รายงานวิจัยฉบับสมบูรณ์ ประจำปีงบประมาณ 2546
โครงการวิจัยย่อยลำดับที่ 7 เรื่อง
การศึกษาและพัฒนากระบวนการเข้าและถอดรหัสสัญญาณวีดิทัศน์ตามมาตรฐาน MPEG-4 เพื่อส่งบนระบบเครือข่ายสื่อสารไร้สาย

1. ผู้รับผิดชอบโครงการ อ.ดร.สุภาวดี อร่ามวิทย์

2. วัตถุประสงค์ของโครงการ

2.1 เพื่อสร้างองค์ความรู้ทางกระบวนการเข้า และถอดรหัสสัญญาณวีดิทัศน์ตามมาตรฐาน MPEG-4 กรรมวิธีการเข้ารหัสสัญญาณวีดิทัศน์แยกส่วน และ การประกอบสัญญาณวีดิทัศน์ แยกส่วน เข้ากับฉากหลังที่เป็นภาพสังเคราะห์ การจัดสรรบิตให้กับวัตถุที่สนใจ รวมถึงคุณ ลักษณะเพิ่มเติมในมาตรฐาน MPEG-4 ว่าด้วยความทนทานต่อความผิดพลาดของ สัญญาณวีดิทัศน์ ในการส่งผ่านเครือข่ายไร้สาย

2.2 เพื่อนำองค์ความรู้ที่ได้มาใช้การสร้างจริงระบบการประชุมสัมมนาทางวีดิทัศน์ ในการส่งผ่านเครือข่ายท้องถิ่น และ ไร้สาย ในส่วนของซอฟต์แวร์

3. ขอบเขตหรือเป้าหมายของโครงการ

3.1 การศึกษาและพัฒนากระบวนการลงรหัสสัญญาณวีดีทัศน์อัตราบิตต่ำตาม มาตรฐาน MPEG-4 โดยจะเน้นที่กรรมวิธีการเข้ารหัสสัญญาณวีดิทัศน์แยกส่วน และ การประกอบสัญญาณ วีดิทัศน์แยกส่วนเข้ากับ ฉากหลังที่เป็นภาพสังเคราะห์

3.2 การนำขั้นตอนวิธีของการเข้ารหัสที่พัฒนาได้ มาประยุกต์ควบคู่กับ วิธีแก้ไขและปกปิดความผิดพลาดเพื่อเพิ่มคุณภาพของสัญญาณวีดีทัศน์ เพื่อทำการส่งบนโครงข่ายสื่อสารไร้สาย ทั้งในแบบจำลองและสถานการณ์จริง

3.3 การสร้างจริงระบบการประชุมสัมมนาทางวีดิทัศน์ ในรูปแบบของซอฟต์แวร์

3.4 การทดสอบและวิเคราะห์ประสิทธิภาพขั้นตอนวิธีที่นำเสนอ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนงานย่อยที่ 1
การรับ-ส่งข้อมูลบนโครงข่าย และ ส่วนเฉพาะสำหรับการรับส่งวีดิทัศน์

ส่วนงานที่ได้ดำเนินการไปแล้ว

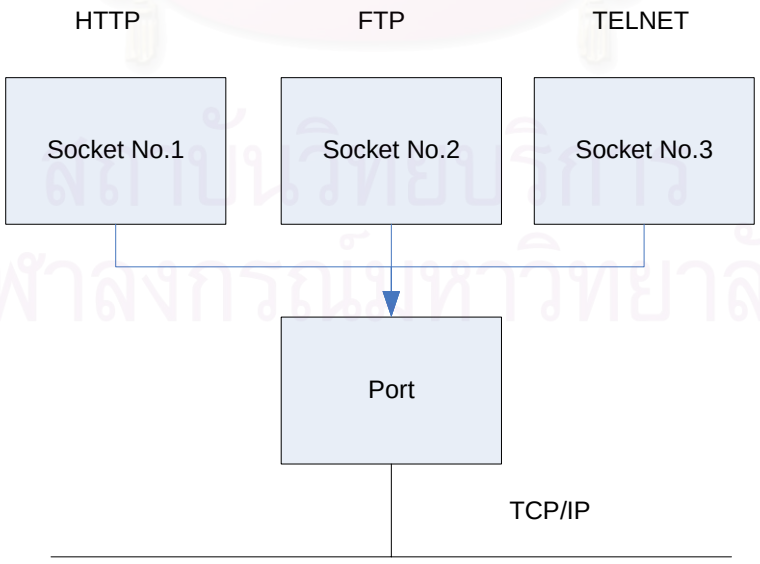
ส่วนงานที่รับผิดชอบในการรับ-ส่งวีดิทัศน์ผ่านโครงข่าย ประกอบด้วยส่วนย่อยหลัก 2 ส่วนได้แก่ ส่วนพื้นฐานการรับ-ส่งข้อมูลบนโครงข่าย และ ส่วนเฉพาะสำหรับการรับส่งวีดิทัศน์ ซึ่งมีรายละเอียดของส่วนย่อยต่างๆ ดังนี้

1. ส่วนพื้นฐานการรับ-ส่งข้อมูลบนโครงข่าย

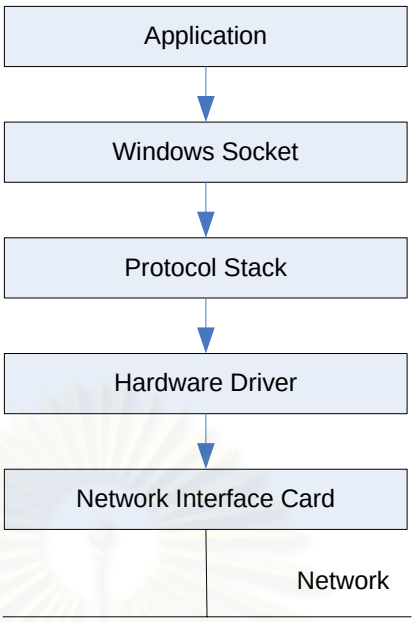
การรับส่งข้อมูลบนเครือข่ายของโปรแกรมใช้การติดต่อผ่าน Application Programming Interface สำหรับ Network บนระบบปฏิบัติการ Windows ชื่อว่า Windows Sockets หรือ Winsock ซึ่งใช้แบบจำลอง “Sockets” สำหรับการสื่อสารข้อมูลบนโครงข่ายที่ได้รับความนิยมในการใช้งาน ซึ่งคิดค้นโดย Berkeley Software Distribution (BSD) UNIX. โดยได้รับการปรับแต่งให้สามารถใช้งานกับ Microsoft Windows และให้ชื่อว่า Windows Sockets 1.1 ในเริ่มแรก จากนั้นมีการพัฒนาต่อมา อีกหลายรุ่น ซึ่งในโครงการนี้ใช้ Windows Socket API version 2.0

Windows Sockets 2 นั้นไม่ใช่โปรโตคอลสำหรับการสื่อสาร แต่เป็นเพียงแค่ interface สำหรับโปรแกรม เพื่อใช้ในการใช้งานการสื่อสารด้วยโปรโตคอลชั้น transport ที่มีใช้กันหลากหลาย และเนื่องจาก Windows Socket ไม่ใช่โปรโตคอลจึงสามารถใช้สื่อสารกันกับระบบปฏิบัตการอื่นทีไม่ใช่ Windows

การใช้งาน Socket ของโปรแกรมต่างๆ นั้นทำงานแยกกันโดยอิสระ Winsock สามารถที่จะสร้างช่องทางสื่อสารขึ้นมาได้หลายๆช่องทาง (ขึ้นอยู่กับเวอร์ชันของ Winsock) และแต่ละช่องทางสื่อสารสามารถที่จะส่งข้อมูลได้โดยไม่ขึ้นกับช่องทางสื่อสารอื่นๆ ดังรูป



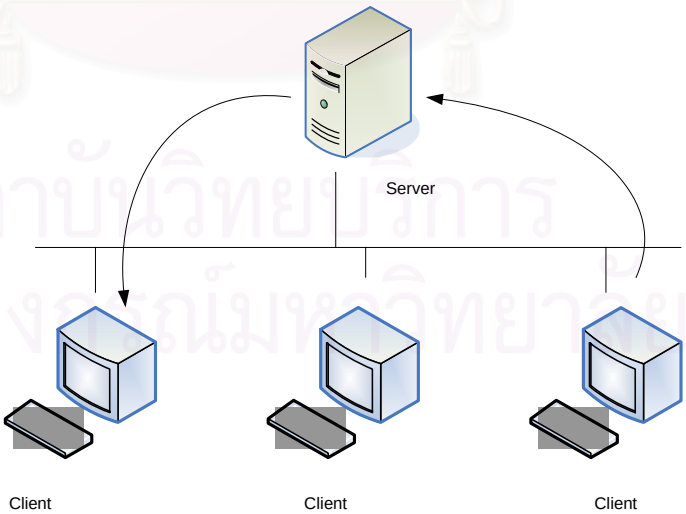
Winsock ทำหน้าที่เป็นตัวกลางในการรับส่งข้อมูลจาก Application กับ TCP/IP จากนั้น TCP/IP จึงส่งข้อมูลลงไปบนระบบอินเตอร์เน็ต ดังรูป



การใช้งาน Windows socket network API นั้น มีรายละเอียดทั้งในด้านการออกแบบ และการโปรแกม ซึ่งการออกแบบแบบจำลองการสื่อสารข้อมูล นั้นทำได้หลายรูปแบบ ตัวอย่างเช่น แบบจำลอง Client/server และแบบจำลอง Peer-To-Peer ซึ่งมีข้อแตกต่างทางโครงสร้างและลักษณะการรับ-ส่งข้อมูลบนโครงข่ายดังนี้

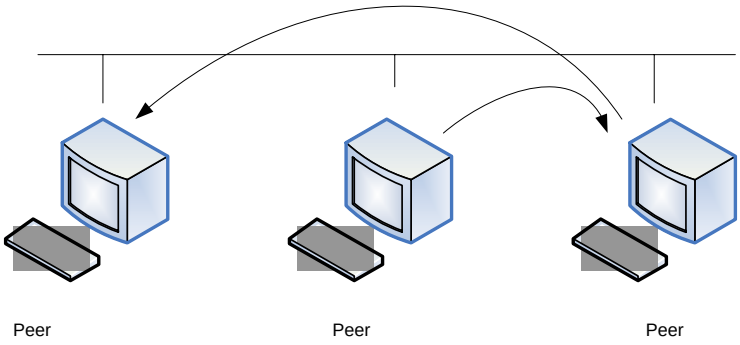
Client/Server

การสื่อสารข้อมูลด้วยแบบจำลอง Client/Server เป็นแบบจำลองที่ได้รับการใช้งานทั่วไป เช่นบริการต่าง ๆ บน internet โดยมีผู้ให้บิรการขอมูลเรียกว่า Server และผู้รับบริการข้อมูลเรียกว่า Client ถ้ามีการสื่อสารระหว่างแต่ละ Client จะต้องผ่านการควบคุมโดย Server ดังรูป



Peer-to-peer

การสื่อสารแบบ peer-to-peer แต่ละ peer สามารถติดต่อรับ-ส่งข้อมูลกันได้โดยไม่ต้องผ่านตัวกลาง และแต่ละ peer มีการควบคุมการรับ-ส่งข้อมูลของตนเอง ดังรูป



การติดต่อสื่อสารข้อมูลระหว่างโปรแกรมมีการใช้งานโปรโตคอล ต่อไปนี้ ได้แก่ส่วน Network Protocol ใช้ Internet Protocol (IP) และส่วนของ Transport Protocol.ใช้ User Datagram Protocol (UDP) และ Transmission Control Protocol (TCP) โดยมีรายละเอียดดังนี้

การใช้งานโปรโตคอล TCP/IP ใช้แบบจำลอง Client/Server โดยมีโปรแกรมที่ทำหน้าที่เป็น Server ควบคุมการทำงานทั้งหมด เริ่มต้นโครงข่ายการสื่อสารด้วย Client ทั้งหมดติดต่อไปยัง Server และร้องขอการรับ-ส่งข้อมูลไประหว่าง client อื่น โดยการส่งข้อมูลไปยัง server เพื่อให้ Server ทำการส่งต่อข้อมูลไปให้ โครงสร้างของระบบจะเก็บไว้ใน server

การใช้งานโปรโตคอล UDP/.IP ใช้แบบจำลอง peer-to-peer โดยทุก peer มีอิสระที่จะทำการรับ-ส่งข้อมูลกับ peer อื่น ๆ ได้ตามต้องการ ขีดจำกัดของการรับ-ส่งข้อมูลขึ้นกับประสิทธิภาพของช่องสื่อสารระหว่างแต่ละ peer

2. ส่วนเฉพาะสำหรับการรับส่งวีดิทัศน์

ส่วนเฉพาะสำรับการรับ-ส่งวีดิทัศน์บนโครงข่ายเป็นการเพิ่มเติมส่วนของการรับ-ส่งข้อมูลให้เหมาะสมกับการใช้งานเฉพาะข้อมูลวีดิทัศน์ โดยแยกได้เป็นส่วนย่อย 2 ส่วนคือ การทำงานกับข้อมูลวีดิทัศน์ และการสื่อสารข้อมูลวีดิทัศน์

การทำงานกับข้อมูลวีดิทัศน์ของโปรแกรมใช้ DirctShow API เดิมเริ่มต้นเรียกว่า Active movie เป็นระบบการทำงานกับสัญญาณวิดีโอของ Microsoft สร้างขึ้นมาเพื่อแก้ไขข้อบกพร่องของระบบเก่าคือ Video for Windows (VFW) ซึ่งเป็น API เดิมที่เคยใช้งานอยู่ และเพื่อให้รองรับการทำงานได้กับคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการต่าง ๆ กัน โดยมีการใช้ Component Object Module (COM) มาเป็นฐานรากในการพัฒนาโปรแกรม ซึ่งทำให้โปรแกรมสามารถเข้าหา Object Linking and Embedding (OLE) ซึ่งสามารถเรียกหา Object ตัวที่เป็นส่วนควบคุมต่าง ๆ เช่น ส่วนควบคุมการขยายข้อมูล ส่วนควบคุมระดับเสียง และส่วนอื่น ๆ มาใช้งานได้ทันที และเนื่องจาก Active Movie มีโครงสร้างบนสถาปัตยกรรมแบบ COM ทำให้ระบบนี้มีข้อได้เปรียบระบบอื่น ที่เป็นระบบที่ทำงานได้บนระบบปฏิบัติการหลายตัวของ Microsoft และสนับสนุนกรทำงานต่อระบบการเข้ารหัสอื่น ๆ ด้วยเช่น MPEG, ระบบเสียง WAV, ทำให้ตัว ActiveMovie ง่ายสำหรับการนำไปประยุกต์ใช้บนระบบอินเทอร์เน็ต และยิ่งไปกว่านั้น ระบบ ActiveMovie ยังรวมอยู่ในโปรแกรม Direct X ซึ่งทำให้สามารถใช้ความได้เปรียบของตัวโปรแกรม สั่งการใช้งานในส่วนของการเร่งภาพและเสียงของการ์ดแสดงผลต่าง ๆ เช่นสามารถใช้เล่นไฟล์ภาพยนตร์แบบ AVI โดยใช้ส่วนของ Direct Draw และ Direct X component มาช่วยในการทำงานของการ์ดแสดงผลแบบต่าง ๆ

การพัฒนาโปรแกรมด้วย DirectShow แบ่งเป็น 2 ประเภทหลักได้แก่การพัฒนาโปรแกรมประยุกต์ และการพัฒนาระบบย่อยการเข้ารหัส

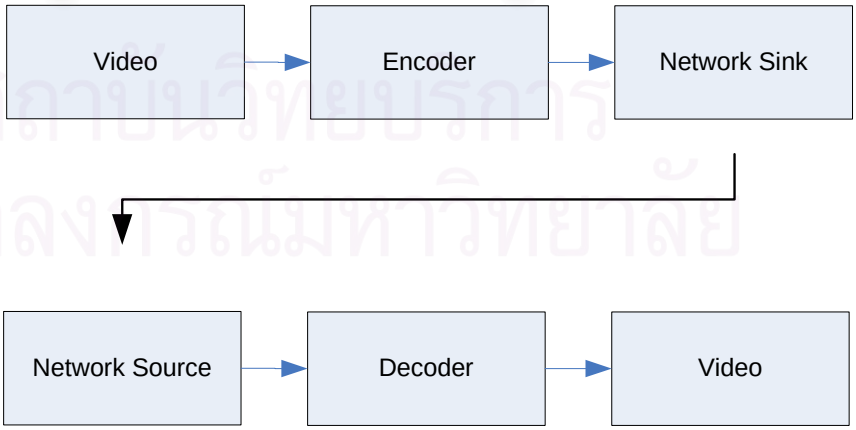
DirectShow มีโครงสร้างหลักย่อยที่สุดเรียกว่า Filter การทำงานกับข้อมูลบน DirectShow จึงเป็นการทำงานได้ทั้งการนำ filter มาใช้งาน และการสร้าง filter สำหรับนำไปใช้งาน สำหรับประเภทของ Filter แยกได้เป็น 3 ประเภทได้แก่

1. Source Filter : ทำหน้าที่สร้างกระแสข้อมูลสำหรับการใช้งาน ซึ่งอาจนำข้อมูลมาจาก capture card, กล้อง หรือช่องสื่อสาร
2. Transform Filter : ทำหน้าที่ประมวลผลข้อมูล เช่นการเข้า/ถอดรหัส
3. Render Filter: ทำหน้าที่แสดงผล/ส่งออกข้อมูลที่ผ่านการประมวลผลแล้วเช่น แสดงผลออกทางจอภาพ หรือส่งออกทางช่องสื่อสาร

การพัฒนาโปรแกรมประยุกต์ทำโดยการเลือก Filter ที่เหมาะสมมาประกอบกันเป็น Filter Graph เพื่อให้โปรแกรมทำงานกับข้อมูล

การพัฒนาเริ่มต้นด้วยการสร้าง Filter สำหรับรับ-ส่งข้อมูลบนโครงข่ายเพื่อนำไปใช้งานในโปรแกรมประยุกต์ โดย filter ที่พัฒนาขึ้นมี 2 แบบได้แก่

1. Network Sink Filter : ทำหน้าที่รับข้อมูลที่ผ่านการเข้ารหัสแล้วส่งออกทางช่องสื่อสาร
2. Network Source Filter: ทำหน้าที่รับข้อมูลจากช่องสื่อสารแล้วส่งข้อมูลนั้นให้ตัวถอดรหัส



การรับ-ส่งวีดิทัศน์จำเป็นต้องมีข้อมูลของ Session/Media Description เพื่อการแสดงผลที่ถูกต้องสำหรับภาคแสดงผล ข้อมูลที่ผ่านการสื่อสารด้วยวิดิทัศน์จำเป็นต้องรู้ลักษณะของข้อมูลวิดิทัศน์ที่รับว่ามี parameter ต่าง ๆ เป็นอย่างไร เช่น ขนาดของภาพ, จำนวนภาพต่อเวลา และชนิดของการเข้ารหัส เป็นต้น

เนื่องจากพบว่า Media description ที่ใช้ในการเชื่อมต่อระหว่างแต่ละ filter มีความเหมาะสมและสามารถนำไปใช้งานได้ จึงเลือกใช้วิธีส่ง structure AM_MEDIA_TYPE ไปเป็น session/media description ซึ่ง MediaType ประกอบด้วยส่วนสำคัญfดังนี้

1. Major Type : ประเภทหลักของข้อมูล เช่น Video, Audio หรือ Text
2. Minor Type: ปรเภทย่อยของข้อมูล เช่น MPEG-Video, DV-Video ในกรณีของ video
3. Format Type/Format Block : คำอธิบายข้อมูลเช่น ขนาดภาพ จำนวนเพรมต่อเวลา และจำนวนบิตต่อเฟรม เป็นต้น

โปรโตคอลที่ใช้ในการรับ-ส่งวีดิทัศน์ระหว่าโปรแกรมคือ Real-Time Transport Protocol (RTP) ซึ่งออกแบบมาเพื่อใช้งานสำหรับการรับส่งข้อมูลผ่านโครงข่าย ที่เหมาะสำหรับโปรแกรมประยุกต์ที่ต้องการรับ-ส่งข้อมูล ณ เวลาจริง ตัวอย่างเช่นข้อมูล audio, video หรือ simulation data บนโครงข่าย multicast หรือ Unicast.

โปรโคคอล RTP ไม่ได้กำหนดเรื่องการจองใช้ทรัพยากรโครงข่าย และไม่มีการรับประกัน quality-of-service สำหรับงานเวลาจริง การรับ-ส่งข้อมูลมีการควบคุมด้วยโปรโตคอลควบคุม RTCP เพื่อให้สามารถเฝ้าดูและตรวจสอบการรับ-ส่งของข้อมูลในระบบโครงข่าย โปรโตคอล RTP และ RTCP ถูกออกแบบมาให้อิสระต่อการใช้ โปรโตคอลชั้นTransport และ Network layer และยังรองรับ translators and mixers ระดับ RTP-layer

การประยุกต์ใช้บางส่วนของโปรโตคอล RTP เช่นการใช้ Timestamp และ Sequence Number มาช่วยในการจัดการกับการรับ-ส่งวีดิทัศน์ โดยการใช้งานจะอยู่ในส่วนของการรับ-ส่งข้อมูลของ Network Filter โดยมีการรับ-ส่งข้อมูลเป็น Packet และการใช้ RTP header สำหรับข้อมูลแต่ละ packet

3. ส่วนเฉพาะสำรับการรับส่งเสียง

การใช้งานการประชุมทางวีดิทัศน์จำเป็นต้องมีการรับ-ส่งเสียงด้วย จึงได้เริ่มต้นพัฒนาการรับ-ส่งเสียงบนโครงข่าย โดยมีการทดลองไปแล้วบางส่วนแต่ยังมีปัญหาเรื่องการเข้าจังหวะกันระหว่างภาพและเสียง

ส่วนงานที่จะดำเนินการต่อไป

1. ส่วนพื้นฐานการรับ-ส่งข้อมูลบนโครงข่าย

การใช้งาน Windows socket network API

ปรับปรุงการทำงานและการใช้งานทรัพยากรของแต่ละ thread การรับ-ส่งข้อมูลภายใน filter ให้มีประสิทธิภาพดีขึ้น และให้เหมาะสมกับการรับ-ส่งข้อมูลแบบหลายจุด

แบบจำลองการสื่อสารข้อมูล

ปรับปรุงโครงสร้างของการสื่อสารให้มีการสื่อสารแบบหลายจุดเกิดขึ้นได้ และมีการควบคุมโครงสร้างของโครงข่าย

2. ส่วนเฉพาะสำรับการรับส่งวีดิทัศน์

การใช้งาน DirectShow API

เปลี่ยนแปลงตัวหรือมีการปรับแต่งตัวเข้า/ถอดหรัสได้ในขณะ Runtime กรณีที่มีการเปลี่ยนแปลงสภาพของโครงข่าย

การรับ-ส่งวีดิทัศน์

จัดการเรื่องการตรวจสอบความผิดพลาดที่เกิดขึ้นจากการสื่อสาร เช่น packet error/loss เพื่อให้ตัวถอดรหัสสามารถดำเนินการแก้ไขหรือปกปิดความผิดพลาดได้

การใช้งานโปรโตคอล RTP

ปรับเข้าสู่การใช้งานโปรโตคอล RTP มากขึ้นโดยเพิ่มการใช้งานในส่วนของ RTCP

3. ส่วนเฉพาะสำรับการรับส่งเสียง

แก้ปัญหาเรื่องการเข้าจังหวะภาพและเสียง และการตรวจสอบความผิดพลาด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนงานย่อยที่ 2

การแยกส่วนภาพวีดิทัศน์โดยอาศัยการลบฉากหลังเพื่อแยกส่วนข้อมูลองค์ประกอบภาพในการเข้ารหัสแบบวัตถุตามมาตรฐานการเข้ารหัส MPEG-4 ซึ่งเป็นระบบแปบปรับตัวได้

แนวคิดและเหตุผลของแผนงาน

ในปัจจุบันมีการพัฒนาเทคโนโลยีทางสื่อประสม (Multimedia) อย่างไม่หยุดยั้ง ซึ่งเทคโนโลยีการบีบอัดข้อมูลวีดิทัศน์ตามมาตรฐาน MPEG-4 ได้เข้ามามีบทบาทอย่างมากในการส่งข้อมูลเชิงสื่อประสม ด้วยคุณสมบัติที่โดดเด่นในการเข้ารหัส/ถอดรหัสเชิงวัตถุ (Object-based CODEC) ทำให้สามารถแยกส่งองค์ประกอบต่างๆของภาพได้อย่างอิสระกัน โดยโครงการวิจัยในส่วนนี้ จะรับผิดชอบในส่วน กระบวนการแยกส่วนภาพวีดิทัศน์ (Video Segmentation) อันเป็นขั้นตอนก่อนหน้าทำหน้าที่แยกส่วนภาพวัตถุต่างๆ หรือองค์ประกอบในภาพที่สนใจออกจากวัตถุนอกความสนใจ เพื่อนำไปแยกเข้ารหัสตามมาตรฐานการเข้ารหัส/ถอดรหัส MPEG-4 และนำไปส่งต่อในช่องสัญญาณต่อไป

มีการวิจัยกระบวนการแยกส่วนภาพวีดิทัศน์กันอย่างต่อเนื่อง ซึ่งโดยทั่วไปสามารถจำแนกวิธีการแยกส่วนภาพวีดิทัศน์ได้เป็น 3 แนวทาง [1] คือ 1) การแยกส่วนภาพภายในเฟรม หรือ เชิงปริภูมิ (Spatial Segmentation) ซึ่งจะทำการแยกส่วนภาพโดยอาศัยข้อมูลเชิงปริภูมิ ได้แก่ ค่าสี และค่าความสว่างของจุดภาพในภาพนั้นๆ เป็นเกณฑ์ในการตัดสินใจ 2) การแยกส่วนภาพโดยอาศัยการจัดกลุ่มสนามการเคลื่อนที่ (Motion Field Clustering) ของภาพวีดิทัศน์ และ 3) การแยกส่วนภาพโดยการตรวจจับการเคลื่อนที่ (Change Detection) ซึ่งอาศัยความแตกต่างของเฟรมภาพเชิงเวลา (Temporal Frame Difference) อย่างไรก็ดีงานวิจัยโดยส่วนใหญ่มักอาศัยข้อมูลของความสว่างและสีภายในภาพเป็นเกณฑ์หลักในการตัดสินใจ ด้วยเหตุผลที่ว่าตามนุษย์แยกวัตถุที่สนใจโดยอาศัยความสว่างและสีเป็นเกณฑ์หลัก [2] เช่นกัน

วัตถุประสงค์ของแผนงาน

1. พัฒนาระบบการแยกส่วนภาพวีดิทัศน์โดยอาศัยการลบฉากหลังเพื่อแยกส่วนข้อมูลองค์ประกอบภาพในการเข้ารหัสแบบวัตถุตามมาตรฐานการเข้ารหัส MPEG-4 ซึ่งเป็นระบบแปบปรับตัวได้
2. พัฒนาโปรแกรมการวัดประสิทธิภาพการแยกส่วนภาพวีดิทัศน์เชิงปริมาณ (Quantitative Performance Evaluation)

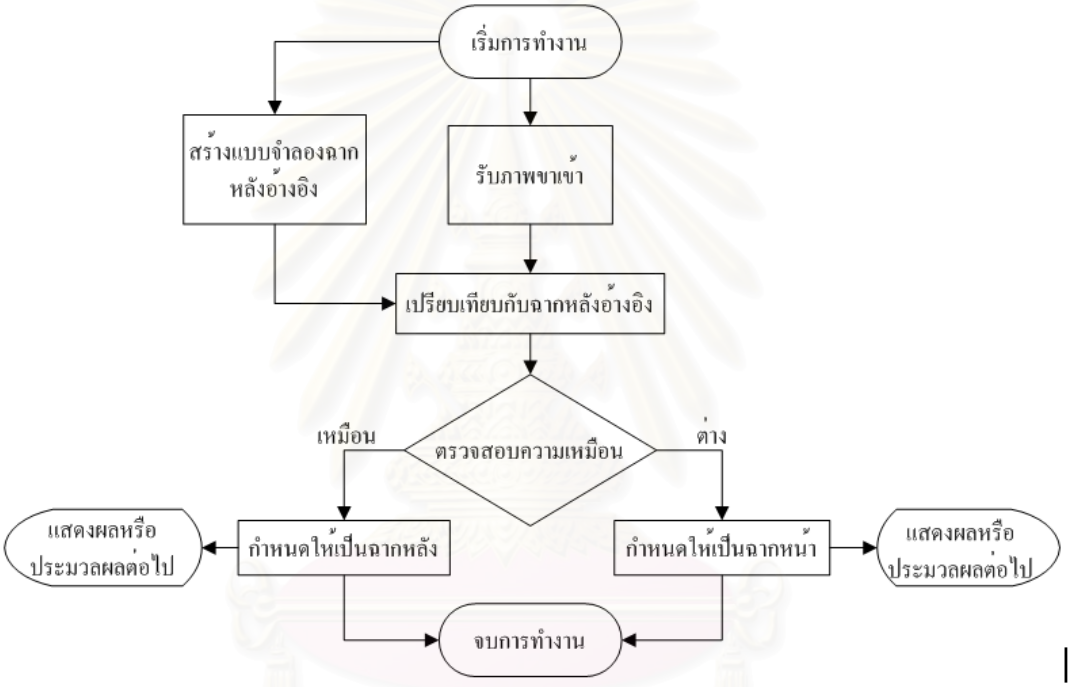
ขอบเขตของแผนงาน

1. พัฒนาระเบียบวิธีการแยกส่วนวัตถุที่เคลื่อนไหวออกจากฉากหลังโดยใช้ วิธีการทางสถิติเชิงพารามเตริกแบบปรับตัวได้ โดยโปรแกรมประยุกต์ที่ได้มีคุณสมบัติดังนี้
   1.1 ทำงานที่ภาพวีดิทัศน์สี 24 บิตต่อจุดภาพ ขนาดกว้าง 320 จุดภาพ สูง 240 จุดภาพ
   1.2 แก้ปัญหาอันเนื่องจากการเปลี่ยนแปลงของฉากหลัง การเปลี่ยนแปลงของความสว่างทั่วทั้งฉาก การเปลี่ยนแปลงของเงาได้
2. พัฒนาโปรแกรมประยุกต์เพื่อใช้วัดประสิทธิภาพเชิงปริมาณ โดยมีผลการทดลองอ้างอิง (Ground Truth) ที่สร้างขึ้นจากการตัดสินใจของมนุษย์เป็นผลการทดลองเปรียบเทียบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

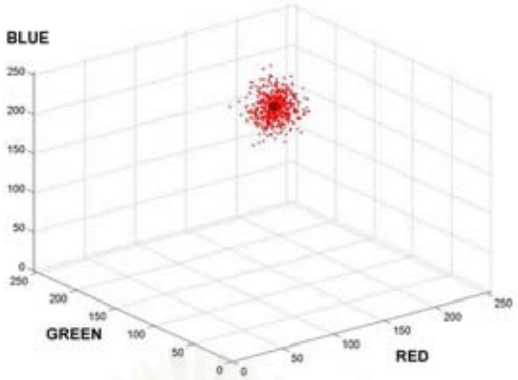
หลักการทำงานพื้นฐานของระบบการแยกส่วนภาพโดยอาศัยการลบฉากหลัง

ระบบการแยกส่วนภาพโดยอาศัยการลบฉากหลังเวลาจริง จะอาศัยหลักการหาความแตกต่างของภาพขาเข้า และภาพฉากหลังที่มีการเก็บไว้อ้างอิง หรือ การสร้างแบบจำลองทางสถิติไว้ซึ่งเป็นที่นิยมมากกว่า โดยตั้งเงื่อนไขว่า ณ จุดภาพที่มีค่าสี “เหมือน” กับภาพอ้างอิงหรือแบบจำลองจะถูกจำแนกให้เป็นฉากหลัง (Background Pixel) ในทางตรงกันข้าม ถ้า ณ จุดภาพที่มีค่าสี “ต่าง” กับภาพอ้างอิงหรือแบบจำลองจะถูกจำแนกให้เป็นจุดภาพฉากหน้า (Foreground Pixel) ซึ่งเงื่อนไขที่จะเป็นตัวกำหนด ความ”เหมือน” และ “ต่าง” ของภาพทั้งสอง จะถูกเรียกว่า “ค่าตัดสิน” (Threshold) และเมื่อทำการประมวลผลครบทุกจุดภาพของภาพขาเข้าปัจจุบัน กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหน้า จะถูกกำหนดให้เป็น “บริเวณฉากหน้า” (Foreground Region) และ กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหลัง จะถูกกำหนดให้เป็น “บริเวณฉากหลัง” (Background Region) โดยสามารถแสดงขั้นตอนการทำงานอย่างง่ายได้ในรูปที่ 1



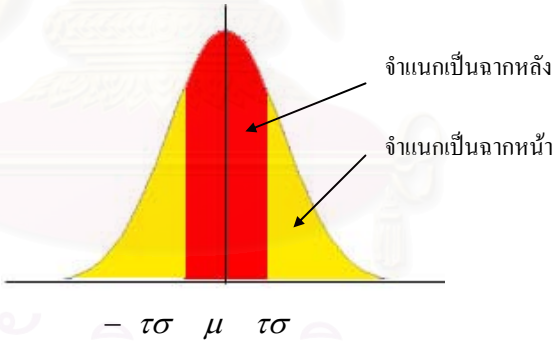
รูปที่ 1 แผนภูมิลำดับการทำงานอย่างง่ายของระบบการแยกส่วนภาพโดยอาศัยการลบฉากหลังเวลาจริง

เมื่อพิจารณาแผนภาพของระบบจะเห็นได้ว่า หัวใจของระบบอยู่ที่ ”*กระบวนการตรวจสอบความเหมือน*” ซึ่งทำงานโดยอาศัยค่าตัดสิน ดังนั้นหน้าที่ของผู้ออกแบบระบบคือการหาค่าตัดสินที่เหมาะสมที่สุดมาใช้ในการจำแนกจุดภาพ และเมื่อพิจารณาย้อนไปถึงการหาค่าตัดสินซึ่งหาได้จาก ”*กระบวนการสร้างแบบจำลองฉากหลัง*” สื่อให้เห็นว่าการสร้างแบบจำลองฉากหลังมีความสำคัญมากเช่นกัน โดยทั่วไป นิยมสร้างแบบจำลองฉากหลังโดยอาศัยระเบียบวิธีทางสถิติ เริ่มจากการเก็บตัวอย่างค่าสี ณ แต่ละจุดภาพ ตามเวลา และพิจารณาการกระจายตัวของข้อมูลจุดสี ซึ่งงานวิจัยส่วนมากมักกำหนดให้ข้อมูลมีการแจกแจงเข้าใกล้แบบจำลองทางทฤษฎี เพื่อความง่ายในการประมวลผล โดยกำหนดให้มีการแจกแจงแบบเกาส์ (Gaussain’s Distribution) ซึ่งตั้งอยู่บนสมมุติฐานที่ว่า เมื่อค่าสี ณ จุดภาพหนึ่งมีค่าคงที่ สัญญาณรบกวนของของสี ณ จุดภาพนั้นในเชิงเวลามีการแจกแจงแบบเกาส์ ซึ่งยืนยันจากการทดลอง ในรูปที่ 2 แสดงรูปแบบการกระจายตัวของค่าสีของจุดภาพหนึ่งตามเวลา และทำการพลอตแบบกระจาย (Scattering Plot) มีการแจกแจงเข้าใกล้แบบเกาส์



รูปที่ 2 การกระจายของค่าสีในหนึ่งจุดภาพในช่วงเวลาที่ทำการประมาณฉากหลังโดยการกระจายของค่าสีมีการแจกแจงใกล้เคียงกับการแจกแจงแบบเกาส์

เมื่อพิจารณาในเชิงเวลา ณ จุดภาพหนึ่งๆ เราจะประมาณฉากหลังให้มีการแจกแจงให้เป็นแบบเกาส์ โดยมีพารามิเตอร์ที่สำคัญได้แก่ ค่าเฉลี่ยเลขคณิต (μ) และ ค่าความแปรปรวน (σ^2) โดย เราใช้ค่าเฉลี่ยเลขคณิตเป็นตัวแทนข้อมูลสี ณ จุดภาพหนึ่งๆ จากภาพทั้งหมดตลอดเวลาที่ทำการสร้างแบบจำลองฉากหลังนั้น และ พิจารณาการกระจายตัวของค่าสีจากค่าความแปรปรวน ดังนั้น จุดภาพที่มีความ “เหมือน“ กับจุดภาพตำแหน่งเดียวกันของฉากหลัง ควรจะมีค่าใกล้เคียงกับค่าเฉลี่ยเลขคณิตที่สุด และมีค่าแตกต่างไปไม่เกินค่าหนึ่งซึ่งยอมรับได้ และโดยทั่วไปมักกำหนดมีค่าความแตกต่างไม่เกิน $|\tau\sigma|$ เป็นฉากหลัง โดยเราเรียก τ ว่า “ค่าตัดสิน” และในทางตรงกันข้ามระบบจะจำแนกจุดภาพที่มีค่าแตกต่างมากกว่า $|\tau\sigma|$ เป็นฉากหน้า ดังแสดงในรูปที่ 3

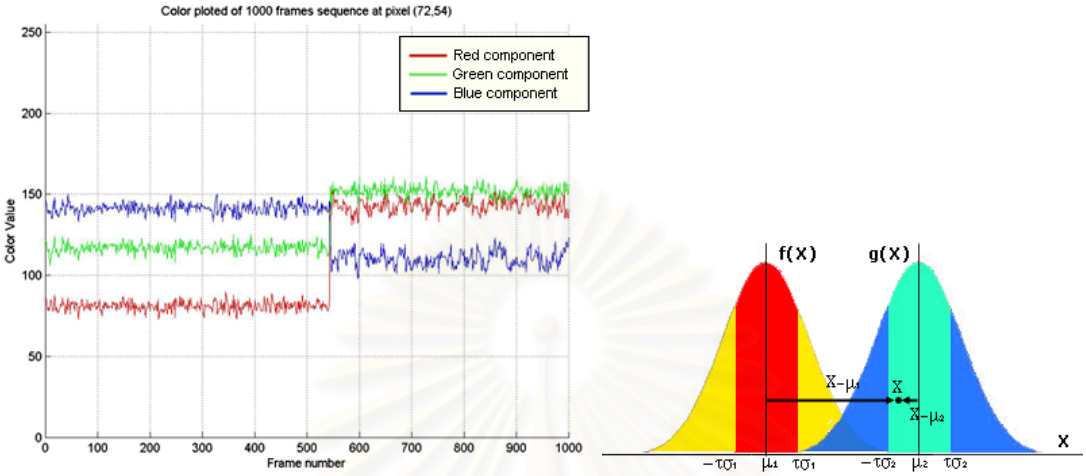


รูปที่ 3 การจำแนกจุดภาพอย่างง่ายโดยใช้แบบจำลองที่มีการแจกแจงแบบเกาส์

วิธีการแยกส่วนภาพโดยอาศัยการลบฉากหลังถูกออกแบบมาให้ใช้กับระบบซึ่งทำงานแบบเวลาจริงโดยเฉพาะ ซึ่งมีข้อดีคือทำงานได้อย่างมีความถูกต้องสูง มีความซับซ้อนน้อยจึงทำงานได้เร็ว แต่อย่างไรก็ตามระบบก็มีความยืดหยุ่นต่ำ เนื่องต้องการการกำหนดค่าก่อนเริ่มทำงาน (Initialization) และทำงานผิดพลาดได้ง่ายเมื่อสภาพแวดล้อมในการทำงานเปลี่ยนแปลง ซึ่งจากการวิเคราะห์สามารถกำหนดข้อจำกัดของระบบการแยกส่วนภาพโดยอาศัยการลบฉากหลังแบบพื้นฐานตามที่กล่าวมาได้ดังนี้

การเปลี่ยนแปลงของความสว่าง (Illumination Changes) ในกรณีนี้หมายถึงการเปลี่ยนแปลงของความสว่างจากแหล่งกำเนิดแสงไม่ว่าจะเป็นแหล่งกำเนิดแสงจริง หรือแหล่งกำเนิดแสงเสมือน ซึ่งจะส่งผลให้คุณลักษณะของแบบจำลองฉากหลังที่ได้ประมาณไว้เปลี่ยนแปลงไปอย่างสิ้นเชิง โดยค่าฉากหลังเมื่อหลังมีการเปลี่ยนแปลงความสว่างจะแตกต่างจากแบบจำลองที่ประมาณไว้เมื่อเริ่มต้นคือ ค่าเฉลี่ยเลขคณิตจะเปลี่ยนแปลง

จะเพิ่มขึ้นหรือลดลงตามค่าความสว่างที่เปลี่ยนแปลงไป และเป็นการเพิ่มค่าความแปรปรวนของแบบจำลองในกรณีระบบที่ปรับตัวได้อีกด้วย [3] ส่งผลให้ระบบที่อาศัยค่าความแปรปรวนเปรียบเทียบกับค่าตัดสิน (Threshold) ทำงานผิดพลาด ซึ่งเราสามารถสังเกตได้จากรูปที่ 4 และผลการประมวลผลในรูปที่ 5



(ก) (ข)

รูปที่ 4 ผลของการเปลี่ยนแปลงความสว่าง (ก) การเปลี่ยนแปลงของค่าสีในเชิงเวลา (ข) เปรียบเทียบการจำแนกเมื่อใช้การกระจายของแบบจำลองฉากหลังเมื่อเริ่มต้น f(x) กับ การกระจายของค่าสีเมื่อมีการเปลี่ยนแปลงความสว่าง g(x)

(ก) (ข) (ค) (ง)

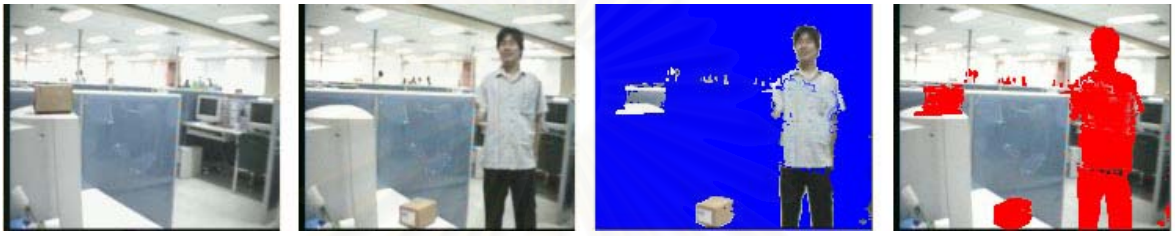
รูปที่ 5 ตัวอย่างที่ระบบทำงานผิดพลาดเนื่องจากการเปลี่ยนแปลงของความสว่างอย่างทั่วฉาก (ก) แบบจำลองฉากหลังที่ได้ทำการประมาณไว้ (ข) ภาพขาเข้าปัจจุบัน (ค) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหน้า (ง) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหลัง

(ก) (ข) (ค) (ง)

รูปที่ 6 ตัวอย่างที่ระบบทำงานผิดพลาดเนื่องจากเงาของวัตถุ (ก) แบบจำลองฉากหลังที่ได้ทำการประมาณไว้ (ข) ภาพขาเข้าปัจจุบัน (ค) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหน้า (ง) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหลัง

เงาของวัตถุ (Object's Shadow) ปัญหาที่เกิดจากเงาคือ การเปลี่ยนแปลงความสว่างเฉพาะท้องถิ่น (Local Illumination Changes) ซึ่งการเปลี่ยนแปลงเกิดจากการเคลื่อนที่ของวัตถุ ระบบจะตีความบริเวณที่เกิดการเปลี่ยนแปลงตามเงาที่ทอดผ่านให้เป็นวัตถุฉากหน้าด้วย โดยสามารถสังเกตตัวอย่างความผิดพลาดในรูปที่ 6 จะเห็นได้ว่า เงาของบุคคลถูกจำแนกให้เป็นฉากหน้า

การย้ายตำแหน่งของวัตถุในฉากหลัง (Relocation of Background Object) ในกรณีที่มีการย้ายตำแหน่งของวัตถุที่อยู่ในฉากหลัง จะทำให้เกิดข้อผิดพลาด 2 ประการ คือ บริเวณที่วัตถุเคยอยู่จะเป็นฉากหลังใหม่ ซึ่งระบบจะตรวจพบการเปลี่ยนแปลงและกำหนดให้เป็นวัตถุฉากหน้าและบริเวณใหม่ที่ย้ายวัตถุไปอยู่ ระบบจะตรวจพบการเปลี่ยนแปลงและกำหนดให้เป็นวัตถุฉากหน้าเช่นกัน ตัวอย่างความผิดพลาดสังเกตได้ในรูปที่ 7



(ก) (ข) (ค) (ง)

รูปที่ 7 ตัวอย่างที่ระบบทำงานผิดพลาดเนื่องจากการย้ายตำแหน่งของวัตถุในฉากหลัง (ก) แบบจำลองฉากหลังที่ได้ทำการประมาณไว้ (ข) ภาพรับเข้าปัจจุบัน (ค) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหน้า (ง) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหลัง

การเปลี่ยนแปลงของฉากหลังซึ่งมีลักษณะเป็นคาบ (Periodic Change in Background) เมื่อฉากหลังมีวัตถุที่เสมือนอยู่ตำแหน่งเดิมแต่มีการเคลื่อนที่ในลักษณะเป็นคาบ จะพบว่าบริเวณดังกล่าวมีการเคลื่อนไหวอยู่ตลอดเวลา เช่นกรณี การพัดโบกของกิ่งไม้ การพลิ้วไหวธง หรือ การแสดงผลบนหน้าจอคอมพิวเตอร์ ล้วนถือเป็นการเคลื่อนไหวแบบคาบทั้งสิ้น ทำให้การจำแนกบริเวณนั้นเป็นฉากหน้าตลอด ในรูปที่ 8

รูปที่ 8 ตัวอย่างที่ระบบทำงานผิดพลาดเนื่องจากการการประมาณแบบจำลองฉากหลังขณะที่มีวัตถุที่เคลื่อนที่อยู่ในฉาก (ก) แบบจำลองฉากหลังที่ได้ทำการประมาณไว้ (ข) ภาพรับเข้าปัจจุบัน (ค) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหน้า (ง) กลุ่มของจุดภาพที่ถูกจำแนกให้เป็นฉากหลัง

ในหลายปีที่ผ่านมา มีการพัฒนาระเบียบวิธีนี้กันอย่างต่อเนื่องเพื่อแก้ปัญหาที่เกิดขึ้นข้างต้น ซึ่งจากการศึกษาและรวบรวม สามารถประมวลผลงานวิจัยที่เกี่ยวข้องได้ โดยเริ่มจาก Jain และคณะ [4] ได้วิจัยระบบซึ่งใช้ความแตกต่างของค่าความเข้มสีระหว่างภาพปัจจุบันและภาพอ้างอิงและเทียบกับค่าตัดสิน นับเป็นระบบอย่างง่ายที่เป็นจุดเริ่มต้นของระบบแยกส่วนภาพโดยอาศัยการลบฉากหลัง ระบบนี้ถูกนำมาใช้เป็นเวลานานนับ

ทศวรรษ จนเมื่อมีการนำเสนอระบบติดตามการเคลื่อนไหวของบุคคล (PFinder : Person Finder) [5] โดย Azerbyjani และคณะซึ่งใช้ความแตกต่างของภาพปัจจุบันกับแบบจำลองฉากหลังทางสถิติแทนภาพอ้างอิง โดยกำหนดให้ค่าสี ณ จุดภาพหนึ่งๆ มีการแจกแจงแบบเกาส์ (Single Gaussian's Distribution) ซึ่งทำให้ระบบทำงานได้อย่างถูกต้องมากขึ้น W^4 (Who What When Where) [6] เป็นระบบต้นแบบอีกระบบหนึ่งซึ่งเป็นที่รู้จักกันอย่างแพร่หลาย ระบบนี้ใช้ข้อมูลของค่าแตกต่างที่มากที่สุดสัมบูรณ์ และค่าแตกต่างที่น้อยที่สุดสัมบูรณ์มาหาค่าตัดสินโดยค่าสีที่อยู่ระหว่างค่าทั้งสองถือว่าเป็นฉากหลัง

จากนั้นความคิดริเริ่มที่จะนำเสนอการสร้างแบบจำลองฉากหลังแบบปรับตัวได้ (Adaptive Background Modeling) เพื่อแก้ปัญหาของการเปลี่ยนแปลงของฉากหลัง โดยผลงานที่โดดเด่นที่สุด เป็นการสร้างแบบจำลองฉากหลังให้มีการแจกแจงเป็นเกาส์แบบผสม (Mixture of Gaussians : MOG) ซึ่งนำเสนอโดย Stauffer และ Grimson [7], [8] โดยระบบจะปรับตัวเมื่อมีการเปลี่ยนแปลงตำแหน่งของวัตถุในฉากหลังได้อย่างดี แต่ถึงกระนั้นระบบยังค่อนข้างซับซ้อนและทำงานได้ช้าเวลาจริง Harville และคณะ [9], [10] ได้ปรับปรุงระบบดังกล่าวให้มีความยืดหยุ่นต่อการเปลี่ยนแปลงของความสว่าง คือใช้ข้อมูลความลึก (Depth) และ ระยะ (Range) มาประกอบการพิจารณา ผลงานที่ออกมาช่วงหลังและเน้นการแก้ปัญหาเรื่องการเปลี่ยนแปลงของฉากหลัง [3], [11], [12] และ [13] จึงมักเป็นการปรับปรุงประสิทธิภาพโดยใช้แบบจำลองที่มีการแจกแจงเกาส์แบบผสมเป็นหลักและใช้คุณสมบัติอื่นๆประกอบการตัดสินใจ ถึงกระนั้นวิธีการข้างต้นไม่ได้ระบุวิธีการแก้ปัญหาเรื่องเงาของวัตถุ

เงาของวัตถุ เป็นประเด็นหนึ่งซึ่งนักวิจัยให้ความสนใจในด้านนี้โดยเฉพาะ วิธีแก้ปัญหาหลักมักทำโดยพิจารณาค่าสีในปริภูมิที่มีการแยกองค์ประกอบทางสี (Chrominance) ออกจากความสว่าง (Luminance) อย่างเป็นอิสระต่อกัน เช่น ปริภูมิ YUV, YCbCr, HLS, HVS นอกจากนี้ Greiffenhagen [14] ได้เสนอการใช้ข้อมูลประกอบของค่าสีในปริภูมิ Normalize-RGB มาช่วยในการตัดสินใจ ซึ่งก็ล้วนแต่เป็นการแยกองค์ประกอบของความสว่างออกจากองค์ประกอบทางสีทั้งสิ้น ผลงานที่โดดเด่นในด้านนี้เป็นวิธีที่นำเสนอโดย T. Horprasert และคณะ [15] โดยระเบียบวิธีใช้ข้อมูลของความเพี้ยนของความสว่าง (Brightness Distortion) และ ความเพี้ยนของสี (Color Distortion) มาประกอบการตัดสินใจ วิธีนี้นอกจากจะมีอัตราการจำแนกได้ (Classification Rate) อยู่ในเกณฑ์ที่สูงแล้วยังมีอัตราการประมวลผลที่สูงด้วย (มากกว่า 50 เฟรมต่อวินาที บนคอมพิวเตอร์ที่มีหน่วยประมวลผล Pentium IV 2.0 GHz) แต่ถึงอย่างไรวิธีดังกล่าวยังไม่สามารถแก้ปัญหาเรื่องการเปลี่ยนแปลงของฉากหลังและความสว่างได้เนื่องจากใช้แบบจำลองแบบสถิตย์ (Static Model)

จากการประมวลวิธีการต่างๆของระเบียบการแยกส่วนภาพโดยอาศัยการลบฉากหลัง มีแนวโน้มที่จะแก้ปัญหาได้มากขึ้นและยืดหยุ่นมากขึ้น อย่างไรก็ดีการนำไปใช้งานจริงมักอยู่ภายใต้เงื่อนไขข้อจำกัดที่เอื้อต่อการทำงานของแต่ละวิธีสูงสุด แบบจำลองเกาส์แบบผสม (Mixture Gaussain's Model) ดูเหมือนจะแก้ปัญหาได้มากที่สุด ในสถานการณ์เวลาจริงที่ไม่คำนึงถึงอัตราการประมวลผล ซึ่งมักใช้ในระบบรักษาความปลอดภัยและสังเกตการณ์บุคคล แต่เมื่อพิจารณาการนำระบบไปประยุกต์ใช้กับระบบสื่อสารเวลาจริง อัตราของเฟรมมีความสำคัญอย่างยิ่งในกระประมวลผลในรูปของความต่อเนื่องของภาพ ซึ่งวิธีที่นำเสนอจะเลือกให้แบบจำลองฉากหลังมีการจำลองเกาส์แบบเดี่ยว (Single Gaussian) โดยนำข้อดีของระบบ [15] มาพิจารณาประกอบการแก้ปัญหาเรื่องเงาและสร้างระบบให้มีการทำงานแบบปรับตัวได้

ทฤษฎีและหลักการ

ปริภูมิของสีและเวกเตอร์ค่าสี

การทำงานของระบบแยกส่วนภาพจะอาศัยข้อมูลของสีเป็นสำคัญ เมื่อพิจารณาระบบกล้องวีดิทัศน์จะรับภาพซึ่งเป็นสัญญาณแอนาลอกและนำไปเปลี่ยนข้อมูลให้เป็นข้อมูลดิจิทัลซึ่งอยู่ในรูปแบบภาพสี RGB 24 บิตต่อจุดภาพ โดยแบ่งเป็นองค์ประกอบของสี แดง เขียว และน้ำเงิน เวกเตอร์ค่าสี ณ จุดภาพหลักที่ i แถวที่ j ของเฟรมที่ n ดังแสดงในสมการที่ (1)

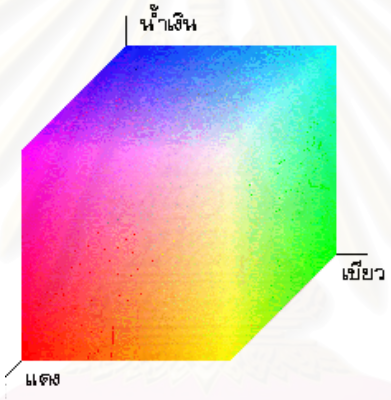
$$\mathbf{X}_{i,j}[n] = (X^{R}_{i,j}[n],\ X^{G}_{i,j}[n],\ X^{B}_{i,j}[n]) \tag{1}$$

โดยที่ $X^{R}_{i,j}[n]$ เป็นค่าความเข้มขององค์ประกอบสีแดง และ $0 \le X^{R}_{i,j}[n] \le 255$

$X^{G}_{i,j}[n]$ เป็นค่าความเข้มขององค์ประกอบสีเขียว และ $0 \le X^{G}_{i,j}[n] \le 255$

$X^{B}_{i,j}[n]$ ค่าความเข้มขององค์ประกอบสีน้ำเงิน และ $0 \le X^{B}_{i,j}[n] \le 255$

ซึ่งแสดงปริภูมิ RGB ได้ในรูปที่ 9



รูปที่ 9 ปริภูมิแสงสีแบบ RGB (สีแดง เขียว และน้ำเงิน)

การสร้างแบบจำลองฉากหลัง

พิจารณากระบวนการจุดภาพ (Pixel Process) ณ จุดภาพ $P(i,j)$ ซึ่งเป็นกระบวนการแบบสุ่ม (Random Process) $\{\mathbf{X}_{i,j}[t]\}$ ให้เป็นกระบวนการแบบเกาส์ (Gaussian Process) จากจุดเวลาจุดใดจุดหนึ่งในลำดับของเซต $\{t_1\ t_2\ ,...,t_n\}$ จากเซตโดเมนเวลา T ฟังก์ชั่นความหนาแน่นร่วม (Joint Density Function) $f_{\mathbf{X}_{i,j}}(\mathbf{X}_{i,j})$ ของเวกเตอร์สุ่ม $\mathbf{X}_{i,j}$ ณ เวลา t ใดๆ คือ

$$f_{\mathbf{X}_{i,j}}(\mathbf{X}_{i,j}) = \frac{|\mathbf{C}_{i,j}|^{-1/2}}{(2\pi)^{n/2}} \exp\{-\frac{1}{2}(\mathbf{X}_{i,j}-\boldsymbol{\mu}_{i,j})^{T}\mathbf{C}_{i,j}^{-1}(\mathbf{X}_{i,j}-\boldsymbol{\mu}_{i,j})\} \tag{2}$$

โดยที่ i เป็นลำดับหลักของจุดภาพที่พิจารณา

j เป็นลำดับแถวของจุดภาพที่พิจารณา

$\mathbf{X}_{i,j}$ เป็นเวกเตอร์ค่าสีที่พิจารณา ณ หลักที่ i แถวที่ j

$\boldsymbol{\mu}_{i,j}$ เป็นเวกเตอร์เวกเตอร์ค่าสีคาดหวัง (Expected Color Vector) ณ หลักที่ i แถวที่ j

$\mathbf{C}_{i,j}$ เป็นเมตริกซ์ค่าสีความแปรปรวนสัมพันธ์ (Covariant Color Matrix) ณ หลักที่ i แถวที่ j

ในเชิงปฏิบัติระบบเริ่มทำงานจากการสร้างแบบจำลองจากการเก็บภาพฉากหลังตัวอย่าง ซึ่งเป็นฉากหลังที่ไม่มีวัตถุที่สนใจอยู่ ขนาดกว้าง W สูง H จุดภาพ เป็นจำนวน N เฟรม เราจะพิจารณาลำดับภาพ (Sequence) โดยสนใจเวกเตอร์ค่าสี $\mathbf{X}_{i,j}[n]$ ณ จุดภาพ $P(i,j)$ เฟรมที่ n เป็นกระบวนการจุดภาพและมีการแจกแจงดังนี้

$$\eta(\mathbf{X}_{i,j}[n],\mathbf{E}_{i,j}[n],\mathbf{C}_{i,j}[n]) = \frac{\left|\mathbf{C}_{i,j}[n]\right|^{-1/2}}{(2\pi)^{N/2}} \exp\{-\frac{1}{2}(\mathbf{X}_{i,j}[n]-\mathbf{E}_{i,j}[n])^T \mathbf{C}_{i,j}^{-1}[n](\mathbf{X}_{i,j}[n]-\mathbf{E}_{i,j}[n])\}$$

$$\text{และ } i \in \{1,2,..W\};\ j \in \{1,2,...,H\}; n \in \{1,2,...N\} \tag{3}$$

โดยที่ n เป็นลำดับเฟรมที่พิจารณา

$\mathbf{E}_{i,j}[n]$ เป็นเวกเตอร์เวกเตอร์ค่าสีเฉลี่ย (Mean Color Vector) ณ หลักที่ i แถวที่ j และเฟรมที่ n

ในที่นี้ เมื่อพิจารณาสร้างแบบจำลองจากข้อมูล N เฟรม กำหนดให้

$$\mathbf{E}_{i,j} = \mathbf{E}_{i,j}[N] = E\{\mathbf{X}_{i,j}[N]\} \tag{4}$$

และ

$$\mathbf{C}_{i,j} = \mathbf{C}_{i,j}[N] = E\{[\mathbf{X}_{i,j}[N]-\mathbf{E}_{i,j}]^T[\mathbf{X}_{i,j}[n]-\mathbf{E}_{i,j}]\} \tag{5}$$

โดยที่ $E\{.\}$ คือตัวปฏิบัติการประมาณ (Estimation Operator)

$\mathbf{E}_{i,j}$ เป็นเวกเตอร์ค่าสีเฉลี่ยสถิตย์ (Static Mean Color Vector)

ในทางปฏิบัติสามารถหา $\mathbf{E}_{i,j}$ และ $\mathbf{C}_{i,j}$ ได้จาก

$$\mathbf{E}_{i,j} = \frac{\sum_{n=1}^{N}\mathbf{X}_{i,j}[n]}{N} \tag{6}$$

$$\mathbf{C}_{i,j} = \mathbf{I}[(\sigma_{i,j}^R)^2\ \ (\sigma_{i,j}^G)^2\ \ (\sigma_{i,j}^B)^2]^T \tag{7}$$

โดยที่ $\mathbf{I}$ เป็นเมตริกซ์เอกลักษณ์

$\sigma_{i,j}^R$ เป็นค่าความแปรปรวนขององค์ประกอบสีแดง ณ หลักที่ i แถวที่ j

$\sigma_{i,j}^G$ เป็นค่าความแปรปรวนขององค์ประกอบเขียว ณ หลักที่ i แถวที่ j

$\sigma_{i,j}^B$ เป็นค่าความแปรปรวนขององค์ประกอบน้ำเงิน ณ หลักที่ i แถวที่ j

จากข้างต้นโดยธรรมชาติของข้อมูลที่เก็บได้ เราได้ตัวแปรของแบบจำลองฉากหลังที่สำคัญ 2 ค่าคือ $\mathbf{E}_{i,j}$ และ $\mathbf{C}_{i,j}$ แต่เนื่องจากข้อมูลทั้งสองยังไม่เพียงพอที่จะทำให้ระบบสามารถจำแนกได้อย่างมีประสิทธิภาพ ดังนั้นเราจึงพิจารณาการแก้ปัญหาเรื่องเงาของวัตถุ โดยจะดัดแปลงระเบียบวิธีจากการนำเสนอของ T. Horprasert และคณะ [15] ซึ่งมีข้อดีดังได้กล่าวไว้ในข้างต้นมาประกอบการพิจารณา

การทำงานของระบบหลังจากหาเวกเตอร์ค่าสีค่าเฉลี่ย และ เมตริกซ์ค่าสีความแปรปรวนสัมพันธ์แล้ว จะกลับมาคำนวณหาค่าความเพี้ยนของความสว่าง (Luminance Distortion) และ ความเพี้ยนของสี (Chrominance Distortion) ซึ่งจะใช้ข้อมูลตัวอย่าง N เฟรมเดิมมาหาค่าทั้งสอง ณ แต่ละจุดภาพ $P(i,j)$

จากหลักการของการมองเห็นของมนุษย์ มักกำหนดให้วัตถุที่มองเห็นมีสีเป็นสีเดิม [2] ไม่ว่าจะมีการเปลี่ยนแปลงแสงสว่างในเชิงเวลา หรือเชิงปริภูมิ ระเบียบวิธีจะอาศัยความเป็นอิสระต่อกันขององค์ประกอบความสว่าง (Luminance) และ ค่าสี (Chrominance) ในพิกัดแบบทรงกลม (Spherical Co-ordinate) โดยค่าสีที่อยู่บนพื้นผิวทรงกลมเดียวกัน หรือมีระยะของรัศมีเท่ากันจะมีองค์ประกอบความสว่างเท่ากัน ในอีกแง่มุมหนึ่ง ค่าสีที่อยู่บนเส้นสี (Chrominance Line) เดียวกันจะมีโทนสีเดียวกันแต่มีความสว่างต่างกัน ดังแสดงในรูปที่ 10

(ก) (ข) (ค)

รูปที่ 10 การพิจารณาโดยแยกองค์ประกอบทางความสว่างและองค์ประกอบทางสี (ก) การพิจารณาในรูปแบบพิกัดทรงกลม (ข) ค่าโทนสีที่ระดับความสว่างเท่ากัน (ค) ค่าความสว่างต่างๆกันที่โทนสีเดียวกัน

ความเพี้ยนของความสว่าง (Luminance Distortion) [α]

ความเพี้ยนของความสว่าง (α) เป็นปริมาณเชิงสเกลาร์ ซึ่งจะมีค่าเมื่อทำให้ค่าสีของผู้สังเกตใกล้เคียงกับค่าสีจริงของวัตถุในแบบจำลองที่สุด โดยค่า α หาได้จากสมการที่ (8)

$$\alpha_{i,j}[n] = \arg_{\psi} \min(\mathbf{X}_{i,j}[n] - \psi \mathbf{E}_{i,j})^2 \tag{8}$$

โดยที่ ψ เป็นตัวแปรหุ่น (Dummy Variable)

ค่า $\alpha_{i,j}[n]$ สามารถใช้แสดงลักษณะของภาพได้จากค่า $\alpha_{i,j}[n]$ ที่ต่างๆกันตามเงื่อนไขต่อไปนี้

$\alpha_{i,j}[n] = 1$ ค่าความสว่างของค่าที่สังเกตุได้เท่ากับค่าความสว่างของค่าสีจริงในแบบจำลอง

$\alpha_{i,j}[n] < 1$ ค่าความสว่างของค่าที่สังเกตุได้น้อยกว่าค่าความสว่างของค่าสีจริงในแบบจำลอง

$\alpha_{i,j}[n] > 1$ ค่าความสว่างของค่าที่สังเกตุได้มากกว่าค่าความสว่างของค่าสีจริงในแบบจำลอง

ในทางปฏิบัติ หา $\alpha_{i,j}[n]$ ได้จากสมการที่ (9)

$$\alpha_{\mathrm{i,j}}[n] = \arg_{\psi} \min\left[\left(\frac{X^R_{i,j}[n] - \psi E^R_{i,j}}{\sigma^R_{i,j}}\right)^2 + \left(\frac{X^G_{i,j}[n] - \psi E^G_{i,j}}{\sigma^G_{i,j}}\right)^2 + \left(\frac{X^B_{i,j}[n] - \psi E^B_{i,j}}{\sigma^B_{i,j}}\right)^2\right] \tag{9}$$

หลังจากทำการหาค่าต่ำสุดจะได้รูปอย่างง่าย ในสมการที่ (10)

$$\alpha_{\mathrm{i,j}} = \frac{\left(\frac{X^R_{i,j}[n]E^R_{i,j}}{(\sigma^R_{i,j})^2} + \frac{X^G_{i,j}[n]E^G_{i,j}}{(\sigma^G_{i,j})^2} + \frac{X^B_{i,j}[n]E^B_{i,j}}{(\sigma^B_{i,j})^2}\right)}{\left(\left[\frac{E^R_{i,j}}{\sigma^R_{i,j}}\right]^2 + \left[\frac{E^G_{i,j}}{\sigma^G_{i,j}}\right]^2 + \left[\frac{E^B_{i,j}}{\sigma^B_{i,j}}\right]^2\right)} \tag{10}$$

ความเพี้ยนของสี (Chrominance Distortion) [λ]

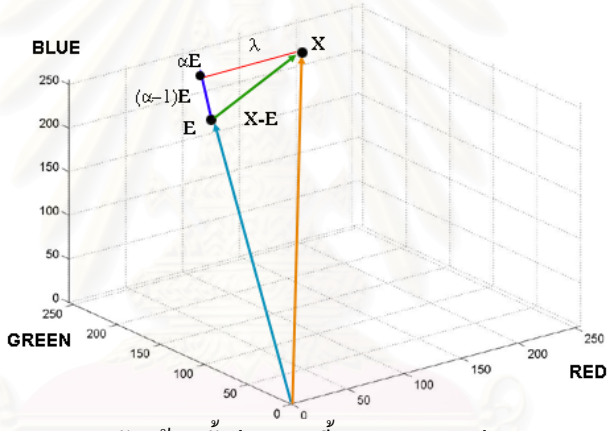
ความเพี้ยนของสี เป็นค่าความแตกต่างของแกนสีของภาพที่ใช้อ้างอิงในแบบจำลองฉากหลังกับสีของภาพที่สังเกตุได้จากภาพขาเข้า โดยจะเป็นระยะทางระหว่าง ค่าสีของภาพที่อ้างอิง กับค่าสีของภาพที่สังเกตุได้ที่ระดับความสว่างเท่ากัน โดยสามารถหาได้จากสมการที่ (11)

$$\lambda_{i,j}[n] = \left\| \mathbf{X}_{i,j}[n] - \alpha_{i,j}[n]\mathbf{E}_{i,j} \right\| \tag{11}$$

ในทางปฏิบัติหา $\lambda_{i,j}[n]$ ได้จากสมการที่ (12)

$$\lambda_{i,j}[n] = \sqrt{\left(\frac{X^R_{i,j}[n] - \alpha_{i,j}E^R_{i,j}}{\sigma^R_{i,j}}\right)^2 + \left(\frac{X^G_{i,j}[n] - \alpha_{i,j}E^G_{i,j}}{\sigma^G_{i,j}}\right)^2 + \left(\frac{X^B_{i,j}[n] - \alpha_{i,j}E^B_{i,j}}{\sigma^B_{i,j}}\right)^2} \tag{12}$$

ความเพี้ยนทั้งสองแนวแกนใช้หลักการของความเป็นอิสระต่อกันเราสามารถแสดงแบบจำลองฉากหลังที่สอดคล้องกับตัวแปรข้างต้นได้ดังในรูปที่ 11



รูปที่ 11 แบบจำลองฉากหลังพร้อมทั้งค่าความเพี้ยนของความสว่างและความเพี้ยนของสี

จากข้างต้นในสมการที่ (8) จะเห็นได้ว่า ความแปรปรวนของ $\alpha_{i,j}[n]$ และ $\lambda_{i,j}[n]$ มีความแตกต่างกันขึ้นอยู่กับ ตำแหน่งจุดภาพ ดังนั้นจึงมีความจำเป็นที่จะเปรียบเทียบที่มาตรฐานเดียวกัน โดยการหาค่าเบี่ยงเบนเชิงความสว่าง $a_{i,j}$ และค่าเบี่ยงเบนเชิงสี $b_{i,j}$ ดังในสมการที่ (13) และ (14)

$$a_{i,j} = RMS(\alpha_{i,j}[n]) = \sqrt{\frac{\sum_{n=1}^{N}(\alpha_{i,j}[n]-1)^2}{N}} \tag{13}$$

$$b_{i,j} = RMS(\lambda_{i,j}[n]) = \sqrt{\frac{\sum_{n=1}^{N}(\lambda_{i,j}[n])^2}{N}} \tag{14}$$

ณ จุดนี้ เราจะได้ ตัวแปรเชิงสถิติของแบบจำลองฉากหลัง 4 ตัวแปร ณ จุด $P(i,j)$ คือ $\{\mathbf{E}_{i,j}, \mathbf{C}_{i,j}, a_{i,j}, b_{i,j}\}$ โดยจะมีทั้งหมด $4 \cdot W \cdot H$ ตัวแปรในการเก็บตัวอย่าง N เฟรม

การประมวลผลเวลาจริงและการจำแนกค่า (Real-time Processing and Classification)

การทำงานเวลาจริงจะประมวลผลภาพทีละเฟรมที่จับจากกล้องวีดิทัศน์ เมื่อรับข้อมูลมาแล้ว จะพิจารณาทีละจุดภาพ $P(i,j)$ จำนวน $W \cdot H$ จุดภาพ โดยทำการลบ (Subtract) เวกเตอร์ค่าสี ณ จุดนั้นๆ $\mathbf{X}_{i,j}[n]$ ด้วยเวกเตอร์ค่าสีเฉลี่ยของแบบจำลอง $\mathbf{E}_{i,j}$ และหาค่าความเพี้ยนของความสว่าง $\alpha_{i,j}[n]$ และความเพี้ยนของสี $\lambda_{i,j}[n]$ ตามสมการที่ (8) และ (11) จากนั้นจะทำให้เป็นค่ามาตรฐาน (Standardization) โดยการนอมัลไลซ์ด้วยค่าเบี่ยงเบนเชิงความสว่าง $a_{i,j}$ และค่าเบี่ยงเบนเชิงสี $b_{i,j}$ เป็นความเพี้ยนของความสว่างมาตรฐาน $\hat{\alpha}_{i,j}[n]$ และ ความเพี้ยนของสีมาตรฐาน $\hat{\lambda}_{i,j}[n]$ ดังแสดงในสมการที่ (15) และ (16)

$$\hat{\alpha}_{i,j}[n] = \frac{\alpha_{i,j}[n] - 1}{a_{i,j}[n]} \tag{15}$$

$$\hat{\lambda}_{i,j}[n] = \frac{\lambda_{i,j}[n]}{b_{i,j}[n]} \tag{16}$$

จากนั้นค่า $\hat{\alpha}_{i,j}[n]$ และ $\hat{\lambda}_{i,j}[n]$ จะถูกใช้ในกระบวนการจำแนกจุดภาพ โดยในกระบวนการจำแนกจะแบ่งกลุ่มของจุดภาพออกเป็น 4 กลุ่มได้แก่

1. ส่วนฉากหลัง (B: Background) ถ้าทั้งความสว่างและสีของจุดภาพปัจจุบันคล้ายคลึงกับจุดภาพ ณ ตำแหน่งเดียวกันของพื้นฉากหลัง
2. ส่วนเงาของวัตถุ (S: Shadow) ถ้าจุดภาพปัจจุบันมีสีคล้ายคลึงแต่มีความสว่างน้อยกว่าจุดภาพตำแหน่งเดียวกันของพื้นฉากหลัง
3. ส่วนแสงสะท้อนจากวัตถุ (H : Highlight) ถ้าจุดภาพปัจจุบันมีสีคล้ายคลึงแต่มีความสว่างมากกว่าจุดภาพตำแหน่งเดียวกันของพื้นฉากหลัง
4. ส่วนวัตถุที่เคลื่อนไหว (F: Foreground) ถ้าทั้งสีและความสว่างของจุดภาพปัจจุบันแตกต่างกับกับจุดภาพตำแหน่งเดียวกันของฉากหลังอย่างสิ้นเชิง

กลุ่มในการจำแนกจะแสดงโดยหน้ากากของการจำแนก (Classification Mask) $M_{i,j}[n]$ โดยจะพิจารณาตามเงื่อนไขจากบนลงล่าง (เงื่อนไข F ไปเงื่อนไข H) ดังแสดงในสมการที่ 17

$$M_{i,j}(n) = \begin{cases} F: \ \hat{\lambda}_{i,j}[n] > \tau_{\lambda} \ \ or \ \ \hat{\alpha}_{i,j}[n] < \tau_{\alpha lo}, else \\ B: \ \hat{\alpha}_{i,j}[n] < \tau_{\alpha 1} \ \ and \ \ \hat{\alpha}_{i,j}[n] > \tau_{\alpha 2} \ , else \\ S: \ \alpha_{i,j}[n] < 0, \ else \\ H: \ otherwise \end{cases} \tag{17}$$

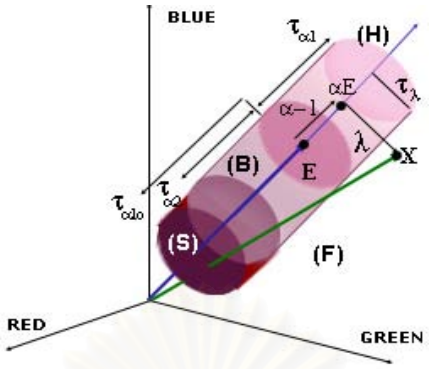
โดยที่ τ_{λ} เป็นค่าตัดสินเชิงสีที่ได้จากค่า $\hat{\lambda}_{i,j}[n]$ ที่ตำแหน่งอัตราความผิดพลาด r หรือกล่าวอีกนัยหนึ่งคือ เป็นค่า $\hat{\lambda}_{i,j}[n]$ ณ ตำแหน่งศตภาค (Percentile) ที่ r ในการแจกแจงค่า $\hat{\lambda}_{i,j}[n]$ ที่เกิดขึ้นทั้งหมดในแบบจำลอง

$\tau_{\alpha 1}$ เป็นค่าตัดสินเชิงความสว่างที่ได้จาก $\hat{\alpha}_{i,j}[n]$ ณ ตำแหน่งอัตราความผิดพลาด r

$\tau_{\alpha 2}$ เป็นค่าตัดสินเชิงความสว่างที่ได้จาก $\hat{\alpha}_{i,j}[n]$ ณ ตำแหน่งอัตราความผิดพลาด 1- r

$\tau_{\alpha lo}$ เป็นค่าตัดสินที่กำหนดขึ้นเพื่อจำกัดความมืดของเงาเพื่อป้องกันการตีความผิดพลาดในกรณีการตีความวัตถุมืด หรือ วัตถุที่มีสีค่อนข้างดำ เป็นเงาของวัตถุ

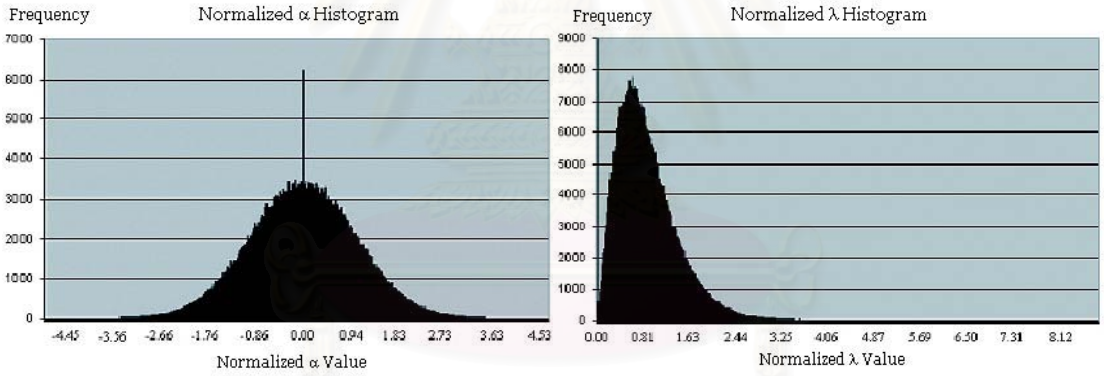
แบบจำลองการจำแนกอธิบายการทำงานข้างต้นได้ในรูปที่ 12



รูปที่ 12 แบบจำลองอธิบายการจำแนกจุดภาพออกเป็น 4 กลุ่มจุดภาพ

การเลือกค่าตัดสินแบบกึ่งอัตโนมัติ (Semi-automatic Threshold Selection)

ในกรณีที่การกระจายของ $\hat{\alpha}_{i,j}[n]$ และ $\hat{\lambda}_{i,j}[n]$ มีการกระจายแบบเกาส์ เราจะสามารถระบุตำแหน่งของค่าตัดสินโดยใช้ส่วนเบี่ยงเบนมาตรฐานได้ แต่ในความเป็นจริงพบว่าการกระจายของทั้งสองตัวแปรไม่อยู่ในรูปแบบเกาส์ โดยจากการทดลองทุกๆลำดับภาพ การกระจายจะอยู่ในรูปแบบ ดังรูปที่ 13 เสมอ [15]



(ก) (ข)

รูปที่ 13 การกระจายตัวของ (ก) ความเพี้ยนของความสว่างมาตรฐาน $\hat{\alpha}_{i,j}[n]$ (ข) ความเพี้ยนของสีมาตรฐาน $\hat{\lambda}_{i,j}[n]$

การหาค่าตัดสินจะกระทำโดยการกำหนด อัตราความผิดพลาด (Error-rate) r โดยผู้ใช้จะเป็นผู้กำหนด และเมื่อการสร้างแบบจำลองของระบบเสร็จสิ้นระบบจะคำนวณหาค่าตัดสินโดยใช้ อัตราความผิดพลาด หาข้อมูลในตำแหน่งศตภาคต่างๆดังนี้

ค่าตัดสินเชิงสี τ_{λ} คำนวณจากค่า $\hat{\lambda}_{i,j}[n]$ ที่ ศตภาค r

ค่าตัดสินเชิงความสว่าง $\tau_{\alpha 1}$ และ $\tau_{\alpha 2}$ คำนวณจากค่า $\hat{\alpha}_{i,j}[n]$ ที่ ศตภาค r และ 1-r

ค่าตัดสินขอบเขตเงา $\tau_{\alpha lo}$ ผู้ใช้เป็นผู้กำหนดค่า $\hat{\alpha}_{i,j}[n]$ ที่ต่ำที่สุดที่ยอมรับได้ว่าเป็นเงา

ผลการทดลอง

หลักจากนั้นจึงพิจารณาทดสอบสอบระเบียบวิธีดังกล่าว โดยใช้โปรแกรมของ T. Horprasert [15] มาปรับปรุงเพื่อสร้างเป็นโปรแกรมต้นแบบในการทดลอง โดยการทดลองใช้ลำดับภาพเปรียบเทียบ (Benchmark

Sequence) จาก T. Horprasert ซึ่งเป็นลำดับภาพการเคลื่อนที่ของบุคคลภายใต้แหล่งกำเนิดแสงหลายแหล่ง [15]ในรูปที่ 14 และ จาก K. Kim ซึ่งเป็นลำดับภาพการเคลื่อนที่ของบุคคลภายใต้แหล่งกำเนิดแสงหลายแหล่งและมีเปลี่ยนแปลงองค์ประกอบของฉากหลัง [16] ในรูปที่ 15 โดยจากผลการทดลองจุดภาพที่ถูกจำแนกให้เป็นวัตถุจะมีหน้ากาก (Mask) สีฟ้า จุดภาพที่ถูกจำแนกให้เป็นเงาจะมีหน้ากากสีแดง

รูปที่ 14 ระเบียบวิธีสามารถจำแนกจุดภาพ และตรวจจับเงาได้อย่างถูกต้อง ถึงแม้การเกิดเงาจะอยู่ภายใต้เงื่อนไขของแหล่งกำเนิดแสงหลายแหล่ง

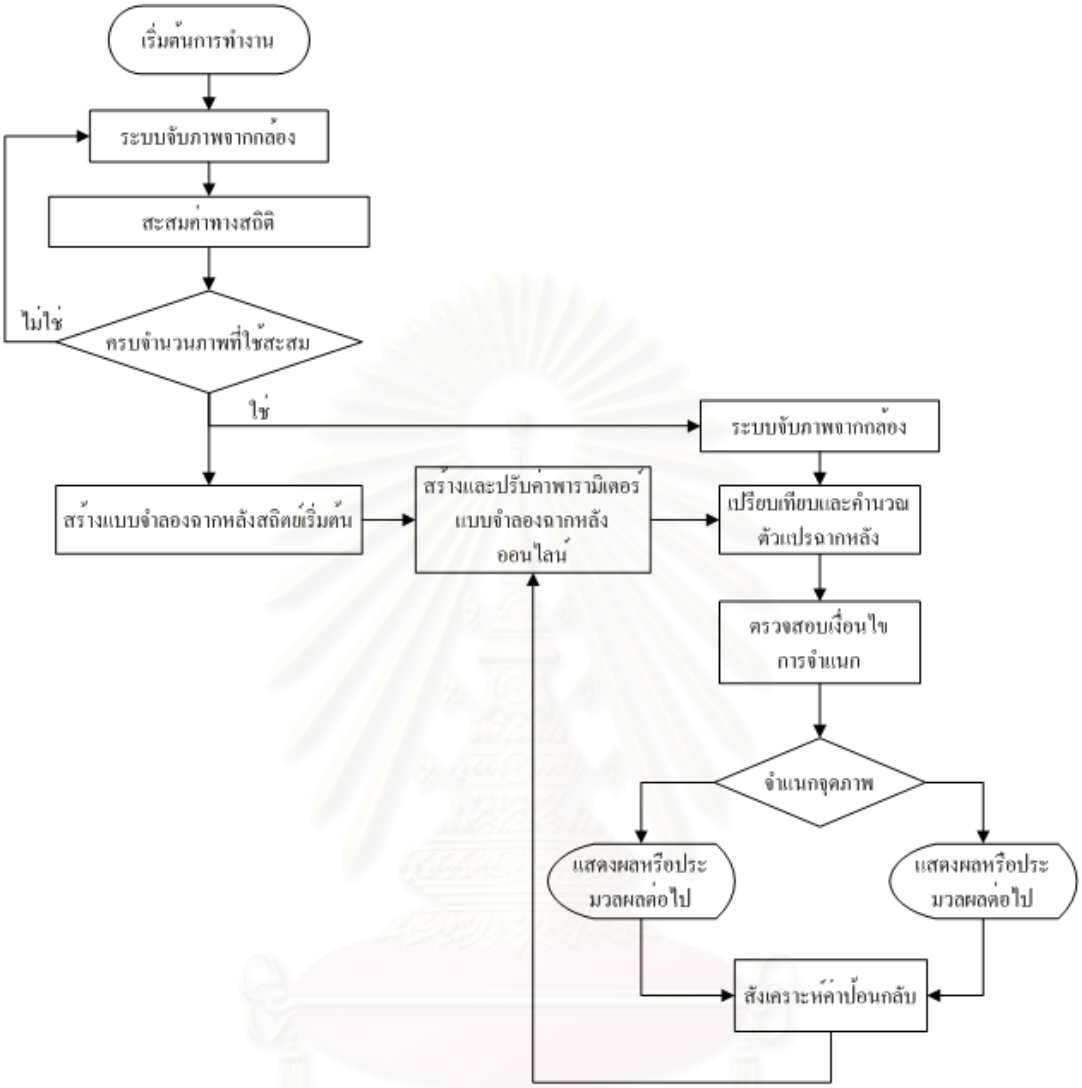
รูปที่ 15 ระเบียบวิธีสามารถจำแนกจุดภาพ และตรวจจับเงาได้อย่างถูกต้อง และระบุตำแหน่งของฉากหลังที่เกิดการเปลี่ยนแปลง

จากการสังเกตผลการทดลองระบบดังกล่าวสามารถจัดการปัญหาที่เกิดจากเงาได้อย่างมีประสิทธิภาพ แต่ในกรณีที่ฉากหลังมีการเปลี่ยนแปลง ระบบจะตรวจพบการเปลี่ยนแปลงและระบุว่าการเปลี่ยนแปลงนั้นเป็นวัตถุเสมอและเมื่อเวลาผ่านไป ฉากหลังมีการเปลี่ยนแปลงมากขึ้น วัตถุที่พบก็จะพบมากขึ้นตามด้วย สมผลให้เกิดความผิดพลาดสะสมและทำให้เกิดปัญหาเมื่อนำไปใช้ในระบบติดตามวัตถุ (Object Tracking System) ระบบจะทำงานผิดพลาด เนื่องจากมีวัตถุที่ระบบตรวจพบมีใช่วัตถุจริงที่สนใจ ยกตัวอย่างเช่น ในรูปที่ 15 จะเห็นได้ว่าระบบจะจำแนกให้บริเวณกล่อง เป็นวัตถุฉากหน้าตลอดไปไม่ว่าเวลาจะผ่านไปนานเท่าใด

แนวคิดของงานนี้

จากที่ได้กล่าวข้างต้นระบบที่พัฒนาขึ้นมีจุดอ่อนในสถานการณ์ที่มีการเปลี่ยนแปลงของฉากหลังรวมทั้งการเปลี่ยนแปลงของความสว่างแบบทั่วฉาก เนื่องจากการสร้างแบบจำลองฉากหลังกระทำครั้งเดียวเมื่อเริ่มต้นระบบระบบดังกล่าวจึงถือว่าเป็น ระบบที่ใช้แบบจำลองฉากหลังแบบสถิตย์ (Static Background Model) ซึ่งไม่ยืดหยุ่นต่อสถานการณ์แวดล้อมที่เปลี่ยนแปลง ดังนั้นระบบที่จะแก้ปัญหาดังกล่าวได้ต้องมีการปรับค่าฉากหลังให้ทันต่อเหตุการณ์เวลาจริง (Real-time Updating) โดยการใช้การสร้างแบบจำลองฉากหลังออนไลน์ (On-line Background Modeling) ซึ่งการป้อนกลับผลการจำแนกกลับไปปรับค่าค่าพารามิเตอร์แบบจำลองฉากหลัง (Background Model Parameters) ดังนั้นวัตถุที่ตรวจพบและไม่มีการเคลื่อนไหวอีกในช่วงระยะเสลาที่กำหนดค่าหนึ่ง จะถูกจำแนกเป็นวัตถุนอกความสนใจ (Non-interesting Object) และกลายเป็นส่วนหนึ่งของแบบจำลองฉากหลังในที่สุด

ระบบที่ออกแบบมีแผนผังการทำงานดังรูปที่ 16



รูปที่ 16 แผนผังการทำงานของระบบที่ได้ออกแบบ

การสร้างแบบจำลองฉากหลังออนไลน์ (On-line Background Modeling)

การสร้างแบบจำลองฉากหลังออนไลน์จะปรับค่าเวลาจริงทุกๆเฟรมหรือทุกๆช่วงระยะเฟรมปรับค่า (Updated Frame Interval) โดยตัวแปรแบบจำลองฉากหลังออนไลน์(On-line Background Model Parameters) จะประกอบด้วย

1. เวกเตอร์ค่าสีเฉลี่ยออนไลน์ $\mathbf{E}_{i,j}[n]$
2. เมตริกซ์ค่าสีความแปรปรวนสัมพันธ์ออนไลน์ $\mathbf{C}_{i,j}[n]$
3. ค่าเบี่ยงเบนเชิงความสว่างออนไลน์ $a_{i,j}[n]$
4. ค่าเบี่ยงเบนเชิงสีออนไลน์ $b_{i,j}[n]$

การปรับค่าออนไลน์(On-line Updating) จะเริ่มต้นด้วยการกำหนดค่าเริ่มต้นจะอาศัยข้อมูลแบบจำลองฉากหลังสถิตที่ได้สร้างไว้เมื่อเริ่มต้น (Initialized Static Background Model) ซึ่งได้แก่ตัวแปรฉากหลัง $\{\mathbf{E}_{i,j}$,

$\mathbf{C}_{i,j}$, $a_{i,j}$, $b_{i,j}$ } เป็นพารามิเตอร์ปรับค่าเริ่มต้น (Initialized Updated Parameters) และทำการปรับค่าตัวแปรแบบจำลองฉากหลังออนไลน์{ $\mathbf{E}_{i,j}[n], \mathbf{C}_{i,j}[n], a_{i,j}[n], \quad b_{i,j}[n]$ } ตามสมการการปรับค่า (Update Equations)ดังนี้

$$\mathbf{E}_{i,j}[n] = \begin{cases} \mathbf{E}_{i,j} & : n = 1 \\ (1-\gamma)\,\mathbf{E}_{i,j}[n-1] + \gamma\,\mathbf{X}_{i,j}[n] & : n > 1 \end{cases} \tag{18}$$

$$\mathbf{C}_{i,j}[n] = \begin{cases} \mathbf{C}_{i,j} & : n = 1 \\ (1-\gamma)\mathbf{C}_{\mathrm{i,j}}[n-1] + \gamma(\mathbf{X}_{\mathbf{i},j}[n] - \mathbf{E}_{\mathrm{i,j}}[n])^T (\mathbf{X}_{\mathbf{i},j}[n] - \mathbf{E}_{\mathrm{i,j}}[n]) & : n > 1 \end{cases} \tag{19}$$

$$a_{i,j}[n] = \begin{cases} a_{i,j} & : n = 1 \\ \sqrt{(1-\gamma)\mathrm{a}_{i,j}^2[n-1] + \gamma(\alpha_{i,j}[n]-1)^2} & : n > 1 \end{cases} \tag{20}$$

$$b_{i,j}[n] = \begin{cases} b_{i,j} & : n = 1 \\ \sqrt{(1-\gamma)\mathrm{b}_{\mathrm{i,j}}^2[\mathrm{n}-1] + \gamma(\lambda_{\mathrm{i,j}}[n])^2} & : n > 1 \end{cases} \tag{21}$$

โดยที่กำหนดให้

γ เป็นอัตราการเรียนรู้ของการปรับค่า (Learning Rate of Adaptation)

ค่าของ γ จะบ่งบอกถึงอัตราเร็วในการปรับค่าของตัวแปรแบบจำลองในขณะปัจจุบันจากตัวแปรแบบจำลองในรอบการคำนวณที่แล้ว ค่า $1/\gamma$ จึงสื่อถึงค่าคงที่เวลา (Time Constant) ในการปรับค่า

สัมประสิทธิ์วิวาซิตี (Vivacity Co-efficient : υ)

จากข้างต้น อัตราการเรียนรู้ของการปรับค่า (γ) บ่งบอกถึง ความเร็วของการปรับค่าของแบบจำลองฉากหลัง ถ้า γ มีค่ามาก ฉากหลังก็จะถูกปรับค่าอย่างรวดเร็วเช่นในกรณีที่มีการเปลี่ยนแปลงของฉากหลัง เช่น การย้ายตำแหน่งเก้าอี้ กล่อง ในสำนักงาน สมการการปรับค่าจะปรับค่าให้วัตถุที่อยู่ตำแหน่งใหม่เป็นวัตถุที่อยู่นอกความสนใจและจำแนกให้เป็นฉากหลังอย่างรวดเร็ว แต่ในอีกแง่มุมหนึ่งในฉากที่มีบุคคลพลุกพล่าน ฉากหลังที่แท้จริงจะถูกรวบกวนให้เปลี่ยนค่าอย่างรวดเร็ว เช่นเดียวกับ กรณีของการปรากฏของบุคคลที่เคลื่อนที่ในฉาก แล้วหยุดนิ่งอยู่ในฉาก เป็นเวลานาน บุคคลนั้นจะถูกกำหนดให้เป็นวัตถุที่อยู่นอกความสนใจและจำแนกให้เป็นฉากหลังอย่างรวดเร็ว ดังนั้น ความต้องการที่สวนทางกันจึงเกิดขึ้นระหว่าง “การปรับค่าฉากหลังบริเวณที่มีการเปลี่ยนแปลงตำแหน่งของวัตถุให้เร็วที่สุด” และ “การปรับค่าฉากหลังบริเวณที่มีวัตถุ หรือ บุคคลที่สนใจซ้อนทับอยู่ให้ช้าที่สุด“ วิธีทั่วไปที่จะแก้ปัญหาคือการหา อัตราการเรียนรู้ที่เหมาะสมที่สุด (Optimal Learning Rate) แต่เนื่องจากความหลากหลายของสถานการณ์ทำให้ระบบไม่มีความยืดหยุ่น สิ่งที่วิทยานิพนธ์นี้เสนอเพื่อแก้ปัญหา คือ *“สัมประสิทธิ์วิวาซิตี (Vivacity Co-efficient)”*

หลักการของการปรับค่าโดยใช้ *สัมประสิทธิ์วิวาซิตี* คือปรับเปลี่ยนอัตราการเรียนรู้ให้สอดคล้องกับการเคลื่อนไหว และ กิจกรรม ณ บริเวณนั้นๆ ในระดับจุดภาพ โดยส่วนที่มีการเคลื่อนไหวประจำให้มีอัตราการเรียนรู้ต่ำ และ บริเวณที่มีการเลื่อนไหวเป็นครั้งคราวหรือ สงบนิ่งให้มีอัตราการเรียนรู้ที่สูง

นิยาม วิวาซิตี (Vivacity) คือ ปริมาณที่บ่งบอกถึงระดับของการเคลื่อนไหว การเคลื่อนที่ หรือ กิจกรรม ณ บริเวณที่พิจารณา

ในระเบียบวิธีที่เสนอนี้ กำหนดให้วิวาซิตีหาได้จาก ค่าเปลี่ยนแปลงสะสมของเวกเตอร์ค่าสี ณ จุดภาพ $P(i,j)$ ที่สนใจ พิจารณาลำดับภาพ RGB ความละเอียด 24 บิตจุดภาพ จำนวน N เฟรม ซึ่งเราจะหาวิวาซิตีที่จุดภาพ $P(i,j)$ ในรูปของ สัมประสิทธิ์วิวาซิตี ได้ดังสมการที่ (22)

$$\upsilon_{i,j} = \sqrt{\frac{\sum_{n=1}^{N} \left\| \mathbf{X}_{i,j}[n] - \mathbf{X}_{i,j}[n-1] \right\|^2}{255^2 N}} \tag{22}$$

เมื่อพิจารณาหาสัมประสิทธิ์วิวาซิตีของลำดับภาพกว้าง W สูง H เพื่อการเปรียบเทียบระหว่างลำดับภาพสามารถหาได้จากสมการที่ (23)

$$V = \sum_{i=i}^{W} \sum_{j=1}^{H} \upsilon_{i,j} \tag{23}$$

เนื่องจากงานวิจัยที่จะนำไปใช้ เป็นระบบแบบปรับค่าได้ ดังนั้นสัมประสิทธิ์วิวาซิตี จะต้องอยู่ในรูปของสมการการปรับค่า ดังสมการที่ (24)

$$\upsilon_{i,j}[n] = \sqrt{(1-\zeta)(\upsilon_{i,j}[n-1])^2 + \zeta \frac{\left\| \mathbf{X}_{i,j}[n] - \mathbf{X}_{i,j}[n-1] \right\|^2}{255^2}} \tag{24}$$

โดยที่ ζ เป็นอัตราการเรียนรู้วิวาซิตี

อัตราการเรียนรู้วิวาซิตีจะช่วยคงค่าวิวาซิตีให้เปลี่ยนแปลงอย่างช้าๆ เพื่อป้องกันการทำงานผิดพลาดในกรณีที่มีวัตถุลวดลายน้อยเคลื่อนที่ในฉาก โดยเราจะกำหนดให้ คงค่าวิวาซิตีเริ่มต้นเป็นศูนย์ ($\upsilon_{i,j}[1] = 0$) ดังนั้น อัตราการเรียนรู้ γ ในสมการการปรับค่า (18), (19), (20) และ (21) จะถูกแทนค่าด้วยสมการที่ (25)

$$\gamma'_{i,j}[n] = (1 - \upsilon_{i,j}[n])\gamma \tag{25}$$

สัมประสิทธิ์วิวาซิตี ไม่ได้มีประโยชน์เพียงในงานวิจัยนี้เท่านั้น แต่ยังสามารถนำไปใช้ในงานวิจัยที่ต้องการระบุระดับของกิจกรรมในฉากต่างๆเป็นเชิงปริมาณ

ช่วงระยะเฟรมปรับค่า (Updated Frame Interval : I_U)

ช่วงระยะเฟรมปรับค่าถูกกำหนดขึ้นเพื่อยืดระยะเวลาในการปรับค่าในกรณีวัตถุที่เคลื่อนที่เร็วและปรับค่าบริเวณฉากหลังด้วยอัตราการเรียนรู้ปกติ หลักการคือ แทนที่ระบบจะทำการปรับค่าทุกเฟรมที่ประมวลผล แต่จะปรับค่าทุกๆ ระยะ I_U เฟรม ดังนั้น จุดภาพที่มีวัตถุนั้นปรากฏอยู่ จะยังไม่ปรับค่าจนเมื่อการประมวลผลผ่านไป I_U เฟรม ดังนั้นที่ผ่านเข้ามาและปรากฏในจุดภาพที่เราสนใจน้อยกว่าช่วงระยะ I_U จะไม่ถูกปรับค่า จนเมื่อ วัตถุอยู่จนครบ I_U ถึงมีการปรับค่า ดังนั้น วิธีดังกล่าวจึงสามารถแยกแยะ สิ่งที่เคลื่อนที่ตลอดเวลา เช่น บุคคลหรือรถในทางสัญจรตลอดเวลา กับรถที่ย้ายตำแหน่งการจอด

การสร้างและพัฒนาระบบ

การสร้างและพัฒนาระบบจะอาศัยโครงสร้างของระบบที่ออกแบบไว้และระเบียบวิธีที่ได้สร้างขึ้น พัฒนาโปรแกรมต้นแบบเพื่อใช้ในการทดลองโดยใช้ Microsoft Visual C++ 6.0 และ Microsoft Visual C++.NET โดยจะออกแบบส่วนติดต่อผู้ใช้ (Graphic User Interface) ให้แสดงข้อมูลวีดิทัศน์แบบเวลาจริงก่อนและหลัง

กระบวนการสำคัญอันได้แก่ ข้อมูลวีดิทัศน์ขาเข้า (Input Video Sequence) ข้อมูลแบบจำลองฉากหลัง (Modelled Background) ข้อมูลหน้ากากสำหรับการจำแนก (Classification Mask) และผลลัพท์การจำแนก (Result or Foreground Object) ซึ่งมีลักษณะดังในรูปที่ 16

รูปที่ 16 ส่วนติดต่อผู้ใช้ของโปรแกรมต้นแบบที่ได้ออกแบบและพัฒนาขึ้น

ผลการทดลอง

ในการทดลองจะพิจารณาใช้ลำดับภาพที่มีรูปแบบเดียวกันคือ RGB 24-bits ขนาดความกว้าง 320 จุดภาพ สูง 240 โดยจะเป็นการทดลองเชิงเปรียบเทียบซึ่งอาศัยการเปรียบเทียบผลการทดลองที่สถานการณ์เดียวกัน หรือเฟรมวีดิทัศน์เดียวกันที่นิยมเรียกกันว่า ลำดับภาพเปรียบเทียบ (Benchmark Sequence) แต่ใช้วิธีการที่แตกต่างกันหรือใช้พารามิเตอร์ต่างกัน ซึ่งเป็นวิธีที่นิยมใช้กันอย่างแพร่หลายในการเปรียบเทียบประสิทธิภาพการค้นพบและแยกส่วนภาพวัตถุ โดยในการทดลองนี้จะทำการเปรียบเทียบระเบียบวิธีเพื่อหาผลของการปรับค่า สัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า ซึ่งทดลองในเงื่อนไขต่อไปนี้

1. ระเบียบวิธีที่ใช้แบบจำลองฉากหลังแบบสถิตย์ ซึ่งไม่มีการปรับค่าของฉากหลังตามเวลา
2. ระเบียบวิธีที่ใช้แบบจำลองฉากหลังแบบปรับค่าได้ แต่ไม่มีการใช้สัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า
3. ระเบียบวิธีที่ใช้แบบจำลองฉากหลังแบบปรับค่าได้ มีการใช้ช่วงระยะเฟรมปรับค่า แต่ไม่มีการใช้การใช้สัมประสิทธิ์วิวาซิตี
4. ระเบียบวิธีที่ใช้แบบจำลองฉากหลังแบบปรับค่าได้ มีการใช้สัมประสิทธิ์วิวาซิตี แต่ไม่มีการใช้ช่วงระยะเฟรมปรับค่า
5. ระเบียบวิธีที่ใช้แบบจำลองฉากหลังแบบปรับค่าได้ มีการใช้ทั้งสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า

(ก) (ข)

รูปที่ 17 (ก) ภาพแบบจำลองสถิตย์เริ่มต้นและ (ข) ภาพขาเข้า (Input Frame)

(ก) (ข) (ค)

รูปที่ 18 ผลการทดลองของระบบในกรณีแบบจำลองสถิตย์ (Static Background Model) (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

(ก) (ข) (ค)

รูปที่ 19 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยไม่ใช้สัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

(ก) (ข) (ค)

รูปที่ 20 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ช่วงระยะเฟรมปรับค่า แต่ไม่ใช้สัมประสิทธิ์วิวาซิตี (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

(ก) (ข) (ค)

รูปที่ 21 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้สัมประสิทธิ์วิวาซิตี แต่ไม่ใช้ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

(ก) (ข) (ค)

รูปที่ 22 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ทั้งสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

การทดลองจะใช้ลำดับภาพเปรียบเทียบสองสถานการณ์ ดังนี้

(ก) สถานการณ์แรก เป็นลำดับภาพห้องทดลองมีบุคคลปรากฏในฉาก ซึ่งเป็นวัตถุที่เราสนใจ โดยบุคคลเดินเข้ามาในฉากแล้วนำกล่อง มาวางไว้บนโต๊ะทำงาน จากนั้นจะนำกระเป๋าสะพายมาวางไว้ข้างๆ กล่อง สถานการณ์นี้จะมีการนำวัตถุทั้งสองซึ่งเป็นวัตถุนอกความสนใจเข้าฉากและออกจากฉากพร้อมทั้งเปลี่ยนตำแหน่งบ่อยครั้ง อีกทั้งการย้ายตำแหน่งของกระดาษบนโต๊ะทำงาน ซึ่งเป็นตัวอย่างของสถานการณ์การย้ายตำแหน่งของวัตถุในฉากหลัง (Relocation of Background Object) ซึ่งมีแบบจำลองสถิตย์เริ่มต้นและภาพขาเข้าดังแสดงในรูปที่ (17) เมื่อกำหนดให้อัตราการเรียนรู้ของการปรับค่า γ เป็น 0.05 อัตราการเรียนรู้วิวาซิตี ζ เป็น 0.1 และ ช่วงระยะเฟรมปรับค่า I_U เท่ากับ 8 เฟรม ได้ผลการทดลองดังแสดงในรูปที่ (18-22)

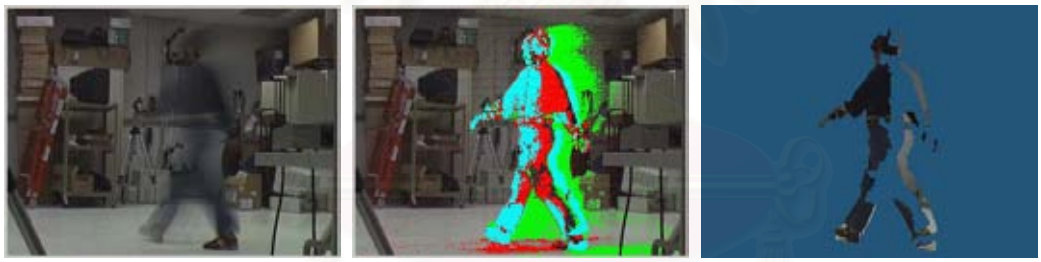
(ข) สถานการณ์ที่สองเป็นลำดับภาพในที่ร่มมีบุคคลปรากฏในฉาก ซึ่งเป็นวัตถุที่เราสนใจ หลังจากนั้นบุคคลดังกล่าวจะทำการปิดไฟในห้องนั้นจำนวนครึ่งหนึ่งของหลอดไฟทั้งหมด ส่งผลให้ความสว่างทั่วฉาก (Global Illumination) เกิดการเปลี่ยนแปลง หลังจากนั้นบุคคลดังกล่าวจะปรากฏในฉากอีกครั้งหนึ่ง ซึ่งลำดับภาพนี้ เป็นตัวอย่างของสถานการณ์การเปลี่ยนแปลงของความสว่าง (Illumination Changed) ซึ่งมีแบบจำลองสถิตย์เริ่มต้นและภาพขาเข้าดังแสดงในรูปที่ (23) เมื่อกำหนดให้อัตราการเรียนรู้ของการปรับค่า γ เป็น 0.1 อัตราการเรียนรู้วิวาซิตี ζ เป็น 0.2 และ ช่วงระยะเฟรมปรับค่า I_U เท่ากับ 3 เฟรมได้ผลการทดลองดังรูปที่ (24)-(28)

(ก) (ข)

รูปที่ 23) ภาพแบบจำลองสถิตย์เริ่มต้นและ (ข) ภาพขาเข้า (Input Frame)

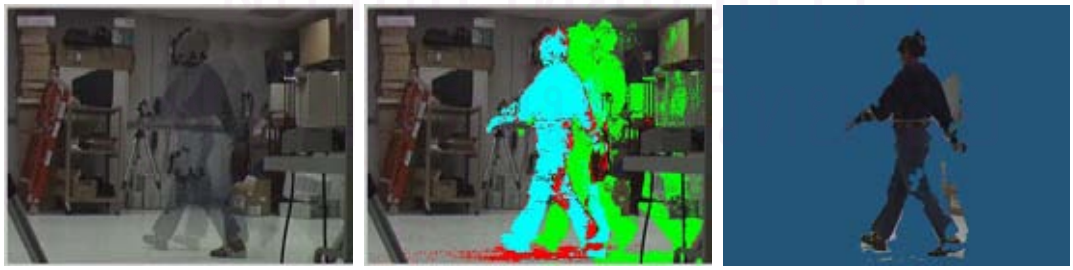
(ก) (ข) (ค)

รูปที่ 24 การทดลองของระบบในกรณีแบบจำลองสถิตย์ (Static Background Model) (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก



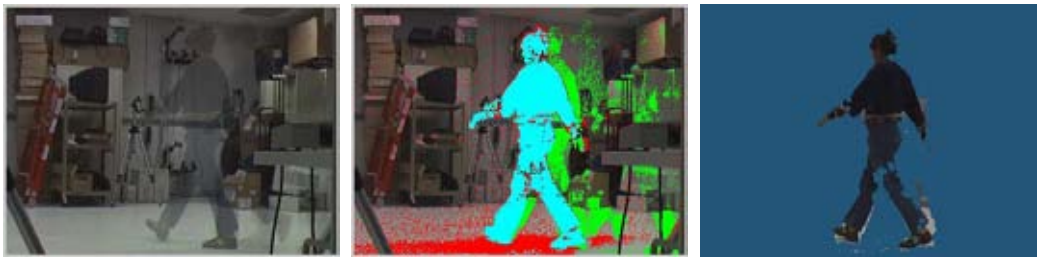
(ก) (ข) (ค)

รูปที่ 25 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยไม่ใช้สัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก



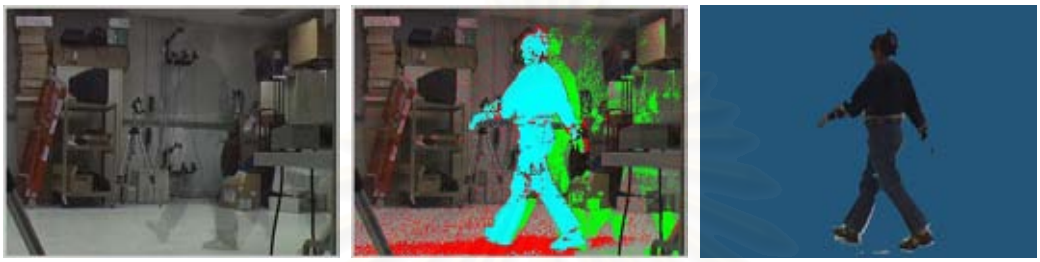
(ก) (ข) (ค)

รูปที่ 26 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ช่วงระยะเฟรมปรับค่า แต่ไม่ใช้สัมประสิทธิ์วิวาซิตี (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก



(ก) (ข) (ค)

รูปที่ 27 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้สัมประสิทธิ์วิวาซิตี แต่ไม่ใช้ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก



(ก) (ข) (ค)

รูปที่ 28 ผลการทดลองของระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ทั้งสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่า (ก) แบบจำลอง (ข) หน้ากากการจำแนก (ค) ผลการจำแนก

วิเคราะห์และสรุปผลการทดลอง

(ก) ในสถานการณ์ของการเคลื่อนย้ายตำแหน่งของวัตถุในฉากหลัง ผลการทดลองชี้ให้เห็นว่า

1. ระบบในกรณีแบบจำลองสถิตย์จะจำแนกกล่องซึ่งเป็นวัตถุที่ถูกเพิ่มเข้ามาในฉากให้เป็นฉากหน้าหรือวัตถุที่สนใจเสมอ ไม่ว่าเวลาจะผ่านไปนานเท่าไหร่ ซึ่งผิดวัตถุประสงค์ของการตรวจจับวัตถุที่เคลื่อนไหว และก่อให้เกิดปัญหาเมื่อนำไปประมวลผลต่อไปดังที่ได้กล่าวไว้ในตอนต้น

2. ระบบในกรณีแบบจำลองปรับค่าได้ โดยไม่ใช้สัมประสิทธิ์วิวาซิตี จะกำหนดให้กล่องซึ่งเปลี่ยนตำแหน่งไปแล้วเป็นระยะเวลาค่าหนึ่งเป็นฉากหลัง แต่เนื่องด้วย ในสถานการณ์ที่แสดงผล บุคคลได้หยุดนิ่งชั่วขณะเพื่อวางกระเป๋าทำให้ระบบที่มีอัตราการปรับค่าที่เท่ากันทุกๆจุดภาพ และมากเพียงพอที่จะเรียนรู้ให้กล่องเป็นฉากหลัง จำแนกบางส่วนของบุคคลเป็นฉากหลังด้วย โดยบุคคลที่อยู่นิ่งจะถูกกำหนดให้เป็นฉากหลังอย่างรวดเร็ว ซึ่งมีความขัดแย้งกันระหว่างความต้องการที่จะมีอัตราการเรียนรู้ที่มากที่สุดเพื่อจะจำแนกวัตถุที่อยู่นิ่งให้กลายเป็นฉากหลังให้เร็วที่สุด และ ความต้องการที่จะมีอัตราการเรียนรู้ที่น้อยที่สุดเพื่อจะคงวัตถุในฉากหน้าไว้ จึงเป็นการยากที่จะหาค่าอัตราการเรียนรู้ของการปรับค่าที่เหมาะสมระบบที่มีอัตราการปรับค่าที่เท่ากันทุกๆจุดภาพ และไม่ยืดหยุ่นต่อสถานการณ์ที่เปลี่ยนไป

3. ระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ช่วงระยะเฟรมปรับค่า แต่ไม่ใช้สัมประสิทธิ์วิวาซิตีจะให้ผลการทดลองที่ดีกว่า ซึ่งระบบจำปรับค่าให้กล่องเป็นฉากหลังได้อย่างรวดเร็ว และคงค่าบุคคลที่เคลื่อนที่ไว้เป็นฉากหน้า เนื่องจากตำแหน่งของบุคคลที่เคลื่อนไหวอยู่เป็นระยะเวลาน้อยกว่าค่า ระยะเฟรมปรับค่า I_U ระบบจึงยังไม่ทำการปรับค่าให้เป็นฉากหลัง ซึ่งตรงกับกรณีนี้ แต่ระบบจะทำงานได้ผิดพลาดในกรณีของวัตถุที่มีความซับซ้อนต่ำ ดังเช่นบุคคลในข้างต้นสวมเสื้อสีขาว ถึงแม้วัตถุจะเคลื่อนที่แต่ค่าสี ณ จุดภาพที่พิจารณาจะเปลี่ยนแปลงน้อยมาก เมื่อค่าสี ไม่เกิดการเปลี่ยนแปลงเกินค่า ระยะเฟรมปรับค่า I_U

ระบบจะจำแนกให้จุดที่พิจารณาเป็นฉากหลังทันที ทำให้เกิดปัญหาในกรณีวัตถุที่สนใจมีขนาดใหญ่เมื่อเทียบกับขนาดภาพ และ มีลวดลายซับซ้อนต่ำ

4. ระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้สัมประสิทธิ์วิวาซิตี แต่ไม่ใช้ช่วงระยะเฟรมปรับค่าจะทำงานได้อย่างน่าพอใจบริเวณบุคคล แต่เนื่องการลดค่า อัตราการเรียนรู้ ในสมการที่ (25) อัตราการเรียนรู้ใหม่ $(1-\upsilon_{i,j}[n])\gamma$ จะมีค่าน้อยกว่าหรือเท่ากับอัตราการเรียนรู้เดิม γ เสมอ ผลของการเคลื่อนไหวทำให้อัตราการเรียนรู้ลดลง จะเห็นได้จากความเร็วที่ระบบจำแนกให้กล่องเป็นฉากหลังจะช้ากว่ากรณีข้างต้น ระบบจึงต้องการอัตราการเรียนรู้ที่มากกว่า กรณีที่ไม่ใช้สัมประสิทธิ์วิวาซิตี

5. แบบจำลองปรับค่าได้ โดยใช้ทั้งสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่าจะทำงานได้อย่างน่าพอใจ เนื่องจาก ช่วงระยะเฟรมปรับค่าจะช่วยคงค่าฉากหน้าที่สนใจไว้ และ ปรับค่ากล่องให้เป็นฉากหลังอย่างรวดเร็ว ดังกรณีข้างต้นและ สัมประสิทธิ์วิวาซิตี จะลดอัตราการปรับค่า ในบริเวณที่มีการเคลื่อนไหวทุกๆเฟรม ซึ่งจะช่วงคงค่าฉากหลัง ไม่ให้ถูกปรับค่าเร็วเกินนัก ในกรณี วัตถุที่สนใจอยู่เกิน I_U

(ข) ในสถานการณ์ของการเปลี่ยนแปลงความสว่าง ผลการทดลองชี้ให้เห็นว่า

1. เมื่อความสว่างลดน้อยลงระบบจะตรวจพบการเปลี่ยนแปลงของเวกเตอร์ค่าสี ณ ทุกจุดภาพ แบบจำลองสถิตย์จะจำแนกจุดภาพที่องค์ประกอบของความสว่างลดลงแต่ค่าสีคงที่ซึ่งสอดคล้องในเงื่อนไขในสมการที่ (17) ให้เป็นเงา ไม่ว่าเวลาจะเปลี่ยนไปเท่าใด ตราบที่ความสว่างไม่เปลี่ยนแปลง ระบบก็จะจำแนกฉากหลังให้เป็นเงาเสมอไป ดังแสดงในรูปที่ 24 (ข) ในกรณีนี้ถึงแม้จะมีการจำแนกจุดภาพซึ่งควรจะเป็นฉากหลังให้เป็นเงา แต่ผลที่ได้อยู่ในระดับที่น่าพอใจ ดังในรูปที่ 24 (ค)

2. ระบบในกรณีแบบจำลองปรับค่าได้ โดยไม่ใช้สัมประสิทธิ์วิวาซิตีจะสามารถปรับค่าแบบจำลองฉากหลังให้มีความสว่างลดลงตามภาพขาเข้าที่มีความสว่างลดลงได้ แต่อัตราการเรียนรู้ที่มากเพื่อการปรับตัวทางแสงจะทำให้บางส่วนของวัตถุถูกจำแนกเป็นฉากหลัง ซึ่งสถานการณ์ของภาพขาเข้าในรูปที่ (23) เครื่องแต่งกายของบุคคลมีลวดลายน้อยมาก ระบบจึงไม่ตรวจพบการเปลี่ยนแปลงบริเวณภายในของภาพบุคคล ผลที่ได้จึงอยู่ในระดับที่ไม่น่าพอใจ

3. ระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้ช่วงระยะเฟรมปรับค่า แต่ไม่ใช้สัมประสิทธิ์วิวาซิตีจะปรับตัวตามการเปลี่ยนแปลงของความสว่างได้ดี ก็ต่อเมื่อมีค่า I_U น้อย โดยในรูปที่ 26 ใช้ค่า I_U เป็น 3 เฟรมจะเห็นได้ว่า บางส่วนของวัตถุยังคงถูกจำแนกให้เป็นฉากหลังเนื่องจากในสถานการณ์ข้างต้นถึงแม้บุคคลจะเคลื่อนไหวอยู่ตลอดเวลา แต่การที่เครื่องแต่งกายของบุคคลมีลวดลายน้อย ทำให้ระบบไม่ตรวจพบการเปลี่ยนแปลงเกินค่าระยะเฟรมปรับค่า I_U ทำให้บุคคลถูกจำแนกให้เป็นฉากหลังอย่างรวดเร็ว ค่า I_U ที่น้อยเกินไป จึงไม่สามารถแก้ปัญหาอันเนื่องมาจากการเปลี่ยนแปลงของความสว่างได้ ทางเลือกจึงอยู่ที่การเพิ่มค่า I_U แต่ระบบจำปรับตัวเข้าสู่ระดับความสว่างใหม่ได้ช้ากว่า

4. ระบบในกรณีแบบจำลองปรับค่าได้ โดยใช้สัมประสิทธิ์วิวาซิตี แต่ไม่ใช้ช่วงระยะเฟรมปรับค่าจะทำงานได้อย่างน่าพอใจบริเวณบุคคล จะปรับตัวเข้าสู่ระดับความสว่างใหม่ได้ช้ากว่า 2 กรณีข้างต้น เนื่องจากคุณสมบัติที่ได้กล่าวไว้ในข้อ ก-4 แต่ให้ผลที่อยู่ในระดับที่น่าพอใจ

5. แบบจำลองปรับค่าได้ โดยใช้ทั้งสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่าจะทำงานได้อย่างน่าพอใจ แต่มีอัตราการปรับตัวเข้าสู่ระดับความสว่างใหม่ได้ช้ากว่า กรณี ข—4 เนื่องจากผลของสัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรม จะทำให้อัตราการเรียนรู้โดยเฉลี่ยลดลง ไม่กระทบต่อผลการจำแนกฉากหน้า เนื่องจากระบบจะจำแนกส่วนที่ยังไม่ปรับค่าเข้าสู่ระดับความสว่างใหม่ให้เป็นเงา

จากการทดลองข้างต้นชี้ให้เห็นว่า สัมประสิทธิ์วิวาซิตี และ ช่วงระยะเฟรมปรับค่ามีผลต่อการทำงานของระบบเป็นอย่างมาก ในบทบาทที่คงค่าพารามิเตอร์ของแบบจำลองฉากหลังไม่ให้เปลี่ยนเร็วเกินไปเมื่อมีการเปลี่ยนแปลงในฉากจากลำดับภาพขาเข้า อนึ่งการกำหนดค่าพารามิเตอร์ควบคุมอันได้แก่ อัตราการเรียนรู้ของการปรับค่า γ อัตราการเรียนรู้วิวาซิตี ζ และ ช่วงระยะเฟรมปรับค่า I_U มักต้องคำนึงถึงสภาวะแวดล้อมที่ใช้งาน สถานการณ์ที่มีการเปลี่ยนแปลงในฉากมาก ต้องกำหนดให้ค่าอัตราการเรียนรู้ของการปรับค่า γ และอัตราการเรียนรู้วิวาซิตี ζ มีค่ามาก และ ช่วงระยะเฟรมปรับค่า I_U มีค่าน้อย

โดยจากการทดลองการเปลี่ยนแปลงของค่าช่วงระยะเฟรมปรับค่าจะมีผลต่อความเร็วของการปรับตัวของฉากหลังมากกว่าสัมประสิทธิ์วิวาซิตี เนื่องจากการทำงานของค่าช่วงระยะเฟรมจะมีการปรับตัวที่อัตราการเรียนรู้เท่ากับ อัตราการเรียนรู้เริ่มต้น γ ณ จุดภาพที่ไม่มีการเปลี่ยนแปลงเกินค่า I_U และเป็นการปรับตัวที่อัตราเท่ากันทุกจุดภาพ ในขณะที่สัมประสิทธิ์วิวาซิตี ทำงานโดยที่อัตราการเรียนรู้เท่ากับ $(1-\upsilon_{i,j}[n])\gamma$ ซึ่งน้อยกว่าและเป็นการปรับตัวที่ขึ้นอยู่กับระดับการเปลี่ยนแปลงค่าสีของแต่ละจุดภาพ ดังนั้นเราจึงควบคุมอัตราการเรียนรู้การปรับค่าของระบบโดยการปรับค่า I_U เป็นการปรับค่าแบบหยาบ (Coarse Adjustment) และปรับค่าโดยใช้สัมประสิทธิ์วิวาซิตีเป็นการปรับค่าแบบละเอียด (Fine Adjustment)

การวัดประสิทธิภาพการทำงานของระบบ

หลังจากที่ได้พัฒนาโปรแกรมต้นแบบเพื่อทำการทดสอบและยืนยันระเบียบวิธีที่ได้นำเสนอ เพื่อแสดงถึงประสิทธิภาพการทำงาน เราจึงต้องกำหนดบรรทัดฐานในการวัดประสิทธิภาพ การวัดประสิทธิภาพของระบบแยกส่วนภาพเวลาจริงนั้น ต้องคำนึงถึงวัตถุประสงค์ของระบบเป็นหลัก ซึ่งจะแบ่งรูปแบบของการวัดประสิทธิภาพออกเป็น 2 กรณีคือ

1. **การวัดประสิทธิภาพเชิงคุณภาพ** ในกรณีนี้การวัดประสิทธิภาพจะเน้นถึงความสามารถในการจัดการและแก้ปัญหาปัญหาที่ได้กล่าวไว้ตอนต้น รวมทั้ง ความทนทานต่อสัญญาณรบกวน (Robustness to Noise) และ ความซับซ้อนในการคำนวณ (Computational Complexity)
2. **การวัดประสิทธิภาพเชิงปริมาณ** ในกรณีนี้การวัดประสิทธิภาพจะเน้นถึงความสามารถในการทำงานตามวัตถุประสงค์ของระบบ อันได้แก่ อัตราการค้นพบ (Detection Rate) และ อัตราการฟ้องความผิดพลาด (False Alarm Rate) การวัดแบบ PDR ที่เสนอโดย [16] รวมทั้งอัตราเฟรม (Frame-rate)

การวัดประสิทธิภาพการแยกส่วนภาพมักกระทำโดยการเปรียบเทียบกับภาพอ้างอิง (ground Truth Result) ซึ่งถือว่าเป็นผลการทดลองในอุดมคติ (Ideal Result) และผลการทดลองตามอุดมคตินี้ จะถือว่ามีความใกล้เคียงกับกระบวนการแยกส่วนภาพด้วยตาของมนุษย์มากที่สุด โดยการสร้างภาพอ้างอิงนี้มักให้มนุษย์เป็นผู้จำแนกกลุ่มจุดภาพจากภาพด้านเข้า ซึ่งในการทดลองนี้ทำโดยการลากเส้นแบ่งขอบเขตในโปรแกรมประยุกต์ด้านการประมวลผลภาพทั่วไป โดยผลที่ผ่านการจำแนกโดยตามนุษย์ซึ่งถือว่าวัตถุที่สนใจ คือบุคคลในภาพและสิ่งของที่บุคคลถือดังรูปที่ 1

รูปที่ 29 (ก) ภาพด้านเข้าต้นแบบซึ่งนำมาผ่านระบบการแยกส่วนภาพ (ข) ผลการแยกส่วนภาพอ้างอิง ซึ่งทำการแยกส่วนโดยตามนุษย์และจำแนกโดยใช้การแบ่งในโปรแกรมประยุกต์ด้วนการประมวลผลภาพทั่วไป

อย่างไรก็ดี เนื่องจากภาพผลการทดลองอ้างอิงที่สร้างขึ้นนี้ ขึ้นอยู่กับระบบการมองเห็นของมนุษย์แต่ละคนซึ่งให้ผลไม่เท่ากัน ดังนั้นอาจทำการหาภาพอ้างอิงจากผู้ทดลองที่มีจำนวนมากขึ้น ซึ่งจะให้ผลการทดลองทางสถิติที่น่าเชื่อถือกว่า

การวัดประสิทธิภาพเชิงปริมาณ

เนื่องจากการเปรียบเทียบเชิงจิตพิสัย (Subjective Comparison) สามารถบอกประสิทธิภาพได้ในระดับหนึ่งแต่ไม่สามารถเปรียบเทียบผลการทดลองจากต่างๆวิธีกันได้อย่างเด่นชัด และขึ้นอยู่กับผู้วัดเนื่องจากความสามารถในการแยกแยะด้วยตาของแต่ละบุคคลไม่เท่ากัน การเปรียบเทียบเชิงวัตถุวิสัย จึงน่าจะเป็นทางเลือกที่ให้ผลน่าเชื่อถือมากกว่า เนื่องจากมีผลเป็นเชิงปริมาณ นำมาใช้ในการเปรียบเทียบระหว่างผลการทดลองจากหลายๆวิธีได้

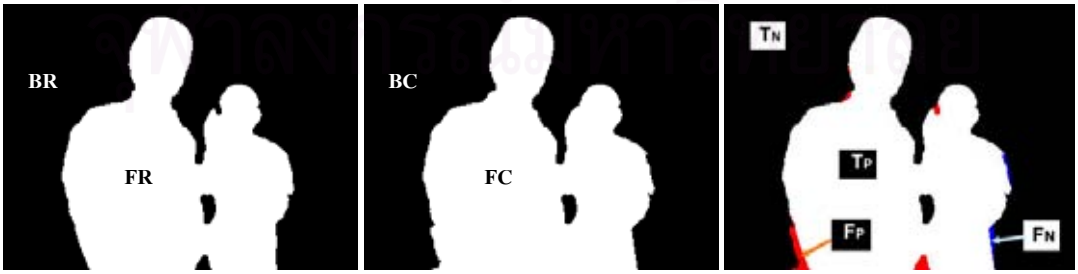
นิยามปัญหา

FR เป็นเซตของจุดภาพของผลการทดลองอ้างอิงซึ่งถูกจำแนกให้เป็นฉากหน้า (Reference Foreground)

FC เป็นเซตของจุดภาพของผลการทดลองที่พิจารณาซึ่งถูกจำแนกให้เป็นฉากหน้า (Considered Foreground)

BR เป็นเซตของจุดภาพของผลการทดลองอ้างอิงซึ่งถูกจำแนกให้เป็นฉากหลัง (Reference Background)

BC เป็นเซตของจุดภาพของผลการทดลองที่พิจารณาซึ่งถูกจำแนกให้เป็นฉากหลัง (Considered Background)



รูปที่ 30 เซตการจำแนก (ก) ผลการทดลองอ้างอิง (ข) ผลการทดลองที่พิจารณา (ค) ประสิทธิภาพการจำแนก เซตการจำแนกถูกและจำแนกผิด

T_p เป็นเซตของจุดภาพที่ผลการทดลองที่พิจารณาจำแนกตรงกับผลการทดลองอ้างอิงให้เป็นฉากหน้า (True Positive)

T_N เป็นเซตของจุดภาพที่ผลการทดลองที่พิจารณาจำแนกตรงกับผลการทดลองอ้างอิงให้เป็นฉากหลัง (True Negative)

F_P เซตของจุดภาพที่ผลการทดลองที่พิจารณาจำแนกเป็นฉากหน้า แต่ผลการทดลองอ้างอิงจำแนกเป็นฉากหลัง (False Positive)

F_N เซตของจุดภาพที่ผลการทดลองที่พิจารณาจำแนกเป็นฉากหลัง แต่ผลการทดลองอ้างอิงจำแนกเป็นฉากหน้า (False Negative)

ซึ่งดัชนีที่ใช้บ่งวัดได้แก่

อัตราการจำแนกถูกแบบบวก (RT_p) เป็นจำนวนจุดภาพที่ผลการทดลองที่พิจารณาจำแนกตรงกับผลการทดลองอ้างอิงให้เป็นฉากหน้าจากจุดภาพทั้งหมด (กว้างWxสูงH)

อัตราการจำแนกถูกแบบลบ (RT_N) เป็นจำนวนจุดภาพที่ผลการทดลองที่พิจารณาจำแนกตรงกับผลการทดลองอ้างอิงให้เป็นฉากหลัง จากจุดภาพทั้งหมด (กว้างWxสูงH)

อัตราการจำแนกผิดแบบบวก (RF_p) เป็นจำนวนจุดภาพที่พิจารณาจำแนกเป็นฉากหน้า แต่ผลการทดลองอ้างอิงจำแนกเป็นฉากหลังจากจุดภาพทั้งหมด (กว้างWxสูงH)

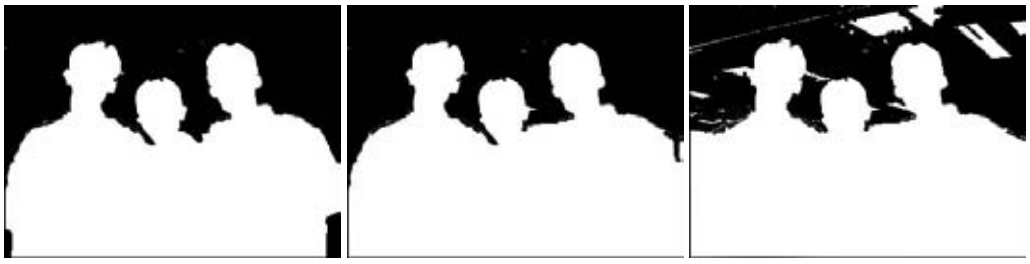
อัตราการจำแนกถูกแบบลบ (RF_N) เป็นจำนวนจุดภาพที่ผลการทดลองที่พิจารณาจำแนกเป็นฉากหลัง แต่ผลการทดลองอ้างอิงจำแนกเป็นฉากหน้าจากจุดภาพทั้งหมด (กว้างWxสูงH)

จากดัชนีชี้วัดประสิทธิภาพทั้ง 4 สามารถกล่าวได้ว่า ระบบที่มีประสิทธิภาพสูงจะมี อัตราการจำแนกถูกสูง และมีอัตราการจำแนกผิดต่ำแต่การนำเอาอัตราทั้ง 4 มาเปรียบเทียบกันโดยตรงไม่สามารถทำได้เนื่องจากค่าทั้ง 4 ขึ้นอยู่กับขนาดของภาพ โดยถึงแม้บริเวณที่เราสนใจมีการจำแนกเหมือนกัน แต่ขนาดของภาพต่างกัน ค่าอัตราทั้ง 4 ก็มีค่าต่างกันดังนั้น จึงมีผู้เสนอให้ใช้ อัตราการค้นพบ (Detection Rate : DR) และ อัตราการฟ้องความผิดพลาด (False Alarm Rate : FAR) ดังนี้

$$Detection\ \ Rate: DR = \frac{RT_p}{RT_p + RF_N} \tag{1}$$

โดยที่ DR สามารถบ่งชี้ประสิทธิภาพของการค้นพบฉากหน้า ว่าผลการทดลองจากวิธีที่พิจารณาสามารถค้นพบฉากหน้าได้ ครบถ้วนเพียงใด ถ้า DR มาก บ่งชี้การค้นพบฉากหน้าได้อย่างครบถ้วน

ถึงอย่างไรก็ดี เพียงลำพังDR ไม่สามารถชี้ประสิทธิภาพได้ทั้งหมด เนื่องจากผลการทดลองที่มีพื้นที่ ที่ซ้อนทับกับผลการทดลองอ้างอิงเท่ากัน จะมีค่า DR เท่ากัน แต่ตามนุษย์จะสามารถสังเกตเห็นส่วนเกินที่ไม่ใช่พื้นที่ที่ต้องการได้ ซึ่งคือ F_p นั่นเอง ดังในรูปที่ 3



(a) (b) (c)

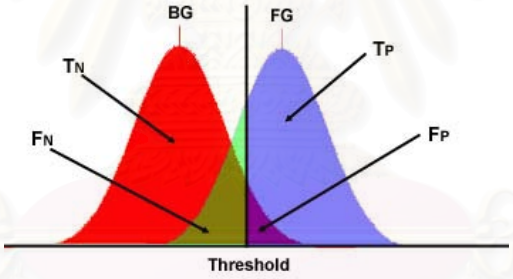
(d) (e)

รูปที่ 31 ผลการแยกส่วนภาพซึ่งได้ค่าอัตราการค้นพบใกล้เคียงกัน (a) ผลการทดลองอ้างอิง (a) ผลการทดลองสมมุติที่ 1 (b) ผลการทดลองสมมุตที่ 2 (d) ผลการวัดประสิทธิภาพของผลการทดลองสมมุติที่ 1 (e) ผลการวัดประสิทธิภาพของผลการทดลองสมมุติที่ 2

จากที่กล่าวข้างต้น เปรียบเสมือนว่าอัตราการค้นพบ (Detection rate) เป็นปริมาณที่ใช้บ่งชี้ความแม่นยำในการจำแนกฉากหน้า ซึ่งละเลยฉากหลังไว้ จึงต้องมีปริมาณซึ่งชี้วัดความแม่นยำของระบบ ซึ่งจำแนกฉากหลังนั่นคือ อัตราการฟ้องความผิดพลาด (False Alarm Rate : FAR)

$$False\ Alarm\ \ Rate: FAR = \frac{RF_p}{RT_p + RF_P}$$

โดยที่ FAR สามารถบ่งชี้ประสิทธิภาพของการค้นพบฉากหน้าโดยไม่จำแนกฉากหลังมาปะปนด้วย ถ้า FAR มาก บ่งชี้การจำแนกฉากหลังผิดเป็นฉากหน้า ซึ่งระบบจะมีประสิทธิภาพน้อยกว่าระบบที่มี FAR น้อย



รูปที่ 32 การจำแนกฉากหน้าและฉากหลัง โดยผลการจำแนกจะถูกวัดประสิทธิภาพและถูกแบ่งออกเป็น 4 กลุ่ม

การพัฒนาโปรแกรมประยุกต์เพื่อวัดประสิทธิภาพ

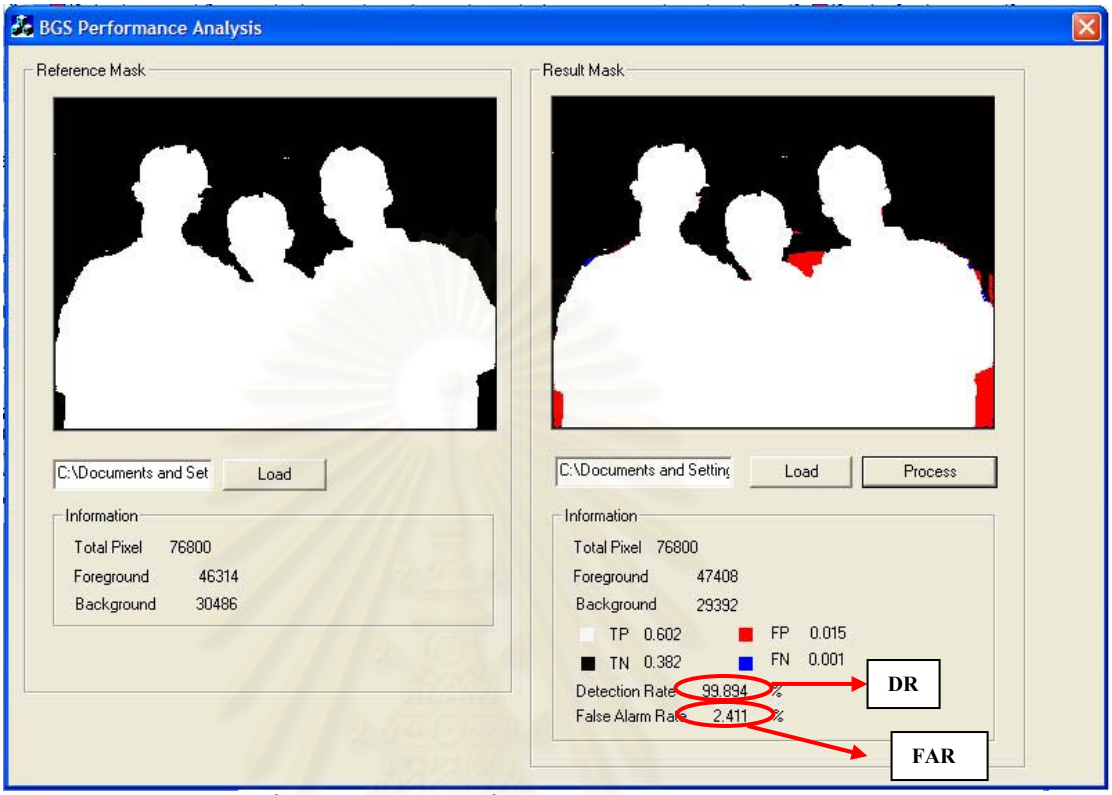
หลังจากอาศัยหลักการข้างต้น สามารถสร้างโปรแกรมวัดประสิทธิภาพของระบบได้โดย พัฒนาจาก Microsoft Visual C++ ซึ่งมีส่วนติดต่อผู้ใช้ดังนี้

การใช้งานโปรแกรมประยุกต์

การใช้งานจะอาศัยภาพหน้ากากของผลการทดลองอ้างอิงและหน้ากากของผลการทดลองจากแต่ละวิธีที่ต้องการวัดประสิทธิภาพโดยหน้ากากนี้จะมีรูปแบบเป็นแฟ้มข้อมูล บิตแมพ 24 บิตต่อจุดภาพ ซึ่งมีค่าสีเป็น สีดำ (0,0,0) ณ บริเวณที่จำแนกเป็นฉากหลัง และสีขาว (255,255,255) ณ บริเวณที่จำแนกเป็นฉากหน้า โดยเริ่มต้นผู้ทดลองจะต้องทำการโหลดข้อมูลหน้ากากของผลการทดลองอ้างอิงเพื่อนำมาใช้เปรียบเทียบกับผลการทดลองวิธีต่างๆ จากนั้นจะทำการโหลดภาพหน้ากากผลการทดลองที่พิจารณาเข้ามา และกดปุ่ม process เพื่อประมวลผล

ผลการประมวลผลที่ได้ในภาพจะมีพื้นที่การประมวลผล 4 พื้นที่ คือ T_p (ขาว), T_N (ดำ) , F_P (แดง), F_N (น้ำเงิน) ตามหลักการที่ได้กล่าวมาแล้ว

จากนั้นโปรแกรมจำทำการคำนวณ RT_p, RT_N ,RF_p ,RF_N และ ค่า DR รวมทั้ง FAR ตามต้องการ ซึ่งสามารถนำค่านี้มาใช้เปรียบเทียบกับแต่ละวิธี



รูปที่ 33 ปรแกรมประยุกเพื่อวัดประสิทธิภาพเชิงปริมาณ

วิเคราะห์การวัดประสิทธิภาพ

วิธีการวัดประสิทธิภาพเชิงประมาณให้เป็นแน่นอนและเป็นรูปธรรมมากกว่าวิธีเชิงคุณภาพ ซึ่ง มักวัดโดยใช้จิตพิสัย (Subjective Test) ซึ่งไม่สามารถบ่งชี้เป็นปริมาณเพื่อใช้เปรียบได้ อย่างไรก็ดี จุดอ่อนของวีธีการนี้อยู่ที่การสร้างหน้ากากผลการทดลองอ้างอิงซึ่ง ยังใช้มนุษย์เป็นผู้สร้างขึ้น ซึ่งตาแต่ละคนมีความสามารถจำแนกไม่เท่ากัน ดังนั้นจึงมีความเหมาะสมที่จะนำวิธี PDR[]สร้างแบบจำลองอ้างอิงซึ่งจะนำมาพิจารณาต่อไป

ผลสัมฤทธิ์ของแผนงานวิจัย

1. ได้โปรแกรมประยุกต์การแยกส่วนภาพโดยใช้วิธีการทางสถิติเชิงพาราเมตริกแบบปรับตัวได้ และระบบสามารถแก้ปัญหาอันเนื่องจากการเปลี่ยนแปลงของฉากหลัง การเปลี่ยนแปลงของความสว่างทั่วทั้งฉาก การเปลี่ยนแปลงของเงาได้

2. ได้โปรแกรมประยุกต์การวัดประสิทธิภาพเชิงปริมาณ โดยมีผลการทดลองอ้างอิง (Ground Truth) ที่สร้างขึ้นจากการตัดสินใจของมนุษย์เป็นผลการทดลองเปรียบเทียบ

3. ผลพลอยได้จากการพัฒนาโปรแกรมประยุกต์ได้องค์ความรู้และไลบรารีการประมวลผลภาพและวีดิทัศน์เพื่อนำไปใช้พัฒนาในงานที่เกี่ยวข้องต่อไป

4. บทความทางวิชาการ การประชุมทางวิชาการ International Conference on Multimedia and Expo 2004 ณ กรุงไทเป ประเทศไต้หวัน

เอกสารอ้างอิง

[1] J. Zaletelj and J. F. Tasič, “Video Object Segmentation Based on Edge Tracking,” In *Proceedings ICIP 2001*, pp. 813-816, 2001.

[2] A.C. Hurlbert, "The Computation of Color", *Technical Report*, MIT Artificial Intelligence Laboratory, 1995.

[3] O. Javed, K. Shafique and M. Shah, “A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information,” In *IEEE Workshop on Motion and Video Computing*, Orlando, 2002.

[4] R. Jain, D. Militzer, and H. Nagel. “Separating Nonstationary from Stationary Scene Components in a Sequence of Real World TV-Images,” In *IJCAI* 1977, pp. 612–618, 1977.

[5] C. Wren, A Azarbayejani, T. Darrel, and A. Pentland. “ Pfinder, Real Time Tracking of the Human Body,” In *IEEE Trans. on PAMI*, 19(7), 1997.

[6] I. Haritaoglu, D. Harwood, and L.S. Davis. “W4: Who? When? Where? What? a Real-time System for Detecting and Tracking People,” In *Proc. the thrid IEEE Int’l Conf. Automatic Face and Gesture Recognition)*, pp. 222–227. IEEE Computer Society Press, 1998.

[7] C. Stauffer and W. Grimson. “Learning Patterns of Acitivty using Real-time Tracking,” In *IEEE Trans. on PAMI*, 22(8):747–757, 2000.

[8] C. Stauffer and W. Grimson, “Adaptive Background Mixture Models for Real-time Tracking,” In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 246-252, 1999.

[9] G. Gordon, T. Darrell, M. Harville, J. Woodfill “ Background Estimation and Removal based on Range and Color,” In *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[10] M. Harville. “ A Framework for High-Level Feedback to Adaptive Per-Pixel Mixture of Gaussian Models,” In *Proceedings of European Conference on Computer Vision*, 2002.

[11] X. Gao, T.E. Boult, F. Coetzee, and V. Ramesh. “ Error Analysis of Background Subtraction,” In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2000.

[12] X. Gao, J. Zhou, “Adaptive Background Estimation for Real-time Traffic Monitoring,” In *Proceeding of IEEE Intelligent Transportation Systems Conference*, 2001

[13] L. Liyuan and L. Maylor. “Integrating Intensity and Texture Differences for Robust Change Detection,” *IEEE Trans. on Image Processing*, pp.105–112, 2002.

[14] M. Greiffenhagen, V. Ramesh, and H. Nieman. “ Thesystematic Design and Analysis of a Vision System: A Case Study in Video Surveillance,” In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2001.

[15] T. Horprasert, D. Harwood and L. S. Davis, “A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection,” In *Proc. IEEE ICCV'99 FRAME-RATE Workshop*, Kerkyra, Greece, 1999

[16] T. H. Chalidabhongse, K. Kim, D. Harwood and L. Davis, "A Perturbation Method for Evaluating Background Subtraction Algorithms", In I*EEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนงานที่ 3
การเข้ารหัสสัญญาณวีดิทัศน์แยกส่วน และ การประกอบสัญญาณวีดิทัศน์แยกส่วนเข้ากับฉากหลังที่เป็นภาพสังเคราะห์

ส่วนงานที่ได้ดำเนินการในส่วนนี้

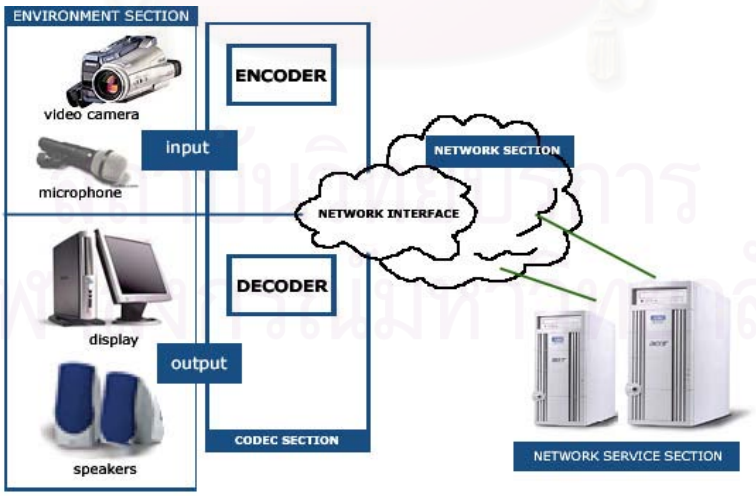
1. ศึกษาการแยกส่วนวัตถุบริเวณที่สนใจ (มนุษย์) เพื่อนำมาทำการเข้ารหัสเชิงวัตถุตามมาตรฐาน MPEG-4
2. สร้างแบบจำลองที่ใช้ในการตรวจจับสีผิวของมนุษย์
3. ศึกษาการแยกองค์ประกอบต่าง ๆ ที่สำคัญบนใบหน้ามนุษย์ เช่น ตา, จมูก, ปาก ฯลฯ
4. ศึกษาและสร้างแบบจำลองที่ใช้สำหรับการติดตามวัตถุ เมื่อมีการเคลื่อนที่ไปยังตำแหน่งต่าง ๆ

กล่าวนำ

มาตรฐานการเข้ารหัสวีดิทัศน์ MPEG-4 จะมีส่วนที่เพิ่มเติมจากมาตรฐานก่อนหน้านี้ (MPEG-2) คือ สามารถทำการเข้ารหัสเชิงวัตถุ ตัวอย่างเช่น ทำการเข้ารหัสวัตถุเคลื่อนไหว (moving object) แยกจากฉากหลัง (background) ซึ่งการเข้ารหัสเชิงวัตถุนี้ทำให้สอดคล้องกับการนำไปใช้งานสื่อสารวีดิทัศน์มากขึ้น

ซึ่งในงานวิจัยนี้จะดำเนินงานในส่วนกระบวนการก่อนหน้า (Pre-processing) การเข้ารหัส คือ การแยกส่วนภาพจากวีดิทัศน์ออกเป็นวัตถุ เพื่อนำไปเข้ารหัสเฉพาะบริเวณ โดยมีวัตถุประสงค์เพื่อนำไปประยุกต์ใช้งานเกี่ยวกับการประชุมสัมมนาวีดิทัศน์ทางไกล (Video Conference) อาทิเช่น การแยกบริเวณส่วนที่เคลื่อนไหวออกจากฉากหลัง, การแยกองค์ประกอบที่สำคัญบริเวณใบหน้ามนุษย์

โดยแบ่งเป็นส่วนต่าง ๆ ที่นำเสนอดังนี้ ทฤษฎีและหลักการ, สิ่งที่นำเสนอในงานวิจัย, ผลการทดลองและสรุปผล



รูปที่ 1 โครงสร้างของระบบการประชุมผ่านสัญญาณวิดีทัศน์

ทฤษฎีและหลักการ

1.ปริภูมิสี

ปริภูมิสี มีมากมายหลายแบบ แต่ละแบบมีข้อดีข้อเสียแตกต่างกันไป และมีความเหมาะสมในการนำมาใช้งานลักษณะต่าง ๆ กัน ปริภูมิสีที่นิยมใช้และมักพบบ่อย คือ ปริภูมิสี RGB ซึ่งปริภูมิสีอื่น ๆ ส่วนมากจะได้จากการแปลง RGB โดยคำนึงถึงวัตถุประสงค์ที่นำมาใช้งาน เช่น ปริภูมิสี CMYK ซึ่งให้แหล่งกำเนิดสีอยู่ที่การซึมซับของหมึกพิมพ์บนกระดาษนำมาใช้สำหรับเครื่องพิมพ์, ปริภูมิสี L*a*b สร้างเพื่อใช้เป็นมาตรฐานในการวัดสี ครอบคลุมทุกสีใน RGB และ CMYK เป็นต้น

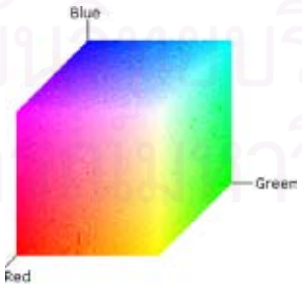
เมื่อพิจารณาระบบการมองเห็นของมนุษย์ กล่าวคือ มนุษย์มีความสามารถในการรับรู้ทางความสว่างและสีได้ดีไม่เท่ากัน โดยสายตามนุษย์จะมีความไวต่อความสว่างมากกว่าสี ในที่นี้จึงขอกล่าวรายละเอียดเฉพาะปริภูมิสีที่มีการแยกระหว่างความสว่าง (Luminance) และสี (Chrominance) ออกจากกัน และปริภูมิสีพื้นฐาน RGB รวม 4 ปริภูมิสี คือ

- ปริภูมิสี RGB
- ปริภูมิสี RGB นอร์แมลไลซ์
- ปริภูมิสี HSV
- ปริภูมิสี YCbCr

1.1 ปริภูมิสี RGB

ปริภูมิสี RGB เป็นปริภูมิสีพื้นฐานหรือปริภูมิสีหลัก โดยค่าแต่ละจุดภาพจะเกิดจากการรวมกันของสเปคตรัมของแสงสีแดง (R) เขียว (G) และ น้ำเงิน (B) ในสัดส่วนความเข้มที่แตกต่างกัน โดยมีการกำหนดค่าความเข้มของสีแดง เขียว และน้ำเงิน เป็นค่าตั้งแต่ 0-255 (ทั้งหมด 256 ระดับ แทนด้วยค่า 8 บิต) ซึ่งที่ค่า 0 จะมีความเข้มมากสุด ไล่ลำดับไปจนถึง 255 ซึ่งมีความเข้มน้อยสุด ค่า RGB นี้เป็นค่าที่โทรทัศน์, กล้องวีดิทัศน์ และจอคอมพิวเตอร์ใช้แสดงผล โดยจะเป็นปริภูมิเวกเตอร์บน R^3 ดังรูปที่ 2

ข้อเสียของปริภูมิสี RGB คือ ความสว่างและสี จะถูกรวมเข้าด้วยกันหรือเป็นสหสัมพันธ์กัน ซึ่งไม่สอดคล้องกับระบบการมองเห็นของสายตามนุษย์



รูปที่ 2 ปริภูมิสี RGB

1.2 ปริภูมิสี RGB นอร์แมลไลซ์ (rgb)

ปัญหาของปริภูมิสี RGB คือ การรวมกันของความสว่างและสี จึงเกิดเป็นปริภูมิสี RGB นอร์แมลไลซ์ ซึ่งมีวัตถุประสงค์เพื่อตัดความสนใจขององค์ประกอบทางความสว่างทิ้งไป โดยหากค่าจุดภาพในปริภูมิสี RGB

นั้นมีสีเดียวกันแต่ความสว่างต่างกัน จะถูกกำหนดเป็นค่าเดียวกันในปริภูมิสี RGB นอร์แมลไลซ์ และแทนเป็นสัญลักษณ์ rgb แต่ละค่าสามารถหาได้จากสมการที่ 1-3

$$r = \frac{R}{(R+G+B)} \quad (1)$$

$$g = \frac{G}{(R+G+B)} \quad (2)$$

$$b = \frac{B}{(R+G+B)} \quad (3)$$

ตัวอย่าง เช่น 1. ค่าในปริภูมิสี RGB (255,0,0) ทำการแปลงไปสู่ rgb จะได้ (1,0,0)

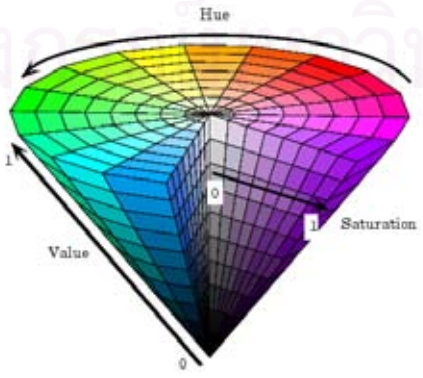
และ 2. ค่าในปริภูมิสี RGB (100,0,0) ทำการแปลงไปสู่ rgb จะได้ (1,0,0)

ซึ่งทั้งสองค่านี้ในปริภูมิสี RGB จะมีเฉพาะองค์ประกอบของสีแดง ที่มีค่าความเข้มของแสงแตกต่างกัน ส่วนในปริภูมิสี rgb องค์ประกอบทางแสงจะถูกละทิ้งไปทำให้ 2 ค่านี้ถือเป็นค่าเดียวกัน ดังรูปที่ 3

รูปที่ 3 ความแตกต่างของแสงในปริภูมิสี rgb จะถูกละทิ้ง

1.3 ปริภูมิสี HSV

ค่าในปริภูมิสี HSV แยกเป็นค่า H และ S คือ องค์ประกอบของสี ส่วนค่า V คือ ความเข้มของแสง เป็นปริภูมิสีที่สร้างขึ้นตามพื้นฐานการมองเห็นสีตามสายตาของมนุษย์ โดยตัว H ย่อจาก Hue เป็นสีที่สะท้อนมาจากสีของวัตถุ ซึ่งแตกต่างกันตามความยาวของคลื่นแสงที่มากระทบวัตถุและสะท้อนกลับมาที่ตาเรา จะถูกแทนด้วยค่า 0 ถึง 360 องศา ตัว S ย่อจาก Saturation เป็นค่าความเข้มข้นและความจางของสี คือ สัดส่วนที่มีอยู่ในสีเทา วัดเป็นค่าเปอร์เซ็นต์ คือจาก 0 (สีเทา) ถึง 100 (สีมีความอิ่มตัวเต็มที่) และ ตัว V ย่อจาก Value เป็นค่าความสว่างและความมืดของสี วัดเป็นค่าเปอร์เซ็นต์ คือจาก 0 (ความมืด) ถึง 100 (ความสว่าง) ซึ่งการแปลงจาก RGB ไปสู่ HSV ทำได้โดยสมการที่ 4-7



รูปที่ 4 ปริภูมิสี HSV

$$H = \begin{cases} H_1 & ; B \le G \\ 360^\circ - H_1 & ; B > G \end{cases} \quad (4)$$

ซึ่งหา H1 ได้จาก

$$H_1 = \cos^{-1}\left\{ \frac{0.5[(R-G)+(R-B)]}{\sqrt{(R-G)(R-G)+(R-B)(G-B)}} \right\} \quad (5)$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (6)$$

$$V = \frac{\max(R,G,B)}{255} \quad (7)$$

1.4 ปริภูมิสี YCbCr

ปริภูมิสี YCbCr จะแบ่งข้อมูลของสีออกในลักษณะเดียวกับการรับรู้ของมนุษย์ กล่าวคือ จะแยกระหว่างองค์ประกอบทางความสว่างและสีออกจากกัน โดยมีค่า Y เป็นค่าความสว่างมีค่าตั้งแต่ 16–235 ส่วนค่า Cb และ Cr คือ องค์ประกอบของสี ซึ่งเป็นค่าความแตกต่างของ สีน้ำเงินกับค่าความสว่าง (B-Y) และ สีแดงกับค่าความสว่าง (R-Y) ตามลำดับ ปริภูมิสี YCbCr เป็นปริภูมิสีดิจิทัล ซึ่งถูกกำหนดเป็นมาตรฐานโดย BT.601 ซึ่งแปลงจากองค์ประกอบ YUV แอนะล็อก และถูกนำมาใช้สำหรับการเข้ารหัสในมาตรฐานการบีบอัดข้อมูลวีดิทัศน์สำหรับการสื่อสาร รวมถึง MPEG-4

สำหรับการแปลงจากปริภูมิสี RGB ไปสู่ YcbCr ทำได้โดยใช้สมการที่ 8

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0980 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (8)$$

และการแปลงกลับจากปริภูมิสี YCbCr ไปสู่ RGB ทำได้โดยใช้สมการที่ 9

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.1643 & -0.0003 & 1.5960 \\ 1.1643 & -0.3920 & -0.8130 \\ 1.1643 & 2.0171 & -0.0001 \end{bmatrix} \begin{bmatrix} Y-16 \\ Cb-128 \\ Cr-128 \end{bmatrix} \quad (9)$$

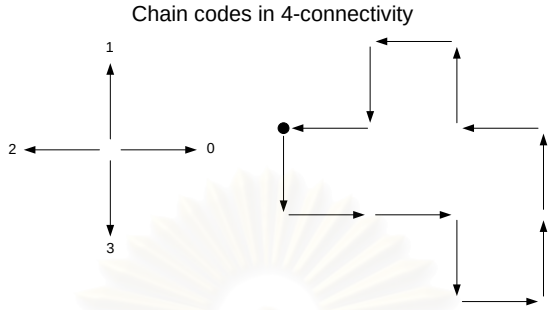
2. การวิเคราะห์หาบริเวณที่เชื่อมต่อกัน (Connected Component Analysis)

การตรวจหาขอบเขตของวัตถุด้วยวิธีแบบ Chain Code จะมีการกำหนดรหัสไว้ก่อนล่วงหน้า โดยแต่ละรหัสจะแสดงถึงการเปลี่ยนแปลงทิศทางของจุดภาพซึ่งเป็นค่าเดียวกันในบริเวณข้างเคียง และเมื่อเราทำการเข้ารหัสภาพที่เป็นไบนารี (มีเพียงค่า 0 กับ 1) จะสามารถพิจารณาได้ว่ามีจุดภาพกลุ่มใดบ้างที่เป็นวัตถุเดียวกันจากความสัมพันธ์ของรหัส การกำหนดรหัสจะมีทั้งแบบ 4 จุดเชื่อมต่อ และ 8 จุดเชื่อมต่อ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ส่วนค่าอนุพันธ์จะสามารถนำมาใช้ในการอธิบายรูปร่างของวัตถุ ซึ่งหาได้จากการเปลี่ยนแปลงการหมุนของลูกศรเป็นทิศ 90 หรือ 45 องศา และ ตามเข็ม หรือ ทวนเข็มนาฬิกา ดังแสดงในรูปที่ 5

ตัวอย่าง Chain Code แบบ 4 จุดเชื่อมต่อ



รูปที่ 5 Chain Code แบบ 4 จุดเชื่อมต่อ โดยมีรหัสคือ 3,0,0,3,0,1,1,2,1,2,3,2 และค่าอนุพันธ์เป็น 1,0,3,1,1,0,1,3,1,1,3,1

3. การทำไดเลชัน (Dilation) และ การทำอีโรชัน (Erosion)

3.1 การทำไดเลชัน

นิยาม คือ เซตของจุด p ที่เกิดจาก x+b โดยที่ x เป็นสมาชิกของ X และ b เป็นสมาชิกของ B ดังสมการที่ 10

$$X \oplus B = \{p \in \varepsilon^2 : p = x + b, x \in X \ \& \ b \in B\} \qquad (10)$$

โดยที่ X คือ รูปภาพที่ทำการไดเลชัน

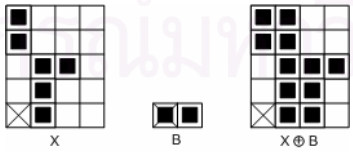
B คือ ขอบเขตในการทำไดเลชัน

ตัวอย่าง การทำไดเลชัน รูปที่ 6

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

$$B = \{(0,0),(1,0)\}$$

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



รูปที่ 6 การทำไดเลชัน

3.2 การทำอีโรชัน

นิยาม คือ เซตของจุด p ทุกจุดจากรูปภาพทดสอบ ผลลัพธ์ของการทำอีโรชันคือ จุด p เหล่านั้นที่ทำให้ p+b เป็นสมาชิกของ X สำหรับทุก ๆ b ที่เป็นสมาชิกของ B แสดงในสมการที่ 11

$$X\Theta B = \{p \in \varepsilon^2 : p + b \in X _ for _ every _ b \in B\} \quad (11)$$

โดยที่ X คือ รูปภาพที่ทำการอีโรชัน

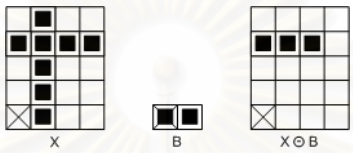
B คือ ขอบเขตในการทำอีโรชัน

<u>ตัวอย่าง</u> การทำอีโรชัน รูปที่ 7

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

$$B = \{(0,0),(1,0)\}$$

$$X\Theta B = \{(0,3),(1,3),(2,3)\}$$



รูปที่ 7 การทำอีโรชัน

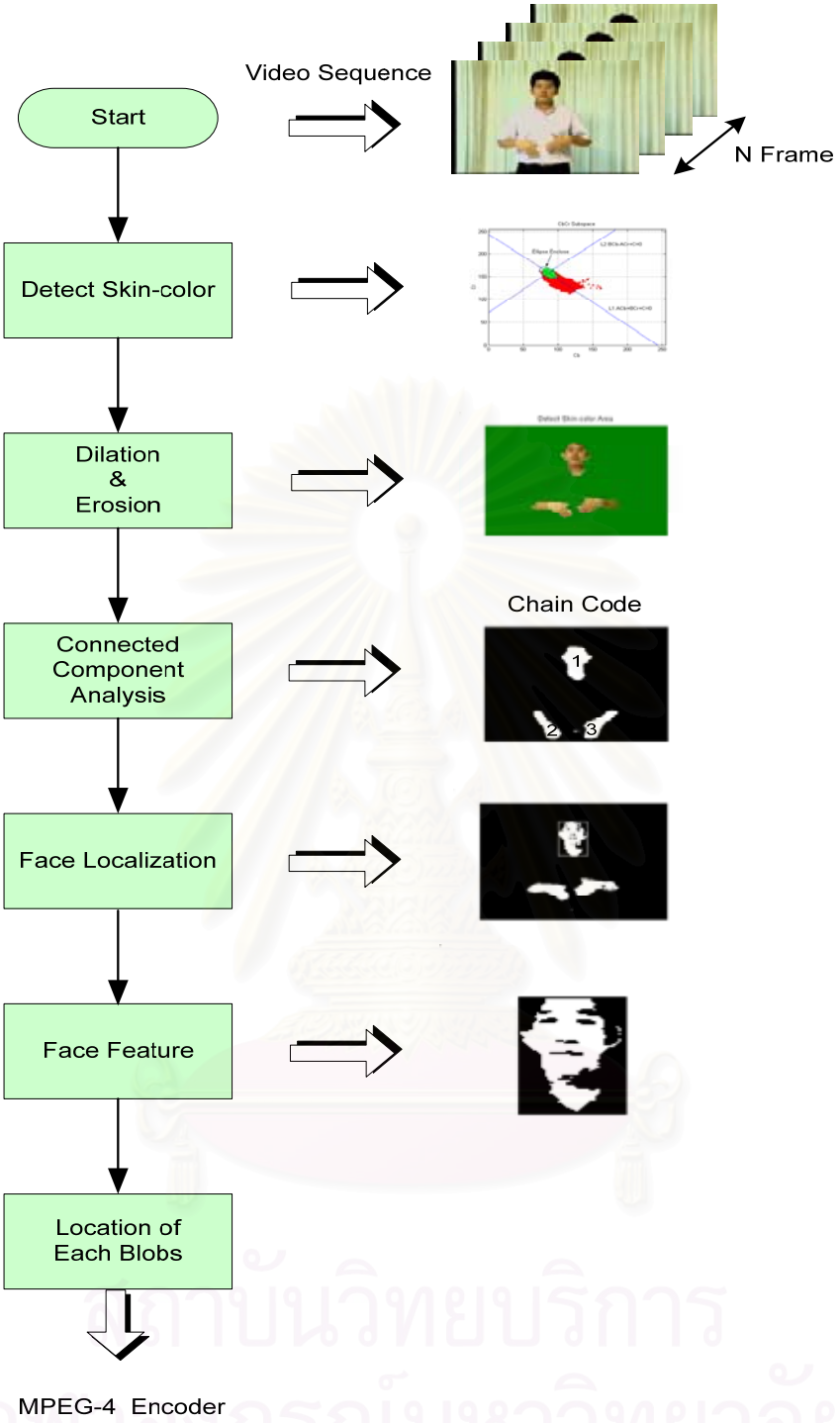
สิ่งที่นำเสนอในแผนงานวิจัย

1. สถาปัตยกรรมของระบบ

การทำงานของระบบการตรวจจับ และ แยกส่วนภาพออกเป็นเชิงวัตถุบนใบหน้ามนุษย์ แสดงดังรูปที่ 8 ประกอบไปด้วย กระบวนการในการตรวจจับสีผิวมนุษย์ ซึ่งแบบจำลองที่เลือกใช้คือ แบบจำลองวงรี (Elliptical model) ในปริภูมิสีย่อย CbCr หลังจากนั้นก็นำมาผ่านกระบวนการทำไดเลชันและอีโรชัน ด้วยตัวกระทำที่มีความเหมาะสม เพื่อให้มาส์กที่ได้มีความสมบูรณ์มากยิ่งขึ้น แล้วทำการวิเคราะห์บริเวณเชื่อมต่อกันเพื่อแยกเป็นแต่ละกลุ่ม (blobs) ด้วยวิธีแบบ Chain code

ในส่วนการหาองค์ประกอบที่สำคัญบนใบหน้า (ตา, คิ้ว, จมูก และ ปาก) จะนำตำแหน่งของใบหน้าที่ได้จากขั้นตอนแรกมา แล้วใช้วิธีตัดสินความแตกต่างจากองค์ประกอบทางความสว่าง โดยอาศัยสมมติฐานที่ว่าบริเวณส่วน ตา, คิ้ว, จมูก และ ปาก มีความสว่างที่น้อยกว่าส่วนสีผิวบนใบหน้า เพื่อแยกและนำแต่ละส่วนไปทำการเข้ารหัสเชิงวัตถุตามมาตรฐาน MPEG-4 ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



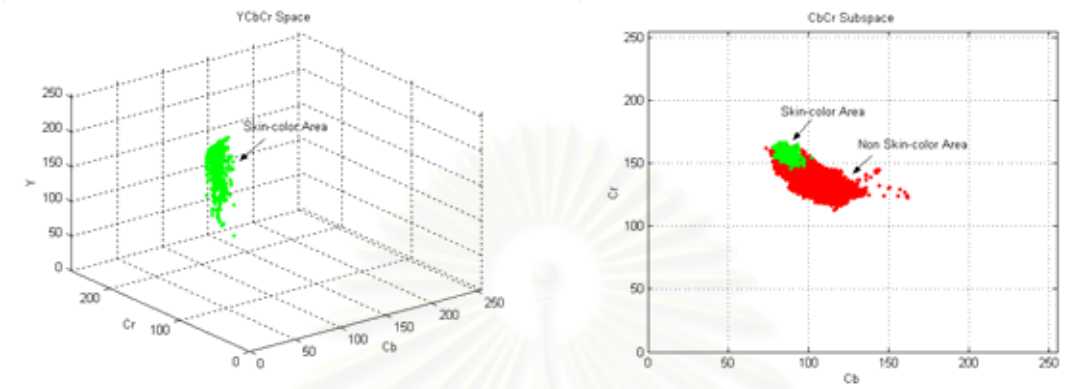
รูปที่ 8 สถาปัตยกรรมของระบบตรวจจับ และ แยกส่วนต่าง ๆ ที่สำคัญบนใบหน้ามนุษย์

2. การกระจายตัวของจุดภาพบริเวณสีผิวในปริภูมิย่อย CbCr

ในงานวิจัยที่ผ่านมาจะทำการเก็บข้อมูลค่าของสีผิวในปริภูมิสีต่าง ๆ แล้วแยกโดยกำหนดค่าสูงสุด ต่ำสุดเป็นค่าในการตัดสินบริเวณส่วนสีผิว ใช้เส้นตรงหลาย ๆ เส้นในการแยกบริเวณส่วนสีผิว, ใช้โพลิโนเมียล ในการกระชับข้อมูลส่วนสีผิว, หรือ ใช้วงรีในการแบ่งค่าตัดสินบริเวณส่วนสีผิว

สำหรับในปริภูมิย่อย CbCr เมื่อเรานำค่า CbCr ของจุดภาพแต่ละจุดมาเขียนลงบนระนาบ CbCr ใน รูปที่ 9 โดยแยกส่วนบริเวณสีผิว กับบริเวณอื่น ๆ ของรูปภาพ จะสังเกตเห็นได้ว่าค่า CbCr ของจุดภาพที่เป็นสีผิว

จะเกาะกลุ่มกันอยู่บริเวณหนึ่ง เป้าหมาย คือ ต้องทำการแยกบริเวณส่วนนี้ออกมา ซึ่งวิธีการที่นำเสนอจะใช้แบบจำลองวงรี เช่นเดียวกัน แต่ระบบตรวจจับบริเวณส่วนสีผิวที่สร้างขึ้น จะสามารถนำไปใช้งานร่วมกับระบบที่มีการปรับตัวได้ เนื่องจากได้มี การนำเสนอกรรมวิธีที่ใช้หาขอบเขตของวงรีเพื่อกระชับกับข้อมูลส่วนที่เป็นสีผิวเพิ่มเติมอีกด้วย



รูปที่ 9 การกระจายของส่วนสีผิวในปริภูมิ YCbCr และ ปริภูมิย่อย CbCr

การหาจุดศูนย์กลางและจุดยอดของวงรี

$$h = \mu_{C_b}; k = \mu_{C_r}$$
$$X1 = \mu_{C_b} - \tau(SD_{C_b}); Y1 = \mu_{C_r} + \tau(SD_{C_r}) \quad (12)$$
$$X2 = \mu_{C_b} + \tau(SD_{C_b}); Y2 = \mu_{C_r} - \tau(SD_{C_r})$$

h, k คือ จุดศูนย์กลางของวงรีซึ่งได้จากค่าเฉลี่ยของ Cb และ Cr ตามลำดับ

X1, Y1 และ X2, Y2 คือ จุดยอดของวงรีบนแกนเอก

$$L_1 : AC_b + BC_r + C = 0; L_2 : BC_b - AC_r + D = 0 \quad (13)$$

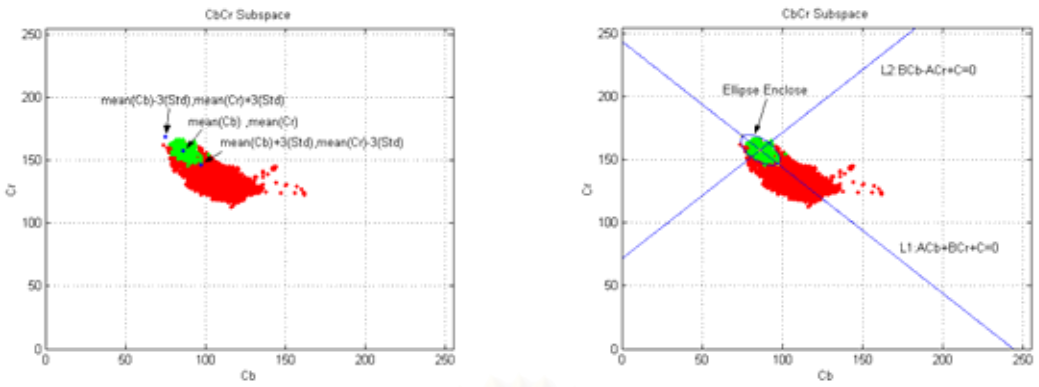
L1 คือ แกนเอกของวงรี และ L2 คือ แกนโทของวงรี

$$\frac{(AC_b + BC_r + C)^2}{a^2} + \frac{(BC_b - AC_r + D)^2}{b^2} = 1 \quad (14)$$

a คือ ความยาวแกนเอกของวงรี

b คือ ความยาวแกนโทของวงรี

ซึ่งในที่นี้เราจะใช้สมการที่ 14 ซึ่งเป็นสมการวงรี ในการกำหนดขอบเขตล้อมรอบบริเวณส่วนที่เป็นสีผิว



รูปที่ 10 การกำหนดจุดศูนย์กลางและจุดยอดวงรี จากค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน

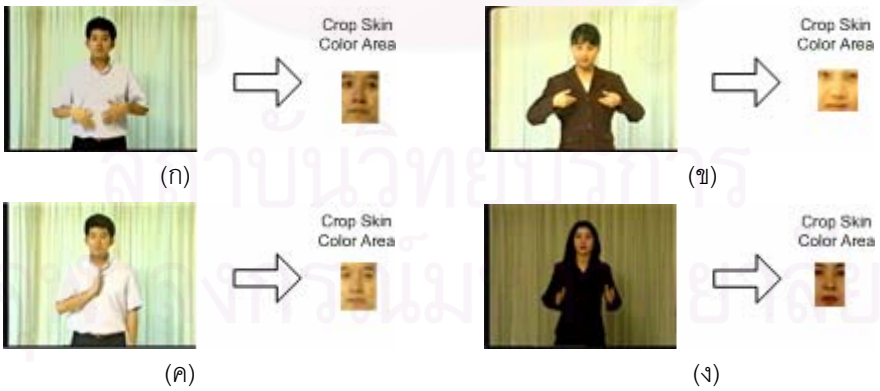
ในการตัดสินลักษณะว่าจุดภาพนั้นว่าเป็นสีผิวหรือไม่ จะตรวจสอบจากค่า CbCr ของจุดภาพ หากพบว่าอยู่ภายในวงรี จะกำหนดจุดภาพนั้นเป็นสีผิว ดังแสดงในสมการที่ 15

$$Pixel_is_skin = \begin{cases} 1, if\,(enclose_with_ellipse) \\ 0, else \end{cases} \quad (15)$$

ผลการทดลอง

ลำดับวีดิทัศน์ที่นำมาใช้ทำการทดสอบมีขนาด 240 x 320 จุดภาพ ในอัตรา 15 เฟรมต่อวินาที

- ลำดับวีดิทัศน์ที่ 1 จำนวน 139 เฟรม ในรูปที่ 12 (ก)
- ลำดับวีดิทัศน์ที่ 2 จำนวน 82 เฟรม ในรูปที่ 12 (ข)
- ลำดับวีดิทัศน์ที่ 3 จำนวน 74 เฟรม ในรูปที่ 12 (ค)
- ลำดับวีดิทัศน์ที่ 4 จำนวน 110 เฟรม ในรูปที่ 12 (ง)



รูปที่ 11 ภาพที่ใช้ทำการทดลองขนาด 240 x 320 จุดภาพ และ
บริเวณสีผิวที่ถูกตัดออกมาเพื่อใช้เป็นค่าอ้างอิง

1. การทดลองแยกบริเวณส่วนสีผิวโดยใช้แบบจำลองชนิดไม่มีตัวแปร (Non Parametric Model)

การทดลองแยกบริเวณส่วนสีผิวโดยใช้แบบจำลองชนิดไม่มีตัวแปร มีจุดประสงค์เพื่อทดสอบว่าปริภูมิสีใดให้ค่าความถูกต้องของการแยกส่วนสีผิวมากกว่ากัน ทำโดยเลือกภาพหนึ่งเฟรมจากลำดับวีดิทัศน์ทดสอบ จากนั้นทำการตัดบริเวณส่วนสีผิวบริเวณใบหน้ามาใช้เป็นค่าอ้างอิง และเก็บค่าข้อมูลบริเวณสีผิวไว้ ทำการ

เปรียบเทียบค่าแต่ละจุดภาพกับข้อมูลบริเวณส่วนสีผิวที่เก็บไว้ หากค่า ณ จุดภาพนั้นมีค่าตรงกับค่าอ้างอิงที่เก็บไว้ แบบจำลองชนิดไม่มีตัวแปรก็จะตีความจุดนั้นเป็นสีผิว ซึ่งทำการทดลองในปริภูมิย่อย 3 ปริภูมิ คือ CbCr (รูปที่ 12 หลักที่ 2), r-g (รูปที่ 12 หลักที่ 3) และ HS (รูปที่ 12 หลักที่ 4) จากผลการทดลองการหาบริเวณส่วนสีผิวโดยใช้แบบจำลองชนิดไม่มีตัวแปร ผลของมาสก์ที่ได้ ในปริภูมิย่อย CbCr จะให้ผลที่ดีกว่า HS และ rg

ส่วนรูปที่ 12 หลักที่ 1 คือ การแยกส่วนสีผิวโดยใช้สายตา และใช้เป็นค่าอ้างอิงเพื่อเปรียบเทียบความถูกต้องกับผลของมาสก์จากการหาส่วนสีผิวในปริภูมิสีทั้งสาม ดังแสดงในตารางที่ 1-4 โดยที่ ค่าในปริภูมิย่อย rg และ HS จะถูกนอร์แมลไลซ์ด้วยค่า 255 เพื่อให้เท่ากับค่าในปริภูมิย่อย CbCr

(ก-1) (ก-2) (ก-3) (ก-4)

(ข-1) (ข-2) (ข-3) (ข-4)

(ค-1) (ค-2) (ค-3) (ค-4)

(ง-1) (ง-2) (ง-3) (ง-4)

รูปที่ 12 การแยกบริเวณส่วนสีผิวด้วยสายตา (ก-1) - (ง-1) และผลจากการแยกโดยใช้แบบจำลองชนิดไม่มีตัวแปรในปริภูมิย่อย CbCr (ก-2) - (ง-2), rg (ก-3) - (ง-3) และ HS (ก-4) - (ง-4) ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

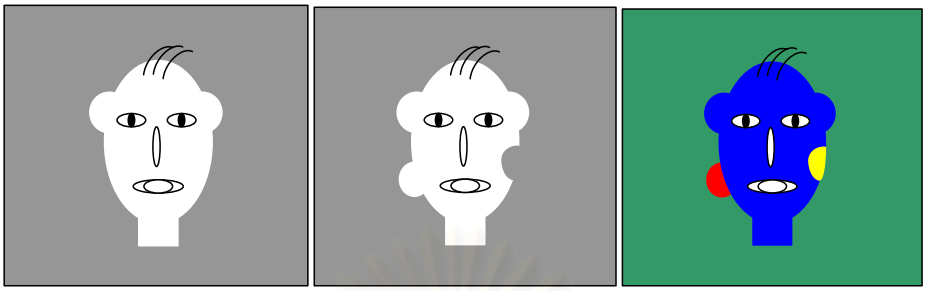
<u>การวัดประสิทธิภาพของแบบจำลองต่าง ๆ ที่ใช้ในการตรวจจับสีผิว</u>

ตัวแปรที่สำคัญที่นำมาใช้สำหรับการวัดประสิทธิภาพของแบบจำลองจะมีอยู่ด้วยกัน 4 ตัวแปร (อธิบายความหมายแต่ละตัวในรูปที่ 13) คือ

- TP ย่อจาก “True Positive” มีหน่วยเป็น จุดภาพ
- TN ย่อจาก “True Negative” มีหน่วยเป็น จุดภาพ

- FP ย่อจาก “False Positive” มีหน่วยเป็น จุดภาพ

- FN ย่อจาก “False Negative” มีหน่วยเป็น จุดภาพ



(ก) (ข) (ค)

รูปที่ 13 ความหมายของตัวแปรทั้ง 4 ตัวแปร สามารถอธิบายได้ดังนี้

(ก) ภาพอ้างอิงซึ่งได้จากการแยกส่วนสีผิวด้วยสายตามนุษย์

(ข) ผลลัพธ์ที่ได้จากแบบจำลองตรวจจับสีผิว

(ค) TP (ส่วนบริเวณสีผิวที่ตรวจจับได้ถูกต้อง)

TN (ส่วนบริเวณฉากหลังที่ตรวจจับได้ถูกต้อง)

FP (ส่วนบริเวณฉากหลังแต่แบบจำลองให้ผลเป็นค่าสีผิว)

FN (ส่วนบริเวณสีผิวแต่แบบจำลองให้ค่าเป็นฉากหลัง)

ค่า DR ย่อจาก “Detection Rate” คือ อัตราความถูกต้องในการตรวจจับ มีหน่วยเป็น (%) เปอร์เซ็นต์ สามารถหาได้จากสมการที่ 16

$$DR = \frac{TP}{TP + FN} \times 100\% \qquad (16)$$

ค่า FAR ย่อจาก “False Alarm Rate” คือ อัตราการเกิดความผิดพลาด มีหน่วยเป็น (%) เปอร์เซ็นต์ สามารถหาได้จากสมการที่ 17

$$FAR = \frac{FP}{TP + FP} \times 100\% \qquad (17)$$

ปริภูมิสี	TP (จุดภาพ)	TN (จุดภาพ)	FP (จุดภาพ)	FN (จุดภาพ)	DR (เปอร์เซ็นต์)	FAR (เปอร์เซ็นต์)
CbCr	3531	71671	60	1538	69.66	1.67
rg	3321	71503	228	1748	65.52	6.42
HS	2715	71485	246	2354	53.56	8.31

ตารางที่ 1 ค่าความถูกต้องในแต่ละปริภูมิสีของแบบจำลองชนิดไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตาของลำดับวีดีทัศน์ที่ 1

ปริภูมิสี	TP (จุดภาพ)	TN (จุดภาพ)	FP (จุดภาพ)	FN (จุดภาพ)	DR (เปอร์เซ็นต์)	FAR (เปอร์เซ็นต์)
CbCr	2113	73150	73	1464	59.07	3.34
rg	2310	67693	5530	1267	64.58	70.54
HS	1738	69394	3829	1839	48.59	68.78

ตารางที่ 2 ค่าความถูกต้องในแต่ละปริภูมิสีของแบบจำลองชนิดไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตาของลำดับวีดีทัศน์ที่ 2

ปริภูมิสี	TP (จุดภาพ)	TN (จุดภาพ)	FP (จุดภาพ)	FN (จุดภาพ)	DR (เปอร์เซ็นต์)	FAR (เปอร์เซ็นต์)
CbCr	3075	72115	64	1546	66.54	2.04
rg	3298	71938	241	1323	71.37	6.81
HS	2660	72043	136	1961	57.56	4.86

ตารางที่ 3 ค่าความถูกต้องในแต่ละปริภูมิสีของแบบจำลองชนิดไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตาของลำดับวีดีทัศน์ที่ 3

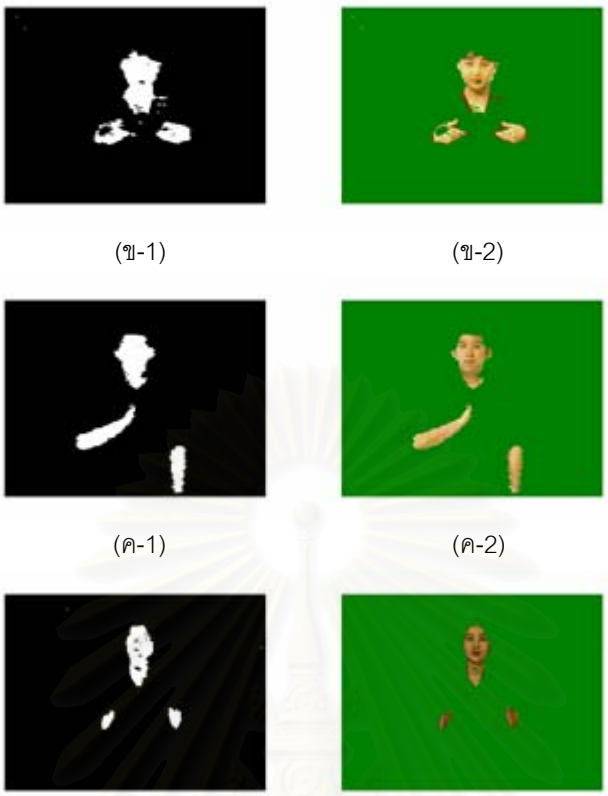
ปริภูมิสี	TP (จุดภาพ)	TN (จุดภาพ)	FP (จุดภาพ)	FN (จุดภาพ)	DR (เปอร์เซ็นต์)	FAR (เปอร์เซ็นต์)
CbCr	1570	74390	16	824	65.58	1.01
rg	1403	68266	6140	991	58.60	81.40
HS	1335	68126	6280	1059	49.64	82.47

ตารางที่ 4 ค่าความถูกต้องในแต่ละปริภูมิสีของแบบจำลองชนิดไม่มีตัวแปรเทียบกับการแยกส่วนสีผิวด้วยสายตาของลำดับวีดีทัศน์ที่ 4

2. การทดลองหาบริเวณส่วนสีผิว โดยใช้แบบจำลองวงรี ในปริภูมิย่อย CbCr

(ก-1)

(ก-2)



(ข-1) (ข-2)

(ค-1) (ค-2)

(ง-1) (ง-2)

รูปที่ 14 ผลการแยกบริเวณส่วนสีผิวโดยใช้แบบจำลองวงรี

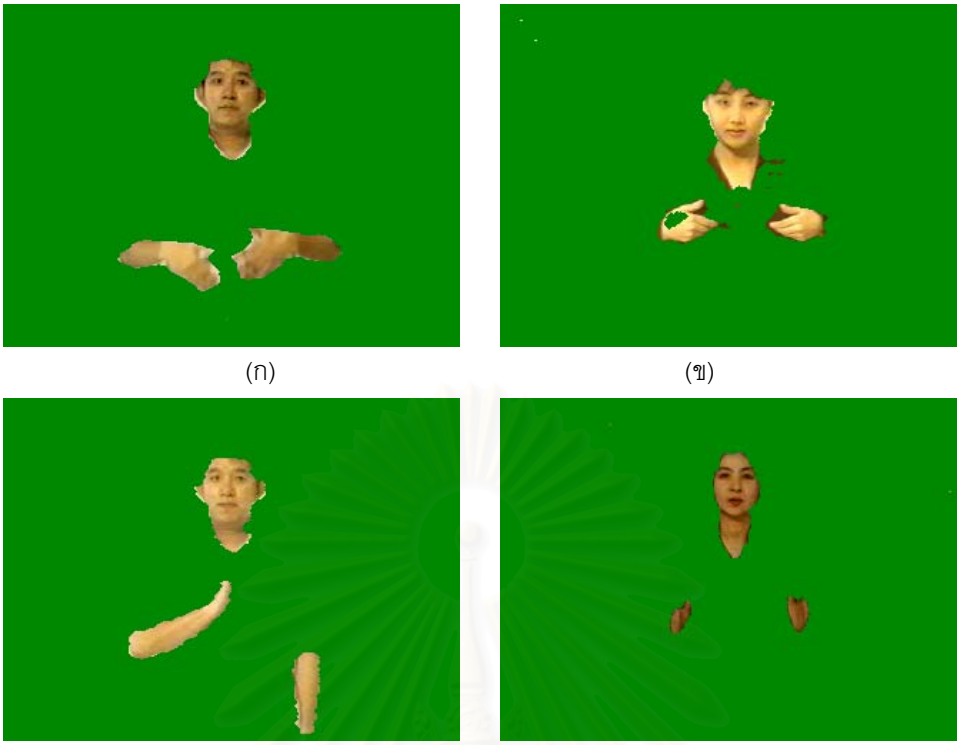
ปริภูมิสี	TP (จุดภาพ)	TN (จุดภาพ)	FP (จุดภาพ)	FN (จุดภาพ)	DR (เปอร์เซ็นต์)	FAR (เปอร์เซ็นต์)
1	4799	71417	314	270	94.67	6.14
2	3220	72584	639	357	90.02	16.56
3	3870	72028	151	751	83.75	3.76
4	2044	74381	25	350	85.38	1.21

ตารางที่ 5 ค่าความถูกต้องของแบบจำลองวงรีเทียบกับการแยกส่วนสีผิวด้วยสายตา

3. ผลจากการทำไดเลชันและอีโรชัน

การทำไดเลชันและอีโรชัน สำหรับวีดิทัศน์ทดสอบจะใช้ตัวกระทำดังสมการที่ 18

$$operator = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad (18)$$



(ก) (ข)

(ค) (ง)

รูปที่ 15 ผลจากการนำบริเวณส่วนสีผิวที่ถูกแยก มาทำไดเลชัน และ อีโรชัน

4. ผลการแยกองค์ประกอบที่สำคัญบริเวณใบหน้า (เช่น ตา คิ้ว จมูก และ ปาก)

4.1 วิธีที่ 1 [วิธีที่นำเสนอโดย R. S. Hsu]

ระเบียบวิธีของ R. S. Hsu นี้ จะทำการตรวจจับส่วนสำคัญบนใบหน้า 2 อย่าง คือ ตาและปาก ซึ่งแยกเป็นแต่ละส่วนดังนี้

4.1.1 การตรวจจับบริเวณตา

เราจะสามารถทำการแยกบริเวณตามนุษย์ ซึ่งจะเป็นส่วนตาจะมีค่า Cb ที่สูง และ Cr ที่ต่ำ โดยใช้สมการที่ 19

$$EyeMapC = \frac{1}{3}\{(Cb^2) + (\widetilde{Cr}^2) + (\frac{Cb}{Cr})\} \quad (19)$$

โดยที่ ค่า Cb^2 , $\widetilde{Cr}^2$ และ $\frac{Cb}{Cr}$ จะถูกนอร์แมลไลซ์ให้อยู่ในช่วง 0 ถึง 255

ค่า $\widetilde{Cr}$ คือ ค่า $255 - Cr$

(ก) (ข)

รูปที่ 16 รูปภาพที่ใช้ทำการทดลอง

(ก) (ข)

รูปที่ 17 ผลจากการตรวจจับบริเวณตา

4.1.2 <u>การตรวจจับบริเวณปาก</u>

ค่าสีบริเวณปาก จะประกอบไปด้วยองค์ประกอบสีแดงที่มาก และ องค์ประกอบสีน้ำเงินน้อยกว่าส่วนอื่น ๆ บนใบหน้า ซึ่งสามารถทำการตรวจจับโดยใช้สมการที่ 20 และ 21

$$MouthMap = Cr^2 * (Cr^2 - \eta * \frac{Cr}{Cb})^2 \quad (20)$$

$$\eta = 0.95 * \frac{\frac{1}{n}\sum_{(x,y)\in\Re} Cr(x,y)^2}{\frac{1}{n}\sum_{(x,y)\in\Re} \frac{Cr(x,y)}{Cb(x,y)}} \quad (21)$$

โดยที่ ค่า Cr^2 และ $\frac{Cr}{Cb}$ จะถูกนอร์แมลไลซ์ให้อยู่ในช่วง 0 ถึง 255

ค่า n คือ จำนวนจุดภาพบริเวณใบหน้า

ค่า η คือ อัตราส่วนของค่าเฉลี่ย Cr^2 เทียบกับ ค่าเฉลี่ย $\frac{Cr}{Cb}$

(ก) (ข)

รูปที่ 18 ผลจากการตรวจจับบริเวณปาก

ในการทดลองหัวข้อ 4.1 นี้ เมื่อเราทำการทดลองตามสมการที่ 21-23 ที่บริเวณตาและปาก จะถูกแยกออกมาจากส่วนอื่น (เห็นได้ว่ามีค่าใกล้เคียงสีขาวมากว่าส่วนอื่น ๆ) แต่มีข้อจำกัดว่ารูปที่นำมาใช้ทดลองต้องผ่านกระบวนการปรับค่าสมดุลขาว (White balance) เพื่อให้เกิดค่าแสงและสีที่เหมาะสม หากใช้รูปจากลำดับวีดิทัศน์ทั้ง 4 (รูปที่ 12) จะให้ผลที่คลาดเคลื่อนเนื่องจากไม่ได้ผ่านกระบวนการในการปรับสมดุลขาว ดังนั้นเพื่อให้กระบวนการในการตรวจจับนี้สามารถทำได้กับรูปโดยทั่วไปจึงได้เสนอวิธีในหัวข้อที่ 4.2

4.2 วิธีที่ 2 [ใช้วิธีตรวจสอบจากองค์ประกอบความสว่าง Y]

จากผลการแยกบริเวณส่วนสีผิวที่ได้โดยใช้แบบจำลองวงรี ในปริภูมิย่อย CbCr (รูปที่ 15) หากเราต้องการแยกองค์ประกอบที่สำคัญบนใบหน้า คือ ส่วนตา, คิ้ว, จมูก และ ปาก ซึ่งส่วนต่าง ๆ เหล่านี้ มีค่าสีที่แตกต่างกับสีผิวบนใบหน้าและมือ โดยเมื่อพิจารณาองค์ประกอบทางแสง (Y) จะมีค่าต่ำ ๆ เพราะค่าสีมีค่าใกล้เคียงกับสีดำ กระบวนการในขั้นตอนนี้จึงทำการหาค่าเฉลี่ยขององค์ประกอบทางแสงเฉพาะจุดภาพบริเวณส่วนสีผิว เพื่อกำหนดค่าทดสอบ (threshold) ตามสมการที่ 22 และนำมาทำการสร้างมาสก์ใหม่ โดยจุดภาพที่มีค่า Y น้อยกว่าค่าทดสอบ จะถือว่าเป็นส่วนตา คิ้ว จมูก และ ปาก

$$Threshold = Mean(Y) - \beta * Std(Y) \quad (22)$$

β คือ ค่าถ่วงน้ำหนักในที่นี้ใช้ค่าเท่ากับ 1.5

$Mean(Y)$ คือ ค่าเฉลี่ยขององค์ประกอบทางแสง (Y) ของจุดภาพบริเวณส่วนสีผิว

$Std(Y)$ คือ ค่าเบี่ยงเบนมาตรฐานขององค์ประกอบทางแสง (Y) ของจุดภาพบริเวณส่วนสีผิว

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

(ก) (ข)

(ค) (ง)

รูปที่ 19 ผลที่ได้จากการเปรียบเทียบกับค่าทดสอบ และทำการกำหนดอาณาเขตบริเวณใบหน้า

หมายเหตุ สังเกตว่ารูป 19 (ข) ซึ่งให้ผลไม่ชัดเจนสามารถปรับปรุงให้ดีขึ้นโดยปรับค่า β จากสมการที่ 22

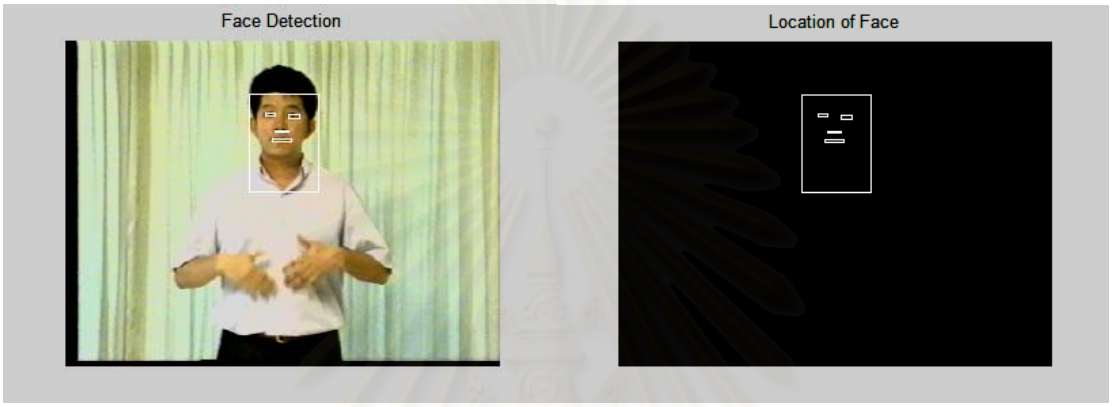
กระบวนการในขั้นตอนต่อไป คือ นำมาสก์ ภายในกรอบสี่เหลี่ยมบริเวณใบหน้ามาทำการ อีโรชัน ด้วยสมการที่ 23 และ จากนั้นก็นำมาผ่านการวิเคราะห์บริเวณเชื่อมต่อกัน เพื่อกำหนดองค์ประกอบต่าง ๆ (ตา, คิ้ว, จมูก และปาก) ซึ่งผลการทดลองแสดงในรูปที่ 20 และ 21

$$Erosion_operator = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \quad (23)$$

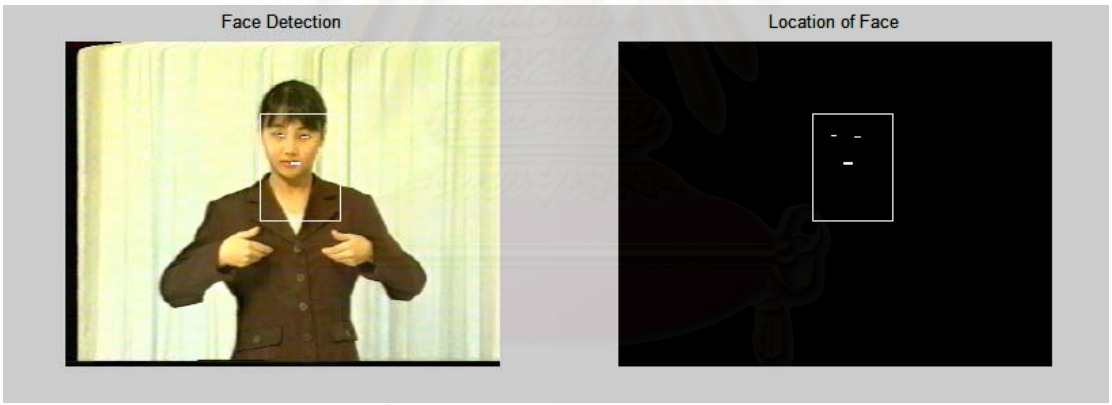
(ก) (ข)

(ค) (ง)

รูปที่ 20 ขอบเขตบริเวณใบหน้า และ คอมพลีเมนต์ (Complement) ที่ผ่านการทำอีโรชัน

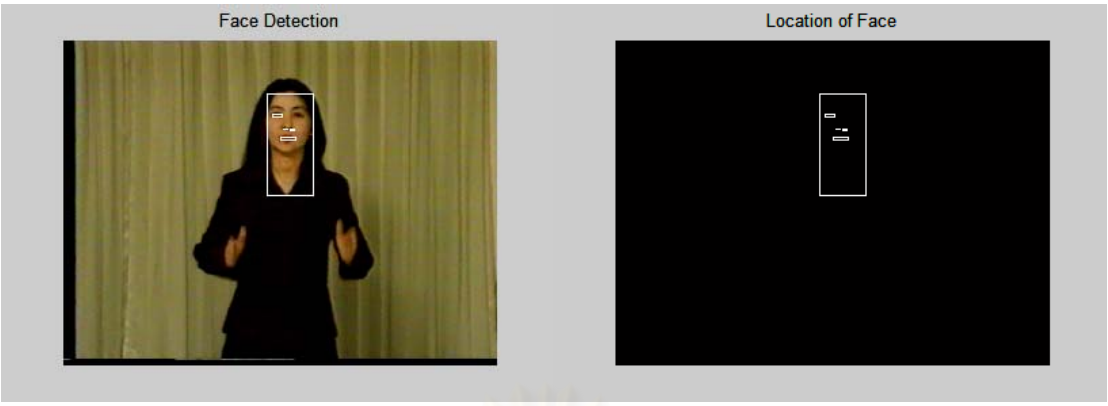


(ก)



(ข)

(ค)



(ง)

รูปที่ 21 องค์ประกอบต่างๆ บนใบหน้า

จากผลการทดลองในรูปที่ 21

- รูปที่ 21 (ก) และ 21 (ค) ให้ผลที่ค่อนข้างถูกต้อง ในส่วนของตา, จมูก และปาก
- ในรูปที่ 21 (ข) มีความผิดพลาดในส่วนของจมูกเนื่องจากกระบวนการตรวจจับโดยใช้องค์ประกอบความสว่างในสมการที่ 22 ซึ่งอาจทำการแก้ไขโดยเพิ่มส่วนประมวลผลในการปรับค่า β
- ส่วนในรูปที่ 21 (ง) มีความผิดพลาดที่ตาข้างขวา เนื่องจากการทำการวิเคราะห์บริเวณเชื่อมต่อกัน สาเหตุเพราะตาของผู้แปลภาษามืออยู่ใกล้กับส่วนผมซึ่งเป็นสีดำ

สรุป

งานวิจัยในส่วนนี้ได้นำเสนอเกี่ยวกับระบบการตรวจจับ และ แยกส่วนต่าง ๆ ที่สำคัญบนร่างกายมนุษย์ ซึ่งประกอบไปด้วย กระบวนการในการแยกส่วนบริเวณสีผิวของมนุษย์ โดยแบบจำลองที่ใช้เป็นแบบวงรี และพิจารณาในปริภูมิย่อย CbCr รวมถึงเสนอวิธีการในการหาขอบเขตของวงรี เพื่อสามารถนำไปประยุกต์ใช้ร่วมกับระบบที่มีการปรับตัวได้ในค่าสีผิว กระบวนการถัดมา คือ การหาองค์ประกอบต่าง ๆ ที่สำคัญบนใบหน้า เช่น ตา, คิ้ว, จมูก และ ปาก เพื่อมาทำการเข้ารหัสเชิงวัตถุตามมาตรฐาน MPEG-4 และสามารถผนวกใช้ร่วมกับระบบการประชุมสัมมนาวีดิทัศน์ได้

เอกสารอ้างอิง

[1] R.-L. Hsu, M. Abdel-Mottaleb, A. K. Jain, “Face Detection in Color Images,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.2, No.5, pp.696-706, May 2002.

[2] J. Fritsch, S. Lang, M. Kleinehagenbrock, G. A. Fink and G. Sagerer, “Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection,” *IEEE Int. Workshop on Robot and Human Interactive Communication*, September 2002.

[3] J. Yang and A. Waibel, “A Real-Time Face Tracker,” *Proc. of Third Workshop on Applications of Computer Vision*, pp.142-147,1996.

[4] S. Kawato, and J. Ohya, “Automatic Skin-color Distribution Extraction for Face Detection and Tracking,” *ICSP2000: The 5th Int. Conf. on signal processing*, vol.2, pp.1415-1418, August 2000.

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People," *Proc. Computer Vision and Pattern Recognition*, 2000.

[6] C. R. Wren, A. Azarbayejani, T. Darrel, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.780-785, July 1997.

[7] M. Sonka, V. Hlavac, and R. Boyle, "Image Processing Analysis and Machine Vision," *2nd Ed., Brooks/Cole Publishing Company*, 1999.

[8] Z. M. Hafed, "Object tracking Principles and challenges to keeping on eye on things," *IEEE Potentials*, August/September 1999.

[9] S. Lu, S. Igi, H. Matsuo and Y. Nagashima. "Towards a Dialogue System Based on Recognition and Synthesis of Japanese Sign Language," *Proc. Beifield Gesture Workshop 1997*, 1997.

[10] U. M. Erdem, and S. Sclaroff, "Automatic Detection of Relevant Head Gestures in American Sign Language Communication," *Proc. ICPR*, August 2002.

[11] A. Hadid, M. Pietikainen and B. MartinKauppi "Color-Based Face Detection Using Skin Locus Model Hierarchical Filtering," *Proc .IEEE Pattern Recognition*, Vol. 4, pp.196-200, 2002.

[12] C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," *IEEE Transactions on Multimedia*, Vol. 1, No. 3, September 1999.

[13] M. C. Shin, K. I. Chang and L. V. Tsap "Does Colorspace Transformation Make Any Difference on Skin Color," *IEEE Workshop on Applications of Computer Vision*, 2002.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย