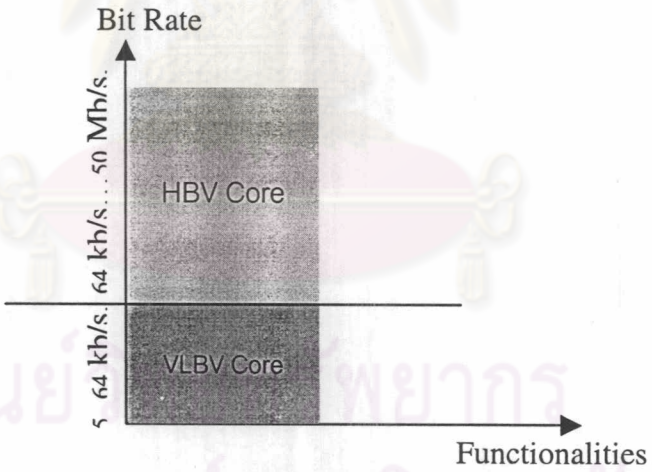
บทที่ 2
ความรู้พื้นฐาน

เนื้อหาในบทนี้กล่าวถึงทฤษฎีต่างๆ ที่เกี่ยวข้องกับวิทยานิพนธ์ คือ หลักการทำงานพื้นฐานของการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 การควบคุมอัตราบิตที่ใช้ (Test Model 5) การเข้ารหัสแบบสเกลาเบิล (Scalable or layered video coding) และ การเข้ารหัสแบบ FGS (Fine Granularity Scalability)

2.1 มาตรฐานการเข้ารหัสสัญญาณวิดีโอ MPEG-4 [7][8]

มาตรฐาน MPEG-4 เป็นมาตรฐานของ ISO (International Standard Organization) สำหรับการเข้ารหัสสัญญาณเสียง ภาพและวิดีโอ ที่อัตราบิตต่ำ โดยมุ่งเน้นการประยุกต์ใช้งานทางด้านการเก็บข้อมูล การถ่ายทอดกระจายข่าวสาร และการสื่อสารปฏิสัมพันธ์ โดยผ่านทางโครงข่าย IP, ATM, โทรศัพท์มือถือ, PSTN หรือ Narrow band ISDN



รูปที่ 2.1 โครงสร้างของมาตรฐาน MPEG-4 แบ่งประเภทตามอัตราบิต และฟังก์ชันการใช้งาน [7]

การเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 แบ่งแยกออกเป็นประเภทตามอัตราบิตและฟังก์ชันการใช้งานเป็น 2 ประเภท ดังรูปที่ 2.1 คือ ประเภทแรกเข้ารหัสวิดีโออัตราบิตต่ำมาก (VLBV: Very Low Bit Rate Video) โดยกำหนดค่าอัตราบิตไว้ที่ระหว่าง 5-64 Kbps. เพื่อรองรับฟังก์ชันการใช้งานที่ความละเอียดของภาพต่ำ และอัตราเฟรมต่ำ (0-15 Hz) เหมาะสำหรับอุปกรณ์สื่อสาร และประเภทที่สองเข้ารหัสวิดีโออัตราบิตสูง (HBV: Higher Bit Rate Video) สูงถึง 50

Mbps เพื่อรองรับฟังก์ชันการใช้งานที่ความละเอียดของภาพสูง และอัตราเฟรมสูง เหมาะสำหรับอุปกรณ์ TV โดยคุณภาพอยู่ในระดับโทรทัศน์ดิจิตอล

2.2 แบบจำลองเพื่อทดสอบสัญญาณวิดีโอตามมาตรฐาน MPEG-4 (MPEG-4 Video Verification Model) [3]

ในเดือนมกราคม ปี ค.ศ. 1996 MPEG Video Group ได้กำหนดแบบจำลองของสัญญาณวิดีโอตามมาตรฐาน MPEG-4 ในรุ่นแรกขึ้น ได้มีข้อกำหนดต่างๆ เพื่อให้ได้ประสิทธิภาพของการเข้ารหัสและกำหนดสำหรับการทำงานในแบบ content-based รวมไปถึงความทนทานต่อความผิดพลาด แบบจำลองได้รองรับคุณลักษณะต่างๆ ดังนี้

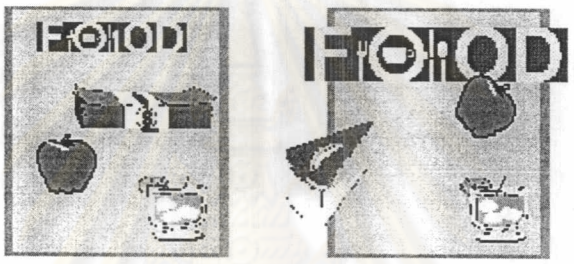
- การเข้ารหัสหลายระนาบวัตถุ (Video Object Plane : VOP) และเป็นภาพแบบไม่เจาะจงรูปร่าง (arbitrary shape) เพื่อรองรับการทำงานแบบ content-based ลำดับภาพที่ใช้ในแบบจำลองเป็นลำดับภาพแบบที่ไม่เจาะจงรูปร่าง และตำแหน่งของ VOP เทียบกับหน้าต่างอ้างอิงก็เปลี่ยนไปตามเวลา
- การเข้ารหัสรูปร่างและข้อมูลของความโปร่งแสง (transparency information) ของแต่ละ VOP โดยการเข้ารหัสแบบไบนารี
- รองรับการเข้ารหัสแบบ I, P, B ซึ่งรายละเอียดจะกล่าวในหัวข้ออื่นถัดไป
- รองรับอัตราเฟรมของลำดับภาพทั้งในแบบคงที่ และแบบปรับความเร็วได้ โดยอัตราเฟรมนั้นจะขึ้นอยู่กับการนำไปประยุกต์ใช้งาน
- รองรับการแบ่งขนาดของบล็อกเป็นบล็อกขนาด 8*8 จุดภาพ (Pixel Element : pel) หรือแบ่งเป็นมาโครบล็อกขนาด 16*16 จุดภาพ
- การเข้ารหัสพื้นผิวของภาพในแบบ I, P และ B-VOP โดยใช้ DCT (Discrete cosine transform)
- การปรับระดับการแสดงผลในเชิงตำแหน่ง และเชิงเวลา (Temporal and Spatial Scalability) สำหรับ VOP แบบไม่กำหนดรูปร่าง
- มีความสามารถในการใช้ได้กับมาตรฐานการเข้ารหัสแบบเก่า ทั้ง MPEG-1/2 และ H.261/263 ถ้าลำดับภาพที่ใช้เป็นแบบลำดับชั้น (layer) เดียว

2.3 ข้อกำหนดสำหรับหน้าที่ของ content-based [9]

กรรมวิธีการเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 สามารถรองรับฟังก์ชันต่างๆ ที่มีอยู่แล้วในมาตรฐาน MPEG-1 และ MPEG-2 รวมทั้งเพิ่มความสามารถในการบีบอัดข้อมูลลำดับ

ภาพที่มีรูปแบบสัญญาณขาเข้า (input formats) ได้หลายรูปแบบ หลายอัตราเฟรม (frame rates) หรือหลาย อัตราบิต (bit rates) นอกจากนั้นยังมีความสามารถด้าน Content Based Scalability สำหรับทั้งภาพนิ่งและสัญญาณวิดีโอ เพื่อให้ผู้ใช้สามารถดำเนินการกับวัตถุได้

แนวคิดของหน้าที่แบบ content-based แสดงดังรูปที่ 2.2 เป็นตัวอย่างของรูปภาพที่มีหลายวัตถุ แล้วทำการถอดรหัสแบบแยกแต่ละวัตถุออกจากกัน และยังสามารถรวมวัตถุเหล่านั้นเข้าด้วยกันใหม่ เป็นการดำเนินการกับภาพดั้งเดิมด้วยการปรับเปลี่ยนแบบง่ายๆ กับบิตสตรีม ทางด้านรับสามารถสร้างภาพเดิมกลับคืนมาได้ด้วยการถอดรหัสแต่ละวัตถุ แล้วแสดงภาพด้วยขนาดวัตถุที่เท่าเดิม และตำแหน่งเดิม ดังในรูปที่ 2.2(ก) หรือจากการดำเนินการของผู้ใช้ทำให้สามารถเปลี่ยนการแสดงภาพได้ ดังตัวอย่างในรูปที่ 2.2(ข) บางวัตถุไม่ได้แสดงผล ส่วนวัตถุอื่นได้รับการปรับแต่งขนาด และหมุนวัตถุ ในตัวอย่างยังมีการเพิ่มวัตถุใหม่ขึ้นมาจากภาพเดิมอีกด้วย



(ก) (ข)

รูปที่ 2.2 ตัวอย่างของความสามารถในการจัดการกับวัตถุในฉาก [9]
(ก) ภาพต้นแบบ (ข) ภาพที่ผ่านการจัดการ

สัญญาณวิดีโอที่เข้ารหัสนี้ไม่จำเป็นต้องเป็นพื้นที่สี่เหลี่ยม (rectangular region) โดยวัตถุทางกายภาพเดียวกันในเฟรม เรียกว่า วัตถุวิดีโอ (Video Objects : VO) ข้อมูลของรูปร่าง การเคลื่อนที่ และลักษณะพื้นผิวของ VOP ของวัตถุวิดีโอเดียวกันจะถูกเข้ารหัสและส่งไปในชั้นของวัตถุวิดีโอ (Video Object Layer : VOL) เดียวกัน นอกจากนี้ ข้อมูลที่จำเป็นในการระบุแต่ละ VOL จะถูกส่งไปที่ด้านรับด้วย ทำให้สามารถถอดรหัสแต่ละ VOP แยกจากกันได้

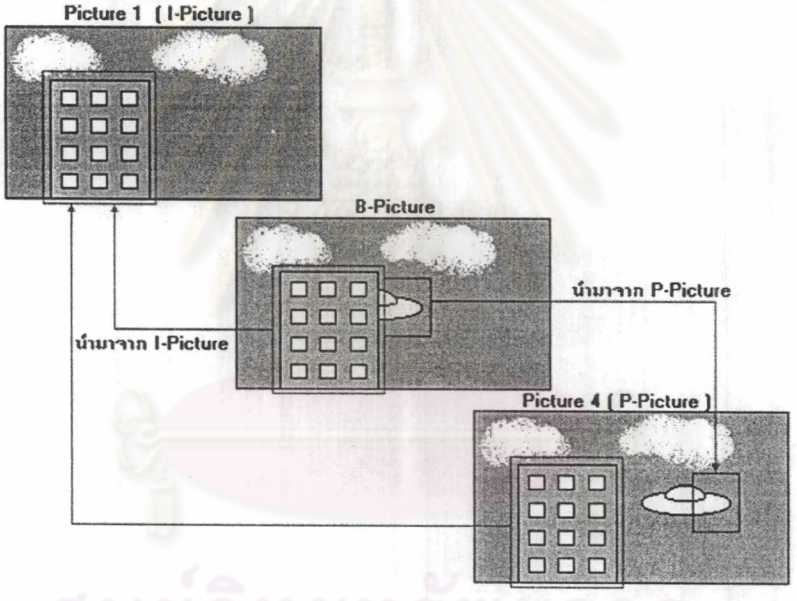
2.4 ชนิดของการเข้ารหัสสัญญาณวิดีโอแบบบีบอัดข้อมูล

การเข้ารหัสสัญญาณวิดีโอแบบบีบอัดข้อมูลจะแบ่งได้สองประเภทตามชนิดของข้อมูล คือ

1. การเข้ารหัสภายในเฟรม (Intra-Frame Coding) เฟรมที่ถูกเข้ารหัสเรียกว่า I เฟรม เป็นการเข้ารหัสโดยไม่ต้องใช้ข้อมูลของภาพจากเฟรมอื่นเลย ดังนั้น ในลำดับภาพวิดีโอเฟรมแรกจะเป็น I เฟรมเสมอ ซึ่งมีอัตราการบีบอัดข้อมูลต่ำที่สุด

2. การเข้ารหัสระหว่างเฟรม (Inter-Frame Coding) ซึ่งจะใช้กระบวนการประมาณและการชดเชยการเคลื่อนที่ (motion estimation/compensation) เพื่อช่วยในการลดข้อมูลที่มีความซ้ำซ้อนทางเวลา มี 2 ประเภทดังนี้

   - เฟรมแบบ P (Predictive Coded Picture) เฟรมประเภทนี้สามารถถอดรหัสได้โดยใช้ข้อมูลจากเฟรมก่อนหน้า ซึ่งเฟรมที่แสดงไปก่อนหน้านี้ เรียกว่า เฟรมอ้างอิง โดยสามารถเป็นเฟรมแบบ I หรือเป็นเฟรมแบบ P ด้วยกันก็ได้
   - เฟรมแบบ B (Bidirectionally Predicted Picture) เป็นเฟรมที่ใช้ข้อมูลจากเฟรมอื่นๆ มาช่วยเหมือนกับเฟรม P แต่แตกต่างกันตรงที่ เฟรมแบบ B สามารถใช้ข้อมูลของ เฟรมอนาคตมาใช้ได้ เฟรมประเภทนี้จึงให้ผลของการบีบอัดข้อมูลที่สูงที่สุด

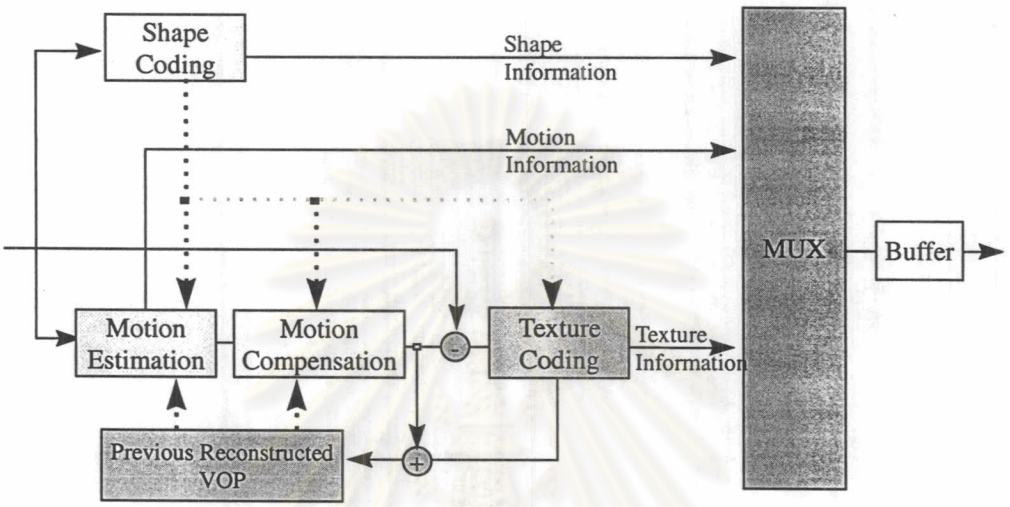


รูปที่ 2.3 ตัวอย่างการเข้ารหัสเฟรม I B และเฟรม P

รูปที่ 2.3 แสดงตัวอย่างการเข้ารหัสภาพที่มีการชดเชยการเคลื่อนที่ ภาพของจานบินที่ถูกบังด้วยตึกในเฟรมที่ 1 เริ่มปรากฏขึ้นในเฟรมที่ 2 ถ้าภาพเฟรมที่ 2 ถูกเข้ารหัสด้วยเฟรม B สามารถนำภาพตึกมาจากเฟรมที่ 1 ได้ และภาพบริเวณส่วนหัวของจานบินก็สามารถนำมาจากเฟรมที่ 4 ได้เช่นกัน เมื่อพิจารณาถึงเฟรมที่ 4 ถ้าสมมุติให้เข้ารหัสแบบเฟรม P เฟรมที่ 4 นี้ก็สามารถใช้ภาพตึกของเฟรมที่ 1 ได้ ซึ่งมีเพียงภาพของจานบินเท่านั้นที่ต้องเข้ารหัสจริงๆ

2.5 โครงสร้างของการเข้ารหัสสัญญาณวิดีโอของ MPEG-4

การเข้ารหัสสัญญาณวิดีโอตามมาตรฐาน MPEG-4 ประกอบไปด้วย การเข้ารหัสการเคลื่อนที่ รูปทรง และพื้นผิว (Motion, Shape and Texture Coding) ดังแสดงในรูปที่ 2.4



รูปที่ 2.4 ภาพแสดงโครงสร้างพื้นฐานตัวเข้ารหัสสัญญาณวิดีโอของ MPEG–4

ข้อมูลรูปทรง ข้อมูลการเคลื่อนที่ และข้อมูลพื้นผิวสำหรับวัตถุเดียวกันจะถูกเข้ารหัสในชั้นเดียวกัน แต่ถ้ามีหลายวัตถุจะถูกเข้ารหัสข้อมูลแต่ละวัตถุแยกชั้นกัน เพื่อให้สามารถถอดรหัสวัตถุแยกจากกันได้ โดยข้อมูลในแต่ละชั้น MPEG-4 ใช้กรรมวิธีการเข้ารหัสที่เหมือนกัน

2.5.1 การประมาณการเคลื่อนที่ (Motion Estimation)

การประมาณการเคลื่อนที่เป็นการหาความแตกต่างระหว่างจุดภาพในตำแหน่งที่สอดคล้องกันระหว่างเฟรมสองเฟรมที่ต่อเนื่องกัน ส่งผลให้สามารถทำการเข้ารหัสเพียง "ค่าของความแตกต่าง" ระหว่างสองเฟรมนั้นแทนที่จะต้องเข้ารหัสด้วยภาพทั้งภาพ ซึ่งถ้าค่าความแตกต่างดังกล่าวมีค่าน้อยเท่าใดก็ทำให้การบีบอัดสัญญาณภาพยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น โดยทั่วไป เราสามารถสมมุติว่าค่าของจุดภาพในภาพที่มีการเคลื่อนที่ของวัตถุได้เลื่อนตำแหน่งจากจุดหนึ่งในเฟรมหนึ่ง ไปยังจุดอื่นในเฟรมถัดไป ถ้าเราสามารถหาตำแหน่งที่เลื่อนไปได้ เราก็สามารถลดค่าความแตกต่างลงไปได้ วิธีนี้เรียกว่า การชดเชยการเคลื่อนที่ (motion compensation) ซึ่งในสัญญาณวิดีโอส่วนใหญ่ เฟรมที่อยู่ติด ๆ กันมักจะมีความคล้ายคลึงกันเป็นอย่างมาก หรือกล่าวได้ว่า ลำดับภาพมีความซ้ำซ้อนกันทางเวลา ยกตัวอย่างเช่น ภาพของผู้

ประกาศข่าวมีฉากหลังที่ซ้ำซ้อนนั่นเอง การชดเชยการเคลื่อนที่จึงเกิดขึ้นเพื่อลดความซ้ำซ้อนในส่วนนี้ โดยการจัดกลุ่มให้กับจุดภาพ (กลุ่มของจุดภาพนี้ เรียกว่าบล็อก) ค่าของตำแหน่งของบล็อกที่เลื่อนไปนี้จะแสดงในรูปของตัวเลขสองตัว ซึ่งแสดงถึงการเคลื่อนที่ในแนวแกนนอนและแนวแกนตั้งตามลำดับ เรียกว่า เวกเตอร์การเคลื่อนที่ (motion vector)

ส่วนใหญ่แล้วขนาดของบล็อกที่ใช้ในการชดเชยการเคลื่อนที่ มีขนาด 8 x 8 หรือ 16 x 16 จุดภาพ ดังในรูปที่ 2.5 แสดงถึงบล็อกข้อมูลขนาด 8 x 8 ในตำแหน่งที่สอดคล้องกันระหว่างเฟรมสองเฟรมที่อยู่ลำดับติดกัน โดยบล็อกข้อมูลในรูปที่ 2.5(ก) แสดงถึงภาพในเฟรมอ้างอิงและบล็อกข้อมูลในรูปที่ 2.5(ข) แสดงถึงภาพในเฟรมถัดไป จากตัวอย่าง ถ้าคำนวณหาความแตกต่างของจุดภาพระหว่างบล็อกในบริเวณตำแหน่งที่สอดคล้องกันตรงๆ จะได้ค่าความแตกต่างไม่เป็นศูนย์ทั้งหมด แต่ถ้าลองเปรียบเทียบความแตกต่างโดยใช้บล็อกข้อมูลที่อยู่ในกรอบเส้นประของเฟรมอ้างอิงพบว่าได้ค่าความแตกต่างเป็นศูนย์ทุกจุดภาพภายในเฟรม ค่าของตำแหน่งที่เลื่อนไปนี้เรียกว่า เวกเตอร์การเคลื่อนที่นั่นเอง นั่นคือ เวกเตอร์การเคลื่อนที่ในแนวแกนนอนและแกนตั้ง ก็คือ สองจุดภาพทางด้านขวาและหนึ่งจุดภาพทางด้านล่าง ตามลำดับ

2.5.2 การชดเชยการเคลื่อนที่ (Motion Compensation)

ด้วยสมมติฐานว่า แต่ละบล็อกของข้อมูลภาพประกอบจากจุดภาพที่มีความเกี่ยวข้องกันมาก ทำให้มีความน่าจะเป็นที่การเคลื่อนที่ภายในเฟรม จะทำให้จุดภาพส่วนใหญ่ภายในบล็อกเคลื่อนที่ไปด้วยระยะทางและทิศทางที่สอดคล้องกัน หรือกล่าวได้ว่า เฟรมหนึ่ง ๆ มีความน่าจะเป็นที่เหมือนกับเฟรมก่อนหน้า ซึ่งเป็นจริงสำหรับลำดับภาพที่มีการเคลื่อนที่ไม่มาก

การเปลี่ยนแปลงเพียงบางส่วนของภาพทำให้สามารถทำนายภาพจากเฟรมอดีตได้โดยตรง เป็นการลดความซ้ำซ้อนทางเวลาของภาพในแบบ P และ B ดังที่ได้กล่าวมาข้างต้น ซึ่งในทางทฤษฎีแล้ว การเก็บข้อมูลเพียงแค่ค่าความแตกต่างระหว่างเฟรมปัจจุบันกับเฟรมที่ทำนายได้จะใช้เนื้อที่น้อยกว่าการเก็บข้อมูลทั้งภาพ

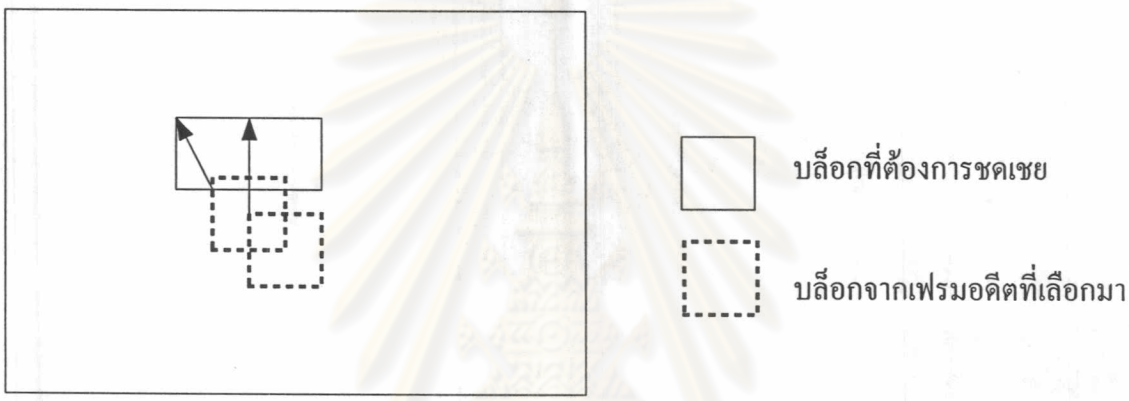
40	42	43	47	51	47	54	53	52	49
39	41	45	45	50	45	58	56	50	48
40	42	43	50	43	45	50	56	56	53
41	43	43	53	48	50	61	53	63	56
42	54	48	50	58	53	53	48	68	50
43	44	45	48	56	56	50	45	68	61
47	46	48	53	50	61	61	45	61	50
48	49	50	53	56	56	58	53	63	61
50	51	50	48	56	56	63	53	61	61

(ก) ค่าของจุดภาพจากเฟรมอ้างอิง

45	45	50	45	58	56	50	48
43	50	43	45	50	56	56	53
43	53	48	50	61	53	63	56
48	50	58	53	53	48	68	50
45	48	56	56	50	45	68	61
48	53	50	61	61	45	61	50
50	53	56	56	58	53	63	61
50	48	56	56	63	53	61	61

(ข) ค่าของจุดภาพในตำแหน่งเดิมของเฟรมถัดไป

รูปที่ 2.5 ตัวอย่างการหาเวกเตอร์การเคลื่อนที่ในการประมาณการเคลื่อนที่



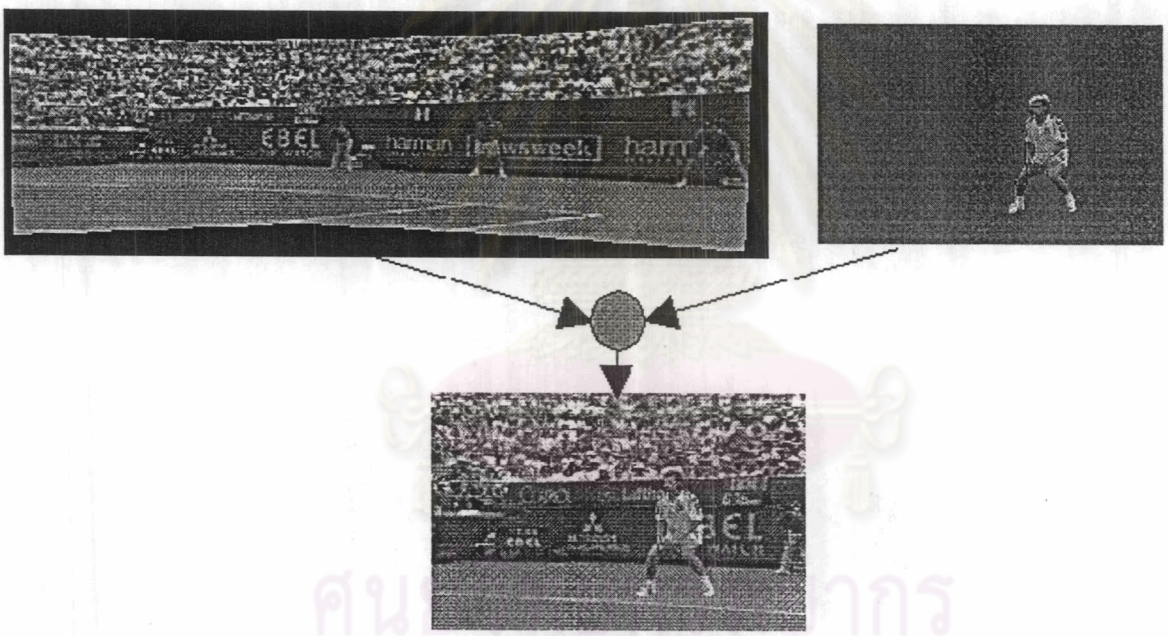
รูปที่ 2.6 ลักษณะการชดเชยการเคลื่อนที่

รูปที่ 2.6 แสดงตัวอย่างของการชดเชยการเคลื่อนที่ โดยบล็อกสี่เหลี่ยมที่เป็นจุดประแทนบล็อกของภาพจากเฟรมอดีตซึ่งเป็นบล็อกที่เลือกมาเพื่อสร้างเฟรมประมาณขึ้น ส่วนบล็อกที่เป็นเส้นทึบแทนบล็อกปัจจุบัน จะเห็นได้ว่า บล็อกที่นำมาจากเฟรมอดีตมีความเป็นไปได้ที่มีพื้นที่ซ้อนทับกัน แต่บล็อกในเฟรมที่ทำการชดเชยการเคลื่อนที่ไม่ซ้อนทับกัน คือ อยู่คนละตำแหน่งกัน

การชดเชยการเคลื่อนที่แบ่งได้เป็นสองแบบ ได้แก่ การชดเชยการเคลื่อนที่แบบไปข้างหน้า (forward motion compensation) และการชดเชยการเคลื่อนที่แบบไปข้างหลัง (backward motion compensation) การชดเชยการเคลื่อนที่แบบไปข้างหน้าอาศัยการเข้ารหัสโดยอ้างอิงจากเฟรมก่อนหน้านี้ในทางลำดับเวลา ส่วนแบบไปข้างหลังนั้นตรงกันข้าม คือ ใช้เฟรมที่อยู่ถัดไปเป็นเฟรมอ้างอิงในการทำนาย เทคนิคทั้งสองแบบนี้สามารถใช้ร่วมกันหรือแยกกันได้ ดังในกรณีของภาพแบบ B ต้องใช้การชดเชยการเคลื่อนที่ทั้งสองแบบ ทั้งตัวเข้ารหัสและตัวถอดรหัสในระบบวิดีโอแบบบีบอัดมีความสามารถในการทำการชดเชยการเคลื่อนที่ สรุปได้ว่า

- การประมาณการเคลื่อนที่เป็นกระบวนการของการประมาณเวกเตอร์การเคลื่อนที่ในขณะที่ทำการเข้ารหัส
- การชดเชยการเคลื่อนที่เป็นกระบวนการที่ใช้เวกเตอร์การเคลื่อนที่ในการสร้างภาพประมาณกลับขึ้นมา

เทคนิคการชดเชยการเคลื่อนไหวเชิงกว้าง (Global Motion Compensation) โดยการใช้ Sprite ซึ่งเป็นเครื่องมือสำหรับตัดฉากภาพนิ่งในลักษณะพื้นหลังมุมกว้าง (Panorama Background) ทำให้มีการส่งข้อมูลของรูปพื้นหลังเพียงครั้งเดียว หลังจากนั้นในแต่ละเฟรมที่มีการใช้พื้นที่ของภาพพื้นหลังคนละส่วนกัน ก็จะส่งมาเฉพาะข้อมูลที่บอกถึงการเคลื่อนที่ไปของมุมมอง (Viewing Point of Camera) แล้วทำการสร้างพื้นหลังในส่วนนั้นใหม่ โดยใช้ข้อมูลที่ส่งมาในครั้งแรกสุด ไม่ต้องส่งข้อมูลมาใหม่ ทำให้สามรถลดข้อมูลที่จะต้องส่งลงได้ ตัวอย่างแสดงในรูปที่ 2.7



รูปที่ 2.7 ตัวอย่างการเข้ารหัสแบบ Sprite [10]

2.5.3 การเข้ารหัสรูปทรง (Shape Coding) [11]

ส่วนนี้เป็นส่วนสำคัญในการแสดงรูปทรงของวัตถุ เทคนิคที่ใช้ประกอบไปด้วย quadtree, chain code และ polygonal approximation ซึ่งจะแทนรูปทรงของวัตถุด้วยมาสก์ไบนารี การเข้ารหัสรูปทรงของวัตถุเป็นปัญหาสำคัญในการเข้ารหัสบนพื้นฐานของหน่วยวัตถุ อัตราบิตที่ใช้ในการเข้ารหัสรูปทรงขึ้นกับจำนวนของวัตถุ และยังขึ้นกับความซับซ้อนของขอบรูปทรง ซึ่งมีรายละเอียดค่อนข้างมาก ซึ่งสามารถจะลดความสับสนได้โดยใช้การเข้ารหัสรูปทรงประกอบกับข้อมูลรอบข้าง

(Content Based Shape Coding) การเข้ารหัสรูปทรงจะแทนขอบเขตของวัตถุด้วยรูปหลายเหลี่ยม (polygons) ซึ่งจะประมาณรูปทรงด้วยเส้นขอบที่มีจำนวนมุมน้อยที่สุด โดยยังคงรักษาลักษณะของรูปทรงเดิมอยู่ สำหรับวิธีการ Content Based Shape Coding จะมีความผิดเพี้ยนออกมาจากรูปทรงได้โดยที่สามารถครอบคลุมจุดภาพทั้งหมดในบริเวณที่สนใจได้ ขณะที่จะยอมให้ความผิดเพี้ยนเข้าไปในรูปทรงน้อยที่สุดเพื่อไม่ให้เกิดจุดภาพที่มีความกำกวมขึ้นในการแยกส่วนของภาพวัตถุ

การเข้ารหัสรูปทรง (Shape Coding) ที่ใช้กันมีดังนี้

- Binary Alpha Blocks (BAB)
  - CAE (Content-based Arithmetic Encoding) ซึ่งเป็นวิธีเข้ารหัสรูปทรงที่ใช้ใน MPEG-4
- 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

2.5.4 การเข้ารหัสพื้นผิว (Texture Coding)

ข้อมูลความผิดพลาดที่เหลือหลังจากการชดเชยการเคลื่อนที่ (residual errors) จะมีขั้นตอนการเข้ารหัสแบ่งได้เป็น 3 ขั้นตอน คือ ส่วนแสดงข้อมูลด้วยสัมประสิทธิ์การแปลง ส่วนการควอนไทซ์สัมประสิทธิ์ของการแปลง และส่วนเข้ารหัสเอนโทรปีสำหรับค่าสัมประสิทธิ์จากการควอนไทซ์

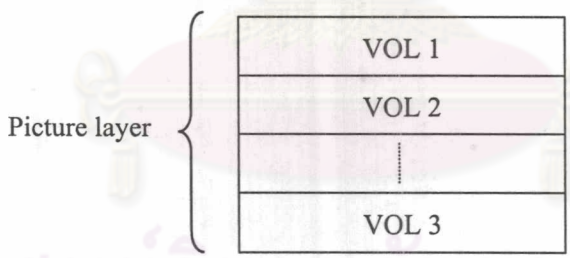
สำหรับส่วนของการแปลงมีหลายแบบด้วยกัน มาตรฐานส่วนใหญ่นิยมใช้การแปลงแบบ Discrete Cosine Transform (DCT) กับบล็อกขนาด NxN ซึ่ง DCT จัดเป็นการแปลงรองจากออปติมัล (suboptimal)

ส่วนการแปลงเวฟเลต (wavelet transform) คือ วิธีการเข้ารหัสแบบออปติมัล (optimal) สำหรับเข้ารหัสวัตถุรูปทรงต่างๆ โดยข้อมูลจะถูกเข้ารหัสข้อมูลโดยใช้กรรมวิธี Zerotree ซึ่งใช้กับการเข้ารหัสภาพนิ่ง [12]

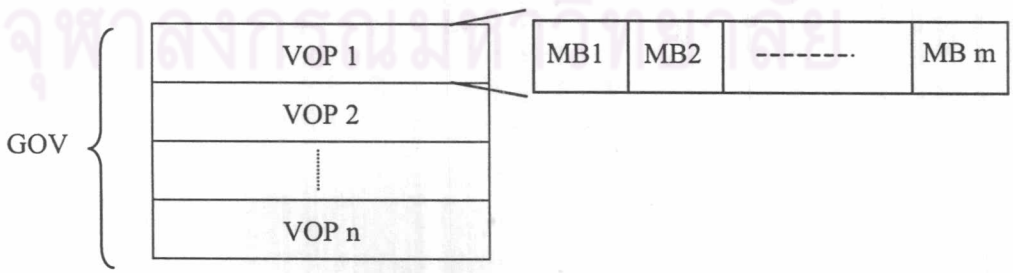
2.6 รูปแบบการจัดเรียงสัญญาณวิดีโอมาตรฐาน MPEG-4 [13]

สัญญาณวิดีโอที่จะถูกทำการเข้ารหัสจะจัดเรียงเป็นลำดับชั้นต่างๆ โดยจะมีข้อมูลส่วนหัวบอกว่าเป็นลำดับชนิดใด ทั้งนี้เพื่อตัวถอดรหัสจะสามารถถอดรหัสสัญญาณวิดีโอในเฟรมต่างๆ ได้ถูกต้อง โดยจะแบ่งเป็นลำดับชั้นต่างๆ ดังแสดงในรูปที่ 2.8 ดังนี้

- ชั้นบล็อก (Block layer) ประกอบไปด้วยจุดภาพขนาด 8 x 8 จุดภาพ
- ชั้นมาโครบล็อก (MB: Macroblock layer) แต่ละมาโครบล็อกจะประกอบไปด้วยจุดภาพขนาด 16 x 16 จุดภาพ
- ชั้นระนาบวัตถุ (VOP: Video Object Plane layer) แต่ละระนาบจะประกอบไปด้วยมาโครบล็อกที่เรียงต่อกันตั้งแต่มาโครบล็อกแรกถึงมาโครบล็อกสุดท้าย
- ชั้นกลุ่มระนาบวัตถุ (GOV: Group of VOP) แต่ละกลุ่มประกอบไปด้วยหลาย VOP ที่เรียงต่อกัน
- ชั้นวัตถุวิดีโอ (VOL : Video Object Layer) ประกอบไปด้วย VOP ของวัตถุวิดีโอเดียวกันที่เรียงต่อกัน
- ชั้นรูปภาพ (Picture layer) จะประกอบไปด้วยชั้นวัตถุวิดีโอตั้งแต่หนึ่งชั้นขึ้นไป (ถ้าภายในเฟรมมีเพียงวัตถุเดียว ชั้นรูปภาพนี้ก็ประกอบด้วยชั้นวัตถุวิดีโอเพียงชั้นเดียว)

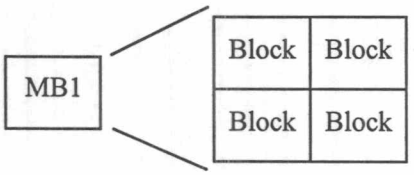


ก. ชั้นวัตถุวิดีโอ



ข. ชั้นกลุ่มระนาบวัตถุ ชั้นมาโครบล็อก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ค. ชั้นบล็อก

รูปที่ 2.8 การจัดเรียงสัญญาณวิดีโอตามมาตรฐาน MPEG-4

2.7 การควบคุมอัตราบิต TM5 (Test Model 5) [14]

การเข้ารหัสที่นำเสนอในวิยานิพนธ์นี้ ที่ชั้นฐานเข้ารหัสด้วยอัตราบิต 32 กิโลบิตต่อวินาที ตลอดทุกการทดลอง โดยใช้การควบคุมอัตราบิตที่เรียกว่า TM5

การควบคุมอัตราบิต TM5 ทำโดยการปรับค่าควอนไทซ์ของแต่ละมาโครบล็อก กระบวนการทำงานแบ่งเป็น 3 ขั้นตอนดังนี้

1. การจัดสรรบิต (Bit Allocation) : ทำหน้าที่ประมาณจำนวนบิตที่จะใช้ในการเข้ารหัสภาพเฟรมถัดไป ซึ่งจะต้องจัดสรรก่อนทำการเข้ารหัส ดังนี้

   - การประมาณความซับซ้อน (Complexity Estimation)

   หลังจากเข้ารหัสเฟรม I, P หรือ B แล้ว ต้องปรับค่าความซับซ้อนเชิงกว้าง (Global Complexity Measure) “ X_i, X_p หรือ X_b ” ดังนี้

$$X_i = S_i Q_i,$$
$$X_p = S_p Q_p,$$
$$X_b = S_b Q_b,$$

โดย Si, Sp, Sb คือ จำนวนบิตที่ถูกกำหนดโดยการเข้ารหัสเฟรมนี้

Qi, Qp, Qb คือ ค่าเฉลี่ยของค่าควอนไทซ์ทั้งเฟรม

และกำหนดค่าเริ่มต้นของพารามิเตอร์ต่างๆ ดังนี้

$$X_i = (160 * bit_rate)/115$$
$$X_p = (64 * bit_rate)/115$$
$$X_b = (42 * bit_rate)/115$$

โดย bit_rate คือ อัตราบิต (bits/s)

- การกำหนดจำนวนบิตของภาพ (Picture Target Setting)

จำนวนบิตสำหรับเฟรมถัดไปใน GOP (Group of Pictures) “Ti, Tp หรือ Tb” คำนวณจาก

$$T_i = \max\left\{\frac{R}{1+\frac{N_p X_p}{X_i X_p}+\frac{N_b X_b}{X_i X_b}}, \frac{bit_rate}{8\times picture_rate}\right\}$$

$$T_p = \max\left\{\frac{R}{N_p+\frac{N_b K_p X_b}{K_b X_p}}, \frac{bit_rate}{8\times picture_rate}\right\}$$

$$T_b = \max\left\{\frac{R}{N_b+\frac{N_p K_b X_p}{K_p X_b}}, \frac{bit_rate}{8\times picture_rate}\right\}$$

โดย K_p และ K_b เป็นค่าคงที่ เท่ากับ 1.0 และ1.4 ตามลำดับ

R คือ จำนวนบิตที่เหลือภายใน GOP ซึ่งปรับค่าหลังจากเข้ารหัส ดังนี้

$$R = R - S_{i,p,b}$$

หลังจากเข้ารหัสเฟรมแรก (เฟรม I) ใน GOP แล้ว กำหนด

$$R = G + R$$

และ $$G = bit_rate * N / picture_rate$$

โดย N คือ จำนวนเฟรมใน GOP

$R = 0$ ที่เฟรมแรกของลำดับภาพ

N_p และ N_b คือ จำนวนเฟรม P และเฟรม B ที่เหลือใน GOP ปัจจุบันที่เข้ารหัสอยู่

2. การควบคุมอัตรา (Rate Control) : การควบคุมในที่นี้หมายถึงการควบคุมอัตราบิตของบัฟเฟอร์ โดยขั้นตอนนี้ทำหน้าที่กำหนดค่าอ้างอิงของพารามิเตอร์ควอนไทซ์สำหรับแต่ละบล็อก

- คำนวณค่าบัฟเฟอร์ที่ใช้ไป (Fullness of Virtual Buffer) ก่อนเข้ารหัสมาโครบล็อกที่ j (j>=1) ดังนี้

$$d_j^i = d_0^i + B_{j-1} - \left(\frac{T_i \times (j-1)}{MB_cnt}\right)$$

หรือ

$$d_j^p = d_0^p + B_{j-1} - \left(\frac{T_p \times (j-1)}{MB_cnt} \right)$$

หรือ

$$d_j^b = d_0^b + B_{j-1} - \left(\frac{T_b \times (j-1)}{MB_cnt} \right)$$

ขึ้นอยู่กับชนิดของเฟรม

โดย d_0^i, d_0^p, d_0^b คือ ค่าเริ่มต้นของบัฟเฟอร์ที่ใช้ไปในแต่ละเฟรม

B_j คือ จำนวนบิตที่ใช้เข้ารหัสมาโครบล็อกทั้งหมดที่เข้ารหัสมาจนถึงมาโครบล็อกที่ j

MB_cnt คือจำนวนมาโครบล็อกในเฟรม

d_j^i, d_j^p, d_j^b คือ ค่าบัฟเฟอร์ที่ใช้ไปแล้วขณะที่เข้ารหัสมาโครบล็อกที่ j ในแต่ละเฟรม และเมื่อ j = MB_cnt กำหนด $d_j^i, d_j^p, d_j^b = d_0^i, d_0^p, d_0^b$ เพื่อเข้ารหัสเฟรมถัดไปของเฟรมชนิดเดียวกัน

- คำนวณค่าอ้างอิงควอนไทซ์ Q_j (Reference Quantization Parameter)สำหรับมาโคร บล็อกที่ j ดังนี้

$$Q_j = \left(\frac{d_j \times 31}{r} \right)$$

โดย

$$r = 2 \times \frac{bit_rate}{picture_rate}$$

กำหนดค่าเริ่มต้นของบัฟเฟอร์ ดังนี้

$$d_0^i = 10 \times \frac{r}{31}$$
$$d_0^p = K_p \times d_0^i$$
$$d_0^b = K_B \times d_0^i$$

3. การปรับค่าควอนไทซ์ (Adaptive Quantization) : ทำหน้าที่ปรับค่าควอนไทซ์ในแต่ละมาโครบล็อก

- คำนวณค่าการเปลี่ยนแปลงทางพื้นที่ (Spatial Activity Measure) สำหรับมาโครบล็อกที่ j

$$act_j = 1 + \min(vblk_1, vblk_2, ..., vblk_8)$$

โดย

$$vblk_n = \frac{1}{64} \times \sum_{k=1}^{64} \left(p_k^n - P_mean_n \right)^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

และ

$$P_mean_n = \frac{1}{64} \times \sum_{k=1}^{64} P_k^n$$

และ P_k คือ ค่าในบล็อกต้นแบบ (original 8*8 block) บล็อกที่ n

- หาค่า Normalise actj

$$N_act_j = \frac{(2 \times act_j) + avg_act}{act_j + (2 \times avg_act)}$$

โดย avg_act คือ ค่าเฉลี่ยของ actj ในเฟรมที่เข้ารหัสเฟรมที่แล้ว เรากำหนดค่าเริ่มต้นเท่ากับ 400

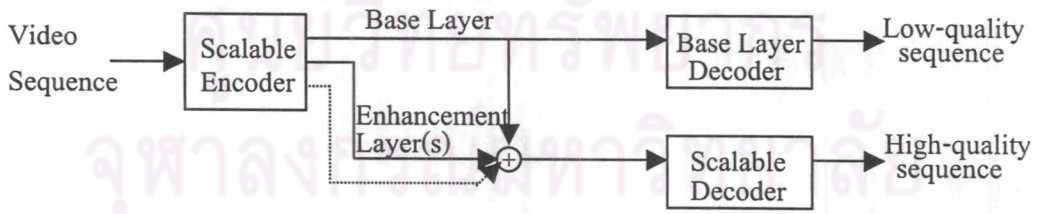
- ค่า $mquant_j$ ที่ได้ คือ

$$mquant_j = Q_j \times N_act_j$$

โดย Q_j คือ ค่าอ้างอิงควอนไทซ์ที่ได้จากขั้นตอนที่ 2

2.8 การเข้ารหัสแบบสเกลาเบิล (Scalable or layered video coding) [15][17]

หลักการพื้นฐานของการเข้ารหัสแบบสเกลาเบิล คือ วิดีโอถูกเข้ารหัสเป็นหลายชั้น ซึ่งแต่ละชั้นประกอบด้วยข้อมูลของวิดีโอแต่ละส่วนกัน การที่จะสามารถรองรับการปรับระดับการแสดงผล (Scalability) ได้นั้น ตัวถอดรหัสต้องสามารถเลือกถอดรหัสบางชั้น เพื่อปรับระดับการแสดงผลวิดีโอให้ได้คุณภาพในระดับหนึ่ง โดยคุณภาพในที่นี้หมายถึง ความละเอียดทางพื้นที่ (Spatial Resolution) ความละเอียดทางเวลา (Temporal Resolution) และการปรับค่าการควอนไทซ์ (Quantization Scale)



รูปที่ 2.9 ระบบการเข้ารหัสแบบสเกลาเบิล (Scalable or layered coding system) [17]

แผนภาพในรูปที่ 2.9 แสดงระบบการเข้ารหัสแบบสเกลาเบิล ภาพวิดีโอถูกเข้ารหัสด้วยคุณภาพต่ำที่ชั้นฐาน (Base Layer) ข้อมูลที่เหลือซึ่งเป็นความแตกต่างระหว่างคุณภาพหยาบๆ ของชั้นฐานกับภาพต้นฉบับ ถูกเข้ารหัสในชั้นเอนแฮนส์เมนท์ (Enhancement Layer) เป็นชั้นเดียวหรือลำดับชั้น (จากรูป 2.9 การเข้ารหัสชั้นเดียวแทนด้วยเส้นทึบ Enhancement Layer เส้นเดียว ถ้าเข้ารหัสเป็นลำดับชั้น แทนด้วยเส้นทึบและเส้นประแสดงให้เห็นถึงการมีมากกว่าหนึ่งชั้น) เมื่อถอด

รหัสข้อมูลชั้นฐาน ผลที่ได้คือ วิดีโอคุณภาพต่ำ และคุณภาพของวิดีโอที่ถูกถอดรหัสถูกปรับปรุงโดยการถอดรหัสข้อมูลชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์

การเข้ารหัสแบบสเกลาเบิล ช่วยปรับปรุงความยืดหยุ่นของระบบการส่งผ่านสัญญาณวิดีโอ ซึ่งการเข้ารหัสวิธีนี้จะลดประสิทธิภาพการบีบอัดข้อมูล แต่อย่างไรก็ตาม วิธีนี้ยังคงมีประสิทธิภาพมากกว่าการเข้ารหัสและการส่งลำดับภาพวิดีโอแบบ multiple ซึ่งถอดรหัสให้ได้ความละเอียดและคุณภาพที่ระดับต่างๆ

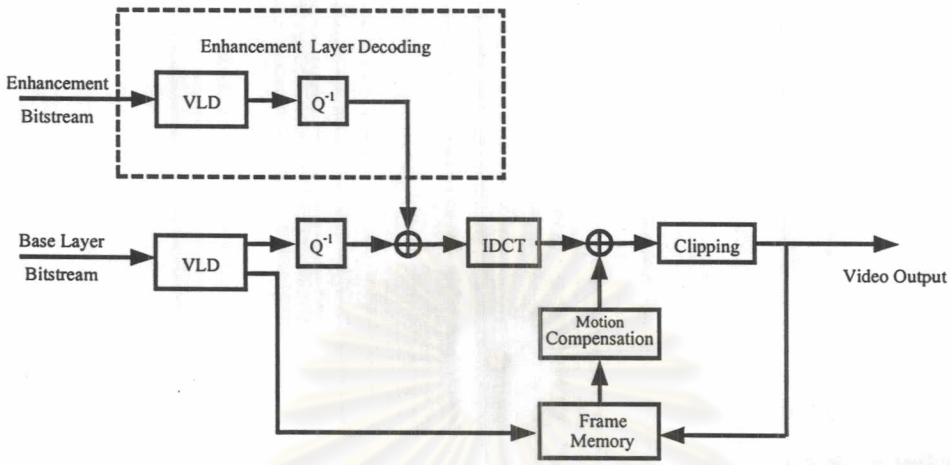
การเข้ารหัสแบบสเกลาเบิลสามารถปรับปรุงสมรรถนะในการเกิดความผิดพลาดขณะส่ง โดยทั่วไปข้อมูลแต่ละชั้นมีความทนทานต่อการเกิดความผิดพลาดต่างกัน ส่วนใหญ่ชั้นฐานมีความสำคัญในการถอดรหัสมากที่สุด และชั้นเอนแฮนส์เมนท์มีความสำคัญน้อยกว่า โดยชั้นฐานสามารถถูกส่งด้วยลำดับความสำคัญที่สูงหรืออัตราความผิดพลาดต่ำ ขณะที่ชั้นเอนแฮนส์เมนท์สามารถทนต่อการส่งที่ลำดับความสำคัญที่ต่ำกว่าหรืออัตราความผิดพลาดที่สูงกว่าได้ การเข้ารหัสแบบนี้สามารถช่วยลดผลของความผิดพลาดที่เกิดขึ้น ชั้นเอนแฮนส์เมนท์ที่ลำดับความสำคัญต่ำจะถูกกระทบโดยความผิดพลาดก่อนชั้นอื่น แต่ทำให้เกิดความเสียหายเพียงเล็กน้อย ส่วนชั้นฐานจะไม่ถูกกระทบยกเว้นกรณีที่อัตราความผิดพลาดสูง

การเข้ารหัสแบบสเกลาเบิลสามารถส่งผ่านโครงข่ายต่างชนิดกันได้ ซึ่งคุณภาพของการบริการที่ได้รับก็ต่างกัน โดยทุกชั้นข้อมูลถูกเข้ารหัสและถูกส่งจากต้นทาง (Source) ตัวอย่างเช่น ถ้าโครงข่ายย่อยไม่สามารถส่งบิตข้อมูลด้วยอัตราบิตสูง ดังนั้นมันควรสามารถปรับระดับการแสดงผลโดยการถอดรหัสข้อมูลชั้นล่างเท่านั้น ในส่วนต่อไปนี้ อธิบายถึงการเข้ารหัสแบบสเกลาเบิลวิธีต่างๆ

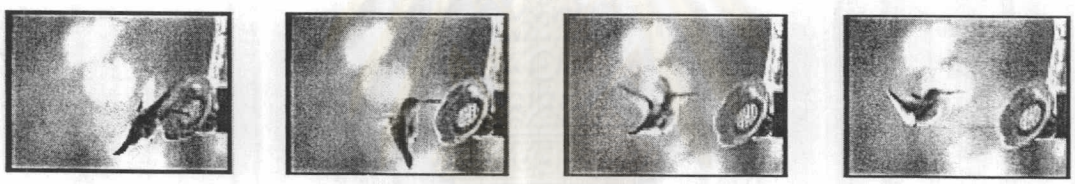
2.8.1 การเข้ารหัสแบบ Signal-to-Noise Ratio (SNR) Scalability

Signal-to-Noise Ratio (SNR) Scalability คือ เทคนิคหนึ่งที่เข้ารหัสสัญญาณวิดีโอเป็นสองชั้นโดยอัตราเฟรมและความละเอียดเดียวกันทั้ง 2 ชั้น แต่ค่าควอนไทซ์ต่างกัน รูปที่ 2.10 แสดงโครงสร้างตัวถอดรหัสแบบ SNR Scalability ตามมาตรฐานการเข้ารหัสสัญญาณวิดีโอ MPEG2 โดยบิตข้อมูลที่ชั้นฐานถูกถอดรหัสด้วย Variable Length Decoding (VLD) ที่ชั้นฐาน แล้วถูกแปลงควอนไทซ์ย้อนกลับเพื่อสร้างสัมประสิทธิ์ DCT กลับคืนมา ส่วนบิตข้อมูลที่ชั้นเอนแฮนส์เมนท์ ถูกถอดรหัสด้วย VLD ที่ชั้นบน และสร้างสัมประสิทธิ์ DCT ส่วนเหลือของเอนแฮนส์เมนท์ (enhancement residues) โดยแปลง ควอนไทซ์ย้อนกลับในชั้นบน ดังนั้นจะได้สัมประสิทธิ์ DCT ที่มีความถูกต้องสูงเมื่อรวมสัมประสิทธิ์ DCT ที่สร้างกลับคืนมาที่ชั้นฐาน กับสัมประสิทธิ์ DCT ส่วนเหลือของเอนแฮนส์เมนท์ สัมประสิทธิ์ DCT ที่มีความถูกต้องสูงถูกนำมาแปลง DCT ย้อนกลับ และรวมกับบล็อกที่ถูกชดเชยการเคลื่อนที่จากเฟรมก่อนหน้า ได้เป็นภาพวิดีโอ สังเกตคุณภาพที่แตกต่างกัน

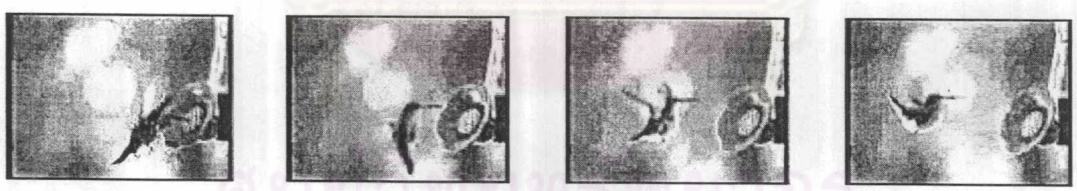
ของภาพชั้นฐานกับภาพชั้นเอนแฮนส์เมนท์ในรูปที่ 2.11 ภาพชั้นฐานมีคุณภาพต่ำกว่าภาพที่ได้จากชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์



รูปที่ 2.10 โครงสร้างตัวถอดรหัสแบบ SNR Scalability



ก) ภาพชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์



ข) ภาพชั้นฐาน

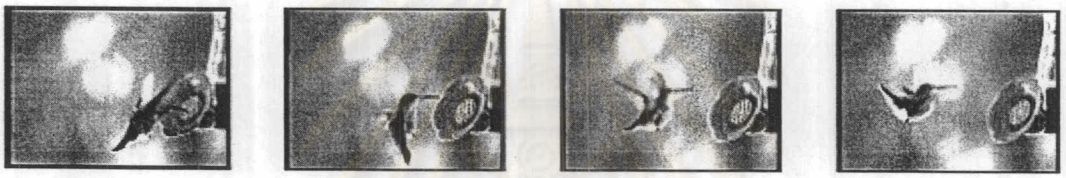
รูปที่ 2.11 คุณภาพของการเข้ารหัสแบบ SNR Scalability [2]

ในตัวถอดรหัส SNR Scalability นั้น ข้อมูลที่ชั้นเอนแฮนส์เมนท์ถูกใช้ในการทำนายการเคลื่อนที่ด้วย ดังนั้นผลที่ตามมา คือ

1. ถ้าตัวเข้ารหัสใช้ข้อมูลที่ชั้นเอนแฮนส์เมนท์ในการทำนายการเคลื่อนที่ และตัวถอดรหัสได้รับข้อมูลที่ชั้นเอนแฮนส์เมนท์นั้น ดังนั้นประสิทธิภาพการเข้ารหัสจะสูง

2. ถ้าตัวเข้ารหัสใช้ข้อมูลที่ชั้นเอนแฮนส์เมนท์ในการทำนายการเคลื่อนที่ และตัวถอดรหัสไม่ได้รับข้อมูลที่ชั้นเอนแฮนส์เมนท์นั้น ทำให้เกิด drift error ที่ชั้นฐาน และประสิทธิภาพการเข้ารหัสจะต่ำ
3. ถ้าตัวเข้ารหัสไม่ใช้ข้อมูลที่ชั้นเอนแฮนส์เมนท์ในการทำนายการเคลื่อนที่ และตัวถอดรหัสได้รับข้อมูลที่ชั้นเอนแฮนส์เมนท์นั้น ทำให้เกิด drift error ที่ชั้นเอนแฮนส์เมนท์และประสิทธิภาพการเข้ารหัสต่ำ
4. ถ้าตัวเข้ารหัสไม่ใช้ข้อมูลที่ชั้นเอนแฮนส์เมนท์ในการทำนายการเคลื่อนที่ และตัวถอดรหัสไม่ได้รับข้อมูลที่ชั้นเอนแฮนส์เมนท์นั้น ผลที่ได้เหมือนกับใช้ชั้นฐานชั้นเดียว

2.8.2 การเข้ารหัสแบบ Temporal Scalability



ก) ภาพชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์

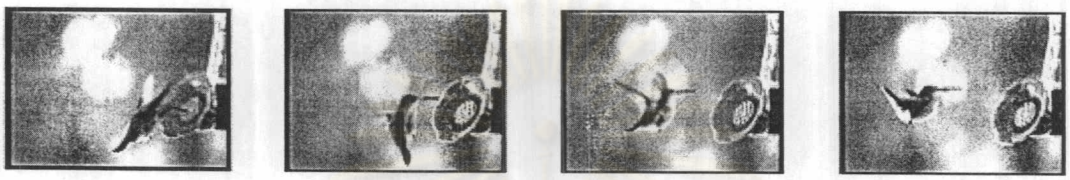
ข) ภาพชั้นฐาน

รูปที่ 2.12 คุณภาพของการเข้ารหัสแบบ Temporal Scalability [2]

Temporal Scalability เป็นเทคนิคที่แบ่งการเข้ารหัสสัญญาณวิดีโอเป็นสองชั้นที่ความละเอียดเท่ากัน แต่อัตราเฟรมต่างกัน โดยที่ชั้นฐานถูกเข้ารหัสด้วยอัตราเฟรมต่ำกว่า และชั้นเอนแฮนส์เมนท์เข้ารหัสด้วยอัตราเฟรมที่สูงกว่า ภาพที่ได้จาการเข้ารหัสแบบนี้ดูได้จากตัวอย่างในรูปที่ 2.12 ซึ่งแสดงให้เห็นว่าภาพชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์มีจำนวนเฟรมมากกว่าชั้นฐาน แต่ขนาดเท่ากันทั้งสองชั้น

2.8.3 การเข้ารหัสแบบ Spatial Scalability

Spatial Scalability เป็นเทคนิคที่แบ่งการเข้ารหัสสัญญาณวิดีโอเป็นสองชั้นที่อัตราเฟรมเท่ากัน แต่ความละเอียดต่างกัน โดยที่ชั้นฐานเข้ารหัสด้วยความละเอียดต่ำกว่าชั้นเอนแฮนส์เมนท์ และเพื่อถอดรหัสภาพที่ความละเอียดที่สูงขึ้นกว่าชั้นฐาน ดังนั้นภาพที่สร้างกลับคืนมาจากชั้นฐานจะถูกเพิ่มขนาด (up-sampling) แล้วรวมกับข้อมูลชั้นเอนแฮนส์เมนท์ ภาพที่ได้แสดงดังรูปที่ 2.13



ก) ภาพชั้นฐานรวมกับชั้นเอนแฮนส์เมนท์

ข) ภาพชั้นฐาน

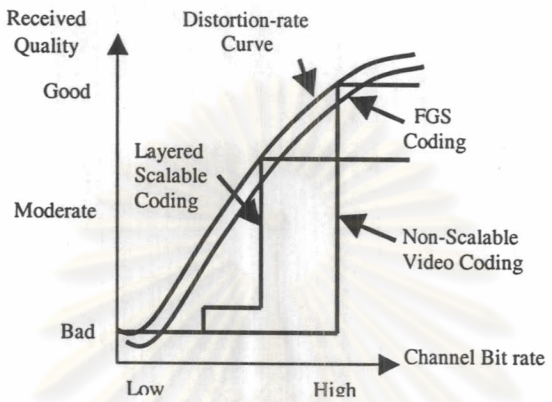
รูปที่ 2.13 คุณภาพของการเข้ารหัสแบบ Spatial Scalability [2]

2.9 การเปรียบเทียบประสิทธิภาพการเข้ารหัสแบบต่างๆ [15]

รูปที่ 2.14 แสดงประสิทธิภาพของการเข้ารหัสสัญญาณวิดีโอด้วยวิธีต่างๆ โดยแกนนอนแสดงอัตราบิตของช่องสัญญาณ แกนตั้งแสดงคุณภาพของวิดีโอที่ได้รับโดยผู้ใช้ มี distortion-rate curve แสดงขอบเขตบนของคุณภาพสำหรับเทคนิคการเข้ารหัสต่างๆ ที่อัตราบิตต่างๆ กราฟเส้นแรกคือ การเข้ารหัสแบบ nonscalable เป็นการเข้ารหัสสัญญาณวิดีโอแบบชั้นเดียว โดยถ้าอัตราบิตของช่องสัญญาณเท่ากับอัตราบิตที่เข้ารหัสสัญญาณวิดีโอ คุณภาพที่ได้รับจะดีที่สุด อย่างไรก็ตาม ถ้าอัตราบิตของช่องสัญญาณต่ำกว่าอัตราบิตที่เข้ารหัสสัญญาณวิดีโอ คุณภาพที่ได้รับจะเลวลงมาก นอกจากนั้น ถ้าอัตราบิตของช่องสัญญาณสูงกว่าอัตราบิตที่เข้ารหัสสัญญาณวิดีโอ คุณภาพที่ได้รับก็ไม่ได้ดีขึ้น

การเข้ารหัสแบบสเกลาเบิลนั้น ที่ชั้นเอนแฮนส์เมนท์ถูกเข้ารหัสด้วย run-level entropy encoding ดังนั้นการส่ง การรับและการถอดรหัสบิตข้อมูลเอนแฮนส์เมนท์ต้องกระทำได้ครบถ้วนจึงจะเพิ่มคุณภาพวิดีโอได้ ถ้าถูกลดทอนจะไม่เพิ่มคุณภาพวิดีโอ นี่เป็นเหตุผลที่เทคนิคการเข้ารหัสนี้มีกราฟประสิทธิภาพเป็น 2 ขั้นบันได ขั้นแรกเป็นคุณภาพของชั้นฐาน ขั้นที่สองเป็นคุณภาพของชั้น

ฐานรวมกับชั้นเอนแฮนส์เมนท์ ดังแสดงในรูปที่ 2.14 ความแตกต่างระหว่าง FGS กับการเข้ารหัสสเกลาเบิลวิธีอื่น คือถึงแม้ว่า FGS จะเข้ารหัสเป็น 2 ชั้น แต่บิตข้อมูลเอนแฮนส์เมนท์สามารถถูกตัดทอนตามแบนด์วิดท์ได้ กลายเป็นการจัดการเพิ่มประสิทธิภาพบางส่วนเท่าที่สามารถถอดรหัสได้สำหรับแต่ละเฟรม ในขณะที่สเกลาเบิลวิธีอื่นทำไม่ได้ ดังนั้นกราฟประสิทธิภาพของ FGS จึงเป็นเส้นโค้งต่อเนื่อง

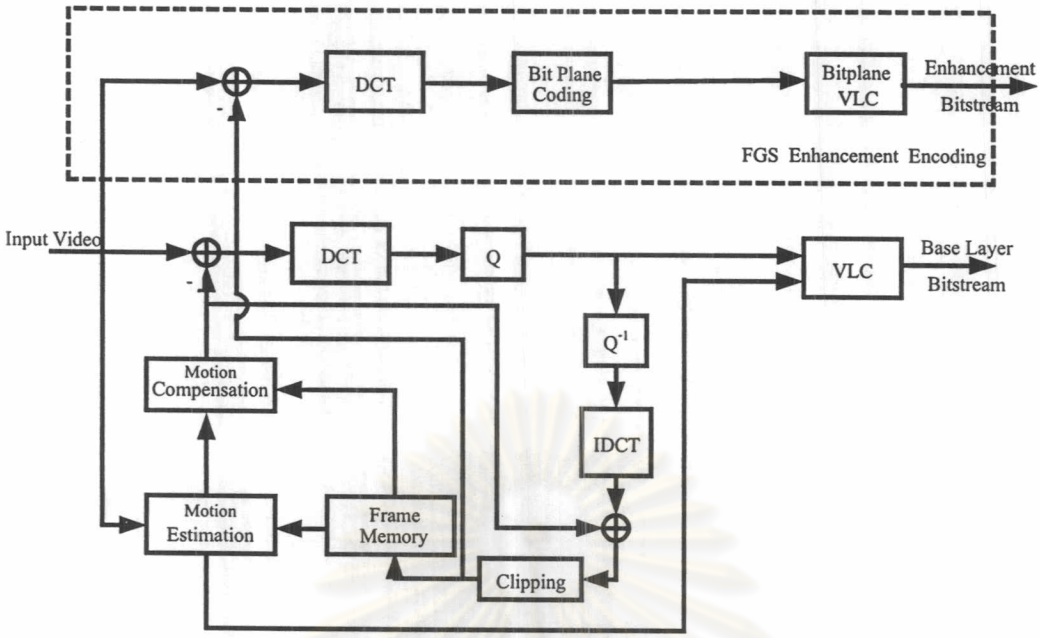


รูปที่ 2.14 กราฟประสิทธิภาพของการเข้ารหัสสัญญาณวิดีโอแบบ scalability [15]

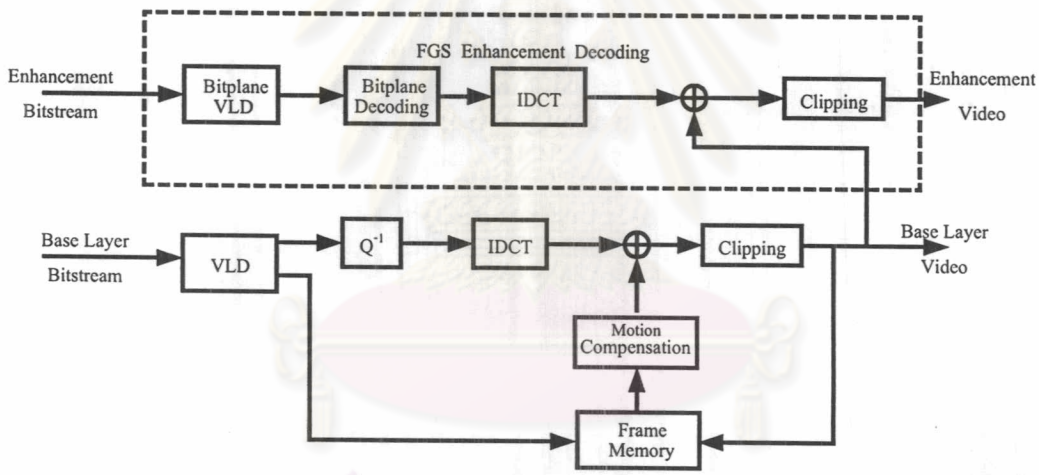
2.10 การเข้ารหัสแบบ FGS (Fine Granularity Scalability) [15]

โครงสร้างการเข้ารหัส FGS แสดงดังรูปที่ 2.15 แบ่งเป็น 2 ชั้น คือ ชั้นฐานและชั้นเอนแฮนส์เมนท์ โดยที่ชั้นฐานนั้นจะมีโครงสร้างเหมือนกับการเข้ารหัสตามมาตรฐาน MPEG-4 ส่วนชั้นเอนแฮนส์เมนท์จะนำภาพที่ถูกสร้างกลับมาซึ่งเกิดการสูญเสียจากการควอนไทซ์ไปลบออกจากภาพจริง เพราะฉะนั้นสิ่งที่ได้ คือส่วนเหลือที่เป็นความผิดพลาดเนื่องจากการควอนไทซ์ (quantized error residue) ค่าส่วนเหลือที่เป็นความผิดพลาดเนื่องจากการควอนไทซ์นี้ถูกนำไปแปลง DCT และเข้ารหัสแบบระนาบบิต (bitplane coding) ซึ่งข้อดีของการเข้ารหัสแบบระนาบบิต คือลดจำนวนบิตลงร้อยละ 20 เนื่องจากการเข้ารหัสแบบระนาบบิตนี้มีตารางการเข้ารหัสแบบปรับความยาว (VLC table) เพื่อบีบอัดข้อมูลทั้งหมด 4 ตารางที่แตกต่างกันสำหรับแต่ละระนาบบิต มีผลทำให้มีประสิทธิภาพในการบีบอัดสูงขึ้น [16]

โครงสร้างตัวถอดรหัสดังแสดงในรูปที่ 2.16 ก็มีสองชั้นเช่นกันคือชั้นฐานและชั้นเอนแฮนส์เมนท์ โดยที่ชั้นเอนแฮนส์เมนท์จะนำภาพที่ได้จากชั้นฐานมารวมกับผลที่ได้จากชั้นเอนแฮนส์เมนท์ ดังนั้นคุณภาพของภาพที่ได้ก็จะดีกว่าที่ชั้นฐาน



รูปที่ 2.15 โครงสร้างตัวเข้ารหัส FGS



รูปที่ 2.16 โครงสร้างตัวถอดรหัส FGS

2.10.1 การเข้ารหัสสัมประสิทธิ์ DCT แบบระนาบบิต(Bit-Plane Coding of the DCT Coefficients)

การเข้ารหัสสัมประสิทธิ์ DCT แบบระนาบบิต มีขั้นตอนดังนี้

1. หาค่าสูงสุดของค่าสัมบูรณ์ (absolute values) ของส่วนเหลือ (residues) ทั้งหมดในแต่ละ VOP (Video Object plane) และหาว่าต้องใช้จำนวนบิต (N) ต่ำที่สุดเท่าไร โดยหาจากการแปลงค่าสูงสุดของค่าสัมบูรณ์ให้เป็นเลขฐานสอง (binary format) ซึ่ง N ก็คือจำนวนของระนาบบิตสำหรับ VOP นั้น

2. ภายในแต่ละบล็อก (8x8) แทนค่าสัมบูรณ์ 64 ค่าให้เป็นเลขฐานสองและ N ระนาบบิต โดยหนึ่งระนาบบิตถูกกำหนดเป็นอาเรย์ 64 บิต
3. เริ่มจากระนาบบนสุดเรียกว่า most signficant bit plane (MSB plane) เขียนให้อยู่ในรูป (RUN,EOP) ซึ่งมีส่วนประกอบ 2 ส่วน ดังนี้ 3.1 RUN คือจำนวนบิตศูนย์ที่ต่อเนื่องกันก่อนหน้าบิตหนึ่ง 3.2 EOP ซึ่งจะให้ค่าเป็น 1 เมื่อพบ 1 บิตสุดท้ายของระนาบ และถ้าระนาบถัดลงมาจาก MSB plane ประกอบด้วยศูนย์ทั้งหมดก็แทนด้วยสัญลักษณ์ ALL-ZERO ซึ่งเมื่อเข้ารหัส VLC แล้วรหัสที่ได้จะสั้นกว่ารหัสอื่น

สมมติว่ามีค่าสัมบูรณ์หลังจากทำการสแกนแบบซิกแซก (zigzag scan) เป็นดังนี้

10, 0, 6, 0, 0, 3, 0, 2, 2, 0, 0, 2, 0, 0, 1, 0, ... 0, 0

ค่าสูงสุดในบล็อกคือ 10 ซึ่งก็คือ 1010 จำนวนบิตที่ต่ำที่สุดก็คือ 4 เมื่อเขียนทุกค่าเป็นเลขฐานสองโดยใช้ 4 บิต จะได้ 4 ระนาบบิต ดังนี้

1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... 0, 0 (MSB-plane)

0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... 0, 0 (Second MSB-plane)

1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, ... 0, 0 (Third MSB-plane)

0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, ... 0, 0 (Fourth MSB-plane or LSB-plane)

เมื่อแปลงบิตของ 4 ระนาบบิตให้อยู่ในรูป (RUN,EOP) จะเป็นดังนี้

(0, 1) (MSB-plane)

(2, 1) (Second MSB-plane)

(0, 0), (1,0), (2,0), (1,0), (0, 0), (2, 1) (Third MSB-plane)

(5, 0), (8, 1) (Fourth MSB-plane or LSB-plane)

ดังนั้นในตัวอย่างนี้ก็จะมีทั้งหมด 10 (RUN,EOP) สัญลักษณ์

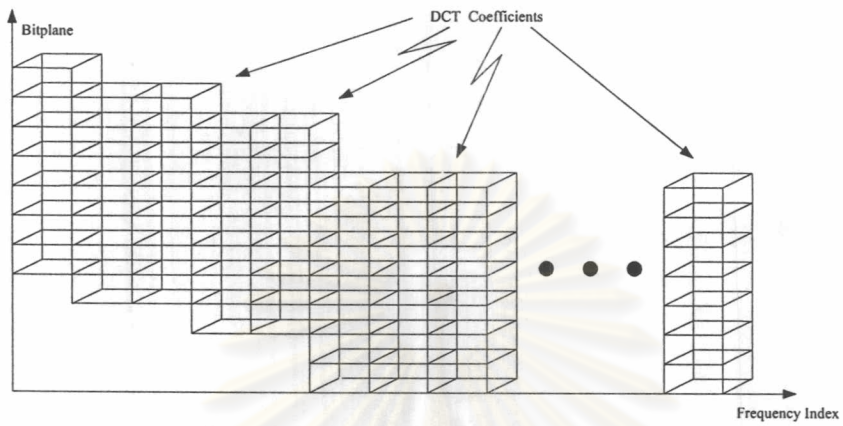
2.10.2 คุณลักษณะพิเศษของ FGS

1.) Frequency Weighting

เป็นที่ทราบกันดีว่า สัมประสิทธิ์ DCT ที่ความถี่ต่ำมีความสำคัญมากกว่าที่ความถี่สูง ดังนั้นคุณภาพก็จะดีกว่าถ้าที่ชั้นเอนแฮนส์เมนท์สามารถส่งบิตข้อมูลของสัมประสิทธิ์ DCT ที่ความถี่ต่ำไปได้ครบหรือไม่ถูกตัดทิ้งไปเนื่องจากแบนด์วิดท์ของช่องสัญญาณมีไม่พอที่จะส่งบิตสตรีมของชั้นเอนแฮนส์เมนท์ทั้งหมด แนวความคิดของ Frequency Weighting จึงมีอยู่ว่า ในเมื่อช่องสัญญาณส่งให้ได้ไม่หมดก็ทำการส่งความถี่ต่ำไปก่อน โดย กรรมวิธีคือ สัมประสิทธิ์ DCT ที่มีความถี่ต่ำๆ จะได้รับการถ่วงน้ำหนักโดยคูณค่าคงที่เข้าไปทำให้ค่าที่ได้มากกว่าค่าเดิม เมื่อนำมาวาดเป็นกราฟจะเห็นว่าสัมประสิทธิ์ DCT ที่มีความถี่ต่ำๆถูกยกขึ้นไปมากกว่าที่ความถี่สูงๆ ดังรูปที่ 2.17 นั่น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

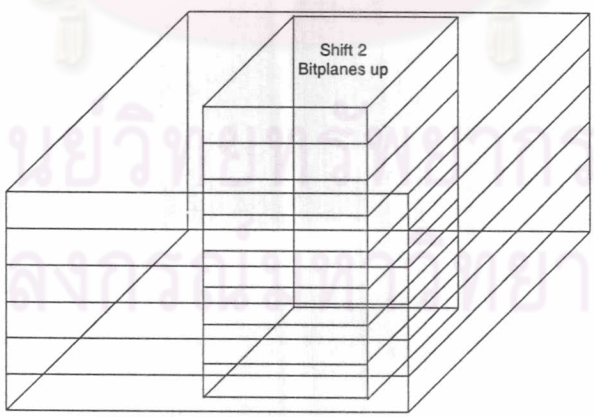
หมายความว่า ในการส่งออกช่องสัญญาณจริง สัมประสิทธิ์ DCT ที่มีความถี่ต่ำๆจะได้รับการส่งออกไปก่อน เพราะฉะนั้นถ้าในช่วงหลัง ช่องสัญญาณเกิดมีแบนด์วิดท์ให้ไม่พอที่จะส่งข้อมูลได้ทั้งหมด ภาพที่ได้ก็ยังมีคุณภาพที่ดีกว่า



รูปที่ 2.17 bit shift สำหรับ frequency weighting

2.) Selective Enhancement

ในกรรมวิธี Selective Enhancement ก็ใช้หลักการเดียวกับ Frequency Weighting เพียงแต่เลือกส่วนของภาพว่าบริเวณใดควรส่งก่อน ก็จะได้รับการถ่วงน้ำหนักโดยคูณค่าคงที่เข้าไปทำให้ค่าที่ได้มากกว่าค่าเดิม ทำให้บริเวณนั้น ๆ ได้ส่งออกไปก่อนนั่นเอง ดังแสดงในรูปที่ 2.18

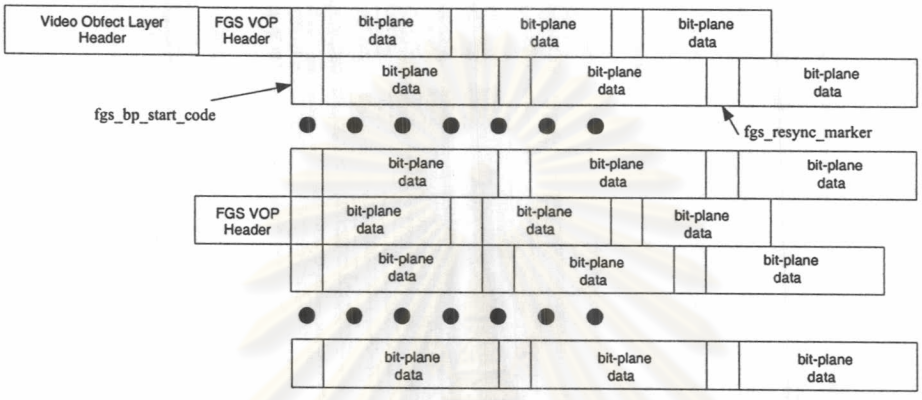


รูปที่ 2.18 selective enhancement

3.) Error Resilience

จากหลักการของ FGS นั้นจะพบความจริงว่า เมื่อเกิดความผิดพลาดภายในชั้นเอนแฮนส์เมนท์แล้ว ความผิดพลาดนั้น ๆ จะไม่แพร่กระจายหรือมีผลกับภาพต่อ ๆ ไป เพราะว่าภาพในชั้น

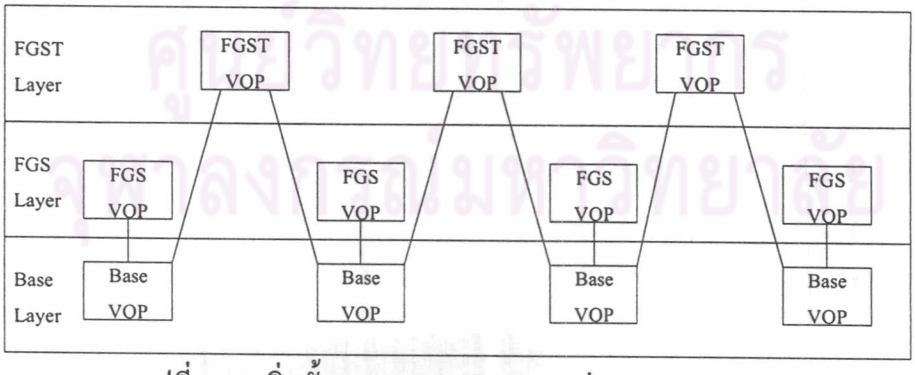
เอนแฮนส์เมนท์ไม่ได้นำมาใช้ในการประมาณหรือใช้เป็นภาพอ้างอิงใด ๆ อย่างไรก็ตามถ้าต้องการเพิ่มความยืดหยุ่นในการจำกัดขอบเขตของความผิดพลาดก็สามารถเพิ่มบิตพิเศษเข้าไปได้ ในกรณีนี้ได้เพิ่ม fgs_resync_marker เข้าไปเพื่อจำกัดขอบเขตของความผิดพลาด คือทำให้ทราบว่าบิตที่ผิดนั้นเกิดจาก DCT ตำแหน่งใด รูปที่ 2.19 แสดงให้เห็นว่าควรจะใส่ fgs_resync_marker เข้าไปที่ตำแหน่งใด



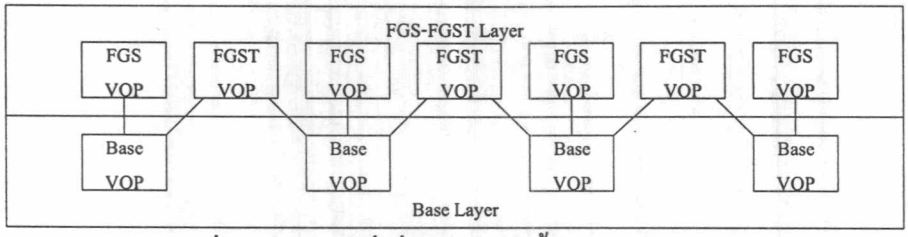
รูปที่ 2.19 โครงสร้าง FGS bitstream ที่มี resynchronization markers

4.) FGS Temporal Scalability (FGST)

ในการที่จะเพิ่มประสิทธิภาพของ FGS ให้คุณภาพของภาพดีขึ้น สามารถที่จะเพิ่มความซับซ้อนเข้าไปที่ชั้นเอนแฮนส์เมนท์ โดยเพิ่มกระบวนการชดเชยและการประมาณจากชั้นฐาน เพื่อเพิ่มจำนวนเฟรมในชั้นเอนแฮนส์เมนท์ก็จะทำให้ภาพที่ได้รับนั้นมีคุณภาพของภาพดีขึ้น การเพิ่มจำนวน เฟรมอาจกระทำได้ 2 แบบ ดังรูปที่ 2.20 และ รูปที่ 2.21



รูปที่ 2.20 เพิ่มชั้น enhancement FGST ต่อจาก FGS layer



รูปที่ 2.21 FGST ที่เพิ่มเข้าไปในชั้นเอนแฮนส์เมนท์

รูปที่ 2.20 เป็นการสร้าง FGST ขึ้นมาอีกชั้นหนึ่งแต่รูปที่ 2.21 ได้เพิ่ม FGST เข้าไปที่ชั้นเอนแฮนส์เมนท์ FGS ทั้ง 2 แบบต่างกันตรงที่ถ้าเป็นรูปที่ 2.20 จะมีบิตสตรีมเป็นของตัวเอง แต่รูปที่ 2.21 จะต้องมีสิ่งที่บอกว่า ส่วนไหนคือ FGS และ FGST

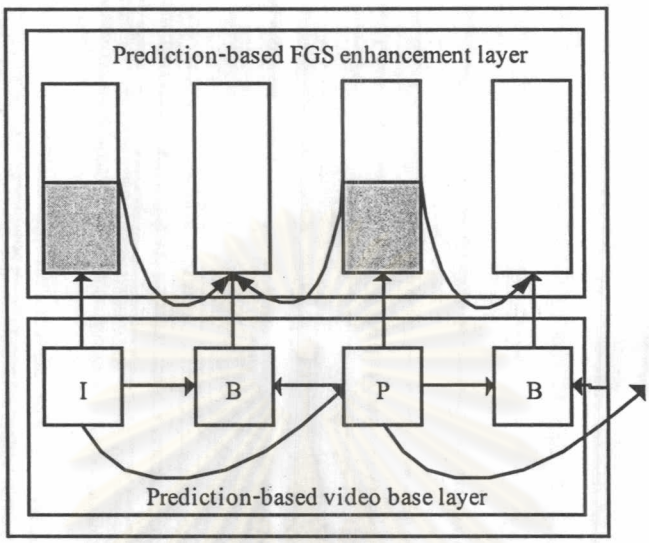
2.11 การเข้ารหัสแบบ MC-FGS (Motion Compensation FGS) [6][18][19]

กรรมวิธีที่เรียกว่า Motion Compensation FGS มีหลักการสำคัญคือ เพิ่มกระบวนการทาง Motion Compensation เข้าไปในชั้น FGS ได้มีการเสนอ 2 กรรมวิธีคือ two-loop MC-FGS และ single-loop (or one loop) MC-FGS

2.11.1 การเข้ารหัสแบบ two-loop MC-FGS

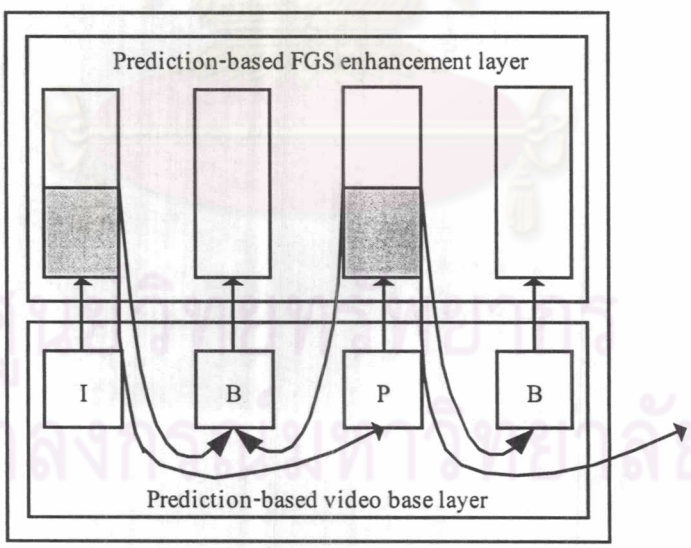
มีหลักการว่า จะนำเวคเตอร์การเคลื่อนที่ (motion vector) ของชั้นฐานไปช่วยในกระบวนการชดเชยการเคลื่อนที่ ซึ่งกระบวนการชดเชยการเคลื่อนที่นั้น (ถูกเพิ่มเติมขึ้น) มีอยู่ภายในชั้นเอน-แฮนส์เมนท์ (แต่เดิมไม่มี) อินพุตของชั้นเอนแฮนส์เมนท์เป็นผลต่างของ error residue ที่เกิดจากการควอนไทต์ การจัดเก็บภาพในหน่วยความจำภาพ (frame memory) จะเก็บเพียงบางระนาบบิต เช่น M ระนาบบิต โดยที่การเลือกจำนวน M จะพิจารณาจากอัตราการเข้ารหัส (coding gain) และความผิดพลาดจากการทำนาย (prediction drift) ถ้ากำหนดให้ M มีค่ามากๆ จะยิ่งเป็นการเพิ่มอัตราการเข้ารหัส เพราะจะทำให้การชดเชยการเคลื่อนที่ถูกต้องมากขึ้น แต่ยิ่งค่า M มากขึ้นเท่าไร จะมีผลต่อตัวถอดรหัส คือ ถ้าได้รับระนาบบิตน้อยกว่า M ระนาบ จะทำให้เกิดการ drift มากขึ้น พิจารณาในเชิงความทนทานต่อการสูญหายของข้อมูล (packet loss resilience) จะพบว่า ถ้าเกิดความผิดพลาดที่เฟรม I บน ชั้นเอนแฮนส์เมนท์ จะทำให้ความผิดพลาดนี้แพร่ไปบนเฟรม B และเฟรม P ด้วย และถ้าเกิดความผิดพลาดที่เฟรม P บนชั้นเอนแฮนส์เมนท์แล้ว จะทำให้ความผิดพลาดนี้แพร่ไปบนเฟรม B และเฟรม P ด้วยเช่นกัน อย่างไรก็ตามความผิดพลาดจะแพร่ไปไม่เกิน 2 เฟรม บนชั้นเอนแฮนส์เมนท์

โครงสร้างของ 2-loop MC-FGS แสดงได้ดังรูปที่ 2.22 จะเห็นได้ว่าที่ชั้นเอนแฮนส์เมนท์ได้มีการนำเอากระบวนการชดเชยมาใช้ แต่จำกัดไว้เพียงบางระนาบบิต



รูปที่ 2.22โครงสร้างของ 2-loop MC-FGS

2.11.2 การเข้ารหัสแบบ single-loop MC-FGS

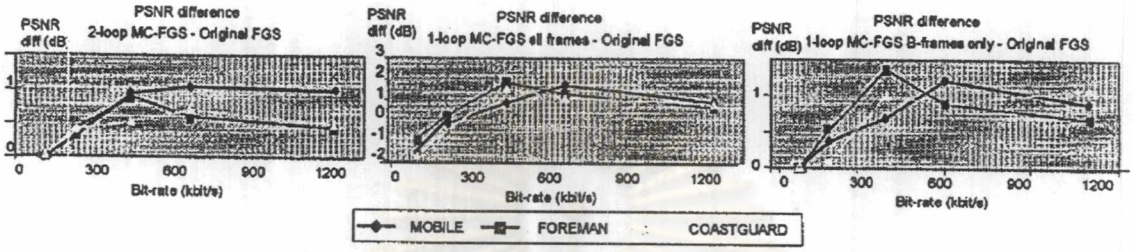


รูปที่ 2.23 โครงสร้างของ 1-loop MC-FGS

โครงสร้างของ single-loop MC-FGS แสดงดังรูปที่ 2.23 จะเห็นว่าได้นำบางระนาบบิตของชั้นเอนแฮนส์เมนท์ มาทำการชดเชยร่วมกับชั้นฐานนั้น ทำให้ความถูกต้องของการประมาณสูงขึ้น ส่งผลให้ประสิทธิภาพการเข้ารหัสสูงตาม แต่โครงสร้างนี้ต้องระมัดระวังอย่างยิ่งในการนำไปใช้เพราะว่า โครงสร้างนี้มีผลต่อการเกิดบิตผิดพลาดค่อนข้างสูง เนื่องจากถ้าเกิดความผิดพลาดบน

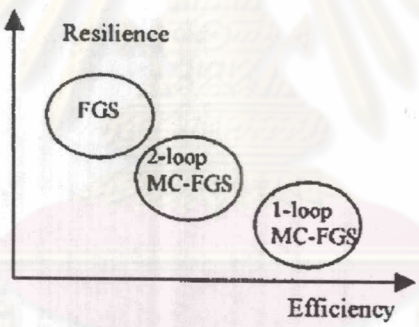
ชั้นเอนแฮนส์เมนท์แล้วจะแพร่ต่อไปยังชั้นฐานด้วย อย่างไรก็ดี ถ้าลองดูที่ตัวเข้ารหัสจะพบว่า มีการทำงานซับซ้อนน้อยกว่า 2-loop MC-FGS เพราะฉะนั้น อาจจะนำไปใช้กับ การส่งวิดีโอแบบเวลาจริงได้

ผลการจำลองระบบดังรูปที่ 2.24 แสดงให้เห็นว่า PSNR ของ 1-loop และ 2-loop ดีกว่าของ FGS ประมาณ 1-2 dB



รูปที่ 2.24 การเปรียบเทียบ PSNR ของทั้ง 3 แบบ [19]

เมื่อนำมาเปรียบเทียบจะพบว่าความทนทานต่อความผิดพลาดของ FGS สูงที่สุด รองลงมาคือ two loop และ single loop แต่ประสิทธิภาพการเข้ารหัสจะกลับกัน ดังรูปที่ 2.25



รูปที่ 2.25 เปรียบเทียบผลของการเข้ารหัสทั้ง 3 แบบ [18]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย