รายงานโครงการวิจัยประจำปีงบประมาณ 2545

เรื่อง

การแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด

Target Segmentation and Detection from Compressed Video

โดย

ดร. ธนารัตน์ ชลิดาพงศ์

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ 10520

พ.ศ. 2546

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานบริการและแผนพัฒนา
คณะเทคโนโลยีสารสนเทศ
วันที่ - 9 เม.ย. 2546
เวลา 16.30 น.

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง
ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

รายงานโครงการวิจัยประจำปีงบประมาณ 2545

เรื่อง

การแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด

Target Segmentation and Detection from Compressed Video

00995
b11397421
2122 1797

โดย

ดร. ธนารัตน์ ซลิดาพงศ์

วัน เดือน ปี 10 ต.ค. 2549
เลขทะเบียน H000995
เลขเรียกหนังสือ R TA 1637
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."
ธ 2475
ฉ. 2

คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ 10520

พ.ศ. 2546

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด

Target Segmentation and Detection from Compressed Video

ดร. ธนารัตน์ ชลิดาพงศ์

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เนื้อเรื่องย่อ

รายงานฉบับนี้กล่าวถึงผลการศึกษาและพัฒนาอัลกอริทึมเพื่อใช้ในการแยกแยะและตรวจจับการเคลื่อนไหวของเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด โดยใช้เทคนิคการลบฉากหลัง (Background Subtraction) และเทคนิค Quantization เพื่อใช้ในการสร้างรูปแบบจำลองของฉากหลัง (Background Model) อัลกอริทึมที่พัฒนานี้เรียกว่า Codebook-based Background Subtraction (CB-BGS) โดยหลักการในการสร้างแบบจำลองจะบันทึกและจัดหมวดหมู่ของค่าพิกเซลที่เหมือนกันตลอดระยะเวลาที่ทำการ model ฉากหลัง ซึ่งมีลักษณะคล้ายคลึงกับวิธีที่ใช้โดยทั่วไปในการบีบอัดวีดิทัศน์ ทำให้อัลกอริทึมนี้ใช้งานได้ดีทั้งกับวีดิทัศน์ที่มีความละเอียดสูง และวีดิทัศน์ที่ถูกบีบอัดและมีความละเอียดต่ำ นอกจากนี้โดยการใช้เทคนิคนี้ทำให้เราสามารถ model ฉากหลังจากค่าในอดีตที่ยาวนานได้ซึ่งจะทำให้แบบจำลองมีความถูกต้องแม่นยำมากขึ้น ทั้งยังสามารถ model การเคลื่อนไหวของฉากหลัง (dynamic background) ได้อีกด้วย

Abstract

This report describes a new algorithm for background modeling and background subtraction based on a quantization technique. Sample background values are quantized into an array of codebooks which represent a compressed background model. This allows us to capture structural background variation due to periodic-like motion over a long period of time under limited memory. We have applied the method to real image sequences of both indoor and outdoor scenes. It is shown that the method is suitable for both stationary and moving backgrounds. The method is also very robust with respect to video quality (both raw and variably compressed videos). It also allows the presence of moving foreground objects in the scene during the initial training period. That the background model can be initialized in unconstrained scenes is useful to many applications.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด

Target Segmentation and Detection from Compressed Video

ดร. ธนารัตน์ ชลิดาพงศ์

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เนื้อเรื่องย่อ

รายงานฉบับนี้กล่าวถึงผลการศึกษาและพัฒนาต่อยอดอัลกอริทึมเพื่อใช้ในการแยกแยะและตรวจจับการเคลื่อนไหวของเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด โดยใช้เทคนิคการลบฉากหลัง (Background Subtraction) และเทคนิค Quantization เพื่อใช้ในการสร้างรูปแบบจำลองของฉากหลัง (Background Model) อัลกอริทึมที่พัฒนานี้เรียกว่า Codebook-based Background Subtraction (CB-BGS) โดยหลักการในการสร้างแบบจำลองจะบันทึกและจัดหมวดหมู่ของค่าพิกเซลที่เหมือนกันตลอดระยะเวลาที่ทำการ model ฉากหลัง ซึ่งมีลักษณะคล้ายคลึงกับวิธีที่ใช้โดยทั่วไปในการบีบอัดวีดิทัศน์ ทำให้อัลกอริทึมนี้ใช้งานได้ดีทั้งกับวีดิทัศน์ที่มีความละเอียดสูง และวีดิทัศน์ที่ถูกบีบอัดและมีความละเอียดต่ำ นอกจากนี้โดยการใช้เทคนิคนี้ทำให้เราสามารถ model ฉากหลังจากค่าในอดีตที่ยาวนานได้ซึ่งจะทำให้แบบจำลองมีความถูกต้องแม่นยำมากขึ้น ทั้งยังสามารถ model การเคลื่อนไหวของฉากหลัง (dynamic background) ได้อีกด้วย

1. บทนำ

ความสามารถในการแยกแยะและตรวจจับการเคลื่อนไหวของเป้าหมายจากรูปภาพหรือวีดิทัศน์ เป็นงานพื้นฐานที่มีความสำคัญอย่างมากต่อเทคโนโลยีการมองเห็นภาพของคอมพิวเตอร์ และสามารถถูกนำมา ประยุกต์ใช้ได้ในหลายระบบ เช่น ระบบลาดตระเวนหรือระบบรักษาความปลอดภัยอัตโนมัติ (surveillance systems) [1, 2] ระบบตรวจสอบสภาพจราจร (traffic monitoring systems) [3, 4] ระบบตรวจจับและจดจำใบหน้าผู้คน (face detection and recognition systems) [5] ระบบคอมพิวเตอร์เพื่อช่วยเหลือคนพิการทางสายตา (machine vision for visual impairs) ระบบการประชุมเสมือน (virtual teleconference systems)เป็นต้น ประสิทธิภาพของการแยกแยะและตรวจจับซึ่งเป็นการประมวลผลขั้นต้น มีผลอย่างยิ่งต่อประสิทธิภาพการประมวลผลในขั้นต่อๆไปของระบบการมองเห็นภาพและการรับรู้ของคอมพิวเตอร์ ดังนั้นความสามารถในการแยกแยะและตรวจจับเป้าหมายที่สนใจอย่างอัตโนมัติ จึงเป็นองค์ความรู้ที่เป็นพื้นฐานและมีความสำคัญอย่างมากต่อการพัฒนาเทคโนโลยีการมองเห็นภาพของคอมพิวเตอร์

เทคนิคการลบฉากหลังเป็นเทคนิคที่ได้รับการยอมรับและใช้กันอย่างแพร่หลาย เพื่อใช้ในการแยกแยะและตรวจจับการเคลื่อนไหวของเป้าหมายจากรูปภาพหรือวีดิทัศน์ อย่างไรก็ตามถึงแม้ว่าเทคนิคนี้จะถูกพัฒนาและนำมาใช้ในการประมวลผลขั้นต้นในหลายระบบ แต่หลายอัลกอริทึมล้มเหลวเมื่อเจอสถานการณ์ที่มีการเปลี่ยนแปลงของแสง (ดูรูปที่ 1a) หลายอัลกอริทึมไม่มีประสิทธิภาพในการจัดการกับฉากหลังที่มีการเคลื่อนไหว (moving backgroud) (ดูรูปที่1b)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่น [6, 7, 8] เป็นต้น อีกข้อจำกัดหนึ่งที่เป็นปัญหาในหลายอัลกอริทึมกล่าวคือ อัลกอริทึมส่วนใหญ่ทำงานได้มีประสิทธิภาพเฉพาะกับวีดิทัศน์ที่มีความละเอียดสูงเท่านั้น ไม่สามารถทำงานได้อย่างมีประสิทธิภาพวีดิทัศน์ที่ถูกบีบอัด (ดูรูปที่1c) ในขณะที่เทคโนโลยีการสื่อสารได้เจริญรุดหน้าไปมาก โดยเฉพาะเทคโนโลยี Internet และ Multimedia ดังนั้นความต้องการอัลกอริทึมที่สามารถประมวลผลรูปภาพหรือวีดิทัศน์ที่ถูกบีบอัดจึงมีมากขึ้น และนี่คือเหตุผลที่ผู้วิจัยต้องการพัฒนาต่อยอดอัลกอริทึมที่เรียกว่า Codebook-based Background Subtraction (CB-BGS) ซึ่งอัลกอริทึมนี้แต่เดิมเริ่มพัฒนาโดยผู้วิจัยร่วมกับนักวิจัยที่ University of Maryland

(a) (b) (c)

รูปที่1 ตัวอย่างปัญหาและข้อจำกัดของอัลกอริทึมต่างๆที่ใช้เทคนิคการลบฉากหลัง (a)บางอัลกอริทึมไม่สามารถตรวจจับเป้าหมายได้อย่างมีประสิทธิภาพ เมื่ออยู่ภายใต้สภาวะที่แสงมีการเปลี่ยนแปลง ในตัวอย่างนี้คือเงาบนพื้น (b)บางอัลกอริทึมไม่สามารถจัดการกับฉากหลังที่มีการเคลื่อนไหว (moving backgroud) ในตัวอย่างนี้คือต้นไม้ที่เคลื่อนไหวไปมาเพราะแรงลม (c)หลายอัลกอริทึมไม่สามารถทำงานได้อย่างมีประสิทธิภาพกับวีดิทัศน์ที่ถูกบีบอัด

จากการทดลองพบว่าอัลกอริทึมนี้ใช้งานได้ดีทั้งกับวีดิทัศน์ที่มีความละเอียดสูงและวีดิทัศน์ที่ถูกบีบอัดซึ่งมีความละเอียดต่ำ นอกจากนี้ยังสามารถทำงานได้ดีในกรณีที่สภาพแสงมีการเปลี่ยนแปลง อีกทั้งด้วยเทคนิคนี้เราสามารถบันทึกค่าในอดีตที่ยาวนานของฉากหลัง โดยใช้หน่วยความจำที่จำกัดได้ ซึ่งหมายความว่าวิธีการนี้มีประสิทธิภาพในการตรวจจับเป้าหมายสูง ทั้งกรณีที่ฉากหลังไม่เคลื่อนที่ (stationary background) และในกรณีที่ฉากหลังเคลื่อนที่ (moving background) นอกจากนี้ผู้วิจัยได้พัฒนาโปรแกรมซึ่งสามารถประมวลผลอย่างมีประสิทธิภาพเพื่อให้เหมาะสมกับการนำไปประยุกต์ใช้ในระบบประมวลผลที่ตอบสนองทันที (real-time systems)

2. Literature Review

วิธีการหลักที่ใช้ในการแยกแยะและตรวจจับเป้าหมายมีหลายวิธี เช่น stereo-based approach ([11, 12, 13]) contour- or region-based approach ([14, 15, 16]) เป็นต้น อีกวิธีการหนึ่งที่เป็นที่ยอมรับและใช้กันอย่างแพร่หลายคือ background subtraction approach ซึ่งเป็นวิธีการที่เราจะใช้ในโครงการวิจัยนี้ โดยหลักการก็คือการนำรูปภาพที่เราสนใจมาหักลบออกจากรูปภาพที่ใช้เป็นตัวแทนของฉากหลัง (background scene model) ผลที่ได้ก็จะเป็นส่วนเค้าโครงของเป้าหมายที่เราสนใจ วิธีการนี้ยังแบ่งออกเป็นได้อีกหลายวิธี เช่น การแบ่งโดยรูปแบบของการคำนวณหาแบบเพื่อใช้เป็นตัวแทนของฉากหลัง (background scene modeling) สามารถแบ่งออกเป็น 2 รูปแบบหลักๆ คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) **Unimodal** รูปแบบนี้ใช้วิธีคำนวณหาค่ารูปแบบหนึ่งเป็นตัวแทนของพิกเซลนั้น ผลงานวิจัยที่ใช้วิธีการลักษณะนี้เป็นตัวแทนของฉากหลัง ได้แก่ งานวิจัยของ Long and Yang [15] และ Karmann and Brandt [16] ซึ่งเป็นกลุ่มแรกๆที่ใช้เทคนิค background subtraction ในการแยกแยะ และตรวจจับเป้าหมายจากรูปภาพ Long and Yang ใช้ค่าหนึ่งที่เลือกมาจากค่าต่างๆของพิกเซลในอดีต มาเป็นตัวแทนค่าของพิกเซลนั้น ขณะที่ Karmann and Brandt ใช้ค่าที่ได้มาจากการถัวเฉลี่ยตามเวลา (temporal smoothing) มาใช้เป็นตัวแทนค่าของพิกเซล ต่อมาได้มีการเริ่มใช้ข้อมูลรูปภาพ หรือวีดิทัศน์สี และปรับปรุงต่อยอดเทคนิคพื้นฐานเหล่านี้ไปในหลายๆวิธี เช่น ในงานของ Yamada et al [6] Wren et al [7] และในงาน วิจัย [17] ซึ่งเป็นผลงานของผู้วิจัยเอง เป็นต้น ทั้งสามกลุ่มนี้มีลักษณะที่คล้ายกันในแง่ที่ใช้หลักการทางสถิติ เพื่อหาค่าตัวแทนของแต่ละพิกเซลจากค่าของพิกเซลในอดีต แต่มีรายละเอียดทางเทคนิคที่แตกต่างกัน เช่น Yamada et al ใช้การเปลี่ยนค่าสี RGB เป็น YIQ เพื่อแยกคุณลักษณะสีออกจากคุณลักษณะของความสว่าง จากนั้นใช้ค่าเฉลี่ยของ YIQ เป็นตัวแทนของค่าของพิกเซล Wren et al ใช้หลักการสมมุติว่าแต่ละพิกเซลมีการกระจายแบบ single Gaussian และใช้ค่าตัวแปรของการกระจายของแต่ละพิกเซล เพื่อเป็นตัวแทนของพิกเซลนั้น ส่วนในงาน [17] นั้น มีการเสนอวิธีการแบบใหม่ในการแยกคุณลักษณะสีออกจากคุณลักษณะของความสว่าง โดยใช้หลักการ ทางสถิติ และไม่มีการตั้งข้อสมมุติที่ว่าแต่ละพิกเซลมีการกระจายพื้นฐานอย่างใดอย่างหนึ่ง ในงานวิจัยนี้ แต่ละพิกเซลสามารถที่จะมีการกระจายที่แตกต่างกันตามข้อมูลทางสถิติของค่าของพิกเซลในอดีต ยังมีผลงานวิจัยที่ใช้หลักการ unimodal และควรกล่าวถึงอีก เช่นผลงานของ Ridder et al [8] ซึ่งใช้เทคนิค Kalman filter ในการหาค่าตัวแทนของแต่ละพิกเซล วิธีการนี้ถึงแม้จะสามารถจัดการกับปัญหาของการเปลี่ยนแปลงของแสงได้ดีแต่การปรับตัวต่อการเปลี่ยนแปลงค่อนข้างช้า ถึงแม้ว่าการใช้รูปแบบ unimodal ในการแทนค่าฉากหลังนี้ จะทำงานได้ดีและมีประสิทธิภาพในหลายๆ ระบบ แต่ก็เฉพาะในกรณีที่มีฉากหลังที่ไม่เคลื่อนไหว (stationary background) เท่านั้น ในกรณีที่มีฉากหลังที่เคลื่อนไหว (moving background) เช่น ต้นไม้หรือพุ่มไม้ในฉากหลังที่เคลื่อนไหวไปมา วิธีการนี้จะให้ความผิดพลาด ในการตรวจจับค่อนข้างสูงในบริเวณที่มีการเคลื่อนไหว

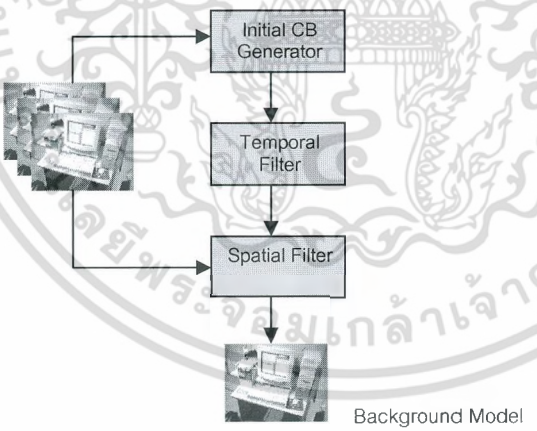
2) **Multimodal** เพื่อที่จะแก้ปัญหาในกรณีฉากหลังที่เคลื่อนไหว นักวิจัยหลายกลุ่มจึงพัฒนาวิธีการที่ใช้มากกว่าหนึ่งรูปแบบ (multimodal) ในการแทนฉากหลัง เช่น ในระบบ W4 [18] ใช้การกระจาย แบบ 2 รูปแบบ (bimodal distribution) เพื่อแทนค่าแต่ละพิกเซล วิธีนี้จะคำนวณการเปลี่ยนแปลงของฉากหลัง แล้วเก็บค่าสูงสุดและต่ำสุดของพิกเซลนั้น พร้อมทั้งค่าความคาดเคลื่อนเชิงเวลาสูงสุด (maximum temporal deviation) ในงานวิจัยของ Stauffer and Grimson [19] ได้ใช้หลักการความน่าจะเป็น ในการสร้างรูปแบบของแต่ละพิกเซลที่สามารถปรับตัวเองได้ (adaptive modeling) โดยสมมติการกระจายของแต่ละพิกเซลเป็นแบบ K Gussians ค่าคงที่และตัวแปรของแต่ละการกระจายจะคำนวณ ได้จากค่าของพิกเซลของฉากหลัง ในขณะตรวจจับถ้าค่าพิกเซลไม่อยู่ในการกระจายเหล่านี้ ก็จะถูกตีความว่าเป็นเป้าหมาย วิธีการนี้ใช้งานได้ดีโดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงของแสงอยู่ตลอดเวลา แต่จะล้มเหลวในการตรวจจับเป้าหมายที่มีการเคลื่อนตัวอย่างช้าๆ ปัญหานี้ถูกเรียกว่า *blending problem* ต่อมา Elgammal et al [20] ได้เสนอวิธีการสร้างรูปแบบ แบบไม่มีปัจจัยกำหนด (non-parametric model) วิธีการนี้ใช้ค่าของพิกเซลในอดีต มาใช้ในการสร้างรูปแบบของแต่ละพิกเซล จากนั้นก็จะประมาณค่าความน่าจะเป็นที่ค่าของพิกเซลที่สังเกตุได้จะเป็นฉากหลังหรือเป้าหมาย ในงานวิจัยนี้ได้มีการกล่าวถึงวิธีการรวมรูปแบบระยะสั้น (short-term model) และ รูปแบบระยะยาว (long-term model) เพื่อเพิ่มประสิทธิภาพของการตรวจจับด้วย สำหรับรายงานฉบับนี้ซึ่งเป็นการพัฒนาต่อยอดวิธีการใหม่ ที่เรียกว่า Codebook-based

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Background Subtraction (CB-BGS) มีหลักการคล้าย กับงานของ Elgammal et al ในแง่ที่ว่าเป็นแบบ ไม่มีปัจจัยกำหนด แต่หลักการ ในรายละเอียดของการสร้างรูปแบบของฉากหลังนั้นมีความแตกต่างกัน กล่าวคือในงานวิจัยชิ้นนี้ได้มีการนำเทคนิค Quantization มาใช้ในการหาค่าตัวแทนของฉากหลัง และด้วยเทคนิคนี้เราสามารถบันทึกค่าในอดีตที่ยาวนานของฉากหลัง โดยใช้หน่วยความจำที่จำกัดได้ ซึ่งหมายความว่าวิธีการนี้ จะมีประสิทธิภาพในการตรวจจับเป้าหมายสูง ทั้งกรณีที่ฉากหลังไม่เคลื่อนที่ (stationary background) และในกรณี ที่ฉากหลังเคลื่อนที่ (moving background) นอกจากนี้ CB-BGS ยังสามารถทำงานได้ดี กับวีดิทัศน์ที่มีคุณภาพที่หลากหลาย ไม่ว่าจะเป็นวีดิทัศน์ที่มีคุณภาพสูง เช่น สัญญาณวีดิทัศน์ความละเอียดสูงแบบ digital หรือ กับวีดิทัศน์ที่มีคุณภาพต่ำ เช่น สัญญาณวีดิทัศน์ที่ถูกบีบอัด (low-bandwidth compressed video) อีกด้วย

3. การสร้างรูปแบบฉากหลัง (Background Modeling)

เพื่อให้อัลกอริทึมสามารถทำงานได้ดีในสภาวะแสงที่เปลี่ยนแปลง และมีฉากหลังที่สามารถเคลื่อนไหวได้ เราจึงนำหลักการของ Vector Quantization (VQ) มาใช้ในการสร้างรูปแบบฉากหลัง โดยหลักการของ VQ เป็นการ mapping ค่าต่างๆในเซทที่ใหญ่ไปยังค่าในเซทที่เล็กกว่า [21] ใน CB-BGS นี้ เรานำเทคนิค VQ นี้มาใช้เพื่อเข้ารหัสค่าพิกเซลของฉากหลัง ดังนั้นลำดับค่าพิกเซลต่างๆตลอดช่วงเวลาที่เราใช้ในการเรียนรู้เพื่อที่จะสร้างรูปแบบฉากหลัง จะถูกนำมาเข้ารหัสเป็นค่า codebooks ต่างๆ ซึ่งโดยหลักการของ VQ ทำให้ขนาดเซทของ codebook ย่อมเล็กกว่าขนาดเซทของค่าของพิกเซลดั้งเดิมมาก ด้วยเหตุนี้ทำให้เราสามารถเข้ารหัสค่าของฉากหลังอันยาวนานภายใต้หน่วยความจำที่จำกัดได้ รายละเอียดและขั้นตอนในการสร้างรูปแบบฉากหลังนั้น มีดังนี้ (ดูรูปที่2)



รูปที่2 ขั้นตอนในการสร้างรูปแบบฉากหลังของ CB-BGS

I. Generating Initial Codebooks

ในขั้นตอนนี้เราจะทำการเข้ารหัสค่าพิกเซลของฉากหลังเบื้องต้น ให้ T_i แทน training set ของค่าของพิเซลที่ i ของฉากหลังซึ่งประกอบไปด้วย N ค่า เท่ากับจำนวนเฟรมที่เราใช้ในการเรียนรู้เพื่อสร้างรูปแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$T_i = \{x_{(i,1)}, x_{(i,2)}, x_{(i,3)}, \ldots, x_{(i,N)}\}$$

และให้ C_i แทน codebook ของพิเซลตำแหน่งที่ i ของฉากหลังซึ่งประกอบไปด้วยค่าต่างๆซึ่งเราจะเรียกว่า codeword จำนวน M ค่า โดย M เป็นเลขจำนวนเต็มเพื่อใช้แทนขนาดของ codebook

$$C_i = \{w_{(i,1)}, w_{(i,2)}, w_{(i,3)}, \ldots, w_{(i,M)}\}$$

ในระหว่างการเรียนรู้ฉากหลังเพื่อสร้างรูปแบบ ค่าของพิกเซลแต่ละตัว ($x_{(i,n)}$) จะถูกนำมาเปรียบเทียบกับค่าของ codeword ($w_{(i,m)}$)ใน codebook ของพิกเซลตำแหน่งนั้น เพื่อที่จะหาว่า codeword ใดมีค่าใกล้เคียงและสามารถใช้เป็นตัวแทนค่าของพิกเซลนั้นได้ ให้ Q แทน function ที่ใช้ในการหาว่า codeword ตัวใดสามารถเป็นตัวแทนค่าของพิกเซล $x_{(i,n)}$ได้

$$Q(x_{(i,n)}) = \hat{w}_{(i,m)}$$

โดย $\hat{w}_{(i,m)}$ คือ codeword แรกที่ซึ่งมีค่าความแตกต่างระหว่างตัวมันเองกับค่าของพิกเซลน้อยกว่าค่าคงที่ตัวหนึ่ง

$$d(X_{(i,n)}, \hat{w}_{(i,m)}) < \varepsilon_1$$

เมตริคที่ใช้วัดค่าความคลาดเคลื่อนหรือค่าความแตกต่างนั้น เราใช้การวัด square-error

ในกรณีที่ไม่มี codeword ที่สามารถเป็นตัวแทนค่าพิกเซลภายใต้เงื่อนไขดังกล่าว (ค่าความแตกต่างระหว่าง codeword กับค่าของพิกเซล มากกว่าค่าคงที่) codeword ตัวใหม่จะถูกสร้างขึ้นและถูกเพิ่มเข้าไปไว้ใน codebook ของพิกเซลนั้น รวมทั้งขนาดของ codebook จะถูกปรับค่า

ในกรณีที่พบ codeword ที่สามารถเป็นตัวแทนค่าพิกเซล ($\hat{w}_{(i,m)}$) ค่าของของ codeword จะถูกปรับโดยรวมข้อมูลของพิกเซลล่าสุดเข้าไปด้วย

II. Temporal Filtering

หลังจากขั้นตอนที่1 เราจะได้ codebook แรก ที่มาจากการเข้ารหัสค่าพิกเซลของฉากหลังที่ได้มาตลอดระยะเวลาที่ใช้ในการเรียนรู้ฉากหลัง ซึ่งอาจจะรวมรหัสของค่าที่เป็นค่าของเป้าหมายที่มีการเคลื่อนไหวในวีดิทัศน์ หรือค่าของสัญญาณรบกวน (noise) ด้วย ดังนั้นเราจึงต้องขจัด codeword ที่เป็นตัวแทนของค่าเหล่านี้ออกจาก codebook โดยการใช้ข้อมูลเชิงเวลา มาเพื่อคัดแยก codeword ดังกล่าวออกไป

จากการสังเกตพบว่า ค่าของพิกเซลของฉากหลังมีคุณลักษณะบางอย่างที่แตกต่างจากค่าของพิกเซลของเป้าหมายที่มีการเคลื่อนไหวในวีดิทัศน์ หรือค่าของสัญญาณรบกวน (noise) คุณลักษณะดังกล่าวคือการเกิดขึ้นหรือปรากฏซ้ำๆ (periodic-like occurance) ดังนั้นเราจึงกำหนดค่าที่ได้จากการคำนวณเชิงเวลาจากค่าของพิกเซล โดยให้ชื่อว่า maximum negative run-length หรือ λ ซึ่งคือค่าระยะเวลานานที่สุดที่ codeword นั้นไม่ได้ถูกปรับค่า หลังจากนั้นเรา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

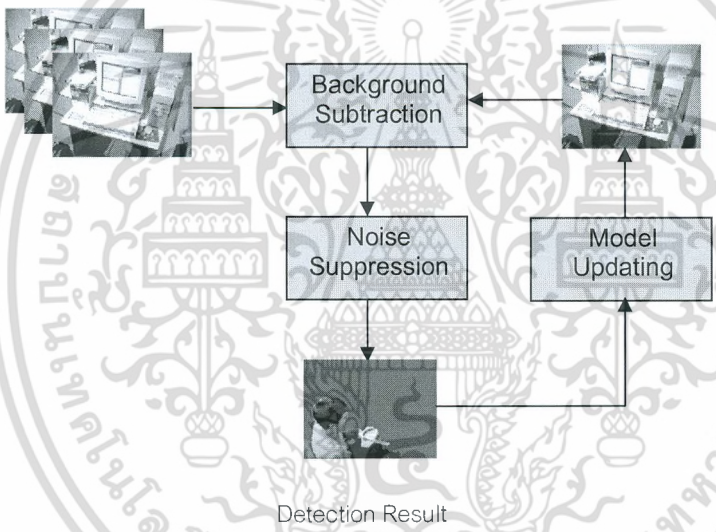
จึงทำการคัดแยก codeword ของเป้าหมายที่มีการเคลื่อนไหวในวีดิทัศน์ และ codeword ของสัญญาณรบกวนออกจาก codebook ผลลัพธ์ที่ได้ก็จะเป็น codebook ที่ประกอบไปด้วย codeword ของค่าของฉากหลังเท่านั้น

III. Spatial Filtering

วัตถุประสงค์ของขั้นตอนนี้เพื่อเป็นการปรับปรุง codebook ที่ได้จากขั้นที่แล้ว ซึ่งอาจจะคัดแยก codeword ที่เป็นของฉากหลังบางชนิดออกไป เนื่องจากในวีดิทัศน์ที่ถูกบีบอัดนั้น อาจมีเหตุการณ์ที่ค่าของฉากหลังเปลี่ยนแปลงบางชั่วขณะเนื่องจากวัตถุที่เคลื่อนที่ ดังนั้นในขั้นตอนนี้เราจะนำค่าเหล่านั้นมาเข้ารหัสอีกครั้งโดยใช้ข้อมูลเชิงพื้นที่ (spatial information) เพื่อเป็นเงื่อนไขในการเข้ารหัส

4. การลบฉากหลัง (Background Subtraction)

ขั้นตอนในการลบฉากหลังและปรับค่ารูปแบบฉากหลัง (ดูรูปที่3)

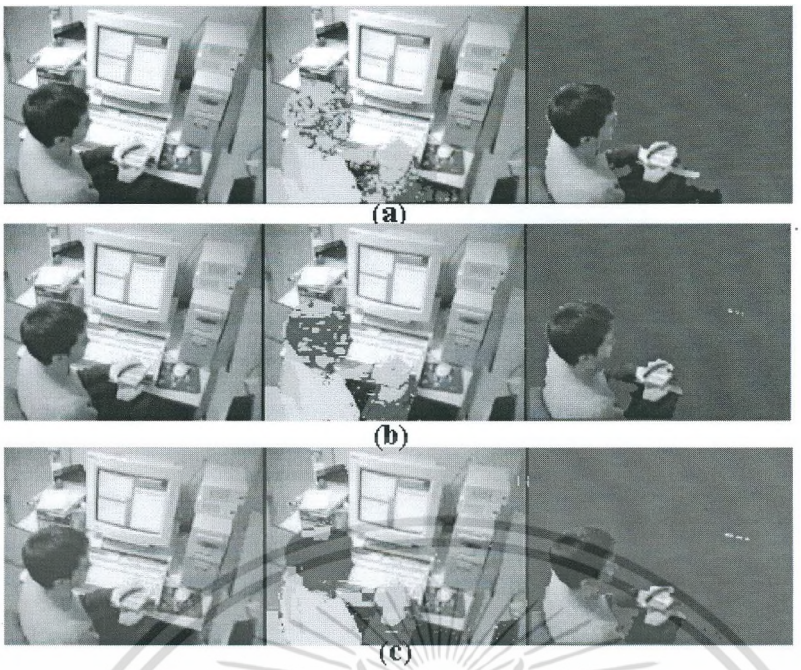


รูปที่3 ขั้นตอนในการลบฉากหลังและปรับค่ารูปแบบฉากหลัง

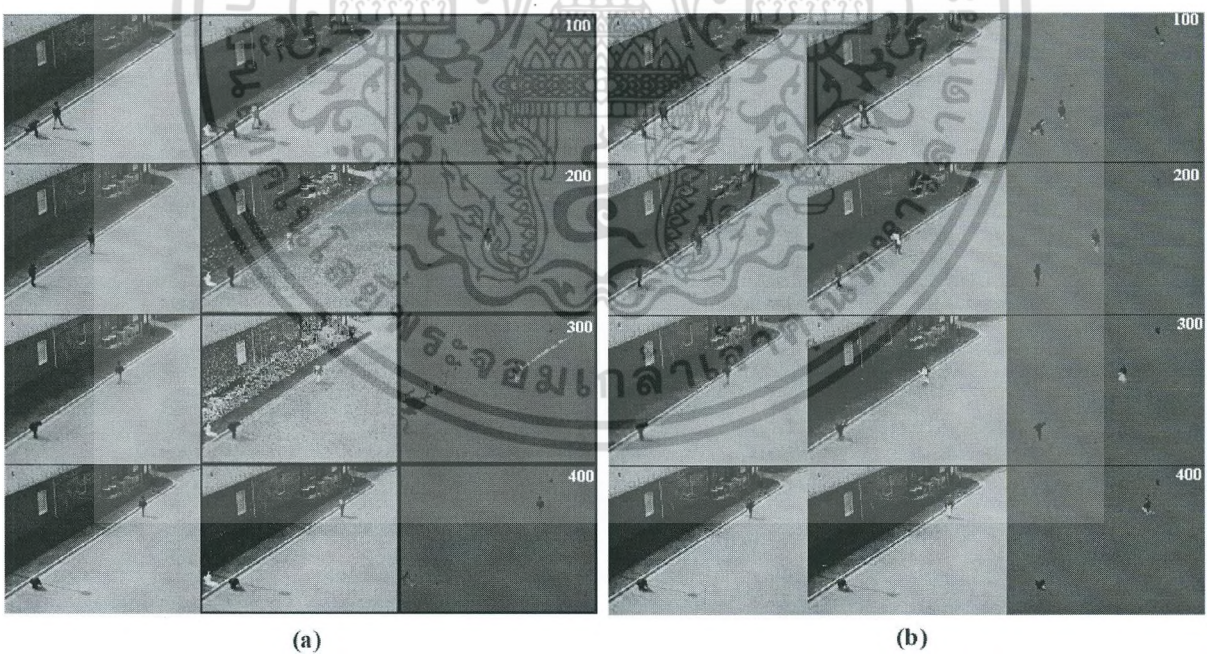
5. ผลการทดลอง

รูปที่4 และ 5 เป็นภาพแสดงผลจากการแยกแยะและตรวจจับเป้าหมายโดยใช้อัลกอริทึมที่ได้พัฒนาดังกล่าว ในรูปที่4 เป็นภาพจากวิดีทัศน์ที่จับภาพชายคนหนึ่งทำงานในห้องทำงานที่มีแสงสว่างค่อนข้างคงที่จากหลอดไฟฟลูออเรสเซนท์ ภาพตามแนวนอน (a,b,c) เป็นภาพในแต่ละเฟรมในวิดีทัศน์ ภาพในคอลัมน์ทางซ้ายคือภาพต้นฉบับของแต่ละเฟรม ภาพในคอลัมน์กลางคือผลจากการแยกแยะพิกเซลว่าแต่ละพิกเซลถูกจัดไว้ในประเภทใด ส่วนภาพในคอลัมน์ขวาคือผลลัพธ์ที่ได้จากการแยกแยะเป้าหมายออกจากสิ่งแวดล้อม ในรูปที่5 เป็นภาพจากวีดิทัศน์ที่จับภาพกลางแจ้ง มีสภาพแสงที่แปรเปลี่ยนอยู่ตลอดเวลา เนื่องจากแสงพระอาทิตย์และการเคลื่อนตัวของเมฆบนท้องฟ้า ทั้งสองวีดิทัศน์เป็นวิดีทัศน์ที่ถูกบีบอัดในรูปแบบ MPEG-1 ที่ขนาด 100 Mbit/s

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4 ผลการแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด สภาพแสงค่อนข้างคงที่ภายในห้อง



รูปที่5 ผลการแยกแยะและตรวจจับเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด สภาพแสงค่อนข้างแปรเปลี่ยนภายนอกอาคาร เนื่องจากแสงพระอาทิตย์และเมฆที่เคลื่อนตัวไปมาบนท้องฟ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. สรุป

รายงานฉบับนี้กล่าวถึงการพัฒนาต่อยอดอัลกอริทึมCB-BGSเพื่อใช้ในการแยกแยะและตรวจจับการเคลื่อนไหวของเป้าหมายจากวีดิทัศน์ที่ถูกบีบอัด โดยใช้เทคนิคการลบฉากหลัง (Background Subtraction) และเทคนิค Quantization เพื่อใช้ในการสร้างรูปแบบจำลองของฉากหลัง (Background Model) อัลกอริทึมนี้ใช้งานได้ดีทั้งกับวีดิทัศน์ที่มีความละเอียดสูง และวีดิทัศน์ที่ถูกบีบอัดและมีความละเอียดต่ำ นอกจากนี้เราสามารถ model ฉากหลังจากค่าในอดีตที่ยาวนานได้ซึ่งจะทำให้แบบจำลองมีความถูกต้องแม่นยำมากขึ้น ทั้งยังสามารถ model การเคลื่อนไหวของฉากหลัง (dynamic background) ได้อีกด้วย

การพัฒนาอัลกอริทึมนี้ยังไม่สมบูรณ์นักเนื่องจากเวลาอันจำกัดของผู้วิจัย ผู้วิจัยมีแนวคิดที่จะพัฒนาต่อโดยเน้นขั้นตอนหลังการประมวลผลเพื่อปรับปรุงประสิทธิภาพ เพิ่มความถูกต้องแม่นยำในการตรวจจับและแยกแยะเป้าหมายต่อไป

เอกสารอ้างอิง (References)

[1] Pece A E C. People Tracking in Surveillance Applications. *Proc. 2nd IEEE Int'l Workshop OnPerformance Evaluation of Tracking and Surveillance (PETS 2001)* 2001.

[2] Greiffenhagen M, Ramesh V, Niemann H. The Systematic Design and Analysis Cycle of a Vision System: A Case Study in Video Surveillance. *Proc. IEEE Conf. On Computer Vision and Pattern Recognition* 2001, Vol2:704-711.

[3] Beymer D, McLauchlan P, Coifman B, Malik J. A Real-time Computer Vision System for Measuring Traffic Parameters. *Proc. IEEE Conf. On Computer Vision and Pattern Recognition* 1997.

[4] Setchell C. Applications of Computer Vision to Road-traffic Monitoring. *Ph.D. Thesis, Dept. of Computer Science, University of Bristol (CS-EXT-1997-118)* 1997.

[5] Zhou S, Krueger V, Chellappa R. Face Recognition from Video: A CONDENSATION Approach. *Proc. Int'l Conf. On Automatic Face and Gesture Recognition (FG02)* 2002.

[6] Yamada M, Ebihara K, Ohya J. A New Robust Realtime Method for Extracting Human Silhouettes from Color Images. *Proc.IEEE Int'l Conf. Automatic Face and Gesture Recognition* 1998.

[7] Wren C, Azarbayejani A, Darrell T, Pentland A. Pfinder: Real-time Tracking of Human Body. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 1997, 19(7):780-785.

[8] Ridder C, Munkelt O, Kirchner H. Adaptive Background Estimation and Foreground Detection Using Kalman Filtering. *Proc. Int'l Conf. On Recent Advances in Mechatronics* 1995.

[9] Zhu S, Yuille A. Region Competition: Unifying snakes, region growing, and bayer/mdl for multiband image segmentation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 1996, 18(9):884-900.

[10] Eveland C, Konolige K, Bolles R C. Background Modeling for Segmentation of Video-Rate Stereo Sequences. *Proc Computer Vision and Pattern Recognition* 1998.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

[11] Yezzi A J, Soatto S. Stereoscopic Segmentation. *Proc.IEEE Int'l Conf. On Computer Vision* 2001, Vol1:59-66.

[12] Kass K, Witkin A, Terzopoulos D. Snakes: Active Contour Models. *Int'l Journal of Computer Vision* 1987, 1:321-331.

[13] Boykov Y Y, Jolly M. Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in n-d Images. *Proc.IEEE Int'l Conf. On Computer Vision* 2001, Vol1:105-112.

[14] Tan K. Ahuja N. Selecting Objects with Freehand Sketches. *Proc.IEEE Int'l Conf. On Computer Vision* 2001, Vol1:337-344.

[15] Long W, Yang Y. Stationary Background Generation: An Alternative to the Difference of Two Images. *Patern Recognition* 1990, 23(12):1351-1359.

[16] Karmann K, Brandt A. Moving Object Recognition Using an Adaptive Backgrond Memory. *Time-Varying Image Processing and Moving Object Recognition,* Elsevier Science Publishers 1990, 289-296.

[17] Horprasert T, Harwood D, Davis L S. A Statistical Approach for Realtime Robust Background Subtraction and Shadow Detection. *Proc.IEEE Int'l Conf. On Computer Vision's Frame Rate Workshop* 1999.

[18] Haritaoglu I, Harwood D, Davis L S. W4: Who? When? Where? What? A Real-time System for Detecting and Tracking People. *Proc. Int'l Conf. On Face and Gesture Recognition* 1998.

[19] Stauffer C, Grimson W. Adaptive Background Mixture Models for Real-time Tracking. *Proc.Computer Vision and Pattern Recognition* 1999, 246-252.

[20] Elgammal A, Harwood D, Davis L S. Non-parametric Model for Background Subtraction. *Proc. European Conf. On Computer Vision* 2000.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้