 4.8	กราฟแสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณที่เหลือ หลังจากผ่านกรรมวิธีสันปันน้ำแล้ว.....	70
รูปที่ 4.9	กราฟแสดงจำนวนบริเวณทั้งหมดที่ได้จากการระบุบริเวณให้กับจุดภาพ.....	70
รูปที่ 4.10	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Akiyo ใน 50 เฟรมแรก.....	72
รูปที่ 4.11	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Claire ใน 50 เฟรมแรก.....	72
รูปที่ 4.12	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Erik ใน 50 เฟรมแรก.....	73
รูปที่ 4.13	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Hall Monitor ใน 50 เฟรมแรก.....	73
รูปที่ 4.14	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Mother & Daughter ใน 50 เฟรมแรก.....	74
รูปที่ 4.15	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Table Tennis ใน 50 เฟรมแรก.....	74

สารบัญภาพ (ต่อ)

	หน้า
รูปที่ 4.16	การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Trevor ใน 50 เฟรมแรก..... 75
รูปที่ 4.17	แสดงค่าจุดเปลี่ยนที่มีการปรับตัวโดยอัตโนมัติ สำหรับลำดับภาพ Akiyo..... 76
รูปที่ 4.18	แสดงค่าจุดเปลี่ยนที่มีการปรับตัวโดยอัตโนมัติ สำหรับลำดับภาพ Mother & Daughter..... 76
รูปที่ 4.19	แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพในเฟรมที่ 50 77 (ก) Akiyo (ข) Mother & Daughter (ค) Hall Monitor (ง) Erik
รูปที่ 4.20	แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Table Tennis ในเฟรมที่ 5 เมื่อค่าจุดเปลี่ยนเท่ากับ 1..... 77
รูปที่ 4.21	รูปแสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Suzie (ก) เฟรมที่ 5 (ข) เฟรมที่ 50..... 78
รูปที่ 4.22	การปฏิบัติการแบบปิด-เปิด สำหรับการตรวจรู้การเปลี่ยนแปลงของภาพ Akiyo เฟรมที่ 50 โดยใช้องค์ประกอบโครงสร้างขนาด 5x5 จุดภาพ 78
รูปที่ 4.23	การปฏิบัติการแบบปิด-เปิด และการกำจัดบริเวณขนาดเล็ก สำหรับการตรวจรู้การเปลี่ยนแปลง ของภาพ Mother&Daughter ในเฟรมที่ 50..... 79
รูปที่ 4.24	การกำจัดบริเวณที่มีขนาดเล็กในการตรวจรู้การเปลี่ยนแปลงของภาพ Claire เฟรมที่ 50 เมื่อ ใช้การตรวจรู้แบบบล็อกขนาด 4x4 จุดภาพ และค่าจุดเปลี่ยน เท่ากับ 0.2..... 79
รูปที่ 4.25	แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Mother & Daughter ที่ไม่ได้ใช้หน่วยความจำ..... 80
รูปที่ 4.26	แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Mother & Daughter ที่ใช้หน่วยความจำ..... 80
รูปที่ 4.27	ลำดับภาพ Akiyo (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 81
รูปที่ 4.28	ลำดับภาพ Claire (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 82
รูปที่ 4.29	ลำดับภาพ Erik (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 82
รูปที่ 4.30	ลำดับภาพ Hall Monitor (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 83
รูปที่ 4.31	ลำดับภาพ Mother & Daughter (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 83
รูปที่ 4.32	ลำดับภาพ Suzie (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 84
รูปที่ 4.33	ลำดับภาพ Table Tennis (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 84
รูปที่ 4.34	ลำดับภาพ Trevor (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ..... 85
รูปที่ 4.35	ผลการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Mother & Daughter..... 85

สารบัญภาพ (ต่อ)

รูปที่ 4.36	รูปภาพแสดงจำนวนบริเวณที่ได้จากการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Mother & Daughter ในเฟรมที่ 1 ถึง 100.....	86
รูปที่ 4.37	ผลการแยกส่วนภาพเชิงเวลาของลำดับภาพ Mother & Daughter.....	86
รูปที่ 4.38	รูปแสดงมาสก์ของภาพวัตถุของลำดับภาพ Mother & Daughter เฟรมที่ 33 35 37 ... 49.....	87
รูปที่ 4.39	รูปแสดงภาพวัตถุของลำดับภาพ Mother & Daughter จากการแยกส่วนภาพวัตถุด้วยข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลาของเฟรมที่ 33 35 37 ... 49.....	87
รูปที่ 4.40	ภาพวัตถุที่ได้ของลำดับภาพ Erik เฟรมที่ 41 ถึงเฟรมที่ 49.....	88
รูปที่ 4.41	ภาพวัตถุที่ได้ของลำดับภาพ Akiyo เฟรมที่ 9 13 17 ... 41.....	89
รูปที่ 4.42	ภาพวัตถุที่ได้ของลำดับภาพ Claire เฟรมที่ 11 15 19 ... 43.....	89
รูปที่ 4.43	ภาพวัตถุของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10.....	90
รูปที่ 4.44	ภาพวัตถุของลำดับภาพ Erik เฟรมที่ 2 ถึง 10 จากผลของ COST211.....	91
รูปที่ 4.45	ภาพวัตถุของลำดับภาพ Erik เฟรมที่ 2 ถึง 10 จากกรรมวิธีที่เสนอ.....	91
รูปที่ 4.46	ภาพวัตถุของลำดับภาพ Mother & Daughter เฟรมที่ 41 ถึง 49 จากผลของ COST211.....	92
รูปที่ 4.47	ภาพวัตถุของลำดับภาพ Mother & Daughter เฟรมที่ 41 ถึง 49 จากกรรมวิธีที่นำเสนอ.....	92
รูปที่ 4.48	ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10 จากผลของ COST211.....	93
รูปที่ 4.49	ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10 จากกรรมวิธีที่เสนอ.....	93
รูปที่ 4.50	ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 32 ถึง 40 จากผลของ COST211.....	94
รูปที่ 4.51	ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 32 ถึง 40 จากกรรมวิธีที่เสนอ.....	94
รูปที่ 4.52	การเปรียบเทียบผลการแยกส่วนภาพวัตถุระหว่างผลจากกรรมวิธีที่เสนอ และมาสก์อ้างอิงของลำดับภาพ Erik (ก) มาสก์อ้างอิง (ข) ผลที่ได้.....	95
รูปที่ 4.53	รูปภาพแสดงจำนวนจุดภาพที่แตกต่างกันระหว่างผลที่ได้และมาสก์อ้างอิงของลำดับภาพ Erik ใน 50 เฟรม.....	95
รูปที่ 4.54	รูปภาพแสดงจำนวนจุดภาพที่แตกต่างกันระหว่างผลของ COST211 และมาสก์อ้างอิงของลำดับภาพ Erik ใน 50 เฟรม.....	96
รูปที่ 4.55	ร้อยละของจำนวนจุดภาพที่แตกต่างกันระหว่างผลที่ได้และมาสก์อ้างอิงเมื่อเทียบกับขนาดของมาสก์อ้างอิงของลำดับ Erik.....	97

สารบัญภาพ (ต่อ)

หน้า

รูปที่ 4.56	ร้อยละของจำนวนจุดภาพที่แตกต่างกันระหว่างผลของ COST211 และมาสก์อ้างอิงเมื่อเทียบกับขนาดของมาสก์อ้างอิงของลำดับ Erik.....	97
รูปที่ 4.57	การเปรียบเทียบผลการแยกส่วนภาพวัตถุระหว่างผลจากกรรมวิธีของผลที่ได้และมาสก์อ้างอิงของลำดับภาพ Hall Monitor (ก) มาสก์อ้างอิง (ข) ผลที่ได้.....	98
รูปที่ 4.58	รูปภาพแสดงจำนวนจุดภาพที่แยกส่วนขาดไประหว่างผลที่ได้และผลของ COST211 ของลำดับภาพ Hall Monitor ใน 50 เฟรม.....	99
รูปที่ 4.59	แสดงร้อยละของจำนวนจุดภาพที่แยกส่วนขาดไปของผลที่ได้และผลของ COST211 เมื่อเทียบกับมาสก์อ้างอิงของลำดับภาพ Hall Monitor.....	100
รูปที่ 4.60	รูปภาพแสดงความต่อเนื่องของภาพวัตถุของลำดับภาพ Akiyo Claire Erik และ Mother & Daughter เฟรมที่ 1 ถึง 50.....	101
รูปที่ 4.61	รูปภาพแสดงจำนวนจุดภาพของมาสก์ของลำดับภาพ Mother&Daughter เฟรมที่ 1 ถึง 100.....	102
รูปที่ 4.62	ภาพที่ได้จากการแยกส่วนเชิงพื้นที่และมาสก์ของภาพวัตถุในเฟรมที่ 16 ของลำดับภาพ Mother & Daughter (ก) ภาพบริเวณจากการแยกส่วนเชิงพื้นที่ (ข) มาสก์ของภาพวัตถุ.....	102
รูปที่ ก.1	ภาพต้นแบบของลำดับภาพ Claire เฟรมที่ 2 ถึง 10.....	112
รูปที่ ก.2	การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Claire เฟรมที่ 2 ถึง 10.....	112
รูปที่ ก.3	บริเวณที่ได้จากการแยกส่วนเชิงพื้นที่ของลำดับภาพ Claire เฟรมที่ 2 ถึง 10.....	113
รูปที่ ก.4	ภาพวัตถุของลำดับภาพ Claire เฟรมที่ 2 ถึง 10.....	113
รูปที่ ก.5	ภาพต้นแบบของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10.....	114
รูปที่ ก.6	การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10.....	114
รูปที่ ก.7	บริเวณจากการแยกส่วนเชิงพื้นที่ของลำดับภาพ Table Tennis เฟรมที่ 17 ถึง 25.....	115
รูปที่ ก.8	ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 17 ถึง 25.....	115
รูปที่ ก.9	ภาพต้นแบบของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10.....	116
รูปที่ ก.10	บริเวณจากการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10.....	116
รูปที่ ก.11	การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10.....	117
รูปที่ ก.12	ภาพวัตถุของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10.....	117
รูปที่ ก.13	ลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42.....	118
รูปที่ ก.14	การแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42.....	118

สารบัญภาพ (ต่อ)

หน้า

รูปที่ ก.15	การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42.....	119
รูปที่ ก.16	ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42.....	119



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บัญชีคำศัพท์

การเข้ารหัสช่องสัญญาณ	Channel Coding
การเข้ารหัสแหล่งกำเนิดสัญญาณ	Source Coding
การเข้ารหัสรูปทรง	Shape Coding
การเข้ารหัสการเคลื่อนที่	Motion Coding
การเข้ารหัสโดยการแปลง	Transform Coding
การเข้ารหัสพื้นผิว	Texture Coding
การเข้ารหัสเลขคณิต	Arithmetic Coding
การเข้ารหัสสัญญาณวิดีโอ	Video Coding
การเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของคอนเทนต์	Content based video coding
การเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ	Object based video coding
การเข้ารหัสเอนโทรปี	Entropy Coding
การควอนไทซ์	Quantization
การชดเชยการเคลื่อนที่	Motion Compensation
การชดเชยการเคลื่อนที่โดยรวม	global motion compensation
การบีบอัดข้อมูล	Data Compression
การประมาณการเคลื่อนที่	Motion Estimation
การประมาณเกรเดียนต์	Gradient Approximation
การประชุมทางวิดีโอ	Video Conferencing
การแปลงโคไซน์แบบไม่ต่อเนื่อง	Discrete Cosine Transform ย่อว่า DCT
การแปลงเวฟเลต	Wavelet Transform
การแยกส่วนมากเกินไป	oversegmentation
การแยกส่วนภาพวัตถุ	Object Segmentation
การแยกส่วนภายในเฟรม	intra-frame segmentation
การแยกส่วนระหว่างเฟรม	inter-frame segmentation
การแยกส่วนเชิงสัญญาณ	morphological segmentation
การแยกส่วนเชิงพื้นที่	spatial segmentation
การแยกส่วนเชิงเวลา	temporal segmentation
การแยกส่วนเชิงพื้นที่-เวลา	spatio-temporal segmentation

การติดตามการเคลื่อนที่	motion tracking
การตรวจรู้การเปลี่ยนแปลง	Change Detection
การเติบโตของบริเวณ	Region Growing
การรวมตัวของบริเวณ	region merging
การลดความซ้ำซ้อน	redundancy reduction
การสื่อสารมัลติมีเดีย	multimedia communication
การไหลทางแสง	optical flow
กระบวนการขั้นต้น	pre processing
กระบวนการภายหลัง	post processing
กรรมวิธี	algorithm
กรรมวิธีสันปันน้ำ	watershed algorithm
กล้องจับภาพ	camera
กลุ่มของเฟรม	Group of Frames ย่อว่า GOF
ค่าความผิดพลาดสัมบูรณ์เฉลี่ย	Mean Absolute Error
	ย่อว่า MAE
ค่าแตกต่างสัมบูรณ์เฉลี่ย	Mean Absolute Difference
ค่าเบี่ยงเบนมาตรฐาน	Standard Deviation ย่อว่า SD
ค่าจุดเปลี่ยน	threshold
ความกว้างแถบ	bandwidth
ความซ้ำซ้อนทางพื้นที่	spatial redundancy
ความซ้ำซ้อนทางเวลา	temporal redundancy
จุดภาพ	pixel
ดิจิทัล	digital
วงจรรองเชิงเส้นฐาน	morphological filter
ตัวปฏิบัติการแบบเปิด	opening operator
ตัวปฏิบัติการแบบปิด	closing operator
ทรัพยากรในการคำนวณ	computation cost
บล็อกที่เปลี่ยนแปลง	Change Detection Block
	ย่อว่า CDB
ปริทัศน์วรรณกรรม	literature review
โปรเจกชัน	projection

พารามิเตอร์	parameter
พื้นหลังที่ไม่ถูกครอบคลุม	uncovered background
ฟังก์ชันบนพื้นฐานของคอนเทนต์	content-based functionality
มาสก์ของการตรวจรู้การเปลี่ยนแปลง	Change Detection Mask ย่อว่า CDM
มาสก์ของภาพวัตถุ	Object Mask ย่อว่า OM
รหัสลูกโซ่	chain code
ระนาบวัตถุของสัญญาณวิดีโอ	Video Object Plane ย่อว่า VOP
ระบบการสื่อสารเคลื่อนที่	mobile communication system
รอยหยักของบล็อกเทียม	blocking artifact
ลำดับภาพ	image sequence
วัตถุเคลื่อนที่แข็งเกร็ง	rigid moving object
กรรมวิธีการจับคู่บล็อก	Block Matching Algorithm ย่อว่า BMA
สัญญาณวิดีโอ	Video Signal
สัญญาณวิดีโอแบบธรรมชาติ	Natural Video
สนามการเคลื่อนที่	motion field
อัตราบิต	bit rate
อัตราส่วนสัญญาณต่อสัญญาณรบกวน	Signal to Noise Ratio ย่อว่า SNR
องค์ประกอบโครงสร้าง	Structure Element ย่อว่า SE

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ปัจจุบันการสื่อสารโทรคมนาคมได้เข้ามามีบทบาทในสังคมมากขึ้น และมีแนวโน้มที่จะทวีความสำคัญมากขึ้นเรื่อยๆ ดังเห็นได้จากการขยายตัวของระบบการสื่อสารในประเทศ และความพยายามในการสร้างระบบสื่อสารที่มีประสิทธิภาพมากขึ้น

ตั้งแต่อดีต การสื่อสารโทรคมนาคมได้พัฒนาขึ้นจากการสื่อสารผ่านทาง โทรเลข โทรศัพท์ โทรสาร จนกระทั่งมาถึงการสื่อสารโดยใช้ข้อมูลภาพ ซึ่งนำมาใช้กันมากในระบบมัลติมีเดีย (multimedia) เมื่อเราพิจารณาการพัฒนาของระบบสื่อสารโทรคมนาคมให้ดีแล้ว จะเห็นได้ว่าระบบที่เกิดขึ้นในช่วงระยะหลังจะมีปริมาณข่าวสารมากกว่าในระบบเดิม โดยทั่วไปแล้วการส่งข่าวสารที่มีปริมาณข่าวสารสูงจำเป็นต้องส่งผ่านช่องสัญญาณที่มีอัตราเร็วสูง ซึ่งการส่งข่าวสารที่ต้องการช่องสัญญาณที่มีอัตราความเร็วสูงกว่าเดิมก็เท่ากับเป็นการเพิ่มต้นทุนของการส่ง ซึ่งจะทำให้ราคาของค่าบริการสูงขึ้นด้วย

ดังนั้นความพยายามที่จะส่งสัญญาณภาพและสัญญาณวิดีโอเพิ่มเข้าไปโดยใช้ระบบการสื่อสารที่มีอยู่ จึงได้เกิดขึ้น แต่ยังไม่สามารถนำมาให้บริการได้จริง เพราะต้องใช้ต้นทุนในการส่งสูงมาก จึงได้มีการพัฒนาเทคโนโลยีของการเข้ารหัสข่าวสารแบบต่าง ๆ เพิ่มขึ้นมากมาย ข่าวสารอย่างหนึ่งที่มีข้อมูลมากที่สุดก็คือ สัญญาณวิดีโอ ดังนั้นการเข้ารหัสสัญญาณวิดีโอจะทำให้สามารถลดอัตราบิตในการส่งข่าวสารลงได้มาก และยังช่วยลดการใช้แถบความถี่ในช่องสัญญาณให้น้อยลง ทำให้สามารถใช้ช่องสัญญาณส่งข่าวสารต่าง ๆ ได้หลายชนิดพร้อมกัน และใช้ช่องสัญญาณได้อย่างคุ้มค่าที่สุด และถ้าสามารถพัฒนากรรมวิธีในการลดปริมาณของข้อมูลสัญญาณวิดีโอได้มากขึ้นเท่าไร ก็จะสามารถใช้ความจุของช่องสัญญาณที่มีอยู่อย่างจำกัดได้อย่างคุ้มค่ามากขึ้นเท่านั้น

เนื่องจากการเข้ารหัสสัญญาณวิดีโอสำหรับใช้ส่งผ่านระบบการสื่อสารเคลื่อนที่ จำเป็นที่จะต้องเข้ารหัสในอัตราบิตต่ำมาก จึงจำเป็นต้องใช้กรรมวิธีการเข้ารหัสสัญญาณที่เหมาะสมเพื่อให้ได้คุณภาพและอัตราบิตตามต้องการ จากแนวโน้มในการวิจัยการเข้ารหัสสัญญาณวิดีโออัตราบิตต่ำ การเข้ารหัสบนพื้นฐานของหน่วยวัตถุ (Object Based Coding) สำหรับสัญญาณ วิดีโอจะสามารถลดความซ้ำซ้อนทางเวลา (Temporal Redundancy) ของข้อมูลได้มากและยังเหมาะสมที่จะใช้กับระบบการสื่อสารมัลติมีเดียอีกด้วย

เนื่องจากมาตรฐาน MPEG-4 ซึ่งเป็นมาตรฐานการเข้ารหัสสัญญาณวิดีโอที่อัตราบิตต่ำ (Low bit rate video coding) และเป็นการเข้ารหัสบนพื้นฐานของหน่วยวัตถุ ที่มีข้อได้เปรียบทั้งในการด้านอัตราการบีบอัดข้อมูลและฟังก์ชันทางด้านมัลติมีเดีย จึงเป็นมาตรฐานการเข้ารหัสข้อมูลที่มีแนว

โน้มจะเป็นพื้นฐานสำหรับการส่งสัญญาณวิดีโอและข้อมูลต่าง ๆ ผ่านระบบสื่อสารเคลื่อนที่ได้อย่างมีประสิทธิภาพสำหรับระบบการสื่อสารมัลติมีเดียในอนาคต

1.1 ความรู้เบื้องต้นเกี่ยวกับการเข้ารหัสสัญญาณวิดีโอ

การเข้ารหัสสัญญาณวิดีโอนั้น มีกรรมวิธีที่ได้มีการนำเสนอขึ้นมา 2 แนวทางด้วยกัน คือ

1.1.1 การแทนรูปคลื่น (waveform representation)

เทคนิคที่ใช้สำหรับวิธีนี้ จะใช้วิธีการแบ่งภาพเป็นบล็อก และใช้การเข้ารหัสแบบทำนายล่วงหน้า (predictive coding) การชดเชยการเคลื่อนที่ (motion compensation) เพื่อลดความซ้ำซ้อนทางเวลา (temporal redundancy) และการเข้ารหัสโดยการแปลง (transform coding) เพื่อลดความซ้ำซ้อนทางพื้นที่ (spatial redundancy) มาตรฐานที่ใช้วิธีการเข้ารหัสแบบนี้ได้แก่ MPEG-1 ในอัตราบิต 1.6 Mbit/s, MPEG-2 ในอัตราบิต 10 Mbit/s และ H.263 ในอัตราบิตต่ำกว่า 64 kbit/s

ข้อดีของวิธีนี้ คือ มีประสิทธิภาพการบีบอัดข้อมูลในอัตราบิตช่วงกว้างขึ้นอยู่กับการใช้งาน ข้อเสียที่เห็นได้ชัด คือ จะเกิดรอยหยักของบล็อกเทียม (blocking artifacts) ที่การเข้ารหัสอัตราบิตต่ำ และไม่มีฟังก์ชันทางด้านการแสดงและการจัดการกับภาพวัตถุ

1.1.2 การแยกเป็นวัตถุมูลฐาน (decomposition into visual primitives)

เทคนิคนี้สามารถแก้ปัญหาของวิธีแบบแบ่งเป็นบล็อก โดยจะแยกภาพออกเป็นส่วนย่อยๆ ที่มีความหมาย มาตรฐานที่กำหนดให้มีการจัดการบนพื้นฐานของหน่วยวัตถุนี้ ได้แก่ MPEG-4 ซึ่งกรรมวิธีในการแยกส่วนภาพวัตถุอาจจำแนกเป็น 2 กลุ่มด้วยกัน ดังนี้

1.1.2.1 วิธีการวิเคราะห์และสังเคราะห์บนพื้นฐานของหน่วยวัตถุ (Object Based Analysis-Synthesis Approach)

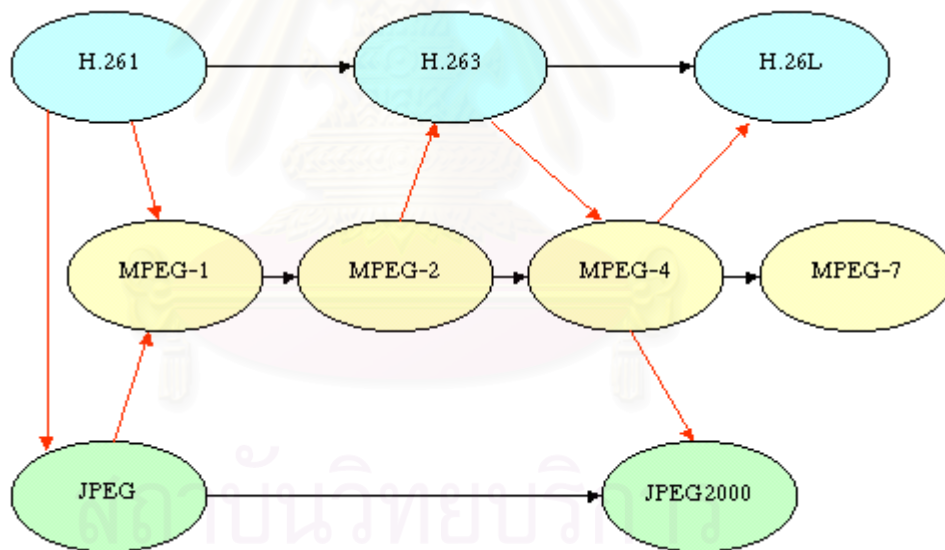
จะแบ่งเป็นบริเวณที่มีการเคลื่อนที่ต่อเนื่องกัน โดยที่แต่ละบริเวณจะถูกแทนด้วยข้อมูลของรูปทรง (shape) การเคลื่อนที่ (motion) และพื้นผิว (texture) สำหรับเฟรมแต่ละเฟรมจะถูกสังเคราะห์ขึ้นจาก 2 ส่วน คือ พารามิเตอร์ที่แสดงบริเวณของเฟรมก่อนหน้า และบริเวณที่มีความคลาดเคลื่อนเกิดขึ้น (motion failure) ซึ่งจะแสดงได้ด้วยข้อมูลรูปทรงและสีของบริเวณที่มีการเปลี่ยนแปลงนี้ ในการส่งรหัสข้อมูลจะต้องส่ง 2 ส่วนด้วยกัน คือ พารามิเตอร์ของการเคลื่อนที่และรูปทรงของทั้งบริเวณเดิมและบริเวณใหม่ และข้อมูลของสีพื้นผิวที่ต้องส่งไปเพิ่มเฉพาะบริเวณที่เกิดขึ้นใหม่

1.1.2.2 วิธีการทางคณิตศาสตร์ (Mathematical Morphology)

ในเฟรมแรกจะแบ่งแยกเป็นบริเวณที่มีความเชื่อมต่อเนื่องกันโดยใช้กรรมวิธีสันปันน้ำ (watershed) ส่วนในเฟรมถัดไปจะใช้การเข้ารหัสแบบคาดการณ์ล่วงหน้า (predictive coding) โดยใช้การโปรเจกการเคลื่อนที่ไปข้างหน้า (forward motion projection) เพื่อนำทาง (tracking) การเคลื่อนที่ของแต่ละบริเวณ โดยจะส่งข้อมูลของการเคลื่อนที่ที่ พื้นผิว และรูปทรงของบริเวณนั้น ส่วนการลดจำนวนของการส่งข้อมูลรูปทรงและพื้นผิวจะใช้ predictive coding ซึ่งสามารถลดข้อมูลส่วนที่ซ้ำซ้อนได้มากเช่นกัน

1.2 มาตรฐานในการเข้ารหัสสัญญาณวิดีโอ

หน่วยงานที่ทำหน้าที่จัดทำมาตรฐานการเข้ารหัสสัญญาณวิดีโอมีอยู่ 2 หน่วยงานหลัก [1] ได้แก่ ITU-T (International Telecommunication Union) และ MPEG (Motion Picture Expert Group) และได้กำหนดมาตรฐานเกี่ยวกับการเข้ารหัสสัญญาณวิดีโอดังนี้



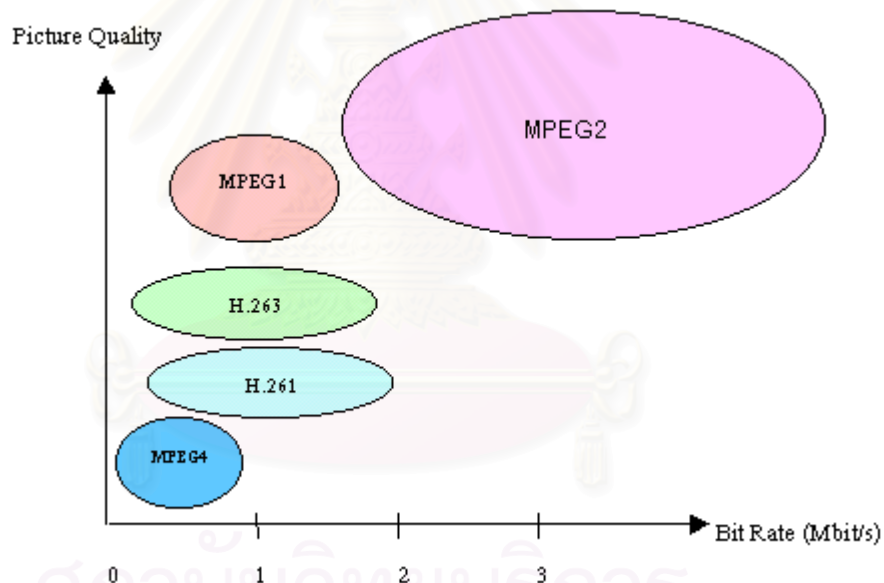
รูปที่ 1.1 การพัฒนาของมาตรฐานการเข้ารหัสสัญญาณวิดีโอ

1.2.1 มาตรฐานของ ITU-T

ITU-T เป็นมาตรฐานที่เกี่ยวข้องกับการสื่อสารโทรคมนาคมของนานาชาติ มาตรฐานการเข้ารหัสสัญญาณวิดีโอได้แก่ มาตรฐาน ITU-T ตระกูล H (H-series) จัดทำขึ้นโดยมีจุดประสงค์มุ่งเน้นในการใช้งานกับเครื่องโทรศัพท์ภาพ (video phone) หรือ การประชุมทางวิดีโอ (video conferencing) มาตรฐานต่าง ๆ ในตระกูล H แสดงไว้ดังตารางที่ 1.1

ตารางที่ 1.1 มาตรฐานการเข้ารหัสสัญญาณวิดีโอของ ITU-T

หมายเลขมาตรฐาน	ปีที่กำหนดใช้	วัตถุประสงค์
H.261	ค.ศ 1990	เพื่อการจัดการระบบการประชุมทางไกลผ่านระบบ ISDN (ISDN Video Conferencing)
H.262 (ร่วมกับ ISO/IEC)	ค.ศ 1995	เพื่อการส่งกระจายสัญญาณ boardcast และพัฒนาการเก็บข้อมูล มาตรฐานเหมือน MPEG 2
H.263	ค.ศ 1996	เพื่อจัดการและเข้ารหัสภาพเคลื่อนไหวผ่านระบบโทรศัพท์เคลื่อนที่
H.263+	ค.ศ 1998	พัฒนามาจาก H.263 โดยมีฟังก์ชันการทำงานมากขึ้น
H.263++	ค.ศ 2000	พัฒนาขึ้นจาก H.263+ และเพิ่มฟังก์ชันการทำงาน
H.26L(H.263L)	ค.ศ 2002	ใช้กับงานสำหรับอินเทอร์เน็ตและการเก็บข้อมูลและให้บริการสัญญาณวิดีโอ



รูปที่ 1.2 คุณภาพและอัตราบิตตามข้อกำหนดของมาตรฐานการเข้ารหัสสัญญาณวิดีโอ

1.2.2 มาตรฐานของ MPEG

มาตรฐาน MPEG เป็นมาตรฐานซึ่งตั้งขึ้นโดยคณะกรรมการ MPEG ปัจจุบันเป็น มาตรฐานของ ISO (International Standard Organization) เริ่มทำงานตั้งแต่ปีคริสต์ศักราช 1988 โดยมีเป้าหมายเพื่อที่จะให้สามารถส่งและบันทึกสัญญาณเสียงและภาพได้เร็วและต่อเนื่อง มาตรฐาน MPEG ได้รับการยอมรับจาก ISO IEC/JTC SC29 ในปี 1991 และมีการพัฒนาเรื่อยมา ดังแสดงในตารางที่

ตารางที่ 1.2 มาตรฐานการเข้ารหัสสัญญาณวิดีโอของ MPEG

หมายเลขมาตรฐาน	ปีที่กำหนดใช้	วัตถุประสงค์
MPEG-1	ค.ศ. 1992	เพื่อเข้ารหัสข้อมูลภาพเคลื่อนไหวสำหรับมัลติมีเดียและการเก็บข้อมูลในสื่อเก็บข้อมูล เช่น แผ่นซีดีรอม (Video CD)
MPEG-2	ค.ศ. 1995	พัฒนาจาก MPEG-1 เพื่อการส่งกระจายสัญญาณสำหรับโทรทัศน์ระบบดิจิทัลความละเอียดสูง (HDTV), โทรทัศน์ผ่านดาวเทียมและเคเบิล รวมทั้งพัฒนาการเก็บข้อมูลให้เก็บได้มีคุณภาพสูงขึ้นและมีคุณสมบัติเพิ่มขึ้น เช่น DVD
MPEG-4	ค.ศ. 1999 (เวอร์ชัน 1) ค.ศ. 2000 (เวอร์ชัน 2)	เพื่อการส่งสัญญาณวิดีโอแบบอัตราบิตต่ำมาก (Very low bit rate) และนำไปประยุกต์ใช้งานทางอินเทอร์เน็ตแบบโต้ตอบกันได้ (Interactive) มุ่งเน้นในการบีบอัดข้อมูลและการใช้งานทางมัลติมีเดีย
MPEG-7	ค.ศ. 2001	เพื่อพัฒนามาตรฐานสู่การค้นหาและจัดเก็บข้อมูลในรูปแบบของมัลติมีเดีย

รายละเอียดของมาตรฐาน MPEG

MPEG-1 เป็นมาตรฐานการเข้ารหัสสัญญาณภาพเคลื่อนไหวและเสียง สำหรับการเก็บข้อมูลและดึงข้อมูลกลับมาใช้บนสื่อเก็บข้อมูล มีอัตราบิตได้สูงถึง 1.6 Mbps MPEG-1 ได้รับการจัดทำขึ้นเมื่อ ค.ศ. 1988 และพัฒนาจนเป็นมาตรฐานในปี ค.ศ. 1994 มีความเหมาะสมสำหรับการใช้งานเก็บข้อมูลและการสื่อสารผ่านช่องสัญญาณที่มีอัตราบิต 1-5 Mbps เช่น เครือข่าย ท้องถิ่น (LAN) เป็นต้น มีรูปแบบการจัดเก็บแบบ SIF (Source Input Format) ขนาดภาพแนวนอนมีได้ถึง 768 จุดภาพและขนาดในแนวตั้งมีเส้นภาพได้ถึง 576 เส้น

MPEG-2 เป็นมาตรฐานที่พัฒนาเพื่อรองรับการใช้งานในขอบข่ายกว้างขึ้นและมีคุณภาพดีขึ้น สามารถส่งข้อมูลที่อัตราบิตได้สูงถึง 2-15 Mbps มุ่งเน้นสนับสนุนการใช้งานด้านเก็บข้อมูล การส่งข้อมูลแบบกระจาย (broadcast) จากเครื่องส่งเครื่องหนึ่งไปสู่เครื่องรับหลาย ๆ เครื่อง และการใช้งานด้านการสื่อสารซึ่งรวมถึงโทรศัพท์ภาพและการประชุมทางวิดีโอด้วย

MPEG-4 มีการเริ่มพัฒนามาตรฐานในเดือนพฤศจิกายน ปี ค.ศ. 1993 โดยเริ่มแรกมีจุดมุ่งหมายเพื่อพัฒนามาตรฐานการเข้ารหัสให้ได้อัตราการบีบอัดข้อมูลสูงๆ ซึ่งกำหนดค่าอัตราบิตไว้ที่ต่ำกว่า 64 Kbps เพื่อใช้งานในการสื่อสารข้อมูลที่อัตราบิตต่ำมาก หลังจากนั้นได้มีการขยายมาตรฐานเพิ่มเติม รวมทั้งปรับปรุงเครื่องมือ กรรมวิธีและฟังก์ชันการใช้งานมากขึ้น และยังเพิ่มความสามารถในการโต้ตอบกับผู้ใช้ เพื่อให้ควบคุมและจัดการกับหน่วยภาพและเสียงได้อย่างมีประสิทธิภาพ มาตรฐาน MPEG-4 เวอร์ชัน 1 ได้พัฒนาเป็นมาตรฐานเรียบร้อยเมื่อเดือนธันวาคม ปี ค.ศ. 1998 ซึ่งมุ่งนำไปใช้งาน 3 ด้านด้วยกัน ได้แก่ โทรทัศน์ระบบดิจิทัล การใช้งานด้านกราฟฟิก และการสื่อสารมัลติมีเดียแบบโต้ตอบ (Interactive multimedia) ส่วนมาตรฐาน MPEG-4 เวอร์ชัน 2 พัฒนาขึ้นเพื่อเพิ่มความสามารถและประสิทธิภาพให้กับเวอร์ชันแรก

MPEG-7 เริ่มโครงการขึ้นในปี 1996 มีชื่อเต็มว่า “ Multimedia Content Description Interface ” โดยกำหนดเพื่อใช้อธิบายถึงลักษณะของข้อมูลมัลติมีเดีย และแบบแผนของการอธิบาย (description schemes) ซึ่ง MPEG-7 กำหนดรูปแบบภาษาที่ใช้ เรียกว่า Description Definition Language (DDL) หน่วยวัตถุของภาพและเสียง (audiovisual object) จะสามารถกำหนดดัชนี (index) และค้นหาข้อมูลโดยใช้ดัชนีนี้ หน่วยวัตถุสามารถเป็นได้ทั้งภาพนิ่ง ภาพกราฟฟิก สัญญาณเสียง สัญญาณวิดีโอ หรือข้อมูลที่มีสื่อเหล่านี้ผสมกันอยู่ ซึ่งอาจอยู่ในรูปแบบของมาตรฐานอื่น เช่น MPEG-1 MPEG-2 หรือ MPEG-4 ได้อีกด้วย

1.3 ปรีทัศน์วรรณกรรม (Literature Review) ของการแยกส่วนภาพวัตถุเคลื่อนที่ในลำดับภาพ

มาตรฐาน MPEG-4 ได้กำหนดกรรมวิธีในการเข้ารหัสสัญญาณวิดีโอโดยอยู่บนพื้นฐานของหน่วยวัตถุ และขั้นตอนที่สำคัญขั้นต้นก่อนการเข้ารหัสสัญญาณวิดีโอแต่ละหน่วยวัตถุ คือ ขั้นตอนการแยกส่วนของภาพวัตถุในลำดับภาพ หลักพื้นฐานของการแยกส่วนภาพวัตถุสำหรับสัญญาณวิดีโอ คือ การแยกส่วนภาพวัตถุที่เคลื่อนที่และภาพพื้นหลังที่อยู่นิ่งแยกจากกัน

การแยกส่วนของภาพวัตถุเคลื่อนที่ในลำดับภาพมีความสำคัญอย่างมากในการใช้งานด้าน มัลติมีเดียและการเข้ารหัสสัญญาณวิดีโอ กล่าวคือ สัญญาณภาพจะถูกแยกส่วนในรูปของภาพวัตถุ เพื่อให้มีการบีบอัดข้อมูลที่มีประสิทธิภาพสูง การเข้ารหัสรูปทรงและการเข้ารหัสพื้นผิวของแต่ละภาพวัตถุจะแยกจากกันทำให้ได้ประสิทธิภาพการบีบอัดข้อมูลที่สูง จากแนวความคิดนี้ภาพวัตถุที่ถูกแยกส่วนนั้นอาจจะไม่เป็นรูปร่างลักษณะที่มีความหมาย ปัจจุบันการใช้งานด้าน มัลติมีเดีย เช่น การได้กลับคืนมา (retrieval) และการประกอบกัน (composition) ของสัญญาณภาพและวิดีโอบนพื้นฐานของคอนเทนต์ (content based image/video) ต้องการภาพวัตถุที่แยกส่วนออกมาแล้วมีความหมายตามการมองของมนุษย์ ถึงแม้ว่าเทคนิคการแยกส่วนภาพวัตถุจะถูกนำเสนอขึ้นมาจำนวนมาก แต่กรรมวิธีการแยกส่วนภาพวัตถุอย่างอัตโนมัติสำหรับการใช้งานทั่วไป ยังไม่ประสบความสำเร็จมากนัก หัวข้อนี้จะกล่าวถึงงานวิจัยและความสำคัญของงานวิจัยของการแยกส่วนภาพวัตถุเคลื่อนที่

1.3.1 การแยกส่วนบนพื้นฐานความไม่ต่อเนื่องของการไหลทางแสง (optical flow discontinuities)

ช่วงแรกของงานวิจัยการแยกส่วนพยายามที่จะแยกส่วนภาพเป็นวัตถุที่เคลื่อนที่โดยใช้การวัดบริเวณท้องถิ่น (local measurement) [Potter J. L., 1975] Potter ได้ใช้การวัดอัตราเร็วในการแยกส่วน โดยอยู่บนสมมติฐานที่ว่า “ทุกๆส่วนของวัตถุมีความเร็วเท่ากันขณะกำลังเคลื่อนที่อยู่” ซึ่งใช้การวัดการเคลื่อนที่ของขอบวัตถุ การวัดการเคลื่อนที่จะกำหนดจุดอ้างอิงเพื่อให้เห็นความแตกต่างระหว่างภาพของวัตถุเคลื่อนที่กับจุดอ้างอิง จุดอ้างอิงจะแบ่งเป็น 3 คลาส คือ ส่วนตัววัตถุ ส่วนเงา และพื้นหลัง จุดต่างๆ จะถูกรวมกลุ่มภายในคลาสบนพื้นฐานของการวัดการเคลื่อนที่ [Potter J. L., 1977] อีกงานวิจัยหนึ่ง Potter จะกำหนดการประมาณความเร็วของจุดภาพโดยใช้วิธีการเข้าคู่เทมเพลต (template matching) วิธีการเข้าคู่เทมเพลตจะให้ข้อมูลความเร็วที่ถูกต้องกว่าสำหรับฉากที่มีความซับซ้อน ข้อได้เปรียบของการใช้เทมเพลต คือ จะไม่ขึ้นกับลักษณะของวัตถุ เนื่องจากลักษณะของเทมเพลตไม่ได้มีทุกรูปแบบ ดังนั้นถ้าไม่มีการวิเคราะห์เชิงพื้นที่ (spatial analysis) จะไม่สามารถแยกส่วนภาพวัตถุทั้งหมดได้

Spoerri และ Ullman [Spoerri A., Ullman S., 1987] นำเสนอการตรวจรู้ขอบเขตการเคลื่อนที่ก่อนจะมีการเคลื่อนที่ต่อเนื่อง โดยการทดสอบความไม่ต่อเนื่องของการไหล (flow discontinuity tests) ทั้งนี้ข้อมูลที่ใช้ คือ ฮิสโตแกรมการไหลที่สร้างจากจุดข้างเคียงแต่ละจุดภาพ วิธีนี้เหมาะกับลำดับภาพที่ประมาณการเคลื่อนที่ได้ยาก แต่ยังคงเห็นขอบเขตการเคลื่อนที่ ข้อได้เปรียบของวิธีการทดสอบการไหล คือ สามารถตรวจรู้ขอบโดยไม่ต้องรู้การเคลื่อนที่ของวัตถุได้ ซึ่งจะช่วยในการแยกส่วนของวัตถุเคลื่อนที่ต่อไป

Overington [Overington I., 1987] ใช้ส่วนประกอบของการไหลคำนวณลักษณะที่ขอบของวัตถุเพื่อหาความต่อเนื่อง และความไม่ต่อเนื่อง สำหรับตรวจรู้วัตถุที่เคลื่อนที่ในภาพ

Thompson [Thompson W. B., 1985] ใช้การตรวจรู้ขอบของการเคลื่อนที่บนภาพเพื่อหาขอบเขตของวัตถุด้วยการขยายหลักการของการตรวจรู้ขอบบนพื้นฐานของความเข้ม โดยวิธีนี้เกรเดียนต์ของการไหลที่ราบเรียบแบบเกาส์ (Gaussian smoothed flow) และ nonmaxima suppression จะถูกใช้เพื่อตรวจรู้การเคลื่อนที่ในแต่ละส่วนประกอบของการไหลของภาพ

Clocksins [Clocksin W. F., 1980] เสนอการใช้ตัวปฏิบัติการลาปลาเซียน (Laplacian operator) สำหรับตรวจรู้การเปลี่ยนแปลงของรูปทรงในสนามความเร็วที่สร้างขึ้น ความไม่ต่อเนื่องของการไหลสามารถตรวจรู้ได้ด้วยตัวปฏิบัติการลาปลาเซียน ส่วน Schunck [Schunck B. G., 1989] จะใช้วิธีที่คล้ายกันโดยใช้การตรวจรู้ขอบของการเคลื่อนที่บนสนามการไหลที่ราบเรียบ กรรมวิธีทั้งสองนี้จะมีปัญหาเหมือนกัน คือ การแยกส่วนมากเกินไป (over segmentation) จากการตรวจรู้ขอบเกรเดียนต์บนพื้นฐานของความเข้ม สามารถแก้ไขได้โดยการรวมข้อมูลเชิงพื้นที่ เช่น ข้อมูลความเข้มสีและลักษณะของพื้นผิว ดังนั้นการแยกส่วนมากเกินไปในสนามการเคลื่อนที่ที่สามารถแก้ไขได้ด้วยการแยกส่วนเชิงพื้นที่

ขณะที่การแยกส่วนภาพบนพื้นฐานของการหาความไม่ต่อเนื่องของการไหลยังมีการดำเนินการวิจัยต่อไป แต่ยังไม่ประสบผลตามที่คาดไว้ จุดสำคัญ คือ สนามการไหลจะมีลักษณะทางสถิติเหมือนกับความเข้มของภาพ ซึ่งจำเป็นต้องใช้ข้อมูลระดับสูงและกฎเกณฑ์เพื่อช่วยในการวิเคราะห์

1.3.2 การแยกส่วนภาพวัตถุบนพื้นฐานของการตรวจรู้การเปลี่ยนแปลง (Segmentation based on change detection)

วิธีการไหลทางแสง หรือการจับความเร็วจะถูกคำนวณที่ทุกๆ เฟรมของลำดับภาพ แต่เนื่องจากจำนวนของจุดภาพส่วนใหญ่จะไม่มีเคลื่อนที่หรือมีการเคลื่อนที่ทั้งหมด วิธีที่มี ประสิทธิภาพดีกว่า คือ การวิเคราะห์บนพื้นที่ที่มีการเปลี่ยนแปลง นอกจากนั้นการแยกส่วนของวัตถุที่เคลื่อนที่ปกติจะใช้กระบวนการหลายขั้นตอนหรือกระบวนการซ้ำ การกำจัดบริเวณที่ไม่เกี่ยวข้องจึงสามารถป้องกันความผิดพลาดและลดบริเวณที่พิจารณาได้ ในส่วนของพื้นหลังส่วนใหญ่จะอยู่นิ่งหรือมีการ

เคลื่อนที่แบบง่าย ๆ จึงสามารถกำจัดบริเวณนี้ได้โดยวิธีดูความแตกต่างพื้นฐาน หรือความแตกต่างของการขดเซกการเคลื่อนที่ ซึ่งจะใช้วิธีนี้กับการแยกส่วนสำหรับการเข้ารหัสเชิงวัตถุและสำหรับฟังก์ชันแบบคอนเทนต์ (content based functionalities) การแยกส่วนภาพโดยใช้การตรวจรู้การเปลี่ยนแปลงจะหลีกเลี่ยงการคำนวณความแตกต่างเกรเดียนต์ในกระบวนการประมาณการไหลที่ไม่แน่นอนได้ กรรมวิธีนี้เริ่มด้วยการตรวจรู้การเปลี่ยนแปลงเพื่อแยกส่วนระหว่างบริเวณที่เปลี่ยนแปลงและไม่เปลี่ยนแปลงของเฟรมภาพที่อยู่ติดกัน แล้ววัตถุที่เคลื่อนที่จะถูกแยกจากบริเวณที่มีการเปลี่ยนแปลงอีกชั้นตอนหนึ่ง การตัดสินใจของแต่ละตำแหน่งจะอยู่บนพื้นฐานของค่าความแตกต่างระหว่างเฟรม (Frame Diffence : FD)

ปกติจะใช้ค่าความแตกต่างระหว่างเฟรมสำหรับหลายจุดภาพตามขนาดของหน้าต่างมากกว่าจุดภาพเดียว ซึ่งขึ้นกับสัญญาณรบกวนในภาพ (noise) และจะถูกเปรียบเทียบกับค่าจุดเปลี่ยน (T_{ch}) ถ้าค่าความแตกต่างระหว่างเฟรมเกินค่าจุดเปลี่ยน มาส์กของการตรวจรู้การเปลี่ยนแปลง จะมีค่าเป็น 1 และจะมีค่าเป็น 0 เมื่อไม่เกินค่าจุดเปลี่ยนนี้ ซึ่งประสิทธิภาพของตัวตรวจรู้การเปลี่ยนแปลงขึ้นกับเกณฑ์ 2 อย่าง คือ การเลือกค่าจุดเปลี่ยน และเกณฑ์ที่เหมาะสมสำหรับกำจัดบริเวณที่มีขนาดเล็กออก

Jain [Jain R., 1981] ได้ทำการศึกษาเกี่ยวกับเทคนิคของการตรวจรู้การเปลี่ยนแปลง และพยายามที่จะแยกส่วนวัตถุเคลื่อนที่แบบแข็งเกร็ง (rigid moving object) โดยการวิเคราะห์ความแตกต่างของภาพอย่างต่อเนื่อง หลักการสำคัญ คือ การสร้างส่วนของฉากพื้นหลังทั้งหมดจากลำดับภาพ ซึ่งถ้าสามารถสร้างฉากพื้นหลังได้ทั้งหมด วัตถุที่เคลื่อนที่ก็จะสามารถแยกส่วนออกมาได้โดยการเปรียบเทียบแต่ละเฟรมกับฉากพื้นหลัง แต่วิธีนี้ไม่เหมาะสมในการทำงานทั่วไปเนื่องจากมีสมมติฐานที่ว่าจะต้องเป็นพื้นหลังที่อยู่นิ่ง วัตถุต้องเคลื่อนที่แบบแข็ง และการเคลื่อนที่ต้องไปในทิศทางเดียว ซึ่งเป็นข้อจำกัดของวิธีนี้ นอกจากนี้การสร้างฉากพื้นหลังจำเป็นต้องใช้ลำดับภาพยาวมากในการวิเคราะห์การเคลื่อนที่

Wang และ Adelson [Wang J., Adelson E. H., 1994] พยายามคิดค้นวิธีในการแบ่งชั้นของภาพวัตถุที่แตกต่างกันจากลักษณะของการเคลื่อนที่ในฉากพาโนรามา

Jain [Jain R., 1979] ได้เสนอการแยกส่วนภาพวัตถุที่เคลื่อนที่ในฉากแบบไดนามิก (dynamic scenes) การแยกส่วนภาพนี้ประกอบด้วย 2 ขั้นตอน ซึ่งใช้กับแต่ละคู่ของเฟรมที่ต่อเนื่องกัน ขั้นแรก คือ การแบ่งบริเวณของภาพซึ่งแตกต่างกันเป็น 3 ส่วน คือ บริเวณพื้นหลังที่ถูกครอบคลุม บริเวณพื้นหลังที่ไม่ถูกครอบคลุม และ บริเวณทั้งที่ถูกครอบคลุมและไม่ครอบคลุม จากนั้นขั้นถัดมา คือ ขั้นตอนการพิสูจน์ ที่จะพิสูจน์ส่วนมาส์กว่าเป็นส่วนวัตถุที่เคลื่อนที่จริงหรือไม่

Jayaramamurthy และ Jain [Jayaramamurthy S. N., Jain R., 1983] ได้เสนอวิธีการแยกส่วนภาพวัตถุบนพื้นฐานของความแตกต่างของภาพ โดยขั้นแรกใช้ความแตกต่างของความเข้มตรวจสอบบริเวณที่เคลื่อนไหว ค่าจุดเปลี่ยนที่เลือกมาจะมีค่าเป็นร้อยละ 10 ของค่าความเข้มสูงสุดที่พบในเฟรม ขั้นตอนถัดมาจะใช้การแปลงแบบเฮาท์ (Hough transform) เพื่อกำหนด พารามิเตอร์การเคลื่อนที่ให้สอดคล้องกับแต่ละบริเวณที่เคลื่อนไหว สุดท้ายค่าความเข้มที่เปลี่ยนแปลงและพารามิเตอร์การเคลื่อนที่จะถูกรวมเพื่อสร้างมาส์กของวัตถุเคลื่อนที่ ซึ่งเป็นการรวมข้อดีของวิธีเฉพาะที่และโดยรวมเข้าด้วยกัน

วิธีการตรวจรู้การเปลี่ยนแปลงโดยตั้งค่าจุดเปลี่ยนแบบพื้นฐานจะมีความผิดพลาดในสภาพโดยรวม ด้วยเหตุนี้จึงมีการพัฒนาการตรวจรู้เป็นลำดับขั้นหรือกรรมวิธีแบบรีแลกเซชัน (relaxation algorithm)

Aach [Aach T., Kaup A., 1993] เสนอวิธีที่ใช้แบบจำลองทางสถิติบนพื้นฐานของการตรวจรู้การเปลี่ยนแปลง เริ่มด้วยการคำนวณระดับความแตกต่างของความเข้ม โดยคำนวณค่าจุดเปลี่ยนด้วยการทดสอบนัยสำคัญ (significance test) บนสมมติฐานของความแตกต่างของความเข้มที่ถูกจำลองด้วยสัญญาณรบกวนแบบเกาส์จากกล้องจับภาพ (Gaussian camera noise) มาส์กของการตรวจรู้การเปลี่ยนแปลงจะถูกตรวจสอบด้วยวิธีรีแลกเซชันบนเกณฑ์ของค่าเฉลี่ยค่าสัมบูรณ์ของความแตกต่าง (MAD) ซึ่งกรรมวิธีนี้จะแก้ปัญหาการเกิดโคโรนา (corona) แต่มาส์กของวัตถุยังมีส่วนของบริเวณพื้นที่เล็กๆ ซึ่งจำเป็นที่จะต้องกำจัดส่วนนี้ต่อไป

1.3.3 การแยกส่วนสำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ (Segmentation for Object Based Video Coding)

กรรมวิธีนี้จะใกล้เคียงกับกรรมวิธีการแยกส่วนระนาบวัตถุของสัญญาณวิดีโอ (VOP) โดยการรวมกันของเทคนิคการตรวจรู้การเปลี่ยนแปลง (change detection) และการแยกส่วนการเคลื่อนที่แบบพารามิเตอร์ (parametric motion segmentation)

Thoma และ Bierling [Thoma R.; 1989] ได้เสนอกรรมวิธีการเข้ารหัสโดยใช้การประมาณการชดเชยการเคลื่อนที่ (motion compensating interpolation) ซึ่งการตรวจรู้การเปลี่ยนแปลงปรับปรุงจากกรรมวิธีของ Hotter และ Thomas [Hotter M.; 1988]

กรรมวิธีการตรวจรู้การเปลี่ยนแปลงนี้มีกระบวนการ 3 ขั้นตอน
ขั้นแรก จะตั้งค่าจุดเปลี่ยนโดยใช้เกณฑ์ของค่าเฉลี่ยค่าสัมบูรณ์ของความแตกต่างระหว่างเฟรม (mean absolute frame difference) เริ่มต้นด้วยการตั้งค่าจุดเปลี่ยนไว้ที่ระดับ $3/256$ ของความเข้มสูงสุด

ขั้นที่สอง จะใช้มีเดียฟิลเตอร์ (median filter) เพื่อให้ขอบเขตของบริเวณที่เปลี่ยนแปลงราบเรียบขึ้น

ขั้นสุดท้าย จะกำจัดส่วนของมาสก์การเปลี่ยนแปลงที่มีขนาดเล็กกว่าขนาดที่กำหนด หลังจากนั้น ค่าจุดเปลี่ยนเริ่มต้นจะถูกปรับเปลี่ยนไปตามค่าเบี่ยงเบนมาตรฐาน (standard deviation) ของสัญญาณรบกวนบนบริเวณที่ไม่เปลี่ยนแปลง กระบวนการนี้จะทำซ้ำต่อไปจนกระทั่งระบบมีเสถียรภาพ การแยกส่วนของวัตถุที่เคลื่อนที่ คือ การแยกส่วนบริเวณพื้นหลังทั้งที่ถูกครอบคลุมและไม่ถูกครอบคลุมออกจากมาสก์ของการเปลี่ยนแปลงบนพื้นฐานของสนามการเคลื่อนที่ก่อนหน้านี้

Hotter และ Thomas [Hotter M., Thoma R.; 1988] ได้เสนอกรรมวิธีการแยกส่วนภาพที่มีโครงสร้างเป็นลำดับชั้น ซึ่งแต่ละบริเวณที่ถูกแยกจะอธิบายด้วยเซตของพารามิเตอร์การเคลื่อนที่

ขั้นแรก การตรวจรู้การเปลี่ยนแปลงจะกระทำกับภาพ 2 เฟรมที่ต่อเนื่องกัน เพื่อแยกส่วนระหว่างบริเวณที่มีการเปลี่ยนแปลงและไม่เปลี่ยนแปลง ซึ่งกระบวนการตรวจรู้การเปลี่ยนแปลงจะมี 3 ขั้นตอนตามที่อธิบายไว้ข้างต้น แต่ละบริเวณการเปลี่ยนแปลงที่ติดต่อกันจะแสดงเป็นวัตถุที่เคลื่อนที่วัตถุหนึ่ง และจะถูกประมาณการเคลื่อนที่โดยพารามิเตอร์ 8 ตัวของแบบจำลองการเคลื่อนที่ ซึ่งวิธีที่ใช้ถูกเสนอโดย Tsai และ Huang [Tsai R. Y., Huang T. S., 1981] ส่วนสนามการกระจัด (displacement field) สามารถอธิบายได้จากแบบจำลองการประมาณการเคลื่อนที่ โดยสนามการกระจัดใช้เพื่อตรวจรู้วัตถุที่เคลื่อนที่แยกจากพื้นหลังที่ไม่ถูกครอบคลุม หรือพื้นหลังที่ถูกครอบคลุมด้วยส่วนที่เปลี่ยนแปลง

ขั้นถัดมาของการแยกส่วนแบบลำดับชั้น คือ การส่งผ่านพารามิเตอร์เพื่อการชดเชยการเคลื่อนที่ในบริเวณที่เปลี่ยนแปลง กระบวนการนี้จะมีการทำซ้ำเป็นลำดับชั้นจนกระทั่งทุกวัตถุที่เคลื่อนที่ที่ถูกอธิบายได้ด้วยพารามิเตอร์การแมปปิง (mapping parameters) เทคนิคของการแยกส่วนนี้จะแยกส่วนได้เฉพาะพื้นหลังและฉากหน้าซึ่งเป็นวัตถุที่เคลื่อนที่เพียงวัตถุเดียวเท่านั้น ดังนั้นเพื่อที่จะแสดงวัตถุที่เคลื่อนที่ต่างกันได้ กรรมวิธีนี้ได้ถูกปรับปรุงโดย Mech และ Wollborn ในโครงร่างของ MPEG-4 แนวความคิดที่จะแทนบริเวณเคลื่อนที่เป็นระดับชั้นที่แตกต่างกันได้ปรับปรุงโดย Wang และ Adelson [Wang J., Adelson H. E.; 1994] และ Borshukov นอกจากนี้เพื่อที่จะแทนฉากของภาพเป็นเลเยอร์ (layer) กรรมวิธีนี้เหมาะกับการเข้ารหัสบนพื้นฐานของหน่วยวัตถุ (object based coding) ซึ่งจะเน้นที่อัตราการบีบอัดข้อมูลโดยที่ไม่ได้พิจารณาความถูกต้องของขอบภาพ ส่วนสำหรับการแยกส่วนภาพวัตถุบนพื้นฐานของคอนเทนต์ (content based object segmentation) ซึ่งข้อมูลของขอบภาพเป็นส่วนที่จำเป็น จึงใช้ข้อมูลเชิงพื้นที่ (spatial information) ร่วมด้วย ปัญหาที่เสนอแนวทางแก้ไขไว้โดย Mech และ Wollborn [Mech R., Wollborn M.; 1998]

Diehl [Diehl N. ; 1991] ได้ปรับปรุงวิธีของ Hotter และ Thomas ให้เป็นกรรมวิธีเชิงพื้นที่-เวลา (spatio-temporal) ทั้งส่วนของขอบและพื้นผิวจากภาพเดียวกันและจากภาพที่ต่อเนื่องกันถูกใช้เพื่อแยกฉากภาพเป็นวัตถุต่างๆ ซึ่งกรรมวิธีที่ใช้ส่วนใหญ่จะคล้ายกับของ Hotter และ Thomas แต่ในทุกๆ ขั้นตอน ผลของการแยกส่วนภาพจะถูกแก้ไขโดยการแยกส่วนภาพภายในเฟรม สำหรับการแยกส่วนภาพระหว่างเฟรมจะควบคุมการรวมกันของบริเวณคล้ายกับการที่ใช้ข้อมูลการเคลื่อนที่เป็นเกณฑ์หลักในการแยกส่วน ในการใช้การแยกส่วนระหว่างเฟรม แต่ละบริเวณจะกำหนดให้เป็นบริเวณเคลื่อนที่ และบริเวณประชิด (adjacent regions) ที่มีรูปแบบเหมือนกันจะรวมกันเป็นวัตถุเคลื่อนที่ การใช้การแยกส่วนภายในเฟรมและเพิ่มหลักการรวมกันจึงเป็นส่วนที่ได้รับการปรับปรุง แต่ยังไม่มียุทธศาสตร์ในการเลือกค่าเริ่มต้นของพารามิเตอร์การประมาณและการเชื่อมต่อของขอบวัตถุ

การแยกส่วนสำหรับการเข้ารหัสบนพื้นฐานของหน่วยวัตถุ (object based coding) จะให้ผลการแยกส่วนวัตถุเป็นบริเวณที่มีการเคลื่อนที่เหมือนกัน ซึ่งแทนด้วยเซตของพารามิเตอร์การเคลื่อนที่แบบเดียวกัน บริเวณนี้จะแสดงเฉพาะส่วนที่มีการเคลื่อนที่เหมือนกัน แต่จะไม่แสดงถึงขอบเขตของวัตถุที่แท้จริง Diehl [Diehl N.; 1991] จะใช้การแยกส่วนเชิงพื้นที่ร่วมด้วย โดยมีจุดประสงค์หลักเพื่อรวมบริเวณเล็ก ๆ เข้ากับบริเวณหลัก และเติมส่วนที่ขาดหายของบริเวณหลัก แต่ไม่สามารถแสดงขอบเขตที่แท้จริงและแยกวัตถุที่เคลื่อนที่ต่างกันได้ นอกจากนี้ยังไม่แสดงภาพวัตถุที่มีความหมาย เพราะการเคลื่อนที่ของวัตถุที่แท้จริงจะไม่เกาะกลุ่มกัน วิธีนี้จะใช้สำหรับตรวจจู่การเคลื่อนที่ของวัตถุซึ่งเป็นกรรมวิธีที่ง่ายและเหมาะสมสำหรับการเข้ารหัสสัญญาณวิดีโอโดยคำนึงถึงอัตราบิตที่ใช้

การให้ได้ภาพวัตถุที่มีความหมายจำเป็นต้องใช้เทคนิคการแยกส่วนสำหรับฟังก์ชันเชิงคอนเทนต์ (content based functionalities)

1.3.4 การแยกส่วนแบบอัตโนมัติสำหรับฟังก์ชันเชิงคอนเทนต์ (Automatic Segmentation for Content Based Functionalities)

ปัจจุบันการแยกส่วนภาพวัตถุสำหรับฟังก์ชันเชิงคอนเทนต์มีความจำเป็นอย่างมากเพื่อให้สามารถแยกส่วนภาพของสัญญาณวิดีโอเป็นภาพวัตถุที่มีความหมาย ตามข้อกำหนดของมาตรฐาน MPEG-4

Mech และ Wollborn [Mech R., Wollborn M., 1998] ได้เสนอแบบแผนการแยกส่วนภาพโดยอยู่บนพื้นฐานของกรรมวิธีของ Hotter และ Thomas [Hotter M., Thoma R., 1988] กรรมวิธีนี้แบ่งเป็น 4 ขั้นตอน

ขั้นแรก การเคลื่อนที่ของกล้องจับภาพจะถูกประมาณและชดเชยการเคลื่อนที่โดยใช้แบบจำลองการเคลื่อนที่ 8 พารามิเตอร์

ขั้นที่สอง การตัดฉาก (scene cut) และการเลื่อนของกล้องจับภาพ (camera panning) จะถูกตรวจรู้โดยการประเมินค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ระหว่างสองเฟรมที่ต่อเนื่องกัน แล้วจะนำเฉพาะบริเวณพื้นหลังของเฟรมก่อนหน้ามาพิจารณา ส่วนกรณีที่เกิดการตัดฉาก ขั้นตอนนี้จะทำใหม่อีก

ขั้นที่สาม มาตรฐานของการเปลี่ยนแปลง (CDM) จะถูกสร้างขึ้นโดยใช้ข้อมูลระหว่างเฟรมสองเฟรมที่ต่อเนื่องกัน โดยใช้เทคนิครีแลกเซชัน (relaxation technique) และค่าจุดเปลี่ยนท้องถิ่น (local threshold) ซึ่งจะพิจารณาสถานะของจุดภาพข้างเคียง ส่วนหน่วยความจำจะถูกนำมาใช้สำหรับเก็บมาตรฐานการเปลี่ยนแปลง เพื่อสร้างบริเวณภาพวัตถุที่มีเสถียรภาพ

ขั้นที่สี่ มาตรฐานของวัตถุเคลื่อนที่ (OM) จะได้จากการกำจัดพื้นหลังที่ไม่ถูกครอบคลุม (uncovered background) จาก CDM แล้วปรับแต่งบริเวณขอบของภาพ การประมาณการเคลื่อนที่และการสร้างมาตรฐานของวัตถุ จะอยู่บนพื้นฐานของกรรมวิธีของ Hotter และ Thomas ส่วนการตรวจรู้การเปลี่ยนแปลงถูกปรับปรุงมาจากกรรมวิธีของ Aach [Aach T., 1993] ซึ่งมีแนวคิดสำคัญ คือ การสร้างมาตรฐานของวัตถุเริ่มแรกจากการตรวจรู้การเปลี่ยนแปลง และนำมาตรฐานนี้ไปใช้ในการติดตามการเคลื่อนที่ของวัตถุที่สนใจด้วย

การสร้างมาตรฐานของวัตถุเริ่มแรก จะใช้การชดเชยการเคลื่อนที่โดยรวม เพื่อชดเชยการเคลื่อนที่ของกล้องจับภาพ หลังจากนั้นจึงใช้การตรวจรู้การเปลี่ยนแปลง กรรมวิธีนี้สามารถให้กล้องจับภาพเลื่อน (panning) และขยายภาพ (zooming) ได้ ในการประมาณการเคลื่อนที่ทั้งเฟรม จุดภาพที่ใกล้ขอบภาพทั้งซ้ายและขวาน้อยกว่า 10 จุดภาพจะใช้เป็นจุดอ้างอิง ซึ่งจะสมมุติว่าไม่มีการเคลื่อนที่ที่ใกล้กับขอบภาพนั้น

Meier [Meier T., Ngan K. N., 1997] เสนอแบบแผนการแยกส่วนระนาบภาพวัตถุของสัญญาณวิดีโอด้วยกระบวนการคล้ายกับ [Mech R., Wollborn M., 1998] โดยเริ่มด้วยการประมาณการเคลื่อนที่โดยใช้การจับคู่บล็อก แล้วจำลองการเคลื่อนที่ด้วยแบบจำลอง 6 พารามิเตอร์ มาตรฐานของภาพวัตถุเริ่มแรกจะได้จากการรวมกันของผลการแยกส่วนเชิงเวลาด้วยการปฏิบัติการเชิงสัดส่วนกับการแยกส่วนเชิงพื้นที่โดยใช้ตัวปฏิบัติการแคนนี่ (Canny operator) และแทนที่การใช้หน่วยความจำในการติดตามการเคลื่อนที่ของวัตถุด้วยการใช้ Hausdorff distance และกระบวนการปรับตัวของแบบจำลอง ส่วนการสร้างระนาบของวัตถุได้จากกระบวนการภายหลัง (post processing) เพราะพารามิเตอร์บางค่าในกระบวนการแยกส่วนการเคลื่อนที่จำเป็นต้องกำหนดจากผู้ใช้อยู่ ซึ่งทำให้ยังไม่เป็นการแยกส่วนแบบอัตโนมัติ

Neri [Neri A., Colonnese S., 1998] สร้างการแยกส่วนของแต่ละเฟรมในลำดับภาพด้วยการประมวลผลเป็นกลุ่มของเฟรม (GOF) $\{f_{k-1}, i = 0, \dots, n\}$ จำนวนของเฟรม (n) ขึ้นกับความเร็วของวัตถุเคลื่อนที่ ขึ้นตอนส่วนใหญ่คล้ายกับกรรมวิธีของ Mech [Mech R., Wollborn M., 1998] โดยหลังจากการขจัดเซกการเคลื่อนที่โดยรวมแล้ว จะมีกระบวนการ 3 ขั้นตอน คือ การตรวจรู้พื้นที่ที่เปลี่ยนแปลง การตรวจรู้การเคลื่อนที่ และการปฏิบัติตามข้อบังคับ (regularization) โดยจะทำการตรวจรู้การเปลี่ยนแปลงกลุ่มของเฟรมด้วย Higher Order Statistics (HOS) เพื่อหลีกเลี่ยงความแปรปรวนของความเข้มแสงเนื่องจากจุดสัญญาณรบกวน มาสก์ของวัตถุเคลื่อนที่ที่จะสร้างจากการประมาณการเคลื่อนที่บน HOS map ที่ไม่รวมพื้นที่ที่ไม่ถูกครอบคลุม (uncovered background) และในขั้นตอนของการปฏิบัติตามข้อบังคับ บริเวณที่ระบุว่าอยู่หนึ่งที่อยู่ภายในบริเวณที่ระบุว่าเคลื่อนที่จะถูกกำหนดเป็นบริเวณที่เคลื่อนที่ หรือฉากหน้า (foreground) กรรมวิธีนี้ไม่จำเป็นต้องใช้การตัดฉากและการติดตามการเคลื่อนที่

การแยกส่วนเชิงเวลาในกรรมวิธีของ Choi [Choi J. G., 1997] ได้มาจากการตรวจรู้การเคลื่อนที่บนพื้นฐานของลักษณะทางสถิติ บริเวณที่เปลี่ยนแปลงจะถูกวางบนบริเวณที่ได้จากการแยกส่วนเชิงพื้นที่ ถ้าบริเวณที่ได้จากการแยกส่วนเชิงพื้นที่แต่ละบริเวณมีจุดภาพส่วนใหญ่อยู่บนบริเวณมาสก์ของการเปลี่ยนแปลง บริเวณนั้นทั้งหมดจะถูกกำหนดให้เป็นฉากหน้า ถ้ามีเพียงส่วนน้อยจะถูกกำหนดให้เป็นพื้นหลัง

จนถึงในปัจจุบัน กรรมวิธีต่าง ๆ ที่พิจารณากันอยู่ยังไม่มีการใช้ข้อมูลความลึกของภาพ เพราะว่าการใช้งาน อย่างเช่น การเข้ารหัส ฟังก์ชันเชิงคอนเทนต์ หรือการติดตามวัตถุ จะวิเคราะห์เฉพาะภาพในฉาก 2 มิติ ซึ่งกล้องจับภาพจะมีการหมุน (rotation) การขยาย (zooming) และการเลื่อนที่ (translation) เท่านั้น จึงละเลยความลึกได้ แต่ข้อมูลความลึก (depth information) จะมีความจำเป็นเพื่อช่วยในการวิเคราะห์การเคลื่อนที่ของกรณีที่วัตถุเกิดการบังกัน (occlusion)

ดังนั้น Pardas [Pardas M., 1997] จึงได้เสนอแบบแผนการแยกส่วนโดยใช้ข้อมูลความต่อเนื่องของการเคลื่อนที่และข้อมูลระดับความลึก โดยใช้กระบวนการ 4 ขั้นตอน คือ การประมาณการเคลื่อนที่ การเปรียบเทียบพารามิเตอร์การเคลื่อนที่ การคำนวณการซ้อนทับกัน (overlapping computation) และการกำหนดระดับความลึก

กรรมวิธีการแยกส่วนในหัวข้อนี้ เป็นการแยกส่วนภาพแบบอัตโนมัติ ที่สามารถแยกส่วนภาพให้เป็นภาพวัตถุที่มีความหมาย ซึ่งประกอบด้วยพื้นหลังและภาพวัตถุเคลื่อนที่ เหมาะสมสำหรับการใช้งานฟังก์ชันเชิงคอนเทนต์ตามข้อกำหนดของ MPEG-4 เมื่อพิจารณางานวิจัยส่วนใหญ่ในการแยกส่วนของภาพวัตถุเคลื่อนที่จะอยู่บนพื้นฐานของการตรวจรู้การเปลี่ยนแปลงแทนที่การหาสนามการเคลื่อนที่ เพราะว่าความแตกต่างของเฟรมมีความน่าเชื่อถือมากกว่าสนามการเคลื่อนที่

1.4 แนวทางของวิทยานิพนธ์

เนื่องจกงานวิจัยทางด้านการเข้ารหัสสัญญาณวิดีโอ มีแนวโน้มของการพัฒนาการวิจัยทางด้านการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ ซึ่งมาตรฐาน MPEG-4 ได้กำหนดหลักการของการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุเอาไว้ เพื่อให้สามารถใช้สำหรับการสื่อสารมัลติมีเดียและรองรับการพัฒนาการสื่อสารในอนาคตได้ การเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุตามข้อกำหนดของมาตรฐาน MPEG-4 มีขั้นตอนที่สำคัญก่อนเข้าขั้นตอนการเข้ารหัส คือ ขั้นตอนการแยกส่วนภาพวัตถุของสัญญาณวิดีโอ ซึ่งเป็นการแยกส่วนของสัญญาณวิดีโอเป็นภาพวัตถุที่สนใจ และสามารถที่จะแยกส่งภาพวัตถุต่าง ๆ ในการสื่อสารผ่านทางภาพวิดีโอได้ เพื่อสนับสนุนการสื่อสารด้านมัลติมีเดีย นอกจากนี้มาตรฐาน MPEG-4 ยังมีได้กำหนดกรรมวิธีในการแยกส่วนภาพวัตถุนี้ จึงเป็นงานวิจัยที่น่าสนใจในการพัฒนาให้สอดคล้องกับมาตรฐาน MPEG-4 วิทยานิพนธ์ฉบับนี้ได้เสนอกรรมวิธีการแยกส่วนภาพวัตถุของสัญญาณวิดีโอโดยอยู่บนพื้นฐานของการรวมลักษณะสำคัญของสัญญาณวิดีโอ ทั้งข้อมูลเชิงพื้นที่ และข้อมูลเชิงเวลา เนื่องจากกรรมวิธีการแยกส่วนภาพวัตถุมีส่วนที่สามารถปรับปรุงได้หลายส่วนด้วยกัน การแยกส่วนเชิงพื้นที่สามารถสร้างภาพให้มีความเรียบง่ายต่อการแยกส่วนภาพมากขึ้น โดยใช้การปฏิบัติการเชิงสัณฐาน (Morphological Operation) และปรับปรุงการตรวจรู้ขอบเขตของบริเวณด้วยการประมาณเกรเดียนต์เชิงสัณฐาน เพื่อลดทรัพยากรในการคำนวณของขั้นตอนการกำหนดบริเวณให้กับจุดภาพ ส่วนการแยกส่วนเชิงเวลาจะใช้การตรวจรู้การเปลี่ยนแปลง และใช้ผลการแยกส่วนของภาพวัตถุในเฟรมก่อนหน้าเพื่อช่วยในการตรวจรู้วัตถุเคลื่อนที่ต่อเนื่องที่หยุดนิ่ง จากนั้นจึงนำผลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อให้สามารถแยกส่วนภาพวัตถุที่สนใจได้ และภาพวัตถุที่ได้มีความหมาย ซึ่งสนับสนุนคุณสมบัติของฟังก์ชันเชิงคอนเทนตสำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุตามข้อกำหนดของ MPEG-4 และแนวโน้มการพัฒนาของ MPEG-7 ต่อไป

1.5 วัตถุประสงค์ของวิทยานิพนธ์

1.5.1 เพื่อศึกษาการเข้ารหัสสัญญาณวิดีโอที่อัตราบิตต่ำ และมาตรฐาน MPEG-4

1.5.2 เพื่อศึกษาวิธีการแยกส่วนภาพวัตถุในลำดับภาพแบบต่าง ๆ และข้อกำหนดในการแยกส่วนภาพวัตถุของ MPEG-4

1.5.3 เพื่อพัฒนาวิธีการแยกส่วนภาพวัตถุให้ได้ภาพวัตถุที่มีความหมาย และใช้ทรัพยากรการคำนวณในการประมวลผลต่ำสำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ

1.6 ขอบเขตของวิทยานิพนธ์

- 1.6.1 ได้กรรมวิธีที่มีประสิทธิภาพการแยกส่วนภาพวัตถุที่ให้ผลการแยกส่วนภาพที่มีความหมาย มีความคลาดเคลื่อนจากการแยกส่วนภาพเฉลี่ยใน 50 เฟรมไม่เกิน 15 เปอร์เซ็นต์ของขนาดภาพวัตถุ โดยมีผลการแยกส่วนภาพที่เกินน้อยกว่า 10 เปอร์เซ็นต์และส่วนที่ขาดไปน้อยกว่า 5 เปอร์เซ็นต์ เมื่อเปรียบเทียบกับมาส์กของการแยกส่วนภาพวัตถุจากแบบจำลองมาตรฐาน
- 1.6.2 สามารถแยกส่วนภาพเป็นส่วนของวัตถุที่เคลื่อนที่และพื้นหลังที่อยู่นิ่ง
- 1.6.3 นำกรรมวิธีการแยกส่วนภาพวัตถุไปทดสอบกับลำดับภาพมาตรฐานที่มีลักษณะต่าง ๆ กัน ทั้งที่มีวัตถุเคลื่อนที่เพียงวัตถุเดียวและหลายวัตถุ
- 1.6.4 ได้กรรมวิธีการแยกส่วนภาพวัตถุที่ใช้ทรัพยากรการคำนวณต่ำ
- 1.6.5 กรรมวิธีการแยกส่วนภาพวัตถุสอดคล้องตามข้อกำหนดของ MPEG-4
- 1.6.6 การวิเคราะห์การเคลื่อนที่จะพิจารณาเฉพาะการเคลื่อนที่ตำแหน่ง
- 1.6.7 การแยกส่วนภาพวัตถุเป็นแบบอัตโนมัติ

1.7 ประโยชน์ที่คาดว่าจะได้รับ

- 1.7.1 ทราบถึงกรรมวิธีการเข้ารหัสสัญญาณภาพและวิดีโอ รวมทั้งวิธีการแยกส่วนภาพวัตถุจากงานวิจัยที่ผ่านมา
- 1.7.2 สามารถพัฒนาวิธีการแยกส่วนภาพวัตถุในลำดับภาพ เพื่อเป็นพื้นฐานในการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ

1.8 ขั้นตอนและวิธีการดำเนินการ

- 1.8.1 ศึกษาการเข้ารหัสสัญญาณวิดีโอที่อัตราบิตต่ำ และมาตรฐาน MPEG-4
- 1.8.2 ศึกษาขั้นตอนการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ
- 1.8.3 ค้นคว้าและศึกษาบทความเกี่ยวกับวิธีการแยกส่วนภาพวัตถุในลำดับภาพ
- 1.8.4 วิเคราะห์ข้อได้เปรียบและข้อเสียเปรียบของกรรมวิธีต่าง ๆ ที่ได้เสนอมานับทศวรรษ
- 1.8.5 พัฒนาแนวทางปรับปรุงและวิธีการแยกส่วนภาพวัตถุสำหรับภาพนิ่งและภาพเคลื่อนไหว
- 1.8.6 เขียนโปรแกรมทดสอบกรรมวิธีที่ได้พัฒนา
- 1.8.7 วิเคราะห์ผลที่ได้จากการทดสอบ และปรับปรุงกรรมวิธีให้มีประสิทธิภาพ
- 1.8.8 สรุปและรวบรวมผลการวิจัย พร้อมทั้งจัดทำวิทยานิพนธ์

1.9 ภาพรวมของวิทยานิพนธ์

สำหรับเนื้อหาของวิทยานิพนธ์นี้แบ่งเป็น 5 บท ดังนี้

บทที่ 1 บทนำ มีเนื้อหาเกี่ยวกับความเป็นมาของวิทยานิพนธ์ ความรู้พื้นฐานของการเข้ารหัสสัญญาณวิดีโอ การพัฒนาของมาตรฐานการเข้ารหัสสัญญาณวิดีโอ ปรัชญาของกรอบการแยกส่วนภาพวัตถุ แนวทางของวิทยานิพนธ์ วัตถุประสงค์ ขอบเขตของงานวิจัย และภาพรวมของวิทยานิพนธ์

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง เนื้อหาในบทนี้จะกล่าวถึงลักษณะของสัญญาณภาพและสัญญาณวิดีโอ หลักการในการเข้ารหัสสัญญาณวิดีโอทั่วไป มาตรฐานการเข้ารหัสสัญญาณวิดีโอพื้นฐานของหน่วยวัตถุ (ISO/MPEG-4) รายละเอียดของกรรมวิธีการเข้ารหัสสัญญาณวิดีโอพื้นฐานของหน่วยวัตถุ พื้นฐานของการประมาณการเคลื่อนที่ งานวิจัยเกี่ยวกับการแยกส่วนภาพวัตถุและการเข้ารหัสรูปทรง กรรมวิธีที่ใช้สำหรับการแยกส่วนภาพวัตถุ

บทที่ 3 กรรมวิธีการแยกส่วนภาพวัตถุที่นำเสนอ ประกอบด้วย การแยกส่วนเชิงพื้นที่ (spatial segmentation) และการแยกส่วนเชิงเวลา (temporal segmentation) รวมทั้งรายละเอียดในแต่ละขั้นตอน ตั้งแต่กระบวนการทำให้ภาพเรียบง่าย (simplification process) การประมาณเกรเดียนต์ (gradient approximation) และกรรมวิธีสันปันน้ำ (watershed algorithm) การตรวจรู้การเปลี่ยนแปลง กระบวนการตัดสินใจบริเวณของภาพวัตถุ นอกจากนั้นยังแสดงลำดับภาพที่ใช้ในการทดสอบ ความเป็นมาและกรรมวิธีที่ใช้ในการเปรียบเทียบมาตรฐาน รวมทั้งเกณฑ์ในการเปรียบเทียบประสิทธิภาพของกรรมวิธีที่เสนอ

บทที่ 4 ผลการวิจัย บทนี้จะเป็นการแสดงผลจากกรรมวิธีการแยกส่วนภาพวัตถุที่นำเสนอในแต่ละขั้นตอน ตั้งแต่การสร้างความเรียบง่ายให้กับภาพ การตรวจรู้ขอบเขตของบริเวณ ผลของการแยกส่วนภาพเชิงพื้นที่ การแยกส่วนภาพเชิงเวลาโดยการตรวจรู้การเปลี่ยนแปลง ระดับค่าจุดเปลี่ยนที่ได้จากการวิเคราะห์การเปลี่ยนแปลงของลำดับภาพ การใช้หน่วยความจำเพื่อช่วยในการแยกส่วนภาพ การวิเคราะห์ลักษณะของลำดับภาพมาตรฐานที่ทดสอบ สุดท้ายเป็นการวิเคราะห์ผลการแยกส่วนภาพเมื่อเปรียบเทียบกับผลของ COST211 และมาสก์อ้างอิง

บทที่ 5 บทสรุปและข้อเสนอแนะ จะสรุปเกี่ยวกับเนื้อหาที่สำคัญในวิทยานิพนธ์ รวมทั้งปัญหาที่เกิดขึ้นและข้อเสนอแนะสำหรับงานวิจัยในอนาคต

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 ลักษณะของสัญญาณวิดีโอ (Video Signal)

สัญญาณภาพดิจิทัล (Digital Image) มีองค์ประกอบอยู่ด้วยกัน 2 ส่วน คือ ส่วนความเข้มแสง (luminance) และส่วนของความเข้มสี (chrominance) ข้อมูลเหล่านี้อยู่ในรูปแบบของความละเอียดของภาพ แสดงเป็นจำนวนจุดภาพต่อความยาวหนึ่งหน่วย เช่น จุดภาพต่อนิ้ว (pixels/inch) หรือ จุดภาพต่อเซนติเมตร (pixels/cm) ค่าความเข้มของแต่ละจุดภาพได้มาจากการควอนไทซ์ระดับความเข้มสี ณ ตำแหน่งของจุดภาพนั้น ๆ โดยทั่วไปแต่ละจุดภาพสามารถแสดงด้วยความเข้มแสงและความเข้มสี ตั้งแต่ 0 (สีดำ) ถึง 255 (สีขาว) ทั้งหมด 256 ระดับหรือ 8 บิต แต่สำหรับความเข้มสีไม่จำเป็นต้องแสดงทุกจุดภาพเนื่องจากความสามารถในการรับรู้สีของระบบการมองเห็นของมนุษย์ต่ำกว่าการรับรู้ความเข้มแสง จึงมีจำนวนจุดบอกความเข้มสีน้อยกว่าความเข้มแสงได้

รูปแบบขององค์ประกอบของสี

มีการกำหนดรูปแบบขององค์ประกอบของสีหลากหลายรูปแบบ แต่ละรูปแบบมีข้อดีข้อเสียแตกต่างกันไป และมีความเหมาะสมในการใช้งานลักษณะแตกต่างกันไป วิทยานิพนธ์ฉบับนี้ให้ความสนใจเฉพาะ 3 รูปแบบเท่านั้น ได้แก่ RGB YUV และ YCbCr

RGB เป็นระบบที่แบ่งข้อมูลของสีออกเป็น R แดง G เขียว และ B น้ำเงิน ซึ่งเป็นสามสีหลักของแสงสี ส่วน YUV กับ YCbCr จะแบ่งข้อมูลของสีออกในลักษณะเดียวกับกระบวนการรับรู้ของประสาทตาคน นั่นคือแบ่งเป็นส่วนของความเข้มของแสงที่เป็นขาวดำและส่วนของแสงสี โดย Y จะเก็บข้อมูลความเข้มของแสง เรียกว่า Luminance ส่วน U, V และ Cb, Cr จะเก็บข้อมูลของแสงสี เรียกว่า Chrominance องค์ประกอบ RGB จะสามารถแปลงไปเป็น YUV และ YCbCr ได้ด้วยสมการการแปลงองค์ประกอบของสีเฉพาะแบบ

การแปลงรูปแบบองค์ประกอบของสีจาก RGB เป็น YcbCr

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$0 \leq Y \leq 255, -128 \leq Cb \leq 128, -128 \leq Cr \leq 128$$

โดยที่ Y คือ ความเข้มของแสง (Luminance) ที่ตามนุษย์มีความไวมาก

Cb และ Cr คือ ความเข้มของสี (Chrominance) ที่ตามนุษย์มีความไวน้อย

การสุ่มค่าความเข้มของแสงและความเข้มของแสงสี

ในภาพของสัญญาณวิดีโอ นั้นแต่ละจุดภาพจะมีข้อมูลของสีอยู่ในรูปองค์ประกอบของความเข้มของแสง (Luminance) และความเข้มของแสงสี (Chrominance) ในรูปแบบของ Y,Cb,Cr หรือ Y,U,V อย่างไรก็ตามเนื่องจากกระบวนการรับรู้ของประสาทตาคน สามารถกำหนดให้ข้อมูลของส่วนความเข้มของสี มีความละเอียดน้อยกว่าส่วนความเข้มของแสงได้ โดยกำหนดให้มีข้อมูลความเข้มของสีเฉพาะในบางจุดภาพ เช่น ในลักษณะจุดภาพเว้นจุดภาพ เป็นต้น ทั้งนี้รูปแบบที่มีใช้กันมีดังนี้

- 4:4:4 sampling ratio: จะมีข้อมูล Luminance และ Chrominance ครบทุกๆ pixel
- 4:2:2 sampling ratio: จะมีข้อมูล Luminance ในทุกๆ จุดภาพ ส่วนข้อมูล Chrominance จะมีจุดภาพเว้นจุดภาพ (ในแนวนอน)
- 4:2:0 sampling ratio: จะมีข้อมูล Luminance ในทุกๆ จุดภาพ และจะลดข้อมูล Chrominance ลงอีก โดยมีข้อมูลหนึ่งจุดทุกๆ 4 จุดภาพ



รูปที่ 2.1 โครงสร้างของสัญญาณภาพในรูปแบบอัตราสุ่ม 4:2:0

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32



$(1+2+9+10)/4$	$(3+4+11+12)/4$	$(5+6+13+14)/4$	$(7+8+15+16)/4$
$(17+18+25+26)/4$	$(19+20+27+28)/4$	$(21+22+29+30)/4$	$(23+24+31+32)/4$

รูปที่ 2.2 ตัวอย่างการแปลงโดยการสุ่มลง (down sampling)

การเปลี่ยนแปลงขนาดของภาพ (Subsampling)

เนื่องจากภาพที่ต้องการแสดงมีแหล่งที่มาจากแหล่งเดียวกัน แต่ต้องการที่จะแสดงในการใช้งานต่างๆกัน ความละเอียดและขนาดของภาพก็แตกต่างกันด้วย ดังนั้นจึงต้องมีการเปลี่ยนแปลงขนาดของภาพเพื่อให้ได้ภาพตามขนาดที่ต้องการ ซึ่งการแปลงจะมีทั้งการแปลงให้ขนาดของภาพใหญ่ขึ้น (up sampling) และการแปลงให้ขนาดของภาพเล็กลง (down sampling) ดังแสดงตัวอย่างการแปลงในรูปแบบที่ 2.2 โดยที่ค่าในแต่ละบล็อกแทนค่าความเข้มของจุดภาพ และในแต่ละบล็อกแทนหนึ่งจุดภาพ

วิดีโอดิจิทัล หรือ ชุดภาพดิจิทัล เป็นกลุ่มของภาพดิจิทัลที่มีลักษณะอยู่ในรูปแบบของหน่วย เฟรมต่อหนึ่งหน่วยเวลา เช่น เฟรมต่อวินาที (frame per second) รูปแบบของสัญญาณวิดีโอได้มีการกำหนดไว้ในตารางที่ 2.1

ตารางที่ 2.1 แสดงรูปแบบของสัญญาณวิดีโอแบบ SIF และ CIF

Horizontal / Vertical Resolution	SIF (30 frames/second)	SIF (25 frames/second)	CIF (30 frames/second)
Y	360 x 242	360 x 288	352 x 288
Cr	180 x 121	180 x 144	176 x 144
Cb	180 x 121	180 x 144	176 x 144
Sampling format	4:2:0	4:2:0	4:2:0

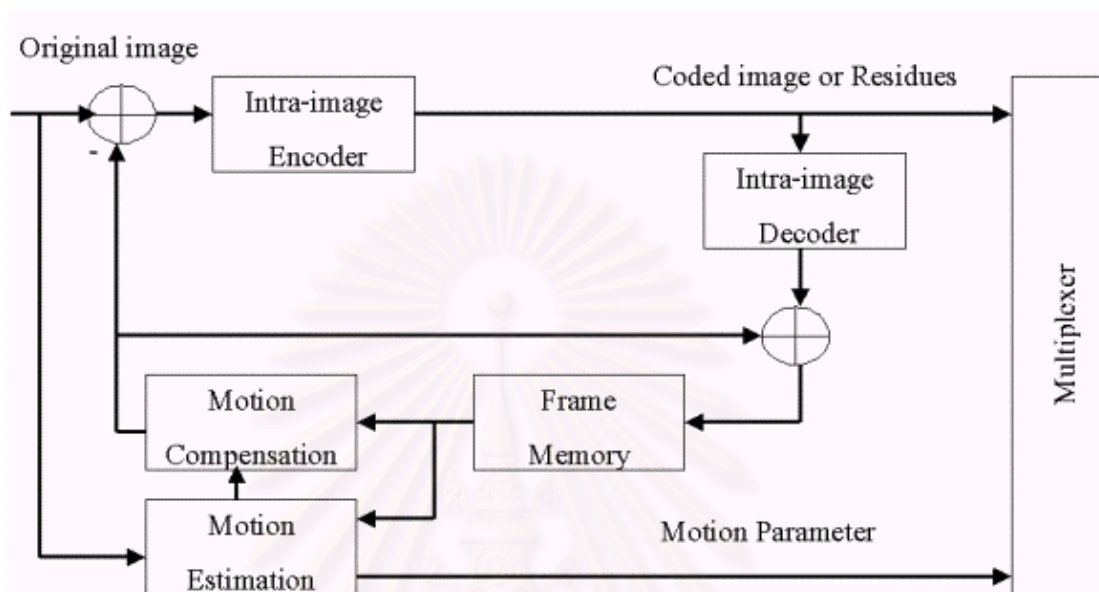
Source Input format (SIF) และ Common Interchange format (CIF)

SIF และ CIF เป็นรูปแบบของสัญญาณวิดีโอดิจิทัล ซึ่งถูกกำหนดโดยมาตรฐาน MPEG-1 และ ITU-T H.261 SIF กำหนดรูปแบบของสัญญาณวิดีโอที่ 30 เฟรมต่อวินาที และที่ 25 เฟรมต่อวินาที ส่วน CIF กำหนดรูปแบบของสัญญาณวิดีโอที่ 30 เฟรมต่อวินาทีเท่านั้น จะเป็นรูปแบบที่ใช้พื้นฐานเดียวกับโทรทัศน์ในระบบ PAL และ NTSC

2.2 การเข้ารหัสสัญญาณวิดีโอ (Video Coding)

การเข้ารหัสสัญญาณวิดีโอจะมุ่งเน้นในงานส่วนใหญ่อยู่ 2 ส่วนคือ ส่วนการเข้ารหัสช่องสัญญาณ (Channel Coding) และการเข้ารหัสแหล่งกำเนิดสัญญาณ (Source Coding) ในส่วนของการเข้ารหัสช่องสัญญาณ (Channel Coding) มีเพื่อตรวจสอบและแก้ไขความผิดพลาดของข้อมูลที่เกิดจากการส่งผ่านช่องสัญญาณ ส่วนการเข้ารหัสแหล่งกำเนิดสัญญาณ (Source Coding) มุ่งเน้นในการบีบอัดข้อมูล (Compression) และการเข้ารหัสลับ (Encryption)

สำหรับการสื่อสาร การบีบอัดข้อมูลเป็นสิ่งจำเป็นเนื่องจากช่องสัญญาณสื่อสารมีอยู่อย่างจำกัด โดยเฉพาะอย่างยิ่งสัญญาณวิดีโอซึ่งมีจำนวนของข้อมูลขนาดใหญ่ โครงสร้างของตัวเข้ารหัสสัญญาณวิดีโอได้แสดงดังรูปที่ 2.3



รูปที่ 2.3 บล็อกไดอะแกรมของการเข้ารหัสสัญญาณวิดีโอแบบที่มีการบีบอัด

การเข้ารหัสสัญญาณวิดีโอแบบบีบอัดข้อมูลจะแบ่งได้ 2 ประเภทตามชนิดข้อมูลดังนี้

2.2.1 การเข้ารหัสภายในเฟรม (Intra-Frame Coding)

ในการเข้ารหัสชนิดนี้จะลดความซ้ำซ้อนทางพื้นที่ (Spatial Redundancy) ของเฟรมภาพ ด้วยการบีบอัดโดยอาศัยข้อมูลภายใน 1 เฟรม โดยใช้เทคนิคต่างๆ

1. การสุ่มข้อมูล (Subsampling) คือ การเข้ารหัสเพียงบางจุดภาพ หรือการเข้ารหัสค่าเฉลี่ยของกลุ่มจุดภาพแทนที่จะเข้ารหัสทุกๆ จุดภาพ เพื่อช่วยลดอัตราบิตในการส่งได้ แต่ความคมของภาพจะลดลง สำหรับจุดภาพที่ไม่ถูกเข้ารหัสก็จะถูกสร้างขึ้นที่ภาครับโดยการประมาณค่ากลาง (Interpolation) ซึ่งมักจะทำให้ภาพมัว

2. การควอนไทซ์แบบหยาบ (Coarse Quantization) เป็นการควอนไทซ์โดยใช้จำนวนบิตต่ำลง โดยจะทำให้มีข้อมูลบางส่วนของภาพหายไป เช่น ภาพขาวดำ ส่วนที่เป็นสีเทาจะถูกควอนไทซ์ด้วยวิธีนี้ เนื่องจากสายตาคคนมักจะไม่เห็นความแตกต่างที่เกิดขึ้น

3. การเข้ารหัสข้อมูลโดยการแปลง (Transform Coding) เป็นการแปลงข้อมูลของภาพจากโดเมนพื้นที่ (Spatial Domain) หนึ่งไปยังโดเมนอื่น เช่น โดเมนความถี่ (Frequency Domain) เพื่อลดข้อมูล เนื่องจากในโดเมนอื่นแสดงลักษณะของภาพที่ต่างออกไป ข้อมูลบางส่วนในโดเมนนั้นอาจจะเป็นส่วนที่ลดได้ ที่นิยมใช้กันคือ การแปลงโคไซน์แบบไม่ต่อเนื่อง (Discrete Cosine

Transform) หลังจากนั้นตัดข้อมูลส่วนความถี่สูงที่มีผลน้อยต่อสายตามนุษย์ออก การแปลงอีกชนิดซึ่งได้รับความนิยมในปัจจุบันเพิ่มขึ้น คือ การแปลงเวฟเลต (Wavelet Transform)

4. การทำเวกเตอร์ควอนไทเซชัน (Vector Quantization) โดยข้อมูลภาพจะถูกแบ่งกลุ่มและแสดงแทนด้วยสัญลักษณ์จากหนังสือรหัส (Codebook) เพื่อทำการส่งเฉพาะสัญลักษณ์แทนเท่านั้นซึ่งช่วยให้ประหยัดบิตในการส่งได้

5. การเข้าเอนโทรปี เป็นการเข้ารหัสข้อมูลแบบไม่มีการสูญเสียข้อมูล โดยมาตรฐาน MPEG จะใช้ชนิดเดียวกันกับการเข้ารหัสภาพนิ่งตามมาตรฐาน JPEG คือ การเข้ารหัสฮัฟแมน (Huffman Coding)

2.2.2 การเข้ารหัสระหว่างเฟรม (Inter-Frame Coding)

เนื่องจากในสัญญาณวิดีโอมีคุณลักษณะว่าเฟรมภาพที่อยู่ติดกัน จะมีความซ้ำซ้อนของข้อมูลภาพ เพราะมีสหสัมพันธ์ทางเวลาเชิงพื้นที่ (spatial-temporal correlation) ดังนั้น การเข้ารหัสระหว่างเฟรม จึงสามารถพิจารณาว่าข้อมูลที่ซ้ำซ้อนกันเหล่านี้เป็นส่วนเกิน และหากรรมวิธีในการลดข้อมูลเหล่านี้ตามคุณสมบัติดังกล่าว โดยข้อมูลที่ทำการลดเป็นข้อมูลส่วนเกินทางเวลาเชิงพื้นที่ (spatial temporal redundancy) เครื่องมือที่ใช้ในการเข้ารหัสมีหลายเทคนิค เช่น

1. การเข้ารหัสแบบสุ่มทางเวลา (Time Subsampling) เป็นการเข้ารหัสไม่ครบทุกเฟรมภาพ แล้วใช้ตัวถอดรหัสช่วยสร้างเฟรมที่หายไปขึ้นมา

2. การเข้ารหัสค่าความแตกต่าง (Different Coding) เป็นวิธีการเข้ารหัสค่าความแตกต่างระหว่างเฟรมติดกัน แทนที่จะเข้ารหัสเฟรมภาพทั้งภาพ เพราะค่าความแตกต่างระหว่างเฟรมจะมีข้อมูลน้อยกว่าภาพทั้งเฟรม เนื่องจากมีความคล้ายกันมากแต่อาจต้องใช้ โอเวอร์เฮด (overhead) ในการบ่งบอกว่า จุดภาพใดที่มีการเปลี่ยนแปลงและจุดใดไม่มีการเปลี่ยนแปลง

3. การเข้ารหัสค่าความแตกต่างบนพื้นฐานของบล็อก (Block-Based Different Coding) เป็นการเข้ารหัสที่คล้ายคลึงกับการเข้ารหัสค่าความแตกต่างระหว่างเฟรม แต่ภาพจะถูกแบ่งเป็นบล็อกก่อน หลังจากนั้นแต่ละบล็อกจะถูกเปรียบเทียบกันแทนการเปรียบเทียบเป็นจุดภาพต่อจุดภาพ

การประมาณการเคลื่อนที่ (Motion Estimation) และการชดเชยการเคลื่อนที่ (Motion Compensation) เป็นเครื่องมือที่ใช้ลดข้อมูลในการส่งสัญญาณภาพวิดีโอ ด้วยการส่งข้อมูลที่จำเป็นในการสร้างภาพไปเท่านั้น และส่วนที่ต้องส่งไป คือ ภาพอ้างอิง หรือ เฟรมอ้างอิง (Reference Frame) เวกเตอร์การเคลื่อนที่ (Motion Vector) และภาพข้อมูลผิดพลาดที่ได้รับการเข้ารหัสภายในเฟรม (Residue with Intra-Frame encoding) ทั้งนี้ข้อมูลภาพแสดงความผิดพลาดน้อยเท่าใด การเข้ารหัสในเฟรมจะบีบอัดได้มากขึ้นเท่านั้น

2.3 มาตรฐาน MPEG-4

มาตรฐาน MPEG-4 เป็นมาตรฐานของ ISO (International Standard Organization) สำหรับการเข้ารหัสสัญญาณเสียง ภาพและวิดีโอ ที่อัตราบิตต่ำ โดยมุ่งเน้นการประยุกต์ใช้งานทางด้านการเก็บข้อมูล การถ่ายทอดกระจายข่าวสาร และการสื่อสารปฏิสัมพันธ์ โดยผ่านทางโครงข่าย IP, ATM, Mobile, PSTN หรือ Narrow band ISDN มาตรฐาน MPEG-4 แบ่งเป็นส่วนหลักๆ ดังนี้ คือ DMIF (Delivery Multimedia Integration Framework) ส่วนของระบบ (System) ส่วนของการเข้ารหัสสัญญาณเสียง (Audio) และส่วนของการเข้ารหัสภาพและสัญญาณวิดีโอ (Visual) ส่วนการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 จะแบ่งลักษณะของสัญญาณได้ 2 ลักษณะ คือ แบบธรรมชาติ (Natural) และแบบสังเคราะห์ (Synthetic)

1. สัญญาณวิดีโอแบบธรรมชาติ (Natural Video)

จุดมุ่งหมายหลักของการเข้ารหัสภาพวัตถุธรรมชาติ (Natural Visual Object) คือ ให้สามารถจัดเก็บ จัดการและขนส่งข้อมูลได้อย่างมีประสิทธิภาพ เพื่อนำไปใช้ในงานด้านมัลติมีเดียได้อย่างเหมาะสม โดยจะต้องมีคุณสมบัติดังนี้

- สามารถบีบอัดภาพและสัญญาณวิดีโอได้อย่างมีประสิทธิภาพ
- สามารถเข้าถึงแต่ละ Visual Object ได้ทุกเวลา (random access)
- สามารถปรับปรุงจัดการกับภาพและสัญญาณวิดีโอให้ดีขึ้นและเหมาะสม
- มีความสามารถในการเข้ารหัสภาพนิ่งและสัญญาณวิดีโอบนพื้นฐานของคอนเทนต์
- มีความสามารถด้าน Content Based Scalability สำหรับทั้งภาพนิ่งและสัญญาณวิดีโอ
- มีความสามารถทางด้าน Spatial, Temporal และ Quality Scalability
- มีความทนทานต่อความผิดพลาดและสามารถแก้ไขความผิดพลาดได้

2. การเข้ารหัสสัญญาณแบบวัตถุสังเคราะห์ (Synthetic Object)

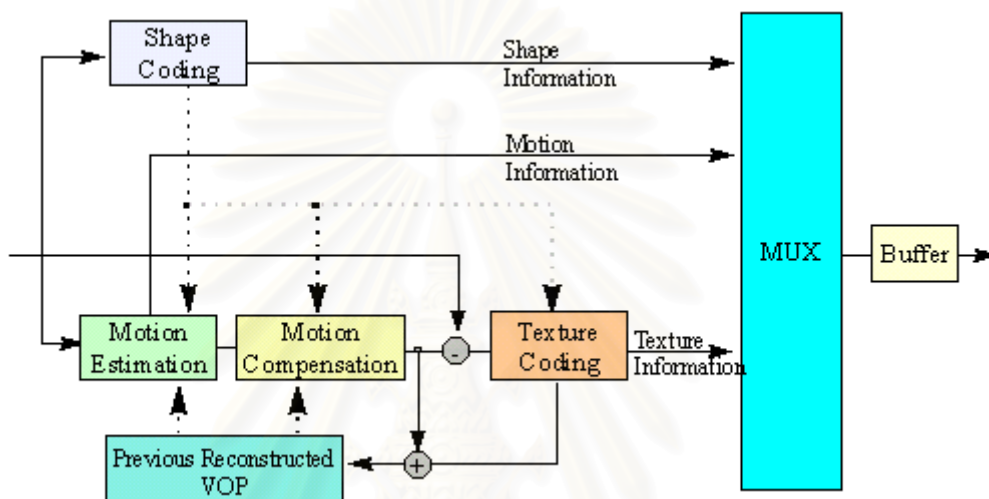
วัตถุสังเคราะห์ (Synthetic Object) จะประกอบขึ้นจากแบบจำลองและพารามิเตอร์ที่ได้การคำนวณของคอมพิวเตอร์มีด้วยกันหลายระดับ ซึ่งอาจแบ่งเป็นกลุ่มได้ดังนี้

- การอธิบายลักษณะเชิงพารามิเตอร์ (Parametric Description) สำหรับการอธิบายเชิงสังเคราะห์ (Synthetic Description) ของส่วนตัวและหน้าของแบบจำลองภาพคน
 - Static/Dynamic Mesh coding with texture mapping
 - Texture coding for view dependent

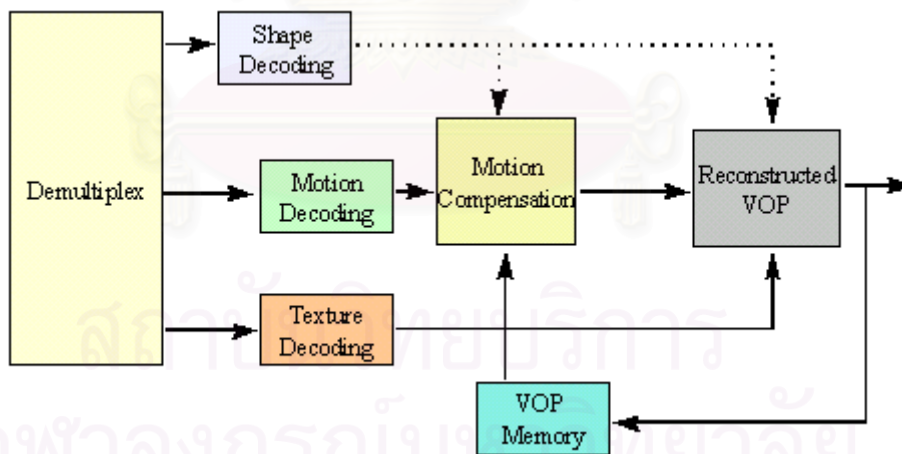
วิทยานิพนธ์ฉบับนี้จะเป็นให้ความสำคัญกับระบบการเข้ารหัสสัญญาณวิดีโอที่ใช้สำหรับสัญญาณวิดีโอแบบธรรมชาติ ตามข้อกำหนดของมาตรฐาน MPEG-4

โครงสร้างของการเข้ารหัสสัญญาณวิดีโอของ MPEG-4

การเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 ก่อนเข้ารหัสตามแบบ Shape Coding, Motion Coding และ Texture Coding จะมีกระบวนการแยกส่วนของภาพวัตถุเป็นวัตถุที่สนใจก่อน



รูปที่ 2.4 ภาพแสดงโครงสร้างพื้นฐานตัวเข้ารหัสสัญญาณวิดีโอของ MPEG-4



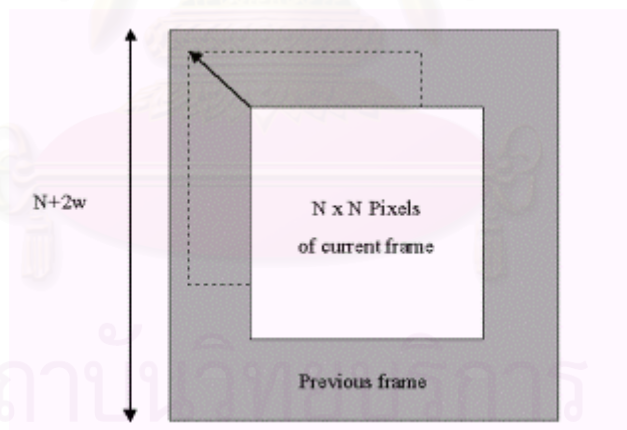
รูปที่ 2.5 ภาพแสดงโครงสร้างพื้นฐานตัวถอดรหัสสัญญาณวิดีโอของ MPEG-4

เทคนิคการชดเชยการเคลื่อนที่ (Global Motion Compensation)

ทำโดยการใช่วิธี Sprite ซึ่งทำเป็นภาพนิ่ง ในลักษณะของ Panorama Background ทำให้มีการส่งข้อมูลของฉากพื้นหลังเพียงครั้งเดียว หลังจากนั้นในแต่ละเฟรมที่มีการใช้พื้นที่ของฉากหลังคนละส่วนกันก็จะส่งเฉพาะข้อมูลที่บอกถึงการเคลื่อนที่ไปของมุมมอง (Viewing Point of Camera) แล้วทำการสร้างในส่วนของฉากพื้นหลังนั้นใหม่โดยใช้ข้อมูลที่ส่งมาในครั้งแรกสุด โดยไม่ต้องส่งข้อมูลมาใหม่ทั้งหมด ทำให้สามารถลดข้อมูลที่จะต้องส่งลงได้มาก

การประมาณการเคลื่อนที่ (Motion Estimation)

สำหรับสัญญาณวิดีโอที่มีการเคลื่อนที่ของวัตถุในลำดับภาพ เพื่อให้หาการเคลื่อนที่ของวัตถุนั้น จะต้องมีการประมาณการเคลื่อนที่ ซึ่งมีอยู่หลายกรรมวิธีด้วยกัน เช่น Full Search Algorithm (FSA), Three Step Search Algorithm (TSSA), Logarithm Search Algorithm (LSA) หรือ Hierarchical Search Algorithm (HSA) ซึ่งแต่ละกรรมวิธีการจับคู่บล็อก (block matching) นั้นจะมีคุณสมบัติที่แตกต่างกัน ทั้งด้านความถูกต้องในการประมาณการเคลื่อนที่ ความซับซ้อนและการคำนวณที่ต้องใช้สำหรับแต่ละกรรมวิธี โดยมีค่าความผิดพลาดที่น้อยที่สุดของการจับคู่บล็อกเป็นค่าที่ใช้ตัดสินใจ



รูปที่ 2.6 รูปแบบในการประมาณการเคลื่อนที่แบบการจับคู่บล็อก

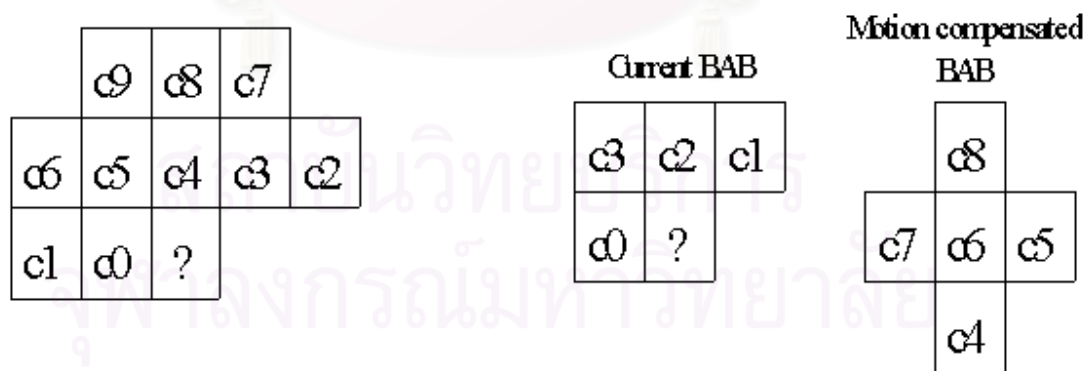
การเข้ารหัสรูปทรง (Shape Coding)

ส่วนนี้เป็นส่วนสำคัญในการแสดงรูปทรงของวัตถุ เทคนิคที่ใช้ประกอบไปด้วย quadtree, chain code และ polygonal approximation ซึ่งจะแทนรูปทรงของวัตถุด้วยมาสก์ไบนารี [19] การเข้ารหัสรูปทรงของวัตถุเป็นปัญหาสำคัญในการเข้ารหัสบนพื้นฐานของหน่วยวัตถุ อัตราบิตที่ใช้ในการเข้ารหัสรูปทรงขึ้นกับจำนวนของวัตถุ และยังขึ้นกับความซับซ้อนของขอบรูป

ทรง ซึ่งมีรายละเอียดค่อนข้างมาก ซึ่งสามารถจะลดความซับซ้อนได้โดยใช้การเข้ารหัสรูปทรง ประกอบกับข้อมูลรอบข้าง (Content Based Shape Coding) การเข้ารหัสรูปทรงจะแทนขอบเขตของวัตถุด้วยรูปหลายเหลี่ยม (polygons) ซึ่งจะประมาณรูปทรงด้วยเส้นขอบที่มีจำนวนมุมน้อยที่สุด โดยยังคงรักษาลักษณะของรูปทรงเดิมอยู่ สำหรับวิธีการ Content Based Shape Coding จะมีความผิดพลาดที่น้อยกว่ารูปทรงที่ได้โดยที่สามารถครอบคลุมจุดภาพทั้งหมดในบริเวณที่สนใจได้ ขณะที่จะยอมให้ความผิดพลาดเข้าไปในรูปทรงน้อยที่สุดเพื่อไม่ให้เกิดจุดภาพที่มีความกำกวมขึ้น ในการแยกส่วนของภาพวัตถุ [9]

การเข้ารหัสรูปทรง (Shape Coding) ที่ใช้กันมีดังนี้

- Binary Alpha Blocks (BAB)
 - CAE (Content-based Arithmetic Encoding)
 - Gray-scale Alpha Blocks
 - การเข้ารหัสชดเชยการเคลื่อนที่ DCT (Motion Compensated DCT Coding)
 - Alpha Block
 - วัตถุที่แยกแยะโดยใช้หลักการสร้างสี่เหลี่ยมล้อมรอบ
 - Intra and Inter Shape Coding
 - การตัดสินใจระดับมาโครบล็อก (Macroblock-level Decision)
 - รูปทรงชดเชยการเคลื่อนที่ (Motion Compensated Shape)
 - การเข้ารหัสรูปทรงแบบไม่มีการสูญเสียหรือมีการสูญเสียข้อมูลได้
- การเข้ารหัสรูปทรงที่ใช้สำหรับ MPEG-4 จะใช้แบบ CAE ดังแสดงในรูปที่ 2.7



(a) INTRA template

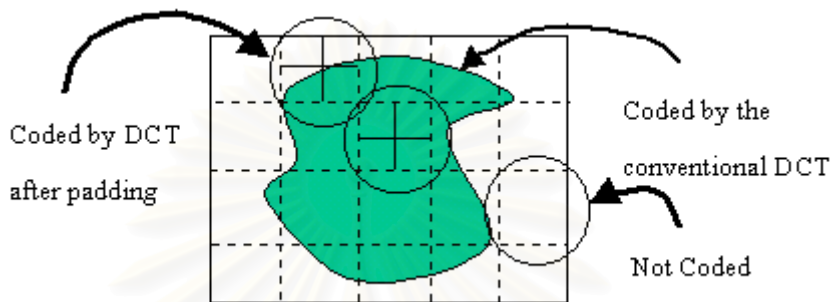
(b) INTER template

รูปที่ 2.7 การเข้ารหัสรูปทรงโดย Content-based Arithmetic Encoding (CAE)

การเข้ารหัสพื้นผิว (Texture Coding)

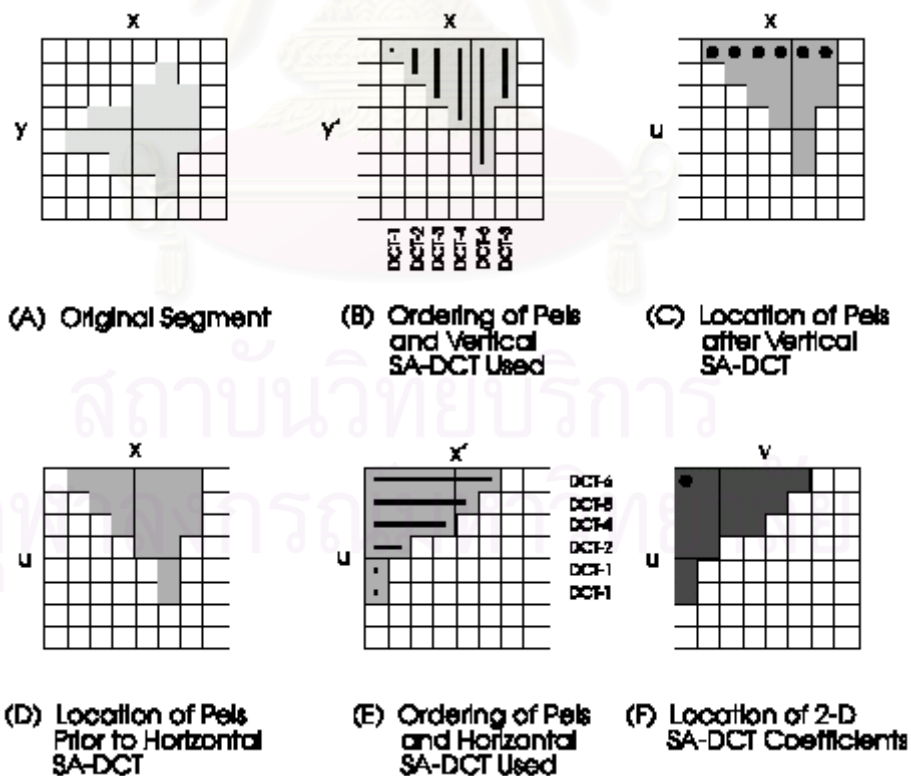
ส่วนของข้อมูลสี (color information) หรือระดับความเข้ม จะมีขั้นตอนการเข้ารหัสแบ่งได้เป็น 3 ขั้นตอน คือ ส่วนแสดงข้อมูลด้วยสัมประสิทธิ์การแปลง ส่วนการควอนไทซ์สัมประสิทธิ์ของการแปลง และส่วนเข้ารหัสเอนโทรปีสำหรับค่าสัมประสิทธิ์จากการควอนไทซ์

สำหรับส่วนของการแปลงมีหลายแบบด้วยกัน มาตรฐานส่วนใหญ่นิยมใช้การแปลงแบบ DCT กับบล็อกขนาด NxN ซึ่ง DCT จัดเป็นการแปลงรองจากออปติมัล (suboptimal)



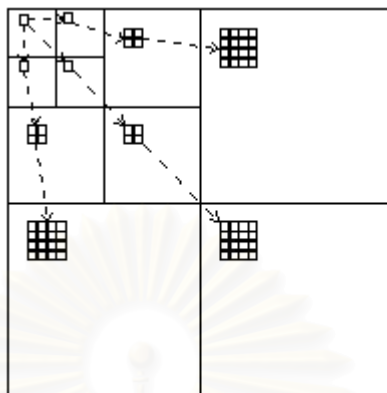
รูปที่ 2.8 ส่วนต่างๆ ที่เข้ารหัสบนพื้นฐานของหน่วยวัตถุ

การเข้ารหัสพื้นผิวของรูปทรงใดๆ จะใช้กรรมวิธี Shape Adaptive DCT (SA-DCT) แสดงดังรูปที่ 2.9



รูปที่ 2.9 กรรมวิธี Shape Adaptive DCT (SA-DCT)

ส่วนการแปลงเวฟเลต (wavelet transform) คือ วิธีการเข้ารหัสแบบออปติมัล (optimal) [Koenen R., 1998] สำหรับเข้ารหัสวัตถุรูปทรงต่างๆ โดยข้อมูลจะถูกเข้ารหัสข้อมูลโดยใช้กรรมวิธี Zerotree ซึ่งใช้กับการเข้ารหัสภาพนิ่ง



รูปที่ 2.10 รูปแบบการเข้ารหัสโดยกรรมวิธี Zerotree

สำหรับขั้นตอนในการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุนั้นมีหลายขั้นตอนด้วยกัน แต่แต่ละขั้นตอนได้มีการทำวิจัยทั้งในส่วนของ การเข้ารหัสรูปทรง (shape coding) การเข้ารหัสพื้นผิว (texture coding) การวิเคราะห์การเคลื่อนที่ (motion analysis) และการแยกส่วนของวัตถุภาพ (object segmentation)

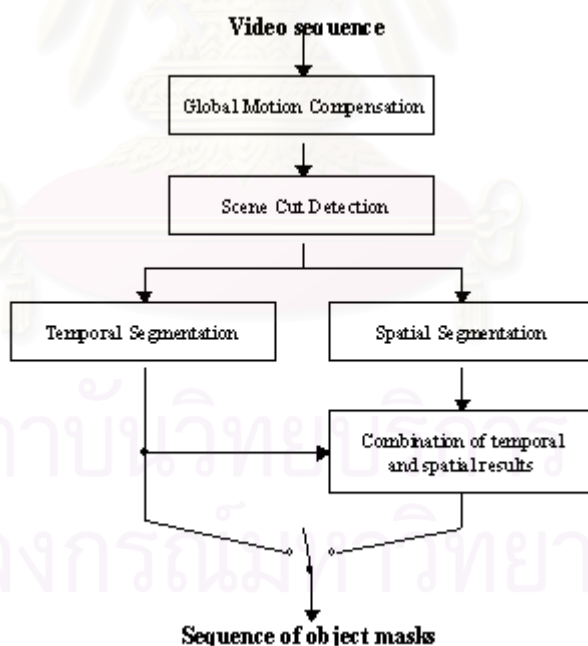
สำหรับงานแต่ละส่วนนั้นมีความสำคัญแตกต่างกันไป แต่สำหรับมาตรฐาน MPEG-4 และการเข้ารหัสสัญญาณบนพื้นฐานของหน่วยวัตถุ จะคำนึงถึงการแยกส่วนของภาพวัตถุที่มีคุณลักษณะแตกต่างกัน และรูปทรงของวัตถุแต่ละวัตถุที่แยกออกมา เพื่อให้สามารถนำเสนอภาพวัตถุ (object representation) ที่มีความหมายและมีคุณสมบัติเป็นภาพวัตถุที่มีความหมาย (semantic meaningful object)

2.4 การแยกส่วนสำหรับการสร้างระนาบวัตถุของสัญญาณวิดีโอตามมาตรฐาน MPEG-4 (Video Object Plane : VOP)

แบบแผนของการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน ISO/IEC 14496-2 ที่รองรับฟังก์ชันบนพื้นฐานของคอนเทนต์ (Content based) ต้องการที่จะอธิบายภาพต่างๆ ในรูปของภาพวัตถุเคลื่อนที่ (video objects) การแยกภาพวัตถุของสัญญาณวิดีโอ ยังต้องให้การเชื่อมต่อกับผู้ใช้ในบริการมัลติมีเดียหลายอย่างเพื่อให้มีความยืดหยุ่นในการเข้าถึงข้อมูล และง่ายต่อการจัดการข้อมูลสัญญาณวิดีโอ ในส่วนการแยกภาพวัตถุนี้จะเป็นกระบวนการขั้นต้น (preprocessing) มีจุดมุ่งหมายที่จะกำหนดภาพวัตถุที่ปรากฏในลำดับภาพโดยอัตโนมัติ และเป็นหัวใจในการเข้ารหัสตามมาตรฐาน MPEG-4 อย่างมีประสิทธิภาพ

การแยกส่วน (segmentation) มีเพื่อแบ่งส่วนภาพหรือสัญญาณวิดีโอในรูปแบบของบริเวณ สำหรับสัญญาณวิดีโอ การแยกส่วนจะแบ่งบริเวณตามความต่อเนื่องของการเคลื่อนที่ เพื่อสร้างมาสก์ของวัตถุ (object masks) ที่ใช้แทนภาพวัตถุเคลื่อนที่ (video object) มีการนำเสนอกรรมวิธีการแยกส่วนภาพวัตถุของสัญญาณวิดีโอที่แตกต่างกันจำนวนมากโดยอยู่บนพื้นฐานของความคล้ายคลึงกันของข้อมูลความเข้มสี (spatial homogeneity) ความต่อเนื่องของการเคลื่อนที่ (motion coherence) หรือกระบวนการรวมกันของข้อมูลทั้งสองส่วน ซึ่งคาดว่าจะสามารถแบ่งกลุ่มของวัตถุที่เคลื่อนที่ ตามลักษณะความคล้ายคลึงกันของความเข้มและความต่อเนื่องของการเคลื่อนที่

กรรมวิธีการแยกส่วนในโครงร่างมาตรฐาน MPEG-4 นี้ มุ่งไปที่การรวมกันที่เหมาะสมของข้อมูลเชิงเวลา (temporal information) และข้อมูลเชิงพื้นที่ (spatial information) มาตรฐาน ISO/IEC 14496 เวอร์ชัน 1 กำหนดเพียงจุดมุ่งหมายและเทคนิคในการแยกส่วนภาพวัตถุจากลำดับภาพ แต่ยังไม่เป็นมาตรฐาน การแบ่งกลุ่มของจุดภาพ (pels) ในลำดับภาพจะแบ่งเป็น 2 กลุ่ม คือ ส่วนวัตถุที่เคลื่อนที่ (moving objects) หรือพื้นหน้า (foreground) และส่วนที่อยู่นิ่ง หรือพื้นหลัง (background) การวิจัยในส่วนนี้ยังดำเนินต่อไปในมาตรฐาน ISO/IEC 14496 เวอร์ชัน 2 โครงสร้างกรรมวิธีการแยกส่วนภาพวัตถุแสดงดังรูปที่ 2.11



รูปที่ 2.11 โครงสร้างของการรวมข้อมูลสำหรับการแยกส่วนภาพวัตถุ

กรรมวิธีการรวมข้อมูลระหว่างเฟรมและข้อมูลภายในเฟรมสำหรับการแยกส่วนภาพวัตถุ

ขั้นตอนแรก ทำการประมาณการเคลื่อนที่ของกล้องจับภาพและทำการชดเชยการเคลื่อนที่โดยรวม (global motion compensation) จากนั้นเป็นการตรวจจับการตัดฉาก (scene cut detection) ซึ่งเป็นกระบวนการขั้นต้นเพื่อกำจัดผลจากการเคลื่อนไหวของกล้องจับภาพ

ขั้นตอนที่สอง ทำการแยกส่วนด้วยข้อมูลเชิงเวลา (temporal) หรือข้อมูลของเวลาเชิงพื้นที่ (spatio-temporal) ของแต่ละภาพขึ้นกับความต้องการในการใช้งาน เนื่องจากปกติการแยกส่วนที่ใช้เฉพาะข้อมูลเชิงเวลาจะมีความซับซ้อนในการคำนวณน้อยกว่าข้อมูลของเวลาเชิงพื้นที่ อย่างไรก็ตามการแยกส่วนด้วยข้อมูลเชิงพื้นที่จะให้ความถูกต้องในการแยกส่วนมากกว่า แต่จะเพิ่มความซับซ้อนในการคำนวณ

สำหรับการแยกส่วนภาพด้วยข้อมูลเชิงเวลา จะมี 2 กรรมวิธีที่อยู่ภายใต้การพิจารณาของ MPEG-4 ซึ่งมาตรฐาน MPEG-4 ในเวอร์ชัน 2 ได้มีการวิจัยเพื่อตัดสินว่ากรรมวิธีใดใช้ได้ดี ส่วนในการแยกส่วนภาพด้วยข้อมูลเชิงพื้นที่มีเพียงกรรมวิธีเดียวที่ถูกพิจารณาอยู่ ขั้นตอนสุดท้ายจะทำการนำผลจากทั้งสองส่วนมารวมกัน

การแยกส่วนเชิงเวลา (temporal segmentation) บนพื้นฐานการตรวจจับการเปลี่ยนแปลง (change detection)

การแยกส่วนภาพวิธีนี้ สามารถแบ่งได้เป็น 2 ขั้นตอนหลัก โดยสมมุติว่าการเคลื่อนไหวของกล้องจับภาพได้รับการชดเชยแล้ว

ขั้นตอนแรก ส่วนหน้ากากการเปลี่ยนแปลงระหว่างเฟรมที่ตามต่อกันมาจะถูกตรวจรู้ตำแหน่งของจุดภาพที่มีการเปลี่ยนแปลงค่าความเข้มเนื่องจากการเคลื่อนที่ของวัตถุจะถูกกำหนดเป็นจุดที่เปลี่ยนแปลง เริ่มแรกส่วนหน้ากากจะสร้างจากการตั้งค่าจุดเปลี่ยนโดยรวม (global thresholding) ของความแตกต่างระหว่างเฟรม หลังจากนั้นขอบเขตของภาพที่เปลี่ยนแปลงจะถูกทำให้ราบเรียบโดยเทคนิครีแลกเซชัน (relaxation technique) โดยใช้ค่าจุดเปลี่ยนที่ปรับค่าได้ในบริเวณท้องถิ่น (local adaptive thresholds) นอกจากนั้นวิธีการตรวจรู้การเปลี่ยนแปลงที่มีการปรับค่าจุดเปลี่ยนได้และการใช้หน่วยความจำสำหรับเก็บหน้ากากของภาพวัตถุในเฟรมก่อนหน้า จะถูกใช้เพื่อให้ได้ขอบเขตของภาพวัตถุที่ชัดเจนขึ้น สุดท้ายหน้ากากของภาพวัตถุ (object mask) จะถูกทำให้ราบเรียบและบริเวณที่มีขนาดเล็กเกินไปจะถูกกำจัด

ขั้นตอนที่สอง หน้ากากของภาพวัตถุจะถูกสร้างมาจากการกำจัดบริเวณพื้นหลังที่ไม่ถูกครอบคลุม (uncovered background) ดังนั้นข้อมูลการเคลื่อนที่ของจุดภาพภายในบริเวณที่มีการเปลี่ยนแปลงจะถูกใช้ การเคลื่อนที่ของจุดภาพนี้จะถูกประมาณโดยการจับคู่บล็อกแบบลำดับ

ขั้น (hierarchical block matching) ส่วนหน้ากากของการเปลี่ยนแปลง (change detection mask) ในขั้นตอนแรกจะถูกพิจารณาด้วยเวกเตอร์การเคลื่อนจากการจับคู่บล็อก ถ้าทั้งจุดแรกและจุดสุดท้ายของเวกเตอร์การเคลื่อนที่อยู่ภายในบริเวณหน้ากากของการเปลี่ยนแปลง จุดภาพนั้นจะเป็นจุดภาพพื้นหน้า ส่วนถ้ามีจุดหนึ่งไม่อยู่ในบริเวณที่เปลี่ยนแปลง จุดภาพนั้นจะเป็นจุดภาพพื้นหลัง

การแยกส่วนเชิงเวลาโดยใช้โมเมนต์ลำดับสูงและการติดตามการเคลื่อนที่ (Temporal segmentation using higher order moments and motion tracking)

กรรมวิธีนี้จะแยกส่วนภาพของเฟรมปัจจุบัน (f_k) โดยการประมวลผลกลุ่มของเฟรมก่อนหน้า (f_{k-i} , $i=0, \dots, n$) จำนวนของเฟรมขึ้นกับพื้นฐานของความเร็ววัตถุ สำหรับแต่ละเฟรมจะแบ่งขั้นตอนเป็น 3 ขั้นตอน

ขั้นแรก ความแตกต่างของแต่ละเฟรมจะเทียบกับเฟรมแรกของกลุ่ม เพื่อจะตรวจจับบริเวณที่เปลี่ยนแปลงขึ้นกับ การเคลื่อนที่ของวัตถุ บริเวณพื้นหลังที่ไม่ถูกครอบคลุม (uncovered background) และสัญญาณรบกวน (noise) สามารถตรวจรู้ได้จากสมการที่ 2.1

$$d_{k,j}(x,y) = f_{k,j}(x,y) - f_{k,n}(x,y), j = 0, \dots, n-1 \quad (2.1)$$

ขั้นที่สอง เพื่อที่จะกำจัดความแปรปรวนของความเข้มเนื่องจากสัญญาณรบกวน จะใช้การทดสอบทางสถิติอันดับสูงขึ้น (higher order statistic test : HOS) สำหรับแต่ละจุดภาพ โดยโมเมนต์ลำดับที่ 4 ของความแตกต่างระหว่างเฟรมจะถูกคำนวณบนวินโดว์ขนาด 3×3 จุดภาพ ซึ่งจะเปรียบเทียบกับเซตของค่าจุดเปลี่ยนที่ปรับค่าได้ และค่าโมเมนต์จะเป็นศูนย์เมื่อมีค่าต่ำกว่าค่าจุดเปลี่ยน ขั้นตอนนี้ทำเพื่อแยกบริเวณที่เปลี่ยนแปลงระหว่างพื้นหลังที่ไม่ถูกครอบคลุม และวัตถุที่เคลื่อนที่ สำหรับแต่ละจุดภาพ การเคลื่อนที่จะถูกประเมินบนวินโดว์ขนาด 3×3 จุดภาพ ถ้าการเคลื่อนที่ไม่เป็นศูนย์ จะให้จุดภาพนั้นเป็นบริเวณที่เคลื่อนที่ ขั้นตอนนี้จะทำซ้ำจนกระทั่ง $j = n-2$ ส่วนจุดภาพที่แสดงการเคลื่อนที่เป็นศูนย์จะเป็นบริเวณอยู่นิ่ง

ขั้นสุดท้าย คือ กรรมวิธีตรวจสอบ (regularization) จะใช้พิกเซลบริเวณที่อยู่นิ่ง และบริเวณที่เคลื่อนที่อีกครั้ง ส่วนของสัญญาณรบกวนและโฮล (hole) จะใช้วงจรรองเชิงสัญญาณเพื่อสร้างผลของภาพวัตถุให้ราบเรียบขึ้น

การแยกส่วนเชิงพื้นที่ (spatial segmentation) บนพื้นฐานของกรรมวิธีสันปันน้ำ (watershed algorithm)

ขั้นตอนแรก ภาพจะถูกจัดการความเข้มสีให้ง่ายต่อการแยกส่วนภาพ โดยใช้วงจรรองเชิงสัญญาณ (morphological filters) วงจรรองนี้จะกำจัดบริเวณที่มีขนาดเล็กกว่าขนาดที่กำหนด แต่จะรักษาขอบเขตของภาพวัตถุเอาไว้

ขั้นตอนที่สอง การตรวจจับความไม่ต่อเนื่องของความเข้มสีภายในเฟรมโดยใช้ตัวปฏิบัติการเกรเดียนต์เชิงสัณฐาน (morphological gradient operator)

ขั้นต่อไป ภาพที่ได้จะนำไปผ่านกรรมวิธีสันปันน้ำ (watershed algorithm) เพื่อแยกส่วนภาพเป็นบริเวณที่มีความเข้มสีคล้ายคลึงกัน สำหรับปัญหาความกำกวมของขอบเขตจากการคำนวณเกรเดียนต์ภาพในเฟรม จะใช้ข้อมูลของสีในการคำนวณเกรเดียนต์ โดยเลือกค่าที่มากที่สุดของค่าเกรเดียนต์ในองค์ประกอบสี Y, Cr และ Cb ที่มีค่าน้ำหนักเป็น a, b และ c ตามลำดับ

ขั้นสุดท้าย กระบวนการตัดสินขอบเขตจะใช้ตัดสินใจบริเวณที่ยังไม่ถูกกำหนดเข้าไปกับบริเวณที่มีลักษณะเหมือนกันที่สุด โดยพิจารณาตามเกณฑ์ของการแยกส่วนภาพ เช่น ความแตกต่างของค่าความเข้ม จากนั้นจะใช้กรรมวิธีการรวมตัวของบริเวณ (region merging) สุดท้ายการแยกส่วนภาพเชิงพื้นที่จะให้บริเวณที่มีขอบเขตชัดเจน สามารถนำไปใช้ในการแยกส่วนร่วมกับข้อมูลเชิงเวลาเพื่อกำหนดตำแหน่งขอบเขตของวัตถุเคลื่อนที่ ดังนั้นวัตถุเคลื่อนที่จะสามารถแสดงเป็นบริเวณที่มีความหมาย (semantically meaningful regions) และมีขอบเขตชัดเจนยิ่งขึ้น

2.5 ความรู้พื้นฐานเกี่ยวกับการประมาณการเคลื่อนที่ (Motion Estimation)

โดยทั่วไป ส่วนใหญ่การประมาณการเคลื่อนที่ (Motion Estimation) จะทำควบคู่กับการชดเชยการเคลื่อนที่ (Motion Compensation) โดยการประมาณการเคลื่อนที่ จะทำงานด้านส่งสัญญาณเพื่อทำการเข้ารหัส จากชุดภาพออกมาเป็นเวกเตอร์การเคลื่อนที่ และการชดเชยการเคลื่อนที่ จะทำงานด้านรับสัญญาณเพื่อทำการถอดรหัสโดยการสร้างภาพประมาณ (Estimated Picture) จากภาพอ้างอิง (Reference Picture) และเวกเตอร์การเคลื่อนที่ แต่ในส่วนของด้านส่งก็จะมี การชดเชยการเคลื่อนที่ได้เช่นเดียวกัน เพื่อช่วยในการหาค่าความเพี้ยนของภาพที่ประมาณ

เทคนิคการประมาณการเคลื่อนที่ สามารถแบ่งวิธีหลักได้ 3 วิธี ดังนี้ [Rao K. R., Hwang J. J., 1996]

1. วิธีสมการการไหลของแสง (Optical Flow Equation Method)
2. วิธีเพลรีเคอร์ซีฟ (Pel Recursive Method)
3. วิธีการเข้าคู่บล็อก (Block Matching Method)

2.5.1 วิธีกรไหลของแสง (Optical Flow Method)

สมการการไหลของแสงตั้งสมมติฐานให้ ความเข้มของจุดภาพในภาพ จะมีการเปลี่ยนไปตามการเคลื่อนที่ โดยให้ $S_c(x_1, x_2, t)$ เป็นการกระจายความเข้มทางด้านพื้นที่และเวลา โดยค่านี้จะต้องสอดคล้อง ตามสมการการไหลของแสง (2.2) ดังนี้

$$\frac{\partial Sc(X;t)}{\partial X1} v1(X;t) + \frac{\partial Sc(X;t)}{\partial X2} v2(X;t) + \frac{\partial Sc(X;t)}{\partial t} = 0 \quad (2.2)$$

กำหนดให้

$$v1(X;t) = \frac{dX1}{dt} \text{ และ } v2(X;t) = \frac{dX2}{dt} \quad (2.3)$$

X เป็นค่าความเข้มแสงของภาพ ประกอบด้วย X1 และ X2

X1, X2 เป็นค่าความเข้มแสงของภาพ ทางแนวแกน X และ Y ตามลำดับ

วิธีนี้มีเป้าหมายในการหาค่า v1 และ v2 ซึ่งเป็นเวกเตอร์การเคลื่อนที่ (Motion Vector)

2.5.2 วิธีเพลรีเคอร์ซีฟ (Pel-Recursive Method)

ในขบวนการรีเคอร์ซีฟ (Recursion) จะปรับค่าเวกเตอร์การเคลื่อนที่ไปจนเข้าสู่ค่าที่ทำให้ความผิดพลาดต่ำใกล้เคียงศูนย์ โดยปรับค่าตามสมการ (2.4)

$$\hat{d}_a(X,t) = \hat{d}_{a-1}(X,t) + u_{a-1}(X,t) \quad (2.4)$$

กำหนดให้

$\hat{d}_a(X,t)$ เป็นค่าเวกเตอร์การเคลื่อนที่ในลำดับการคำนวณ (Iteration) ที่ a

$\hat{d}_{a-1}(X,t)$ เป็นค่าเวกเตอร์การเคลื่อนที่ในลำดับการคำนวณ (Iteration) ที่ a-1

$u_{a-1}(X,t)$ เป็นค่าปรับปรุงจากลำดับการคำนวณที่ a-1 ไปยัง a

การปรับค่าข้างต้นมุ่งเน้นในการหาค่าเวกเตอร์การเคลื่อนที่ที่ให้ค่าของผลต่างของเฟรม (Displaced Frame Different : DFD) ต่ำสุด

$$e(X, d_{a-1}) = Sc(X + d_{a-1}(x,t); t + \Delta t) - Sc(X,t) \quad (2.5)$$

เมื่อได้ค่าความเพี้ยนจากสมการดังกล่าวแล้วก็จะนำค่า ผลต่างเฟรมมาหาค่า $u_{a-1}(X,t)$ ดังนี้

$$u_{a-1}(X,t) = -\frac{\delta}{2} \nabla_d [e^2(X, d_{a-1})] \quad (2.6)$$

∇_d เป็น ตัว ปฏิบัติการเกรเดียนต์ (Gradient Operator)

δ เป็น ค่าช่วงก้าว ของแต่ละลำดับการคำนวณ ถ้าช่วงกว้างเกินไป จะลู่เร็วแต่ไม่ค่อยเสถียร และจะผิดพลาดสูง ถ้าช่วงสั้นจะลู่ช้าแต่จะเสถียรและมีค่าความผิดพลาดต่ำ ตามหลักการลดค่าตามเกรเดียนต์ (Gradient Descent)

2.5.3 กรรวิธีกรเข้าคู่บล็อก (Block Matching Algorithm: BMA)

วิธีกรนี้เป็นวิธีที่ได้รับความนิยมใช้กันมากในมาตรฐานทั่วไป ในการหาค่าเวกเตอร์การเคลื่อนที่ ไม่ว่าจะเป็ MPEG 1, MPEG 2 และ มีแนวโน้มใน MPEG 4 ด้วย หรือจะเป็น H.261/262/263/263+/263++ เนื่องด้วยวิธีกรนี้ใช้ความซับซ้อนในการคำนวณต่ำกว่าวิธีอื่น และ มีความสมดุลระหว่างความซับซ้อนในการคำนวณและคุณภาพการประมาณที่ดี

กรรวิธีกรเข้าคู่บล็อกอยู่บนสมมติฐานที่ว่า จุดภาพภายในบล็อกจะมีการเคลื่อนที่ในแนวระนาบเท่ากันจากเฟรมหนึ่งไปยังเฟรมถัดไป

กรรวิธีกรเข้าคู่บล็อกจึงเหมาะสมในการวิเคราะห์การเคลื่อนที่ของวัตถุในลำดับภาพ เนื่องจากใช้ทรัพยากรในการคำนวณที่ต่ำและมีคุณภาพการวิเคราะห์ที่ดีพอสมควร

2.5.3.1 การประมาณการเคลื่อนที่แบบบล็อก (Block Motion Estimation)

วิธีกรประมาณการเคลื่อนที่โดยการจับคู่บล็อก (Block Matching) เป็นวิธีที่แก้ปัญหาเรื่องความไวต่อสัญญาณรบกวน เนื่องจากจะพิจารณาการเปลี่ยนแปลงของความเข้มเป็นกลุ่มบล็อกแทนที่จะพิจารณาเป็นจุดภาพ

$$\sum_{x \in R} (I(x, t) - I(x + d(x), t + 1)) = 0 \quad (2.7)$$

โดยสมมุติว่าจุดภาพทุก ๆ จุดในบล็อกมีการเคลื่อนที่เหมือนกัน การประมาณการเคลื่อนที่แบบบล็อกจะพิจารณาความราบเรียบท้องถิ่น (local smoothness) ซึ่งจะให้ผลดีกว่าการพิจารณาความราบเรียบโดยรวม (global smoothness) ที่ใช้โดย Horn และ Schunck [...] การจับคู่บล็อกจะเลือกให้ค่าความแตกต่างของเฟรมที่เคลื่อนที่ (Displaced Frame Difference) น้อยที่สุด

$$DFD(x, d(x)) = \sum_{x \in R} (I(x, t) - I(x + d(x), t + 1)) \quad (2.8)$$

เกณฑ์ในการจับคู่บนพื้นฐานของ DFD ได้แก่ ค่าเฉลี่ยกำลังสองของความแตกต่าง (MSD) , ค่าเฉลี่ยสัมบูรณ์ของความแตกต่าง (MAD) และเกณฑ์บนพื้นฐานของฟังก์ชันแบบโลรองตซ์ (Lorentzian function)

$$L(d) = \sum_{x \in R} \log(1 + ((x, t) - I(x + d(x), t + 1))^2 / 2\sigma^2) \quad (2.9)$$

เกณฑ์การจับคู่บล็อกแบบง่าย เรียกว่า Pixel Difference Classification (PDC) เสนอโดย Ghavani และ Mills [GM90] ที่มีนิยามดังนี้

$$PDC(d) = \sum_{x \in R} T(x, d) \quad (2.10)$$

$$T(x, d) = \begin{cases} 1, & |I(x, t) - I(x + d(x), t + 1)| \leq T \\ 0, & otherwise \end{cases} \quad (2.11)$$

โดยที่

โดยที่ T เป็นค่าจุดเปลี่ยน ถ้า $T = 1$ แสดงว่าจุดภาพในบล็อกจะเป็นจุดภาพที่คู่กัน และบล็อกที่มีค่า PDC มากที่สุดจะถูกเลือกเป็นบล็อกที่คู่กัน

ข้อได้เปรียบของการประมาณการเคลื่อนที่โดยวิธีการจับคู่บล็อก คือ จะเป็นวิธีที่ง่ายต่อการนำมาใช้งานจริงและใช้ทรัพยากรในการคำนวณต่ำ ดังนั้นมาตรฐานการเข้ารหัสสัญญาณวิดีโอจึงใช้วิธีการนี้ในกระบวนการประมาณการเคลื่อนที่ นอกจากนี้ยังเป็นกรรมวิธีที่มีความเสถียร (robust) และถูกต้องมากกว่าวิธีดิฟเฟอเรนเชียล (differential approach) แต่มีข้อเสีย คือ สนามการเคลื่อนที่จะไม่เกาะกลุ่มกันเนื่องจากการพิจารณาเป็นบล็อก จึงเป็นวิธีที่เหมาะสมกับการเข้ารหัสมากกว่าการวิเคราะห์การเคลื่อนที่ นอกจากนี้ยังอาจเกิดบล็อกเทียม (block artifacts) ในการเข้ารหัส ซึ่งสามารถแก้ไขได้โดยการใช้แบบจำลองเลียนแบบ (affine model) หรือเพิ่มความถูกต้องโดยใช้การจับคู่บล็อกแบบแปรผันขนาดของบล็อกได้

2.5.4 วิธีพารามิเทริกซ์ (Parametric Methods)

วิธีการนี้เป็นการโปรเจกชันการเคลื่อนที่ของวัตถุ 3 มิติ เป็นการเคลื่อนที่แบบ 2 มิติบนระนาบของภาพ ซึ่งจะอธิบายการเคลื่อนที่ใน 3 มิติ โดยใช้ฟังก์ชันพื้นผิว 3 มิติ (3D surface function) และการโปรเจกชันทางสัญญาณ (projection geometry)

เมื่อพิจารณาจุดบนวัตถุที่เคลื่อนที่ใน 3 มิติ ให้ตำแหน่งของจุดที่เวลา t และ $t+1$ เป็น $X(t) = (X, Y, Z)^T$ และ $X' = X(t+1) = (X', Y', Z')^T$ ตามลำดับ จะนิยามการกระจัดของจุดใน 3 มิติได้ดังนี้

$$D(X) = X(t+1) - X(t) = X' - X \quad \text{หรือ} \quad X' = X + D(X) \quad (2.12)$$

เมื่อพิจารณาในระนาบของภาพ 2 มิติ ให้ตำแหน่งของจุดที่เวลา t และ $t+1$ เป็น $x(t) = (x, y)^T$ และ $x' = x(t+1) = (x', y')^T$ ตามลำดับ จะนิยามการกระจัดของจุดใน 2 มิติได้ดังนี้

$$d(x) = x(t+1) - x(t) = x' - x \quad \text{หรือ} \quad x' = x + d(x) \quad (2.13)$$

เวกเตอร์การกระจัดของทุกจุดในบริเวณหรือภาพจากสนามการเคลื่อนที่ สามารถจำลองขึ้นโดยการแปลง 2 มิติที่อธิบายโดยเซตของพารามิเตอร์ หรือเวกเตอร์พารามิเตอร์ T ซึ่งเรียกว่าพารามิเตอร์การแมปปิง (mapping parameters)

เซตของพารามิเตอร์ที่ใช้แทนการเคลื่อนที่ ขึ้นกับ 3 ปัจจัยด้วยกัน

(1) การเคลื่อนที่ใน 3 มิติของวัตถุ ซึ่งจะจำลองเหมือนการเคลื่อนที่ใน 3 มิติที่อธิบายด้วยเมทริกซ์การหมุน R (rotation matrix) และเมทริกซ์การเคลื่อนย้าย D (translation matrix)

$$X' = RX + D$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (2.18)$$

- (2) แบบจำลองทางคณิตศาสตร์ที่อธิบายการโปรเจกชันจาก 3 มิติลงบนระนาบ 2 มิติ ซึ่งทำโดย central หรือ parallel projection
- (3) การประมาณลักษณะพื้นผิวของวัตถุ ซึ่งอาจเป็นพื้นผิวเรียบ (planar) หรือแบบพาราโบลา (parabolic) แบบจำลองที่ใช้ทั่วไปในการวิเคราะห์การเคลื่อนที่ คือ
1. พื้นผิวระนาบภายใต้การโปรเจกชันแบบขนานหรือแบบตั้งฉาก (planar surface under parallel (orthographic) projection)

ภายใต้การโปรเจกชันแบบขนาน ใน 3 มิติและระนาบ 2 มิติมีความสัมพันธ์ดังนี้

$$(x, y) = (X, Y) \text{ และ } (x', y') = (X', Y') \quad (2.19)$$

$$\text{พื้นผิวเรียบแทนด้วย } aX + bY + cZ = 1 \quad (2.20)$$

โดยการรวมสมการ (2.18) , (2.19) และ (2.20) จะได้แบบจำลอง 6 พารามิเตอร์

$$\begin{aligned} x' &= a_1x + a_2y + a_3 \\ y' &= a_4x + a_5y + a_6 \end{aligned} \quad (2.21)$$

แบบจำลองนี้จะถูกเรียกว่า แบบจำลองการเคลื่อนที่ที่เลียนแบบ (affine motion model)

$$\text{โดยที่ } T = (a_1, a_2, a_3, a_4, a_5, a_6)^T$$

แบบจำลองนี้จะเหมาะสมและให้ผลที่ดีสำหรับกรณีที่ระยะจากกล้องจับภาพถึงวัตถุมี

ความยาวกว่าขนาดของวัตถุมาก

2. พื้นผิวระนาบภายใต้การโปรเจกชันแบบเพอร์สเปกทีฟหรือแบบศูนย์กลาง (planar surface under central (perspective) projection) สำหรับการมองที่เป็นจริงมากขึ้น จะมีการโปรเจกชันดังนี้

$$(x, y) = \left(f \frac{X}{Z}, f \frac{Y}{Z} \right), \quad (x', y') = \left(f' \frac{X'}{Z'}, f' \frac{Y'}{Z'} \right) \quad (2.22)$$

โดย f เป็นความยาวโฟกัสของกล้องจับภาพ และรวมสมการ (2.18) , (2.20) จะได้แบบจำลอง 8 พารามิเตอร์

$$x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1}, \quad y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \quad (2.23)$$

$$\text{โดยที่ } T = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)^T$$

3. พื้นผิวพาราโบลาลภายใต้การโปรเจกชันแบบขนาน (parabolic surface under parallel (orthographic) projection) การโปรเจกชันแบบขนานของพื้นผิวแบบพาราโบลา

$$Z = aX^2 + bXY + cY^2 + dX + eY + g \quad (2.24)$$

การเคลื่อนที่ตามสมการ (2.18) จะให้แบบจำลองควอดราติก 12 พารามิเตอร์

$$\begin{aligned} x' &= a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6 \\ y' &= b_1x^2 + b_2y^2 + b_3xy + b_4x + b_5y + b_6 \end{aligned} \quad (2.25)$$

โดยที่ $T = (a_1, a_2, a_3, a_4, a_5, a_6, b_1, b_2, b_3, b_4, b_5, b_6)^T$ มี 12 พารามิเตอร์

2.6 งานวิจัยที่เกี่ยวข้องทางการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ (Object Based Video Coding)

2.6.1 การพัฒนางานวิจัยทางการเข้ารหัสรูปทรง (Shape Coding)

ส่วนนี้เป็นส่วนสำคัญในการแสดงรูปทรงของวัตถุ (object representation) และสามารถช่วยในการลดข้อมูลในการเข้ารหัสแบบ Object Based ซึ่งมีงานวิจัยที่พัฒนาสำหรับการเข้ารหัสรูปทรงดังนี้ [Kim J., Evans B. L., 1998]

ในปี 1961 Freeman ได้ใช้รหัสลูกโซ่ (chain code) ในการเข้ารหัสรูปทรงแบบไม่มีการสูญเสีย (lossless contour)

ในปี 1991 Lu และ Durham ได้พัฒนารหัสลูกโซ่ขึ้นโดยการใช้การเข้ารหัสเลขคณิต (arithmetic code) สามารถปรับปรุงการเข้ารหัสได้ดีกว่า differential chain coding ถึงร้อยละ 25

ในปี 1994 ได้ใช้ predictive polygons และ spline code บนแนวทางของแบบจำลอง 2 มิติ (source model of flexible 2D)

ในปี 1997 O' Connell ได้ปรับปรุง lossy shape code โดยใช้ vertex-based chain code ซึ่งเหมาะสมสำหรับ non-predictive binary shape coding

ในปี 1998 ได้มีการปรับปรุง vertex-based เพื่อสร้างรหัสลูกโซ่ในการเข้ารหัส intra-binary shape code และขยายการประมาณแบบรูปหลายเหลี่ยม (polygon) เป็นการเข้ารหัสระหว่างกัน (inter-code) โดยการใช้การประมาณการเคลื่อนที่ของรูปทรงทั้งแบบมีการสูญเสียและไม่มีการสูญเสีย

2.6.2 การวิเคราะห์การเคลื่อนที่และการแยกส่วนภาพวัตถุ

การวิเคราะห์การเคลื่อนที่แบบขนาดบล็อกคงที่ จะมีข้อมูลที่สามารถชดเชยการเคลื่อนที่ได้เฉพาะบริเวณเล็กๆ แต่ยังได้ผลไม่เป็นที่น่าพอใจสำหรับบริเวณพื้นที่ขนาดใหญ่หรือข้อมูลภาพทั้งเฟรม จึงมีการเสนอวิธีการแก้ไขโดยใช้กรรมวิธีที่ขนาดของบล็อกแปรผันได้ (variable block size algorithm) และปรับแต่งได้ด้วยการชดเชยการเคลื่อนที่โดยรวม (global motion compensation) การแบ่งส่วนบล็อกภายใน (inner block segmentation) และแบบจำลองของการเคลื่อนที่ (motion models) ซึ่งจะทำให้อัตราบิดดีขึ้น ขณะที่ค่าอัตราส่วนสัญญาณสูงสุดต่อสัญญาณรบกวน (PSNR) ยังเท่าเดิม เมื่อเปรียบเทียบกับวิธีที่ใช้ขนาดบล็อกคงที่ (fixed block size) โดยจะพิจารณาเฉพาะการชดเชยการเคลื่อนที่ของการเลื่อนตำแหน่ง (translation) เท่านั้น

ส่วนข้อได้เปรียบสำหรับการวิเคราะห์บนพื้นฐานของบล็อก (Block Based) คือ สามารถนำไปประยุกต์ใช้ (implement) และทำงานแบบเวลาจริงได้ โดยมีประสิทธิภาพการเข้ารหัสที่ดี เหมือนกับวิธีบนพื้นฐานของการแยกส่วน (Segmentation Based)

ส่วนแบบขนาดของบล็อกแปรผันได้ (Variable Block Size) จะสามารถเปลี่ยนขนาดของบล็อกตามรูปแบบของการเคลื่อนที่ และรูปร่างของวัตถุ มีข้อดีที่ประสิทธิภาพจะดีกว่าแบบขนาดของบล็อกคงที่ (Fixed Block Size) เมื่อเปรียบเทียบค่า PSNR และอัตราบิต แต่มีข้อเสีย คือ จะเกิดขอบเทียม (artifacts) กับลำดับภาพที่ถอดรหัสมา เพราะการใช้ขนาดของบล็อกที่ไม่เท่ากัน ซึ่งจะทำให้เห็นบริเวณขอบของการเคลื่อนที่ ปัญหานี้อาจแก้ไขโดยปรับปรุงคุณภาพของบริเวณนั้น ด้วยวิธีการแยกส่วนการเคลื่อนที่ (motion segmentation methods) ดังต่อไปนี้

(1) การแยกส่วนโดยใช้ Gray Level Thresholding

จะมีข้อได้เปรียบ คือ ทำให้ขอบของการเคลื่อนที่ (motion boundary) กลับคืนมาได้จากเฟรมที่สร้างขึ้นมาใหม่ โดยไม่ต้องเพิ่มส่วนหัว (overhead) ที่ทำให้อัตราบิตเพิ่มขึ้นสำหรับการเข้ารหัสรูปทรง (shape coding) แต่มีข้อเสียที่ใช้ไม่ได้สำหรับบริเวณที่มีความแตกต่างกันต่ำเกินไป

(2) การค้นหาขอบเขตการเคลื่อนที่อย่างละเอียด

โดยพิจารณาการเคลื่อนที่โดยรวม (global motion) ที่เกิดจากการเคลื่อนที่ของกล้องจับภาพและปัญหาในการประมาณการเคลื่อนที่ที่ซับซ้อน จะแสดงขอบเขตของการเคลื่อนที่ได้ชัดเจน แต่ต้องค้นหาในหลายมิติ จึงใช้การคำนวณสูงไม่เหมาะกับการใช้งานจริง

(3) การประมาณการเคลื่อนที่โดยใช้การแปลงแบบเฮาท์ (Hough Transform)

ใช้สำหรับประมาณการเคลื่อนที่แบบขนานและหาค่าพารามิเตอร์ของการเคลื่อนที่เพื่อการแยกส่วนภาพ

จากงานวิจัยเหล่านี้ จึงได้มีการเสนอวิธีปรับปรุงประสิทธิภาพของการขจัดขอบเขตการเคลื่อนที่โดยใช้โครงสร้างแบบชั้น (layered structure) และการประมาณการเคลื่อนที่แบบลำดับชั้น (hierarchical motion estimation)

2.6.3 งานวิจัยทางการแยกส่วนภาพวัตถุ (Object Segmentation)

สำหรับการแยกส่วนภาพวัตถุสำหรับสัญญาณวิดีโอ จะแบ่งตามขั้นตอนเริ่มแรกได้ 2 แบบ

1. แบบที่ผู้ใช้ช่วยกำหนดขอบเขตเบื้องต้นให้กับวัตถุ (User Assisted หรือ Active System)
2. แบบอัตโนมัติ (Automatic Segmentation)

2.6.3.1 ระบบแบบผู้ช่วยกำหนดขอบเขตเบื้องต้น (User Assisted System)

ระบบแบบนี้จะมีกระบวนการอยู่ 2 ขั้นตอนด้วยกัน คือ ขั้นตอนแรกผู้ใช้จะต้องกำหนดขอบเขตของวัตถุที่ต้องการแยกส่วนก่อนในเฟรมแรกของลำดับภาพ โดยใช้การเชื่อมโยงผ่านอุปกรณ์คอมพิวเตอร์ เช่น เมาส์ (mouse) หรือ ปากกาแสง (optical pen) ซึ่งขอบเขตที่ได้จะไม่ละเอียด จึงต้องผ่านกระบวนการเพื่อให้ได้ขอบของภาพวัตถุถูกต้องยิ่งขึ้น เช่น Snake Algorithm [Zhong D., Chang S. F., 1998] จากนั้นจึงนำผลการแยกส่วนภาพวัตถุในขั้นต้นผ่านขั้นตอนของกระบวนการติดตามการเคลื่อนที่ของวัตถุ (object tracking) ในกระบวนการขั้นถัดไป

นอกจากนั้นกรรมวิธีที่เรียกว่า การแยกส่วนแบบกึ่งอัตโนมัติ (Semi Automatic Segmentation) [Kwak J., 1999] จะใช้การเชื่อมโยงกับผู้ใช้ผ่านทางกราฟฟิก GUI (graphical user interface) ในการกำหนดกรอบของเฟรมแรกที่ต้องการผ่านกระบวนการ แล้วจึงใช้กรรมวิธีสันปันน้ำ (Watershed Algorithm) เพื่อทำการแยกส่วนเชิงพื้นที่ (Spatial Segmentation) จากนั้นจึงผ่านขั้นตอนของการใช้ข้อมูลการเคลื่อนที่ร่วมด้วย สำหรับกรรมวิธีสันปันน้ำจะทำให้ขอบเขตของวัตถุถูกต้องขึ้น เป็นวิธีพื้นฐานของกรรมวิธีการเติบโตของบริเวณ (Region Growing Algorithm) โดยเริ่มจากจุดที่ตั้งไว้ (markers) แล้วจึงค่อยๆ เคลื่อนเข้าไปที่ขอบของวัตถุที่มีลักษณะของบริเวณที่แตกต่างกัน ทำให้ได้ขอบเขตของภาพวัตถุนั้น

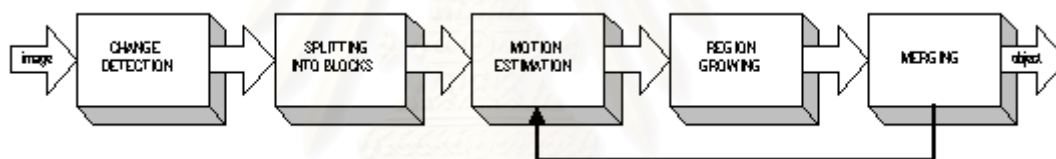
2.6.3.2 ระบบการแยกส่วนภาพวัตถุแบบอัตโนมัติ (Automatic Segmentation)

เนื่องจากการแยกส่วนของภาพวัตถุสำหรับตัวเข้ารหัสของมาตรฐาน MPEG-4 ยังไม่ได้กำหนดกรรมวิธีในการแยกส่วนที่ชัดเจน มีเพียงข้อกำหนดให้ชาวสสารสามารถแบ่งแยกได้ที่ตัวถอดรหัสเท่านั้น สำหรับสัญญาณวิดีโอแบบธรรมชาติ (Natural Video) และการใช้งานแบบเวลาจริงจะต้องใช้วิธีการแยกส่วนภาพที่มีคุณสมบัติสามารถทำงานได้อย่างอัตโนมัติ [Vass J., Palaniappan K., Zhuang X., 1998] ระบบแบบนี้ไม่จำเป็นต้องให้ผู้ใช้เข้าไปยุ่งเกี่ยวระหว่างกระบวนการและไม่ต้องรู้ข้อมูลที่ผ่านมาก่อนด้วย

2.6.4 การแยกส่วนการเคลื่อนที่โดยการเติบโตของบริเวณ (Region Growing)

วิธีการเติบโตของบริเวณมีโครงสร้างเป็นลำดับขั้น จากบริเวณที่เล็กสู่บริเวณที่ใหญ่กว่า บริเวณที่เปลี่ยนแปลงหรือมีการเคลื่อนที่ของวัตถุจะถูกแยกส่วนออกเป็นบล็อกขนาดคงที่ ซึ่งมีขนาดเล็กพอที่จะแยกความแตกต่างของวัตถุที่สนใจและใหญ่เพียงพอที่สามารถหาพารามิเตอร์การเคลื่อนที่ได้ ส่วนของวัตถุที่เชื่อมต่อกันจะรวมเข้าเป็นวัตถุเดียวกันโดยอยู่ในเกณฑ์ความคลาดเคลื่อน (error criterion) และพารามิเตอร์การเคลื่อนที่ในบริเวณที่กำหนดแล้ว (labeled regions) จะพิจารณาจุดภาพที่ยังไม่ได้กำหนด (unlabeled pixels) ที่อยู่ติดกัน ซึ่งบริเวณนี้จะเติบโตใหญ่ขึ้นถ้าอยู่ในเกณฑ์ของความคลาดเคลื่อน กระบวนการรวมตัว (merging) และการเติบโตของ

บริเวณ (region growing) จะมีการประมาณค่าพารามิเตอร์การเคลื่อนที่ใหม่ โดยที่ค่าจุดเปลี่ยน (threshold) ซึ่งกำหนดเกณฑ์ความคลาดเคลื่อนจะปรับค่าไป และบริเวณนี้จะขยายจนกระทั่งขนาดของบริเวณที่ได้ใหม่นี้มีขนาดตามที่กำหนด ส่วนของบริเวณที่มีการเปลี่ยนแปลงจะแบ่งเป็น 2 บริเวณ คือ บริเวณที่เคยปรากฏขึ้นแล้ว หรือ Motion Compliance (MC) และบริเวณที่ปรากฏขึ้นใหม่ หรือ Motion Failure (MF) ซึ่งบริเวณ Motion Compliance (MC) เป็นบริเวณที่การเคลื่อนที่ของวัตถุสามารถชดเชยได้จากแบบจำลองการเคลื่อนที่ ส่วนบริเวณ Motion Failure (MF) จะไม่สามารถชดเชยได้ด้วยแบบจำลองการเคลื่อนที่ บริเวณ MF นี้รวมถึงบริเวณพื้นหลังที่ถูกครอบคลุมและไม่ถูกครอบคลุมในภาพและวัตถุที่ปรากฏขึ้นมาใหม่ด้วย ค่าการเข้ารหัส (coding cost) หรือ อัตราบิตขึ้นอยู่กับค่าพารามิเตอร์ของรูปทรงทั้งสองบริเวณ พารามิเตอร์การเคลื่อนที่ของบริเวณ MC และข้อมูลสีของบริเวณ MF จะมีอัตราบิตของการเข้ารหัสค่อนข้างสูง ดังนั้นขนาดของบริเวณ MF และจำนวนของบริเวณ MC จึงควรจะให้มีความเล็กเท่าที่ความผิดพลาดจะยอมรับได้ เพราะส่วนข้อมูลของการเข้ารหัสรูปทรงจะเพิ่มขึ้น โดยขึ้นกับความซับซ้อนและรายละเอียดของรูปทรงด้วย [9] ขั้นตอนการแยกส่วนภาพวัตถุโดยกรรมวิธีการเติบโตของบริเวณแสดงดังรูปที่ 2.12



รูปที่ 2.12 โครงสร้างการแยกส่วนภาพวัตถุโดยกรรมวิธีการเติบโตของบริเวณ (Region Growing)

2.6.5 การแยกส่วนเชิงพื้นที่-เวลา (Spatio-Temporal Segmentation)

การแยกส่วนภาพวัตถุต้องการให้ได้ผลลัพธ์ที่ดีเพื่อใช้ในการประมาณการเคลื่อนที่ของวัตถุ และการประมาณการเคลื่อนที่ที่ถูกต้องก็จำเป็นสำหรับการแยกส่วนภาพวัตถุที่ดีด้วย ดังนั้นทั้งสองส่วนนี้จึงต้องทำไปพร้อมๆ กันจึงจะได้ผลดี ดังนั้นจึงมีงานวิจัยค้นคว้ากรรมวิธีใหม่ๆ ที่จะแยกส่วนภาพวัตถุให้ได้ผลดี

การแยกส่วนโดยใช้ข้อมูลของความเข้มและตัวปฏิบัติการเชิงสัณฐาน (morphological operator) มีข้อเสีย คือ ภาพที่ได้จากการแยกส่วนจะขึ้นอยู่กับความเข้มของจุดภาพที่เป็นเนื้อเดียวกันเท่านั้น ส่วนการประมาณการเคลื่อนที่โดยใช้ค่าเฉลี่ยของการเคลื่อนที่โดยรวม และความหนาแน่นของการไหลทางแสง (dense optical flow) จะใช้แบ่งแยกกลุ่มของพารามิเตอร์การเคลื่อนที่ แต่ผลของการประมาณการเคลื่อนที่ที่ยังไม่สามารถที่จะแสดงบริเวณขอบของวัตถุที่ต้องการได้

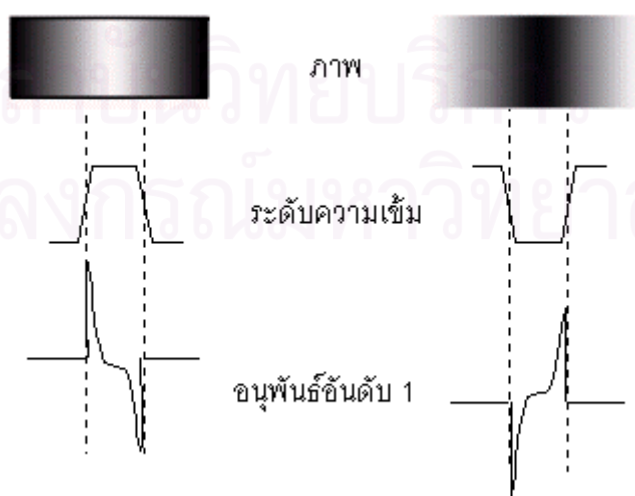
[Choi J. G., 1997]

เพื่อที่จะแก้ปัญหาเหล่านั้น ได้มีการนำเสนอเทคนิคการแยกส่วนภาพเชิงพื้นที่-เวลา (Spatio-Temporal) [Dufaux F., Moscheni F., 1995] โดยมีขั้นตอนพิจารณาเป็นลำดับดังนี้ เริ่มตั้งแต่การแยกส่วนโดยใช้ข้อมูลทางความเข้ม การแยกส่วนโดยใช้ข้อมูลทางการเคลื่อนที่ และการแยกส่วนโดยใช้ข้อมูลทั้งสองส่วนควบคู่กันไป เนื่องจากการใช้ข้อมูลการแยกส่วนทั้งทางความเข้มและการเคลื่อนที่ไปพร้อม ๆ กัน จึงจำเป็นต้องใช้การคำนวณในแต่ละบริเวณซ้ำ ๆ กัน เพื่อกำหนดจุดภาพให้เป็นบริเวณต่าง ๆ กัน ซึ่งวิธีนี้จะต้องใช้การคำนวณค่อนข้างมากในการหาค่าความเข้มเฉลี่ยแทนค่าของบริเวณที่เปลี่ยนไป จึงได้มีการปรับปรุงการคำนวณค่าเฉลี่ยซ้ำๆกันนี้ โดยการรวมกันครั้งละหลายจุดภาพเข้าในบริเวณแล้วจึงค่อยคำนวณค่าเฉลี่ยของแต่ละบริเวณ ทำให้ใช้การคำนวณที่ลดลงเนื่องจากจุดภาพที่อยู่ใกล้ ๆ กันจะมีคุณสมบัติของความเข้มและการเคลื่อนที่คล้ายกัน [Choi J. G., 1997] การปรับปรุงนี้แม้จะลดทรัพยากรในการคำนวณไปได้บ้าง แต่ก็ยังใช้การคำนวณค่อนข้างสูงอยู่

จากงานวิจัยข้างต้นนี้ การใช้ทั้งข้อมูลความเข้มและข้อมูลการเคลื่อนที่ของวัตถุในลำดับภาพ ทำให้ได้การแยกส่วนภาพวัตถุที่ดี แต่เนื่องจากการใช้ข้อมูลทั้งสองรวมกันในแต่ละบริเวณแล้วจึงทำให้การเติบโตของบริเวณต้องใช้การคำนวณค่อนข้างมาก จึงไม่เหมาะที่จะนำมาใช้ในกระบวนการแยกส่วนภาพวัตถุเพื่อทำการเข้ารหัสสัญญาณวิดีโออัตราบิตต่ำบนพื้นฐานของหน่วยวัตถุ อย่างไรก็ตามวิธีนี้ทำให้เกิดแนวคิดในการกำหนดค่าจุดเปลี่ยนและค่าน้ำหนักของข้อมูลทั้งสองให้เหมาะสมกับลักษณะของสัญญาณภาพวิดีโอ

2.7 การตรวจรู้ขอบ (Edge Detection)

ขอบ (edge) ของวัตถุ คือ ขอบเขต (boundary) ระหว่าง 2 บริเวณที่มีค่าความเข้มแตกต่างกัน โดยการพิจารณาขอบของภาพนั้นเราจะพิจารณาจากข้อมูลระดับความเข้ม ดังรูปที่ 2.13



รูปที่ 2.13 แสดงถึงระดับความเข้มของขอบ และอนุพันธ์อันดับ 1 ของระดับความเข้ม

2.7.1 การตรวจรู้ขอบด้วยค่าเกรเดียนต์

จะเห็นว่าการตรวจรู้ขอบนั้นเราทำได้โดยพิจารณาขนาดของอนุพันธ์อันดับ 1 ของระดับความเข้ม ซึ่งทำโดยใช้ตัวปฏิบัติการเกรเดียนต์ (Gradient Operator: ∇) จากการวิเคราะห์พบว่าเมื่อนำเกรเดียนต์มาใช้กับค่าความเข้มในแต่ละจุดภาพจะได้ความสัมพันธ์ดังนี้

$$G_x = (Z_7 + 2Z_8 + Z_9) - (Z_1 + 2Z_2 + Z_3) \quad (2.26)$$

$$G_y = (Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7) \quad (2.27)$$

โดยที่ G_x คือ ค่าเกรเดียนต์ตามแนวแกน x

G_y คือ ค่าเกรเดียนต์ตามแนวแกน y

Z_i คือ ระดับความเข้มของจุดภาพในบริเวณ 3×3 จุดภาพที่จุดภาพ $i = 1, 2, \dots, 9$

การหาค่าเกรเดียนต์ของระดับความเข้มทำได้โดยใช้มาสก์ขนาด 3×3 จุดภาพดังรูปที่

2.14 ผ่านจุดภาพทั้งหมดบนภาพที่พิจารณา

Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

มาสก์ขนาด 3×3

-1	-2	-1
0	0	0
1	2	1

มาสก์สำหรับหาค่า G_x

-1	0	1
-2	0	2
-1	0	1

มาสก์สำหรับหาค่า G_y

รูปที่ 2.14 บริเวณภาพขนาด 3×3 จุดภาพ และมาสก์ขนาด 3×3 สำหรับใช้หาค่า G_x และ G_y

2.7.2 การตรวจรู้ขอบแบบแคนนี่ (Canny)

เป็นวิธีการหนึ่งในการหาขอบภาพ ประกอบด้วย 4 ขั้นตอนย่อย ๆ ดังนี้

1. การกรองความถี่ด้วยตัวกรองแบบไบโนเมียล

$$f(x, y, n) = 1/16 \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.26)$$

โดยที่ $n = 2$ เท่าของความแปรปรวนในการกระจายแบบเกาส์เซียน

2. การคำนวณเกรเดียนต์โดยใช้มาสก์แบบ Prewitt

3. การหาจุดภาพที่มีค่าเกรเดียนต์สูงสุดเฉพาะถิ่น

กำหนดขนาดเกรเดียนต์ของจุดภาพ เป็น $p_1 \dots p_9$ ดังนี้

$p_1 \ p_2 \ p_3$

$p_4 \ p_5 \ p_6$

$p_7 \ p_8 \ p_9$

โดยให้ p_i, p_j และ p_5 อยู่ในแนวเส้นตรงเดียวกัน โดย p_5 อยู่ระหว่าง p_i และ p_j ในที่นี้จะลากเส้นตรงได้ 4 แนว

จุดภาพ 5 เป็นจุดสูงสุดเฉพาะถิ่นก็ต่อเมื่อ $p_i < p_5$ และ $p_j < p_5$

ถ้าจุดภาพ 5 เป็นจุดสูงสุดเฉพาะถิ่น และ $p_5 > T_L$ (ค่าจุดเปลี่ยนค่าต่ำ) ก็จะใช้จุดภาพ 5 นี้ในการคำนวณขั้นสุดท้าย

4. การตั้งค่าจุดเปลี่ยน (Threshold) แบบ Hysteresis

ขั้นตอนที่ใช้เลือกจุดภาพที่เป็นขอบภาพ มีหลักการดังนี้

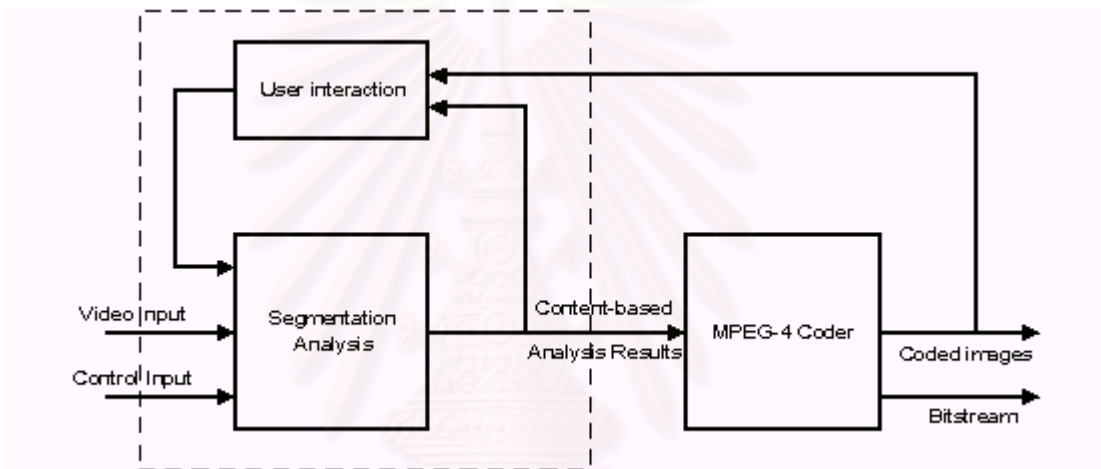
1. จุดภาพที่มีค่าเกรเดียนต์ $> T_H$ (ค่าจุดเปลี่ยนค่าสูง) จะถือว่าเป็นขอบ
2. ทุก ๆ จุดภาพที่ประกอบกันเป็นลายเส้นขอบเส้นหนึ่ง จะต้องมีย่านอย่างน้อยหนึ่งจุดภาพในลายเส้นนั้นที่มีค่าเกรเดียนต์ มากกว่า ค่าจุดเปลี่ยน
3. จุดภาพที่มีค่าเกรเดียนต์ น้อยกว่า ค่าจุดเปลี่ยน และไม่สอดคล้องกับ ข้อ 2. จะถูกตัดทิ้ง

บทที่ 3

กรรมวิธีที่นำเสนอ

3.1 ความสำคัญของการแยกส่วนภาพวัตถุในการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ

การเข้ารหัสสัญญาณวิดีโอตามข้อกำหนดของมาตรฐาน MPEG-4 จำเป็นต้องมีการแยกส่วนภาพวัตถุเป็นกระบวนการขั้นต้น เพื่อให้ได้ภาพวัตถุแยกเป็นอิสระต่อกันสำหรับการเข้ารหัสของ MPEG-4 โครงสร้างของการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 แสดงดังรูปที่ 3.1



รูปที่ 3.1 โครงสร้างของการแยกส่วนภาพวัตถุและการเข้ารหัส MPEG-4 สำหรับการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ (Object Based Video Coding)

ขั้นตอนการแยกส่วนภาพวัตถุนั้น จะต้องสามารถแยกภาพวัตถุแต่ละภาพออกมาให้มีความหมายในตัวเอง และมีฟังก์ชันบนพื้นฐานของคอนเทนต์ มีการเสนอกรรมวิธีจำนวนมากเพื่อมาใช้ในการแยกส่วนภาพวัตถุนั้น แต่ยังไม่มีการวิธีใดที่สามารถแยกส่วนภาพวัตถุได้อย่างแม่นยำและทำงานโดยอัตโนมัติ [Zhang D., Lu G.; 1999] งานวิจัยนี้จึงมีจุดมุ่งหมายเพื่อพัฒนากกรรมวิธี การแยกส่วนภาพที่เหมาะสมสำหรับการใช้งานเฉพาะอย่างหรือการใช้งานทั่วไปอย่างมีประสิทธิภาพ

การแยกส่วนภาพวัตถุนั้น จากการศึกษพบว่าการใช้การแยกส่วนภาพเชิงพื้นที่ร่วมกับการแยกส่วนภาพเชิงเวลาจะให้ผลการแยกส่วนภาพที่ดี โดยแต่ละขั้นตอนมีจุดประสงค์เพื่อสนับสนุนซึ่งกันและกัน ข้อมูลเชิงพื้นที่ใช้สำหรับการกำหนดขอบเขตของภาพวัตถุที่ชัดเจน ส่วนข้อมูลเชิงเวลาใช้ในการระบุตำแหน่งของวัตถุที่เคลื่อนที่และความต่อเนื่องของการเคลื่อนที่ กรรมวิธีในแต่ละขั้นตอน จะอธิบายในหัวข้อถัดไป

3.2 แนวความคิด

กรรมวิธีการแยกส่วนภาพวัตถุของลำดับภาพในปัจจุบัน มีจุดประสงค์เพื่อตรวจรู้วัตถุที่เคลื่อนที่ สามารถแบ่งวิธีการหลักได้ 3 วิธี คือ การใช้ข้อมูลของความเข้ม (intensity based) , การใช้ข้อมูลของการเคลื่อนที่ (motion based) และ การใช้ข้อมูลการเคลื่อนที่และข้อมูลของความเข้มไปด้วยกัน

วิธีการพิจารณาข้อมูลของความเข้ม จะใช้ข้อมูลความเข้มภายในเฟรมและข้อมูลความเข้มระหว่างเฟรม โดยปกติจะอยู่ในรูปแบบของการแยกส่วนความเข้มแสง และการตรวจรู้การเปลี่ยนแปลง โดยจะสามารถแยกส่วนภาพวัตถุที่เคลื่อนที่ออกจากบริเวณของพื้นหลังที่อยู่นิ่งได้ แต่ไม่สามารถที่จะแยกส่วนภาพวัตถุที่มีการเคลื่อนที่แตกต่างกันได้ [Alatan A. A., Onural L., Wollborn M.; 1998]

ส่วนวิธีการพิจารณาข้อมูลของการเคลื่อนที่ จะได้จากกระบวนการที่ใช้ข้อมูลของความเข้ม การใช้ข้อมูลของการเคลื่อนที่ในการแยกส่วนภาพ ในแต่ละเฟรมจะสามารถแยกเป็นบริเวณจำนวนหนึ่งในแต่ละบริเวณมีการเคลื่อนที่เหมือนกัน ถึงแม้ว่าวิธีนี้จะสามารถระบุตำแหน่งของวัตถุได้ แต่ยังเกิดปัญหาจากการประมาณการเคลื่อนที่ที่ต้องใช้ข้อมูลหลายเฟรม จึงทำให้ขอบเขตของภาพวัตถุที่ได้ไม่แม่นยำ ด้วยเหตุนี้จึงเกิดวิธีที่ใช้การประมาณการเคลื่อนที่และการแยกส่วนภาพวัตถุไปด้วยกัน ซึ่งจะมีประสิทธิภาพมาก แต่ก็ยังไม่เป็นที่สนใจ เนื่องจากต้องใช้การคำนวณที่ซับซ้อนค่อนข้างสูง

ดังนั้นการลดทรัพยากรที่ใช้ในการคำนวณที่ต้องใช้ในการแยกส่วนภาพวัตถุโดยใช้ข้อมูลของความเข้มและการเคลื่อนที่ไปด้วยกัน จึงเป็นงานวิจัยที่น่าสนใจ แนวทางหนึ่ง คือ การลดการคำนวณที่ต้องทำไปพร้อมกัน ด้วยการแยกกระบวนการในแต่ละขั้นตอนออกจากกัน แล้วนำผลในแต่ละกระบวนการที่ได้มารวมกันโดยใช้ขั้นตอนการตัดสินใจระบุบริเวณของภาพวัตถุ

การรวมกันของคุณลักษณะหลายอย่าง

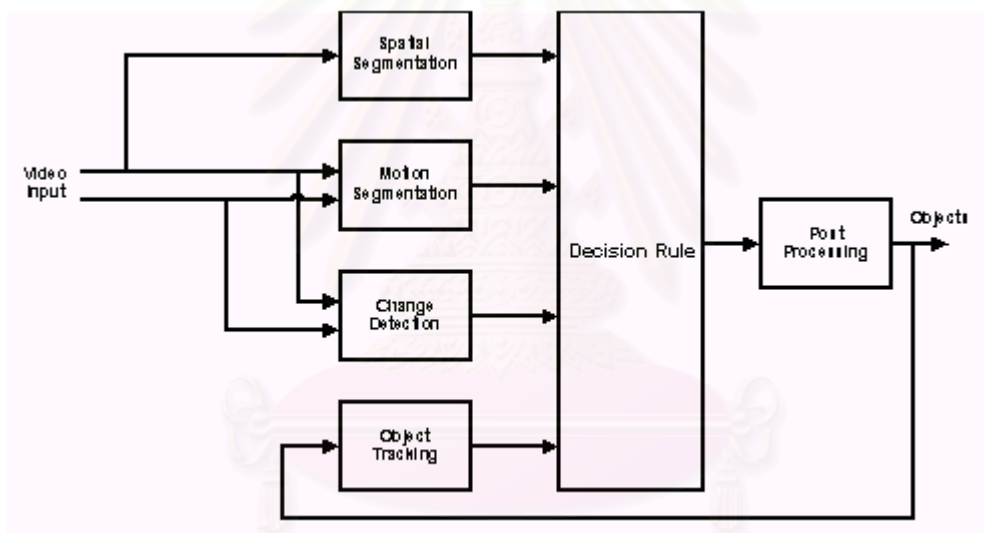
เนื่องจากสัญญาณวิดีโอมีคุณลักษณะที่แตกต่างกันอยู่หลายอย่างด้วยกัน คือ สัญญาณ วิดีโอที่มีการเคลื่อนที่ของวัตถุเร็ว เช่น ลำดับภาพรถวิ่ง หรือเรือแล่น และสัญญาณวิดีโอที่มีการเคลื่อนที่ของวัตถุช้า เช่น ลำดับภาพของการประชุมทางวิดีโอ หรือโทรศัพท์ภาพ นอกจากนี้บางภาพยังมีรายละเอียดค่อนข้างซับซ้อน และบางภาพมีรายละเอียดน้อย ส่วนการกระจายความเข้มแสงของลำดับภาพก็เป็นอีกปัจจัยหนึ่งด้วย ดังนั้นข้อมูลที่ใช้สำหรับการแยกส่วนภาพวัตถุของสัญญาณวิดีโอแต่ละอย่างจึงมีความสำคัญที่แตกต่างกัน กรรมวิธีการแยกส่วนภาพวัตถุจึงควรมีการวิเคราะห์ลักษณะของลำดับภาพและควรมีความสามารถที่จะปรับความสำคัญของข้อมูลให้

เหมาะสมกับลำดับภาพแต่ละประเภท นอกจากนั้นยังควรที่จะปรับขั้นตอนการประมวลผลเพื่อเพิ่มคุณภาพของภาพวัตถุหรือเพิ่มความเร็วในการคำนวณได้ด้วย และภาพวัตถุที่ได้จะต้องมีความหมายต่อการมองเห็น และสนับสนุนฟังก์ชันบนพื้นฐานของคอนเทนต์ตามข้อกำหนดของ มาตรฐาน MPEG-4

3.3 กรรรมวิธีของการแยกส่วนภาพวัตถุสำหรับสัญญาณวิดีโอ

กรรรมวิธีการแยกส่วนภาพวัตถุนี้จะใช้ทั้งข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา ซึ่งประกอบด้วยข้อมูลของความเข้มแสง ข้อมูลของการเคลื่อนที่ และข้อมูลการเปลี่ยนแปลงความเข้ม เป็นข้อมูลที่น่ามาวิเคราะห์และใช้ร่วมกันในการแยกส่วนภาพวัตถุของลำดับภาพ โดยข้อมูลที่ได้จากการวิเคราะห์ในแต่ละส่วนจะมีคุณลักษณะและความสำคัญที่แตกต่างกัน

กระบวนการแยกส่วนภาพวัตถุประกอบด้วยขั้นตอนดังต่อไปนี้



รูปที่ 3.2 โครงสร้างของกรรรมวิธีการแยกส่วนภาพวัตถุโดยการรวมกันของคุณลักษณะหลายอย่าง

3.3.1 กระบวนการขั้นต้น (Preprocessing)

ในกรณีที่ภาพมีการเคลื่อนที่เนื่องจากกล้องจับภาพจะทำการประมาณการเคลื่อนที่และการชดเชยการเคลื่อนที่โดยรวม (Global Motion Compensation) โดยจะพิจารณาจุดภาพในบริเวณพื้นหลังของภาพก่อนหน้าเปรียบเทียบกับจุดภาพปัจจุบัน การหาเวกเตอร์การเคลื่อนที่ที่ใช้การค้นหาแบบทั้งหมด เพื่อค้นหาภายในบริเวณพื้นที่ขนาด 16X16 จุดภาพ และชดเชยการเคลื่อนที่โดยรวม เพื่อให้สามารถวิเคราะห์การเคลื่อนที่ของวัตถุภายในเฟรมได้ สำหรับกรณีที่พื้นหลังไม่มีการเคลื่อนที่เนื่องจากกล้องจับภาพอยู่นิ่งจะไม่จำเป็นต้องใช้ขั้นตอนนี้

3.3.2 การแยกส่วนภาพเชิงพื้นที่ (Spatial Segmentation)

กรรมวิธีการแยกส่วนภาพเชิงพื้นที่จะอยู่บนพื้นฐานของกรรมวิธีเชิงสัญญาณ ซึ่งจะแยกส่วนภาพทั้งภาพเป็นบริเวณที่มีความเข้มคล้ายคลึงกัน โดยบริเวณที่แตกต่างกันจะแบ่งด้วยเส้นขอบเขต ซึ่ง Salembier [Salembier P., Pardas M.; 1994] ได้เสนอวิธีสำหรับการแยกส่วนเชิงพื้นที่ซึ่งอยู่บนพื้นฐานของกรรมวิธีเชิงสัญญาณ โดยใช้ตัวกรองเชิงสัญญาณ (morphological filter) การประมาณค่าเกรเดียนต์ (gradient approximation) และกรรมวิธีสันปันน้ำ (watershed algorithm) เนื่องจากกรรมวิธีสันปันน้ำมักให้ผลการแยกส่วนมากเกินไป (oversegmentation) จึงใช้การสร้างภาพให้เรียบง่ายขึ้นเพื่อแก้ปัญหานี้ และเพื่อลดการคำนวณลงจะหลีกเลี่ยงการใช้กระบวนการรวมบริเวณ (region merging) โดยลดจำนวนจุดภาพบริเวณขอบของการตรวจรู้ขอบเขตของบริเวณ

กรรมวิธีการแยกส่วนภาพเชิงพื้นที่ ประกอบด้วยขั้นตอนดังต่อไปนี้

1. การสร้างภาพให้มีรูปแบบที่เรียบง่ายขึ้นต่อการแยกส่วนภาพ (image simplification)
2. การตรวจรู้ขอบเขตของบริเวณโดยวิธีการประมาณเกรเดียนต์ (gradient approximation)
3. การกำหนดบริเวณให้กับจุดภาพโดยกรรมวิธีสันปันน้ำ (watershed algorithm)

3.3.2.1 การสร้างภาพให้มีรูปแบบที่เรียบง่ายขึ้น (Image Simplification)

การทำให้ภาพมีรูปแบบที่เรียบง่ายขึ้น โดยการกรองบริเวณที่มีรายละเอียดของภาพมากในบริเวณที่สนใจ เพื่อให้สามารถแยกส่วนภาพที่ต้องการได้ง่ายขึ้น ด้วยการในตัวกรองเชิงสัญญาณ ที่ประกอบด้วยตัวปฏิบัติการแบบเปิดและแบบปิด (opening-closing operator)

การปฏิบัติการเชิงสัญญาณ

การปฏิบัติการเชิงสัญญาณโดยทั่วไปจะอยู่บนพื้นฐานของตัวปฏิบัติการ 2 กระบวนการด้วยกัน คือ ตัวปฏิบัติการไคเลชัน และ ตัวปฏิบัติการอีโรชัน

นิยาม ตัวปฏิบัติการไคเลชัน

$$\text{Dilation: } \delta_n(f(x)) = \text{Max} \{f(x-y), y \in M_n\} \quad (3.1)$$

นิยาม ตัวปฏิบัติการอีโรชัน

$$\text{Erosion: } \varepsilon_n(f(x)) = \text{Min} \{f(x+y), y \in M_n\} \quad (3.2)$$

โดยที่ $f(x)$ เป็นสัญญาณภาพ และ M_n เป็นองค์ประกอบโครงสร้างขนาด $n \times n$

องค์ประกอบโครงสร้าง (structuring element) เป็นเมทริกซ์ที่ประกอบด้วยค่า 1 และ 0 คือ บริเวณของจุดภาพข้างเคียง (pixel's neighborhood) สำหรับการกระทำไคเลชันหรืออีโรชัน ซึ่งมีขนาดต่าง ๆ กันได้ จุดภาพตรงกลางจะแทนจุดภาพที่สนใจ และส่วนประกอบของเมทริกซ์ที่เป็น 1 จะเป็นบริเวณของจุดภาพข้างเคียง

ตัวอย่างการดำเนินการเชิงสัณฐาน (morphological operation)

ภาพ

2 2 0 0 0 0 0 0 0 0

2 2 0 0 1 1 1 0 0 0

0 0 0 0 1 1 1 0 0 0

0 0 3 0 1 1 1 0 0 0

0 0 0 0 0 0 0 0 0 0

องค์ประกอบโครงสร้าง (structure element: SE)

องค์ประกอบโครงสร้างขนาด 3x3 จุดภาพ

1 1 1

1 1 1

1 1 1

อีโรชัน (erosion) ถ้าทุกจุดภาพขององค์ประกอบโครงสร้างเป็น 1 เมื่อเทียบกับภาพแล้ว จุดภาพที่พิจารณาจะเป็น 1 มิฉะนั้นจุดภาพที่พิจารณาจะมีค่าเป็น 0

ตัวอย่างอีโรชันของภาพ

0 0 0 0 0 0 0 0 0 0

0 0 0 0 1 1 1 0 0 0

0 0 0 0 1 1 1 0 0 0

0 0 0 0 1 1 1 0 0 0

0 0 0 0 0 0 0 0 0 0

ไดเลชัน (dilation) ถ้ามีบางจุดภาพขององค์ประกอบโครงสร้างเป็น 1 เมื่อเทียบกับภาพแล้ว จุดภาพที่พิจารณาจะมีค่าเป็น 1 มิฉะนั้นจุดภาพที่พิจารณาจะมีค่าเป็น 0

ตัวอย่างไดเลชันของภาพ

2 2 2 1 1 1 1 1 0 0

2 2 2 1 1 1 1 1 0 0

2 3 3 3 1 1 1 1 0 0

0 3 3 3 1 1 1 1 0 0

0 3 3 3 1 1 1 1 0 0

การปฏิบัติการเชิงสัณฐานแบบเปิดและแบบปิด

การปฏิบัติการแบบปิด (closing operation) คือ การไดเลชันแล้วตามด้วยการอีโรชัน

การปฏิบัติการแบบเปิด (opening operation) คือ การอีโรชันแล้วตามด้วยการไดเลชัน

$$\text{Morphological opening: } \gamma_n(f) = \delta_n(\epsilon_n(f)) \quad (3.3)$$

การปฏิบัติการแบบเปิด จะกระทำการอีโรชัน แล้วตามด้วยการไดเลชัน ด้วยองค์ประกอบโครงสร้างเดียวกัน ซึ่งจะกำจัดส่วนสว่าง (จุดภาพที่มีค่าเท่ากับ 1) ที่อยู่ในบริเวณที่มีมืด (จุดภาพที่มีค่าเท่ากับ 0) โดยเปรียบเทียบกับองค์ประกอบโครงสร้าง ถ้าบริเวณที่สว่างที่พิจารณา อยู่ภายในองค์ประกอบโครงสร้าง จะถูกกำจัดออก

$$\text{Morphological closing : } \varphi_n(f) = \epsilon_n(\delta_n(f)) \quad (3.4)$$

การปฏิบัติการแบบปิด จะกระทำการไดเลชัน แล้วตามด้วยการอีโรชัน ด้วยองค์ประกอบโครงสร้างเดียวกัน ซึ่งจะกำจัดส่วนมืด ที่อยู่ภายในบริเวณที่สว่าง โดยเปรียบเทียบกับองค์ประกอบโครงสร้าง

การปฏิบัติการแบบเปิด/ปิดเชิงสัจฐานนี้จะทำให้ภาพมีความราบเรียบมากขึ้นโดยการกำจัดส่วนประกอบของจุดภาพที่สว่างในบริเวณที่มีมืด และกำจัดส่วนประกอบของจุดภาพที่มีมืด ในบริเวณที่สว่าง ที่มีขนาดเล็กกว่าขนาดขององค์ประกอบโครงสร้าง ตัวกรองชนิดนี้สามารถทำให้ภาพมีความราบเรียบก่อนการแยกส่วนภาพได้ แต่ไม่สามารถรักษาขอบเขตของภาพวัตถุให้สมบูรณ์ได้ จึงต้องมีการปรับปรุงตัวกรองโดยให้สามารถรักษาขอบเขตของภาพวัตถุได้ด้วย

การปฏิบัติการด้วยตัวปฏิบัติการแบบปิด/เปิดโดยการสร้างใหม่ (morphological opening / closing by reconstruction)

ตัวปฏิบัติการนี้ใช้สำหรับกรณีที่ภาพมีค่าความเข้มที่แปรเปลี่ยนขึ้นๆลงๆไม่ราบเรียบ ส่วนประกอบความสว่างและส่วนประกอบความมืดควรจะนำมาทำให้ราบเรียบขึ้นโดยที่ยังรักษาขอบเขตของวัตถุ ซึ่งตัวกรองเชิงสัจฐานแบบเปิด/ปิดโดยการสร้างใหม่ (morphological opening / closing by reconstruction) จะกรองบริเวณที่มีขนาดเล็กกว่าขนาดที่กำหนดไว้ จึงยังรักษาขอบเขตของภาพวัตถุไว้

นิยามตัวปฏิบัติการเชิงสัจฐาน (morphological operator)

การปฏิบัติการ Geodesic นิยามด้วยฟังก์ชันอ้างอิง (r)

การปฏิบัติการ Geodesic ขนาดเป็นหนึ่ง มีนิยามดังนี้

$$\text{Geodesic dilation of size one : } \delta^{(1)}(f,r) = \text{Min} \{ \delta_1(f,r) \} \quad (3.5)$$

$$\text{Geodesic erosion of size one : } \epsilon^{(1)}(f,r) = -\delta^{(1)}(-f,-r) \quad (3.6)$$

ส่วนการปฏิบัติการ Geodesic ด้วยขนาดเป็นอนันต์ จะเรียกว่า การปฏิบัติการโดยการสร้างใหม่ (reconstruction)

Reconstruction by dilation :

$$\gamma^{(rec)}(f,r) = \delta^{(\infty)}(f,r) = \dots \delta^{(1)}(\dots \delta^{(1)}(f,r) \dots, r) \quad (3.7)$$

Reconstruction by erosion :

$$\varphi^{(rec)}(f, r) = \mathcal{E}^{(\infty)}(f, r) = \dots \mathcal{E}^{(1)}(\dots \mathcal{E}^{(1)}(f, r) \dots, r) \quad (3.8)$$

การปฏิบัติการด้วยตัวปฏิบัติการแบบเปิด(ปิด)โดยการสร้างใหม่ของอีโรชัน(ไดเลชัน)

Opening by reconstruction of erosion : $\gamma^{(rec)}(\mathcal{E}_n(f), f) \quad (3.9)$

Closing by reconstruction of dilation : $\varphi^{(rec)}(\mathcal{D}_n(f), f) \quad (3.10)$

ภาพที่ได้จากกระบวนการแบบเปิด/ปิดโดยการสร้างใหม่จะมีความราบเรียบมากกว่าภาพต้นแบบ และยังรักษาข้อมูลขอบของภาพวัตถุอยู่ แต่ยังมีปัญหาสำหรับวัตถุที่เคลื่อนที่อยู่ จึงต้องแก้ปัญหานี้โดยการทำให้ราบเรียบในแนวแกนเวลาด้วย โดยการทำให้ฟังก์ชันอ้างอิงราบเรียบด้วยตัวปฏิบัติการแบบเปิด/ปิดที่มีขนาดเล็ก

การปฏิบัติการด้วยตัวกรองเชิงสัณฐานแบบเปิด/ปิดโดยการสร้างใหม่บางส่วน

(morphological opening / closing by partial reconstruction)

ตัวปฏิบัติการด้วยตัวกรองโดยการสร้างใหม่บางส่วน สามารถแก้ปัญหตามแนวแกนเวลาทำให้ภาพมีความราบเรียบ ซึ่งนิยามได้ดังนี้

Opening by partial reconstruction :

$$\gamma^{(rec)}(\mathcal{E}_n(f), \gamma_k(f)) \quad ; 0 \leq k \leq n \quad (3.11)$$

Closing by partial reconstruction :

$$\varphi^{(rec)}(\mathcal{D}_n(f), \varphi_k(f)) \quad ; 0 \leq k \leq n \quad (3.12)$$

โดยที่ n คือ ขนาดขององค์ประกอบโครงสร้าง และ k คือ ตัวปรับขนาดของการสร้างใหม่

ถ้า $k = n$ จะไม่มีการสร้างใหม่ และ ถ้า $k = 0$ จะเป็นการสร้างใหม่เต็มที่

ตัวกรองเชิงสัณฐานแบบเปิด/ปิดโดยการสร้างใหม่บางส่วน (morphological opening / closing by partial reconstruction) [Salembier P., Pardas M.; 1994] จะทำให้บริเวณแต่ละส่วนมีความชัดเจนยิ่งขึ้นและช่วยในการแยกส่วนบริเวณในขั้นถัดไปด้วย ซึ่งขนาดของส่วนประกอบโครงสร้างที่ใช้ จะขึ้นกับขนาดและรูปแบบของภาพ ซึ่งภาพขนาด (QCIF) 144x176 จุดภาพจะใช้องค์ประกอบโครงสร้างขนาด 5x5 จุดภาพ และสำหรับฟังก์ชันอ้างอิงขนาด 3x3 จุดภาพ

การประเมินผลการสร้างความเรียบง่ายและการแยกส่วนของภาพ

1. ความราบเรียบ (Flatness)

พารามิเตอร์นี้จะวัดความแปรปรวนของภาพซึ่งจะบ่งชี้ประสิทธิภาพในการสร้างความราบเรียบของภาพ ถ้ามีค่าความแปรปรวนต่ำจึงจะดี

2. การกำหนดตำแหน่งของขอบ (Edge Localization)

พารามิเตอร์นี้จะวัดจำนวนจุดภาพที่แตกต่างกันระหว่างภาพที่ได้จากการแยกส่วนและภาพอ้างอิงต่อพื้นที่ของภาพวัตถุจากการแยกส่วน ซึ่งจะเป็นการบ่งชี้การรักษาข้อมูลขอบของภาพวัตถุ ถ้ามีค่าความแตกต่างก็น้อยยิ่งจะดี

3.3.2.2 การประมาณเกรเดียนต์ (Gradient Approximation)

ภาพที่ผ่านกระบวนการขึ้นต้นมาแล้วจะถูกนำมาหาค่าเกรเดียนต์โดยใช้ morphological gradient operator [Kim, M., Choi, J. G., Ho, Y. S., 1999] ในบริเวณส่วนที่คล้ายคลึงกันจะมีค่าเกรเดียนต์น้อย ส่วนในบริเวณขอบของภาพวัตถุที่มีความแตกต่างของบริเวณจะให้ค่าเกรเดียนต์มาก Morphological Gradient Operator (MGO) สามารถแสดงได้ดังนี้

$$3.3.2.2.1 \text{ Morphological gradient : } G(f) = \delta(f, SE) - \varepsilon(f, SE) \quad (3.13)$$

ค่าเกรเดียนต์ เท่ากับ ค่าโดเลชันของภาพลบด้วยค่าอีโรชันของภาพ ดังสมการที่ (3.13) ค่านี้จะมีค่าเป็นบวกเสมอและใช้ระบุบริเวณขอบของภาพวัตถุ ซึ่งสามารถใช้กรรมวิธีสันปันน้ำในการแยกส่วนภาพตามลักษณะขอบของภาพวัตถุได้

โดยที่ f คือ ความเข้มของภาพ และ SE คือ องค์ประกอบโครงสร้าง

สำหรับภาพที่มีความเบี่ยงเบนของความเข้มน้อย บริเวณข้างเคียงอาจทำให้ขอบเขตของภาพวัตถุกำกวมได้ เนื่องจากมีค่าเกรเดียนต์น้อย เพื่อแก้ปัญหานี้จะใช้ข้อมูลสีในการคำนวณเกรเดียนต์และให้ค่าน้ำหนักถ่วงสำหรับแต่ละส่วนประกอบของสี โดยจะเลือกค่าที่มากที่สุดของค่าเกรเดียนต์นี้เป็นขอบเขตของภาพวัตถุ

$$G_i = \max\{w_1 G_y, w_2 G_{cr}, w_3 G_{cb}\} \quad (3.14)$$

โดยที่ G_i คือ ค่าเกรเดียนต์ของขอบภาพ

G_y คือ ค่าเกรเดียนต์ขององค์ประกอบ Y

G_{cr} คือ ค่าเกรเดียนต์ขององค์ประกอบ Cr

G_{cb} คือ ค่าเกรเดียนต์ขององค์ประกอบ Cb

w_1, w_2, w_3 คือ ค่าน้ำหนักถ่วงขององค์ประกอบ Y, Cr, Cb ตามลำดับ

เนื่องจากกรรมวิธีสันปันน้ำจะทำให้เกิดการแยกส่วนมากเกินไป ดังนั้นค่าเกรเดียนต์ของภาพจึงต้องตั้งค่าจุดเปลี่ยนจำกัดเอาไว้ ถ้าค่าเกรเดียนต์นี้มีค่าน้อยกว่าค่าจุดเปลี่ยนจะให้มีค่าเป็น 0 ทั้งหมด ทำให้บริเวณเกรเดียนต์ระดับต่ำ ๆ จะราบเรียบทั้งหมด ส่วนที่เหลือจะให้ค่าเท่าเดิม

จากงานวิจัยของ Yo Sung Ho [Kim M., Choi J. G., Ho Y. S.; 1999] ค่าจุดเปลี่ยนเท่ากับ 10 และค่าน้ำหนักเท่ากับ 1,2,2 ตามลำดับ จะให้ผลที่ดี เนื่องจากจะทำให้ค่าเกรเดียนต์ของความเข้มสีมีค่าใกล้เคียงกับค่าเกรเดียนต์ของความเข้มแสงมากขึ้น

3.3.2.2 การปฏิบัติการเกรเดียนต์แบบหลายชั้น (Multiscale Gradient Operation)

เมื่อทำการปฏิบัติการเกรเดียนต์แล้ว จะเกิดเส้นขอบขึ้นสูงและความกว้างของขอบจะขึ้นกับขนาดขององค์ประกอบโครงสร้าง ถ้าใช้องค์ประกอบโครงสร้างขนาดใหญ่จะเกิดความกว้างของขอบใหญ่ แต่ถ้าใช้ขนาดเล็กจะตรวจรู้เส้นขอบได้ยาก และเมื่อผ่านกระบวนการสร้างความราบง่ายของภาพแล้วจะเกิดการสูญเสียของข้อมูลขอบด้วย ดังนั้นเพื่อที่จะลดการแยกส่วนภาพมากเกินไปและรักษาขอบเขตของภาพวัตถุรวมทั้งรายละเอียดของภาพไว้ จะเพิ่มการปรับขนาดขององค์ประกอบโครงสร้างเข้าไปในตัวปฏิบัติการเกรเดียนต์ ซึ่งจะเรียกว่า ตัวปฏิบัติการเกรเดียนต์แบบหลายชั้น โดยการเพิ่มขนาดขององค์ประกอบโครงสร้างขึ้นแกนละ 2 จุดภาพในแต่ละรอบการคำนวณ

ตัวปฏิบัติการเกรเดียนต์หลายชั้น (multiscale gradient operator) [Wang D.; 1997]

$$MG(f) = \frac{1}{n} \sum_{i=1}^n [\mathcal{G}(f, SE_i) - \mathcal{G}(f, SE_{i-1})]; \quad 1 \leq i \leq n \quad (3.15)$$

โดยที่ SE_i เป็นองค์ประกอบโครงสร้างขนาด $(2i+1) \times (2i+1)$

n เป็นลำดับชั้นของการประมาณเกรเดียนต์

ตัวปฏิบัติการนี้จะสร้างเส้นขอบขนาดสองจุดภาพ ทำให้เกิดจุดภาพที่ไม่ระบุบริเวณเพียงเล็กน้อยเท่านั้น แต่จะมีปัญหาในกรณีที่ตั้งค่าจุดเปลี่ยนของการตัดค่าเกรเดียนต์ให้มีค่าต่ำสำหรับภาพที่มีข้อมูลขอบหนาแน่น จะเกิดขอบจำนวนมาก ดังนั้นจึงต้องตั้งค่าจุดเปลี่ยนให้สูงขึ้นเพื่อรักษาขอบเขตของบริเวณเอาไว้ในการใช้ตัวปฏิบัติการนี้

3.3.2.3 กรรรมวิธีสันปันน้ำ (Watershed Algorithm)

หลังจากได้แนวขอบเขตของภาพวัตถุแล้ว จะต้องกำหนดจุดภาพทั้งหมดของภาพให้เป็นบริเวณที่มีความคล้ายคลึงกันของความเข้มแสง ซึ่งจะใช้หลักการของการปันน้ำ โดยเริ่มจากจุดภาพที่ต่ำที่สุดของแต่ละบริเวณ แล้วขยายบริเวณเพิ่มขึ้นไปจนถึงขอบของสันปันน้ำ แล้วจึงกำหนดบริเวณให้กับจุดภาพของบริเวณนั้น

3.3.3 การแยกส่วนภาพเชิงเวลา (Temporal Segmentation)

3.3.3.1 การวิเคราะห์การเคลื่อนที่ของลำดับภาพ (Motion Analysis)

การประมาณการเคลื่อนที่เป็นเครื่องมือช่วยในการวิเคราะห์แนวการเคลื่อนที่ของจุดภาพแต่ละเฟรม โดยวิธีที่รวดเร็วที่สุดและได้ผลที่ยอมรับได้ คือ วิธีการจับคู่บล็อก (BMA)

วิธีการจับคู่บล็อกนี้ มีกรรมวิธีในการค้นหาคู่บล็อกที่มีความคล้ายกันมากที่สุดของภาพในเฟรมที่ต่อเนื่องกันอยู่หลายวิธี แต่วิธีที่ให้ความถูกต้องในการวิเคราะห์การเคลื่อนที่ที่ดีที่สุด คือ กรรมวิธีแบบค้นหาทั้งหมด

กรรมวิธีแบบค้นหาทั้งหมด (Full Search Algorithm) มีวิธีการดังนี้

ค่าเวกเตอร์การเคลื่อนที่ที่เป็นไปได้ทั้งหมดจะถูกนำมาคิดค่าความเพี้ยนต่อบล็อกเพื่อเปรียบเทียบหาค่าที่ต่ำสุด โดยให้ขนาดของหน้าต่างการค้นหาเวกเตอร์การเคลื่อนที่ทั้งแกน x และ แกน y จาก $-w$ ถึง $+w$ ซึ่งจะมีขนาดหน้าต่างการค้นหาขนาด $(2w+1) \times (2w+1)$ จุดภาพ

การวัดค่าความเพี้ยนต่อบล็อกจะใช้ค่าความผิดพลาดสัมบูรณ์เฉลี่ย และจะได้เวกเตอร์การเคลื่อนที่หนึ่งค่า สำหรับบล็อกขนาด 8×8 จุดภาพ สำหรับภาพขนาด QCIF (144×176) จะได้เวกเตอร์ทั้งหมด 396 ค่า เวกเตอร์การเคลื่อนที่นี้จะช่วยในการวิเคราะห์ภาพวัตถุ และใช้ในการแยกวัตถุที่เคลื่อนที่ทิศทางแตกต่างกัน

ในการวิเคราะห์ภาพวัตถุ เวกเตอร์ที่ได้จากการประมาณการเคลื่อนที่จะใช้ในการกำจัดพื้นที่ส่วนพื้นหลังที่ไม่ถูกรวมคลุมของภาพ จากข้อมูลการตรวจจับการเปลี่ยนแปลงความเข้ม โดยจะสร้างมาสก์ของวัตถุขึ้นต้นจากการเปลี่ยนแปลงของภาพ จากนั้นจะพิจารณาเฉพาะจุดภาพที่มีการเปลี่ยนแปลง ถ้าทั้งจุดเริ่มต้นและจุดสิ้นสุดของเวกเตอร์การเคลื่อนที่จากการประมาณอยู่ในบริเวณที่มีการเปลี่ยนแปลง จุดภาพนั้นจะกำหนดให้เป็นส่วนฉากหน้า (foreground) แต่ถ้าจุดใดจุดหนึ่งไม่อยู่จะกำหนดให้เป็นของส่วนพื้นหลัง (background)

3.3.3.2 การตรวจจับการเปลี่ยนแปลง (Change Detection)

การตรวจจับการเปลี่ยนแปลงจะสังเกตจากการเปลี่ยนแปลงของภาพในเฟรมที่ต่อเนื่องกัน โดยจะพิจารณาจากจุดภาพที่การเปลี่ยนแปลงความเข้มระหว่างเฟรมซึ่งเกิดจากการเคลื่อนที่ของวัตถุในลำดับภาพ

การตรวจจับการเปลี่ยนแปลงในวิทยานิพนธ์นี้ใช้การตรวจจับ 2 แบบด้วยกัน คือ

(1) กรรมวิธีการตรวจจับการเปลี่ยนแปลงเป็นบล็อกใช้ค่า MAD สำหรับการตัดสินใจ

การตรวจจับการเปลี่ยนแปลงนี้จะกำหนดเป็นบล็อกของจุดภาพขนาด $M \times N$ และพิจารณาการเปลี่ยนแปลงจากค่าแตกต่างสัมบูรณ์เฉลี่ย (MAD) ระหว่างเฟรมสองเฟรมที่ต่อเนื่องกันโดยใช้ค่า MAD เปรียบเทียบกับค่าจุดเปลี่ยน (T) ถ้าค่า MAD ของบล็อกมีค่ามากกว่า T จะให้จุดภาพภายในบล็อกทั้งหมดเป็นจุดภาพที่มีการเปลี่ยนแปลง เรียกบล็อกนี้ว่า บล็อกของการตรวจจับการเปลี่ยนแปลง (CDB)

ค่าแตกต่างสัมบูรณ์เฉลี่ย หาได้จากสมการ (3.16)

$$MAD(i, j) = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N |X(m, n) - X^R(m+i, n+j)| \quad (3.16)$$

โดยที่ $X(m, n)$ เป็นค่าความเข้มของภาพปัจจุบันที่จุด (m, n)
 $X^R(m, n)$ เป็นค่าความเข้มของภาพก่อนหน้าที่จุด (m, n)
 $M \times N$ เป็นขนาดของบล็อกที่ตรวจจับการเปลี่ยนแปลง

$$CDB = \begin{cases} 1 & \text{if } MAD \geq T \\ 0 & \text{if } MAD < T \end{cases} \quad (3.17)$$

โดยที่ CDB คือ ค่าความเข้มของจุดภาพภายในบล็อก

T คือ ค่าจุดเปลี่ยน

สำหรับการตรวจรู้การเปลี่ยนแปลงของจุดภาพเพื่อหาマスクของวัตถุที่เปลี่ยนแปลง จะต้องตั้งค่าจุดเปลี่ยนทั่วไป (global threshold) คือ ค่าความแตกต่างระหว่างเฟรม และใช้ค่าจุดเปลี่ยนท้องถิ่น (local threshold) [Mech R., Wollborn M., 1997] เพื่อปรับแต่งความเรียบของส่วนขอบที่เปลี่ยนแปลงไป ซึ่งจะสังเกตได้ว่าบริเวณของภาพวัตถุในเฟรมก่อนหน้านี้จะเป็นบริเวณของマスクของการตรวจรู้การเปลี่ยนแปลงนี้ด้วย และสุดท้ายจะต้องกำจัดส่วนที่ตรวจรู้ผิดพลาดซึ่งมีขนาดเล็กที่บริเวณพื้นหลังออกไป เพื่อให้มีเฉพาะマスクของภาพวัตถุเท่านั้น

(2) กรรรมวิธีการตรวจรู้การเปลี่ยนแปลงของแต่ละจุดภาพ

สามารถแบ่งได้เป็น 2 ขั้นตอน คือ [Mech R., Wollborn M., 1997]

ขั้นตอนแรก มาส์กของการตรวจรู้การเปลี่ยนแปลง (CDM) ได้จากการพิจารณาทุกๆ จุดภาพที่ค่าความเข้มของแสงมีการเปลี่ยนแปลงไปเนื่องจากการเคลื่อนที่ของวัตถุ

$$CDM_i(x, y) = \begin{cases} 1 & \text{if } d_i \geq T_i \\ 0 & \text{if } d_i < T_i \end{cases} \quad (3.18)$$

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

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

T คือ ค่าจุดเปลี่ยนที่มีการปรับตัวอย่างอัตโนมัติตามการเปลี่ยนแปลงความเข้มของภาพ

โดยให้ค่าเริ่มต้นของค่าจุดเปลี่ยนมีค่าเท่ากับ 10 เปอร์เซ็นต์ของค่าความเข้มสูงสุดในภาพ

T_i มีค่าเท่ากับ ค่าความแตกต่างของความเข้มระหว่างเฟรมที่ i และเฟรมที่ i-1 เฉลี่ย

เพื่อที่จะให้ได้บริเวณของภาพวัตถุที่มีเสถียรภาพ CDM จะถูกปรับให้สัมพันธ์กับマスクของภาพวัตถุ (OM) ของเฟรมที่แล้ว โดยจะให้ทุกจุดภาพภายใน OM ของเฟรมก่อนหน้านี้ เป็นจุดภาพที่เปลี่ยนแปลงของ CDM ของเฟรมปัจจุบัน ซึ่งอยู่บนสมมติฐานที่ว่าทุกๆ จุดภาพของマスクของภาพวัตถุในเฟรมก่อนหน้านี้จะเป็นจุดภาพของマスクของการตรวจรู้การเปลี่ยนแปลงในเฟรมปัจจุบัน แต่เพื่อป้องกันการเกิดความผิดพลาดขยายเพิ่มขึ้น จึงต้องมีการตรวจสอบโดยเก็บマスクของภาพวัตถุไว้ในหน่วยความจำ (memory) โดยจุดภาพของ OM ในเฟรมก่อนหน้านี้จะเปลี่ยนแปลงก็ต่อเมื่อมีการเปลี่ยนแปลงใน CDM อย่างน้อยใน 1 เฟรมของหน่วยความจำ จึงต้องมีการเก็บマスクในหน่วยความจำ ซึ่งจะมีการปรับหน่วยความจำ ตามสมการ

$$MEM_{(k)}(x, y) = \begin{cases} L & \text{if } CDMs_{(k)}(x, y) = 1 \\ \text{Max}(0, MEM_{(k-1)}(x, y) - 1) & \text{if } CDMs_{(k)}(x, y) = 0 \end{cases} \quad (3.19)$$

โดยที่ L คือ ความยาวของหน่วยความจำ (เฟรม)

$$CDMu_{(k)}(x, y) = CDMs_{(k)}(x, y) \vee \begin{cases} OM_{(k-1)}(x, y), & \text{if } MEM_{(k)}(x, y) > 0 \\ 0 & \text{if } MEM_{(k)}(x, y) = 0 \end{cases} \quad (3.20)$$

โดยที่ CDMu คือ CDM ที่ปรับแล้วโดยตัวปฏิบัติการ OR (\vee)

หลังจากนั้น 마스크ของการตรวจรู้การเปลี่ยนแปลงสุดท้ายจะทำให้มีรูปแบบที่ง่ายโดยใช้ตัวปฏิบัติการเชิงสัณฐานแบบปิด (morphological closing operator) และกำจัดบริเวณที่มีจำนวนจุดภาพน้อยกว่าจำนวนที่กำหนดออกไป ซึ่งอยู่บนสมมติฐานที่ว่าบริเวณที่มีขนาดเล็กไม่ควรจัดเป็นภาพวัตถุ และมีแนวโน้มที่จะเป็นจุดรบกวนได้มาก

ขั้นตอนที่สอง พื้นหลังที่ไม่ถูกรอบคลุม (UBG) ในส่วนของ 마스크ของการตรวจรู้การเปลี่ยนแปลง (CDM) จะถูกกำจัดออก แล้วจะได้ 마스크ของทุกๆ จุดภาพที่เป็นจุดภาพของวัตถุที่เคลื่อนที่ เรียกว่า 마스크ของภาพวัตถุ (OM)

การสร้าง 마스크ของภาพวัตถุจาก 마스크ของการตรวจรู้การเปลี่ยนแปลง จะกำจัดบริเวณของพื้นหลังที่ไม่ถูกรอบคลุมออก โดยใช้ข้อมูลของการเคลื่อนที่ของจุดภาพภายในบริเวณที่เปลี่ยนแปลง ถ้าจุดภาพที่มีเวกเตอร์ตรงข้ามกับเวกเตอร์ของจุดภาพนอก CDM จะถูกตั้งสมมติฐานให้เป็นบริเวณของ UBG ดังนั้นจะสามารถแบ่งบริเวณ CDM ได้ 2 ส่วน คือ บริเวณที่เป็นส่วนของพื้นหลังที่ไม่ถูกรอบคลุม (UBG) และบริเวณของภาพวัตถุ (OM)

3.3.4 การนำทางการเคลื่อนที่ของวัตถุ (Object Tracking)

หลังจากที่ผ่านกระบวนการการแยกส่วนภาพวัตถุในเฟรมหนึ่งแล้ว จะได้ 마스크ของภาพวัตถุ ซึ่งจะนำมาใช้เป็นข้อมูลในการแยกส่วนภาพวัตถุในเฟรมถัดไป โดยจะนำข้อมูลของ 마스크ของภาพวัตถุที่ได้เข้ามารวมใช้ในกระบวนการแยกส่วนภาพของเฟรมปัจจุบัน ซึ่งจะแทนด้วย 마스크ของการขดเซยการเคลื่อนที่ ในส่วนนี้นอกจากจะช่วยในการแยกส่วนภาพแล้วยังช่วยในการนำทาง (tracking) การเคลื่อนที่ของแต่ละวัตถุและบอกสถานะของวัตถุในส่วนของวัตถุที่ปรากฏขึ้นใหม่ได้อีกด้วย

3.3.5 กฎการตัดสินใจ (Decision Rule)

กระบวนการนี้จะพิจารณาข้อมูลของมาสก์ในแต่ละส่วนในกระบวนการตัดสินใจ ซึ่งนำเอาคุณสมบัติที่แตกต่างกันมาใช้ ในขั้นตอนการแยกส่วนโดยใช้ข้อมูลเชิงพื้นที่จะให้จำนวนบริเวณที่แตกต่างกันเป็นจำนวนมาก แต่จะให้ข้อมูลส่วนของขอบภาพวัตถุที่ค่อนข้างแม่นยำ ขณะที่ข้อมูลเชิงเวลาจะให้ข้อมูลเพิ่มเติมสามารถบอกการเปลี่ยนแปลงและการเคลื่อนที่ของภาพวัตถุ ข้อมูล

ของการเปลี่ยนแปลงความเข้มจะให้ข้อมูลส่วนของบริเวณขอบไม่ชัดเจน แต่จะระบุบริเวณที่มีการเคลื่อนที่ของวัตถุและบริเวณที่ไม่มีเคลื่อนที่หรือเป็นส่วนของพื้นหลังได้ ส่วนข้อมูลของการวิเคราะห์การเคลื่อนที่จะสามารถระบุตำแหน่งและทิศทางของแต่ละวัตถุได้ กระบวนการตัดสินใจนี้จะใช้ข้อมูลเชิงพื้นที่ในการกำหนดขอบเขตของบริเวณ ซึ่งจะสร้างมาสก์ของวัตถุโดยการรวมบริเวณของแต่ละส่วนความเข้มที่ได้จากการแยกส่วนเชิงพื้นที่เข้าด้วยกันตามข้อมูลเชิงเวลา

การพิจารณาสร้างมาสก์ของภาพวัตถุ จะใช้ทั้งบริเวณที่ได้จากข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา แต่เนื่องจากการแยกส่วนเชิงพื้นที่ จะทำให้ขอบเขตมีความถูกต้องมากขึ้น จึงใช้เป็นบริเวณหลักสำหรับทำมาสก์ให้สมบูรณ์สำหรับแสดงขอบเขตของภาพวัตถุ ซึ่งมีขั้นตอนดังนี้

[Alatan Aydin Alatan, Tuncel E., Onural L., 1997]

กรณีจุดภาพภายในบริเวณที่ได้จากการแยกส่วนเชิงพื้นที่ถูกตรวจสอบว่าเป็นฉากหน้า (foreground) ในมาสก์ของการตรวจรู้การเปลี่ยนแปลง มากกว่าครึ่งหนึ่งของทั้งบริเวณพื้นที่นั้น ทุกจุดภาพของบริเวณพื้นที่นั้นจะกำหนดให้เป็นฉากหน้า ส่วนกรณีจุดภาพภายในบริเวณจากการแยกส่วนเชิงพื้นที่ถูกตรวจสอบว่าเป็นฉากหน้าซึ่งอยู่ในมาสก์ของการตรวจรู้การเปลี่ยนแปลง น้อยกว่าครึ่งหนึ่งของบริเวณพื้นที่นั้น ทุกจุดภาพของบริเวณพื้นที่นั้นจะกำหนดให้เป็นพื้นหลัง

3.3.6 กระบวนการภายหลัง (Postprocessing)

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

มาสก์ของภาพวัตถุที่ได้จะมีค่าเป็นจำนวนบวก ซึ่งแต่ละค่าจะแทนวัตถุที่แตกต่างกัน ส่วนพื้นหลังของภาพจะมีค่าเป็นศูนย์ โดยที่มาสก์ในเฟรมแรกจะมีค่าเป็นศูนย์ทั้งหมดเนื่องจากยังไม่มีการแยกส่วนของภาพเกิดขึ้น

3.4 การประเมินประสิทธิภาพของกรรมวิธีการแยกส่วนภาพวัตถุ

การประเมินประสิทธิภาพของการแยกส่วนภาพวัตถุในวิทยานิพนธ์ฉบับนี้จะเปรียบเทียบกับผลจาก COST211 ซึ่งเป็นกลุ่มวิจัยที่ทำวิจัยหลักทางด้าน การแยกส่วนภาพของลำดับภาพ

3.4.1 ขอบเขตงานของ EUROPEAN COST 211

กลุ่ม COST 211 เป็นกลุ่มผู้ร่วมวิจัยของสมาชิกกลุ่มประเทศยุโรป ซึ่งมีจุดมุ่งหมายหลัก เพื่อที่จะพัฒนาวิธีการสำหรับการแยกส่วนของลำดับภาพ ทั้งแบบอัดโน้มนัดและกึ่งอัดโน้มนัดโดยให้มีการเชื่อมโยงกับผู้ใช้ เพื่อสนับสนุนการประยุกต์ใช้งานใน MPEG-4

แนวทางของการให้บริการและการใช้งานของกลุ่ม COST 211 ประกอบด้วย

- การสื่อสารแบบเวลาจริง เช่น การบริการโต้ตอบกัน หรือ การประชุมทางวิดีโอ
- การให้บริการเข้าถึงข้อมูล เช่น MPEG-4 วิดีโอหรือฮอดิโอเพื่อเก็บไว้บนฐานข้อมูล
- การให้บริการกระจายข่าวโต้ตอบกันได้

3.4.1.1 โครงการทำงานของ COST โดยทั่วไป

กิจกรรมของ COST ยุโรป เริ่มต้นขึ้นโดยกลุ่มสหภาพยุโรป (European Community : EU) โดยการหาความร่วมมือด้านการวิจัยและพัฒนา (R&D) ในกลุ่มประเทศยุโรป การดำเนินงานของ COST นำไปสู่พื้นฐานของงานวิจัยในหัวข้อวิจัยเฉพาะ ที่มีความสำคัญสำหรับการพัฒนาของวงการด้านข่าวสาร (information society)

3.4.1.2 ความเป็นมาของ COST 211

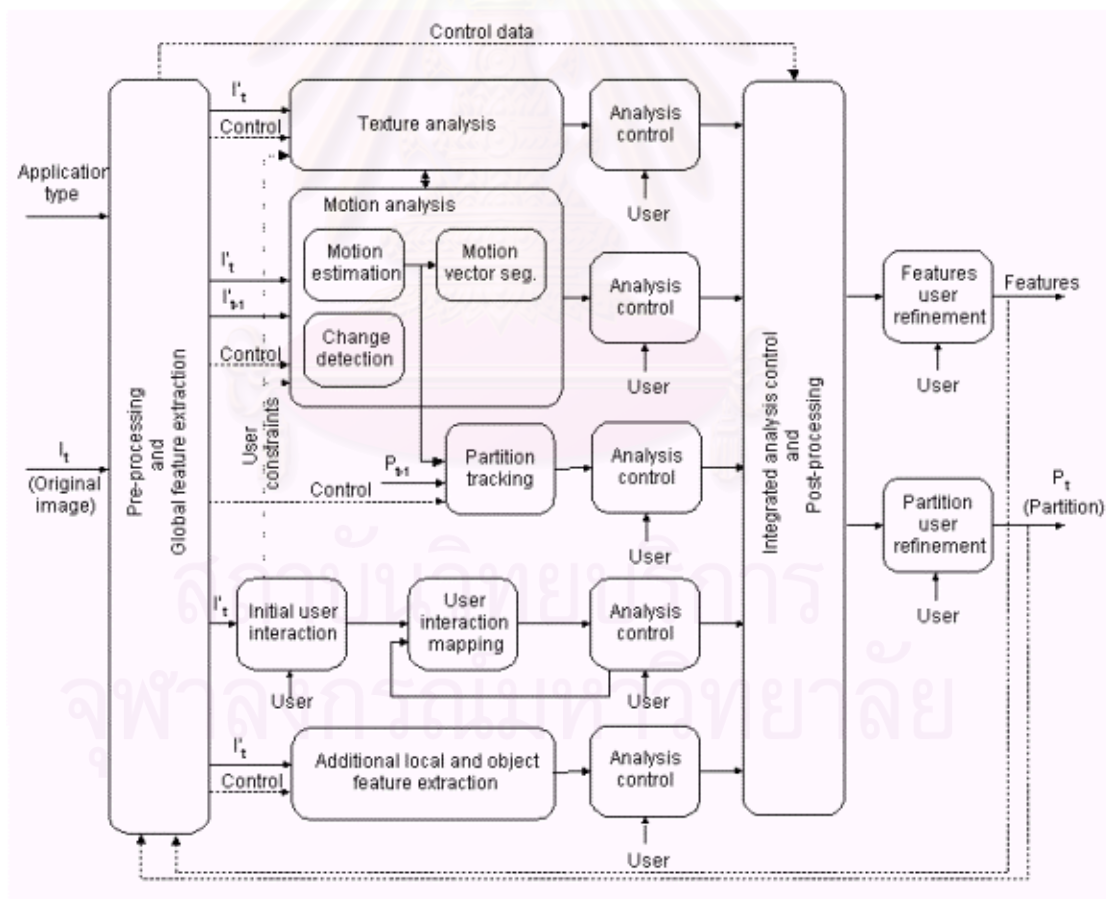
การดำเนินงานของ COST 211^{ter} ได้ดำเนินโครงการต่อเนื่องจาก COST 211 และ COST 211^{bis} โดยการดำเนินการทั้งหมดเกี่ยวกับเทคนิคการลดความซ้ำซ้อนของสัญญาณวิดีโอ โครงการของ COST 211 ที่ดำเนินมาแล้ว คือ การสร้างสรรค์วิธีการในการเข้ารหัสสัญญาณวิดีโอ และเผยแพร่เป็นมาตรฐานต่อไป ซึ่งปัจจุบันมาตรฐานของสัญญาณดิจิทัลวิดีโอได้รับเอาผลจากการดำเนินงานของ COST 211 มาใช้ในสาขาของการเข้ารหัสสัญญาณวิดีโอ

การดำเนินงานของ COST 211 เริ่มขึ้นเมื่อปี 1976 ตอนที่การพัฒนาทางด้านสัญญาณดิจิทัลวิดีโอเพิ่งจะเริ่มเกิด ผลงานของการดำเนินงาน คือ กรรมวิธีการเข้ารหัสสัญญาณวิดีโอสำหรับการประชุมทางวิดีโอเริ่มแรกที่อัตราบิต 2 Mbit/s และนำไปสู่มาตรฐานแรกของ ITU ทางด้านนี้ คือ H.120 ต่อมา COST 211^{bis} ได้ปรับปรุงประสิทธิภาพของกรรมวิธีการเข้ารหัสสำหรับโทรศัพท์ภาพ และการประชุมทางวิดีโอที่อัตราของ ISDN แล้ว ITU ได้ร่างมาตรฐานเป็น H.261 นอกจากนั้น COST 211^{ter} ยังเป็นผู้สนับสนุนหลักของการดำเนินงานใน MPEG (MPEG 1,2,4) และ ITU ด้วย

โครงการของ COST ได้มีความร่วมมือกันมายาวนานมากกว่า 20 ปี ซึ่งมีอิทธิพลในสาขาของการเข้ารหัสสัญญาณวิดีโอ และยังมีการดำเนินงานของ COST 211^{quat} ที่จะดำเนินงานต่อไปอีก 5 ปีเริ่มตั้งแต่ปี 1998 ดังนั้นแบบจำลองและผลการวิเคราะห์ของ COST 211 จึงได้ใช้เป็นแบบจำลองมาตรฐานในการวัดประสิทธิภาพสำหรับกรรมวิธีการแยกส่วนภาพวัตถุที่เสนอขึ้น

3.4.2 การเปรียบเทียบผลการแยกส่วนภาพวัตถุกับแบบจำลองของ COST 211

ปัจจุบันเนื่องจากการกำเนิดขึ้นของมาตรฐาน MPEG-4 และ MPEG-7 การแยกส่วนภาพวัตถุของสัญญาณภาพและวิดีโอจึงมีความสำคัญมากขึ้น แต่สำหรับการแยกส่วนภาพของสัญญาณวิดีโอยังเป็นปัญหาอยู่สำหรับนักวิจัย ดังนั้น COST211 จึงได้สร้างกรรมวิธีพื้นฐานขึ้นเรียกว่า COST AM สำหรับการแยกส่วนภาพวัตถุเคลื่อนที่มีจุดประสงค์เพื่อเปรียบเทียบกรรมวิธีที่แตกต่างกันในการแยกส่วนภาพวัตถุ



รูปที่ 3.3 โครงสร้างแบบจำลองการแยกส่วนภาพวัตถุของ COST 211 (COST AM)

3.4.2.1 ข้อกำหนดของกรรมวิธีการแยกส่วนภาพวัตถุของ COST 211

1. กรรมวิธีที่ใช้ในการแยกส่วนภาพ

กรรมวิธีที่ใช้จะต้องเป็นกรรมวิธีการแยกส่วนแบบอัตโนมัติหรือกึ่งอัตโนมัติ โดยสามารถใส่พารามิเตอร์บางตัวได้ โดยจะต้องระบุค่าอย่างชัดเจน และจะต้องเป็นพารามิเตอร์ชุดเดียวสำหรับลำดับภาพที่ทดสอบ

กรรมวิธีที่เสนออาจจะเหมาะสมสำหรับรูปแบบการใช้งานเฉพาะอย่างหรือลักษณะของลำดับภาพบางรูปแบบได้

2. ลำดับภาพที่ใช้ทดสอบในการแยกส่วน

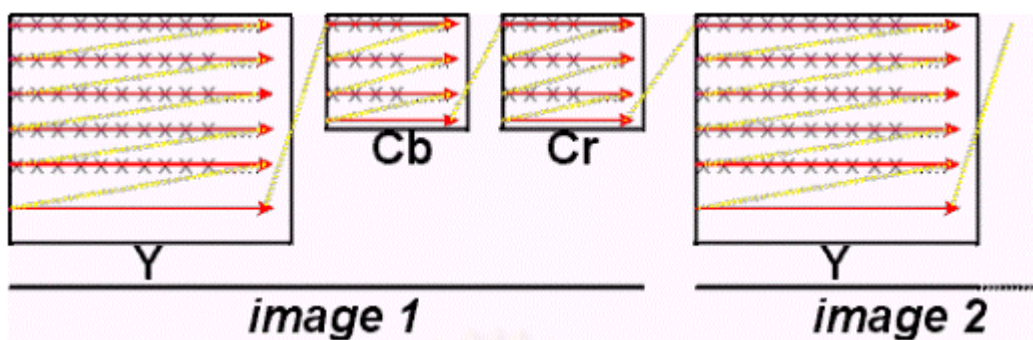
ลำดับภาพสำหรับการทดสอบจาก COST211 ที่มีลักษณะแตกต่างกัน มีดังตารางที่ 3.1

ตารางที่ 3.1 แสดงชื่อและลักษณะสำคัญของลำดับภาพที่ใช้ทดสอบกรรมวิธีการแยกส่วนภาพ

ชื่อของลำดับภาพ	วัตถุเป้าหมาย	ที่มาของลำดับภาพ	ข้อมูลเพิ่มเติม	ลักษณะของภาพ
Mother & Daughter	The mother, the daughter	AT&T	300 เฟรม, 30 Hz	พื้นหลังอยู่นิ่ง
Hall Monitor	The two men	Texas Instruments	300 เฟรม, 30 Hz	พื้นหลังอยู่นิ่ง
Table Tennis	The players, the ball	CCETT	300 เฟรม, 25 Hz	Pan and zoom
Erik	The person	Universitat Hannover	50 เฟรม, 12.5 Hz	พื้นหลังอยู่นิ่ง

ข้อมูลของลำดับภาพจะอยู่ในรูปแบบของ YCbCr (4:2:0) ซึ่งจะมีข้อมูลของความเข้มแสงและข้อมูลความเข้มสีอยู่ในรูปแบบขององค์ประกอบ Y,Cb,Cr ตามลำดับ โดยข้อมูลของความเข้มสีจะมีขนาดเล็กกว่าข้อมูลของความเข้มแสง การเรียงของเพิ่มข้อมูลแสดงดังรูปที่ 3.4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.4 รูปแบบของแฟ้มข้อมูลลำดับภาพ (file format of image sequence)

ส่วนมาส์กของการแยกส่วนภาพ จะอยู่ในรูปแบบ ขนาดเท่ากับขนาดของภาพทดสอบ แต่จะไม่มีส่วนของความเข้มสี ค่าของความเข้มแสงจะเป็น 0 เมื่อจุดภาพนั้นเป็นส่วนของพื้นหลัง และค่าที่เป็นบวกทั้งหมดจะเป็นจุดภาพของวัตถุเคลื่อนที่ โดยที่ค่าที่เท่ากันจะเป็นวัตถุเดียวกัน โดยมาส์กแรกของลำดับภาพจะให้เป็นมาส์กที่ทุกจุดภาพเป็น 0 ทั้งหมด

ลักษณะของลำดับภาพที่ทดสอบ

2.1 ลำดับภาพรูปแบบ CIF (352x288)



Hall Monitor



Mother & Daughter



Erik

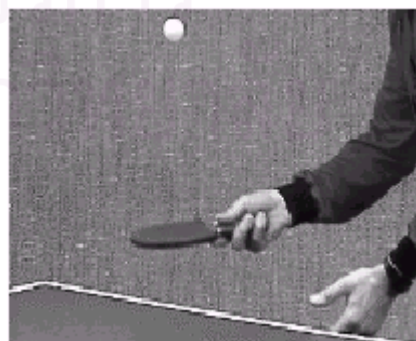


Table Tennis

รูปที่ 3.5 ลำดับภาพทดสอบรูปแบบ CIF

รูปแบบของลำดับภาพทดสอบ จะอยู่ในรูปแบบ CIF 4:2:0 โดยส่วนความเข้มแสงมีขนาด 352 X 288 จุดภาพ และส่วนความเข้มสีมีขนาด 176 X 144 จุดภาพ ลักษณะเป็น RAW file แต่ละจุดภาพมีขนาด 8 บิต (256 ระดับ) เรียงตามลำดับเฟรมและองค์ประกอบ Y, Cb, Cr ตามลำดับ

2.2 ลำดับภาพรูปแบบ QCIF (176x144)



Akiyo



Claire



Suzie



Trevor

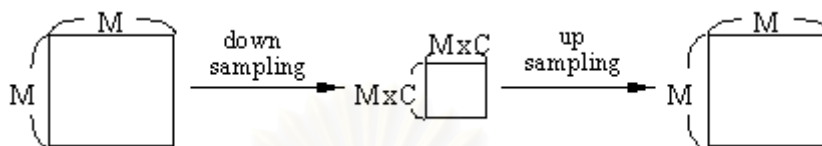
รูปที่ 3.6 ลำดับภาพทดสอบรูปแบบ QCIF

รูปแบบของลำดับภาพทดสอบ จะอยู่ในรูปแบบ QCIF 4:2:0 โดยส่วนความเข้มแสงมีขนาด 176 X 144 จุดภาพ และส่วนความเข้มสีมีขนาด 88 X 72 จุดภาพ ลักษณะเป็น RAW file แต่ละจุดภาพมีขนาด 8 บิต (256 ระดับ) เรียงตามลำดับเฟรมและองค์ประกอบ Y, Cb, Cr ตามลำดับ

2.3 การแปลงขนาดของลำดับภาพ (Size Conversion)

ในกรณีที่ลำดับภาพมีรูปแบบที่แตกต่างกัน และต้องการให้มีข้อมูลและขนาดของภาพเท่ากันเพื่อนำไปประมวลผลและเปรียบเทียบกัน จะต้องมีการแปลงขนาดของข้อมูล โดยเมื่อต้องการแปลงรูปแบบของภาพขนาดใหญ่ไปเป็นขนาดเล็ก จะใช้การสุ่มลง (down sampling) ส่วนเมื่อต้องการแปลงภาพขนาดเล็กไปเป็นขนาดใหญ่ จะใช้การสุ่มขึ้น (up sampling)

ในวิทยานิพนธ์นี้จะใช้ภาพรูปแบบ CIF และ QCIF ดังนั้นการแปลงขนาดของภาพจำเป็นจะใช้การส่มลงและการส่มขึ้นเพื่อลดขนาดและเพิ่มขนาดของภาพ การแปลงขนาดของภาพแสดงดังรูปที่ 3.7



รูปที่ 3.7 การแปลงขนาดของภาพ

3.4.2.2 เกณฑ์ในการประเมินผลการแยกส่วนภาพ [Villegas P. ; 1999]

1. Subjective Evaluation

การประเมินผลเชิงบุคคล จะใช้การสังเกตผลการแยกส่วนภาพของกรรมวิธีที่เสนอ เทียบกับผลจาก COST211 โดยประเมินความถูกต้องเชิงพื้นที่ และความต่อเนื่องเชิงเวลาของมาสก์วัตถุ

2. Objective Evaluation

การประเมินผลเชิงวัตถุ จะใช้เกณฑ์ในการประเมินความถูกต้องเชิงพื้นที่และความต่อเนื่องเชิงเวลาของมาสก์วัตถุเปรียบเทียบกับมาสก์อ้างอิง โดยคุณภาพเชิงพื้นที่ (Spatial quality) จะวัดจำนวนของจุดภาพที่แตกต่างและความคลาดเคลื่อนระหว่างมาสก์ที่ประเมินกับมาสก์อ้างอิง ต่อจำนวนจุดภาพของมาสก์อ้างอิง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิจัย

บทนี้จะเป็นการแสดงผลการวิจัยจากกรรมวิธีการแยกส่วนภาพวัตถุที่นำเสนอในแต่ละขั้นตอน โดยแบ่งเป็นสองส่วน คือ ผลจากการแยกส่วนภาพเชิงพื้นที่ และผลจากการแยกส่วนภาพเชิงเวลา และการวิเคราะห์ลำดับภาพที่ใช้ในการวิจัย การแยกส่วนภาพเชิงพื้นที่จะวิเคราะห์ในแต่ละขั้นตอนตั้งแต่การสร้างความสะดวกสบายให้กับภาพ การตรวจรู้ขอบเขตของบริเวณด้วยวิธีการเชิงสัญญาณ และผลการแยกส่วนภาพโดยกรรมวิธีสันปันน้ำ ส่วนการแยกส่วนเชิงเวลาจะแสดงผลของการแยกส่วนภาพโดยการตรวจรู้การเคลื่อนที่ด้วยการปรับค่าจุดเปลี่ยนโดยอัตโนมัติและการใช้หน่วยความจำช่วยในการแยกส่วนภาพวัตถุเคลื่อนที่ รวมทั้งการนำผลจากการวิเคราะห์ทั้งสองส่วนมาวิเคราะห์ร่วมกัน เพื่อช่วยให้การแยกส่วนภาพวัตถุเคลื่อนที่ที่มีความหมายดีขึ้น ซึ่งจะแสดงผลของการแยกส่วนภาพวัตถุที่มีความหมายเป็นภาพวัตถุที่ต่อเนื่อง สุดท้ายเป็นการวิเคราะห์ผลการแยกส่วนภาพเมื่อเปรียบเทียบกับกรรมวิธีของ COST211 ทั้งในส่วนการประเมินผลเชิงอัตวิสัยและการประเมินผลเชิงวัตถุ

4.1 การแยกส่วนเชิงพื้นที่

4.1.1 การทำให้ภาพมีความเรียบง่าย (Image Simplification)

ส่วนนี้เป็นกระบวนการขั้นต้นเพื่อประมวลผลสัญญาณภาพให้มีความเรียบง่ายขึ้นเพื่อให้การแยกส่วนภาพทำได้ง่ายขึ้น และลดปัญหาการแยกส่วนมากเกินไป

การสร้างความสะดวกสบายให้กับภาพ จะใช้วิธีการเชิงสัญญาณ ด้วยตัวปฏิบัติการแบบเปิด โดยใช้องค์ประกอบโครงสร้างขนาด 3x3 จุดภาพ



(ก)



(ข)



(ค)



(ง)



(จ)

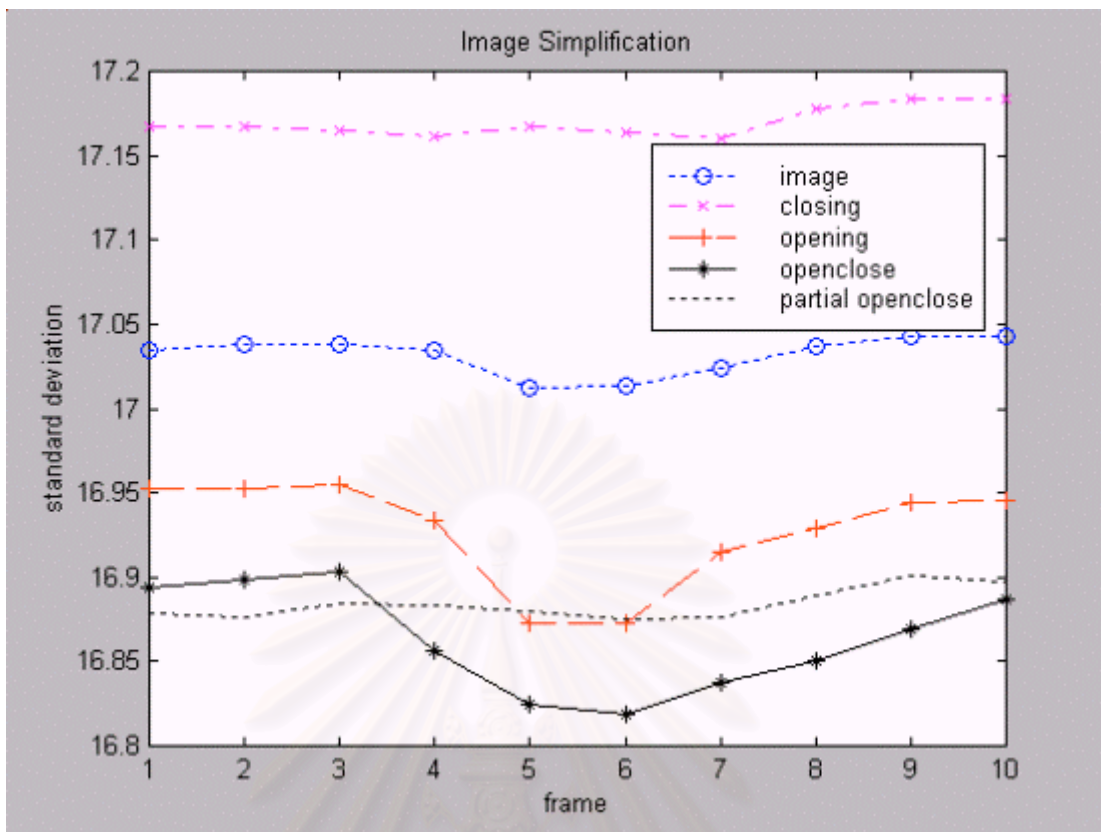
รูปที่ 4.1 การปฏิบัติการเชิงสัญญาณของภาพ Akiyo (ก) การปฏิบัติการแบบเปิด
 (ข) การปฏิบัติการแบบปิด (ค) การปฏิบัติการแบบเปิด-ปิด (ง) การปฏิบัติการแบบปิด-เปิด
 (จ) การปฏิบัติการแบบเปิด-ปิดด้วยการสร้างใหม่บางส่วน

การประเมินสมรรถนะการสร้างควมเรียบง่าย

ความเรียบง่าย จะสังเกตได้จากข้อมูลความเข้มแสงที่มีความต่อเนื่องกัน มีการกระจายของความเข้มแสงสูง ดังนั้นจึงใช้ค่าเบี่ยงเบนมาตรฐานในการวัดความเรียบง่ายของภาพ

โดยจะแสดงค่าเบี่ยงเบนมาตรฐานเฉลี่ยภายใน 10 เฟรมภาพที่ต่อเนื่องกัน เปรียบเทียบวิธีการสร้างควมเรียบง่ายแบบต่าง ๆ ดังนี้

- ภาพต้นแบบ (original image)
- การปฏิบัติการแบบปิด (closing operation)
- การปฏิบัติการแบบเปิด (opening operation)
- การปฏิบัติการแบบเปิด-ปิด (open-close operation)
- การปฏิบัติการแบบเปิด-ปิดด้วยการสร้างใหม่บางส่วน (open-close operation by partial reconstruction)



รูปที่ 4.2 แสดงค่าเบี่ยงเบนมาตรฐานของภาพที่ถูกสร้างให้มีความเรียบง่ายในลำดับภาพ Akiyo เฟรมที่ 1 ถึง 10

ตารางที่ 4.1 แสดงค่าเฉลี่ยของค่าเบี่ยงเบนมาตรฐานของภาพที่ถูกสร้างให้มีความเรียบง่ายของลำดับภาพ Akiyo ใน 10 เฟรมด้วยวิธีต่าง ๆ ดังนี้

วิธีการสร้างความเรียบง่าย	ค่าเบี่ยงเบนมาตรฐานเฉลี่ยของภาพ
ภาพต้นแบบ	17.03
การปฏิบัติการแบบปิด	17.17
การปฏิบัติการแบบเปิด	16.93
การปฏิบัติการแบบเปิด-ปิด	16.86
การปฏิบัติการแบบเปิด-ปิดด้วยการสร้างใหม่บางส่วน	16.88

จากผลการวิจัยในส่วนนี้ จะเห็นว่าการปฏิบัติการแบบเปิด-ปิดจะให้ค่าเบี่ยงเบนมาตรฐานต่ำกว่าวิธีอื่น และได้ภาพที่มีความเรียบง่ายโดยที่ยังรักษาข้อมูลของขอบภาพไว้ได้ จึงเป็นวิธีที่จะนำมาใช้ในการสร้างความเรียบง่ายของภาพสำหรับการแยกส่วนภาพวัตถุในวิทยานิพนธ์นี้

4.1.2 การตรวจรู้ข้อบด้วยวิธีการประมาณเกรเดียนต์

การตรวจรู้ข้อบจะใช้หลังจากการสร้างควมเรียบง่ายให้กับภาพแล้ว เพื่อให้สามารถแยกบริเวณของภาพได้ชัดเจน และลดการเกิดการแยกส่วนมากเกินไป

หลังภาพผ่านการประมาณเกรเดียนต์ด้วยวิธีเชิงสัณฐานแล้ว ค่าเกรเดียนต์จะถูกตัดสินใจด้วยค่าจุดเปลี่ยน ซึ่งมีค่าร้อยละ 10 ของค่าเกรเดียนต์ที่มีค่าสูงที่สุด เพื่อป้องกันปัญหาการแยกส่วนมากเกินไป

4.1.2.1 การประมาณเกรเดียนต์ (Gradient Approximation)

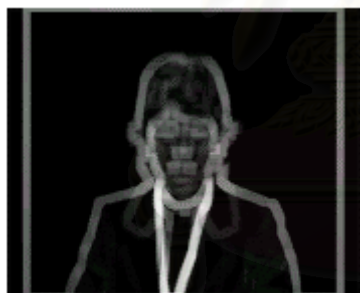
การประมาณค่าเกรเดียนต์ของภาพ ทำได้โดยได้เลขชั้นลบด้วยอีโรชันของภาพ โดยใช้องค์ประกอบโครงสร้างขนาด 5x5 จุดภาพ



(ก)



(ข)



(ค)



(ง)

รูปที่ 4.3 ภาพแสดงขั้นตอนของการประมาณเกรเดียนต์ (ก) การปฏิบัติการได้เลขชั้น

(ข) การปฏิบัติการอีโรชัน (ค) ค่าการประมาณเกรเดียนต์

(ง) ค่าเกรเดียนต์ที่มีค่าสูงกว่าค่าจุดเปลี่ยน

ค่าเกรเดียนต์สูงสุดที่สุดของภาพ เท่ากับ 255

ค่าจุดเปลี่ยนของค่าเกรเดียนต์ เท่ากับ 25.5

จำนวนบริเวณจากขอบเขตของค่าเกรเดียนต์ เท่ากับ 26

จำนวนจุดภาพที่ยังไม่ระบุบริเวณ เท่ากับ 18310

4.1.2.2 การประมาณเกรเดียนต์แบบหลายชั้น (Multiscale Gradient)



รูปที่ 4.4 แสดงการประมาณเกรเดียนต์แบบหลายชั้นของภาพ Claire

(ก) ค่าการประมาณเกรเดียนต์แบบหลายชั้น (ข) ค่าเกรเดียนต์ที่มีค่าสูงกว่าค่าจุดเปลี่ยน

ค่าเกรเดียนต์สูงสุดที่สุดของภาพ เท่ากับ 210

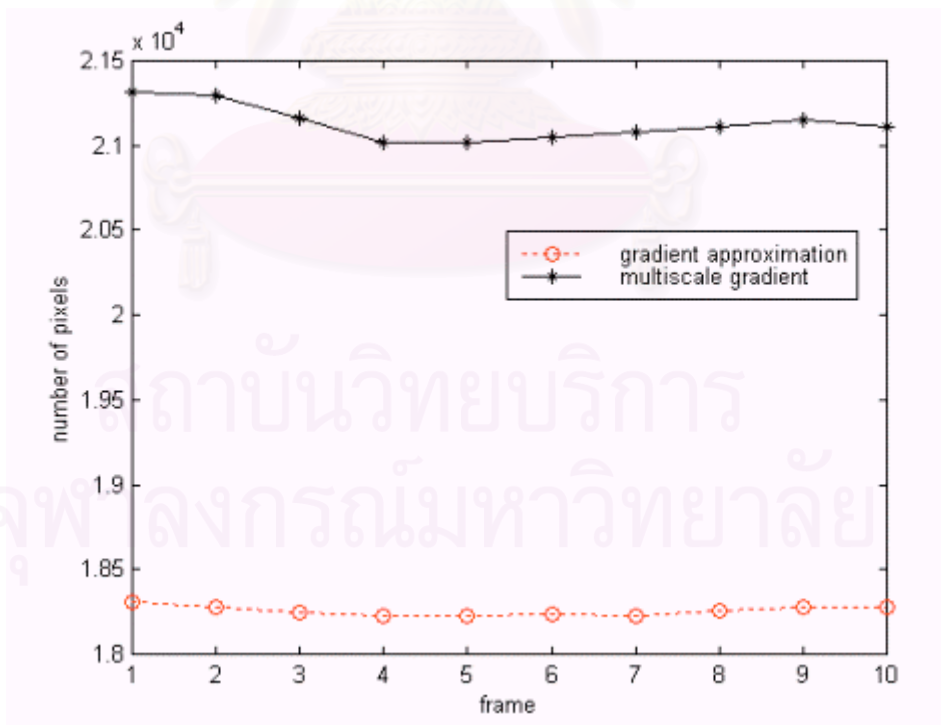
ค่าจุดเปลี่ยนของค่าเกรเดียนต์ เท่ากับ 21

จำนวนบริเวณจากขอบเขตของค่าเกรเดียนต์ เท่ากับ 10

จำนวนจุดภาพที่ยังไม่ระบุบริเวณ เท่ากับ 21311

การวิเคราะห์สมรรถนะของการตรวจรู้ขอบ

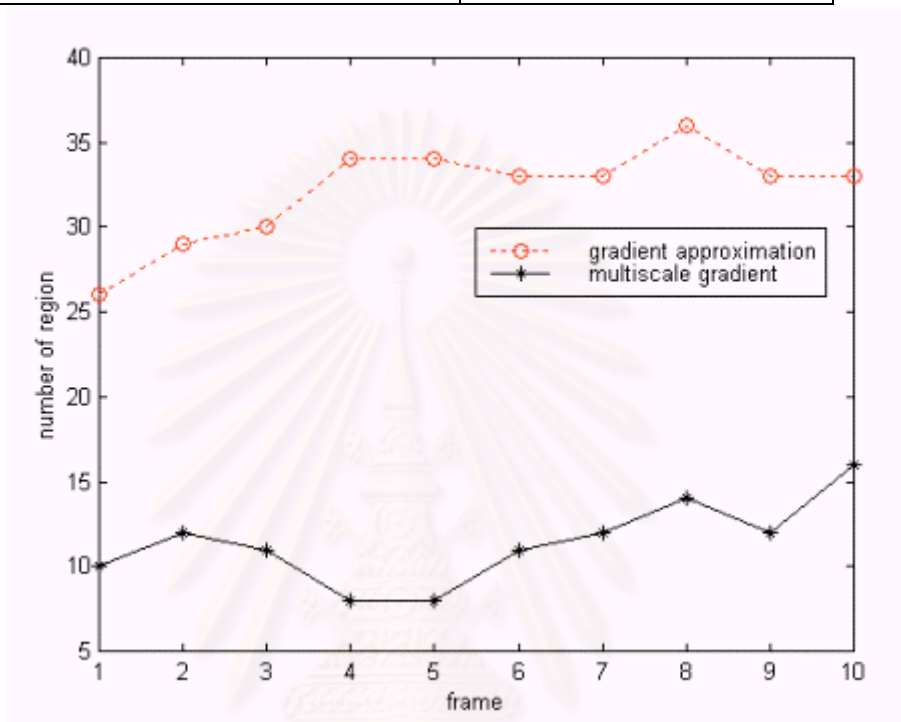
จะใช้การวัดจำนวนบริเวณที่เกิดจากการแยกส่วนภาพ และจำนวนจุดภาพที่ยังไม่ระบุบริเวณ (uncertain pixels)



รูปที่ 4.6 แสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณในเฟรมที่ 1 ถึง 10 ระหว่างการประมาณเกรเดียนต์แบบชั้นเดียว และแบบหลายชั้น

ตารางที่ 4.3 แสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณเฉลี่ยใน 10 เฟรม ระหว่างการประมาณ
เกรเดียนต์แบบขั้นเดียว และแบบหลายขั้น

วิธีการตรวจจุดขอบ	จำนวนจุดภาพ (จุดภาพ)
การประมาณเกรเดียนต์แบบขั้นเดียว	18,256
การประมาณเกรเดียนต์แบบหลายขั้น	21,130



รูปที่ 4.5 แสดงจำนวนบริเวณเฉลี่ยจากการตรวจจุดขอบในเฟรมที่ 1 ถึง 10 ของการประมาณ
เกรเดียนต์แบบขั้นเดียวและแบบหลายขั้น

ตารางที่ 4.2 แสดงจำนวนบริเวณเฉลี่ยจากการตรวจจุดขอบด้วยการประมาณเกรเดียนต์

วิธีการตรวจจุดขอบ	จำนวนบริเวณเฉลี่ย (บริเวณ)
การประมาณเกรเดียนต์แบบขั้นเดียว	32.1
การประมาณเกรเดียนต์แบบหลายขั้น	11.4

การวิเคราะห์ผลการวิจัย

จากการวิเคราะห์เมื่อเปรียบเทียบจำนวนบริเวณที่ได้จากการตรวจจุดขอบด้วยการ
ประมาณเกรเดียนต์แล้ว จะเห็นว่าเมื่อใช้การประมาณเกรเดียนต์แบบหลายขั้นจะเกิดจำนวน
บริเวณโดยเฉลี่ยน้อยกว่าการประมาณเกรเดียนต์แบบขั้นเดียว ซึ่งจะลดปัญหาของการแยกส่วน
มากเกินไป นอกจากนี้ ภาพเกรเดียนต์ที่ได้จากการประมาณเกรเดียนต์แบบหลายขั้นจะมี

ขอบเขตของบริเวณที่มีความกว้างของจุดภาพน้อยกว่าแบบขั้นเดียว จึงใช้วิธีการประมาณเกรเดียนต์แบบหลายชั้นในการตรวจรู้ขอบภาพของวิทยานิพนธ์ฉบับนี้

4.1.3 การใช้กรรมวิธีสันปันน้ำ (Watershed Algorithm)

ส่วนนี้จะแสดงการแยกส่วนภาพเชิงพื้นที่ ซึ่งให้บริเวณของภาพที่ชัดเจนตามข้อมูลความเข้มแสง และแทนค่าความเข้มของแต่ละบริเวณด้วยค่าเฉลี่ยของความเข้มแสงในบริเวณนั้น

หลังจากใช้กรรมวิธีสันปันน้ำ ในการเพิ่มขนาดแต่ละบริเวณจากจุดภาพที่ยังไม่ระบุบริเวณ จะสามารถลดจุดภาพที่ยังไม่ระบุบริเวณได้จำนวนหนึ่ง แต่ยังมีบางจุดภาพที่ไม่สามารถระบุบริเวณให้ได้ เนื่องจากเป็นบริเวณภายในขอบเขตท้องถิ่น (local maxima) จึงกำหนดบริเวณนี้ให้เป็นบริเวณใหม่เพิ่มขึ้น เพื่อให้สามารถแยกส่วนได้ละเอียดยิ่งขึ้น



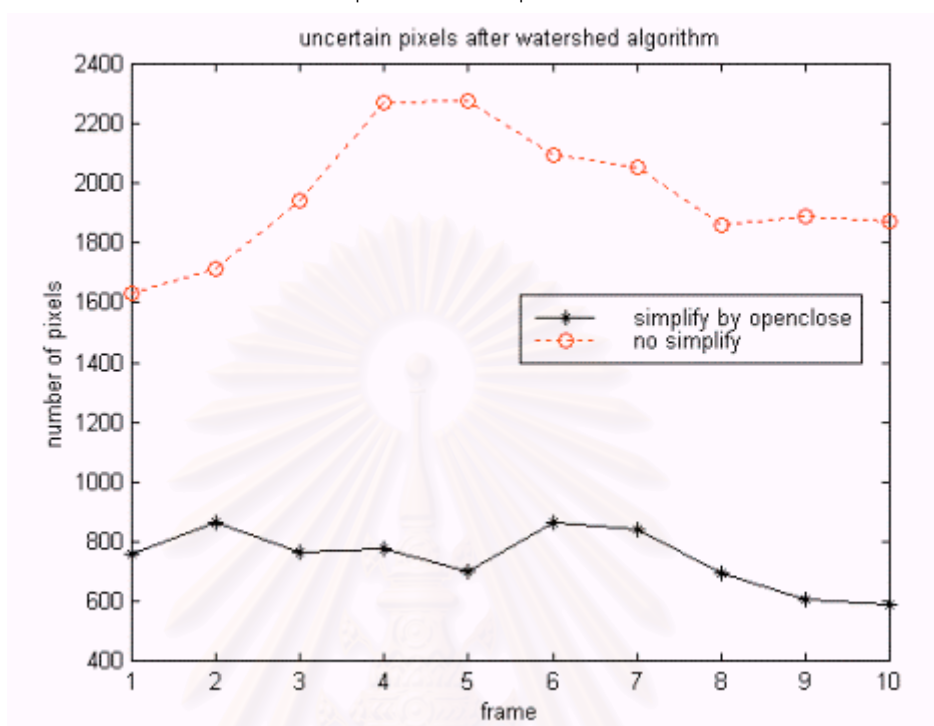
รูปที่ 4.7 ผลการแยกส่วนภาพเชิงพื้นที่โดยใช้กรรมวิธีสันปันน้ำของลำดับภาพ Claire
(ก) เฟรมที่ 1 (ข) เฟรมที่ 10

จำนวนบริเวณจากการแยกส่วนเชิงพื้นที่ในเฟรมที่ 1 เท่ากับ 28 บริเวณ
จำนวนบริเวณจากการแยกส่วนเชิงพื้นที่ในเฟรมที่ 10 เท่ากับ 27 บริเวณ

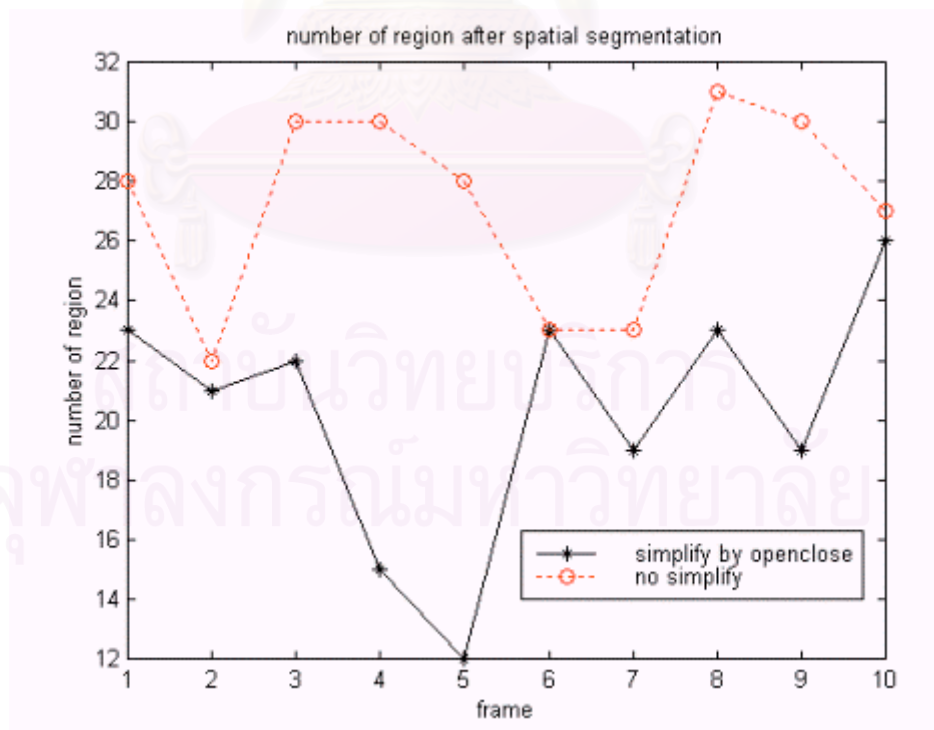
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์การแยกส่วนจากกรรมวิธีส้นปันน้ำ

จะวัดจำนวนจุดภาพที่ยังไม่ระบุบริเวณที่เหลือ หลังจากผ่านกรรมวิธีส้นปันน้ำแล้ว และจำนวนบริเวณทั้งหมดที่ได้จากการระบุบริเวณให้กับจุดภาพ



รูปที่ 4.8 กราฟแสดงจำนวนจุดภาพที่ยังไม่ระบุบริเวณที่เหลือ หลังจากผ่านกรรมวิธีส้นปันน้ำแล้ว



รูปที่ 4.9 กราฟแสดงจำนวนบริเวณทั้งหมดที่ได้จากการระบุบริเวณให้กับจุดภาพ

ตารางที่ 4.4 ผลของการแยกส่วนภาพเชิงพื้นที่ของภาพ Claire โดยกรรมวิธีสันปันน้ำ

ตารางที่ 4.4 (a) ไม่ใช้การสร้างความเรียบงายให้ภาพ

จำนวนจุดภาพที่ยังไม่ระบุเฉลี่ยหลังจากกรรมวิธีสันปันน้ำ	1960 จุดภาพ
จำนวนบริเวณเฉลี่ยของการแยกส่วนภาพเชิงพื้นที่	27.2 บริเวณ

ตารางที่ 4.4 (b) ใช้การปฏิบัติการแบบเปิด-ปิดสร้างความเรียบงายให้ภาพ

จำนวนจุดภาพที่ยังไม่ระบุเฉลี่ยหลังจากกรรมวิธีสันปันน้ำ	744.3 จุดภาพ
จำนวนบริเวณเฉลี่ยของการแยกส่วนภาพเชิงพื้นที่	20.3 บริเวณ

การวิเคราะห์ผลการวิจัย

จะเห็นว่าหลังจากผ่านกรรมวิธีสันปันน้ำเพื่อกำหนดจุดภาพให้กับบริเวณแล้ว ถ้าใช้การสร้างความเรียบงายให้กับภาพแล้ว จะเหลือจำนวนจุดภาพที่ยังไม่ระบุบริเวณน้อยกว่าเมื่อไม่ได้ใช้การสร้างความเรียบงาย ซึ่งแสดงว่าเมื่อมีการสร้างการเรียบงายให้แก่ภาพแล้ว จะช่วยลดจำนวนของจุดภาพที่เกิดจากค่าเกรเดียนต์ของภาพที่มีค่าต่ำ และจะให้จำนวนบริเวณน้อยลง ซึ่งแก้ปัญหาการแยกส่วนมากเกินไปและสามารถลดการคำนวณของขั้นตอนนี้ลงได้

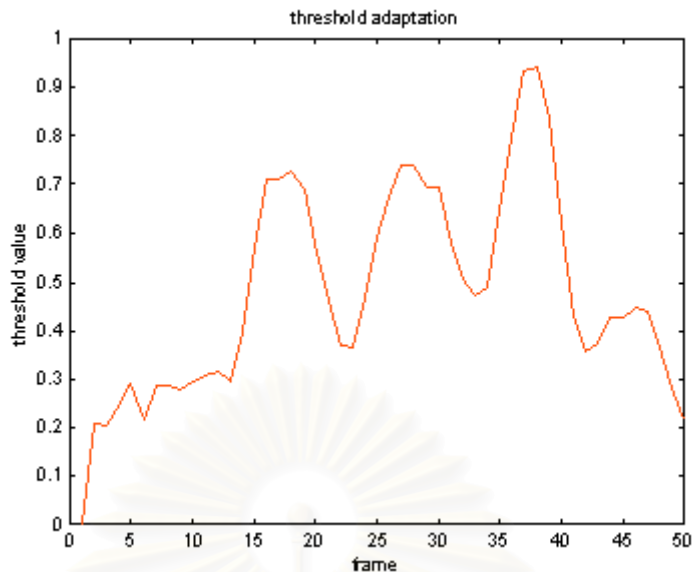
4.2 การแยกส่วนเชิงเวลา (Temporal Segmentation)

การแยกส่วนเชิงเวลา จะใช้ข้อมูลระหว่างเฟรมที่ต่อเนื่องกัน ซึ่งจะสามารถตรวจรู้บริเวณที่มีการเคลื่อนที่ได้ การแยกส่วนนี้จะอยู่บนพื้นฐานของการตรวจรู้การเปลี่ยนแปลงเป็นหลัก เพื่อให้การประมวลผลใช้ทรัพยากรในการคำนวณที่ต่ำกว่าการประมวลผลการเคลื่อนที่ เนื่องจากไม่จำเป็นต้องใช้ข้อมูลของเวกเตอร์การเคลื่อนที่ในการแยกส่วนระหว่างเฟรมหลัง และวัตถุที่เคลื่อนที่ที่มีเพียงวัตถุเดียว สำหรับการประยุกต์ใช้งาน ด้านการประชุมทางวิดีโอ หรือโทรศัพท์ภาพ

4.2.1 การวิเคราะห์การเปลี่ยนแปลงความเข้มแสงระหว่างเฟรม

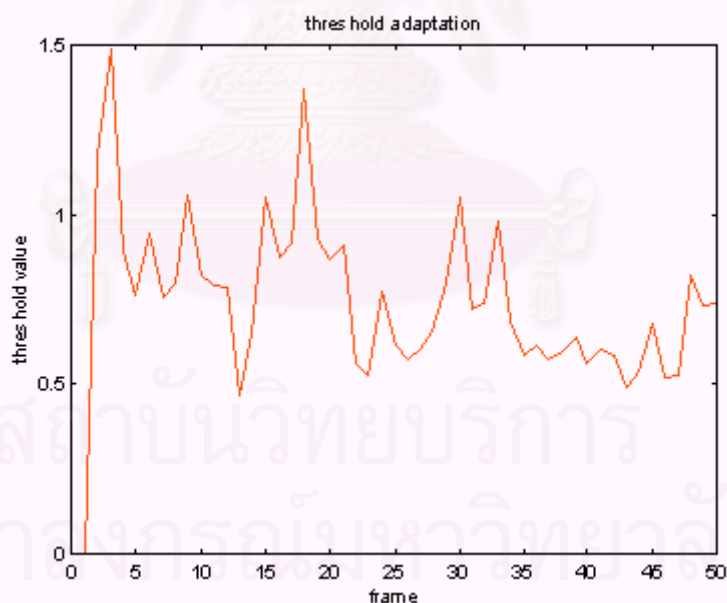
การวิเคราะห์การเปลี่ยนแปลงความเข้มแสง จะใช้ช่วยสำหรับการกำหนดค่าจุดเปลี่ยนแปลงในการตรวจรู้การเปลี่ยนแปลง และทราบถึงลักษณะความเร็วของการเปลี่ยนแปลงในลำดับภาพ

การวิเคราะห์การเปลี่ยนแปลงจะพิจารณาเป็นค่าเฉลี่ยของความเข้มแสงของจุดภาพที่เปลี่ยนแปลงไประหว่างเฟรมที่ต่อเนื่องกัน



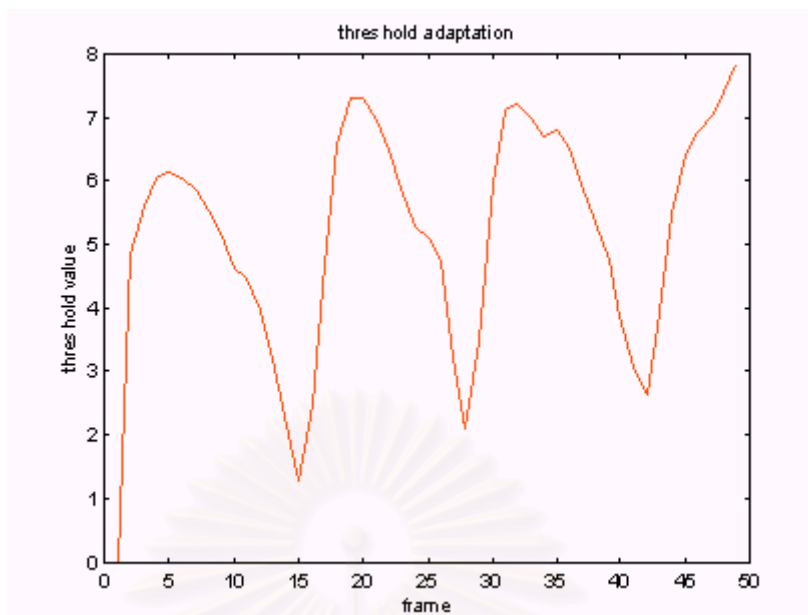
รูปที่ 4.10 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Akiyo ใน 50 เฟรมแรก

จะเห็นว่าลำดับภาพ Akiyo จะมีการเปลี่ยนแปลงความเข้มระหว่างเฟรมน้อยมาก แสดงว่ามีการเคลื่อนที่ค่อนข้างน้อย และควรที่จะใช้ค่าจุดเปลี่ยนค่าต่ำเพื่อให้สามารถตรวจจับการเคลื่อนที่ของวัตถุได้ครบถ้วน ในช่วงเฟรมที่ 1 ถึงเฟรมที่ 50 จะต้องตรวจจับการเปลี่ยนของทุกจุดภาพเนื่องจากค่าการเปลี่ยนแปลงความเข้มเฉลี่ยต่ำกว่า 1 ตลอดช่วงของลำดับภาพ



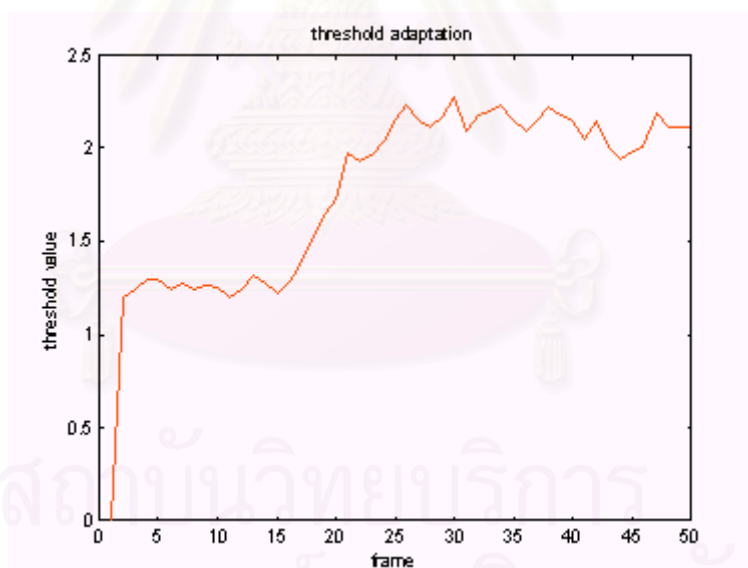
รูปที่ 4.11 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Claire ใน 50 เฟรมแรก

จะเห็นว่าลำดับภาพ Claire จะมีการเคลื่อนที่เล็กน้อยคล้ายภาพ Akiyo แต่การเคลื่อนที่จะไม่สม่ำเสมอจึงต้องมีการปรับเปลี่ยนค่าจุดเปลี่ยนด้วย เพื่อให้สามารถตรวจจับการเคลื่อนที่ของวัตถุได้



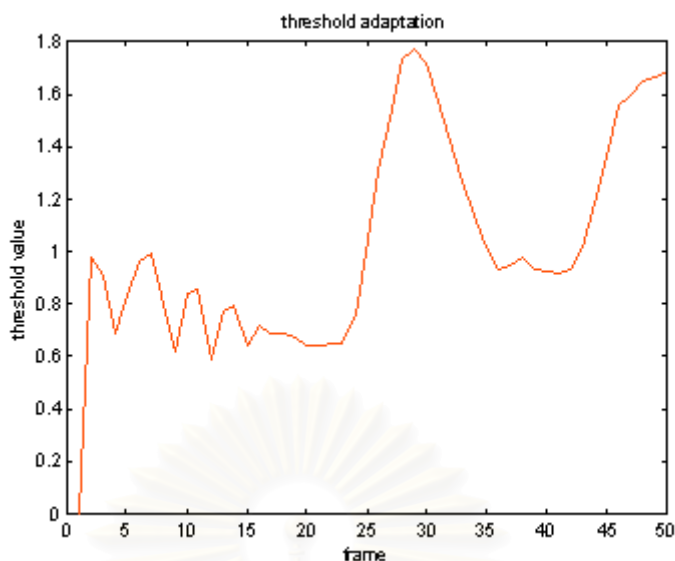
รูปที่ 4.12 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Erik ใน 50 เฟรมแรก

สำหรับภาพ Erik จะเห็นว่ามี การเปลี่ยนแปลงความเข้มค่อนข้างสูง และยังมี การเคลื่อนที่ของวัตถุที่ไม่สม่ำเสมอตลอดลำดับภาพ การเคลื่อนที่ในช่วงแรกจะมีมาก และน้อยลงแล้วมากใหม่ สลับกันไป ซึ่งจะมีการเคลื่อนที่น้อยในช่วงเฟรมที่ 15, 28 และ 42 ตามลำดับ



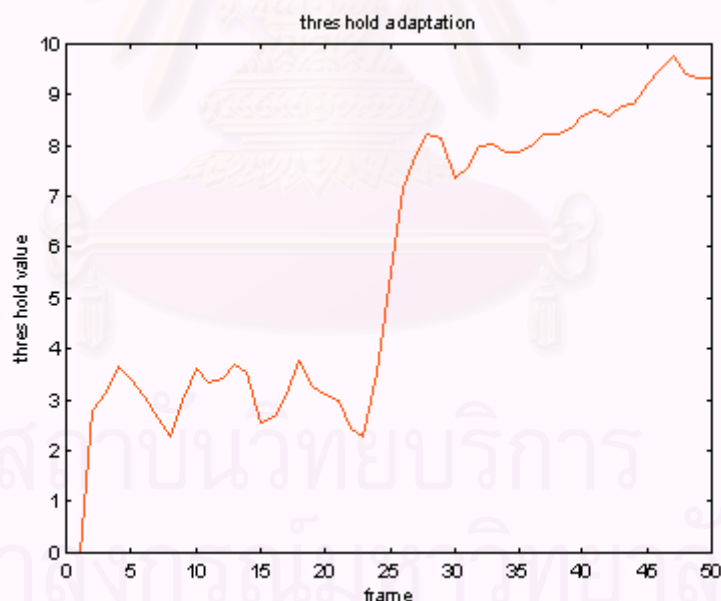
รูปที่ 4.13 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Hall Monitor ใน 50 เฟรมแรก

ในช่วงเฟรมที่ 1-15 ของลำดับภาพ Hall Monitor จะมีการเปลี่ยนแปลงคงที่ เนื่องจาก การเปลี่ยนแปลงจะเกิดจากความแปรปรวนของแสงจากกล้องจับภาพ หลังจากนั้นจะมีการเคลื่อนที่ของวัตถุจึงทำให้การเปลี่ยนแปลงมีค่ามากขึ้น จนกระทั่งการเคลื่อนที่ของวัตถุมีความสม่ำเสมอ



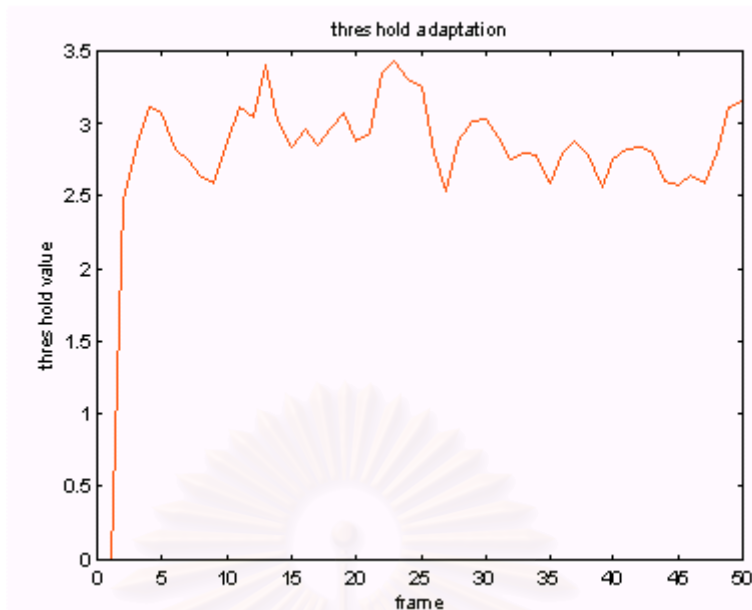
รูปที่ 4.14 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Mother & Daughter ใน 50 เฟรมแรก

การเปลี่ยนแปลงของลำดับภาพ Mother & Daughter จะมีเล็กน้อย เนื่องจากเป็นลักษณะของภาพการประชุมทางวิดีโอคล้ายกับ Akiyo และ Claire แต่จะมีการเปลี่ยนแปลงมากเป็นบางช่วง เนื่องจากบางช่วงมีการเคลื่อนไหวที่เพียงเฉพาะแม่ แต่บางช่วงมีการเคลื่อนไหวที่ทั้งแม่และลูกด้วย จึงทำให้เกิดการเปลี่ยนแปลงมากขึ้น



รูปที่ 4.15 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Table Tennis ใน 50 เฟรมแรก

การเปลี่ยนแปลงของลำดับภาพ Table Tennis จะมีปรับเปลี่ยนรูปแบบค่อนข้างมาก เนื่องจากในช่วงแรกของลำดับภาพ จะมีการเปลี่ยนแปลงจากการเคลื่อนไหวของวัตถุ หลังจากเฟรมที่ 25 เป็นต้นไป จะมีการขยายออกของกล่องจับภาพทำให้การเปลี่ยนแปลงมีสูงขึ้นมาก



รูปที่ 4.16 การเปลี่ยนแปลงความเข้มแสงของลำดับภาพ Trevor ใน 50 เฟรมแรก

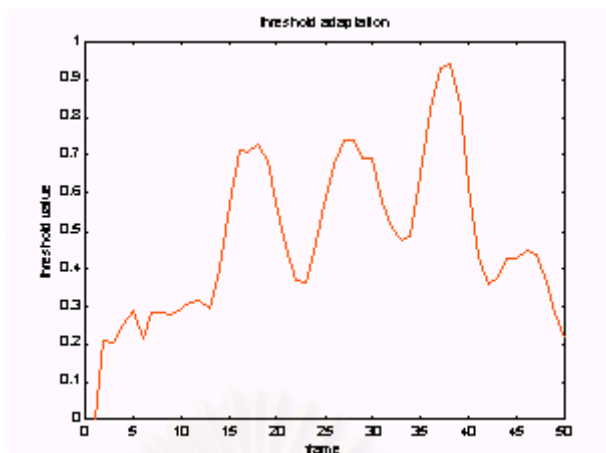
ลำดับภาพ Trevor จะมีการเปลี่ยนแปลงมากพอสมควรเนื่องจากเกิดการเคลื่อนที่ของหลายวัตถุที่แตกต่างกันในลำดับภาพ บางวัตถุที่การเคลื่อนที่ช้ามาก แต่บางวัตถุมีการเคลื่อนที่เร็ว เมื่อวิเคราะห์การเปลี่ยนแปลงโดยเฉลี่ยจะเห็นว่ามีการเปลี่ยนแปลงโดยเฉลี่ยค่อนข้างสม่ำเสมอ

4.2.2 การตรวจรู้การเปลี่ยนแปลง

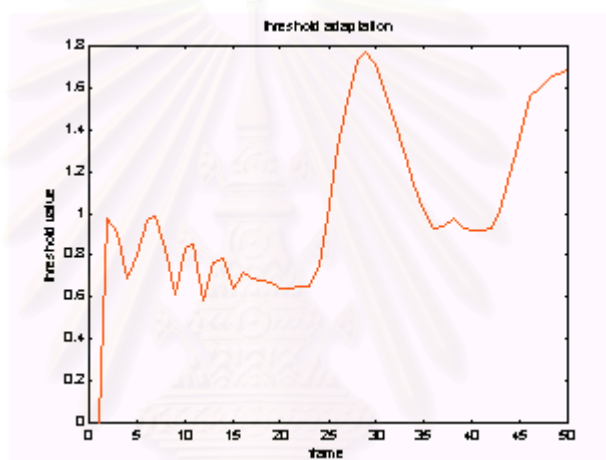
การตรวจรู้ภาพวัตถุที่เคลื่อนที่สามารถทำได้โดยการตรวจรู้การเปลี่ยนแปลง ซึ่งอาจตรวจรู้ทุกจุดภาพ โดยสังเกตการเปลี่ยนแปลงของความเข้ม หรือวิเคราะห์เป็นบริเวณที่มีการเปลี่ยนแปลงเฉลี่ยเกินค่าจุดเปลี่ยนที่กำหนด วิทยานิพนธ์ฉบับนี้ จะแสดงเฉพาะการตรวจรู้การเปลี่ยนแปลงของแต่ละจุดภาพ และการวิเคราะห์การเปลี่ยนแปลงเป็นบล็อกขนาด 4×4 และ 8×8 จุดภาพ โดยค่าจุดเปลี่ยนจะมีการปรับตัวโดยอัตโนมัติ ตามค่าความเข้มที่เปลี่ยนแปลงเฉลี่ยระหว่างเฟรมในปัจจุบันและเฟรมก่อนหน้า

4.2.2.1 การปรับตัวของค่าจุดเปลี่ยนโดยอัตโนมัติ

ค่าจุดเปลี่ยนจะมีการปรับตัวโดยอัตโนมัติ ตามค่าความเข้มที่เปลี่ยนแปลงเฉลี่ยระหว่างเฟรมในปัจจุบันและเฟรมก่อนหน้า โดยค่าจุดเปลี่ยนตั้งต้นจะมีค่าเท่ากับร้อยละ 10 ของค่าความเข้มแสงสูงสุดของภาพในเฟรมแรก และจะมีการปรับตัวสำหรับเฟรมต่อไป



รูปที่ 4.17 แสดงค่าจุดเปลี่ยนที่มีการปรับตัวโดยอัตโนมัติ สำหรับลำดับภาพ Akiyo



รูปที่ 4.18 แสดงค่าจุดเปลี่ยนที่มีการปรับตัวโดยอัตโนมัติ สำหรับลำดับภาพ Mother & Daughter

4.2.2.2 การตรวจรู้การเปลี่ยนแปลงของแต่ละจุดภาพ

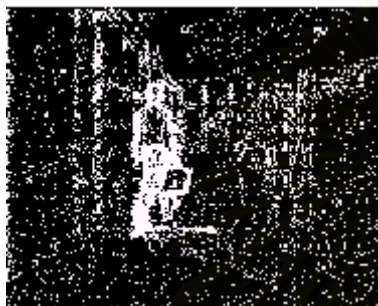
การตรวจรู้การเปลี่ยนแปลงนี้จะใช้ตรวจรู้การเปลี่ยนแปลงของทุกจุดภาพ การตรวจรู้การเปลี่ยนแปลงจะมีการปรับตัวขึ้นกับค่าจุดเปลี่ยน เหมาะสำหรับลำดับภาพที่มีวัตถุเคลื่อนที่ขนาดเล็ก หรือมีการเคลื่อนที่ไม่มากนัก แต่อาจทำให้เกิดสัญญาณรบกวนในบริเวณพื้นหลังได้จำนวนมาก ถ้าค่าจุดเปลี่ยนมีค่าต่ำเกินไป



(ก)



(ข)



(ค)



(ง)

รูปที่ 4.19 แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพในเฟรมที่ 50 (ก) Akiyo

(ข) Mother & Daughter (ค) Hall Monitor (ง) Erik

4.2.2.3 การวิเคราะห์การเปลี่ยนแปลงเป็นบล็อกขนาด 4x4 จุดภาพ

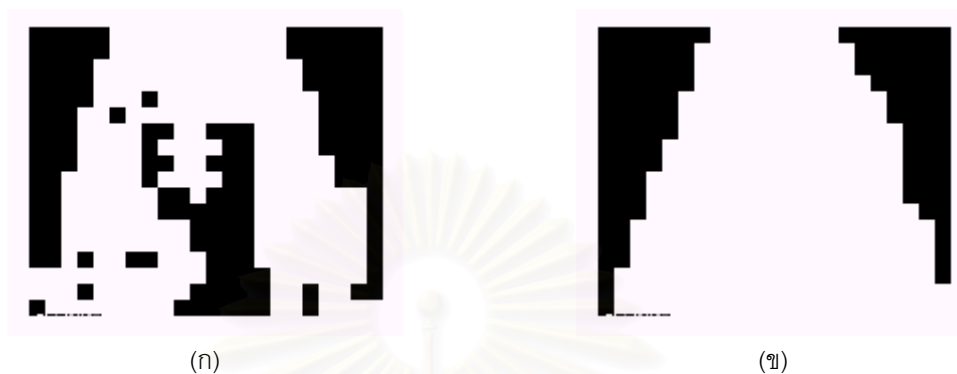
ใช้ตรวจรู้การเปลี่ยนแปลงของจุดภาพเป็นบล็อก เพื่อป้องกันสัญญาณรบกวนในบริเวณพื้นหลัง หรือการเคลื่อนไหวเล็กน้อยของกล้องจับภาพ เหมาะกับฉากที่วัตถุเคลื่อนที่ขนาดเล็ก และมีสัญญาณรบกวนได้เล็กน้อย



รูปที่ 4.20 รูปแสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Table Tennis ในเฟรมที่ 5
เมื่อค่าจุดเปลี่ยน เท่ากับ 1

4.2.2.4 การวิเคราะห์การเปลี่ยนแปลงเป็นบล็อกขนาด 8x8 จุดภาพ

ใช้ตรวจรู้การเปลี่ยนแปลงของจุดภาพเป็นบล็อก เพื่อป้องกันสัญญาณรบกวนในบริเวณพื้นหลัง หรือการเคลื่อนไหวเล็กน้อยของกล้องจับภาพ เหมาะกับฉากที่วัตถุเคลื่อนที่ขนาดใหญ่ และมีสัญญาณรบกวน



รูปที่ 4.21 รูปแสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Suzie (ก) เฟรมที่ 5 (ข) เฟรมที่ 50

4.2.3 การกรองสัญญาณรบกวน และการเติมเต็มภาพวัตถุ

การกรองสัญญาณรบกวน จะใช้ตัวปฏิบัติการแบบเปิด ส่วนการเติมเต็มภาพวัตถุจะใช้ตัวปฏิบัติการแบบปิด ดังนั้นการสร้างสรรค์ของการตรวจรู้การเปลี่ยนแปลง จึงใช้ตัวปฏิบัติการแบบปิด-เปิด และการกำจัดบริเวณที่มีขนาดเล็ก กับผลที่ได้จากการตรวจรู้การเปลี่ยนแปลงของจุดภาพ



รูปที่ 4.22 การปฏิบัติการแบบปิด-เปิด สำหรับการตรวจรู้การเปลี่ยนแปลงของภาพ Akiyo เฟรมที่ 50 โดยใช้องค์ประกอบโครงสร้างขนาด 5x5 จุดภาพ



รูปที่ 4.23 การปฏิบัติการแบบปิด-เปิด และการกำจัดบริเวณขนาดเล็กสำหรับ
การตรวจรู้การเปลี่ยนแปลงของภาพ Mother&Daughter ในเฟรมที่ 50



รูปที่ 4.24 การกำจัดบริเวณที่มีขนาดเล็กในการตรวจรู้การเปลี่ยนแปลงของภาพ Claire
เฟรมที่ 50 เมื่อใช้การตรวจรู้แบบบล็อกขนาด 4x4 จุดภาพ และค่าจุดเปลี่ยน เท่ากับ 0.2

การวิเคราะห์ผลการวิจัย

การกำจัดบริเวณของมาสก์ที่มีขนาดเล็กออกจะช่วยให้การตรวจรู้การเปลี่ยนแปลงและการเคลื่อนที่ของวัตถุมีความถูกต้องสูงขึ้น เนื่องจากจะลดบริเวณที่เกิดจากความแปรปรวนของแสงในส่วนของพื้นหลังลง ให้เหลือแต่ภาพวัตถุเคลื่อนที่ที่สนใจเท่านั้น

4.2.4 การใช้หน่วยความจำช่วยในการตรวจรู้

เนื่องจากการตรวจรู้การเปลี่ยนแปลงจะพิจารณาเฉพาะความแตกต่างระหว่างเฟรมในปัจจุบันและเฟรมก่อนหน้าเพียงเฟรมเดียวเท่านั้น ดังนั้นอาจไม่สามารถตรวจรู้การเคลื่อนที่ของวัตถุในบริเวณที่ไม่มี การเปลี่ยนแปลง หรือการเคลื่อนที่ของวัตถุหยุดนิ่งในช่วงเฟรมที่พิจารณาได้ด้วยเหตุนี้ การแยกส่วนภาพวัตถุเคลื่อนที่ให้มีลักษณะที่มีความหมายจึงจำเป็นต้องนำผลจากการตรวจรู้การเปลี่ยนแปลงในเฟรมก่อนหน้ามาพิจารณาด้วย เพื่อให้สังเกตการเคลื่อนที่ของวัตถุอย่างต่อเนื่องได้

การเปรียบเทียบการตรวจรู้การเปลี่ยนแปลงเมื่อไม่ใช้และใช้หน่วยความจำ แสดงดังรูป



รูปที่ 4.25 แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Mother & Daughter
ที่ไม่ได้ใช้หน่วยความจำ

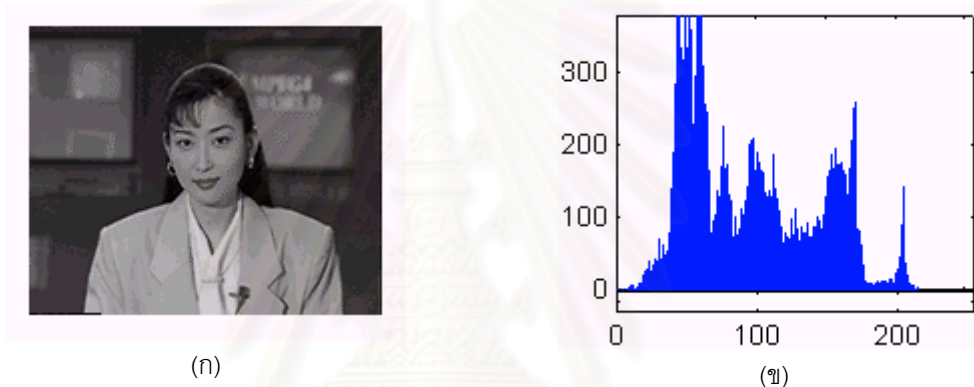


รูปที่ 4.26 แสดงการตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Mother & Daughter
ที่ใช้หน่วยความจำ

จากมาส์กของการเปลี่ยนแปลงจะแสดงให้เห็นว่าการใช้หน่วยความจำช่วยในการติดตามการเคลื่อนที่ที่จะสามารถตรวจรู้การเคลื่อนที่ของภาพวัตถุที่เฟรมที่ไม่มีเปลี่ยนแปลงได้ด้วย เนื่องจากหน่วยความจำจะเก็บผลของการตรวจรู้การเคลื่อนที่ในเฟรมก่อนหน้าเอาไว้ ถ้าในเฟรมปัจจุบันไม่มีการเปลี่ยนแปลงในตำแหน่งของมาส์กที่ตรวจรู้ได้จากเฟรมก่อนหน้า แสดงว่าวัตถุมีการหยุดเคลื่อนที่ แต่จะนำมาส์กของภาพวัตถุก่อนหน้ามาพิจารณาด้วย เพื่อให้สามารถตรวจรู้ภาพวัตถุที่เคลื่อนที่ได้อย่างต่อเนื่อง

4.3 การวิเคราะห์ลำดับภาพ

4.3.1 ลำดับภาพ Akiyo



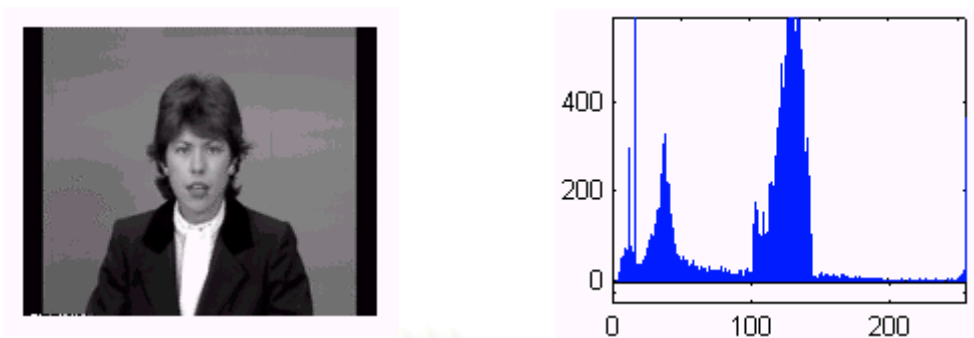
รูปที่ 4.27 ลำดับภาพ Akiyo (n) ภาพเฟรมที่ 1 (x) ฮิสโทแกรมของภาพ

ลักษณะของลำดับภาพ

ภาพจะมีการกระจายของความเข้มแสงที่ค่อนข้างสม่ำเสมอทุกความเข้มแสง โดยพื้นหลังของภาพส่วนใหญ่จะมีค่าความเข้มแสงต่ำ (สีเข้ม) และพื้นหน้าจะมีค่าความเข้มแสงสูง (สีอ่อน) จะมีเฉพาะบริเวณผมที่มีค่าความเข้มแสงต่ำคล้ายพื้นหลัง ทำให้เกิดปัญหาในการแยกส่วนเชิงพื้นที่ของบริเวณนี้

ส่วนการเคลื่อนที่ของลำดับภาพ วัตถุจะมีการเคลื่อนที่ในพื้นที่หลังที่อยู่นิ่ง โดยวัตถุจะมีการเคลื่อนที่ทั้งวัตถุตั้งแต่ส่วนหัวตลอดถึงลำตัวทำให้สามารถแยกส่วนภาพวัตถุได้ง่ายโดยการแยกส่วนเชิงเวลา

4.3.2 ลำดับภาพ Claire



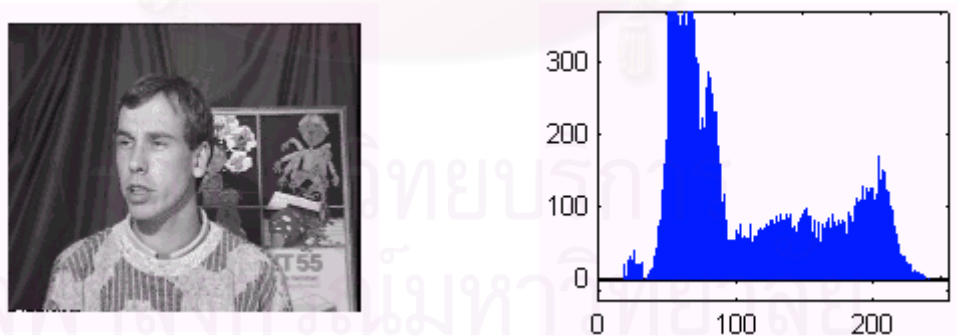
รูปที่ 4.28 ลำดับภาพ Claire (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ

ลักษณะของลำดับภาพ

ภาพจะมีการกระจายของความเข้มแสงเป็นกลุ่ม ๆ ความเข้มแสงส่วนใหญ่จะมีค่าความเข้มแสงต่ำ (สีเข้ม) แต่มีความแตกต่างของแต่ละบริเวณมากพอควร จึงทำให้สามารถแยกส่วนเชิงพื้นที่ได้ง่ายและแยกบริเวณได้จำนวนไม่มากนัก

ส่วนการเคลื่อนที่ของลำดับภาพ วัตถุจะมีการเคลื่อนที่ช้าในพื้นที่หลังที่อยู่นิ่ง โดยวัตถุจะมีการเคลื่อนที่ส่วนใหญ่ในบริเวณศีรษะและปาก บริเวณลำตัวค่อนข้างอยู่นิ่ง ทำให้สามารถแยกส่วนของบริเวณศีรษะได้จากการตรวจจับการเปลี่ยนแปลง แต่ในบริเวณลำตัวจำเป็นต้องใช้ข้อมูลของความเข้มช่วยในการแยกส่วนภาพวัตถุทั้งหมดออกมา

4.3.3 ลำดับภาพ Erik

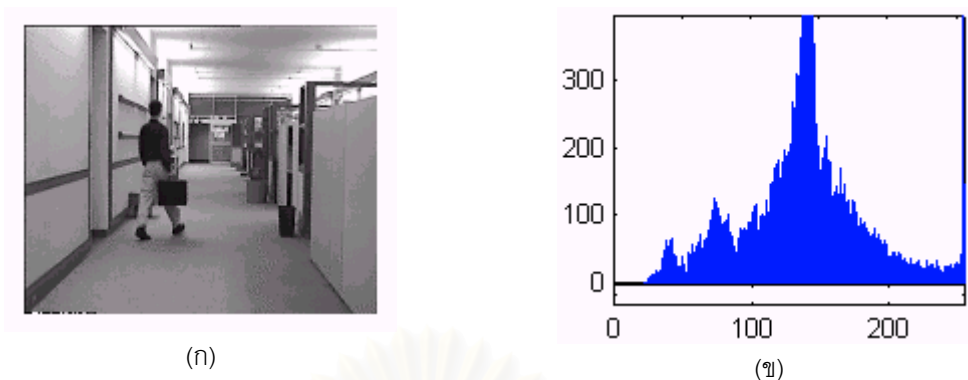


รูปที่ 4.29 ลำดับภาพ Erik (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ

ลักษณะของลำดับภาพ

ความเข้มแสงกระจายตัวสม่ำเสมอ รายละเอียดพื้นหลังของภาพมีพอสมควร และวัตถุที่การเคลื่อนที่ตลอดเวลา เกิดเงาของภาพวัตถุด้วย

4.3.4 ลำดับภาพ Hall Monitor

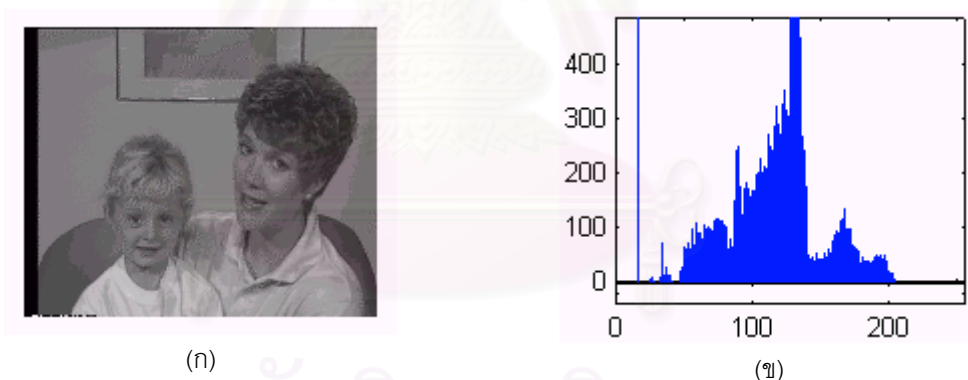


รูปที่ 4.30 ลำดับภาพ Hall Monitor (ก) ภาพเฟรมที่ 50 (ข) ฮิสโทแกรมของภาพ

ลักษณะของลำดับภาพ

ช่วง 10 เฟรมแรกจะไม่มี การเคลื่อนที่ของวัตถุ หลังจากนั้นจะมีวัตถุขนาดเล็กเคลื่อนที่ ในแนวลึก และมีอีกวัตถุเคลื่อนที่สวนกลับมา สามารถแบ่งภาพวัตถุได้ 2 ภาพ

4.3.5 ลำดับภาพ Mother & Daughter



รูปที่ 4.31 ลำดับภาพ Mother & Daughter (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ

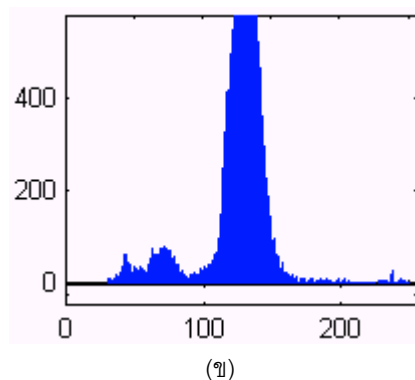
ลักษณะของลำดับภาพ

การกระจายตัวของความเข้มค่อนข้างต่ำ แต่ขอบเขตของภาพวัตถุค่อนข้างชัดเจน ลำดับภาพมีการเคลื่อนที่ของวัตถุแยกกัน 2 วัตถุ โดยมีวัตถุหนึ่งเคลื่อนที่มาก ส่วนอีกวัตถุหนึ่งเคลื่อนที่น้อย และบางเฟรมหยุดนิ่ง

4.3.6 ลำดับภาพ Suzie



(ก)



(ข)

รูปที่ 4.32 ลำดับภาพ Suzie (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ

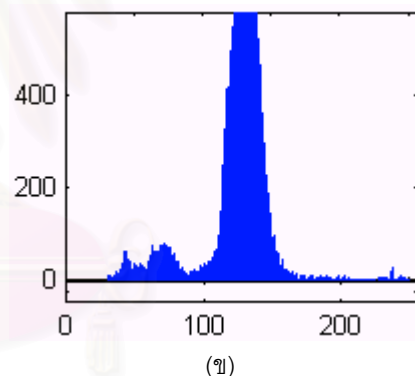
ลักษณะของลำดับภาพ

การกระจายของความเข้มสม่ำเสมอมาก ทำให้ตรวจรู้ขอบของภาพวัตถุเคลื่อนที่ค่อนข้างยาก ภาพวัตถุมีขนาดใหญ่ เนื่องจากเป็นโทรศัพท์ภาพ และการเคลื่อนที่มีมากอยู่ตลอดเวลา

4.3.7 ลำดับภาพ Table Tennis



(ก)



(ข)

รูปที่ 4.33 ลำดับภาพ Table Tennis (ก) ภาพเฟรมที่ 10 (ข) ฮิสโทแกรมของภาพ

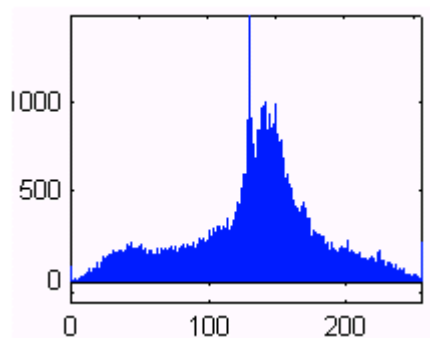
ลักษณะของลำดับภาพ

มีการเคลื่อนที่ของวัตถุ 2 วัตถุบนพื้นหลังที่ไม่ราบเรียบ วัตถุหนึ่งมีขนาดเล็ก และอีกวัตถุหนึ่งมีหลายส่วนประกอบที่แตกต่างกันเชิงความเข้มแสง มีสัญญาณรบกวนเกิดขึ้นเนื่องจากกล้องจับภาพ และมีการขยายภาพออกในช่วงหลังของลำดับภาพ

4.3.8 ลำดับภาพ Trevor



(ก)



(ข)

รูปที่ 4.34 ลำดับภาพ Trevor (ก) ภาพเฟรมที่ 1 (ข) ฮิสโทแกรมของภาพ

ลักษณะของลำดับภาพ

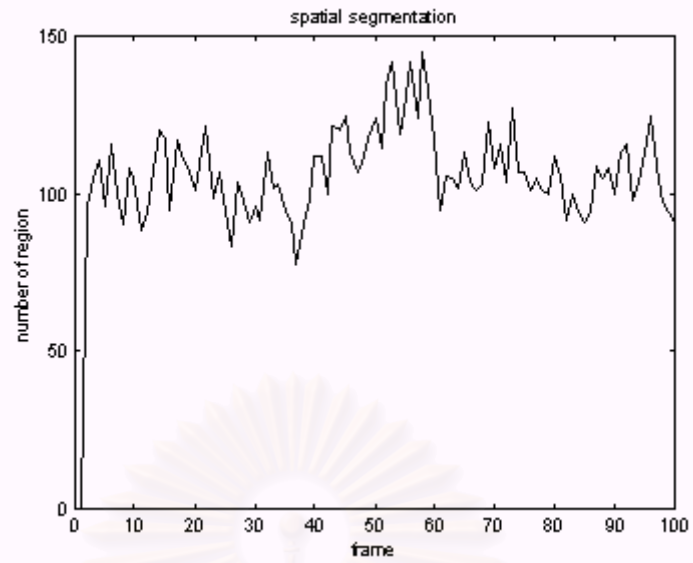
ช่วงแรกมีการเคลื่อนที่ของวัตถุหลายวัตถุในพื้นที่หลังที่อยู่นิ่ง วัตถุมีลักษณะการเคลื่อนที่แตกต่างกัน ซึ่งมีบางวัตถุเคลื่อนที่เข้ามา yak ต่อการตรวจรู้การเคลื่อนที่ และช่วงหลังมีตัดฉากไปอีกภาพหนึ่ง เป็นภาพการประชุมทางวิดีโอ

4.4 การแยกส่วนภาพวัตถุโดยการวิเคราะห์เชิงพื้นที่-เวลา

การนำผลจากการวิเคราะห์จากทั้งการแยกส่วนเชิงพื้นที่ และการแยกส่วนเชิงเวลามาวิเคราะห์ร่วมกันจะช่วยให้ได้ผลการแยกส่วนภาพวัตถุที่มีความหมายดีขึ้น



รูปที่ 4.35 ผลการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Mother & Daughter



รูปที่ 4.36 รูปกราฟแสดงจำนวนบริเวณที่ได้จากการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Mother & Daughter ในเฟรมที่ 1 ถึง 100



รูปที่ 4.37 ผลการแยกส่วนภาพเชิงเวลาของลำดับภาพ Mother & Daughter



รูปที่ 4.38 รูปแสดงมาสก์ของภาพวัตถุของลำดับภาพ Mother & Daughter
เฟรมที่ 33 35 37 ... 49



รูปที่ 4.39 รูปแสดงภาพวัตถุของลำดับภาพ Mother & Daughter จากการแยกส่วนภาพวัตถุ
ด้วยข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลาเฟรมที่ 33 35 37 ... 49

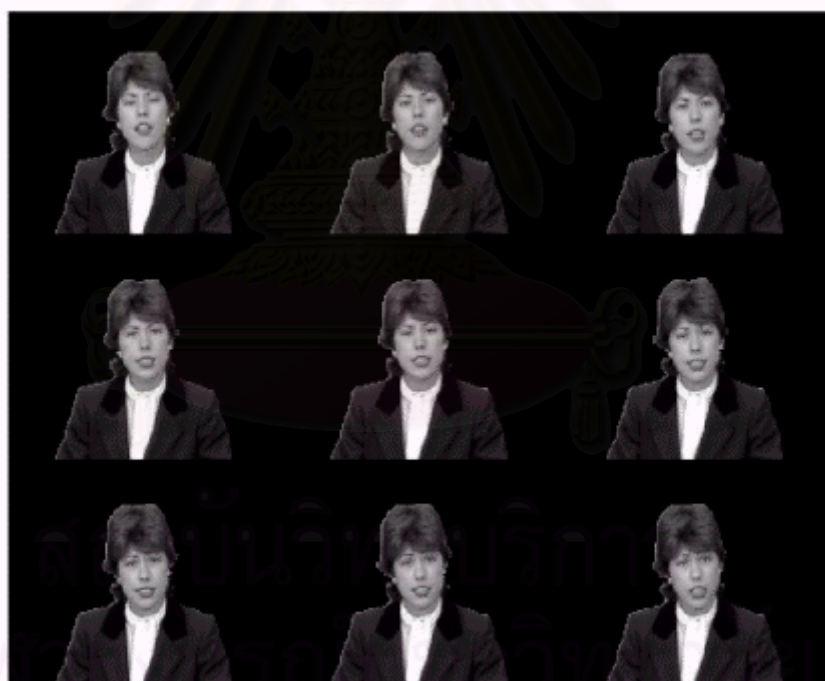
ผลการแยกส่วนเชิงพื้นที่ที่สามารถแยกจุดภาพที่มีความเข้มแสงใกล้เคียงกันเป็นบริเวณเดียวกันได้ แต่ไม่สามารถแยกเป็นภาพวัตถุที่มีความหมายได้ เนื่องจากมีการแยกส่วนมากเกินไปและผลจากมาสก์ที่ได้จากการตรวจรู้การเปลี่ยนแปลงจะเห็นว่าจะสามารถบอกตำแหน่งของวัตถุเคลื่อนที่ได้คร่าวๆ แต่ยังไม่สามารถกำหนดขอบเขตของภาพวัตถุที่ชัดเจนได้ อย่างไรก็ตามผลจากการรวมข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา รวมทั้งการติดตามการเคลื่อนที่จากข้อมูลภาพวัตถุที่ได้จากเฟรมก่อนหน้า จะสามารถกำหนดขอบเขตของภาพวัตถุได้อย่างชัดเจนยิ่งขึ้น แต่จะสังเกตได้ว่าภาพวัตถุที่ถูกแยกส่วนมาได้จะมีส่วนของพื้นหลังที่มีความเข้มใกล้เคียงกับภาพวัตถุมากแยกออกมาด้วย แต่ก็มีลักษณะที่มีความหมาย ส่วนการติดตามการเคลื่อนที่จากผลของภาพวัตถุที่ต่อเนื่องจะแสดงให้เห็นภาพวัตถุที่มีความหมายตั้งแต่เฟรมที่ 15 เป็นต้นไป ส่วนในเฟรมแรกๆ จะยังไม่สามารถตรวจรู้การเคลื่อนที่ของภาพเด็กได้เนื่องจากเด็กอยู่นิ่งตลอดเวลา แต่เมื่อเด็กมีการเคลื่อนที่แล้ว ก็สามารถตรวจรู้ภาพวัตถุได้ทั้งภาพแม่และเด็กได้อย่างมีความหมาย



รูปที่ 4.40 ภาพวัตถุของลำดับภาพ Erik จากกรรมวิธีที่เสนอ เฟรมที่ 41 ถึงเฟรมที่ 49



รูปที่ 4.41 ภาพวัตถุของลำดับภาพ Akiyo เฟรมที่ 9 13 17 ... 41



รูปที่ 4.42 ภาพวัตถุของลำดับภาพ Claire เฟรมที่ 11 15 19 ... 43



รูปที่ 4.43 ภาพวัตถุของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10

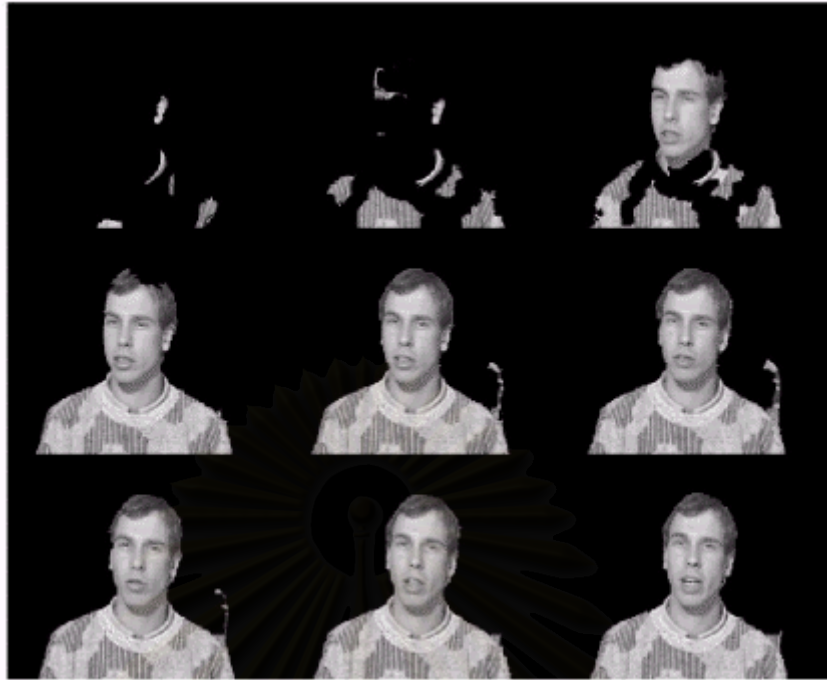
4.5 การวัดสมรรถนะของกรรมวิธีการแยกส่วนภาพวัตถุ

4.5.1 การวัดผลการแยกส่วนภาพเชิงอัตวิสัย (Subjective Test)

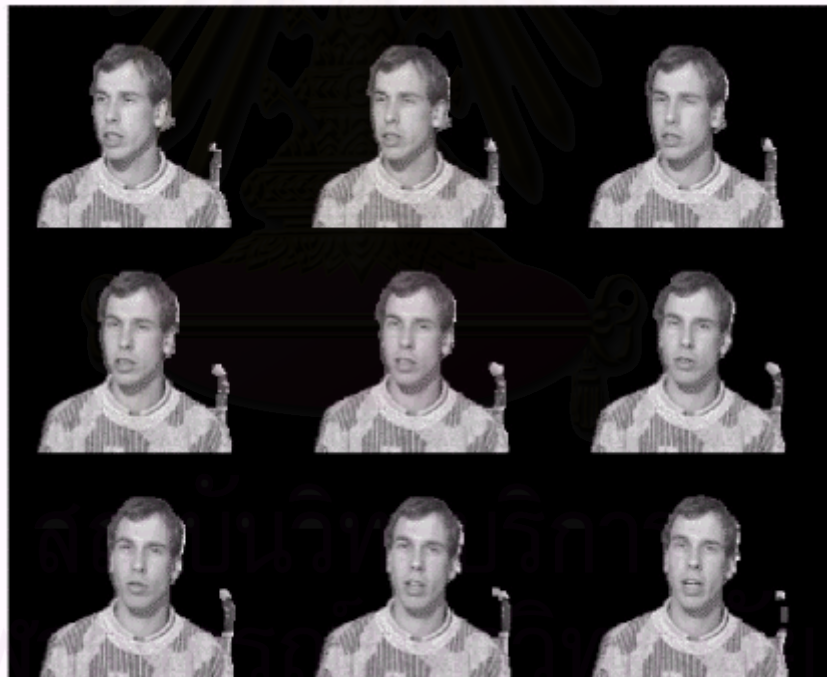
ผลการแยกส่วนภาพวัตถุจะพิจารณาเทียบกับผลของ COST211 โดยพิจารณาลักษณะของภาพที่มีความหมายและความเร็วในการตรวจรู้การเคลื่อนที่ของภาพวัตถุ

ในการเปรียบเทียบกับผลของ COST211 จะเปรียบเทียบผลของลำดับภาพที่ได้จาก COST211 และตรงกับลำดับภาพมาตรฐานที่มีอยู่ ซึ่งได้แก่ลำดับ Erik Mother & Daughter Table Tennis และ Hall Monitor

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.44 ภาพวัตถุของลำดับภาพ Erik เฟรมที่ 2 ถึง 10 จากผลของ COST211



รูปที่ 4.45 ภาพวัตถุของลำดับภาพ Erik เฟรมที่ 2 ถึง 10 จากกรรมวิธีที่เสนอ



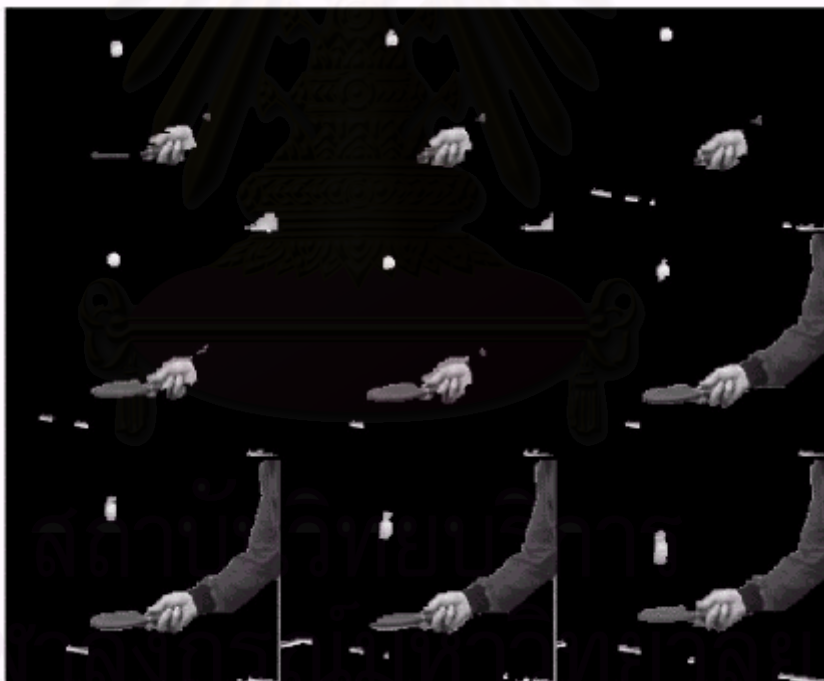
รูปที่ 4.46 ภาพวัตถุของลำดับภาพ Mother & Daughter เฟรมที่ 41 ถึง 49 จากผลของ COST211



รูปที่ 4.47 ภาพวัตถุของลำดับภาพ Mother & Daughter เฟรมที่ 41 ถึง 49 จากกรรมวิธีที่นำเสนอ



รูปที่ 4.48 ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10 จากผลของ COST211



รูปที่ 4.49 ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10 จากกรรมวิธีที่เสนอ



รูปที่ 4.50 ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 32 ถึง 40 จากผลของ COST211



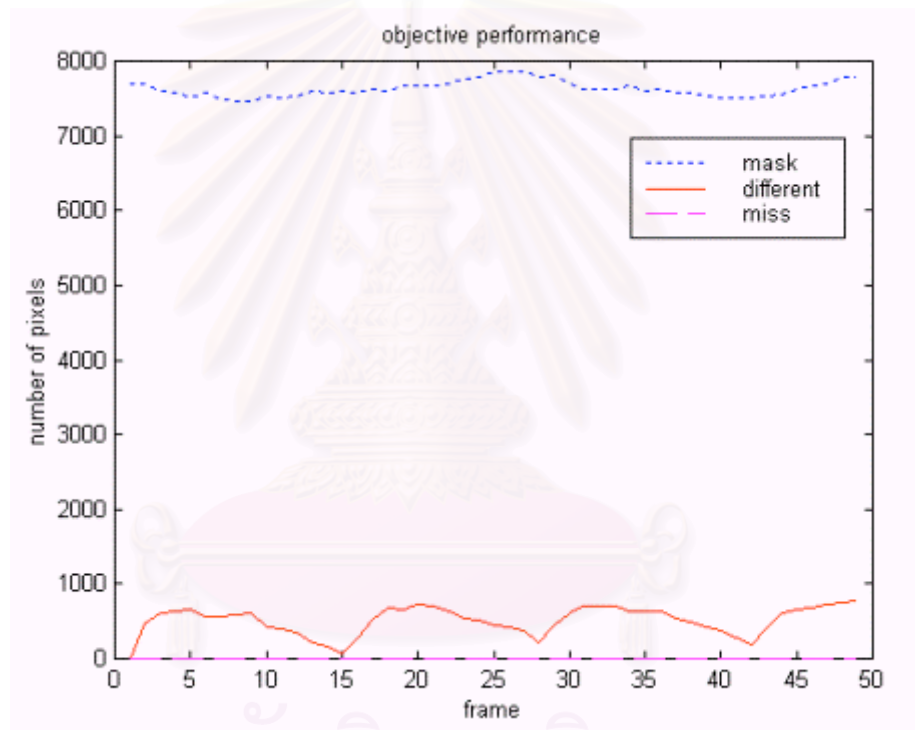
รูปที่ 4.51 ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 32 ถึง 40 จากกรรมวิธีที่เสนอ

4.5.2 การวัดผลการแยกส่วนภาพเชิงวัตถุ (Objective Test)

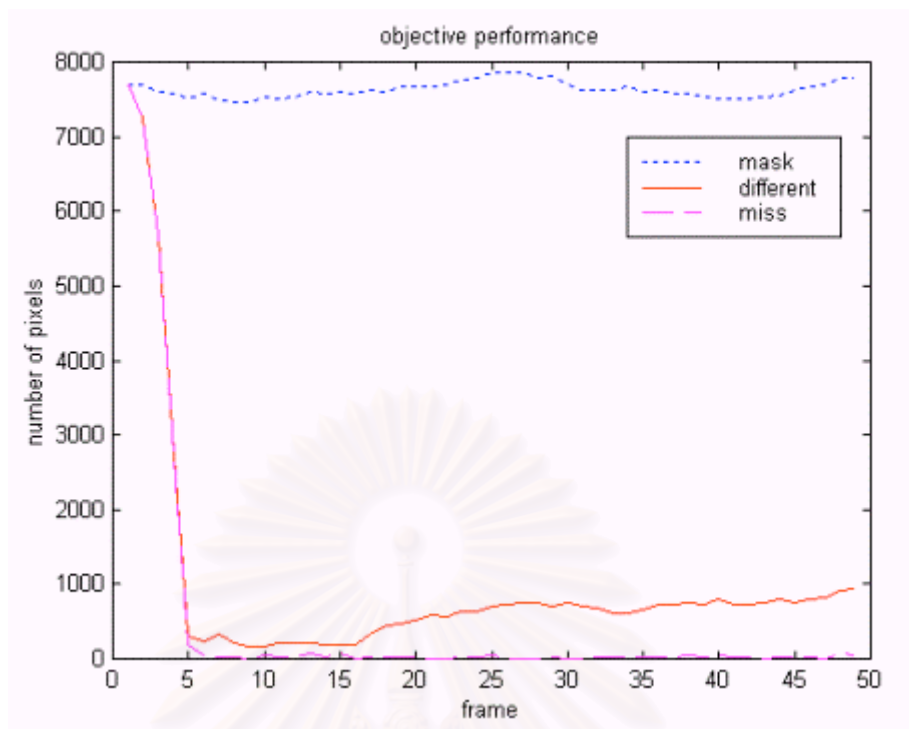
ความถูกต้องของการแยกส่วนภาพวัตถุจะวัดจากจำนวนจุดภาพที่เกิน และจำนวนจุดภาพที่ขาดไปของผลการแยกส่วนภาพวัตถุเคลื่อนที่ เมื่อเทียบกับมาสก์อ้างอิง รวมทั้งร้อยละของจำนวนจุดภาพที่แตกต่างกันเทียบกับจำนวนจุดภาพของมาสก์อ้างอิง



รูปที่ 4.52 การเปรียบเทียบผลการแยกส่วนภาพวัตถุระหว่างผลที่ได้จากกรรมวิธีที่เสนอ และมาสก์อ้างอิงของลำดับภาพ Erik (ก) มาสก์อ้างอิง (ข) ผลที่ได้



รูปที่ 4.53 รูปกราฟแสดงจำนวนจุดภาพที่ต่างกันระหว่างผลที่ได้ และมาสก์อ้างอิงของลำดับภาพ Erik ใน 50 เฟรม

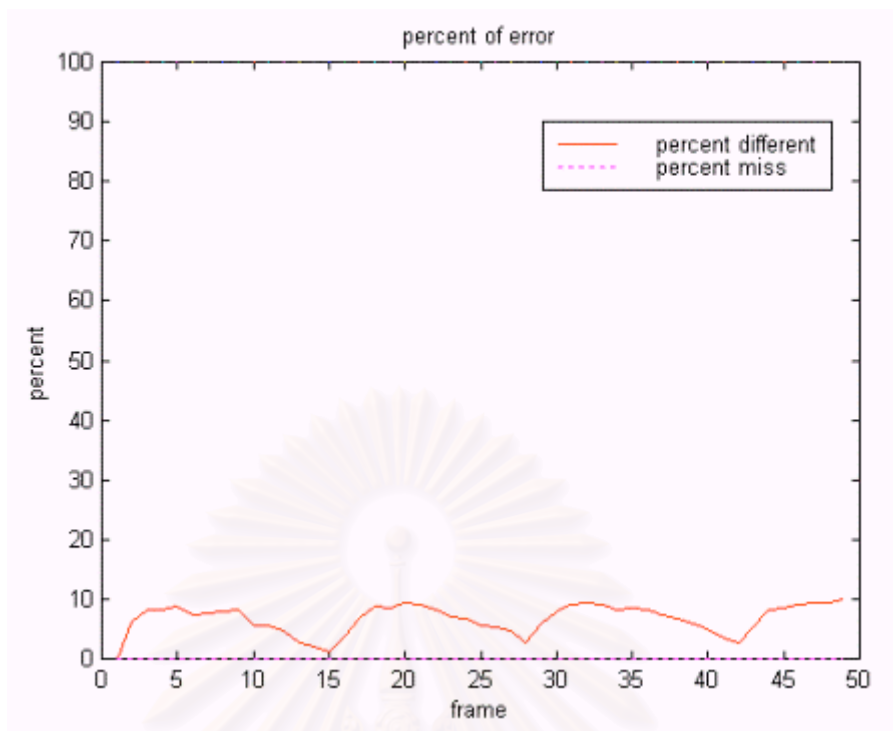


รูปที่ 4.54 รูปกราฟแสดงจำนวนจุดภาพที่ต่างกันระหว่างผลของ COST211 และมาสก์อ้างอิงของลำดับภาพ Erik ใน 50 เฟรม

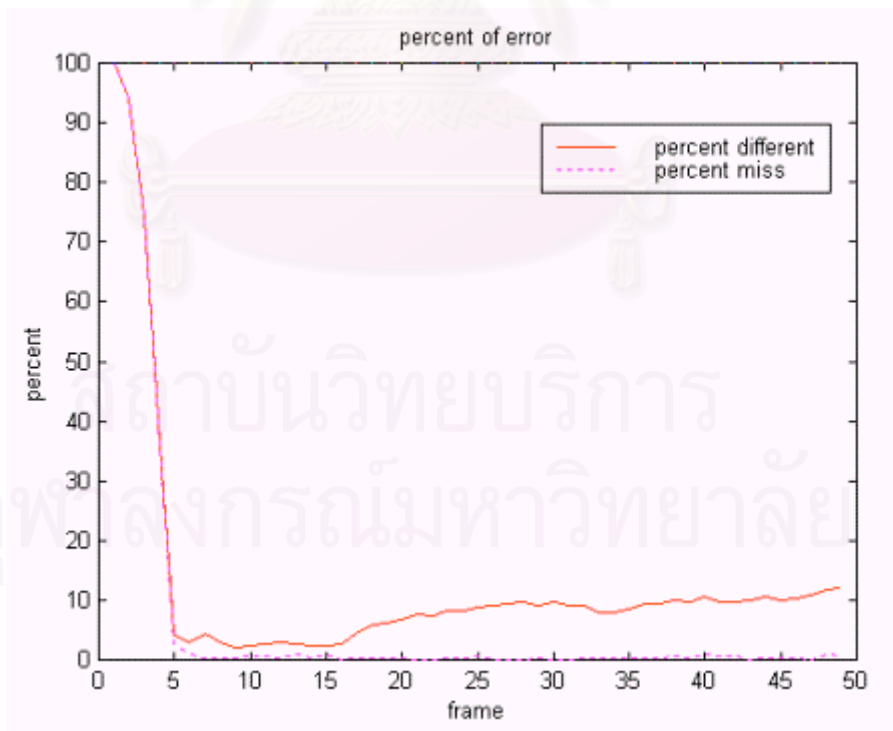
ตารางที่ 4.5 แสดงค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211 เทียบกับมาสก์อ้างอิงของลำดับภาพ Erik ใน 50 เฟรม

ค่าผิดพลาดเฉลี่ย	ผลที่ได้	ผลของ COST211
จุดภาพของมาสก์อ้างอิง (จุดภาพ)	7631	7631
จุดภาพที่แตกต่าง (จุดภาพ)	511	994
จุดภาพที่เกินไป (จุดภาพ)	511	492
จุดภาพที่ขาดไป (จุดภาพ)	0	502

ผลการแยกส่วนภาพวัตถุของลำดับภาพ Erik แสดงให้เห็นถึงภาพวัตถุที่เคลื่อนไหวและมีการเปลี่ยนแปลงของความเข้มแสง จากผลจะเห็นว่าสามารถแยกส่วนทั้งบริเวณของศีรษะและลำตัวได้อย่างชัดเจน เพียงแต่มีการแยกส่วนผิดพลาดในส่วนของเงาเนื่องจากบริเวณที่เกิดเงาจะมีความเข้มแสงเปลี่ยนแปลงไปในแต่ละเฟรม แม้ว่าในส่วนของเงาจะได้ใช้การปฏิบัติการแบบเปิดกำจัดไปได้บ้างแล้ว แต่ก็ยังไม่สามารถกำจัดได้ทั้งหมด ดังนั้นในส่วนของการกำจัดเงาจึงยังมีการวิจัยในหัวข้อนี้อยู่



รูปที่ 4.55 ร้อยละของจำนวนจุดภาพที่แตกต่างกันระหว่างผลที่ได้และมาสก์อ้างอิง เมื่อเทียบกับขนาดของมาสก์อ้างอิงของลำดับ Erik



รูปที่ 4.56 ร้อยละของจำนวนจุดภาพที่แตกต่างกันระหว่างผลของ COST211และมาสก์อ้างอิง เมื่อเทียบกับขนาดของมาสก์อ้างอิงของลำดับ Erik

ตารางที่ 4.6 แสดงร้อยละของค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST 211 เทียบกับมาตรฐานอ้างอิงของลำดับภาพ Erik ใน 50 เฟรม

ค่าผิดพลาดเฉลี่ย	ผลที่ได้	ผลของ COST211
จุดภาพที่แตกต่าง (ร้อยละ)	6.7	12.9
จุดภาพที่เกินไป (ร้อยละ)	6.7	6.4
จุดภาพที่ขาดไป (ร้อยละ)	0	6.5

จากรูปที่ 4.55 และรูปที่ 4.56 จะแสดงให้เห็นความคลาดเคลื่อนของผลที่ได้และผลของ COST211 จะเห็นว่าผลที่ได้จะมีความคลาดเคลื่อนไม่เกินร้อยละ 10 ของขนาดภาพวัตถุและสามารถตรวจรู้ภาพวัตถุได้ครบทั้งวัตถุตลอดลำดับภาพ ซึ่งในบางเฟรมสามารถแยกส่วนภาพวัตถุได้ใกล้เคียงมาตรฐานอ้างอิงมาก ส่วนบางเฟรมเกิดการแยกส่วนภาพวัตถุเกินเนื่องจากผลกระทบของเงาวัตถุ และยังสังเกตได้ว่ากรรมวิธีที่เสนอมีผลการแยกส่วนภาพวัตถุได้ตั้งแต่เฟรมที่สอง ส่วนผลของ COST211 จะสามารถแยกส่วนได้ทั้งภาพวัตถุได้หลังจากเฟรมที่ 5 เป็นต้นไป นอกจากนี้ค่าผิดพลาดเฉลี่ยของผลที่ได้มีค่าต่ำกว่าผลของ COST211 ด้วย

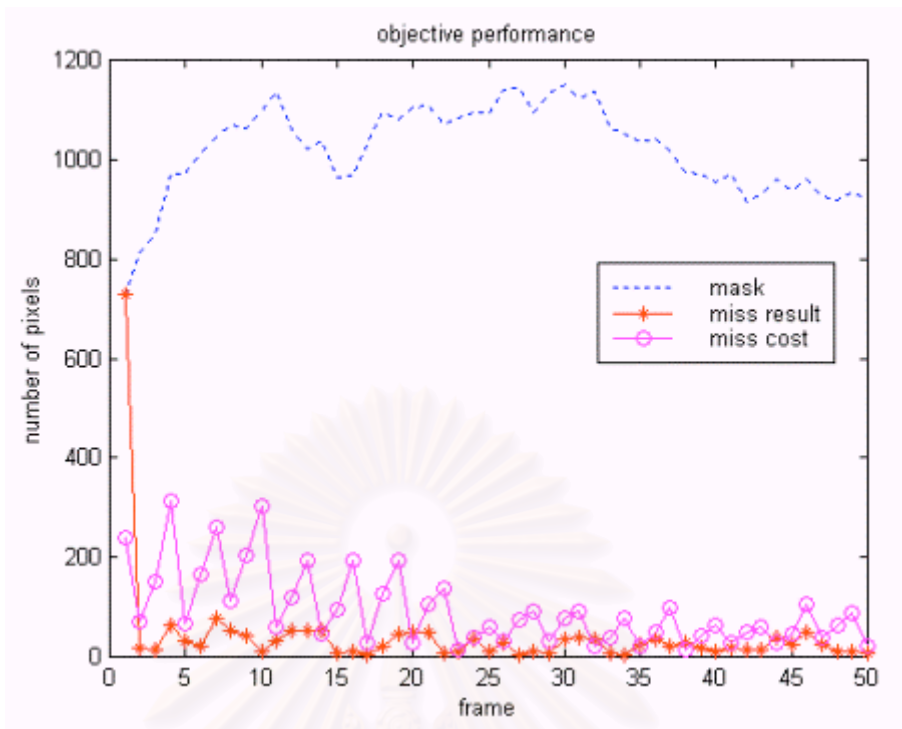
ผลของความแตกต่างเมื่อเปรียบเทียบกับมาตรฐานอ้างอิง จะเห็นว่าหลังจากเฟรมที่ 5 เป็นต้นไป ค่าความผิดพลาดของภาพวัตถุที่ได้ของผลทั้งสองจะไม่เกินร้อยละ 10 ของขนาดภาพวัตถุ เนื่องจากทั้งสองวิธีสามารถรักษาการติดตามการเคลื่อนที่ของภาพวัตถุได้ตลอดลำดับภาพ



(ก)

(ข)

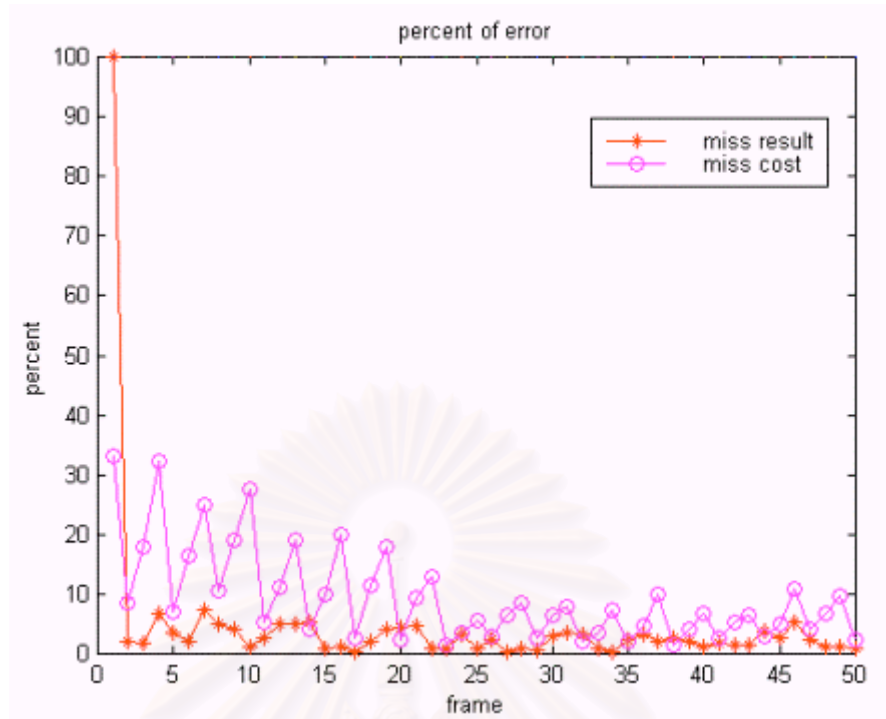
รูปที่ 4.57 การเปรียบเทียบผลการแยกส่วนภาพวัตถุระหว่างผลที่ได้และมาตรฐานอ้างอิงของลำดับภาพ Hall Monitor (ก) มาตรฐานอ้างอิง (ข) ผลที่ได้



รูปที่ 4.58 รูปกราฟแสดงจำนวนจุดภาพที่แยกส่วนขาดไประหว่างผลที่ได้และผลของ COST211 ของลำดับภาพ Hall Monitor ใน 50 เฟรม

ตารางที่ 4.7 แสดงค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211 เทียบกับมาสก์อ้างอิงของลำดับภาพ Hall Monitor ใน 50 เฟรม

ค่าผิดพลาดเฉลี่ย	ผลที่ได้	ผลของ COST211
จุดภาพของมาสก์อ้างอิง (จุดภาพ)	1019	1019
จุดภาพที่แตกต่าง (จุดภาพ)	754	1437
จุดภาพที่เกินไป (จุดภาพ)	714	1343
จุดภาพที่ขาดไป (จุดภาพ)	40	93



รูปที่ 4.59 แสดงร้อยละของจำนวนจุดภาพที่แยกส่วนขาดไปของผลที่ได้และผลของ COST211 เมื่อเทียบกับมาสก์อ้างอิงของลำดับภาพ Hall Monitor

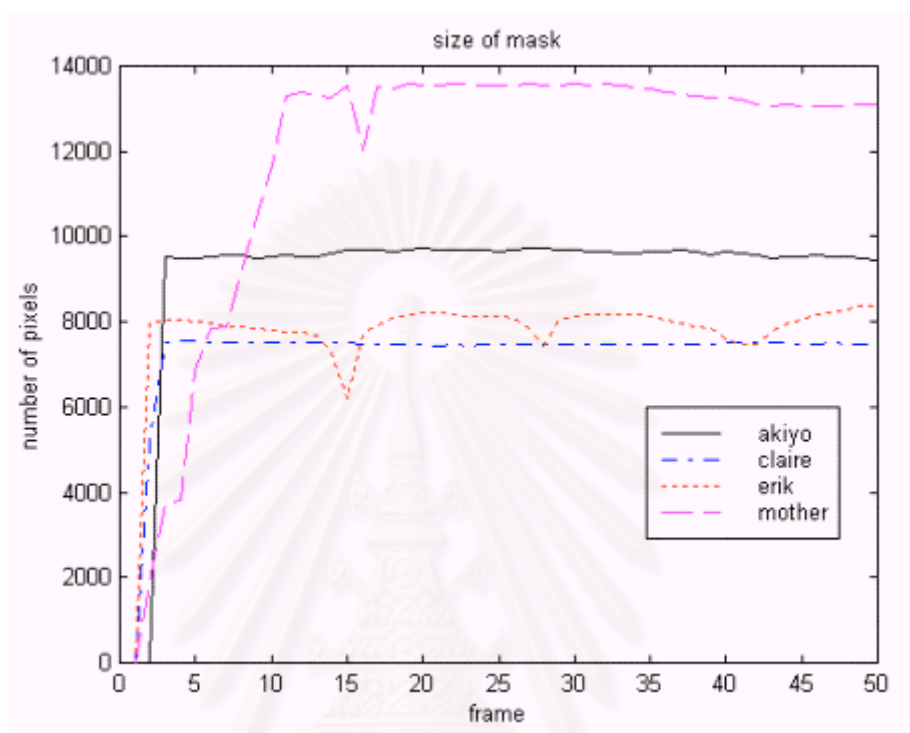
ตารางที่ 4.8 แสดงร้อยละของค่าผิดพลาดเฉลี่ยของผลที่ได้จากการแยกส่วนภาพวัตถุและผลของ COST211 เทียบกับมาสก์อ้างอิงของลำดับภาพ Hall Monitor ใน 50 เฟรม

ค่าผิดพลาดเฉลี่ย	ผลที่ได้	ผลของ COST211
จุดภาพที่แตกต่าง (ร้อยละ)	73.8	139.8
จุดภาพที่เกินไป (ร้อยละ)	69.3	130.5
จุดภาพที่ขาดไป (ร้อยละ)	4.5	9.3

จากรูปที่ 4.59 จะแสดงความผิดพลาดของภาพวัตถุตั้งแต่เฟรมที่ 21 ถึง 50 เนื่องจากลำดับภาพ Hall Monitor ในช่วงเฟรมก่อนนั้นจะไม่มีวัตถุเคลื่อนที่ปรากฏให้เห็น หลังจากเฟรมที่ 20 ภาพวัตถุเคลื่อนที่ที่จะปรากฏออกมา ซึ่งจะเห็นว่าผลที่ได้จะมีความผิดพลาดน้อยกว่าผลของ COST211 โดยแยกส่วนภาพวัตถุขาดไปไม่เกินร้อยละ 5 ของขนาดมาสก์อ้างอิง แต่เนื่องจากวัตถุที่เคลื่อนที่มีขนาดค่อนข้างเล็ก จึงมีการแยกส่วนภาพเกินไปค่อนข้างมาก แต่ผลที่ได้ยังมีการแยกส่วนภาพเกินไปน้อยกว่าผลของ COST211 ผลที่ได้จะสามารถแยกส่วนภาพวัตถุที่เคลื่อนที่ได้ทั้งหมดแต่จะมีส่วนที่แยกส่วนเกินไปบ้าง เนื่องจากมีรายละเอียดของภาพค่อนข้างซับซ้อนและมีการเปลี่ยนแปลงความเข้มแสงที่พื้นหลัง จึงทำให้มีการแยกส่วนพื้นหลังที่ติดกับภาพวัตถุมาด้วย

4.5.3 ความต่อเนื่องของภาพวัตถุ

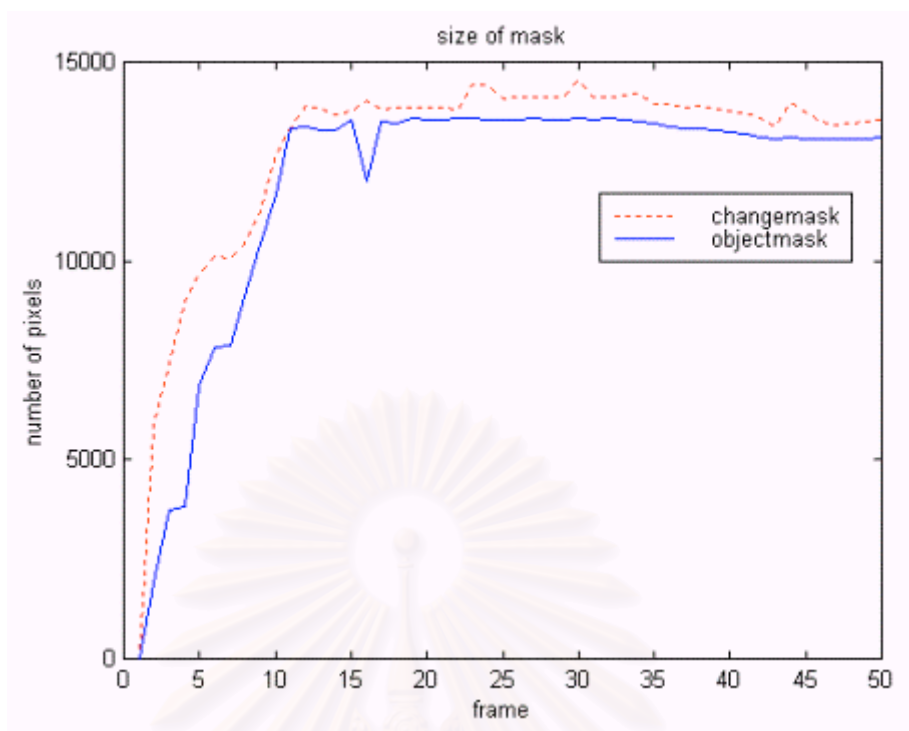
ความต่อเนื่องของการแยกส่วนภาพวัตถุจะวัดจำนวนจุดภาพของภาพวัตถุใน 50 เฟรม ซึ่งความต่อเนื่องที่ดีจะต้องมีขนาดของภาพวัตถุสม่ำเสมอ



รูปที่ 4.60 รูปกราฟแสดงความต่อเนื่องของภาพวัตถุของลำดับภาพ Akiyo Claire Erik และ Mother & Daughter เฟรมที่ 1 ถึง 50

สำหรับลำดับภาพ Akiyo และ Claire จะสังเกตได้ว่าสามารถตรวจรู้ภาพวัตถุเคลื่อนที่ได้ง่ายเนื่องจากเป็นลำดับภาพที่มีพื้นหลังอยู่นิ่งและเกิดสัญญาณรบกวนจากพื้นหลังน้อยมาก ทำให้ภาพวัตถุที่ได้จากการแยกส่วนภาพมีความต่อเนื่อง ซึ่งจากกราฟจะเห็นว่าความต่อเนื่องของภาพวัตถุสำหรับลำดับภาพ Akiyo และ Claire จะดีมากเนื่องจากขนาดของภาพวัตถุจะค่อนข้างคงที่เสมอตลอดลำดับภาพ

ส่วนลำดับภาพ Erik และ Mother & Daughter จะมีบางช่วงของลำดับภาพที่ไม่สามารถตรวจรู้การเคลื่อนที่ของวัตถุได้ ทำให้ความต่อเนื่องของจำนวนจุดภาพลดลงไปบางช่วง แต่ภาพวัตถุส่วนใหญ่ในลำดับภาพจะมีขนาดของภาพวัตถุค่อนข้างสม่ำเสมอ มีเพียงช่วงแรกของลำดับภาพ Mother & Daughter ที่ไม่มีการเคลื่อนที่ของวัตถุ จึงไม่สามารถตรวจรู้การเคลื่อนที่ของบางวัตถุได้ แต่ในช่วงหลังที่ตรวจรู้การเคลื่อนที่ได้แล้ว ภาพวัตถุที่ได้จะมีขนาดสม่ำเสมอ



รูปที่ 4.61 รูปกราฟแสดงจำนวนจุดภาพของมาสก์ของลำดับภาพ Mother&Daughter เฟรมที่ 1 ถึง 100



รูปที่ 4.62 ภาพที่ได้จากการแยกส่วนเชิงพื้นที่และมาสก์ของภาพวัตถุในเฟรมที่ 16 ของลำดับภาพ Mother & Daughter

(ก) ภาพบริเวณจากการแยกส่วนเชิงพื้นที่ (ข) มาสก์ของภาพวัตถุ

จากกราฟแสดงจำนวนจุดภาพของมาสก์จากการแยกส่วนภาพวัตถุในลำดับภาพ Mother & Daughter จะเห็นว่าในช่วงแรกจะยังไม่สามารถตรวจรู้ภาพวัตถุที่เคลื่อนที่ทั้งหมดได้ เนื่องจากภาพวัตถุบางส่วนยังไม่แสดงการเคลื่อนที่ แต่หลังจากเฟรมที่ 10 ไปแล้วภาพวัตถุทั้งหมดทั้งแม่และลูกมีการเคลื่อนที่ทั้งหมดจึงสามารถตรวจรู้และติดตามการเคลื่อนที่ได้ แต่เมื่อสังเกตจากกราฟรูปที่ 4.61 จะเห็นว่าในเฟรมที่ 16 มาสก์ของภาพวัตถุจะมีจำนวนจุดภาพลดลง ซึ่งแสดงผลภาพของเฟรมที่ 16 ให้เห็นในรูปที่ 4.62 จำนวนจุดภาพจะลดลงเนื่องจากการตัดลินใจในเฟรมนี้ผิดพลาด เกิดจากการแยกส่วนเชิงพื้นที่ที่กำหนดบริเวณมือที่ยกขึ้นมาเป็นบริเวณของพื้นหลัง

เพราะส่วนของมือมีความเข้มแสงใกล้เคียงกับพื้นหลังมาก แต่กรรมวิธีการแยกส่วนนี้สามารถใช้การติดตามการเคลื่อนที่เพื่อแก้ไขผลในเฟรมต่อไปได้อย่างดีดังแสดงให้เห็นว่าจำนวนจุดภาพหลังจากนั้นจนถึงเฟรมที่ 100 จะมีจำนวนจุดภาพค่อนข้างคงที่ และให้ผลการแยกส่วนภาพวัตถุเคลื่อนที่ที่ชัดเจนและมีความหมายตลอดทั้งลำดับภาพ



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สรุปและข้อเสนอแนะ

5.1 สรุป

วิทยานิพนธ์นี้สรุปความรู้พื้นฐานและกระบวนการของการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุตามข้อกำหนดของ MPEG-4 และได้เสนอกรรมวิธีการแยกส่วนภาพวัตถุสำหรับสัญญาณวิดีโอ ซึ่งเป็นขั้นตอนที่สำคัญก่อนขั้นตอนการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุ กรรมวิธีที่เสนอนี้มีคุณสมบัติตามข้อกำหนดของมาตรฐาน MPEG-4 ซึ่งเป็นการแยกส่วนของสัญญาณวิดีโอเป็นภาพวัตถุเคลื่อนที่ที่ต่อเนื่อง และสามารถที่จะแยกส่งภาพวัตถุที่ต้องการซึ่งได้จากการแยกส่วนภาพในการสื่อสารทางวิดีโอได้ เพื่อสนับสนุนฟังก์ชันบนพื้นฐานของคอนเทนตีในการสื่อสารด้านมัลติมีเดีย กรรมวิธีการแยกส่วนภาพวัตถุของสัญญาณวิดีโอที่นำเสนอในวิทยานิพนธ์ฉบับนี้ อยู่บนพื้นฐานของการใช้คุณลักษณะสำคัญของสัญญาณวิดีโอ ทั้งข้อมูลเชิงพื้นที่และข้อมูลเชิงเวลา ข้อมูลเชิงพื้นที่ได้มาจากข้อมูลของความเข้มแสง ซึ่งภาพในแต่ละเฟรมจะถูกแยกส่วนภาพเชิงพื้นที่เป็นบริเวณที่มีความเข้มต่อเนื่องกัน เริ่มด้วยการปรับปรุงภาพให้มีความเรียบง่ายต่อการแยกส่วนโดยใช้การปฏิบัติการเชิงสัณฐาน (Morphological Operation) การตรวจรู้ขอบเขตของบริเวณในภาพโดยใช้การประมาณเกรเดียนต์ และการกำหนดบริเวณให้แก่จุดภาพที่มีความเข้มใกล้เคียงกันโดยกรรมวิธีสันปันน้ำ (Watershed Algorithm) ส่วนข้อมูลเชิงเวลานั้นจะได้รับการตรวจรู้การเปลี่ยนแปลงโดยการพิจารณาการเปลี่ยนแปลงด้วยค่าจุดเปลี่ยนที่ปรับตัวได้อัตโนมัติ ซึ่งวิเคราะห์จากค่าการเปลี่ยนแปลงเฉลี่ยของเฟรมที่ต่อเนื่องกัน และพิจารณาร่วมกับข้อมูลที่ได้จากการแยกส่วนภาพในเฟรมก่อนหน้าเพื่อช่วยในการติดตามการเคลื่อนที่ของวัตถุ หลังจากนั้นจึงนำข้อมูลที่ได้ในแต่ละส่วนมาผ่านกระบวนการตัดสินใจเพื่อให้สามารถแยกส่วนภาพวัตถุได้อย่างมีความหมาย โดยใช้ข้อมูลเชิงเวลาในการระบุบริเวณที่มีการเคลื่อนที่และข้อมูลเชิงพื้นที่ในการกำหนดขอบเขตของภาพวัตถุที่แม่นยำ

ผลการวิจัยแสดงให้เห็นว่าการปรับปรุงภาพให้มีความเรียบง่ายโดยใช้การปฏิบัติการแบบเปิด-ปิดจะช่วยลดจำนวนบริเวณที่เกิดขึ้นจากการแยกส่วนภาพเชิงพื้นที่ ซึ่งจะแก้ปัญหาการแยกส่วนมากเกินไปของกรรมวิธีสันปันน้ำ และการประมาณเกรเดียนต์แบบหลายชั้นยังช่วยลดจำนวนของจุดภาพบริเวณขอบที่ยังไม่ระบุบริเวณเมื่อเทียบกับการประมาณเกรเดียนต์แบบชั้นเดียว จึงช่วยลดการคำนวณในการกำหนดบริเวณให้กับจุดภาพที่ยังไม่ระบุบริเวณของกรรมวิธีสันปันน้ำได้ นอกจากนี้การตรวจรู้การเปลี่ยนแปลงที่ใช้การปรับตัวแบบอัตโนมัติของค่าจุดเปลี่ยนตามการวิเคราะห์การเปลี่ยนแปลงของลำดับภาพ ยังสามารถตรวจรู้การเคลื่อนที่ของวัตถุได้ และ

จะใช้ตัวปฏิบัติการแบบปิด-เปิด และการกำจัดบริเวณที่มีขนาดเล็กออก เพื่อเติมเต็มภาพวัตถุและกำจัดสัญญาณรบกวน ทำให้สามารถตรวจวัตถุที่ต้องการได้ สุดท้ายกระบวนการตัดสินใจที่ใช้หน่วยความจำเก็บข้อมูลที่ได้จากการแยกส่วนในเฟรมก่อนหน้านี้ สามารถปรับปรุงผลที่ไม่ได้ใช้หน่วยความจำในเฟรมที่วัตถุเคลื่อนที่หยุดนิ่งได้ด้วย จากผลการวิจัยกรรมวิธีที่เสนอขึ้นสามารถแยกส่วนภาพวัตถุที่เคลื่อนที่ได้อย่างมีความหมาย โดยเมื่อเปรียบเทียบกับผลของ COST211 กรรมวิธีที่เสนอจะสามารถแยกส่วนภาพวัตถุได้องค์ประกอบของภาพครบเร็วกว่าผลของ COST211 และเมื่อเปรียบเทียบกับมาสก์อ้างอิง ผลที่ได้จะมีความคลาดเคลื่อนเฉลี่ยไม่เกินร้อยละ 15 ของขนาดภาพวัตถุ และจะมีผลที่ได้จากการแยกส่วนภาพวัตถุขาดไปไม่เกินร้อยละ 5 ของขนาดภาพวัตถุสำหรับลำดับภาพการประชุมทางวิดีโอที่นำมาทดสอบ

5.2 ข้อเสนอแนะ

ปัญหาระหว่างการดำเนินงานวิจัย

1. ข้อมูลมาตรฐานที่ใช้ในการทำงานวิจัยค่อนข้างหายาก เช่น มาตรฐานและเอกสารเกี่ยวกับการเข้ารหัสสัญญาณวิดีโอและลำดับภาพมาตรฐานที่นำมาทดสอบ เนื่องจากเป็นงานวิจัยยังเริ่มต้นใหม่อยู่สำหรับประเทศไทย
 2. การร่วมงานวิจัยระหว่างสถาบันต่าง ๆ ทั้งในประเทศและต่างประเทศยังมีน้อยอยู่
 3. การหาแบบจำลองการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 ค่อนข้างยาก เนื่องจากการเข้ารหัสสัญญาณวิดีโอบนพื้นฐานของหน่วยวัตถุยังเป็นเทคโนโลยีใหม่อยู่
- งานวิจัยที่ควรได้รับการศึกษาและพัฒนาต่อ
1. การพัฒนากรรมวิธีการแยกส่วนภาพวัตถุที่สามารถตรวจรู้ภาพวัตถุที่มีขนาดเล็กได้ดียิ่งขึ้นโดยการวิเคราะห์ค่าจุดเบี่ยงเบนอย่างอัตโนมัติ
 2. การหากรรมวิธีการแทนรูปร่างของขอบเขตภาพวัตถุที่ได้จากการแยกส่วนภาพให้มีรูปร่างเรียบง่ายขึ้นและยังรักษาบริเวณของภาพวัตถุไว้ได้อย่างครบถ้วน
 3. การใช้เวกเตอร์เคลื่อนที่ที่ได้จากการประมาณการเคลื่อนที่มาช่วยในการแยกส่วนภาพวัตถุเป็นเลเยอร์ของภาพวัตถุที่มีการเคลื่อนที่แตกต่างกัน

รายการอ้างอิง

- Aach, T., Kaup, A., and Mester, R., Statistical Model Based Change Detection in Moving Video, IEEE Transactions on Signal Processing. Vol 31 (March 1993): 165-180.
- Alatan, A. A., Onural, L., and Wollborn, M., Image Sequence Analysis for Emerging Interactive Multimedia Services - The European COST 211 Framework, IEEE Transactions on Circuits and Systems for Video Technology. Vol 8 (November 1998): 802-813.
- Alatan, A. A., Tuncel, E., and Onural, L. A Rule based Method for Object Segmentation in Video Sequences, Proceedings of International Conference on Image Processing. Vol 2 (1997): 522-525.
- Bober, M., and Kittler, J., Video Coding for Mobile Communications – MPEG-4 Perspective, IEE Savoy Place, London. Vol 3 (1996): 1-9.
- Castagno, R., Ebrahimi, T., and Kunt, M., Video Segmentation Based on Multiple Features for Interactive Multimedia Applications IEEE Transactions on Circuits and Systems for Video Technology. Vol 8 (September 1998): 562-571.
- Cho, S. H., and Choi, H. M. Modified Morphological Filter Based Reduction of Oversegmentation in Watershed Transformation, Proceedings of ITC-CSCC. Vol 1 (1999): 269-272.
- Choi, J. G., Kim, M., Lee, M. H., and Ahn, C. (ETRI) ISO/IEC JTC1/SC29/WG11 MPEG97/m2091 Automatic Segmentation Based on Spatio-Temporal Information, Bristol: ISO/IEC, 1997.
- Choi, J. G., Lee, S. W., and Kim, S. D. Spatio-Temporal Video Segmentation Using a Joint Similarity Measure, IEEE Transactions on Circuits and Systems for Video Technology. Vol 7 (April 1997): 279-286
- Cote, G., Erol, B., Gallant, M. and Kossentini, F., H.263+: Video Coding at Low Bit Rates, IEEE Transactions on Circuits and Systems for Video Technology. Vol 8 (November 1998): 849-866.
- Crinon, R., and Sezan, I., Sprite Based Video Coding Using Online Segmentation, IEEE

- Transactions on Acoustics Speech and Signal Processing. Vol 5 (1998): 2981-2984.
- Diehl, N., Object-Oriented Motion Estimation and Segmentation in Image Sequence, IEEE Transactions on Signal Processing on Image Communication. Vol 3 (1991): 23-56.
- Dufaux, F., Moscheni, F., and Lippman, A. Spatio-Temporal Segmentation based on Motion and Static Segmentation, Proceedings of International Conference on Image Processing. Vol 1 (1995): 306-309.
- Georgi, D., Borshukov, Bozdagi, G., Altunbasak, Y., and Tekalp, A. M. Motion Segmentation by Multistage Affine Classification, IEEE Transactions on Image Processing. Vol 6 (November 1997): 1591-1594.
- Gonzalez, R. C. and Woods, R. E. Digital Image Processing, :Addison-Wesley Publishing Company, 1993.
- Hotter, M., and Thoma, R., Image Segmentation Based on Object Oriented Mapping Parameter Estimation, IEEE Transactions on Signal Processing. Vol 15 (1998): 315-334.
- Kim, J. I., and Evans, B. L. Predictive Shape Coding Using Generic Polygon Approximation, Proceedings of IEEE International Symposium on Circuits and Systems. Vol 5 (1998): 277-280.
- Kim, M., Choi, J. G., and Ho, Y. S. A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information IEEE Transactions on Circuits and Systems for Video Technology. Vol 9 (December 1999): 1216-1226.
- Koenen, R. ISO/IEC JTC1/SC29/WG11/ N2459 Overview of the MPEG-4 Standard, Atlantic City: ISO/IEC, 1998.
- Koenen, R. ISO/IEC JTC1/SC29/WG11/ N2502a Final Draft of International Standard Coding of Moving Pictures and Audio. Atlantic City: ISO/IEC, 1998.
- Kwok, S. H., and Constantinides, A. G. A Fast Recursive Shortest Spanning Tree for Image Segmentation and Edge Detection, IEEE Transactions on Image Processing. Vol 6 (February 1997): 328- 332.

- Lau, K. S., and Wade, G. Spatial Spectral Clustering using Recursive Spanning Trees, IEEE Proceedings on Communications, Speech and Vision. Vol 138 (August 1991): 232-238.
- Lin, I. J., and Kung, S. Y. Circular Viterbi Based Adaptive System for Automatic Video Object Segmentation, IEEE Second Workshop on Multimedia Signal Processing. Vol 1 (1998): 541-546.
- Mech, R., and Wollborn, M. A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera, IEEE Transactions on Signal Processing. Vol 66 (1998): 203-217.
- Mech, R., and Wollborn, M., A Noise Robust Method for Segmentation of Moving Objects in Video Sequences, Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing. Vol 4 (1997): 2657-2660.
- Meier, T., Ngan, K. N., Automatic Video Sequence Segmentation Using Object Tracking, Proceedings of TENCON 97 IEEE Region Annual Conference on Speech and Image Technologies for Computing and Telecommunications. (1997): 283-286.
- Meier, T., Ngan, K. N., Segmentation and Tracking of Moving Objects for Content Based Video Coding, IEE Proceedings on Visual Image and Signal Processing. Vol 146 (June 1999): 144-150.
- Meier, T., Ngan, K. N., Video Segmentation for Content Based Coding, IEEE Transactions on Circuits and Systems for Video Technology. Vol 9 (December 1999): 1190-1203.
- Moscheni, F. Spatio-Temporal Segmentation and Object Tracking: An Application to Second Generation Video Coding. Ph.D. Thesis, Electrical and Electronic Engineering, Engineering, Swiss Federal Institute of Technology, 1997.
- Neri, A., Colonnese, S., Russo, G., and Ralone, P., Automatic Moving Object and Background Separation, IEEE Transactions on Signal Processing. Vol 66 (1998): 219-232.
- Rao, K. R., and Hwang, J. J. Techniques and Standards for Image, Video and Audio Coding, : Prentice Hall PTR, 1996.

- Salembier, P., and Pardas, M. Hierarchical Morphological Segmentation for Image Sequence Coding, IEEE Transactions on Image Processing. Vol 3 (September 1994): 639-651.
- Salembier, P., and Serra, J., Morphological Multiscale Segmentation of Images, IEEE Transactions on SPIE. Vol 1 (1992): 620-631.
- Salembier, P., Morphological Multiscale Segmentation for Image Coding, IEEE Transactions on Signal Processing. Vol 38 (July 1994): 359-386.
- Schafer, R., MPEG-4 : A Multimedia Compression Standard for Interactive Applications and Services, IEEE Transactions on Electronics and Communication Engineering. Vol 10 (December 1998): 253-262.
- Sikora, T., MPEG-4 Very Low Bit Rate Video, Proceedings of International Symposium on Circuits and Systems for Video Technology. Vol 1 (June 1997): 1440-1443.
- Sikora, T., and Chiariglione, L., MPEG-4 Video and its Potential for Future Multimedia Services, Proceedings of International Symposium on Circuits and Systems. Vol 1 (June 1997): 1468-1471.
- Sikora, T., The MPEG-4 Video Standard Verification Model, IEEE Transactions on Circuits and Systems for Video Technology. Vol 7 (February 1997): 19-31.
- Tancharoen D., Jitapunkul S., Kittipanya-ngam P., and Siritarangkul N. Video Segmentation Based on Adaptive Combination of Multiple Features Proceedings of International Conference on Visual Communication and Image Processing. Vol 2 (June 2000): 1099-1106.
- Tancharoen D., Jitapunkul S., and Chompun S. Spatial Segmentation Based on Modified Morphological Tools Proceedings of International Conference on Information Technology : Coding and Computing. Vol 1 (April 2001): 478-482.
- Thoma, R., and Bierling, M. Motion Compensating Interpolation Considering Covered and Uncovered Background IEEE Transactions on Signal Processing on Image Communication. Vol 1 (1989): 191-212.
- Vass, J., Palaniappan, K., and Zhuang, X. Automatic Spatio-Temporal Video Sequence Segmentation, Proceedings of International Conference on Image Processing. Vol

1 (1998): 958-962.

Villegas P., Marichal X., Salcedo A. Objective Evaluation of Segmentation Masks in Video Sequences Proceedings of WIAMIS 99 Workshop Berlin. (May 1999): 85-88.

Vincent, L., and Soille, P. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 13 (June 1991): 583-598.

Wak, J. K., Jeon, J. G., and Kim, M. A Semi Automatic Video Segmentation Method for Object Oriented Multimedia Applications, Proceedings of The International Technical Conference on Circuits and Systems, Computers and Communications (ITC-CSCC'99). Vol 1 (1999): 114-117.

Wang, D., A Multiscale Gradient Algorithm for Image Segmentation Using Watersheds. IEEE Transactions on Pattern Recognition. Vol 30 (1997): 2043-2052.

Wang, J. Y. A., and Adelson, E. H. Representing Moving Images with Layers IEEE Transactions on Image Processing. Vol 3 (September 1994): 625-638.

Xia, X., and Shi, Y. Q. A Thresholding Hierarchical Block Matching Algorithm for Motion Estimation, IEEE Proceedings of International Symposium on Circuits and Systems. Vol 2 (1996): 624-627.

Yemez, Y., Sankur, B., Anarim, E. Region Growing Motion Segmentation and Estimation in Object Oriented Video Coding, Proceedings of International Conference on Image Processing. Vol 2 (1996): 521-524.

Zhang, K., and Kittler, J. Image Sequence Coding Using Multiple Level Segmentation and Affine Motion Estimation, IEEE Journal on Selected Area in Communications, Vol 15 (December 1997): 1704-1713.

Zhong, D., and Chang, S. F. AMOS : An Active System For MPEG-4 Video Object Segmentation, Proceedings of International Conference on Image Processing, Vol 2 (1998): 647-651.

ภาคผนวก

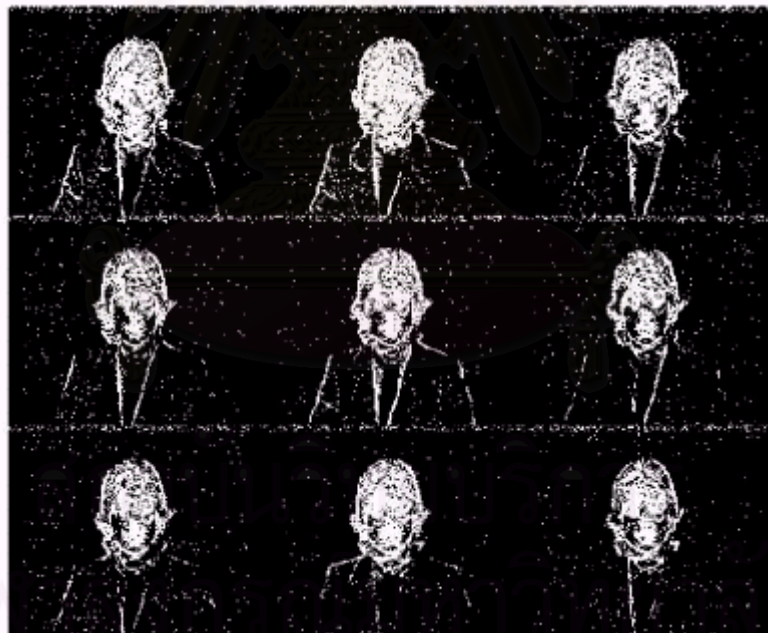


สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก.
ผลการแยกส่วนภาพวัตถุของลำดับภาพมาตรฐาน



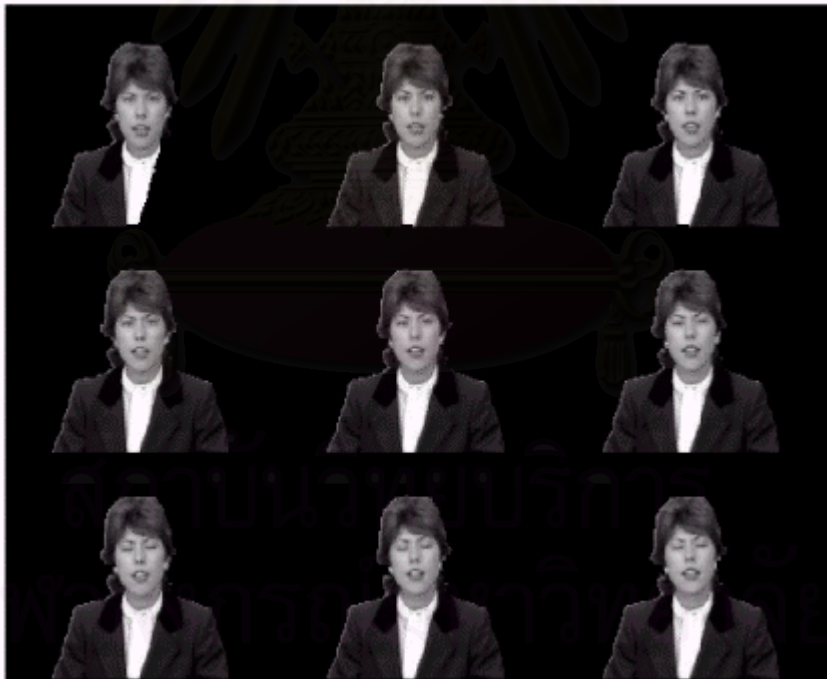
รูปที่ ก.1 ภาพต้นแบบของลำดับภาพ Claire เฟรมที่ 2 ถึง 10



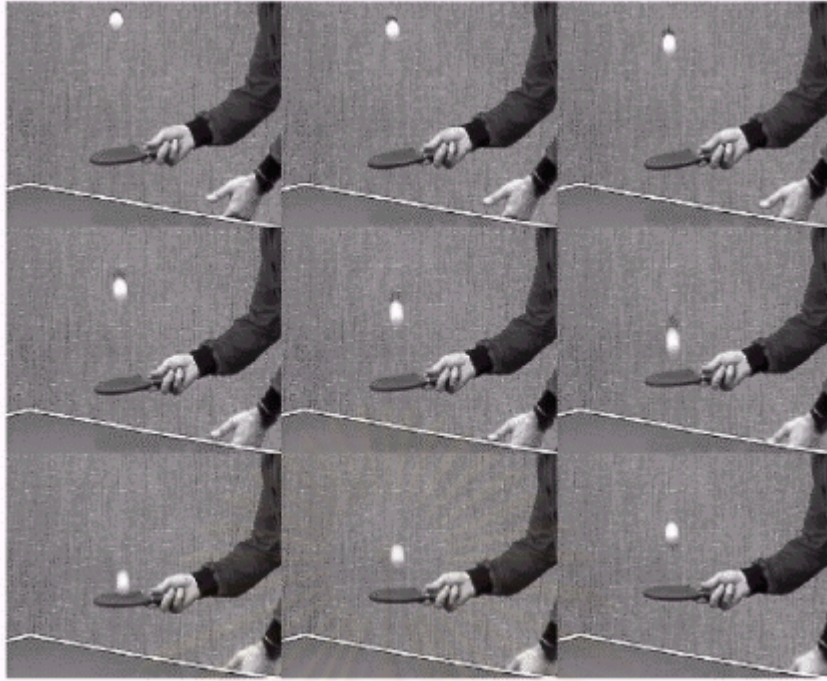
รูปที่ ก.2 การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Claire เฟรมที่ 2 ถึง 10



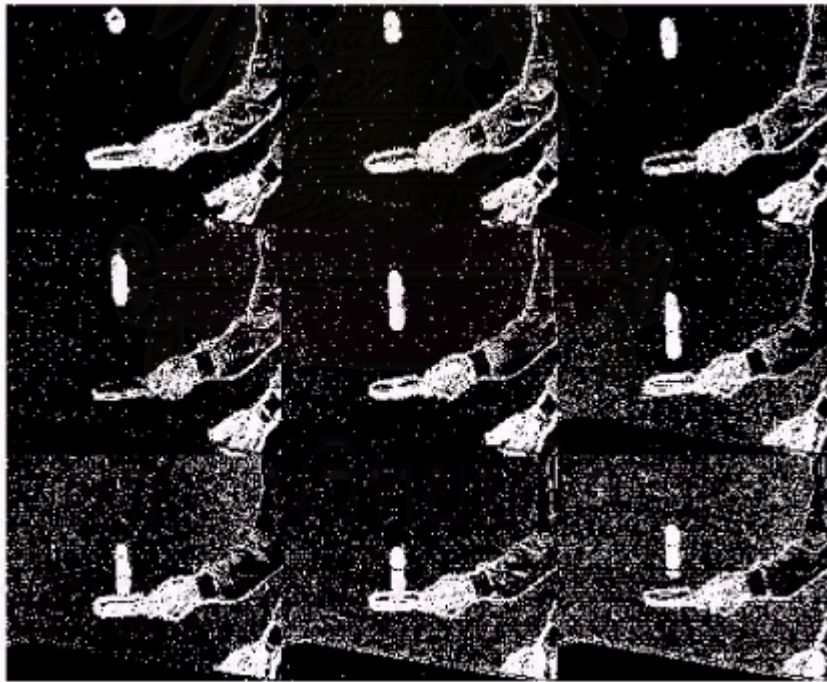
รูปที่ ก.3 บริเวณที่ได้จากการแยกส่วนเชิงพื้นที่ของลำดับภาพ Claire เฟรมที่ 2 ถึง 10



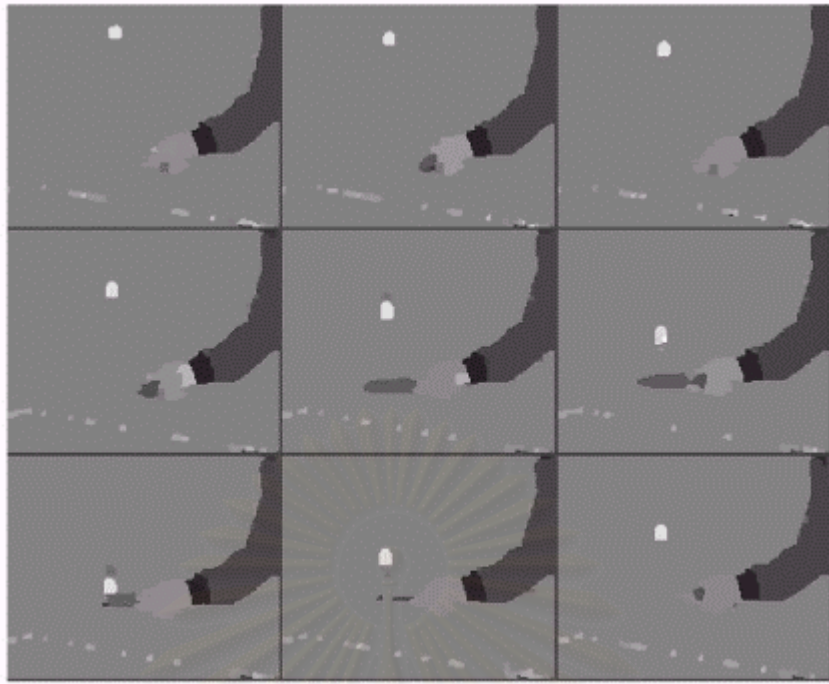
รูปที่ ก.4 ภาพวัตถุของลำดับภาพ Claire เฟรมที่ 2 ถึง 10



รูปที่ ก.5 ภาพต้นแบบของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10



รูปที่ ก.6 การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Table Tennis เฟรมที่ 2 ถึง 10



รูปที่ ก.7 บริเวณจากการแยกส่วนเชิงพื้นที่ของลำดับภาพ Table Tennis เฟรมที่ 17 ถึง 25



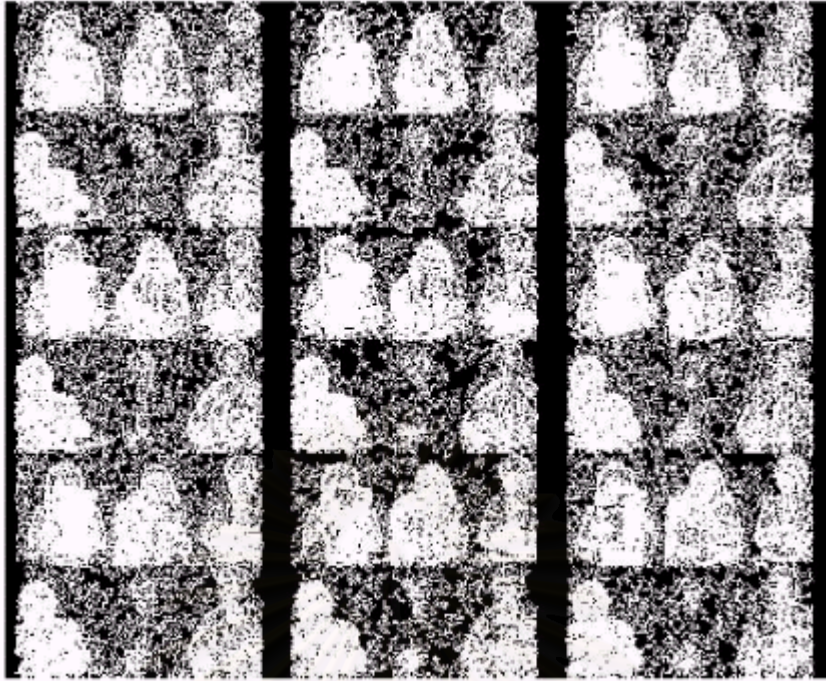
รูปที่ ก.8 ภาพวัตถุของลำดับภาพ Table Tennis เฟรมที่ 17 ถึง 25



รูปที่ ก.9 ภาพต้นแบบของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10



รูปที่ ก.10 บริเวณจากการแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10



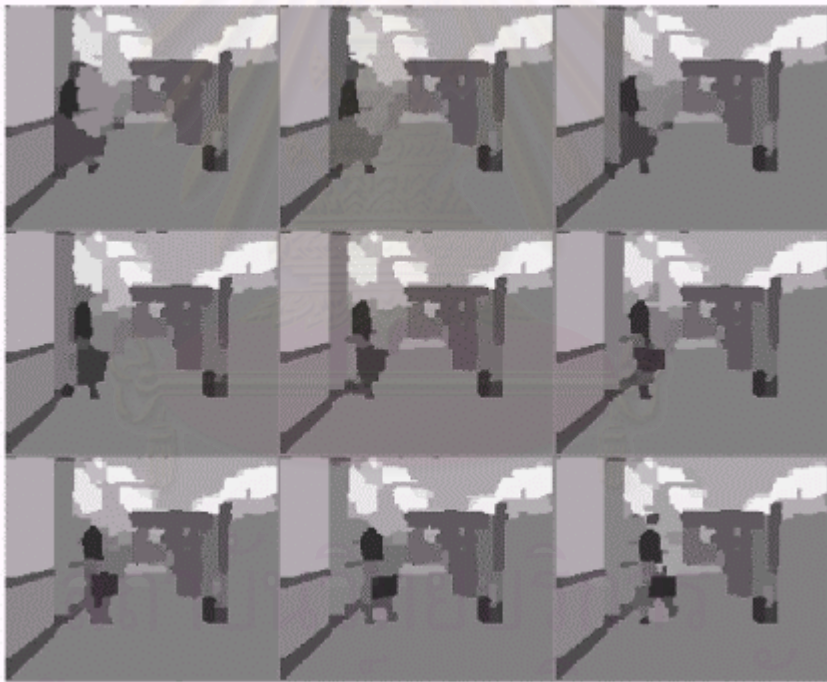
รูปที่ ก.11 การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10



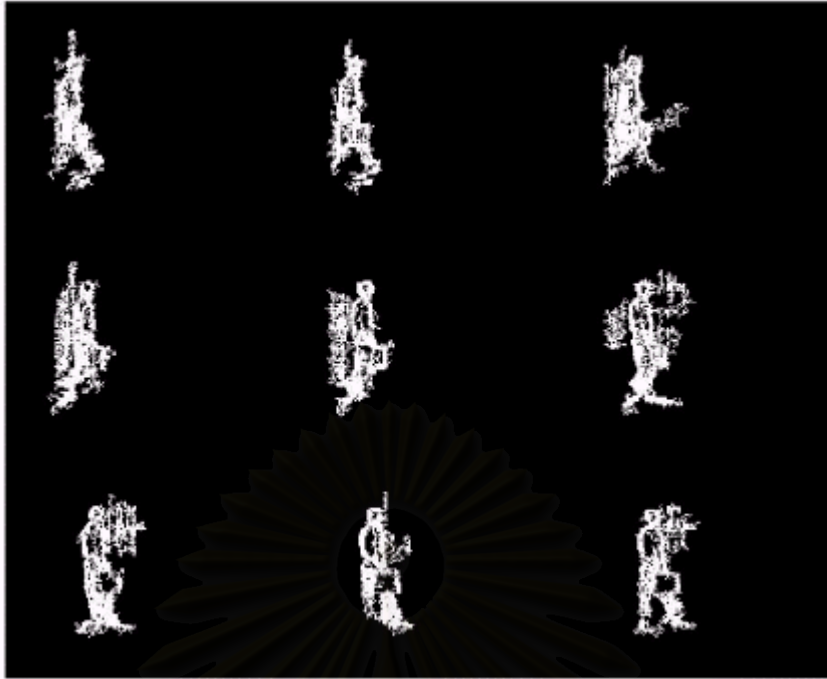
รูปที่ ก.12 ภาพวัตถุของลำดับภาพ Trevor เฟรมที่ 2 ถึง 10



รูปที่ ก.13 ลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42



รูปที่ ก.14 การแยกส่วนภาพเชิงพื้นที่ของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42



รูปที่ ก.15 การตรวจรู้การเปลี่ยนแปลงของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42



รูปที่ ก.16 ภาพวัตถุของลำดับภาพ Hall Monitor เฟรมที่ 26 28 30 ... 42

ภาคผนวก ข.
บทความที่ได้รับการตีพิมพ์

1. Video Segmentation Based on Adaptive Combination of Multiple Features According to MPEG-4, Proceedings of SPIE on Visual Communications and Image Processing (VCIP-2000), June 2000, Perth, Australia.
2. Spatial Segmentation Based on Modified Morphological Tools, Proceedings of IEEE Computer Society on Information Technology : Coding and Computing (ITCC-2001), April 2001, Las Vegas, Nevada, USA.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Video Segmentation Based on Adaptive Combination of Multiple Features according to MPEG-4

D. Tancharoen, S. Jitapunkul, P. Kittipanya-ngam, N. Siritaranukul

Digital Signal Processing Research Laboratory, Department of Electrical Engineering

Chulalongkorn University, Bangkok 10330, Thailand

E-mail: datchakorn@daisy.ee.eng.chula.ac.th, jsomchai@chula.ac.th

ABSTRACT

Video segmentation for object based video coding according to MPEG-4 should be able to segment interested objects in video sequence clearly. This paper presents the object segmentation algorithm which image features are combined to use in segmentation process following to characteristic of video signal. Because the combination of many features in video sequence is a method that can achieve high quality object segmentation. In addition, this algorithm is an adaptive method that many parameters can be adjusted in order to give clearly segmentation. The significant features are used in segmentation process including color, motion vector and change information. A fast shortest spanning tree (SST) algorithm is adapted to use for fast segmenting image boundary. Motion vectors are estimated and searched by thresholding hierarchical block matching (HBM) which want quite low computation and give a few groups of motion vectors. The change information is used to detect moving objects that can separate between moving object and static background. After that, each feature will be considered in segmentation decision process to segment interested objects. Then the post processing refines the final segmentation. The results from many test sequences have good quality and show object boundary clearly.

Keywords: object segmentation, image feature, low bit rate video coding, MPEG-4

1. INTRODUCTION

Nowadays visual information has been used widely in multimedia communication. The requirement is to receive not only multimedia information but also the interactive function. The visual information has to be described by a semantically meaningful representation. This representation derives from the scene content, which is decomposed in terms of the constituting objects.

In very low bit rate video coding using block based coding techniques, texture and motion information are used in order to minimize the required transmission bit rate at specific image quality. However they still are not able to achieve some requirements. The object based video coding according to MPEG-4^{1,2} has been shown that shape information has to be transmitted to present arbitrary shape of moving objects in order to improve the coding efficiency and increase content-based functionalities.

For MPEG-4 video the image sequence input is considered to be an arbitrary shape or a video object plane and the shape and location of an object within a reference window may vary over frame sequence. The coding of standard rectangular image sequences is supported as a special case of the video object plane approach. Very low bit rate video coding supports the coding of rectangular image sequences by motion and texture coding. The additional functionalities were extended by allowing the more flexible coding with using shape and transparency information for each video object plane.³

In order to obtain object based video coding, video sequence should be segmented in terms of meaningful semantic objects and has low computational algorithm. The method which uses the combination of many features including spatial and temporal information then merges them in decision process will give high quality object segmentation and also be suitable for multimedia applications.⁴

2. SEGMENTATION ALGORITHM

This paper proposes the video object segmentation algorithm based on adaptive combination of multiple features including color, motion, change and also previous information.

The block diagram of the segmentation algorithm is shown in figure 1.

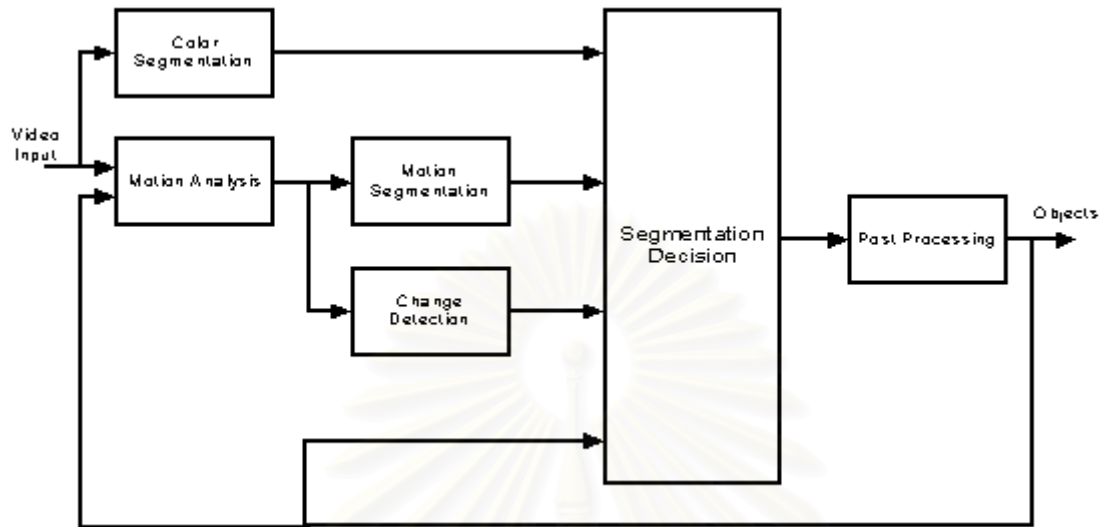


Figure 1.

The block diagram of the segmentation algorithm

2.1 Color analysis

Current frame is divided into a number of regions in color segmentation process. The color information including YUV component is used to segment the image into the regions with coherent intensity. The regions show the boundaries of the semantic objects in the scene and each region belongs to only one semantic object.

The SST (shortest spanning tree) segmentation is a method that takes the advantages of to simple control a number of region.⁵ The boundary of image segmentation can separate image into the regions better than edge detection and give clearly region boundary.

The mapping method from image to graph is the changing from gray level intensity to node weight in graph. The weighed graph has links of the graph that is a set of edge linking between every node. If weight link which has value less than threshold will be set as the link of boundary.

$$\begin{aligned} w_{ij} &= 1 \text{ if } |v_i - v_j| \leq T \\ w_{ij} &= 0 \text{ if } |v_i - v_j| > T \end{aligned} \quad (1)$$

where w_{ij} is the weight of an edge link between nodes i and j , T is the threshold value, v_i and v_j are intensity level of pixel i and j respectively.

This algorithm used a fast shortest spanning Tree⁶ to segment current frame because it has many advantages over the SST such as constructing the spanning tree in a region growing and reducing the computation complexity.

2.2 Motion analysis

In motion analysis, the motion between two consecutive frames will be detected. The Hierarchical Block Matching (HBM) algorithm is used due to acceptable results with quite low computational demand. The mean of absolute difference (MAD) will be used as a criterion to implement block matching. Motion vectors obtained from block matching are grouped in the same direction. By the way, the computational load can be reduced by varying threshold, block size and search range.⁷ This thresholding method gives a few groups of motion vector and can improve computational time.

2.3 Previous information analysis

The previous segmentation has been included in the decision process by inserting tracked object into the current segmentation process. This process creates previous mask on the current segmentation mask and also tended to track the individual object along image sequence by using the decision of motion information.

2.4 Change detection analysis

The changing mask can be detected from two successive frames. A group of pixels whose intensity has been changed more than criterion due to a moving object are labeled as changed block.⁸ This information is used to detect moving object and easy to implement by thresholding the changing intensity in each block. The average mean absolute difference will be used as a criterion to measure intensity changing for each block and image can be divided into changed region and unchanged region.

3. SIGNIFICANCE OF MULTIPLE FEATURES

Each feature has not the same impact to segmentation process, so different weight is required according to different feature. However, motion and intensity information have mostly influence to segmentation process. According to the features of image sequences, if image sequence is sensitive or moves quickly, significance of motion information should be used more than intensity information. On the other hand, if image sequence is static or quite moves slowly, most of weight should be used on intensity information. However if the image has quite low contrast intensity, both of them should be used to support each other. Other features are still considered however they have not much important in video sequence, so a part of weight constantly will be used for other features.

Motion information is quite accurate to locate the position and direction of moving object . On the other hand, spatial information such as color and intensity is reliable in uniform areas but tends to give many segmented regions. Therefore, the combination of them is essential.

3.1. Additional features analysis

The other features can be used in the segmentation scheme including color component, position, texture and motion vectors. Each feature has quite different range values. Color information has different components such as RGB, YUV and YCbCr that have different type of information. The positions are limited by the size of the image. Furthermore, the texture has the big variations and the motion vectors span over a limited interval. All information has to be normalized to use in decision process.

The normalization process is a primary step for a meaningful definition of distance in the vector space. Many features are different in both the range of variation and the level of reliability.⁹ So, the significance of each information should be given in the segmentation process in order to decide the object boundary correctly.

4. SEGMENTATION DECISION

In decision process, all features are combined and used some features to create object mask. For change information, it can eliminate some pixels that are in static background. Then, intensity region will be projected onto motion region.¹⁰ If motion region overlaps on intensity region, decision process will be set that intensity region as foreground. If some parts of these regions are overlapped then intensity region will be considered with percentage thresholding of overlapped region according to video sequence characteristic. The intensity segmentation gives an accurate region boundary; therefore, the most of color information is used to detect object boundary. If both of them do not overlap each other then intensity region will be set as static background.

In post processing, the segmentation output will be eliminated the small regions by morphological filter and the same motion regions will be merged. Finally, boundary smoothing is used to complete the regions and smooth the contour of object.¹¹ The convolution mask is used on the output region and then thresholds the convolution value to smooth the boundary. So, the final segmentation mask will be obtained to detect the interested objects.

5. RESULTS

The proposed method has been tested on MPEG standard test sequences such as Suzie and Claire sequence which are the video phone and video conferencing sequences. In this algorithm, the motion detection was used to segment image sequence which has a large moving object as Suzie sequence then morphological filter and convolution mask were used to smooth the object boundary. In change detection process, each frame was divided into 8x8 block size and the average changing intensity threshold was set to 1 then small region was eliminated by convolution mask. Motion detection and segmented results of Suzie are shown in figure 2(a)-2(d). However for Claire sequence, change information can separated only moving part of Claire head as shown in figure 3(a) and 3(b) because Claire sequence quite has still body. So the body boundary will be detected by using intensity information as shown in figure 3(c) then smoothed by 4x4 mask as shown in figure 3(d). The combination of motion and intensity information of Claire sequence was shown in figure 3(e). By the way, Table Tennis sequence which has fast motion and non-coherent background will be tested. Both of motion detection and intensity information were used to segment interested objects that included several moving object such as arm, hand, racket and ball. In addition, table is segmented by intensity information. Table tennis sequence was divided into 4x4 block size in change detection process because it had many small objects. The change and moving objects of Table Tennis between 1st and 2nd frame were shown in figure 4(c) and 4(d). The block size of 16x16 was used to eliminate small region of background after intensity processing. Many objects could be detected by intensity information such as arms, racket and table but hand and ball cannot be detected using this information because they were in non-uniform region. In the other hand, hand and ball were detected clearly by motion information. So, the combination of them could separate the interested object as shown in figure 4(h). Moreover, the direction of moving objects could be detected by motion vectors as shown in figure 4(i).

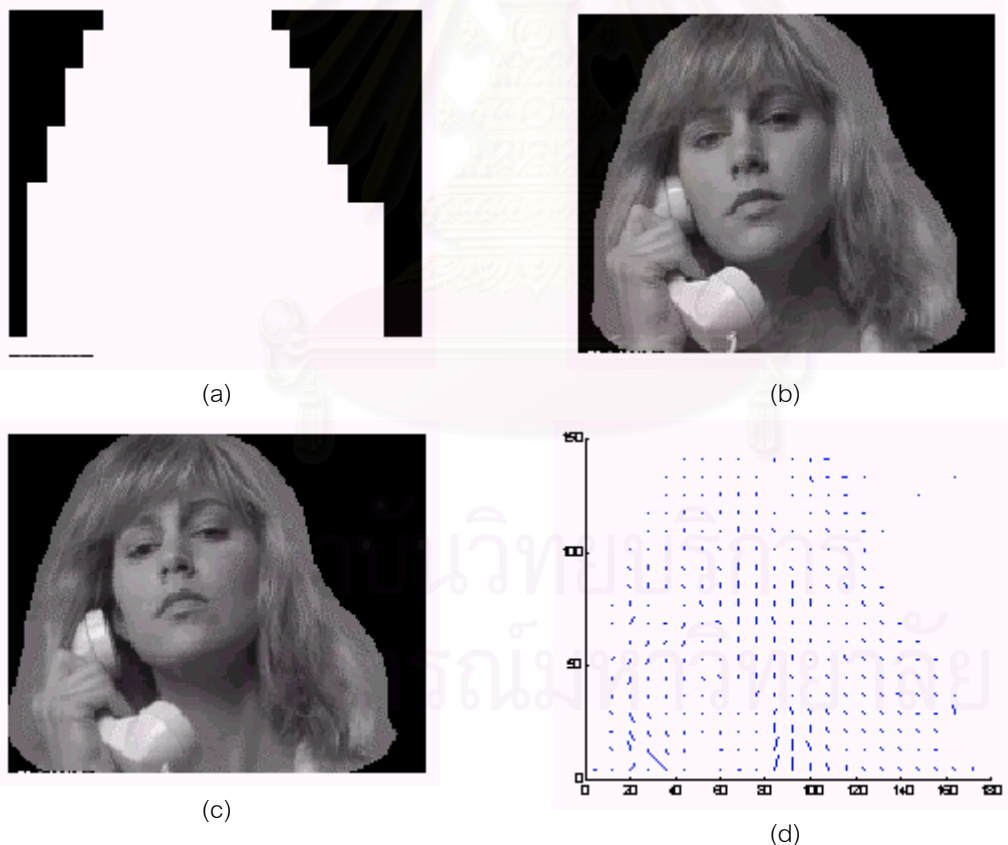


Figure 2. Segmentation results of Suzie sequence (a) change detection of 40th frame (b) 40th frame (c) 45th frame (d) motion vectors between 40th and 42nd frame

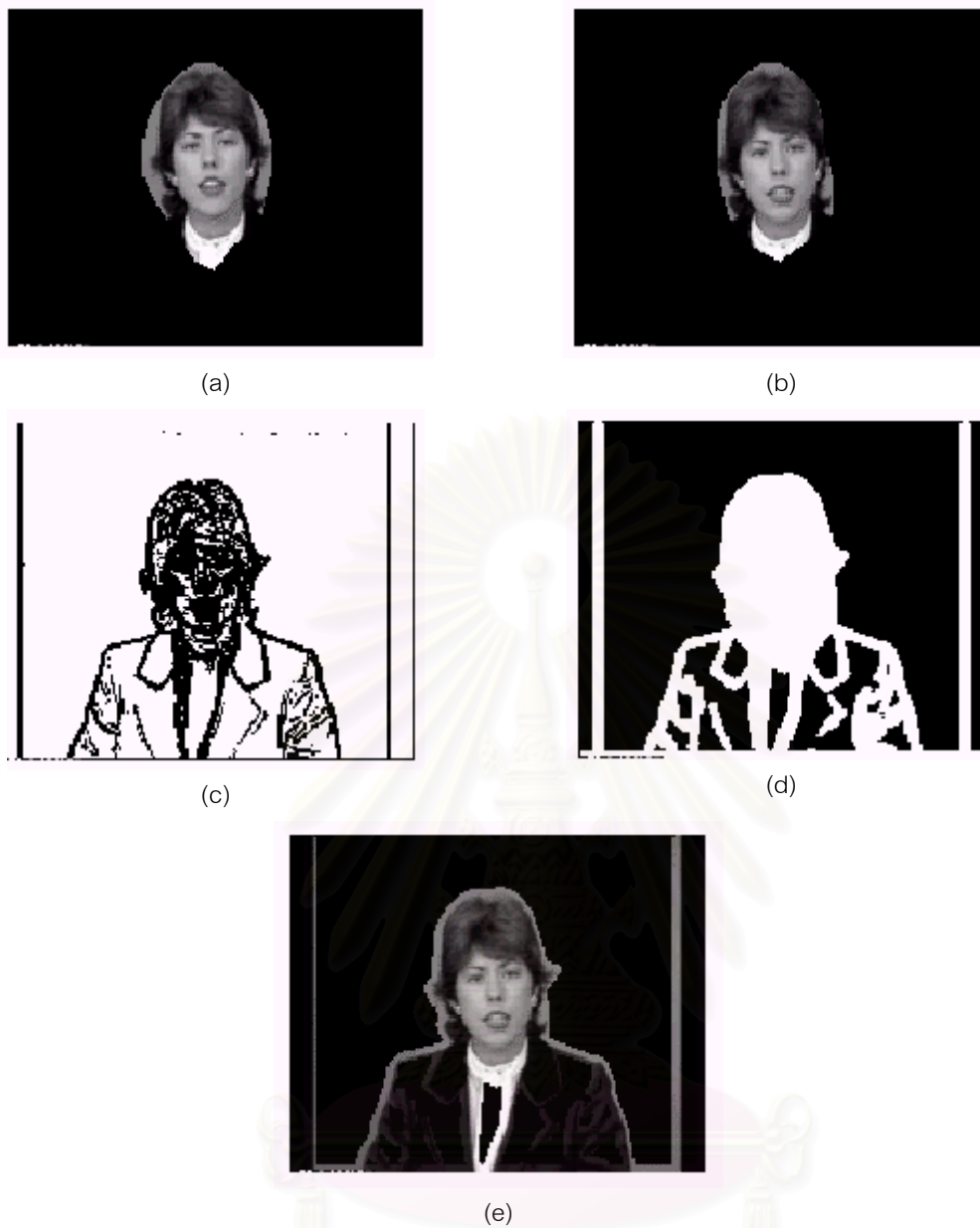
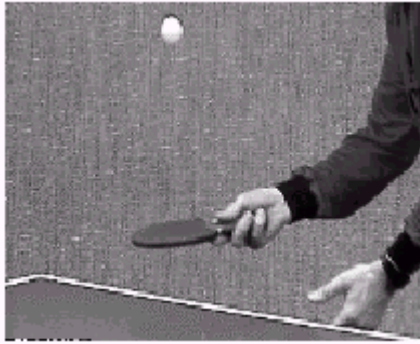


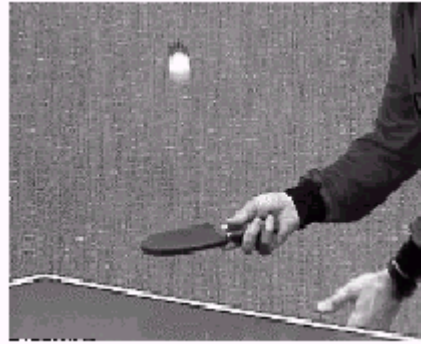
Figure 3. Segmentation results of Claire Sequence

(a) moving object of 15th frame (b) moving object of 20th frame (c) spanning tree of 20th frame
 (d) convolution mask of spanning tree (e) interested object

จุฬาลงกรณ์มหาวิทยาลัย



(a)



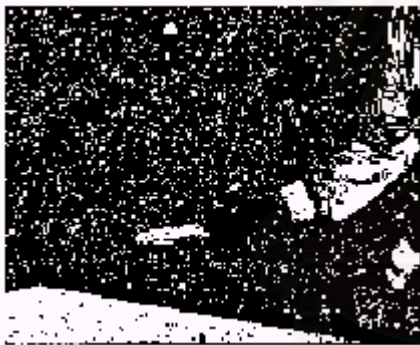
(b)



(c)



(d)



(e)



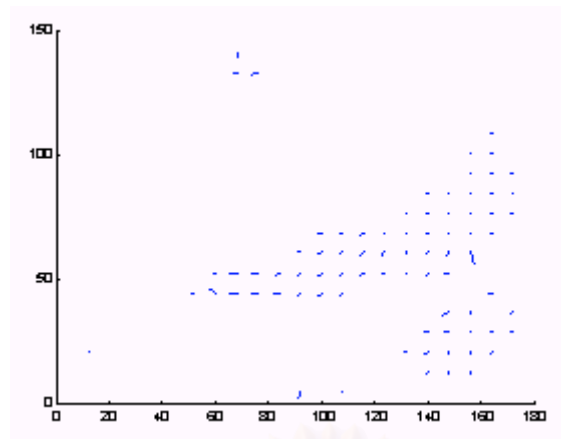
(f)



(g)



(h)



(i)

Figure 4. Segmentation results of Table Tennis sequence (a) 2nd frame (b) 4th frame
 (c) changing of 2nd frame (d) moving object of 2nd frame (e) spanning tree of 2nd frame
 (f) convolution mask of spanning tree (g) final segmentation mask (h) interested objects
 (i) motion vectors between 2nd and 4th frame

6. CONCLUSION

In this video object segmentation algorithm, the multiple features will be used in order to be able to segment interested object for each image sequence. In this proposed method, many parameters can be adjusted both in motion and intensity parameters due to image sequence characteristics to satisfy the interested object. For this reason, many image sequences will be segmented into objects between moving object and static background. Then the object segmentation algorithm also supports content-based interactivity. Moreover, this algorithm quite used low computation in all process to reduce computational complexity and the image can be segmented into the regions that show the semantic meaningful object. So this video object segmentation algorithm should be able to support in object based video coding to satisfy MPEG-4 low bit rate requirement.

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

ACKNOWLEDGEMENTS

The author wishes to thank Dr.Somchai, Dr.Satien and Dr.Wuttipong who gave many advices for doing research and also NSTDA for the TGIST scholarship to support this research. In addition, the author has gained much knowledge from discussions with many co-researchers and the facilities in Digital Signal Processing Research Laboratory, Chulalongkorn University

REFERENCES

- [1] T. Sikora, "The MPEG-4 Video Standard Verification Model", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.7, No.1, Feb. 1997.
- [2] R. Schafer, "MPEG-4: A Multimedia Compression Standard for Interactive Applications and Services", *Electronics and Communication Engineering*, Dec. 1998.
- [3] T. Sikora, "MPEG-4 Very Low Bit Rate Video", *IEEE International Symposium on Circuits and Systems*, June 1997.
- [4] A. Aydin Alatan, L. Onural and M. Wollborn, "Image Sequence Analysis for Emerging Interactive Multimedia Services - The European COST 211 Framework", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.8, pp 802-813, Nov. 1998.
- [5] K. S. Lau and G. Wade, "Spatial Spectral Clustering using Recursive Spanning Trees", *IEEE Proceedings Communications, Speech and Vision*, Vol.138, pp 232-238, Aug. 1991.
- [6] S. H. Kwok and A. G. Constantinides, "A Fast Recursive Shortest Spanning Tree for Image Segmentation and Edge Detection", *IEEE Transactions on Image Processing*, Feb.1997.
- [7] X. Xia and Y.Q. Shi, "A Thresholding Hierarchical Block Matching Algorithm for Motion Estimation", *IEEE International Sym. On Circuits and Systems*, Vol.2, 1996.
- [8] R. Mech, M. Wollborn, "A Noise Robust Method for Segmentation of Moving Objects in Video Sequences", *IEEE Inter. Con. on Speech and Signal Processing*, Vol.4, 1997.
- [9] R. Castagno, T. Ebrahimi and M. Kunt, "Video Segmentation Based on Multiple Features for Interactive Multimedia Applications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.8, pp 562-571, Sep.1998.
- [10] A. Aydin Alatan, E. Tuncel and L. Onural, "A Rule based Method for Object Segmentation in Video Sequences", *Proceedngs. on Image Processing*, Vol.2, 1997.
- [11] I. J. Lin and S. Y. Kung, "Circular Viterbi based Adaptive System for Automatic Video Object Segmentation", *IEEE Second Workshop on Multimedia Signal Processing*, pp 541-546, 1998.

Spatial Segmentation Based on Modified Morphological Tools

D. Tancharoen, S. Jitapunkul, and S. Chompun

*Digital Signal Processing Research Laboratory, Department of Electrical Engineering, Chulalongkorn University,
Bangkok 10330, Thailand*

E-mail: datchakorn@ee.eng.chula.ac.th, jsomchai@chula.ac.th

Abstract

The MPEG-4 visual standard supports the trend in video object segmentation for spatial segmentation process in only one way. Therefore, this paper presents the spatial segmentation process according to MPEG-4. The spatial segmentation includes image simplification, gradient operation and watershed algorithm. This paper presents the morphological tools to use in the segmentation process. Firstly, the morphological filters are used to simplify the image by open-close operation by partial reconstruction while preserve the boundary information. Secondly, the multiscale gradient operator is used to reduce the oversegmentation and also preserve the homogenous region boundary. This simplification and labeling of the region method give the homogenous region and the small number of regions. Finally, the uncertain pixels are reduced for using with the watershed algorithm.

1. Introduction

The MPEG-4 visual standard supports the trend in segmentation research for temporal segmentation in two ways and spatial segmentation in one way [1]. Therefore, the spatial segmentation process is an interesting research. It includes image simplification, gradient operation and watershed detection [2]. Its process splits an image into several homogenous regions in terms of luminance. The different homogeneous regions are distinguished by their boundaries. Then, a morphological segmentation process will be executed using the morphological tools and a watershed algorithm. However, due to the watershed algorithm, it produces oversegmentation problem, many methods are proposed to solve this problem [3].

Also, this paper proposed a technique to overcome the oversegmentation problem. The morphological filters are presented in order to reduce the oversegmentation without the loss of edge. First, a morphological opening or closing is used to simplify the original image by removing the bright or dark components that do not fit in the structuring element [2]. The elementary morphological filters can be used as simplification tools before segmentation process but they cannot preserve the perfect contour information [4]. The morphological filter by reconstruction is presented to improve the contour preservation [5]. The simplified image is much simpler than the original image but there is a problem in the boundary of moving objects. Moreover, when the simplification is applied to the gradient image then a lot of ramp edges are lost if the monoscale

gradient operator was used. However, the ramp edge information can preserve by the multiscale gradient algorithm [6].

The following sections are explained about the methods of the segmentation algorithm.

2. Morphological Segmentation

The morphological segmentation algorithm consists of the following steps: simplification, morphological gradient and watershed detection. The block diagram is shown in Figure 1. The details of each step are explained in the following sections.

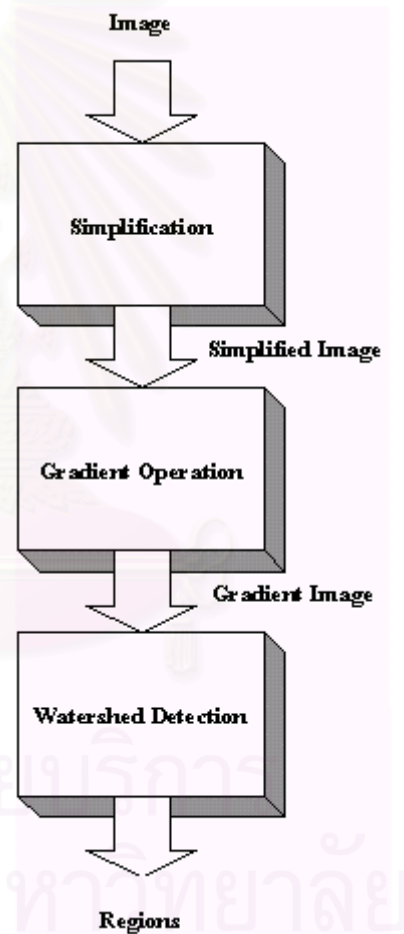


Figure 1. Block diagram of morphological segmentation

2.1 Simplification

The morphological filter is presented in the simplification process. The smoothing of the texture component in the inner region of the interested object is performed to simplify the image for easy segmentation.

Many morphological tools based on two basic operators including erosions and dilations.

The erosion and dilation with flat structuring element are defined as

$$\text{Erosion : } \mathcal{E}_n(f(x)) = \text{Min} \{f(x+y), y \in M_n\} \quad (1)$$

$$\text{Dilation : } \mathcal{D}_n(f(x)) = \text{Max} \{f(x-y), y \in M_n\} \quad (2)$$

While $f(\cdot)$ is an input signal and M_n is a flat structuring element of size n .

Elementary erosion and dilation allow the definition of morphological filters such as the morphological opening and closing.

$$\text{Morphological opening : } \gamma_n(f) = \mathcal{D}_n(\mathcal{E}_n(f)) \quad (3)$$

$$\text{Morphological closing : } \varphi_n(f) = \mathcal{E}_n(\mathcal{D}_n(f)) \quad (4)$$

A morphological opening and closing are used to simplify the original image by removing the bright or dark components that do not fit in the structuring element. If the simplification has to deal with both bright and dark components, an open_close or a close_open operation has to be used.

The reconstruction process that can improve the contour preservation in simplification process is generated from the geodesic transform.

The geodesic transforms are defined with a reference function (r) .

The geodesic dilation and erosion are defined as [4]

$$\text{Geodesic dilation of size one} \quad \mathcal{D}^{(1)}(f, r) = \text{Min} \{f, r\} \quad (5)$$

$$\text{Geodesic erosion of size one} \quad \mathcal{E}^{(1)}(f, r) = -\mathcal{D}^{(1)}(-f, -r) \quad (6)$$

The geodesic dilation (erosion) of infinite size is called reconstruction by dilation (erosion).

$$\text{Reconstruction by dilation : } \gamma^{(rec)}(f, r) = \mathcal{D}^{(rec)}(f, r) = \dots \mathcal{D}^{(1)}(\dots \mathcal{D}^{(1)}(f, r) \dots, r) \quad (7)$$

$$\text{Reconstruction by erosion : } \varphi^{(rec)}(f, r) = \mathcal{E}^{(rec)}(f, r) = \dots \mathcal{E}^{(1)}(\dots \mathcal{E}^{(1)}(f, r) \dots, r) \quad (8)$$

While superscript (rec) is the reconstruction process.

The opening (closing) operation by reconstruction of erosion (dilation) is given by

$$\text{Opening (closing) by reconstruction of erosion (dilation) :} \\ \text{Opening} \quad \gamma^{(rec)}(\mathcal{E}_n(f), f) \quad (9)$$

$$\text{Closing} \quad \varphi^{(rec)}(\mathcal{D}_n(f), f) \quad (10)$$

In the case of opening by reconstruction of erosion, the simplification is performed by the erosion. Then, the reconstruction process restores the contour of the components that have not been removed by the erosion.

In order to improve the contour of the object that is simplified by the morphological filter, the partial reconstruction process is used to be the smoothing filter.

The smoothing filter is considered to solve the problem in the prepreservation of object boundary by using a small opening to smooth the reference function in the simplification process.

The opening and closing by partial reconstruction [5] is used to simplify the original image and also preserve the contour information of the object.

$$\text{Opening, closing by partial reconstruction :} \\ \gamma^{(rec)}(\mathcal{E}_n(f), \gamma_k(f)), \varphi^{(rec)}(\mathcal{D}_n(f), \varphi_k(f)) ; 0 \leq k \leq n \quad (11)$$

The parameter k can be adapted to tune the smoothing filter of the reconstruction process. If k is set as n , there will be no reconstruction and if k is set as zero, there will be full reconstruction.

2.2 Morphological Gradients

Generally, the morphological gradient operators are described as

Morphological gradient

$$G(f) = \mathcal{D}(f, SE_n) - \mathcal{E}(f, SE_n) \quad (12)$$

Where SE represents a square structuring element of size $n \times n$.

When this gradient operator is used, the gradient of the ramp edge becomes high and the width of the edge becomes wide in case of large structuring element. But the ramp edges cannot be obtained easily in case of small structuring element. Moreover, a lot of ramp edges are lost when the simplification using morphological filter is applied to the gradient image which obtained using the monoscale gradient operator. In order to overcome this problem, a multiscale gradient operator is used to divide the region which is defined as

$$MG(f) = \frac{1}{n_i} \sum_{i=1}^n [s(\mathcal{D}(f, SE_i) - s(f, SE_i), SE_{i-1})] \quad (13)$$

where SE_i is a square structuring element of size $(2i+1) \times (2i+1)$.

This operation produces two-pixel width line that coincides with the edge. However, when the multiscale gradient operator is used, it is possible for the watershed lines to be disappeared in the region of dense edges in case of the low threshold value. Therefore, high threshold value is set to produce the watershed line in order to preserve the boundary of region.

3. Watershed Detection

The watershed algorithm constitutes one of the most powerful segmentation tools provided by mathematical morphology. The segmentation addresses the extraction of the different objects as the separation of the partially overlapping objects.

The notion of the watershed algorithm has been introduced by considering the gray level values of an image as the altitude of topography. The watershed lines partition the space by associating a region called a catchment basin to each local minimum [7].

The watershed detection is used to segment the regions from the gradient image. At the gradient stage, the edges of the gradient image that less than threshold value are set to zero in order to remove influence of dense edges. Then, the remained edge is used to divide the image into the region following the watershed algorithm.

4. Experimental Results

4.1 Simplification

In this section, the results of simplified image using the morphological filters were presented. The images at 1st frame of Claire and Akiyo sequence were shown in Figure 2 and 3. The structuring element in closing and opening operation has the size of 5x5 pixels. By the way, the closing operation by partial reconstruction is used by selecting $n = 5$ and $k = 3$. The closing and open_close operation can simplify the image but it has the loss of edge information in the face area of both images. However, the closing operation by partial reconstruction can preserve the region boundary and also simplify the image in complex area.

4.2 Morphological Gradient Operation

The morphological gradient image results were presented in Figure 4 and 5. These results are performed from simplified image. The threshold value of gradient image is set as 10 percent of maximum intensity. In Figure 4, the morphological gradient images are shown. The regions after thresholding from open_close operation has shown that it can avoid the problem of oversegmentation better than closing operation.

However, the gradient approximation gave the uncertain regions around the boundary more than the multiscale gradient operation as shown the result in Figure 5(a). In Figure 5(b), the multiscale gradient image used SE as maximum size 5x5 gave a better result when it is considered in term of the number of regions and uncertain pixels.



(a)



(b)



(c)

Figure 2. Claire sequence at the 1st frame

(a) original image (b) simplified image by closing operation (c) open_close operation (d) closing by partial reconstruction



Figure 2. Claire sequence at the 1st frame

(d) closing by partial reconstruction

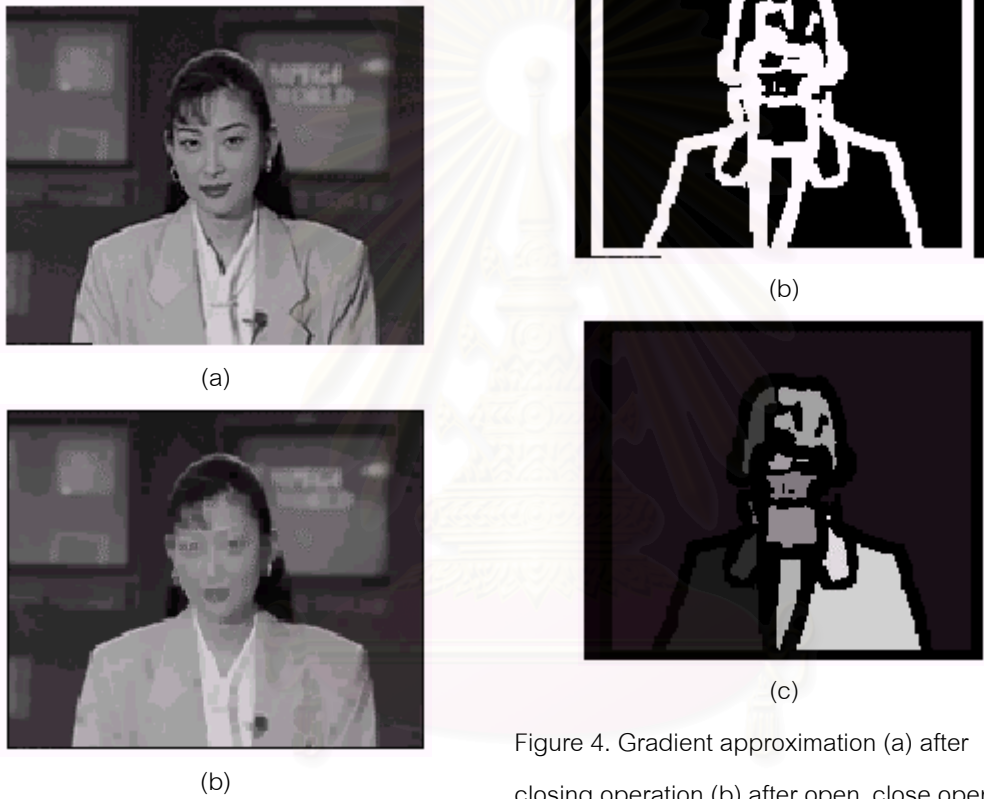


Figure 3. Akiyo sequence at 1st frame (a) original image

(b) simplified image by partial reconstruction

Figure 4. Gradient approximation (a) after closing operation (b) after open_close operation and thresholding (c) region labeling in 18 regions

จุฬาลงกรณ์มหาวิทยาลัย



(a)



(b)

Figure 5. Multiscale gradient operation (a) gradient image after thresholding (b) region labeling after thresholding where maximum size of structuring element is 5x5 and has 10 regions

5. Conclusion

This paper presents the morphological segmentation according to the spatial segmentation step of MPEG-4. The simplification process by morphological partial reconstruction is used to smooth the image in order to solve the oversegmentation problem and also to preserve the interested region boundary. In addition, the multiscale gradient operation is performed to segment the image into homogenous regions that have the small number of regions and uncertain pixels. Finally, this method can reduce the uncertain pixels to assign the regions in the watershed algorithm.

6. Acknowledgements

The author wishes to thank NSTDA for the TGIST scholarship and Dr.Prasit Prapinmongkolkarn for his support in this conference.

7. References

- [1] ISO/IEC JTC1/SC29/WG11, "Final Draft of International Standard Coding of Moving Pictures and Audio," N2502a, Atlantic City, Oct 1998.
- [2] M. Kim, J. G. Choi and Y. S. Ho, "A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Trans.Circuits and Systems for Video Tech.*, vol.9, no.8, Dec 1999.
- [3] S. H. Cho and H. M. Choi, "Modified morphological filter based reduction of oversegmentation in watershed transformation," *in Proceeding of ITC-CSCC*, 1999.
- [4] P. Salembier and J. Serra, "Morphological multiscale segmentation of images," *in Proc. SPIE Visual Commun. Image Processing*, 1992, pp.620-631.
- [5] P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Processing*, vol.3, no.5, Sep 1994.
- [6] D. Wang, "A multiscale gradient algorithm for image segmentation using watersheds," *Pattern Recognition*, vol.30, no.12, pp.2043-2052, 1997.
- [7] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Patt. Anal. Machine Intell.*, vol.13, no.6, June 1991.

ประวัติผู้เขียนวิทยานิพนธ์

นายด้ชกรณ์ ต้นเจริญ เกิดวันที่ 13 กุมภาพันธ์ พ.ศ. 2520 ที่จังหวัดกรุงเทพมหานคร เข้ารับการศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2537 สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ในสังกัดห้องปฏิบัติการวิจัยกรรมวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า ที่จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2541



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย