

## รายงานการวิจัยฉบับสมบูรณ์

การศึกษาเบื้องต้นการนำเทคโนโลยีทางด้าน  
การประมวลผลภาพมาใช้ในระบบรักษาความปลอดภัย  
The preliminary Study of Image Processing Techniques  
for Security System

ผู้วิจัย

ดร.นิคม สุวรรณวร

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย จากเงินรายได้  
มหาวิทยาลัยสงขลานครินทร์ ประจำปีงบประมาณ 2550

## บทคัดย่อ

ระบบการรักษาความปลอดภัยด้วยภาพวิดีโอจากกล้องได้เข้ามามีบทบาทต่อชีวิตประจำวันอย่างจริงจังและมีความสำคัญมากขึ้น การเฝ้าระวังด้วยข้อมูลภาพวิดีโอในปัจจุบันได้เข้าสู่ยุคที่สามแล้ว ข้อมูลภาพวิดีโอแบบดิจิทัลที่ได้จากกล้องแบบไอพีสามารถนำมาใช้ประโยชน์ได้อย่างมาก งานวิจัยนี้ได้มุ่งเน้นไปยังสองหัวข้อสำคัญคือ การศึกษาและพัฒนาเทคนิคการประมวลผลภาพขั้นพื้นฐานเพื่อการวิเคราะห์ภาพจากกล้องวิดีโอวงจรปิด ซึ่งประกอบด้วย การวิเคราะห์วัตถุเคลื่อนไหว การวิเคราะห์หาภาพใบหน้าคนที่มีคุณภาพดีที่สุด การตรวจจับป้ายทะเบียน และการตรวจจับภาพได้ท้องรถ ส่วนที่สองจะเน้นไปที่การพัฒนาและออกแบบระบบซอฟต์แวร์เพื่อการจัดการข้อมูลภาพจากกล้องวิดีโอวงจรปิดเพื่อให้มีความสามารถที่จะเรียกใช้งานการวิเคราะห์ผลภาพที่กล่าวมาแล้วข้างต้นได้อย่างมีประสิทธิภาพ

## **Abstract**

In the domain of security system, the popularity of IP-based cameras for video surveillance system is growing really fast. The digital image sequences obtained from IP-camera could be analyzed by the image processing techniques which are an intelligent approach for the automatic detection and identification system. In this report, we emphasize our research into two directions: the research and development of the image processing techniques for surveillance system, such as object motion detection, face quality measurement for key frame extraction, license plate detection, and under vehicle scanner; and secondly, it aims for the research and development of video surveillance software management. This software manages how to apply the proposed processing techniques to the video sequences obtained from IP-cameras.

# สารบัญ

	หน้า
บทคัดย่อ	i
Abstract	ii
สารบัญ	iii
รายการตาราง	v
รายการภาพ	vi
<b>บทที่</b>	
1. บทนำ	
1.1 ความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ประโยชน์ที่คาดว่าจะได้รับ	2
1.4 ขอบเขตการวิจัย	2
1.5 แนวคิดงานวิจัยโดยภาพรวม	2
2. การตรวจจับการเคลื่อนไหวของวัตถุ	
2.1 ปัญหาและความสำคัญ	5
2.2 ปรัชญาวรรณกรรม	5
2.3 ระเบียบวิธีการวิเคราะห์ภาพเคลื่อนไหวของวัตถุ	6
2.4 ผลการทดลองและวิจารณ์	12
2.5 สรุปผล	16
2.6 บรรณานุกรม	16
3. การวิเคราะห์คุณภาพของภาพใบหน้าสำหรับการค้นหาเฟรมภาพสำคัญ เพื่อใช้ใน ระบบเฝ้าระบบ	
3.1 ปัญหาและความสำคัญ	17
3.2 ปรัชญาวรรณกรรม	18
3.3 ระเบียบวิธีการวิเคราะห์คุณภาพใบหน้า	19
3.4 ผลการทดลองและวิจารณ์	22
3.5 สรุปผล	26
3.6 บรรณานุกรม	27
4. การตรวจจับหาป้ายทะเบียน	
4.1 ปัญหาและความสำคัญ	28
4.2 ปรัชญาวรรณกรรม	28

## สารบัญ (ต่อ)

	หน้า
4.3 ระเบียบวิธีการตรวจจับป้ายทะเบียนและแยกตัวอักษร	29
4.4 ผลการทดลองและวิจารณ์	34
4.5 สรุปผล	37
4.6 บรรณานุกรม	38
5. การตรวจสอบภาพใต้ท้องรถ	
5.1 ปัญหาและความสำคัญ	39
5.2 ปรีทัศน์วรรณกรรม	39
5.3 การออกแบบระบบและระเบียบวิธีการตรวจสอบภาพใต้ท้องรถ	40
5.4 ผลการทดลองและวิจารณ์	51
5.5 สรุปผล	53
5.6 บรรณานุกรม	54
6. ระบบการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิด	
6.1 ปัญหาและความสำคัญ ปรีทัศน์วรรณกรรม	55
6.2 การออกแบบการจัดการกล้องวิดีโอวงจรปิด	56
6.3 ระเบียบวิธีการวิเคราะห์ภาพวิดีโอ	59
6.4 ระเบียบวิธีการอัดและการจัดเก็บภาพวิดีโอ	62
6.5 ผลการทดลองและวิจารณ์	63
6.6 สรุปผล	65
6.7 บรรณานุกรม	66
7. บทสรุปและข้อเสนอแนะ	67
ภาคผนวก	69

## รายการตาราง

ตาราง	หน้า
4.1 แสดงผลการหาดำแหน่งป้ายทะเบียน	32
4.2 แสดงผลการหาดำแหน่งป้ายทะเบียนและแยกตัวอักษร	35
5.1 แสดงผลลัพธ์ของการต่อภาพใต้ห้องรถด้วยวิธีของลูคัส-คาเนส	49
5.2 แสดงผลลัพธ์ของการต่อภาพใต้ห้องรถด้วยวิธีของ CAMSHIFT	52

## รายการภาพ

ภาพ	หน้า
2.1 การวิเคราะห์การเคลื่อนไหวของวัตถุ	6
2.2 การวิเคราะห์การเคลื่อนไหวด้วยการลบพื้นหลังแบบเฟรมติดกัน	7
2.3 การวิเคราะห์การเคลื่อนไหวด้วยการลบพื้นหลังแบบเลือกพื้นที่	9
2.4 การปรับปรุงคุณภาพของวัตถุ	10
2.5 การตรวจจับวัตถุเคลื่อนไหวและการคำนวณประวัติ	10
2.6 การวิเคราะห์คุณสมบัติตำแหน่ง ขนาดและทิศทางของวัตถุ	11
2.7 ไดอะแกรมแสดงระบบโดยรวมการวิเคราะห์การเคลื่อนไหวของวัตถุ	12
2.8 แสดงตัวอย่างการทดสอบกับไฟล์ภาพวิดีโอ	13
2.9 แสดงผลลัพธ์ของโปรแกรมในการตรวจจับบุคคล	14
2.10 แสดงผลลัพธ์ของการทดสอบการตรวจจับจากไฟล์วิดีโอ	15
3.1 ภาพรวมการทำงานของอัลกอริทึม Harr-Like และ Camshift	19
3.2 ผลลัพธ์จากการรวมวิธีการของ Harr-like และ Camshift	20
3.3 กระบวนการวัดดัชนีคุณภาพใบหน้าที่ดีที่สุด	20
3.4 ดัชนีคุณภาพใบหน้าที่ดีที่สุด	21
3.5 ข้อมูลภาพที่ใช้ในการทดสอบ	23
3.6 ลำดับวิดีโอที่ 1 (a) กราฟของ $\kappa_i(I_f)$ , $\eta_{\mu,\sigma}(\kappa_i)$ และ $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.	24
3.7 ลำดับวิดีโอที่ 2 (a) กราฟของ $\kappa_i(I_f)$ , $\eta_{\mu,\sigma}(\kappa_i)$ และ $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.	24
3.8 ลำดับวิดีโอที่ 3 (a) กราฟของ $\kappa_i(I_f)$ , $\eta_{\mu,\sigma}(\kappa_i)$ และ $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.	25
3.9 ลำดับวิดีโอที่ 4 (a) กราฟของ $\kappa_i(I_f)$ , $\eta_{\mu,\sigma}(\kappa_i)$ และ $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.	25
3.10 แสดงโปรแกรมอัลกอริทึมที่ได้นำไปใช้ในโปรแกรม VsMonitor	26
4.1 การหาขอบของภาพด้วยวิธีการ Canny	30
4.2 การปรับแต่งขอบด้วยการใช้เทคนิคทางด้าน Morphological (Close)	30
4.3 การปรับแต่งขอบด้วยการใช้เทคนิคทางด้าน Morphological (Open)	30
4.4 ตำแหน่งของป้ายที่หาได้	31
4.5 การหาดำแหน่งมุมต่างๆ ของป้าย	32
4.6 การแปลงภาพเป็นไบนารี	33

## รายการภาพ (ต่อ)

ภาพ	หน้า	
4.7	การ Projection ตามแนวแกนนอน	34
4.8	การ Projection ตามแนวแกนตั้ง	34
4.9	โปรแกรมที่ทำงานร่วมกับ VsMonitor ได้	37
5.1	ระบบ PPA Vehicle Surveillance System	40
5.2	ระบบ Under Vehicle Surveillance System (UVSS) Integration	40
5.3	ระบบ Und-Aware™ Under-Vehicle-Surveillance-System	40
5.4	ระบบตรวจข้อมูลภาพจากใต้ท้องรถ	41
5.5	เครื่องมือรับภาพใต้ท้องรถ	41
5.6	อุปกรณ์ต้นแบบกล่องรับภาพสะท้อน	42
5.7	ภาพสะท้อนจากกระจกเงาสะท้อนมุม 45 องศา	42
5.8	แสดงไดอะแกรมการต่อภาพ	43
5.9	ตัวอย่างภาพนิ่งทั้ง 2 ภาพ	43
5.10	ส่วนที่เหมือนกันระหว่างทั้ง 2 ภาพ	45
5.11	หาจุดที่เหมือนกันระหว่างภาพ 2 ภาพ	45
5.12	ภาพที่เกิดจากการซ้อนกันของ 2 ภาพ	46
5.13	การติดตั้ง Laser pointer ในกล่องรับภาพใต้ท้องรถ	49
5.14	แสดงจุดอ้างอิงที่สร้างเองโดยใช้ Laser pointers	50
5.15	ติดตามจุดที่สนใจโดยใช้อัลกอริทึม CAMSHIFT	51
6.1	ระบบเฝ้าระวังด้วยกล้องวิดีโอ รุ่นที่สาม	56
6.2	สถาปัตยกรรมระบบ	59
6.3	ตัวอย่างการวิเคราะห์ภาพที่ได้จากกล้องวงจรปิดของ ร.ร. มอ วิทยานุสรณ์	59
6.4	แสดงหน้าต่างการจัดกลุ่มกล้องขนาด 3x4	60
6.5	ภาพวิดีโอจากกล้องใต้ตึกภาควิชาวิศวกรรมคอมพิวเตอร์	64
6.6	ผลการค้นหาวัตถุเคลื่อนไหว	64



# บทที่ 1

## บทนำ

### 1.1 ความสำคัญของปัญหา

ความปลอดภัยในชีวิตและทรัพย์สินเป็นภัยคุกคามที่เกิดขึ้นได้ตลอดเวลาทั้งกับเราเองกับบุคคลรอบข้างหรือแม้แต่กับหน่วยงานหรือระดับสถาบันก็ตาม เช่น ปัญหาอาชญากรรมต่างๆ ปัญหาที่เกี่ยวข้องกับความมั่นคงที่สัมพันธ์เฉพาะพื้นที่ เช่น ความไม่สงบในสามจังหวัดชายแดนภาคใต้ เป็นต้น การป้องกันไม่ให้เกิดเหตุร้ายหรือการที่จะสามารถติดตามแก้ไขปัญหาได้ทันถ่วงทีหากมีเหตุการณ์ใดเกิดขึ้นจึงถือได้ว่าเป็นแนวทางปฏิบัติโดยทั่วไป ในส่วนของมหาวิทยาลัยสงขลานครินทร์ซึ่งเป็นหน่วยงานของรัฐที่ตั้งอยู่ในพื้นที่เสี่ยงได้ตระหนักเป็นอย่างดีถึงเรื่องดังกล่าว ดังจะให้เห็นจากการเพิ่มจุดตรวจเพื่อรักษาความปลอดภัยโดยรอบและการติดตั้งระบบกล้องวิดีโอวงจรปิดซึ่งเป็นเทคโนโลยีสมัยใหม่ทางด้านภาพและวิดีโอมาใช้ประกอบเพื่อช่วยให้การตรวจสอบเป็นไปได้อย่างรวดเร็วหากเกิดเหตุผิดปกติขึ้น จึงถือได้ว่าเป็นแนวทางใหม่ที่จะทำให้การรักษาความปลอดภัยมีประสิทธิภาพมากยิ่งขึ้น

ปัจจุบันระบบรักษาความปลอดภัยด้วยภาพวิดีโอจากกล้องสมัยใหม่ที่ทำงานอยู่บนระบบไอพีและส่งข้อมูลภาพวิดีโอเป็นแบบดิจิทัลได้นั้น เพิ่งได้เข้ามามีบทบาทต่อชีวิตประจำวันของเราอย่างจริงจัง อย่างไรก็ตามระบบที่ใช้อยู่ในปัจจุบันยังมีข้อจำกัดอยู่อีกหลายประการ เช่น ระบบซอฟต์แวร์และระบบกล้องโดยทั่วไปแล้วจะต้องเป็นยี่ห้อเดียวกัน การจัดเก็บข้อมูลภาพวิดีโอไม่สามารถเก็บได้นานซึ่งจะขึ้นอยู่กับความละเอียดของภาพ จำนวนเฟรมต่อวินาทีที่ต้องการบันทึกหรือการบันทึกข้อมูลภาพตลอดเวลาทั้งๆ ที่ไม่มีเหตุการณ์ที่ต้องสนใจ ข้อจำกัดด้วยความสามารถในการตรวจสอบค้นหาข้อมูลที่สนใจ เช่น ต้องใช้เวลานานหรือเครื่องมือที่ใช้ในการค้นหาไม่เพียงพอ ความสามารถในการวิเคราะห์ภาพแบบอัตโนมัติที่มีความฉลาดและเหมาะสมกับปัญหานั้น เช่น การวิเคราะห์หน้า การจดจำป้ายทะเบียน การตรวจสอบได้ท้องรถ และอื่นๆ อย่างไรก็ตาม ข้อจำกัดต่างๆ ตามที่กล่าวมาข้างต้นนั้นจะมีความสัมพันธ์อย่างยิ่งกับราคา หากต้องการระบบที่มีความสามารถครบก็จะทำให้ระบบนั้นมีราคาแพงมาก

ด้วยเหตุผลสำคัญที่ได้กล่าวมาแล้วนั้น การพัฒนาระบบที่มีความสามารถในการทำงานให้ได้ตรงตามความต้องการมากที่สุดและมีราคาถูกลงที่สุดจึงมีความจำเป็นอย่างยิ่ง งานวิจัยนี้จึงเน้นไปที่การศึกษาและพัฒนาระบบเบื้องต้นในสองส่วนสำคัญคือ ระบบซอฟต์แวร์เพื่อการจัดการข้อมูลภาพจากกล้องวิดีโอวงจรปิด และการศึกษาคิดค้นและพัฒนาเทคนิคอัลกอริทึมทางด้านประมวลผลภาพที่จำเป็นเพื่อช่วยให้ระบบกล้องวิดีโอวงจรปิดมีความฉลาดมากขึ้น เช่น การวิเคราะห์วัตถุเคลื่อนไหว การวิเคราะห์ภาพใบหน้าที่ดีที่สุด การวิเคราะห์ภาพได้ท้องรถและ การวิเคราะห์ป้ายทะเบียนรถยนต์ เป็นต้น

## 1.2 วัตถุประสงค์ของการวิจัย

ก) ศึกษาความต้องการของผู้ใช้งานและข้อกำหนดทางเทคโนโลยี อันที่จะสามารถเพิ่มประสิทธิภาพให้กับระบบรักษาความปลอดภัยได้

ข) ศึกษาและพัฒนาระบบโดยภาพรวมถึงการนำเทคโนโลยีทางด้านการประมวลผลภาพมาใช้ในระบบความปลอดภัย โดยพิจารณารวมถึงลักษณะขององค์กรและการจัดการ

ค) ศึกษาและทดสอบเทคนิคการวิเคราะห์ข้อมูลภาพจากกล้องเบื้องต้น เช่น การวิเคราะห์วัตถุเคลื่อนไหว การวิเคราะห์ภาพใบหน้าที่ดีที่สุด การวิเคราะห์ภาพใต้ท้องรถ และการวิเคราะห์ป้ายทะเบียนรถยนต์ เป็นต้น

## 1.3 ประโยชน์ที่คาดว่าจะได้รับ

ก) ผลงานของโครงการ

- รายงานสรุปและขอเสนอการออกแบบระบบที่มีการนำเทคนิคการประมวลผลภาพมาใช้ในระบบรักษาความปลอดภัย
- รายงานทางเทคนิค (Technical report) สรุปอัลกอริทึมการวิเคราะห์ภาพจากกล้องในกรณีต่างๆ ตามวัตถุประสงค์

ข) ผลงานวิจัยพิเศษ

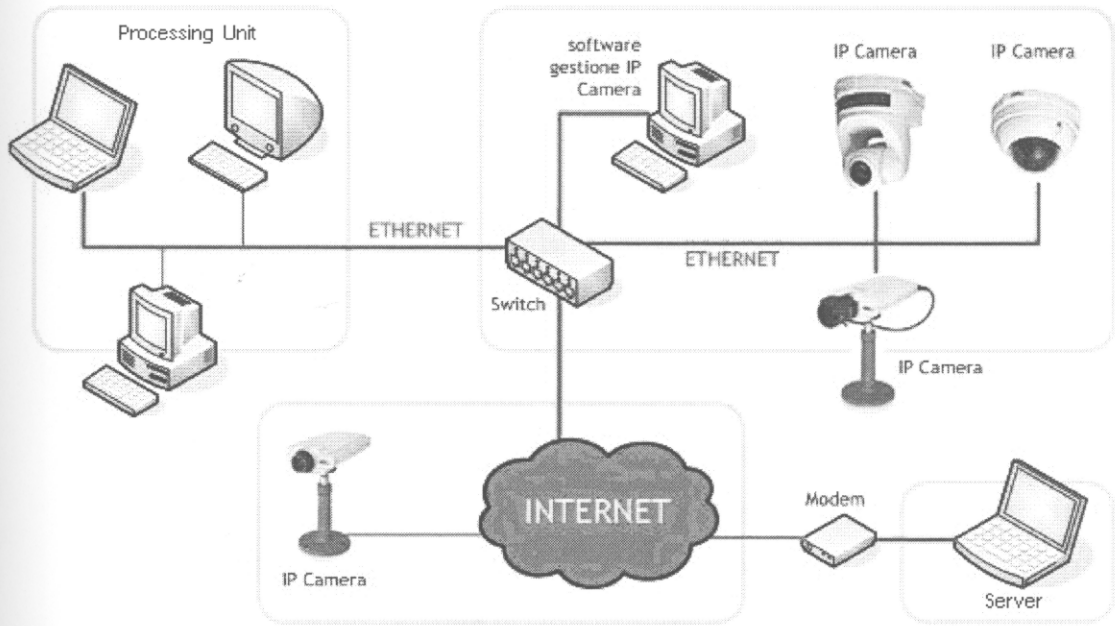
- บทความตีพิมพ์ในที่ประชุมวิชาการ

## 1.4 ขอบเขตการวิจัย

- โครงการนี้จะทำวิจัยในเชิงของการศึกษาและทดสอบเบื้องต้น **ยังไม่รวมถึงการพัฒนาาระบบเพื่อการใช้งานจริง**
- ศึกษาและทดสอบเทคนิคการวิเคราะห์ภาพ **ที่มีอยู่แล้ว** เพื่อมาประยุกต์ใช้ในระบบความปลอดภัยได้

## 1.5 แนวคิดงานวิจัยโดยภาพรวม

ระบบการรักษาความปลอดภัยที่เป็นการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิดในปัจจุบันได้ก้าวเข้าสู่เจนเนอเรชันที่สามแล้ว ตามรูปที่ 1.1 อุปกรณ์วิดีโอรับภาพถูกออกแบบให้สนับสนุนเทคโนโลยีใหม่ที่ทำงานอยู่บนระบบ IP (IP-based cameras) โดยมีข้อมูลภาพวิดีโอเป็นแบบดิจิทัล ภาพวิดีโอในรูปแบบของข้อมูลดิจิทัลนี้จะสามารถส่งผ่านระบบอินเทอร์เน็ตที่ใช้กันอย่างแพร่หลายไปยังที่ใดก็ได้ ในปัจจุบันระบบเฝ้าระวังด้วยกล้องวิดีโอ (Video surveillance system) นี้ได้ถูกนำมาใช้โดยทั่วไปและถือได้ว่าเป็นเรื่องใกล้ตัวมาก โดยเฉพาะอย่างยิ่งการเฝ้าระวังสำหรับหน่วยงานขนาดใหญ่ เช่น มหาวิทยาลัยสงขลานครินทร์ โรงเรียนม.อ.วิทยานุสรณ์ โรงพยาบาล ม.อ. และโรงงานต่างๆ ในสงขลา เป็นต้น



รูปที่ 1.1. ระบบเฝ้าระวังด้วยกล้องวิดีโอ เจนเนอเรชันที่สาม

การได้มาซึ่งข้อมูลภาพวิดีโอซึ่งอยู่ในรูปแบบดิจิทัลทำให้มีข้อดีที่สำคัญคือ การที่จะสามารถนำข้อมูลเหล่านั้นมาประมวลผลแบบอัตโนมัติด้วยเทคนิคการประมวลผลภาพ (Image processing technique) เพื่อลดการทำงานของบุคคลและหวังผลการทำงานที่มีความถูกต้องและมีความเสถียรมากที่สุดเท่าที่จะทำได้ ตัวอย่างเทคนิคที่ใช้อยู่แล้วในปัจจุบัน ได้แก่ การวิเคราะห์การเคลื่อนไหว เป็นต้น ในงานวิจัยนี้จะมุ่งเน้นไปที่สองประเด็นสำคัญ คือ ก) ศึกษาคิดค้นและทดสอบเทคนิคขั้นต้นการวิเคราะห์ภาพอัตโนมัติที่จำเป็นสำหรับระบบการเฝ้าระวังด้วยกล้องวิดีโอ ข) ศึกษาคิดค้นและพัฒนาาระบบสำหรับการจัดการข้อมูลภาพวิดีโอที่เหมาะสมสำหรับระบบเฝ้าระวังด้วยกล้องวิดีโอ

ในรายงานนี้จะนำเสนอเทคนิคการวิเคราะห์ภาพและระบบการจัดการข้อมูลภาพซึ่งจะแบ่งตามบทใน 5 หัวข้อสำคัญดังต่อไปนี้คือ

- 1). การตรวจจับการเคลื่อนไหวของวัตถุ (Object Motion Detection) นำเสนอในบทที่สอง เป็นการศึกษาและพัฒนาเทคนิคการวิเคราะห์การเคลื่อนไหวของวัตถุ การจำแนกวัตถุ การติดตามวัตถุ การหาคุณสมบัติของวัตถุ เช่น ตำแหน่ง ขนาด ทิศทาง และความเร็ว และระบบโดยรวม
- 2). การวิเคราะห์คุณภาพของภาพใบหน้าสำหรับการค้นหาเฟรมภาพสำคัญ (Key Frame Extraction from Video Sequence Using a Face Quality Index) นำเสนอในบทที่สาม (เนื้อหาของบทนี้ได้ตีพิมพ์ใน *proceeding* ของงานประชุมวิชาการ ECTI-CARD 2009) เป็นการศึกษาออกแบบและพัฒนาเทคนิคการค้นหาหน้าในภาพวิดีโอ การวิเคราะห์หาภาพใบหน้าที่ดีที่สุด ณ ช่วงเวลานั้น ๆ ทั้งนี้เพื่อจะใช้งานที่มีความซับซ้อนมากขึ้นในอนาคต เช่น การจดจำใบหน้า เป็นต้น

- 3). การตรวจจับหาป้ายทะเบียน (License Plate Detection) นำเสนอในบทที่สี่ เป็นการศึกษาและออกแบบกระบวนการวิเคราะห์หาตำแหน่งป้ายทะเบียนรถในภาพ และการแยกตัวอักษร เพื่อใช้ในการจดจำต่อไป
- 4). การตรวจสอบภาพใต้ท้องรถ (Under Vehicle Scanner) นำเสนอในบทที่ห้า เป็น การออกแบบอุปกรณ์เพื่อรับภาพใต้ท้องรถโดยใช้กล้องราคาถูกจำนวนหลายกล้อง ที่ต่อกันเป็นอาร์เรย์ และพัฒนาอัลกอริทึมในการต่อภาพจากหลายกล้องให้สามารถ ทำงานได้ตามความจริง
- 5). ระบบการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิด (A Generic Visual Surveillance System) นำเสนอในบทที่สาม (เนื้อหาของบทนี้ได้ตีพิมพ์ใน proceeding ของงาน ประชุมวิชาการ PEC 2008) เป็นการออกแบบและพัฒนาซอฟต์แวร์ระบบรักษา ความปลอดภัยด้วยกล้องวงจรปิด ที่สามารถติดต่อไปยังกล้องแบบไอพีชนิดต่างๆ ที่มีอยู่ในท้องตลาดได้ โดยสามารถนำเทคนิคและอัลกอริทึมตามที่ได้กล่าวมาแล้ว ก่อนหน้านี้มาใช้งานได้

## บทที่ 2

### การตรวจจับการเคลื่อนไหวของวัตถุ Object Motion Detection

#### 2.1 ปัญหาและความสำคัญ

การบันทึกภาพจากกล้องเพื่อเป็นการเฝ้าระวังภัยในสถานที่ต่างๆ มักจะเป็นการบันทึกแบบตลอดเวลา ซึ่งโดยปกติแล้วจะทำให้สามารถบันทึกข้อมูลภาพวิดีโอได้เพียงไม่กี่วันเท่านั้น ทั้งนี้เนื่องมาจากข้อจำกัดของฮาร์ดแวร์ที่ใช้ในการบันทึกซึ่งมีอยู่อย่างจำกัด หากพิจารณาถึงเหตุการณ์หรือสิ่งที่สนใจที่ต้องการบันทึกจริงๆ แล้วยกตัวอย่างมีจำนวนน้อยมากเมื่อเทียบกับการบันทึกส่วนใหญ่ การวิเคราะห์ภาพวิดีโอเพื่อค้นหาหรือตรวจจับเหตุการณ์ซึ่งเกี่ยวข้องกับพฤติกรรมของบุคคลโดยอัตโนมัติ เช่น การโจรกรรม การลอบวางเพลิง หรือ การควบคุมการเข้าออกในสถานที่เขตหวงห้าม จึงมีเป็นความจำเป็นพื้นฐานที่จะต้องมีสำหรับระบบรักษาความปลอดภัยโดยการใช้กล้องวิดีโอวงจรปิด

งานวิจัยนี้จึงเน้นไปที่การวิเคราะห์พื้นฐานเพื่อการตรวจจับการเคลื่อนไหวและติดตามของบุคคล โดยที่จะต้องสามารถนำไปประยุกต์กับระบบการรักษาความปลอดภัยโดยกล้องวิดีโอวงจรปิดได้ การวิเคราะห์นี้จะถือว่าเป็นเทคนิคพื้นฐานที่นำไปสู่การวิเคราะห์ที่ก้าวหน้ามากขึ้น เพื่อนำไปประยุกต์ใช้กับระบบงานอื่น ๆ เช่น ระบบการควบคุมกล้องเพื่อจับภาพบุคคล และระบบระบุตัวตนของบุคคลที่เข้าออกอาคาร เป็นต้น

#### 2.2 ปรัชญาวรรณกรรม

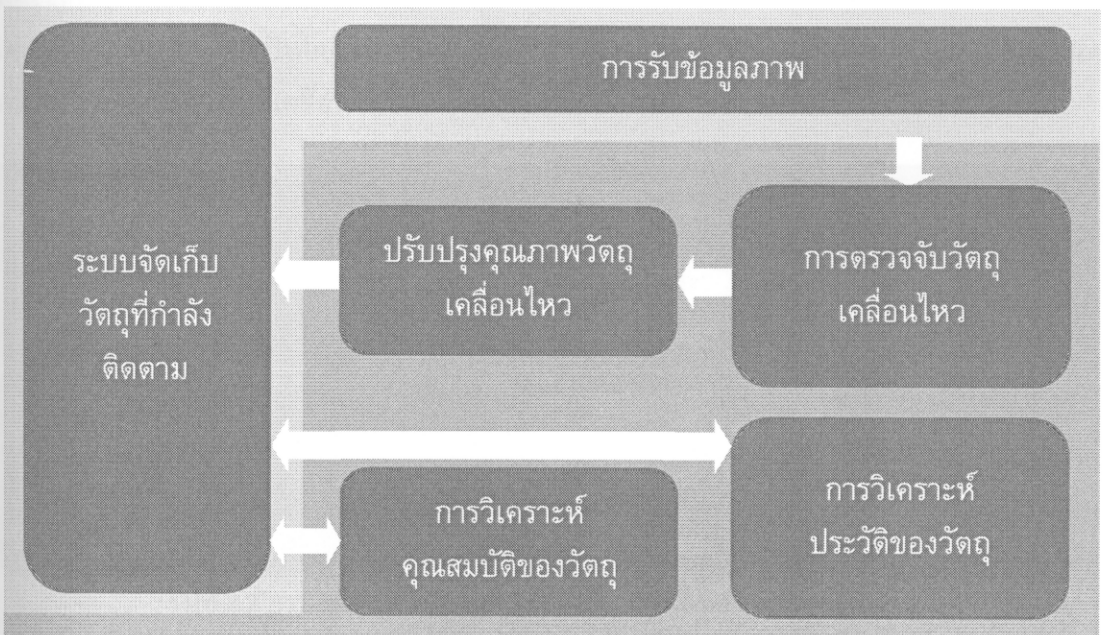
การเฝ้าระวังพื้นที่จากพฤติกรรมของคนถือว่าเป็นหนึ่งในงานวิจัยที่ได้รับความสนใจเป็นอย่างมากทางด้านคอมพิวเตอร์วิชัน [1] ซึ่งมีการนำไปใช้จริงในหลายด้าน เช่น การควบคุมการเข้าออกพื้นที่เฝ้าระวัง การระบุบุคคล ณ ระยะไกล การวิเคราะห์ทางสถิติการไหลของกลุ่มบุคคล การวิเคราะห์ความแออัด การตรวจจับพฤติกรรมที่ผิดปกติ การสังเกตการณ์จากความสัมพันธ์ของกลุ่มหลายตัว [2][3] งานวิจัยส่วนหนึ่งเน้นไปที่การปรับความถูกต้องของอัลกอริทึม เช่น การจัดการเมื่อเกิดการบัง การรวมการวิเคราะห์เพื่อติดตามวัตถุจากสองมิติและสามมิติไว้ด้วยกัน การวิเคราะห์การเคลื่อนไหวร่วมกับไบโอเมตริก การหาความผิดปกติร่วมกับการทำนายพฤติกรรม การสืบค้นข้อมูลวิดีโอด้วยภาพ การเข้าใจพฤติกรรมและภาษาธรรมชาติ การผสมข้อมูลจากเซนเซอร์ต่างๆ และการเฝ้าระวังจากระยะไกล เป็นต้น โดยปกติแล้วโครงสร้างการประมวลผลของระบบเหล่านี้ จะประกอบด้วยขั้นตอนทั่วไปดังนี้คือ การทำโมเดลสภาวะแวดล้อม การตรวจหาการเคลื่อนไหว การจัดกลุ่มวัตถุความเคลื่อนไหว การติดตาม

ความเข้าใจหรือการอธิบายพฤติกรรม การหาเอกลักษณ์ความเป็นมนุษย์ และการรวมผลการวิเคราะห์จากหลายๆ กล้อง เป็นต้น

การหาการเคลื่อนไหวในภาพมีวัตถุประสงค์เพื่อการค้นหาบริเวณในภาพที่เป็นวัตถุเคลื่อนที่ซึ่งอาจจะเป็นมนุษย์หรือไม่ก็ได้ มีวิธีการโดยทั่วไปคือ การหาการเคลื่อนไหวจากการลบด้วยพื้นหลัง ซึ่งปกติจะเหมาะสมกับพื้นหลังนิ่งไม่มีการเคลื่อนไหว การหา Optical flow ซึ่งจะวิเคราะห์เวกเตอร์ของการเคลื่อนไหวของวัตถุนั้น ๆ การจัดแยกวัตถุ เป็นวิธีการที่สำคัญที่จะหาความแตกต่างระหว่างบริเวณของภาพที่กำลังเคลื่อนไหวซึ่ง ซึ่งจะมีสองแนวทางคือ การแยกตามลักษณะรูปร่างของวัตถุและการแยกตามการเคลื่อนไหว ตามลำดับ เมื่อได้วัตถุที่เคลื่อนไหวแล้ว ขั้นตอนต่อไปคือการติดตามวัตถุนั้นไปยังเฟรมอื่นๆ ซึ่งในการติดตามนี้จะใช้คุณสมบัติต่างๆ เช่น ลักษณะของ Region ลักษณะของ active contour ลักษณะสำคัญของวัตถุ features และโมเดลของวัตถุ ในงานวิจัยนี้จะทำการศึกษาวิธีการต่าง ๆ จนถึงขั้นการติดตามเท่านั้น โดยยังไม่รวมถึงการจดจำหรือการเข้าใจพฤติกรรม

### 2.3 ระเบียบวิธีการวิเคราะห์ภาพเคลื่อนไหวของวัตถุ

ส่วนนี้จะอธิบายถึงรายละเอียดการทำงานของระบบการวิเคราะห์ภาพการเคลื่อนไหวของบุคคลภายในภาพวิดีโอ แผนภาพข้างล่างแสดงกระบวนการโดยรวมของระบบ ซึ่งประกอบด้วย การรับข้อมูลภาพ (Image Acquisition) ทั้งภาพจากกล้องและภาพจากไฟล์วิดีโอ การตรวจจับวัตถุเคลื่อนไหวด้วยเทคนิคการลบพื้นหลัง (Background subtraction) การปรับปรุงวัตถุการเคลื่อนไหวด้วยวิธีทาง Morphological (Object Quality Improvement) การวิเคราะห์หาคุณสมบัติของวัตถุ (Object's Characteristics) และระบบโดยรวมการวิเคราะห์ภาพการเคลื่อนไหว (Motion Analysis System)



รูปที่ 2.1 การวิเคราะห์การเคลื่อนไหวของวัตถุ

### 2.3.1. การตรวจจับวัตถุเคลื่อนไหว (Motion detection by background subtraction)

ในส่วนนี้จะเป็นการศึกษาและทดลองการตรวจจับวัตถุเคลื่อนไหวจากภาพที่ได้จากกล้องหรือไฟล์ภาพโดยใช้เทคนิคการลบพื้นหลัง (Background subtraction) แบบต่าง ๆ การหาวัตถุเคลื่อนไหวคือการนำเอาภาพที่ต้องการวิเคราะห์มาหาความแตกต่างกับพื้นหลัง ซึ่งสามารถหาได้หลายวิธี ได้แก่ การหาความแตกต่างจากเฟรมที่ติดกัน (Frame Differencing) การหาพื้นหลังด้วยค่าเฉลี่ย (Averaging background) และการหาพื้นหลังด้วยค่าเฉลี่ยแบบเลือกพื้นที่ (Average background with selectivity) สำหรับงานวิจัยนี้จะเน้นไปที่การวิเคราะห์ภาพขาวดำเป็นหลัก แนวคิดพื้นฐานในการตรวจจับหรือค้นหาวัดวัตถุเคลื่อนไหว  $O$  คือการหาความแตกต่างระหว่างพื้นหลัง  $B_t$  และภาพจากเฟรมใดๆ  $I_t$  ตามสมการดังต่อไปนี้

$$O = \min(|I_t - B_t|, I_{\max})$$

ความถูกต้องของวิธีการนี้จะขึ้นอยู่กับคุณภาพของพื้นหลังซึ่งปกติจะมีสัญญาณรบกวนเกิดขึ้นอันเนื่องมาจากหลายสาเหตุ ได้แก่ การกำหนดค่า Threshold ที่จะกำหนดว่าพิกเซลใดควรเป็นภาพพื้นหลังหรือวัตถุ การเปลี่ยนแปลงอย่างรวดเร็วของความสว่างและค่า contrast ของบริเวณที่สนใจ เป็นต้น

#### การหาวัตถุเคลื่อนไหวจากพื้นหลังที่เป็นเฟรมติดกัน (Frame Differencing)

วิธีนี้เป็นวิธีพื้นฐานที่ง่ายที่สุดที่จะทำให้ได้วัตถุที่สนใจ (Foreground) โดยการกำหนดใช้เฟรมก่อนหน้าเฟรมที่ต้องการวิเคราะห์เป็นพื้นหลัง เนื่องจากเป็นการใช้ภาพขาวดำที่มีแปดบิต ผลลัพธ์ที่ได้จึงมีค่าความสว่างของพิกเซลอยู่ระหว่าง 0 – 255 ซึ่งหากมีค่าน้อยจะหมายความว่าพิกเซลนั้นน่าจะเป็นพื้นหลัง ในทางตรงข้ามหากมีค่ามากแสดงว่าตำแหน่งนั้นน่าจะเป็นวัตถุที่เคลื่อนไหว ซึ่งทำให้จำเป็นอย่างยิ่งที่จำเป็นต้องมี Threshold ค่าหนึ่งเพื่อที่จะใช้ระบุว่าตำแหน่งหรือพิกเซลใดเป็นวัตถุที่เคลื่อนไหว



รูปที่ 2.2 การวิเคราะห์การเคลื่อนไหวด้วยการลบพื้นหลังแบบเฟรมติดกัน

ผลลัพธ์ที่ได้จากการหาความแตกต่างระหว่างเฟรมที่สนใจและพื้นหลังซึ่งเป็นเฟรมที่ติดกัน พบว่าจะมีสัญญาณรบกวนปนอยู่เป็นจำนวนมาก โดยจะทำให้ได้การเคลื่อนไหวเฉพาะตรงขอบของวัตถุเท่านั้น โดยไม่สามารถหาการเคลื่อนไหวภายในวัตถุได้

### การหาวัตถุเคลื่อนไหวจากพื้นหลังด้วยค่าเฉลี่ย (Averaging background)

การหาวัตถุเคลื่อนไหวจากพื้นหลังด้วยค่าเฉลี่ยเป็นอีกวิธีหนึ่งที่ทำให้ผลลัพธ์ค่อยข้างดีและมีความเร็วในการทำงานสูง ซึ่งภาพพื้นหลังจะสามารถหาได้โดยการใช้หลักการของการนำเฟรมจำนวน  $n$  เฟรมมาหาค่าค่าเฉลี่ย (average) หรือค่ากลาง (median) ซึ่งเมื่อนำเฟรมที่ได้จากการหาค่าเฉลี่ยนี้มาหาค่าความแตกต่างกับเฟรมภาพที่ต้องการวิเคราะห์ ก็จะได้วัตถุที่สนใจ อย่างไรก็ตามการใช้วิธีนี้จำเป็นที่จะต้องใช้หน่วยความจำเป็นจำนวนมาก ( $n \times$  จำนวนหน่วยความจำของหนึ่งเฟรม) ซึ่งสามารถแก้ปัญหาได้ด้วยการประมาณการหาค่าเฉลี่ยโดยการใช้ค่า learning rate มาช่วยในการคำนวณ ซึ่งจะได้ ภาพพื้นหลังแบบไม่เลือกพื้นที่ (without selectivity) สมการในการคำนวณหาดังนี้

$$B_{i+1} = \alpha * F_i + (1 - \alpha) * B_i$$

จากสมการดังกล่าวจะพบว่าข้อมูลภาพหลายๆ เฟรมจะไม่ถูกนำมาเก็บไว้ในหน่วยความจำอีกต่อไป โดยการคำนวณจะใช้หน่วยความจำเพียงสามเฟรมเท่านั้น ได้แก่ เฟรมพื้นหลังล่าสุด ( $B_i$ ) เฟรมพื้นหลังปัจจุบัน ( $B_{i+1}$ ) และเฟรมของวัตถุสนใจล่าสุด ( $F_i$ ) ในขณะที่  $\alpha$  คือค่า learning rate โดยปกติแล้วมีค่าเท่ากับ 0.05 ซึ่งหมายความว่า จะกำหนดค่าความเข้มของเฟรมพื้นหลังปัจจุบันจะเท่ากับการนำเอาค่าความเข้มแสงจำนวน 95% จากเฟรมพื้นหลังก่อนหน้านั้นและอีก 5% จากข้อมูลภาพเฟรมใหม่ที่เข้ามา ค่า learning rate จะมีความสำคัญต่อการประมาณค่าพื้นหลังอย่างมาก ซึ่งหากมีค่าน้อยหมายความว่าภาพพื้นหลังจะมีความทนทานต่อสัญญาณรบกวนแต่จะทำให้ไม่สามารถปรับตัวกับการเปลี่ยนแปลงอย่างรวดเร็วของพื้นที่ที่สนใจได้ เช่น การเกิดฝนตกในทันใด การมีเมฆบังที่ทำให้พื้นที่นั้นมีแสงมากหรือน้อยอย่างปัจจุบันทันด่วน เป็นต้น

อัลกอริทึมนี้ยังสามารถทำการเพิ่มประสิทธิภาพได้โดยการเพิ่มเงื่อนไขการปรับค่าบนภาพพื้นหลังแบบเลือกพื้นที่ (With selectivity) ซึ่งมีหลักการง่ายๆ คือให้มีการปรับค่าในเฟรมพื้นหลังใหม่เฉพาะบริเวณพิกเซลที่ถูกจำแนกว่าเป็นพื้นหลัง (background) เท่านั้น ซึ่งสามารถอธิบายได้ตามสมการต่อไปนี้

$$B_{i+1}(x, y) = \alpha F_i(x, y) + (1 - \alpha) B_i(x, y) \quad \text{เมื่อ } F_i(x, y) \text{ เป็นพื้นหลัง และ}$$
$$B_{i+1}(x, y) = B_i(x, y) \quad \text{ถ้า } F_i(x, y) \text{ เป็นวัตถุสนใจ}$$

จากรูปแสดงผลการทดลองการหาวัตถุเคลื่อนไหวด้วยวิธีการหาภาพพื้นหลังจากค่าเฉลี่ยและมีลักษณะเป็นการปรับค่าตามพื้นที่





รูปที่ 2.3 การวิเคราะห์การเคลื่อนไหวด้วยการลบพื้นหลังแบบเลือกพื้นที่

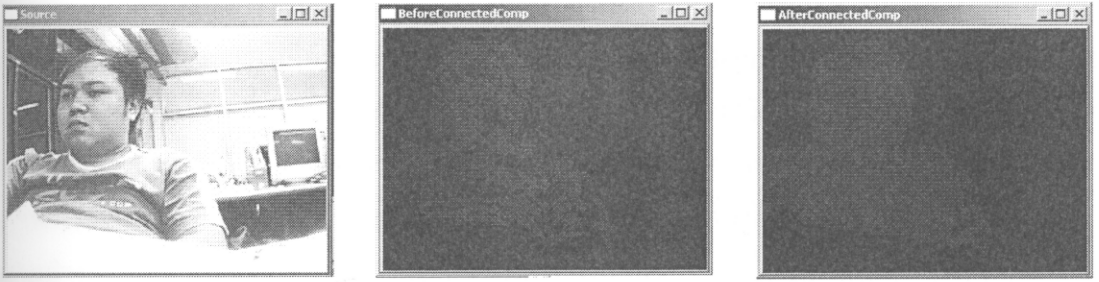
ลักษณะเด่นของวิธีนี้คือ จะได้วัตถุที่กำลังเคลื่อนไหวทั้งหมดโดยไม่มีรบกวนและในกรณีที่วัตถุหยุดการเคลื่อนไหวเป็นการชั่วคราว วิธีการนี้ก็ยังสามารถตรวจจับได้ ทั้งที่ไม่สูญเสียหน่วยความจำและมีการทำงานอย่างรวดเร็ว

### 2.3.2. การปรับปรุงวัตถุการเคลื่อนไหวด้วยวิธีทาง Morphological (Object quality improvement)

การปรับปรุงวัตถุการเคลื่อนไหวโดยกระบวนการนี้จะเป็นการทำให้ภาพของวัตถุที่สนใจมีสัญญาณรบกวนน้อยที่สุด ทั้งนี้เพื่อเพิ่มประสิทธิภาพในการระบุตัววัตถุได้อย่างเที่ยงตรงและแม่นยำมากยิ่งขึ้นนั่นเอง จากผลลัพธ์ของอัลกอริทึมแบบต่างๆ ก่อนหน้านี้จะเห็นว่าภาพวัตถุที่ไ้ยังมีพื้นที่ส่วนที่ไม่ได้ต้องการอยู่มาก ซึ่งจำเป็นต้องหาวิธีการลบส่วนเหล่านี้ออกไปโดยจะใช้หลักการวิเคราะห์พื้นที่ส่วนที่เชื่อมต่อกันหรืออยู่ในบริเวณใกล้เคียงกันโดยใช้วิธีทางด้าน Morphological โดยเทคนิคโอเปอเรชัน Open จะทำหน้าที่กำจัดพื้นที่รบกวนเล็กๆ และเทคนิคโอเปอเรชัน Close จะทำหน้าที่ในการสร้างพื้นที่เพิ่มหรือกลบหลุมอันทำให้สามารถเชื่อมต่อพื้นที่ที่ขาดหายไปหรือพื้นที่ที่ถูกแยกออกจากกันได้

ขั้นตอนต่อไปหลังจากได้ผ่านการลบสัญญาณรบกวนด้วยวิธีทาง Morphological ออกแล้วก็คือ การลบสัญญาณรบกวนที่อาจจะหลงเหลือโดยการหาคอนทราสต์ที่เป็นแบบ Close loop และลบ คอนทราสต์ที่เป็นคอนทราสต์แบบ Open loop หรือคอนทราสต์ที่มีความยาวรอบวัตถุน้อยเกินไป (วัตถุขนาดเล็กมากซึ่งถือว่าเป็นสัญญาณรบกวน)

ผลลัพธ์ที่ได้จากขั้นตอนนี้คือ วัตถุจะถูกนิยามด้วยเส้นกรอบรูปหรือพื้นที่ประมาณการที่เป็นสีเหลี่ยมผืนผ้าล้อมวัตถุนั้น ซึ่งสามารถนำไปใช้ในขั้นตอนต่อไป ได้แก่ การติดตามวัตถุและการระบุว่าวัตถุดังกล่าวเป็นบุคคลหรือไม่



รูปที่ 2.4 การปรับปรุงคุณภาพของวัตถุ

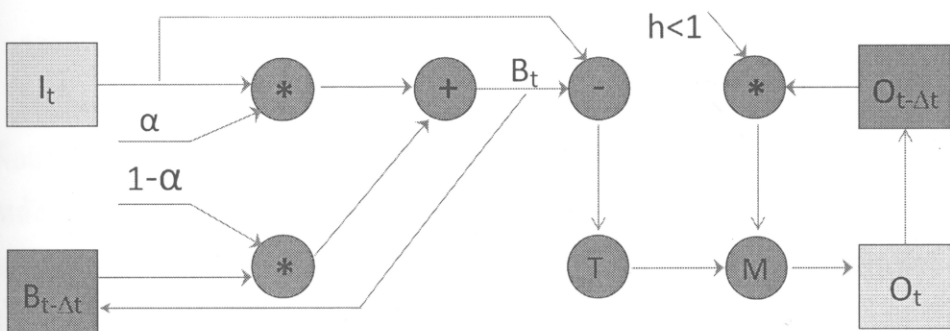
### 2.3.3. การติดตามและวิเคราะห์คุณสมบัติวัตถุ (Object Tracking & Characterization)

ขั้นตอนนี้มีความสำคัญเป็นอย่างยิ่งในการที่จะระบุว่าวัตถุที่ได้ถูกตรวจจับในเวลาหนึ่งๆ นั้นจะไปปรากฏ ณ ตำแหน่งใดในภาพถัดไป ซึ่งจะทำให้ทราบว่าในขณะนั้นว่าวัตถุที่อยู่ในบริเวณดังกล่าวมีการเคลื่อนที่ไปในทิศทางใดอย่างไร โดยปกติทั่วไปการติดตามวัตถุจะนำเอาคุณลักษณะเฉพาะของวัตถุนั้นๆ ไปใช้ เช่น รูปแบบลายหรือสีของวัตถุ ขนาดและทิศทาง เป็นในงานวิจัยนี้ได้คำนึงถึงการติดตามคุณลักษณะขนาด ทิศทางและบริเวณการตรวจพบวัตถุเท่านั้น ซึ่งจะเน้นไปที่การวิเคราะห์การเคลื่อนไหว โดยไม่ได้นำคุณลักษณะทางสีและลวดลายของวัตถุมาใช้

การวิเคราะห์ประวัติการเคลื่อนไหวของวัตถุมีวัตถุประสงค์เพื่อเพิ่มความเสถียรของวิธีการการตรวจจับวัตถุเคลื่อนไหวที่ได้จากการหาความแตกต่างระหว่างเฟรมที่สนใจและเฟรมพื้นหลัง โดยการนำเอาการเคลื่อนไหวของวัตถุก่อนหน้านั้น (ประวัติ) มาใช้พิจารณาประกอบวัตถุ ณ เวลาที่  $t$  ซึ่งจะกลายเป็นประวัติการเคลื่อนไหวของตัวมันเอง ณ เวลาที่  $t+1$  ซึ่งประวัติการเคลื่อนไหวนี้จะถูกเก็บไว้ช่วงเวลาหนึ่ง แล้วค่อยๆ หายไปในที่สุด โดยขึ้นอยู่กับค่าคงที่  $h$  ( $h < 1$ ) ตามสมการดังต่อไปนี้

$$O_t = \max(h \times O_{t-\Delta t}, O)$$

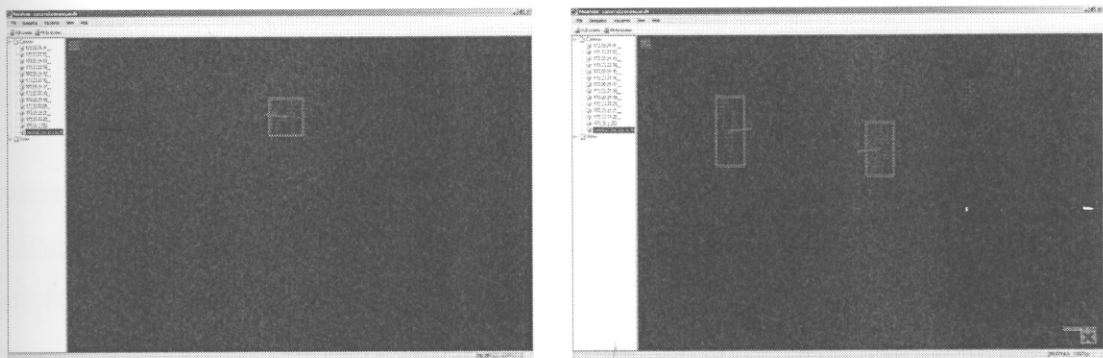
เมื่อ  $O_t$  และ  $O_{t-\Delta t}$  คือประวัติการเคลื่อนไหวของวัตถุปัจจุบันและประวัติก่อนหน้านั้นตามลำดับ  $O$  คือวัตถุเคลื่อนที่ที่ถูกตรวจจับได้ด้วยวิธีการลบพื้นหลัง ไดอะแกรมต่อไปนี้แสดงแผนผังรวมการทำการวิเคราะห์วัตถุด้วยการลบพื้นหลังและการคำนวณหาประวัติการเคลื่อนไหวของวัตถุ



รูปที่ 2.5 การตรวจจับวัตถุเคลื่อนไหวและการคำนวณประวัติ

ภาพประวัติการเคลื่อนที่ของวัตถุที่ได้นั้นสามารถนำมาใช้ในขั้นต่อไปคือ การคำนวณหาทิศทางการเคลื่อนที่ของวัตถุ และการแยกวัตถุ (Segmentation) ออกจากภาพด้วยการนำทิศทางการเคลื่อนที่ที่ได้ก่อนหน้านั้นไปพิจารณา การวิเคราะห์ทิศทางวัตถุจากภาพประวัตินี้สามารถทำได้ด้วยการนำวิธี Gradient มาใช้ จากสมการการคำนวณภาพประวัติพบว่าภาพที่ได้เป็นภาพที่ประกอบด้วยวัตถุที่มีความเข้มแสงลดหลั่นกันลงไปตามค่าคงที่  $h$  และ  $t$  ตามลำดับ ซึ่งหากใช้วิธีการหา Gradient กับภาพนี้แล้วก็จะทำให้ได้ทิศทางการเคลื่อนที่ของวัตถุที่มีทิศตั้งฉากกับขอบวัตถุ ณ เวลา  $t$  และ  $t-1$  ใดๆ

ด้วยกระบวนการดังกล่าวจะทำให้ได้คุณสมบัติของวัตถุ ทั้งขนาด ตำแหน่ง และทิศทางที่สามารถนำไปใช้ในการวิเคราะห์ที่มีความซับซ้อนมากขึ้น อันจะทำให้สามารถทราบถึงพฤติกรรมของวัตถุได้อย่างละเอียดมากขึ้น จากการทดสอบพบว่า เมื่อกำหนดค่าคงที่  $h$  มีค่าเข้าใกล้ 1 จะทำให้ได้ประวัติของวัตถุที่มีความยาวมากขึ้น ซึ่งในกรณีนี้จะทำให้ได้ค่าตำแหน่งและขนาดของวัตถุที่ผิดพลาดและมีความถูกต้องในเรื่องของทิศทางมากยิ่งขึ้น และเมื่อกำหนดให้  $h$  มีค่าเข้าใกล้ 0 ปรากฏว่าจะทำให้ได้ผลลัพธ์ในทางตรงกันข้ามกล่าวคือ ค่าตำแหน่งและขนาดของวัตถุจะมีความถูกต้องมากขึ้นแต่คำนวณทิศทางถูกต้องลดลง นอกจากนี้ยังพบว่าวิธีการนี้จะใช้ไม่ได้เมื่อวัตถุมีลักษณะที่เปลี่ยนรูปได้เช่น ค้อนไฟ เป็นต้น หรือในบางกรณีไม่เปลี่ยนรูป เช่น ท่าทางของมนุษย์ หากภาพนั้นเป็นภาพมูมโกล์ซึ่งจะทำให้ได้ผลการวิเคราะห์ที่ผิดพลาดคือ วัตถุชนิดเดียว (บุคคลหนึ่งคน) แต่ถูกวิเคราะห์ว่าเป็นหลายวัตถุ (มือ ศีรษะ และลำตัว เป็นต้น) ที่มีทิศทางแตกต่างกัน



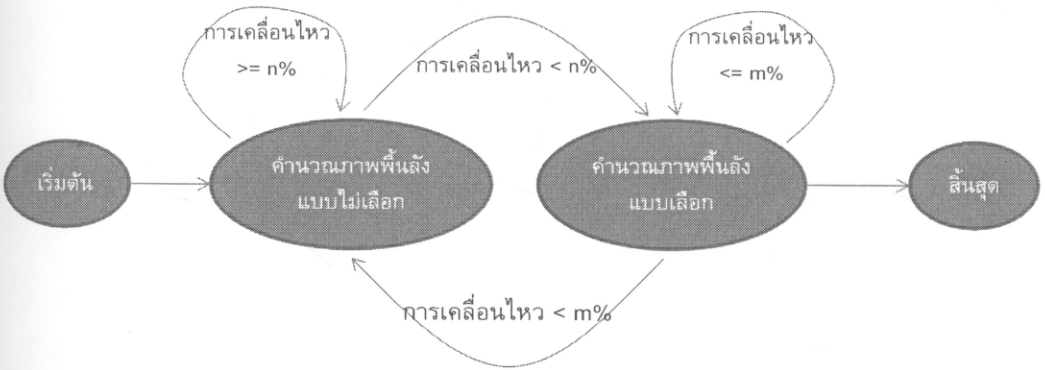
รูปที่ 2.6 การวิเคราะห์คุณสมบัติตำแหน่ง ขนาดและทิศทางของวัตถุ

### 2.3.4. ระบบโดยรวมการวิเคราะห์การเคลื่อนไหว (Motion analysis system)

ในขั้นตอนนี้จะพิจารณาระบบโดยรวมในการวิเคราะห์การเคลื่อนไหวของวัตถุภายในภาพเพื่อแก้ปัญหาข้อจำกัดของอัลกอริทึมที่ได้นำเสนอนั้นคือ การวิเคราะห์นี้ต้องการใช้ภาพพื้นหลังที่หลังที่คุณภาพดีถึงจะได้วัตถุเคลื่อนที่ที่ถูกต้อง ซึ่งภาพพื้นหลังนี้จะต้องถูกคำนวณอย่างต่อเนื่องจากสภาพแวดล้อมนั้น ๆ โดยไม่มีวัตถุเคลื่อนไหวใดๆ เลย ซึ่งในสภาพพื้นที่จริงที่จะใช้งานจะไม่สามารถควบคุมให้ได้ตามที่ต้องการได้ เช่น ทางเข้าออกของสถานที่ต่างๆ ที่การเคลื่อนไหวมีอยู่ตลอดเวลา ซึ่งสามารถแยกการวิเคราะห์ภาพพื้นหลังได้ในกรณี ดังนี้คือ

- ก) ในกรณีที่ภาพพื้นหลังเพิ่งทำการคำนวณ หรือ ในกรณีที่ม็ีวิตตุกำลังเคลื่อนไหวเป็นจำนวนมาก ก็ให้มีการคำนวณพื้นหลังแบบไม่เลือกพื้นที่ (without selectivity) และเมื่อพื้นหลังที่ได้จากการคำนวณมีค่าถูกต้องมากแล้ว ก็ให้เปลี่ยนการคำนวณพื้นหลังเป็นแบบเลือกพื้นที่ (with selectivity) ต่อไป
- ข) ในกรณีที่ภาพพื้นหลังกำลังคำนวณแบบเลือกพื้นที่ (with selectivity) และเมื่อมีการเคลื่อนไหวของพื้นหลังมากหรือมีการเปลี่ยนแปลงสภาวะแวดล้อมอย่างกะทันหัน เช่น การเปลี่ยนของความสว่าง และเกิดฝนตก เป็นต้น ก็มีการเปลี่ยนการคำนวณภาพพื้นหลังกลับเป็นแบบไม่เลือกพื้นที่ใหม่อีกครั้ง

จากแนวคิดข้างต้นสามารถออกแบบระบบการวิเคราะห์การเคลื่อนไหวของวัตถุให้อยู่ในลักษณะเป็นสเตตไดอะแกรม (State diagram) ดังต่อไปนี้คือ



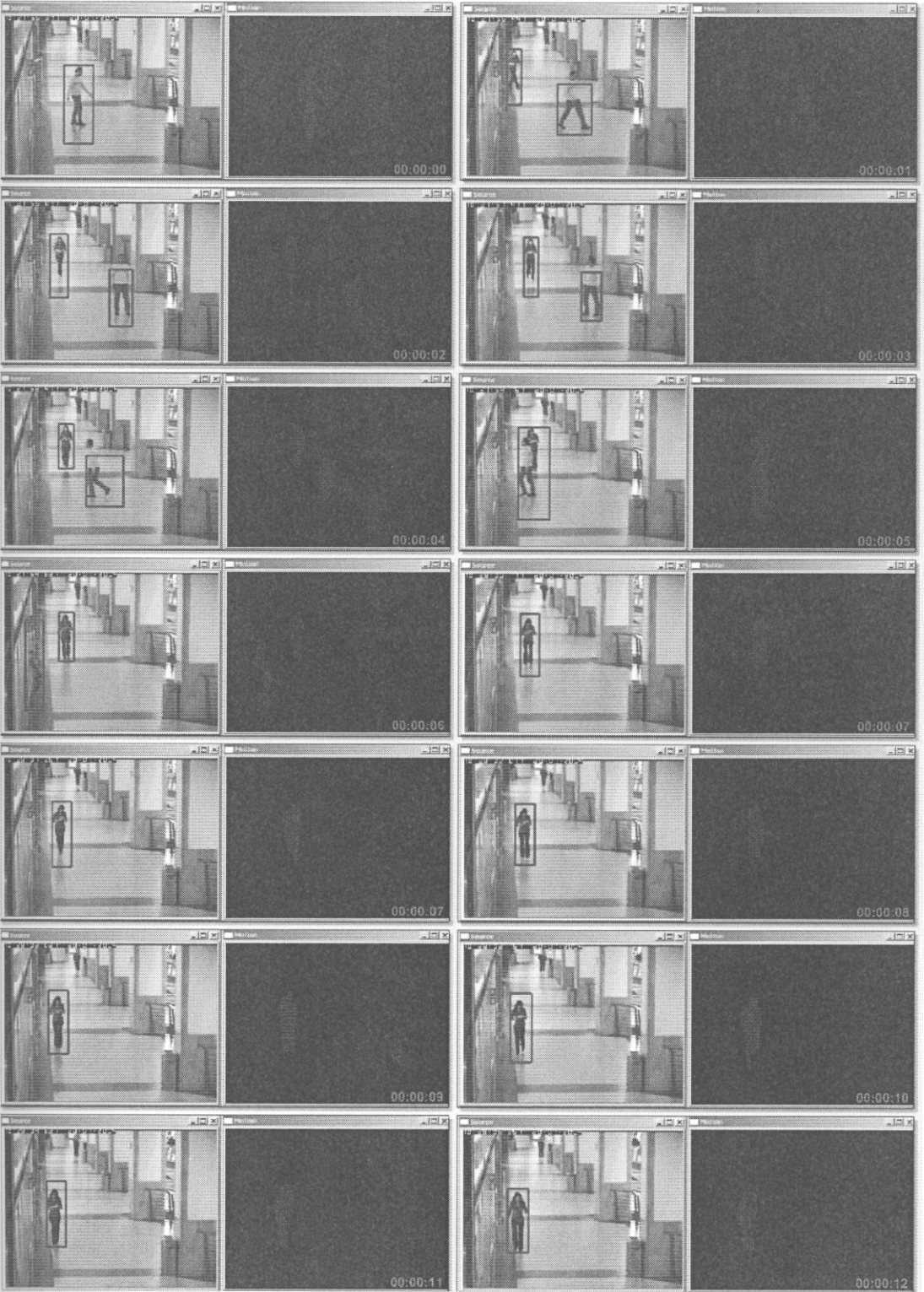
รูปที่ 2.7 ไดอะแกรมแสดงระบบโดยรวมการวิเคราะห์การเคลื่อนไหวของวัตถุ

กำหนดให้  $n$  และ  $m$  คือจำนวนเปอร์เซ็นต์การเคลื่อนไหวที่เกิดขึ้นทั้งหมดเมื่อเทียบกับขนาดของภาพ โดย  $n$  มีค่าน้อยกว่า  $m$  เสมอ ซึ่งโดยปกติแล้วค่าของ  $n$  จะต้องมิต่ำกว่าที่หมายความว่าการคำนวณค่าพื้นหลังจะให้ความสำคัญต่อการเคลื่อนไหวของวัตถุให้น้อยที่สุดเท่าที่จะทำได้ถึงจะเปลี่ยนไปเป็นการคำนวณแบบเลือกพื้นที่ และค่าของ  $m$  จะต้องมิต่ำกว่าที่สมควรที่บ่งชี้ถึงการเคลื่อนไหวที่มากจนเกินไปที่จะทำให้การวิเคราะห์วัตถุเคลื่อนไหวเริ่มผิดพลาด การบวนการนี้คือว่าเป็นหลักการทำงานหลักของโปรแกรม ก่อนที่จะมีการส่งผลลัพธ์ของพื้นหลังที่ได้ไปคำนวณหาวัตถุเคลื่อนที่และคุณสมบัติต่างๆ ตามที่ได้นำเสนอมาแล้วก่อนหน้านี้

## 2.4 ผลการทดลองและวิเคราะห์

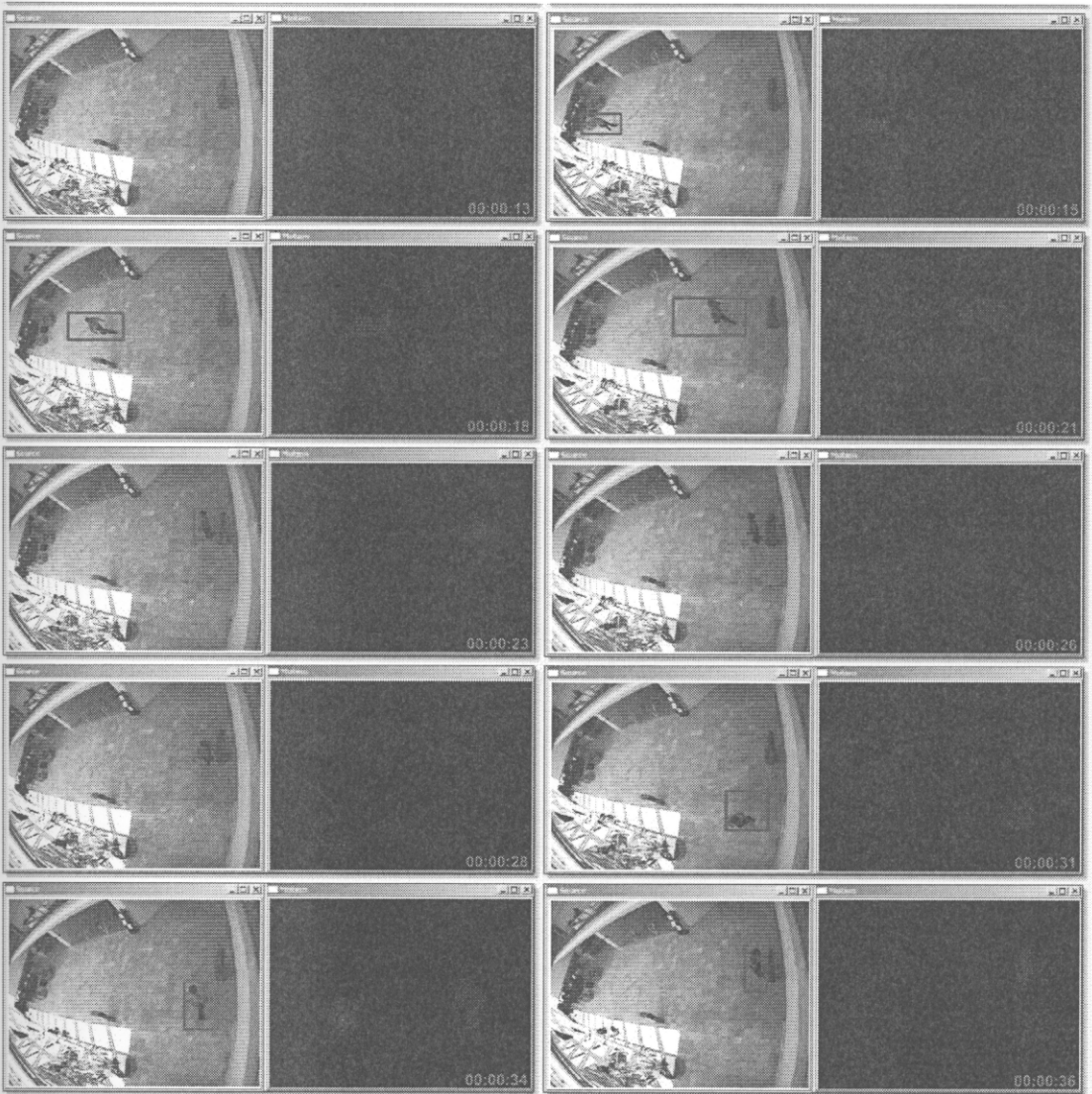
ในการทดสอบการวิเคราะห์ในครั้งนี้ ได้นำไฟล์ฐานข้อมูลจากโปรเจกต์ CAVIAR [http://www-prima.inrialpes.fr/PETS04/caviar\\_data.html](http://www-prima.inrialpes.fr/PETS04/caviar_data.html) มาใช้ซึ่งประกอบ ซึ่งเป็นฐานข้อมูลมาตรฐานที่ในการวัดหรือเปรียบเทียบการทำงานของอัลกอริทึม เช่น บุคคลเดินคนเดียว บุคคลพบกับคนอื่น ๆ มุมมองจากร้านขายของ การต่อสู้ และการวางสิ่งของ เป็นต้น

# ตัวอย่างที่ 1



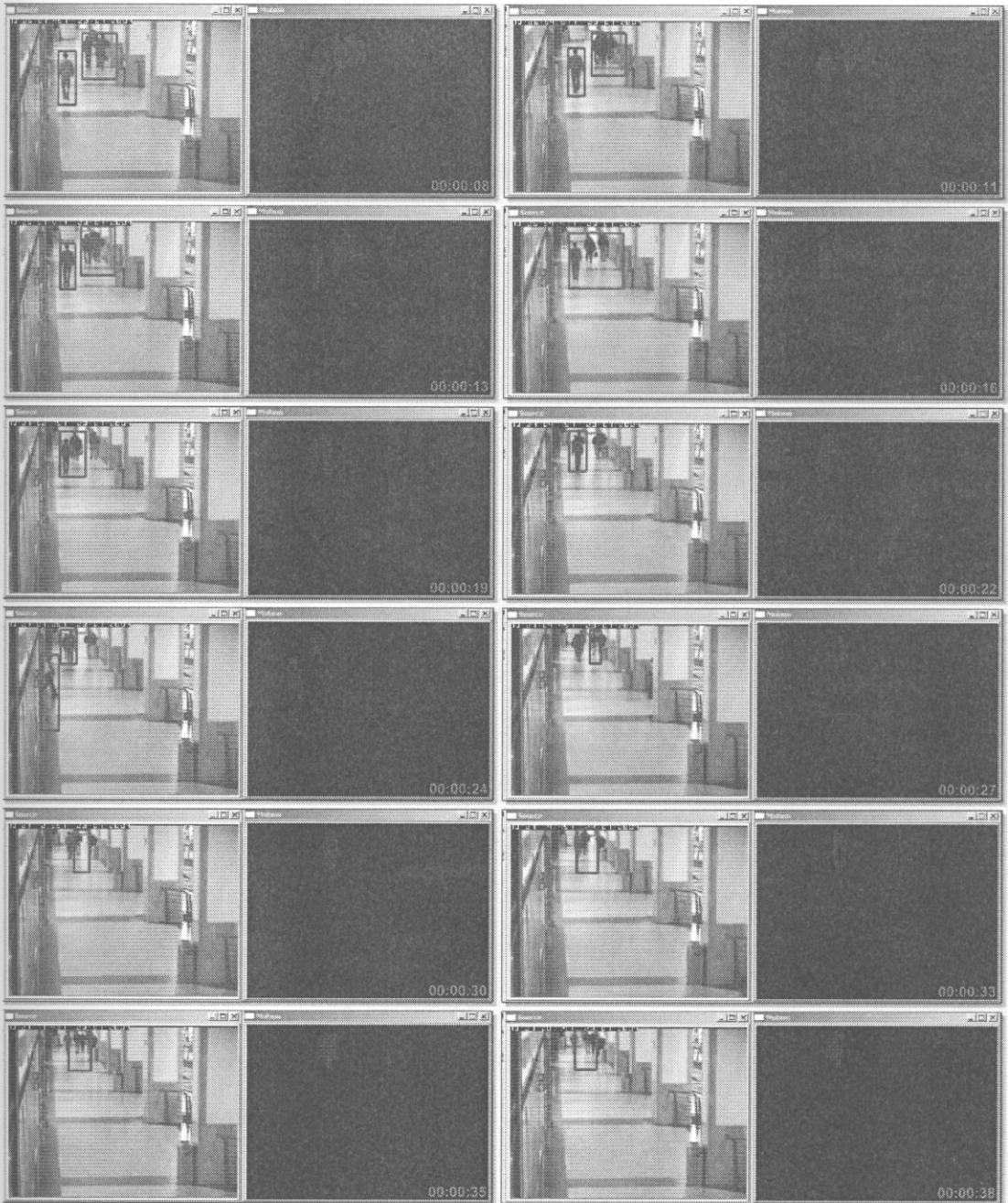
รูปที่ 2.8 แสดงตัวอย่างการทดสอบกับไฟล์ภาพวิดีโอ

## ตัวอย่างที่ 2



รูปที่ 2.9 แสดงผลลัพธ์ของโปรแกรมในการตรวจจับบุคคล

### ตัวอย่างที่ 3



รูปที่ 2.10 แสดงผลลัพธ์ของการทดสอบการตรวจจับจากไฟล์วิดีโอ

จากผลการทดลองพบว่าวิธีการวิเคราะห์ที่ได้นำเสนอนี้สามารถค้นหาวัดตุเคลื่อนไหวได้เป็นอย่างดี อย่างไรก็ตาม ยังไม่มีการติดตามวัตถุที่ดีเท่าที่ควร อันที่จะสามารถแก้ปัญหาต่าง ๆ ได้ เช่น การเกิดการสวนกันของวัตถุที่เคลื่อนไหวจากหลายวัตถุ หรือ กรณีเมื่อเกิดการบังเกิดขึ้น เป็นต้น

## 2.5 สรุปผล

งานวิจัยนี้ได้พัฒนาเทคนิคขั้นพื้นฐานที่มีความสำคัญมากสำหรับการวิเคราะห์ภาพจากระบบกล้องวิดีโอวงจรปิดที่มีความซับซ้อนและนำไปใช้งานได้จริง โดยได้มุ่งเน้นไปที่การนำเทคนิคการลบพื้นหลังมาใช้ โดยได้เพิ่มขั้นตอนต่าง ๆ อันทำให้ผลการวิเคราะห์มีความถูกต้องมากขึ้น นอกจากนี้ได้ทำการออกแบบระบบการค้นหาที่มีลักษณะเป็นการทำงานแบบมีสถานะเพื่อที่จะสามารถปรับการทำงานของระบบให้เหมาะสมกับสิ่งแวดล้อมได้อย่างทันท่วงที วิธีการที่ได้นำเสนอนี้ได้ถูกทดสอบกับฐานข้อมูลมาตรฐาน ซึ่งผลพบว่าได้ผลลัพธ์เป็นที่น่าพอใจ อย่างไรก็ตาม การวิเคราะห์นี้ถือได้ว่ายังก่อนไปทางการวิเคราะห์ขั้นต้นที่จำเป็นต้องมีการปรับแต่งเพื่อให้สามารถแก้ปัญหาต่าง ๆ ที่ซับซ้อนได้มากขึ้น เช่น การติดตามหลายบุคคลที่เดินสวนกัน การเกิดการบังไปชั่วระยะเวลาหนึ่ง หรือการลบเงาที่เกิดขึ้นแหล่งกำเนิดแสง เป็นต้น

## 2.6 บรรณานุกรม

- [1] Aggarwal, J.K. Cai, Q., Human motion analysis: a review, IEEE Proceedings Nonrigid and Articulated Motion Workshop, 16 Jun 1997, p. 90-102
- [2] Weiming Hu Tieniu Tan Liang Wang S. Maybank, A Survey on Visual Surveillance of Object Motion and Behaviors, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Aug. 2004, Volume: 34, Issue: 3, page(s): 334-352, ISSN: 1094-6977
- [3] D. M. Gavrila, The Visual Analysis of Human Movement: A Survey, Computer Vision and Image Understanding: CVIU, 1999, Volume: 73, Issue: 1, page(s): 82-98
- [4] Massimo Piccardi, Background subtraction techniques, Faculty of Engineering, UTS, April 2004



## บทที่ 3

### การวิเคราะห์คุณภาพของภาพใบหน้าสำหรับการค้นหาเฟรมภาพ สำคัญ เพื่อใช้ในระบบเฝ้าระบบ

#### Key Frame Extraction from Video Sequence Using a Face Quality Index Applied to Surveillance System

ในบทนี้จะนำเสนอวิธีการวัดคุณภาพของข้อมูลภาพใบหน้าที่ได้จากการตรวจจับในภาพวิดีโอ การวิเคราะห์ดังกล่าวจะนำมาประยุกต์ใช้ในระบบการรักษาความปลอดภัยข้อมูลภาพใบหน้าของเฟรมใดมีคุณภาพมากกว่าเฟรมอื่นในช่วงเวลาใดๆ จะเรียกว่าการค้นหาเฟรมสำคัญ (Key frame) ซึ่งสามารถหาได้โดยการพิจารณาค่าโควาเรียนเมตริกของภาพใบหน้านั้นๆ เมื่อเปรียบเทียบกับภาพใบหน้าในเฟรมที่ใกล้เคียง ณ ช่วงเวลาหนึ่ง ซึ่งการทดสอบได้ดำเนินการกับข้อมูลกลุ่มหนึ่งซึ่งให้ผลเป็นที่น่าพอใจ

#### 3.1 ปัญหาและความสำคัญ

ในระบบวิดีโอรักษาความปลอดภัย (Surveillance System) การแยกเฟรมสำคัญ (key frame) เป็นประเด็นที่สำคัญในหลายๆ งานวิจัย [1][2] ซึ่งจุดประสงค์ของงานวิจัยนี้ก็เพื่อแก้ปัญหาของวิดีโอที่ไม่สามารถระบุตัวตนได้อันเนื่องมาจากภาพมีความละเอียดต่ำ ซึ่งโดยส่วนใหญ่จะเกิดจากการบีบอัดไฟล์ข้อมูล ถึงแม้ว่ากล้อง IP-camera ที่ใช้งานกันอยู่ในปัจจุบันนั้นมีความละเอียดสูง (1-2M พิกเซล) แต่ในงานประยุกต์ที่ใช้กันอยู่โดยส่วนใหญ่มีการจำกัดแบนด์วิธที่ใช้ในการส่งข้อมูล ดังนั้นจึงไม่สามารถส่งข้อมูลที่มีคุณภาพสมบูรณ์หรือเหมือนต้นฉบับได้ ซึ่งโดยทั่วไปแล้วความละเอียดของภาพจะใช้งานในระดับ VGA หรือ CIF ที่มีการเข้ารหัสแบบ MPEG4 หรือ H.264 เป็นต้น

สำหรับงานวิจัยนี้ในส่วนแรกจะทำการระบุเฟรมที่สำคัญ (key frame) จากภาพที่มีความละเอียดต่ำหรือคุณภาพต่ำจากวิดีโอที่ได้จากกล้อง IP-camera หากเกิดเหตุการณ์สำคัญหรือที่สนใจก็จะบ่งชี้ว่าเฟรมนั้นเป็นเฟรมที่สำคัญ (key frame) ก็จะทำการร้องขอข้อมูลภาพที่มีความละเอียดสูงสุดของเฟรมภาพปัจจุบันมาได้ ถึงแม้ว่าบางครั้งจะมีการลดจำนวนแบนด์วิธแต่ก็ยังสามารถรับภาพที่มีคุณภาพดีที่สุดในช่วงเหตุการณ์นั้นได้ ซึ่งในงานวิจัยนี้จะมุ่งเน้นเฉพาะเทคนิคการวิเคราะห์วิดีโอสำหรับการเลือกเฟรมสำคัญเท่านั้น โดยจุดประสงค์ของวิธีการนี้จะวิเคราะห์หาดัชนีคุณภาพของใบหน้าโดยการคำนวณค่าโควาเรียนเมตริก (Covariance Matrix) ของภาพใบหน้าทำการค้นหาพบ

## 3.2 ปรีทัศน์วรรณกรรม

ในการค้นหาวัตถุและติดตามวัตถุได้ใช้อัลกอริทึมที่สำคัญ 2 อัลกอริทึม ได้แก่ วิธีการ Haar-like และวิธีการ Camshift ซึ่งในงานวิจัยนี้ได้ประยุกต์วิธีการดังกล่าว

### 3.2.1 การค้นหาวัตถุโดยใช้ Haar-Like

เทคนิคในการค้นหาวัตถุมีพื้นฐานจาก Haar-like feature ซึ่งมีผู้เสนอข้อเสนอนี้หลายท่าน แต่ในงานวิจัยนี้ได้มุ่งประเด็นที่สนใจตามวิธีการของ Viola และคณะ [3][5] detection framework โดยจะดำเนินการแบ่งออกเป็นสองขั้นตอนที่สำคัญคือ ในลำดับแรกจะทำการเรียนรู้คุณลักษณะที่เป็นของ Haar-like (คุณลักษณะของรูปสี่เหลี่ยมจำนวน 180,000 รูปที่เหมาะสมกับกับแต่ละภาพในหน้าต่างย่อย) และมีการเรียนรู้ของภาพ positive และภาพ negative ซึ่งอัลกอริทึม Adaboost จะทำการเลือกคุณลักษณะสำคัญออกมา อัลกอริทึมนี้ถูกออกแบบเพื่อที่จะเลือกคุณลักษณะเพียงลักษณะเดียวที่สามารถแยกภาพ positive ที่แตกต่างกันมากที่สุดออกจากภาพ negative โดยใช้ฟังก์ชันการแบ่งประเภท  $h_j(x)$  ที่ประกอบไปด้วยคุณลักษณะเฉพาะ (feature)  $f_j(x)$  และ เทรซโฮลด์  $\theta_j, p_j$  เพื่อแสดงทิศทางที่แตกต่างกัน และ  $x$  คือภาพที่มีขนาด 24x24 pixel

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{อื่นๆ} \end{cases}$$

ในลำดับต่อไป ก็จะเป็นการสร้าง stage เพื่อให้เหมาะสำหรับการเรียนรู้และการตรวจจับได้อย่างถูกต้องและรวดเร็วที่สุดโดยใช้อัลกอริทึม Adaboost

### 3.2.2 การติดตามวัตถุโดยใช้ Camshift

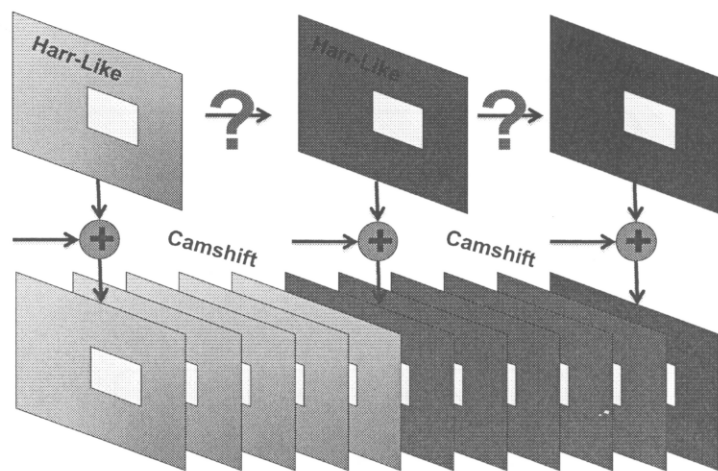
Camshift [4] เป็นอัลกอริทึมที่ใช้ในการติดตามวัตถุ ซึ่งได้มาจากอัลกอริทึม Mean Shift อันเป็นเทคนิคแบบ Non-Parametric ด้วยวิธีนี้ สีของวัตถุจะถูกแทนด้วยการกระจายค่าความน่าจะเป็นที่ถูกติดตามในลำดับวิดีโอโดยค่อยๆ จะทำการเพิ่มค่าเกรเดียนท์ (Gredient) เพื่อที่จะหา Nearest Dominant Peak สำหรับการติดตามวัตถุโดยใช้ Camshift จะทำการอัปเดตการกระจายความน่าจะเป็นด้วย ซึ่งทำการติดตามการเคลื่อนไหวของวัตถุที่การเปลี่ยนแปลงได้ดีกว่า อัลกอริทึมนี้สามารถคำนวณได้ต่อไปเรื่อยๆ โดยในแต่ละวิดีโอเฟรม จะมีการกระจายความน่าจะเป็นของสีด้วยวิธีการฮิสโตแกรมของโมเดลสี HSV ซึ่งจุดศูนย์กลางและขนาดของการติดตามของวัตถุจะถูกคำนวณจากความน่าจะเป็นของสีในภาพโดยใช้จุดเซนทรอยด์และพื้นที่ของวัตถุ ซึ่งขนาดของวัตถุปัจจุบันและตำแหน่งของวัตถุที่ทำการติดตามนั้น จะถูกรายงานและถูกใช้เพื่อกำหนดขนาดและตำแหน่งของหน้าต่างที่จะใช้ในการค้นหาในเฟรมวิดีโอถัดไป ซึ่งกระบวนการนี้จะถูกทำซ้ำสำหรับการติดตามในครั้งต่อไปนั่นเอง

### 3.3 ระเบียบวิธีการวิเคราะห์คุณภาพใบหน้า

ในส่วนนี้จะอธิบายถึงวิธีการที่ได้เสนอในรายละเอียดของการเลือกเฟรมสำคัญ (key frame) จากลำดับของวิดีโอโดยการวัดดัชนีคุณภาพของใบหน้า เทคนิคที่ใช้จะประกอบไปด้วย 2 ขั้นตอน คือ การค้นหาและติดตามใบหน้า และการวัดดัชนีคุณภาพของใบหน้า

#### 3.3.1 การค้นหาและติดตามใบหน้า

จุดประสงค์หลักสำหรับงานในขั้นตอนนี้ คือ เพื่อที่จะหาพื้นที่ของใบหน้าในแต่ละเฟรมของวิดีโอ ซึ่งจะถูกใช้เพื่อสนับสนุนในขั้นตอนการวัดคุณภาพของใบหน้า โดยได้รวม 2 วิธีการภายใต้หลักการของ Viola [3] สำหรับการค้นหาใบหน้า และ Bradski [4] สำหรับการติดตามใบหน้า เทคนิคในการค้นหาใบหน้าได้ใช้วิธีการ Haar-like ซึ่งมีการค้นหาที่แม่นยำในขณะที่ใช้เวลาในการคำนวณน้อย อย่างไรก็ตามในสถานการณ์ที่เป็นจริงสำหรับวิดีโอจากระบบกล้อง CCTV ซึ่งมีความละเอียดของภาพมาก (โดยทั่วไปคือ VGA หรือ CIF) แต่คุณภาพของภาพจะลดลงเมื่อมีการบีบอัดของข้อมูล ดังนั้น อัตราการค้นหาใบหน้าก็ลดลงและจะมีความผิดพลาดมากขึ้น จากความสัมพันธ์ของอัลกอริทึม Camshift จึงใช้วิธีการนี้ทำการติดตามวัตถุที่มีการติดตามแบบค่อยๆ เพิ่มขึ้นของการกระจายของความน่าจะเป็น ซึ่งในสถานการณ์ที่เป็นจริงสีของพื้นหลังหรือพื้นผิวของวัตถุจะเหมือนกันกับพื้นที่ที่ใช้ในการติดตาม ดังนั้น Camshift จะมีการอัปเดตการกระจายความน่าจะเป็นแบบไดนามิก ด้วยเหตุนี้จะทำให้การติดตามพื้นที่ที่ผิดพลาดน้อยลง และมีความแม่นยำที่เพิ่มขึ้นสำหรับการค้นหาและติดตามใบหน้า



รูปที่ 3.1: ภาพรวมการทำงานของอัลกอริทึม Harr-Like และ Camshift

จากวิธีการดังกล่าวจึงได้รวมวิธีการทั้งสอง เพื่อที่จะให้ความแม่นยำในการค้นหาและติดตามใบหน้ามากยิ่งขึ้น โดยกระบวนการทั้งหมดแสดงได้ดังรูปที่ 3.1 ซึ่งมีการทำงานดังนี้ คือ สำหรับแต่ละวิดีโอเฟรม จะทำการค้นหาใบหน้าโดยใช้อัลกอริทึมของ Haar-like ถ้าพบส่วนที่เป็นใบหน้า ฮิสโตแกรมส่วนนั้นจะถูกคำนวณและใช้การกระจายความน่าจะเป็นสำหรับการกำหนดค่าเริ่มต้นของ Camshift หรือเพื่อการปรับปรุงค่า ถ้าไม่มีใบหน้าที่ค้นพบในเฟรมนั้น

พื้นที่ที่ใช้ในการติดตามได้จากการคำนวณของอัลกอริธึม Camshift โดยอัตโนมัติ สำหรับการตรวจสอบความถูกต้องของภาพที่ถูกค้นพบและการติดตามใบหน้าทีพบนั้น สามารถตรวจสอบความถูกต้องได้ง่าย โดยการกรองให้ความถี่ต่ำผ่าน (Low-pass Filter) นั้นเอง

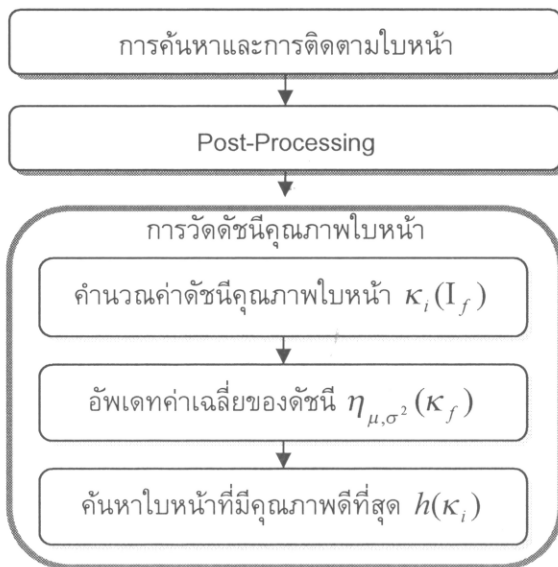
รูปที่ 3.2 ได้แสดงผลลัพธ์ในส่วนของการค้นหาและติดตามใบหน้าตามวิธีการที่นำเสนอ โดยในขั้นตอนส่วนสุดท้ายได้แสดงภาพใบหน้าที่ถูกแยกจากทุกๆ เฟรมในลำดับของวิดีโอ



รูปที่ 3.2: ผลลัพธ์จากการรวมวิธีการของ Harr-like และ Camshift

### 3.3.2 การแยกเฟรมที่สำคัญเพื่อใช้วัดดัชนีคุณภาพของใบหน้า

การวัดคุณภาพของใบหน้าที่มีจุดประสงค์สำหรับการกำหนดเฟรมสำคัญจากวิดีโอ ซึ่งสัมพันธ์กับคุณลักษณะที่ปรากฏให้เห็นที่ดีที่สุดสำหรับคุณลักษณะเฉพาะของใบหน้า ในขั้นตอนนี้สามารถถูกแบ่งเป็น 2 กระบวนการด้วยกัน คือ Post-processing และการวัดดัชนีคุณภาพของใบหน้า ตามที่แสดงกระบวนการดังรูปที่ 3



รูปที่ 3.3: กระบวนการวัดดัชนีคุณภาพใบหน้าที่ดีที่สุด

ในกระบวนการของ Post-processing การค้นหาใบหน้าจากทุกๆ เฟรมถูกปรับให้มีขนาด 150x150 พิกเซล ตาม bilinear interpolation และมีการประยุกต์ใช้ Histogram Equalization เพื่อที่จะลดสัญญาณรบกวนที่เกิดจากจำนวนของ pixel และความเข้มของสีที่

เกิดจากการเปลี่ยนของแสง ซึ่งจะเป็นตัวเชื่อมต่อกับวิธีการของการวัดคุณภาพของใบหน้า หลังจาก Post-processing การวัดคุณภาพของใบหน้าสามารถถูกคำนวณ โดยการกำหนด  $\kappa_i(I_f)$  เป็นดัชนีคุณภาพของใบหน้า ณ เฟรมนั้น ซึ่งสามารถเขียนได้ดังสมการดังนี้

$$\kappa_i(I_f) = \text{Trace}(\Sigma(I_f)) + \text{Trace}(\Sigma(I_f^T))$$

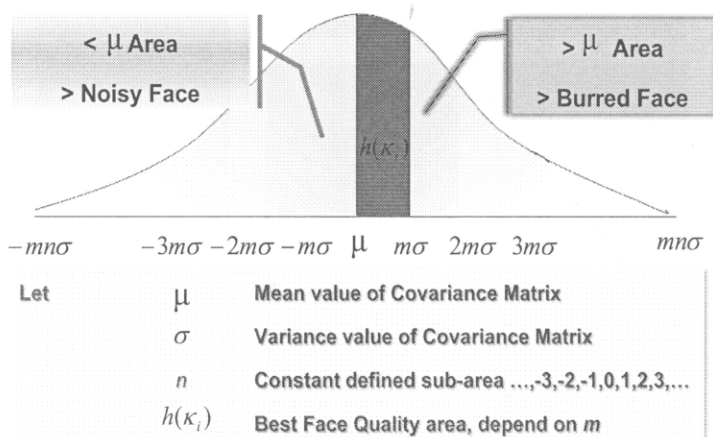
ในที่นี้  $\Sigma$  คือโควาเรียนซ์เมตริก (covariance matrix) ของภาพใบหน้า  $I_f$  ฟังก์ชันนี้จะให้ค่าที่มีค่ามาก ๆ เมื่อภาพเป็นภาพเบลอ และมีสัญญาณรบกวนน้อย ดังนั้น ฟังก์ชันนี้จะแสดงให้เห็นว่ามีแนวโน้มของค่าในแง่ของค่าที่มีสัญญาณรบกวน ภาพที่ชัด และภาพที่มัว โดยไม่มีการแยกส่วนของค่าที่สามารถบ่งชี้ได้ว่าเป็นภาพใบหน้าที่ดีที่สุด ซึ่งในความเป็นจริงแล้วค่านี้จะขึ้นอยู่กับลักษณะทางกายภาพ ณ เวลานั้น เช่น คุณลักษณะของการเคลื่อนที่ของวัตถุ การเปลี่ยนแปลงระหว่างความสว่างและความมืด (Contrast) ในสภาพแวดล้อมและอื่นๆ ดังนั้นวิธีการที่ได้เสนอจะพิจารณาภาพที่มีคุณภาพดีที่สุดในคาบเวลาสั้น ๆ ณ ช่วงเวลาหนึ่งเท่านั้น โดยในขั้นแรกจะคำนวณค่าที่สามารถอ้างอิงและกำหนดช่วงของค่าที่สามารถแทนคุณภาพของใบหน้าที่ดีที่สุดที่มี  $\kappa_i(I_f)$  รอบๆค่าเหล่านั้น จากนั้นจึงได้พิจารณา Gaussian model ที่เป็น low-pass filter  $\eta$  สำหรับการกำหนดค่าอ้างอิง  $\mu$  ซึ่งจำเป็นต้องมีการอัปเดตค่าอย่างต่อเนื่อง

$$\eta_{\mu,\sigma^2}(\kappa_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\kappa_i-\mu)^2}{2\sigma^2}}$$

ภาพที่มีคุณภาพที่ดีที่สุดสามารถถูกค้นพบด้วยฟังก์ชันดัชนี ดังนี้

$$h(\kappa_i) = \begin{cases} 1 & \text{if } \mu_{\kappa_i} + nm\sigma_{\kappa_i} < \kappa_i < \mu_{\kappa_i} + (n+1)m\sigma_{\kappa_i} \\ 0 & \text{อื่นๆ} \end{cases}$$

ซึ่ง  $n$  และ  $m$  คือพารามิเตอร์ที่แทนตำแหน่งในแต่ละวินโดวย่อยที่อ้างอิงค่าเฉลี่ยและขนาดของหน้าต่างย่อย ตามลำดับ พารามิเตอร์เหล่านี้ถูกกำหนดจากการทดลอง ซึ่งรายละเอียดจะอธิบายในส่วนต่อไป รูปที่ 3.4 ได้แสดงดัชนีคุณภาพหน้าที่ดีที่สุด  $h(\kappa_i)$



รูปที่ 3.4: ดัชนีคุณภาพใบหน้าที่ดีที่สุด

### 3.4 ผลการทดลองและวิเคราะห์

เพื่อที่จะประเมินผลการทดลองของระบบที่ได้เสนอ จึงกำหนดพารามิเตอร์  $m$  และ  $n$  และทำการทดสอบประสิทธิภาพของระบบโดยใช้สถานการณ์จริง สำหรับการหาค่าพารามิเตอร์  $m$  และ  $n$  นั้น ในขั้นแรกจะทำการเรียนรู้ลักษณะของใบหน้า 10 แบบ (ทั้งผู้หญิงและผู้ชาย) ในหลายๆ ลักษณะ (มีสัญญาณรบกวนมาก, มีสัญญาณรบกวนน้อย, ชัดเจน, มัวน้อย, มัวมาก) แสดงได้ดังรูปที่ 3.5 เพื่อที่จะกำหนดพารามิเตอร์  $m$  และ  $n$  ซึ่งจากการทดสอบให้ผลว่าค่า  $m$  และ  $n$  ที่เหมาะสมที่สุดสำหรับภาพใบหน้าที่ชัดเจนที่สุด คือ  $n = 0$  และ  $m = 1/15$  นั่นคือ ภาพใบหน้าที่มี  $\sigma_i$  มากกว่า  $\mu_{\kappa_i}$  และน้อยกว่า  $\mu_{\kappa_i} + \sigma_{\kappa_i}/15$  จะถูกพิจารณาว่าเป็นภาพที่มีคุณภาพใบหน้าที่ดีที่สุด



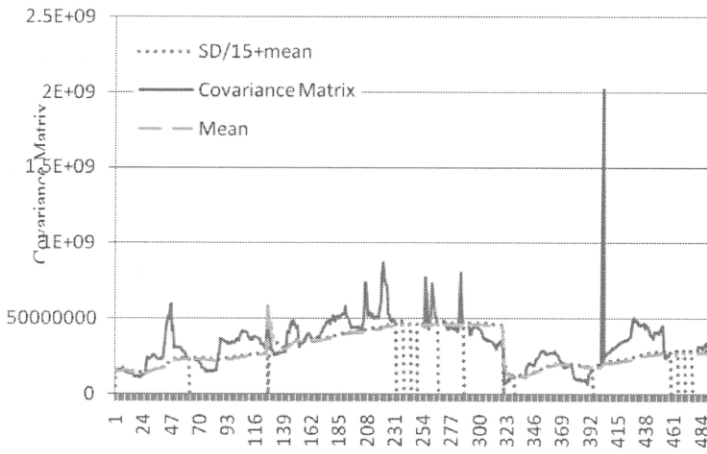


รูปที่ 3.5: ข้อมูลภาพที่ใช้ในการทดสอบ

สำหรับการทดสอบประสิทธิภาพของระบบ จึงได้ทำการสร้างการทดสอบจากกลุ่มของข้อมูลวีดีโอ 4 ชุด ที่มีระยะเวลา 1 นาที และ 30 เฟรมต่อวินาที แต่ละลำดับของวีดีโอจะประกอบไปด้วยลักษณะการทดสอบดังนี้

- 1) ภาพใบหน้าที่มีัวจากการเคลื่อนไหวที่เร็วหรือจากกล้องที่อยู่ไกลมาก
- 2) ภาพที่มีสัญญาณรบกวน
- 3) ภาพใบหน้าที่ชัด

รูปที่ 3.6, 3.7, 3.8 และ 3.9 เป็นกราฟของการวัดดัชนีคุณภาพของหน้า  $\kappa_f(I_f)$  ค่าเฉลี่ยของค่าโควาเรียนเมตริกของใบหน้า  $\eta_{\mu, \sigma^2}(\kappa_f)$  และภาพที่มีคุณภาพดีที่สุดแทนด้วยฟังก์ชัน  $h(\kappa_f)$  จากการทดสอบแสดงว่าค่า  $\kappa_f(I_f)$  จะแตกต่างกันในแต่ละการทดลอง เนื่องจากการทดสอบมีปัจจัยที่ขึ้นอยู่กับพื้นผิวของใบหน้า และสภาพแวดล้อมด้วย ดังนั้นใบหน้าที่มีคุณภาพที่ดีที่สุดสามารถถูกกำหนดขึ้นโดยวิธีการดังกล่าวซึ่งจะทำการค้นหาใบหน้าที่ดีที่สุดเฉพาะช่วงเวลาหนึ่งเท่านั้น จากการทดลองสามารถเลือกส่วนของใบหน้าที่ดีที่สุดได้ ดังนั้นภาพที่มีความัวและมีสัญญาณรบกวนมากจึงไม่แสดงผลว่าเป็นส่วนของใบหน้าที่ดีที่สุด



(a)



(b)

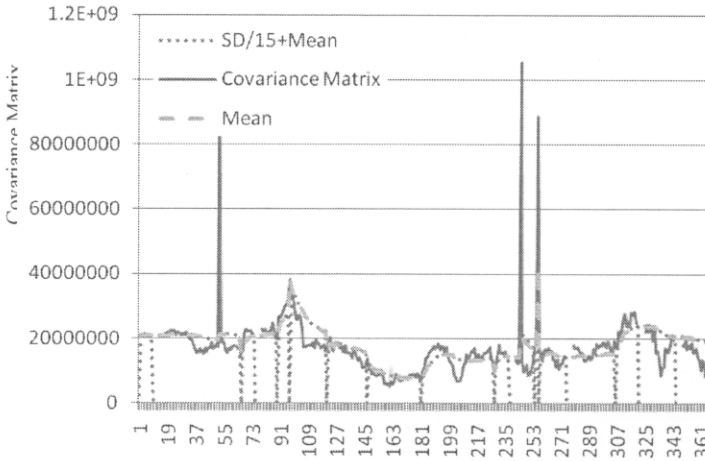


(c)

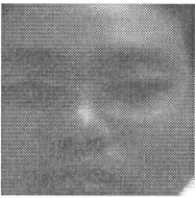


(d)

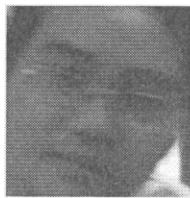
รูปที่ 3.6: ลำดับวิดีโอที่ 1 (a) กราฟของ  $\kappa_i(I_f)$ ,  $\eta_{\mu,\sigma^2}(\kappa_i)$  และ  $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.



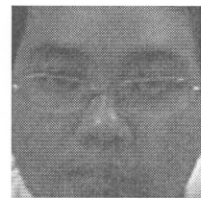
(a)



(b)



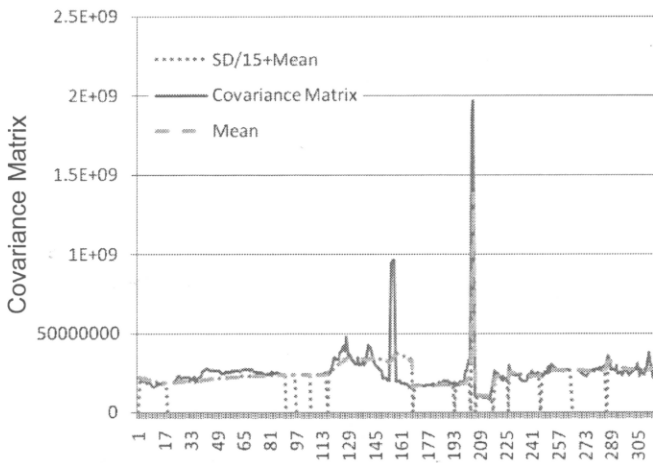
(c)



(d)

รูปที่ 3.7: ลำดับวิดีโอที่ 2 (a) กราฟของ  $\kappa_i(I_f)$ ,  $\eta_{\mu,\sigma^2}(\kappa_i)$  และ  $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.

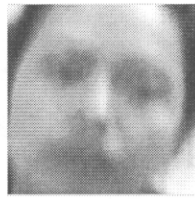




(a)



(b)

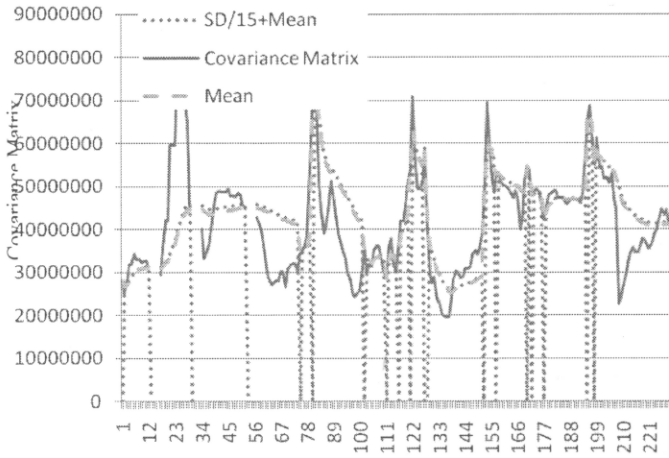


(c)



(d)

รูปที่ 3.8: ลำดับวิดีโอที่ 3 (a) กราฟของ  $\kappa_i(I_f)$ ,  $\eta_{\mu,\sigma}(\kappa_i)$  และ  $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.



(a)



(b)



(c)



(d)

รูปที่ 3.9: ลำดับวิดีโอที่ 4 (a) กราฟของ  $\kappa_i(I_f)$ ,  $\eta_{\mu,\sigma}(\kappa_i)$  และ  $h(\kappa_i)$ ; (b,c) ตัวอย่างเฟรมในลำดับวิดีโอ; และ (d) ตัวอย่างใบหน้าที่ถูกเลือก.

ดังนั้น อัลกอริธึมดังกล่าวจะทำการรวมกันกับ Video Analysis Framework ที่ได้เสนอ ใน [6][7] เพื่อที่จะวิเคราะห์ลำดับของวิดีโอจากกล้อง IP-camera ซึ่งถูกประมวลผลบนมัลติคอร์ และมัลติโพรเซสเซอร์ โปรแกรมของเฟรมเวิร์คเป็นวิธีการที่อยู่ภายใต้ .Net framework เวอร์ชัน 2.0 ประกอบไปด้วย 4 เลเยอร์ คือ Application layer, Application core layer, System core layer และ External module layer.



รูปที่ 3.10: แสดงโปรแกรมอัลกอริทึมที่ได้นำไปใช้ในโปรแกรม VsMonitor

### 3.5 สรุปผล

กระบวนการวัดดัชนีคุณภาพใบหน้าได้ใช้ค่าโควาเรียนเมตริก (Covariance Matrix) โดยประยุกต์กับการแยกเฟรมที่สำคัญ (Key frame) สำหรับดัชนีคุณภาพใบหน้าที่ถูกกำหนด นั้นขึ้นอยู่กับลักษณะของใบหน้าและสภาพแวดล้อมของวิดีโอที่ทำการบันทึก ซึ่งผลลัพธ์ที่ได้ สามารถทำงานได้ตามเวลาจริง (real-time) กับแอปพลิเคชันที่มีอยู่จริง

### 3.6 บรรณานุกรม

- [1] Ntalianis, K.S. and Kollias, S.D., An Optimized Key-Frames Extraction Scheme Based on SVD and Correlation Minimization, ICME 2005, 6-6 July 2005  
Page(s):792 - 795
- [2] Method of selecting key-frames from a video sequence, patent United States  
Patent 7184100
- [3] Viola, P. and Jones, M, Rapid object detection using a boosted cascade of simple features, CVPR2001, I-511- I-518 vol.1
- [4] Gary R Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal, No.Q2. 98
- [5] N. Suvonvorn and Anant Chocksuriwong, Real-Time Face Detection/Identification for Surveillance System, IEEE International Conference on Electronic Design, Malaysia, 1-3 December 2008
- [6] N. Suvonvorn, 2008, A Video Analysis Framework for Surveillance System, IEEE International Workshop on Multimedia Signal Processing, Cairns, Queensland, Australia, 8-10 October 2008.
- [7] N. Suvonvorn, 2008, "A Generic Surveillance System", The Sixth PSU-Engineering Conference, Songkhla, Thailand, 8-9 May 2008

## บทที่ 4

### การตรวจจับป้ายทะเบียน License Plate Detection

#### 4.1 ปัญหาและความสำคัญ

รถยนต์นับเป็นยานพาหนะที่ได้รับความนิยมอย่างแพร่หลายในการเดินทาง การขนส่ง สังกัดได้จากการเพิ่มปริมาณของจำนวนรถยนต์ที่สูงขึ้นในแต่ละปี จำนวนรถเพิ่มมากขึ้นทำให้ยากที่จะควบคุมตรวจสอบ การจัดการดูแลความปลอดภัย เช่น ระบบควบคุมการจราจรภายในอาคาร ระบบป้องกันการโจรกรรมรถยนต์ และรวมทั้งระบบรักษาความปลอดภัยในพื้นที่ชุมชน เป็นต้น ถึงแม้ว่าในปัจจุบันจะมีการตรวจสอบรถที่เข้าออกในสถานที่ต่างๆ โดยการใช้กล้องวงจรปิดเพื่อจับภาพเคลื่อนไหว ซึ่งยังทำได้เพียงเก็บภาพจากกล้องวงจรปิด หรือการใช้คนในการตรวจสอบและทำการเก็บข้อมูล ซึ่งส่งผลให้เกิดความล่าช้าและความผิดพลาดบ่อยครั้ง เมื่อเกิดเหตุร้ายขึ้นหรือต้องการติดตามรถที่มีพฤติกรรมน่าสงสัยจึงทำได้ยากและขาดประสิทธิภาพ ทั้งนี้ขาดเครื่องมือหรือโปรแกรมในการจำแนกป้ายทะเบียนรถยนต์ที่จะสามารถใช้งานได้สอดคล้องกับความต้องการนั่นเอง

งานวิจัยในส่วนนี้จะเน้นไปที่การศึกษาเทคนิคการตรวจสอบป้ายทะเบียนรถยนต์เบื้องต้นที่ใช้กับกล้องวงจรปิดเพื่อให้การทำงานความถูกต้อง แม่นยำ และรวดเร็ว ซึ่งเทคนิคจะประกอบด้วยการนำเอาภาพวิดีโอจากกล้องวงจรปิดมาทำการวิเคราะห์ เพื่อระบุหาตำแหน่งแผ่นป้ายทะเบียนรถยนต์และติดตามเพื่อการปรับปรุงคุณภาพของภาพแผ่นป้ายทะเบียนรถยนต์นั้น แล้วจึงจะทำการตัดภาพบริเวณแผ่นป้ายเพื่อนำมาวิเคราะห์ตัวอักษรในภาพป้ายทะเบียนนั้นต่อไป โดยจะไม่มีกรจดจำตัวอักษร อย่างไรก็ตามงานวิจัยนี้มีข้อจำกัดต่าง ๆ เช่น ภาพที่ได้ต้องเป็นภาพที่สามารถอ่านได้ด้วยสายตาดูปกติ ซึ่งอย่างน้อยจะต้องสามารถมองเห็นตัวเลขและตัวอักษรในภาพได้อย่างชัดเจนและครบถ้วน ข้อมูลภาพต้องมีคุณภาพชัดเจน เช่น ป้ายทะเบียนจะต้องไม่เบลอจนโคลน หรือมีรูปภาพ ลวดลายเป็นพื้นหลัง เป็นต้น

#### 4.2 ปรัชญาวรรณกรรม

อัลกอริทึมการจดจำป้ายทะเบียน (License plate recognition) จากภาพนิ่งหรือวิดีโอปกติแล้วจะประกอบด้วยสามขั้นตอนสำคัญคือ ก) การหาบริเวณที่เป็นแผ่นป้ายทะเบียน (License plate localization) ข) การแยกตัวอักษร (Character segmentation) และ ค) การจดจำตัวอักษร (Optical character recognition) [1][2] การพัฒนาวิธีการในแต่ละขั้นตอนเหล่านี้ให้สามารถทำงานได้อย่างมีประสิทธิภาพถือว่าเป็นงานที่ทำหายมากเนื่องจาก ด้วยข้อกำหนดต่างๆ ที่แตกต่างกันจะทำให้ผลลัพธ์ที่ได้ไม่ถูกต้องเท่าที่ควร เช่น ลักษณะของทะเบียนที่มีหลายรูปแบบ การเปลี่ยนแปลงของแสงที่มีอยู่ตลอดเวลาหรือสภาวะแวดล้อมในการนำเข้าข้อมูลภาพไม่เหมาะสม เป็นต้น ดังนั้นจึงมีงานวิจัยส่วนหนึ่งที่ได้นำเสนอวิธีการภายใต้เงื่อนไขต่าง ๆ เช่น แสดงคงที่ ค่าความเร็วของรถ ลักษณะของถนน และลักษณะภาพพื้นหลัง เป็นต้น บางเทคนิคที่ได้พัฒนามุ่งไปเพื่อใช้งานกับภาพนิ่งหรือภาพวิดีโอเป็นการเฉพาะ วิธีการต่างๆ เหล่านี้ จะมีความ

แตกต่างกันซึ่งสามารถนำมาเปรียบเทียบได้ด้วยคุณลักษณะต่าง ๆ เช่น เวลาในการประมวลผล พลังงานที่ใช้ หรือ ระดับความสามารถในการจดจำได้

การหาตำแหน่งของป้ายทะเบียนจะสามารถทำได้จากคุณลักษณะพิเศษของป้ายเอง เช่น โครงสร้างทางเลขาคณิต อัตราส่วนความกว้างความสูง สี ลักษณะของโทนสีขาวดำ ความถี่ การเปลี่ยนแปลงของความเข้มแสงหรือคุณลักษณะของตัวอักษรที่อยู่ภายในป้าย ค่าความเปลี่ยนแปลงจากการเปลี่ยนแปลงของ gradient อัตราส่วนของตัวอักษร และอัตราการกระจายตัวของช่วงว่างระหว่างตัวอักษร เป็นต้น ด้วยคุณลักษณะดังกล่าวจึงทำให้สามารถหาตำแหน่งของป้ายทะเบียนได้ ซึ่งขั้นตอนต่อมาก็คือการแยกตัวอักษรและจดจำ ตามลำดับ ในการแยกตัวอักษรสามารถทำได้หลายวิธีเช่นกัน ได้แก่ การทำ Projection วิธีทาง Morphology การทำ relaxation labeling และการหา component ที่ติดกันเป็นต้น ในแต่ละวิธีก็จะมีข้อดีข้อเสียแตกต่างกันออกไป สำหรับส่วนขั้นตอนของการจดจำนั้น มีวิธีการต่างๆ หลากหลายมาก เช่น genetic algorithms neural networks fuzzy c-means support vector machine และ Markov processes วิธีการเหล่านี้สามารถแยกออกเป็นวิธีที่ต้องทำซ้ำหรือไม่ทำซ้ำก็ได้ หากเป็นวิธีการที่จะต้องทำซ้ำปกติจะได้รับผลลัพธ์ที่ดีกว่าแต่ก็มีความซับซ้อนมากขึ้นตามลำดับ

ในงานวิจัยนี้จะมุ่งเน้นไปที่ส่วนที่หนึ่งและสองตามลำดับ อันได้แก่ การหาตำแหน่งป้ายและการแยกตัวอักษร ส่วนการจดจำตัวอักษรนั้นจะไม่ได้ศึกษาในงานวิจัยนี้

#### 4.3 ระเบียบวิธีการตรวจจับป้ายทะเบียนและแยกตัวอักษร

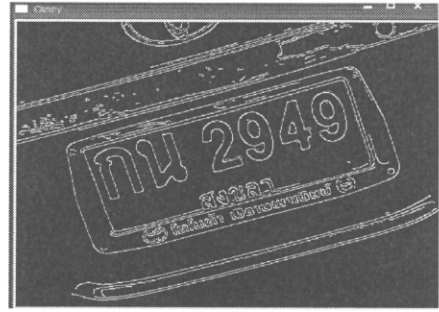
ในส่วนนี้จะนำเสนอรายละเอียดของเทคนิคการตรวจหาแผ่นป้ายทะเบียนและการติดตามแผ่นป้ายทะเบียนรถยนต์ การดำเนินการจะแบ่งการทำงานออกเป็น 5 ส่วนด้วยกันคือ การระบุหาตำแหน่งแผ่นป้ายทะเบียนรถยนต์จากภาพ การตรวจสอบความเอียงของป้ายทะเบียนรถยนต์ การตัดแยกตัวอักษรของป้ายทะเบียนรถยนต์

##### 4.3.1 การระบุหาตำแหน่งแผ่นป้ายทะเบียนรถยนต์

การระบุหาตำแหน่งแผ่นป้ายทะเบียนรถยนต์จะมีขั้นตอนการทำงานดังต่อไปนี้

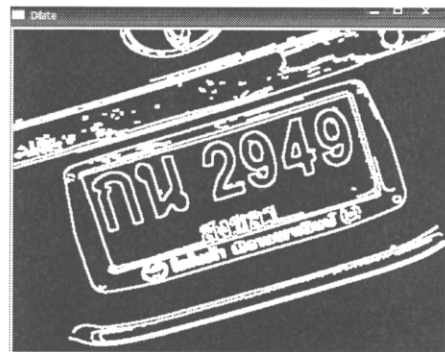
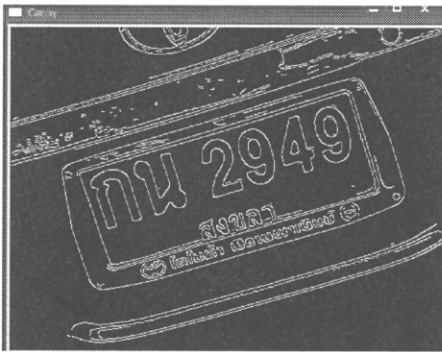
**ขั้นตอนที่ 1:** แปลงภาพที่ต้องการวิเคราะห์ให้เป็นแบบขาวดำ (gray scale) โดยปกติแล้ว ข้อมูลภาพที่ได้รับจากกล้องวิดีโอจะเป็นภาพสี

**ขั้นตอนที่ 2:** การหาขอบของภาพ (edge detection) โดยได้เลือกใช้วิธีการของ Canny ซึ่งจากการศึกษาพบว่า เป็นวิธีที่เหมาะสมในการนำมาใช้งาน เพราะสามารถหาขอบภาพได้ชัดเจน ทำให้ได้ขอบของภาพที่มีรายละเอียดพอที่จะใช้ในขั้นตอนต่อไป ผลลัพธ์จากการใช้วิธีนี้จะขึ้นอยู่กับค่าปรับระดับค่า Threshold ซึ่งมีอยู่สองค่า (strong/weak edge)



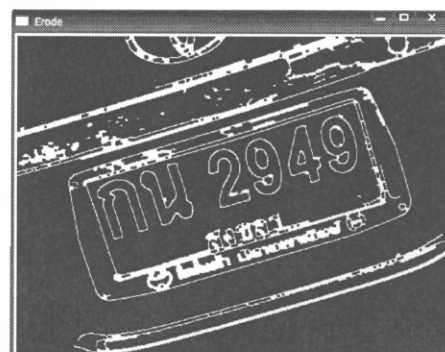
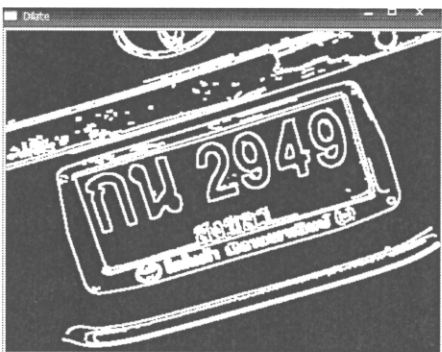
รูปที่ 4.1 การหาขอบของภาพด้วยวิธีการ Canny

เนื่องจากขอบของภาพที่ได้จากวิธีการของ Canny ยังไม่สมบูรณ์มากพอ ได้แก่ ลักษณะของเส้นยังไม่มีความต่อเนื่องมากพอ จึงจำเป็นต้องมีกระบวนการปรับแต่งได้การใช้เทคนิคทางด้าน Morphological การทำ Close โอเปอเรชัน จะทำให้มีการปิดรูเล็กๆ ที่บริเวณเส้นขอบและเชื่อมต่อส่วนของเส้นขอบให้มีความต่อเนื่องกันมากยิ่งขึ้น



รูปที่ 4.2 การปรับแต่งขอบด้วยการใช้เทคนิคทางด้าน Morphological (Close)

หลังจากนั้นก็จะผ่านกระบวนการ Open ซึ่งเป็นกระบวนการในการกัดกร่อน เพื่อกำจัดส่วนที่เป็นส่วนเกินของบริเวณขอบทิ้ง ผลลัพธ์ที่ได้จะทำให้พิกเซลของภาพจะถูกปิดเชื่อมต่อกันมากขึ้น



รูปที่ 4.3 การปรับแต่งขอบด้วยการใช้เทคนิคทางด้าน Morphological (Open)

**ขั้นตอนที่ 3:** เมื่อปรับปรุงส่วนของขอบได้ตามที่ต้องการแล้ว ก็จะทำหารูปร่างลักษณะของขอบที่มีความต่อเนื่องกันแบบ close loop โดยจะใช้วิธี freeman chain code ซึ่งเป็นวิธีการที่จะค้นหา

พิกเซลที่อยู่บนเส้นขอบด้วยการติดตามพิกเซลเหล่านั้นตามทิศทางที่กำหนดทิศทางโดยใช้ตัวเลขโดยขอบของวัตถุที่เราสนใจคือขอบที่มีลักษณะใกล้เคียงสี่เหลี่ยม ซึ่งเป็นลักษณะเด่นของป้ายทะเบียนรถยนต์ที่มีรูปแบบเป็นสี่เหลี่ยมผืนผ้า

**ขั้นตอนที่ 4:** เมื่อได้ขอบที่มีความต่อเนื่องกันแล้ว ก็จะนำไปวิเคราะห์ว่ามีลักษณะเป็นสี่เหลี่ยมหรือไม่ โดยใช้อัลกอริทึม Douglas-Peucker algorithm ซึ่งจะเป็นการประมาณขอบของภาพซึ่งเป็นการเชื่อมต่อของพิกเซลด้วยส่วนของเส้นตรง โดยยังคงคุณสมบัติทางเลขาคณิตของขอบของภาพเอาไว้ให้ได้ใกล้เคียงของเดิมมากที่สุด ในขั้นตอนนี้หากขอบดังกล่าวเป็นขอบของป้าย เราก็จะได้ส่วนของเส้นตรงจำนวนสี่เส้นที่แทนขอบของป้ายนั้น โดยมีจุดเชื่อมต่อ 4 จุดแทนมุมของสี่เหลี่ยม



รูปที่ 4.4 ตำแหน่งของป้ายที่หาได้

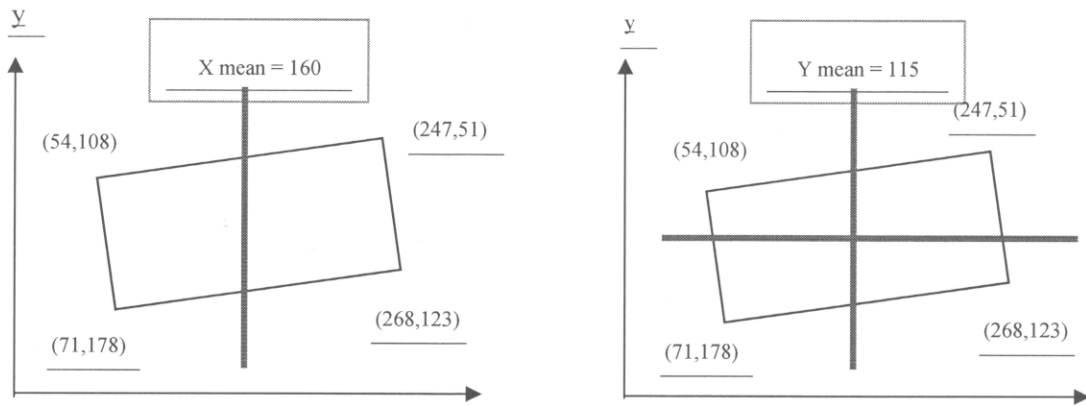
#### 4.3.2 การวิเคราะห์ความเอียงของป้ายทะเบียนรถยนต์

เนื่องจากความเอียงของป้ายจะมีผลทำให้การตัดตัวอักษรในขั้นตอนต่อไปนั้นมีความผิดพลาดอย่างมาก ดังนั้นจึงจำเป็นอย่างยิ่งที่จะต้องแปลงป้ายที่มีลักษณะเอียงหรือบิดให้อยู่ในแนวตรง โดยจะแบ่งการทำงานออกเป็นสองส่วนสำคัญคือ ก) การตรวจสอบจุดที่ได้จากการวิเคราะห์ในขั้นตอนนี้ก่อนหน้า เป็นมุมได้ของป้ายทะเบียน ข) การทำการบิดภาพทางเลขาคณิต (Geometric transformation)

ส่วนที่หนึ่ง จะเป็นการหาว่าจุดที่ได้ทั้ง 4 จุดนั้น อยู่ ณ ตำแหน่งใดของป้าย อันได้แก่ ตำแหน่งขอบซ้ายบน ขอบซ้ายล่าง ขอบขวาบน และขอบขวาล่าง โดยสามารถหาได้ด้วยวิธีการดังต่อไปนี้

**ขั้นตอนที่ 1:** หาค่าเฉลี่ยของตำแหน่ง  $x$  ของทั้ง 4 จุดเพื่อทำการแบ่งจุดที่อยู่ทางซ้ายและทางขวาของป้ายทะเบียนออกจากกัน ซึ่งจะได้ 2 จุดที่อยู่ทางซ้าย และ 2 จุดที่อยู่ทางขวา

**ขั้นตอนที่ 2:** หาค่าเฉลี่ยของตำแหน่ง  $y$  ของทั้ง 4 จุดเพื่อทำการแบ่งจุดที่อยู่ด้านบนและด้านล่างของป้ายทะเบียนออกจากกัน ซึ่งจากขั้นตอนนี้จะได้ จุดที่อยู่ด้านซ้ายบน จุดที่อยู่ด้านซ้ายล่าง จุดที่อยู่ด้านขวาบน และจุดที่อยู่ด้านขวาล่าง ซึ่งเป็นตำแหน่งขอบของป้ายทะเบียนตามลำดับ



รูปที่ 4.5 การหาค่าหมุมต่างๆ ของป้าย

ส่วนที่สอง เมื่อได้มุมทั้งสี่ด้านของป้ายแล้ว ก็นำจุดเหล่านั้นมาใช้ในการทำการบิดภาพให้อยู่ในลักษณะตรงตามที่ต้องการซึ่งรู้ตำแหน่งแน่นอนอยู่แล้ว โดยในการทดลองนี้ได้ใช้ Affine transform สำหรับการบิดภาพทางเลขาคณิต การบิดภาพจะต้องทำการบิดทั้งทางแกน X และแกน Y เป็นการย้ายจุด (x,y) ของภาพเดิมไปยังจุด (x',y') ของภาพใหม่นั้นเอง โดยสมการการบิดภาพจะมีลักษณะดังนี้คือ

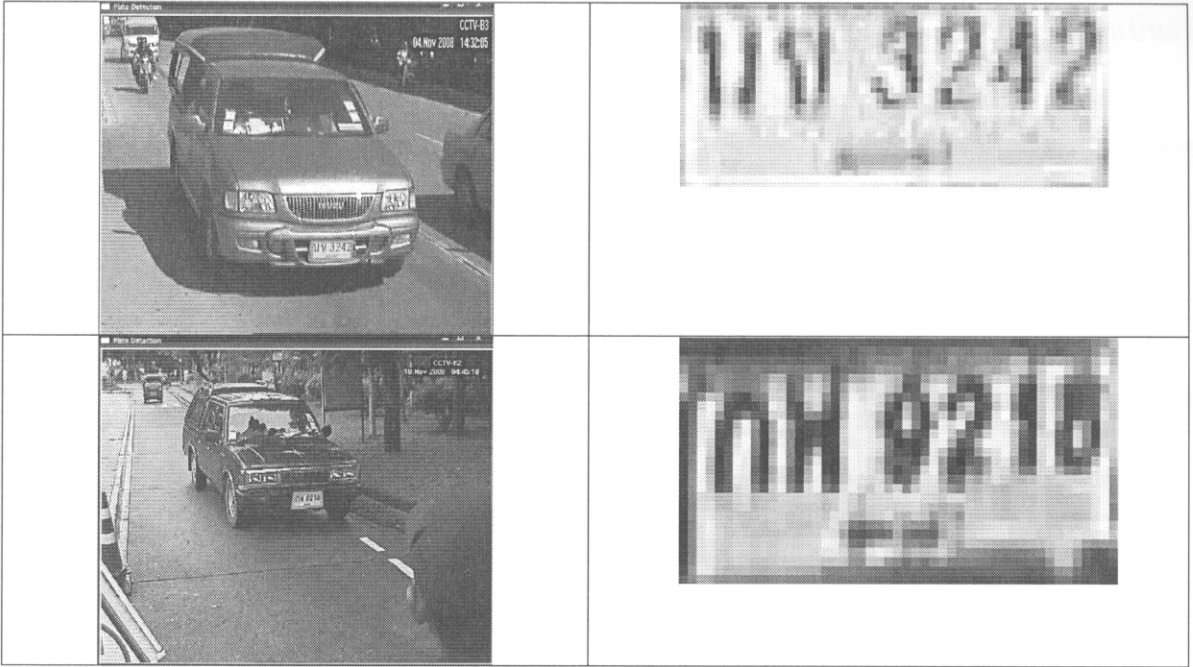
$$y' = a_1y + b_1x + c_1$$

$$x' = a_2y + b_2x + c_2$$

ตารางที่ 4.1 แสดงผลการหาค่าหมุมป้ายทะเบียน





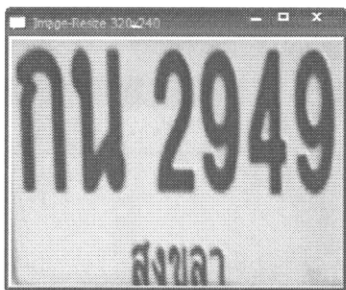


### 4.3.3 การตัดแยกตัวอักษรของป้ายทะเบียนรถยนต์

การตัดแยกตัวอักษรจะมีสองลักษณะที่ใช้เทคนิคเดียวกัน คือ การตัดแบ่งตามแนวนอน และ การตัดแบ่งตามแนวตั้งตามลำดับ

#### การตัดแบ่งในแนวนอน

**ขั้นตอนที่ 1:** นำภาพที่ได้มาทำการแปลงภาพให้อยู่ในรูปของ Binary ซึ่งสามารถทำได้โดยใช้ค่า Threshold ซึ่งเป็นการพิจารณาว่าพิกเซลใดควรเป็นสีขาวโดยมีค่าเท่ากับ 1 หรือ จุดใดควรเป็นสีดำโดยมีค่าเท่ากับ 0 ในที่นี้ส่วนที่เป็นสีขาวคือพื้นหลังของป้ายทะเบียนและส่วนที่เป็นสีดำคือตัวอักษรหรือตัวเลข



รูปที่ 4.6 การแปลงภาพเป็นไบนารี

**ขั้นตอนที่ 2:** นำภาพป้ายทะเบียนแบบ Binary ที่ได้มาทำการโปรเจกชัน สามารถทำได้โดยการนับจำนวนของพิกเซลของภาพในแต่ละแถวที่มีค่าเป็นสีดำ แล้วนำไปแสดงไว้ในอีกระนาบหนึ่ง ซึ่งจะนำผลลัพธ์ที่ได้ไปใช้ในการตัดแบ่งตัวอักษรออกจากกันโดยให้สังเกตจากจำนวนของสีดำในระนาบที่ได้ทำการโปรเจกชันว่ามีอย่างน้อยแค่ไหน หากมาค่ามากแสดงว่าบริเวณนั้นจะเป็นเป็นตัวอักษร หากมาค่าน้อยแสดงว่าเป็นบริเวณระหว่างอักษร ซึ่ง ในทางปกติแล้วค่าสีดำที่นับได้ระหว่างตัวอักษรจะไม่เท่ากับศูนย์ เนื่องจาก

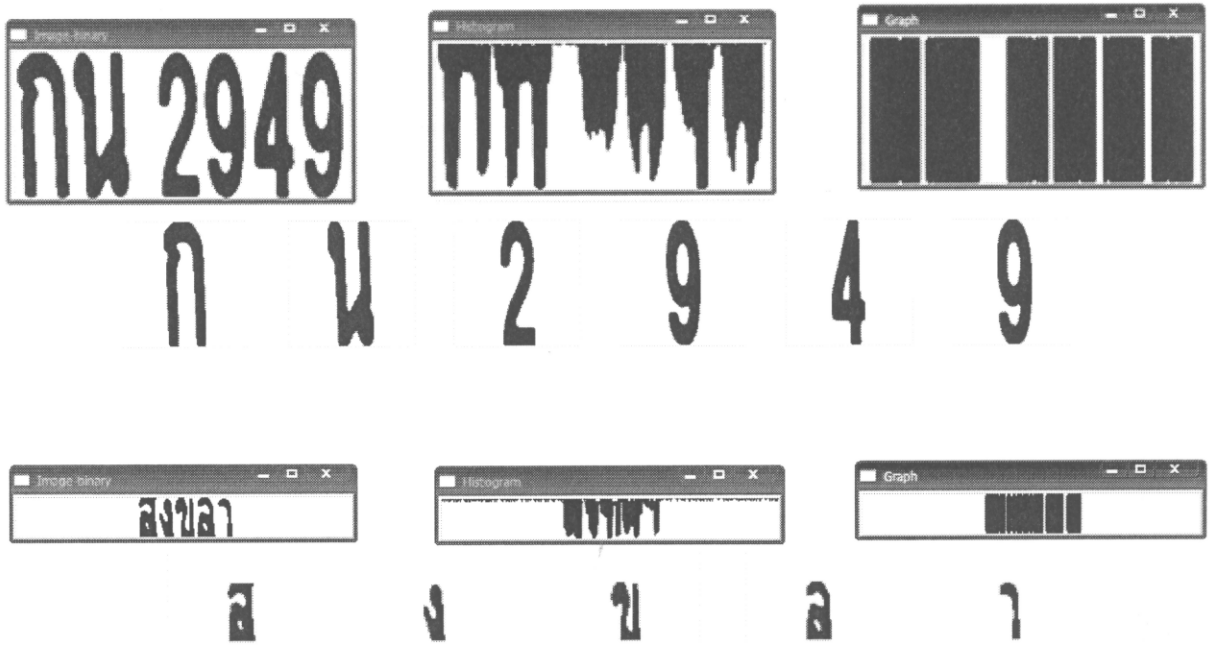
ภาพที่เราพิจารณานั้นจะมีสัญญาณรบกวนอยู่จำนวนหนึ่ง ดังนั้น ในการตัดแบ่งตัวอักษรนั้นจึงจำเป็นต้องกำหนดค่า Threshold ให้อย่างเหมาะสม



รูปที่ 4.7 การ Projection ตามแนวแกนนอน

### การตัดแบ่งในแนวตั้ง

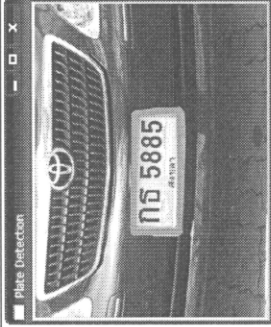

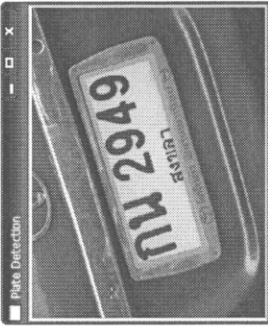
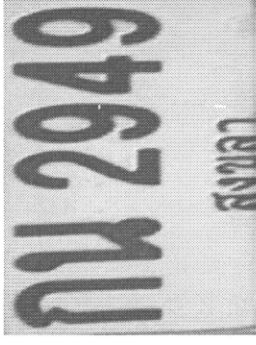
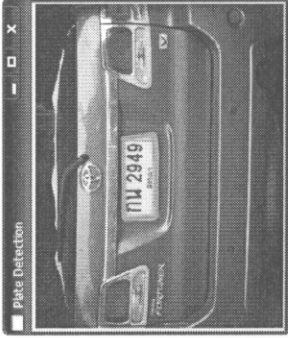

สำหรับขั้นตอนนี้ ให้ทำดำเนินการเช่นเดียวกับการตัดแบ่งตัวอักษรในแนวนอน

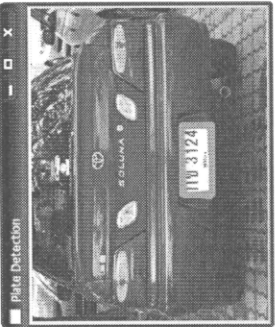
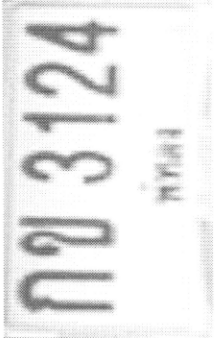
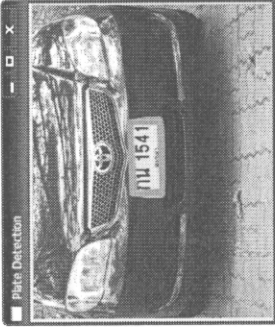
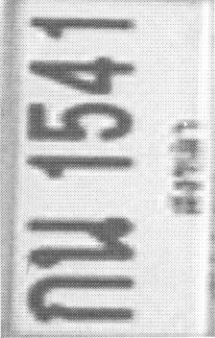

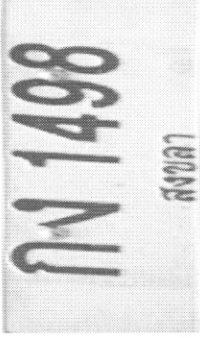


รูปที่ 4.8 การ Projection ตามแนวแกนตั้ง

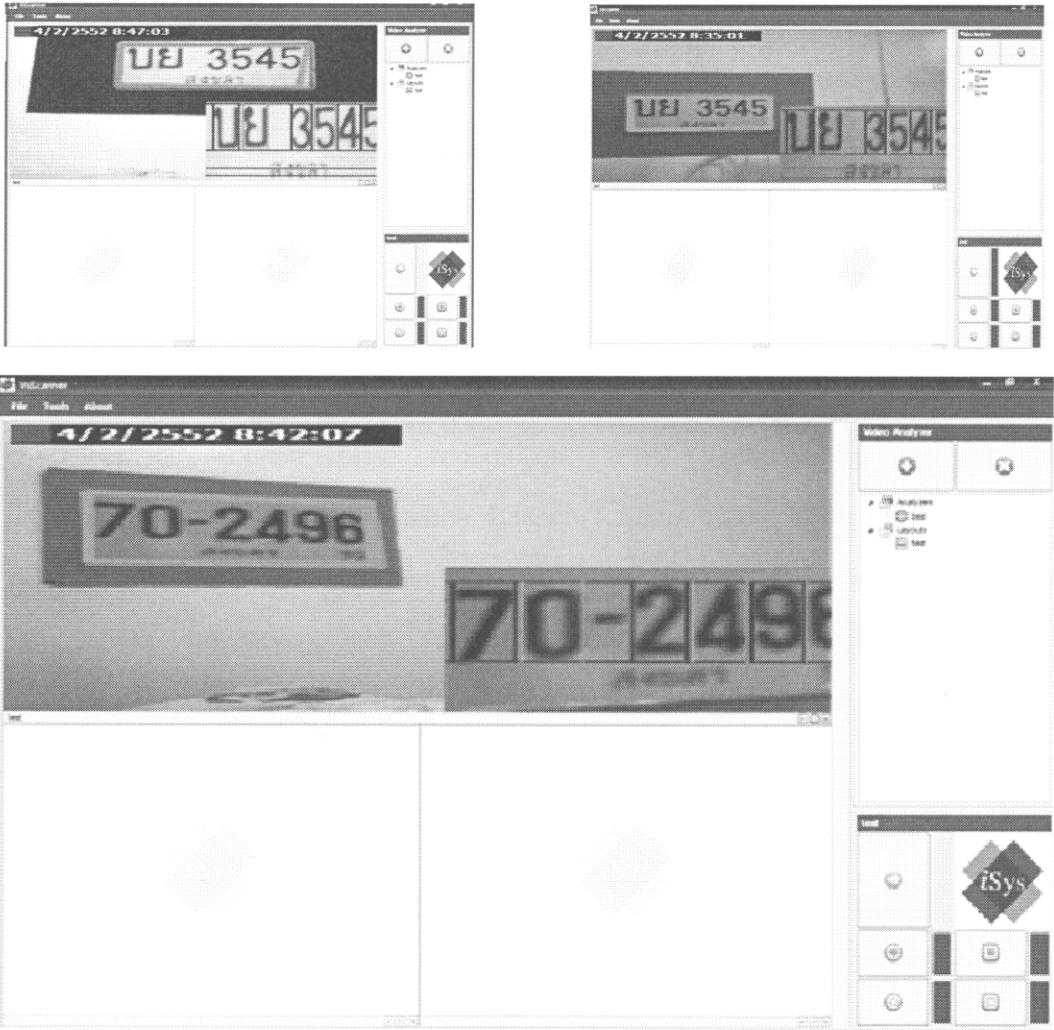
## 4.4 ผลการทดลองและวิเคราะห์

ตารางต่อไปนี้จะแสดงการตัดแบ่งตัวอักษรในลักษณะต่างๆ ซึ่งพบว่าวิธีการที่น่าเสนอนี้สามารถนำมาใช้งานได้ อย่างไรก็ตามผลลัพธ์ตามเทคนิคนี้ยังไม่เสถียรมากนักหากมีการเปลี่ยนแปลงของแสงในสิ่งแวดล้อมนั้นๆ

การประมาณส่วนของป้ายทะเบียนรถยนต์	การตัดภาพและ การปรับความเอียง	การตัดแบ่งในแนวฮอน	การตัดแบ่งในแนวตั้ง
1 		การตัดแบ่งในแนวฮอน <b>กธ 5885</b> สงขลา	การตัดแบ่งในแนวตั้ง <b>ก</b> <b>ธ</b> <b>5</b> <b>8</b> <b>8</b> <b>5</b> ส ข ง ข ๗ ๗ ๗ ๗ ๗ ๗
2 		การตัดแบ่งในแนวฮอน <b>กน 2949</b> สงขลา	การตัดแบ่งในแนวตั้ง <b>ก</b> <b>น</b> <b>2</b> <b>9</b> <b>4</b> <b>9</b> ส ข ง ข ๗ ๗ ๗ ๗ ๗ ๗
3 		การตัดแบ่งในแนวฮอน <b>กน 2949</b> สงขลา	การตัดแบ่งในแนวตั้ง <b>ก</b> <b>น</b> <b>2</b> <b>9</b> <b>4</b> <b>9</b> ส ข ง ข ๗ ๗ ๗ ๗ ๗ ๗

	การประมาณค่าของป้ายทะเบียนรถยนต์	การตัดภาพและการปรับความเอียง	การตัดแบ่งในแนวนอน	การตัดแบ่งในแนวตั้ง
4			<p>กข 3124</p> <p>ก ข ๓ ๑ ๒ ๔</p>	<p>ก ข ๓ ๑ ๒ ๔</p> <p>ก ข ๓ ๑ ๒ ๔</p>
5			<p>กน 1541</p> <p>ก น ๑ ๕ ๔ ๑</p>	<p>ก น ๑ ๕ ๔ ๑</p> <p>ก น ๑ ๕ ๔ ๑</p>
6			<p>กน 1498</p> <p>ก น ๑ ๔ ๙ ๘</p>	<p>ก น ๑ ๔ ๙ ๘</p> <p>ก น ๑ ๔ ๙ ๘</p>

โปรแกรมได้ที่พัฒนานี้ได้นำเอาไปรวมกับโปรแกรม VsMonitor ซึ่งจะช่วยให้สามารถนำภาพมาจากระบบกล้องวิดีโอวงจรปิดมาใช้ในการประเมินผลได้



รูปที่ 4.9 โปรแกรมที่ทำงานร่วมกับ VsMonitor ได้

#### 4.5 สรุปผล

จากการดำเนินการการตรวจหาแผ่นป้ายทะเบียนรถยนต์ด้วยฐานข้อมูลภาพจากกล้องวงจรปิดพบว่าสามารถทำการตรวจหาส่วนที่เป็นแผ่นป้ายทะเบียนรถยนต์ได้ แต่ผลที่ได้ยังไม่ดีมากเพียงพอ ซึ่งพบว่ายังคงมีประมาณอีก 25 เปอร์เซ็นต์ที่ไม่สามารถตรวจหาส่วนของแผ่นป้ายทะเบียนรถยนต์ได้ ซึ่งปัญหาเกิดขึ้นจากสามสาเหตุหลักด้วยกันคือ ก) ปัญหาของขอบที่ไม่สมบูรณ์ ไม่มีความต่อเนื่องมากเพียงพอ ข) ปัญหาความแตกต่างของความเข้มสีระหว่างแผ่นป้ายทะเบียนรถยนต์กับตัวรถยนต์เองที่แตกต่างกันไม่มากเพียงพอ ทำให้ไม่สามารถระบุตำแหน่งป้ายได้ไม่ถูกต้อง ค) ปัญหาคุณภาพของภาพที่ได้จากระบบกล้องวิดีโอวงจรปิดที่ได้ผ่านการบีบอัดมาแล้วจึงทำให้มีผลต่อการวิเคราะห์อย่างชัดเจน ในส่วนของการตัดแบ่งตัวอักษรนั้น พบว่าสามารถตัดแบ่งตัวอักษรได้ค่อนข้างดี ยกเว้นคุณภาพของตัวอักษรในระดับจังหวัดยังมีขนาดเล็กเกินไปจนเกินไป จนทำให้การตัดตัวอักษรไม่ถูกต้องเพียงพอ

## 4.6 บรรณานุกรม

- [1] Anagnostopoulos, C.-N.E., Anagnostopoulos, I.E., Psoroulas, I.D., Loumos, V., Kayafas, E., License Plate Recognition From Still Images and Video Sequences: A Survey, IEEE Transactions on Intelligent Transportation Systems, Volume 9, Issue 3, Date: Sept. 2008, Pages: 377-391
- [2] Arth, C. Limberger, F. and Bischof, H., Real-Time License Plate Recognition on an Embedded DSP-Platform, IEEE Conference on Computer Vision and Pattern Recognition, 17-22 June 2007, On page(s): 1-8, ISBN: 1-4244-1180-7
- [3] Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung and Sei-Wan Chen, Automatic license plate recognition, IEEE Transactions on Intelligent Transportation Systems, Volume 5, Issue 1, Date: March 2004, Pages: 42- 53
- [4] Beatriz Díaz Acosta, Experiments in Image Segmentation for Automatic US License Plate Recognition, Thesis Dissertation (Master of Science), Faculty of the Virginia, Polytechnic Institute and State University, June 18, 2004.

## บทที่ 5

### การตรวจสอบภาพใต้ท้องรถ Under Vehicle Scanner

#### 5.1 ปัญหาและความสำคัญ

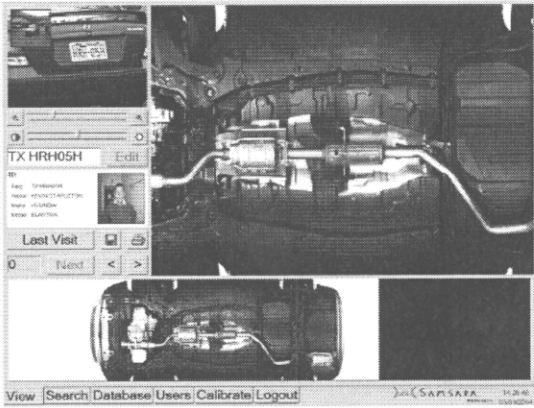
การตรวจสอบใต้ท้องรถเพื่อการป้องกันการลักลอบการซุกซ่อนวัตถุอันตรายหรือสิ่งผิดกฎหมายในบริเวณทางเข้าออกของตัวอาคารหรือสถานที่สำคัญถือว่ามีความสำคัญยิ่งในบางพื้นที่ ซึ่งในปัจจุบันการตรวจสอบส่วนใหญ่ได้มีการนำเอากระบอกเงาสะท้อนมาใช้ โดยการจัดอุปกรณ์ไว้ด้านข้างทางผ่าน ซึ่งทำให้สามารถมองผ่านด้านข้างของตัวรถไปยังใต้ท้องรถได้ แม้ว่าจะเป็นแนวทางหนึ่งที่สามารถทำได้เร็วและมีราคาไม่แพง แต่เมื่อมองถึงประสิทธิภาพการใช้งานได้แล้วยังถือว่าไม่ดีเท่าที่ควร ทั้งนี้เนื่องจากในทางปฏิบัติแล้ว ลักษณะของใต้ท้องรถจะมีส่วนโค้งงออยู่พอสมควร ซึ่งการมองจากด้านข้างในบางกรณีหรือรถบางยี่ห้อไม่สามารถตรวจสอบได้อย่างแน่ชัด

จากปัญหาข้างต้นงานวิจัยนี้จึงเน้นไปที่การนำเอากล้องวงจรปิดมาใช้ในการประกอบการบันทึกภาพใต้ท้องรถตามแนวขวางทั้งคัน โดยผู้มีหน้าที่สามารถตรวจสอบภาพใต้ท้องรถจากจอร์จายานผลได้โดยตรง ระบบดังกล่าวจะประกอบไปด้วยสองส่วนคือ ก) โปรแกรมการถ่ายภาพ และ ข) อุปกรณ์การรับภาพจากท้องรถ ซึ่งการออกแบบอุปกรณ์ดังกล่าวจะใช้ชิ้นอุปกรณ์ราคาถูก และไปเน้นการพัฒนาซอฟต์แวร์สำหรับการประมวลผลภาพเป็นหลัก ในรายงานนี้จะนำเสนอวิธีการทั้งหมดที่ได้ทดสอบแม้ว่าในบางวิธีจะได้ผลลัพธ์จากการทดสอบตามสภาพแวดล้อมจริงไม่ดีนักก็ตาม

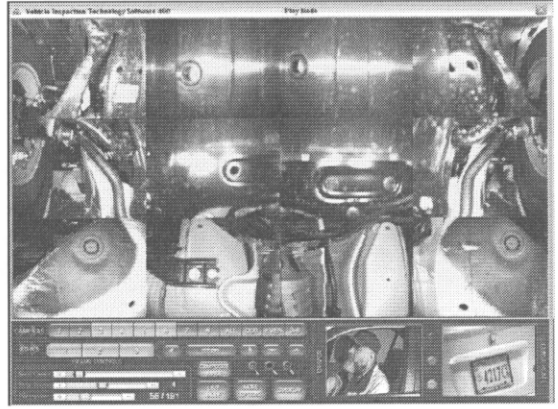
#### 5.2 ทัศนวิสัยวรรณกรรม

ระบบตรวจจับภาพใต้ท้องรถนั้นได้มีผู้พัฒนาไปมากแล้วและเป็นเทคโนโลยีที่มีหลายบริษัทได้พัฒนาเป็นผลิตภัณฑ์ อย่างไรก็ตามเนื่องจากเป็นระบบนี้ยังมีผู้ใช้น้อยและเป็นเทคโนโลยีขั้นสูงที่จะใช้เฉพาะบางพื้นที่เท่านั้น จึงทำให้ระบบดังกล่าวมีราคาค่อนข้างแพง ระบบตรวจใต้ท้องรถ PPA Vehicle Surveillance System [1] ซึ่งเป็นบริษัทของคนไทย ได้เสนอระบบตรวจตราใต้ท้องรถ โดยสามารถแสดงภาพใต้ท้องรถโดยรวมเป็นลักษณะภาพนิ่ง ผู้ขับขี่รถ และทะเบียนรถยนต์ ได้พร้อมๆ กัน ดังรูปที่ 5.1 ระบบ Under Vehicle Surveillance System (UVSS) Integration [2] เป็นระบบที่สามารถทำหน้าที่ทั้งสามอย่างเหมือนกัน รวมถึงความสามารถในการเชื่อมต่อกับระบบอื่นๆ ได้ด้วย เช่น ระบบเปิดปิดประตู เป็นต้น เมื่อรถยนต์แล่นเข้ามาในบริเวณตรวจสอบระบบ จะมีเซนเซอร์ที่จะกระตุ้นให้ระบบการอ่านป้ายทะเบียนอัตโนมัติ Automatic License Plate Reader (ALPR) System ทำงาน ซึ่งจะถอดรหัสตัวเลข

และตัวอักษร ไปแสดงยังหน่วยรักษาความปลอดภัยและเก็บบันทึกเพื่อการค้นหาต่อไป ในขณะที่เดียวกัน ก็จะนำผลการสแกนได้ทั้งรถทั้งภาพมาแสดงผลไว้เช่นกัน ดังแสดงไว้ในรูปที่ 5.2 ระบบ Und-Aware™ Under-Vehicle-Surveillance-System [3] มีความสามารถในลักษณะเดียวกันกับทั้งสองระบบที่กล่าวมาแล้ว แต่แตกต่างกันที่รายละเอียดปลีกย่อยเท่านั้น ดังสามารถ รูปที่ 5.3



รูปที่ 5.1 ระบบ PPA Vehicle Surveillance System



รูปที่ 5.2 ระบบ Under Vehicle Surveillance System (UVSS) Integration



รูปที่ 5.3 ระบบ Und-Aware™ Under-Vehicle-Surveillance-System

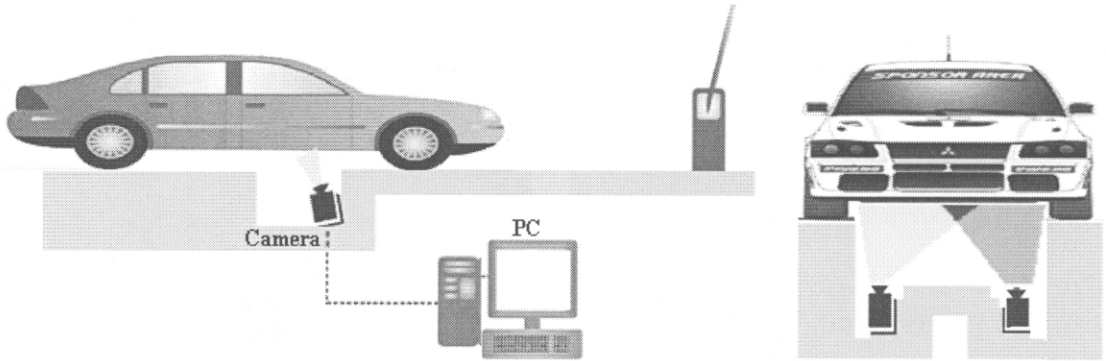
ในบทนี้เราจะสนใจเฉพาะการตรวจสอบภาพใต้ท้องรถเท่านั้น ส่วนการวิเคราะห์ป้ายทะเบียนและการตรวจจับหน้า ได้นำเสนอในบทที่สามและสี่ตามลำดับแล้ว

### 5.3 การออกแบบระบบและระเบียบวิธีการตรวจสอบภาพใต้ท้องรถ

ระบบการตรวจสอบข้อมูลภาพใต้ท้องรถนี้ได้ออกแบบโดยคำนึงถึงการใช้งานเครื่องมือที่มีราคาถูกโดยมุ่งเน้นไปที่การพัฒนาซอฟต์แวร์ที่มีประสิทธิภาพทดแทนเครื่องมือราคาสูง องค์ประกอบหลักของระบบนี้ คือ กล้องที่ใช้ในการรับข้อมูลภาพจากใต้ท้องรถ หากเราเลือกใช้กล้องและเลนส์มุมกว้างที่มีคุณภาพดีแล้ว ก็จะสามารถใช้กล้องเพียงหนึ่งตัวเพื่อตรวจสอบภาพใต้ท้องรถเต็มคันตามความกว้างได้ อย่างไรก็ตาม กล้องและเลนส์ที่มีคุณสมบัติดังกล่าว นั้นมีราคาสูง การประยุกต์ใช้กล้องธรรมดาและไม่มีเลนส์มุมกว้างมาใช้ซึ่งมีราคาถูกจึงเป็นแนวทางที่น่าสนใจ อย่างไรก็ตาม เนื่องจากว่าระยะระหว่างใต้ท้องรถและกล้องนั้นมีความสูงไม่มากนัก จึงทำให้ภาพที่ได้จากกล้องเป็นมุมแคบสามารถมองเห็นได้เพียงพื้นที่บางส่วนของใต้



ห้องรถเท่านั้น หากต้องการภาพทั้งหมด ก็สามารถทำได้ด้วยการใช้อาร์เรย์ของกล้องซึ่งภาพที่ได้จากกล้องหลายตัวจะไม่เชื่อมต่อกัน ซึ่งในงานวิจัยนี้เรามุ่งเน้นไปที่การใช้อัลกอริทึมทางด้านการประมวลผลภาพเพื่อการเชื่อมต่อข้อมูลภาพเหล่านั้น รูปข้างล่างนี้แสดงโครงสร้างโดยรวมของระบบการตรวจสอบภาพใต้ห้องรถ

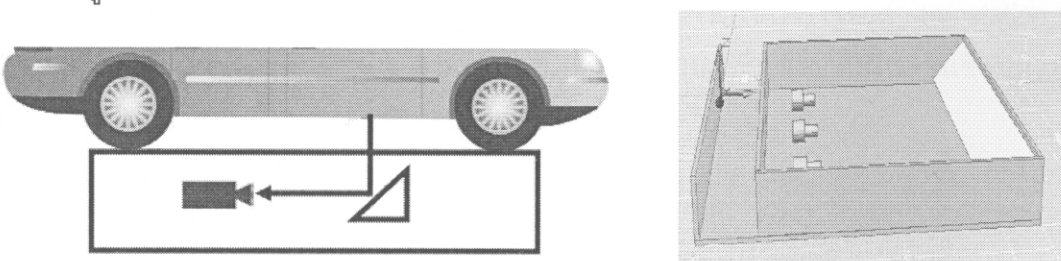


รูปที่ 5.4: ระบบตรวจข้อมูลภาพจากใต้ห้องรถ

การทำงานของระบบจะมีลักษณะดังนี้คือ เมื่อรถเคลื่อนที่เข้าใกล้จุดตรวจสอบ ระบบการรับภาพจากอาร์เรย์ของกล้อง (มีมากกว่าสองกล้อง) ที่อยู่ใต้ห้องรถจะทำการจับภาพได้แล้วส่งภาพเหล่านั้นไปให้แก่คอมพิวเตอร์หลักเพื่อการประมวลผลการเชื่อมต่อภาพตามความเร็วจริงและนำเสนอรายงานนอกทางจอภาพในทันที ซึ่งระบบทั้งสองส่วนนี้จะนำเสนอแยกออกเป็นสองหัวข้อหลักคือ การออกแบบและพัฒนาเครื่องรับภาพใต้ห้องรถและการพัฒนาอัลกอริทึมในการเชื่อมภาพ โดยใช้เทคนิคลูคัส-คานาเนส (Lucas & Kanade Technique) และเทคนิค CAMSHIFT ตามลำดับ

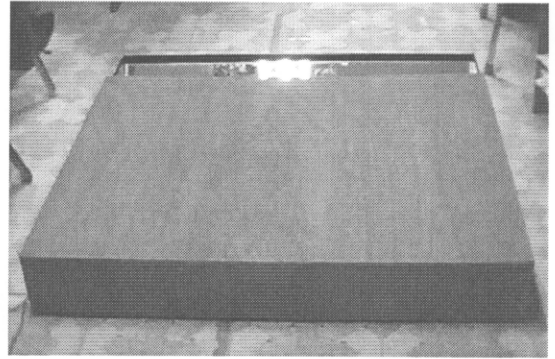
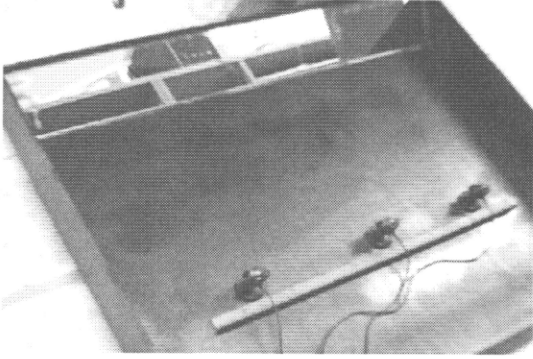
### 5.3.1 การออกแบบและพัฒนาเครื่องมือรับภาพใต้ห้องรถ

การออกแบบเครื่องมือหรือกล้องรับภาพจากใต้ห้องรถมีข้อกำหนดสำคัญอยู่ที่ระยะห่างระหว่างกล้องและรถซึ่งมีค่าอยู่ที่ประมาณ 20-45 เซนติเมตร การที่จะให้กล้องรับภาพจากใต้ห้องรถโดยตรงเพื่อให้ได้ภาพตามแนวขวางทั้งหมด (มากกว่า 1.2 เมตร) ซึ่งจำเป็นต้องใช้กล้องมากกว่า 4 ตัว ทั้งนี้ในการเรียงกันเป็นอาร์เรย์ภาพที่ได้จากกล้องที่อยู่ติดกันจะต้องมีส่วนเหลื่อมกันเพื่อที่จะให้สามารถนำภาพเหล่านั้นมาเชื่อมต่อกันได้ด้วยวิธีทางการประมวลผลภาพ ด้วยข้อกำหนดข้างต้นจึงทำให้มีจำนวนกล้องไม่คงที่ ซึ่งจะขึ้นอยู่กับความกว้างและความสูงของตัวรถ ในการออกแบบระบบรับภาพจึงได้มีแนวคิดที่จะนำภาพที่ได้จากการสะท้อนมาใช้แทนซึ่งแสดงดังรูปต่อไปนี้



รูปที่ 5.5 เครื่องมือรับภาพใต้ห้องรถ

กล่องอุปกรณ์มีการติดตั้งกล่องแนวระนาบเพื่อรับภาพใต้ท้องรถที่ได้จากการสะท้อนผ่านกระจกที่ทำมุม 45 องศากับตัวกล่องและใต้ท้องรถ อุปกรณ์มีขนาดประมาณกว้าง 1 เมตร ยาว 0.85 เมตร ซึ่งภายในติดตั้งกระจกเงาสท้อนยาว 1 เมตร และสูง 0.15 เมตร กล่องรับภาพแบบอาร์เรย์ถูกติดตั้งไว้ที่ระยะ 0.6 เมตรจากกระจกสะท้อน ดังรูป



รูปที่ 5.6 อุปกรณ์ต้นแบบกล่องรับภาพสะท้อน

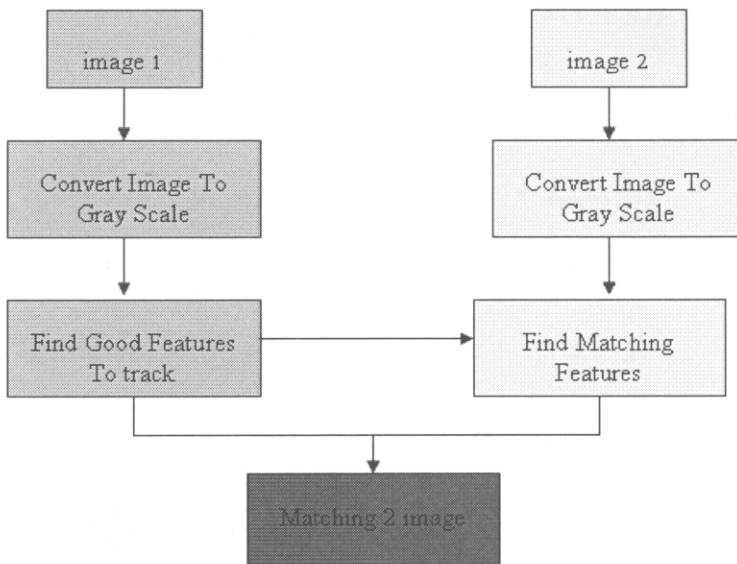
รูปที่ 5.7 แสดงภาพที่ได้จากการกลิ้งที่ผ่านกระจกสะท้อนของกล่องสองตัว โดยการทดสอบกับฉากภาพนำไปไว้เหนือกระจกสะท้อน ณ ระยะต่างๆ



รูปที่ 5.7 ภาพสะท้อนจากกระจกเงาสท้อนมุม 45 องศา

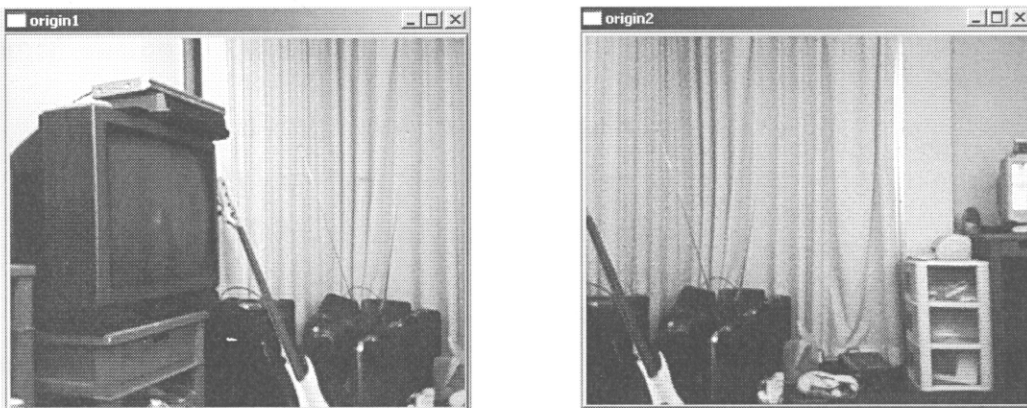
### 5.3.2 การเชื่อมต่อภาพโดยใช้เทคนิคลูคัส-คานาส (Lucas & Kanade Technique)

การเชื่อมต่อภาพจากกล่องสองตัวโดยเทคนิคของ ลูคัส-คานาส นั้นจะมีขั้นตอนดังต่อไปนี้ คือ ก) การค้นหาจุดหรือลักษณะสำคัญของภาพขาวดำจากกล่องทั้งสองโดยใช้วิธี Harris ข) การนำจุดสำคัญที่ได้จากทั้งสองภาพมาทำการหาความสัมพันธ์ (matching) โดยใช้วิธีของ ลูคัส-คานาส ค) ผลลัพธ์ที่ได้จากขั้นตอนก่อนหน้าจะนำไปใช้ในการหาค่า translation ทางเลขาคณิตทั้งในแกน x และ แกน y ตามลำดับ ซึ่งค่าที่คำนวณได้จะนำมาใช้ในการเชื่อมต่อภาพทั้งสองต่อไป ระบบโดยรวมสามารถสรุปได้ดังรูป



รูปที่ 5.8 แสดงไดอะแกรมการต่อภาพ

จากรูปที่ 5.9 แสดงตัวอย่างภาพที่ได้จากกล้องซ้ายมือและขวามือตามลำดับจะเห็นได้ว่าภาพทั้งสองมีความต่อเนื่องกันและมีพื้นที่ที่เหมือนกันระหว่างภาพสองภาพ ซึ่งบริเวณดังกล่าวจะเป็นบริเวณที่จะนำไปใช้ในการวิเคราะห์การหาจุดและและความสัมพันธ์ระหว่างกันต่อไป



รูปที่ 5.9 ตัวอย่างภาพนิ่งทั้ง 2 ภาพ

### ขั้นตอนที่ 1 การวิเคราะห์หาจุดที่สนใจ

วิธีการหาจุดสำคัญของ Harris และ Stephens [4] เป็นการพัฒนาความสามารถมาจากวิธีของ Moravec [5] โดยการพิจารณาค่าความเป็นจุดสำคัญซึ่งมีค่าสัมพันธ์กับทุกทิศทางรอบๆ จุดสำคัญที่สนใจ โดยวิธีการของ Moravec จะให้ความสำคัญเฉพาะในทิศหลักๆ ทั้งแปดเท่านั้น การคำนวณค่าความเป็นจุดสำคัญจะใช้ค่า autocorrelation เป็นตัวกำหนดแต่หากต้องการความรวดเร็วก็สามารถใช้ค่า SSD แทนได้

หากกำหนดให้ภาพขาวดำ  $I$  มีพื้นที่ที่สนใจ  $I(u,v)$  และพื้นที่ขนาดเท่ากันไปในทุกทิศทาง  $(x,y)$  เราสามารถวัดความเหมือนระหว่างสองพื้นที่นั้น  $S$  ได้ด้วยสมการดังต่อไปนี้

$$S(x, y) = \sum_u \sum_v w(u, v) (I(u, v) - I(u+x, v+y))^2$$

หากประมาณการด้วยการขยาย  $I(u+x, v+y)$  โดยวิธีของ Taylor แล้ว จะได้

$$I(u+x, v+y) \approx I(u, v) + I_x(u, v)x + I_y(u, v)y$$

ซึ่งจะได้ว่า

$$S(x, y) \approx \sum_u \sum_v w(u, v) (I_x(u, v)x + I_y(u, v)y)^2,$$

โดยที่  $I_x$  และ  $I_y$  คือ partial derivative ของ  $I$  จะได้ว่า

$$S(x, y) \approx (x \ y)A \begin{pmatrix} x \\ y \end{pmatrix},$$

ซึ่ง  $A$  เป็นเมตริกของ Harris

$$A = \sum_u \sum_v w(u, v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix}$$

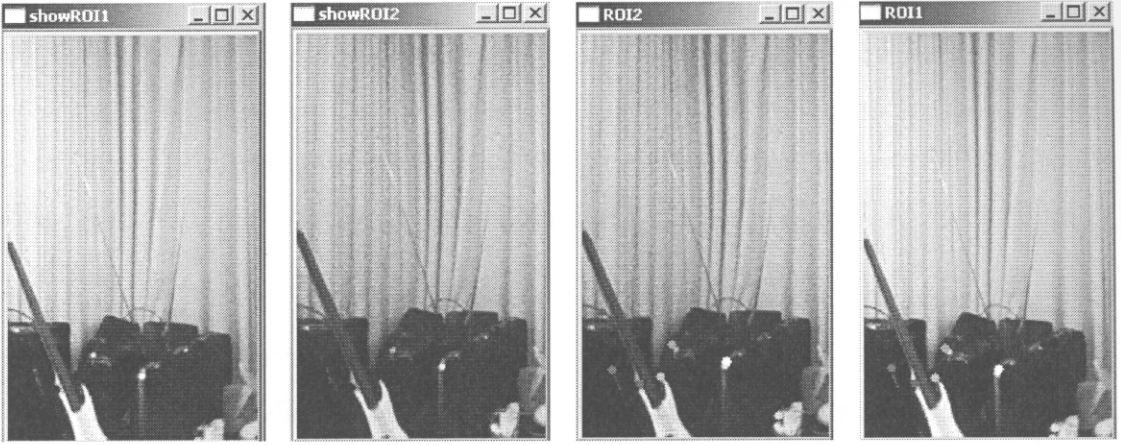
ค่าความเป็นจุดสำคัญสามารถอธิบายได้ด้วยคุณลักษณะการเปลี่ยนแปลงของความเข้มแสงบริเวณโดยรอบตำแหน่งที่สนใจ ซึ่งคุณลักษณะดังกล่าวมีความสัมพันธ์กับค่า eigen values ทั้งสองค่าของเมตริก  $A$  ดังนี้คือ

- ก) ถ้าค่าทั้งสองมีค่ามาก แสดงว่าตำแหน่งนั้นเป็นจุดสำคัญ
- ข) ถ้าค่าใดค่าหนึ่งมีค่ามาก และ อีกค่ามีค่าน้อย แสดงว่าตำแหน่งนั้นคือขอบของภาพ
- ค) ถ้าค่าทั้งสองมีค่าน้อย แสดงว่าไม่ใช่ตำแหน่งที่สนใจ

ซึ่งจากความสัมพันธ์ข้างต้น Harris และ Stephens ได้กำหนดฟังก์ชันการตัดสินใจความเป็นจุดสำคัญหรือไม่ ดังนี้คือ

$$M_c = \lambda_1 \lambda_2 - \kappa (\lambda_1 + \lambda_2)^2 = \det(A) - \kappa \text{trace}^2(A)$$

โดย  $\kappa$  เป็นค่าคงที่ ซึ่งมีค่าเหมาะสมคือ 0.04 รูปที่ 5.10 แสดงบริเวณที่ต้องการวิเคราะห์ และรูปที่ 5.11 แสดงจุดสำคัญที่ได้ที่ได้จากการนำวิธีการของ Harris มาใช้



รูปที่ 5.10 ส่วนที่เหมือนกันระหว่างทั้ง 2 ภาพ

รูปที่ 5.11 หาจุดที่เหมือนกันระหว่างภาพ 2 ภาพ

## ขั้นตอนที่ 2 การหาจุดที่เหมือนกันจากภาพสองภาพ

เทคนิคลูคัส-คาเนส (Lucas & Kanade Technique) [6] [7] เป็นวิธีการประมาณการเคลื่อนไหวกว้างภาพ หรือ ที่เรียกว่า optical flow (ในกรณีนี้จะเป็นการเคลื่อนไหวของจุดสำคัญระหว่างภาพสองภาพที่ได้รับจากวิธีการของ Harris) ในการประมาณค่านี้ กำหนดให้ค่า flow  $(V_x, V_y)$  มีค่าคงที่ภายในหน้าต่างเล็กขนาด  $m \times m$  โดย  $m > 1$  ซึ่งมีพิกเซล  $x, y$  เป็นจุดศูนย์กลาง โดยกำหนดให้พิกเซลในพื้นที่นั้นเป็นพิกเซลที่  $1 \dots n$ ,  $n = m^2$  ตามลำดับ โดยสามารถเขียนเป็นสมการได้ดังนี้

$$I_{x1}V_x + I_{y1}V_y = -I_{t1}$$

$$I_{x2}V_x + I_{y2}V_y = -I_{t2}$$

:

$$I_{xn}V_x + I_{yn}V_y = -I_{tn}$$

ซึ่งสามารถเขียนให้อยู่ในรูปของเมตริก คือ

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \vdots \\ -I_{tn} \end{bmatrix} \quad \text{หรือ} \quad A\vec{v} = -b$$

หากใช้การแก้สมการด้วยวิธีการของ Least square จะได้ว่า

$$\vec{v} = (A^T A)^{-1} A^T (-b) \quad \text{หรือ}$$

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum I_{x_i}^2 & \sum I_{x_i} I_{y_i} \\ \sum I_{x_i} I_{y_i} & \sum I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \end{bmatrix}$$

เพื่อที่ที่ต้องการให้วิธีการของ ลูคัส-คานเนส สามารถตอบสนองการคำนวณ flow ที่มีขนาดหรือระยะทางที่มากพอสมควรสำหรับการเคลื่อนที่ของจุดสำคัญระหว่างภาพที่เปลี่ยนแปลงอย่างรวดเร็ว การนำเทคนิค coarse-to-fine มาใช้ในการนี้จึงมีความจำเป็นอย่างยิ่ง โดยวิธีการดังกล่าวจะทำให้สามารถคำนวณหาค่า flow ซึ่งมีขนาดใหญ่ก่อนแล้วค่อยๆ ปรับไปยังสเกลที่เล็กลง ซึ่งปกติแล้วจะมีการปรับเป็นลักษณะพีรามิดขึ้นซึ่งจะทำให้ได้ flow ที่ได้มีความถูกต้องมากยิ่งขึ้น

### ขั้นตอนที่ 3 การเชื่อมต่อภาพ

การหา Flow หรือการเคลื่อนไหวของจุดสำคัญจากภาพซ้ายไปขวาจะทำให้สามารถหาค่าการ Translation โดยเฉลี่ยทั้งในแนวแกน x และแกน y ได้ ซึ่งค่าทั้งสองสามารถนำมาใช้ในการในการ translation ภาพ ด้วยเมตริกดังต่อไปนี้คือ

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

รูปที่ 5.12 แสดงผลลัพธ์ที่ได้จากการนำภาพสองภาพมาเชื่อมต่อกันโดยใช้เทคนิคของ



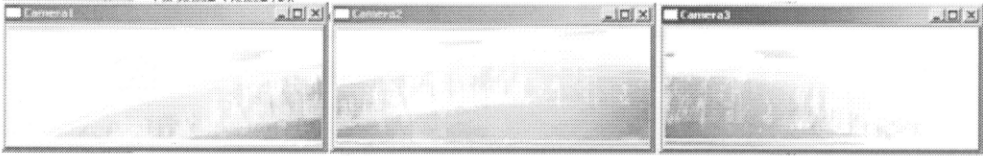
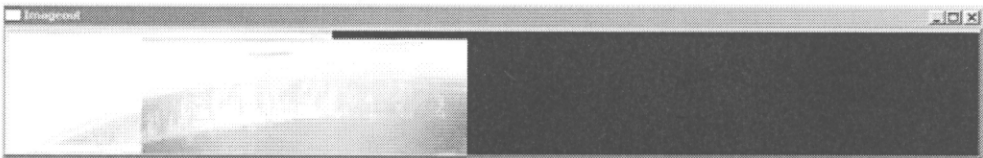
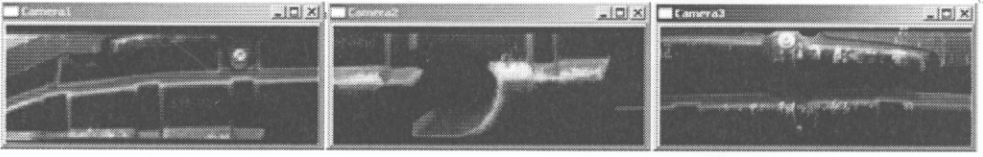
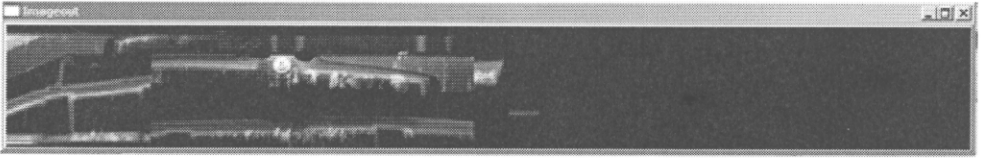
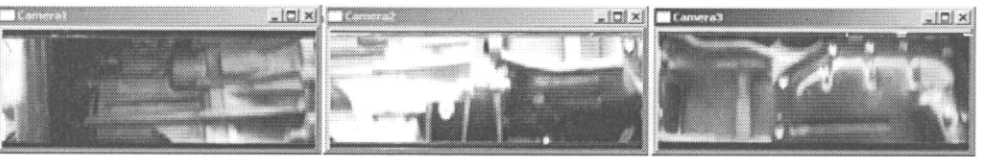
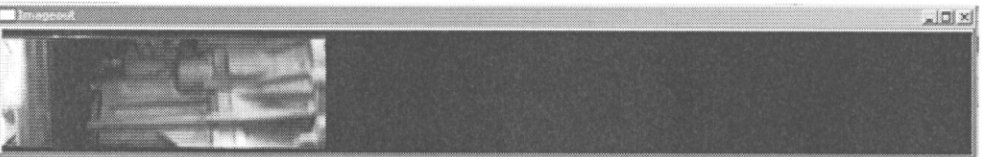
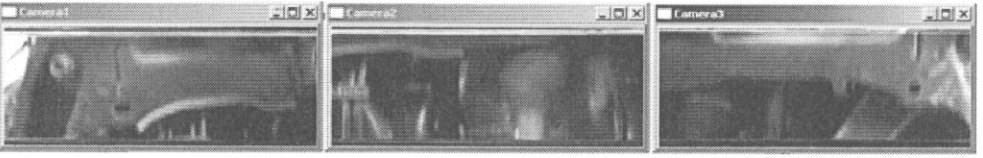
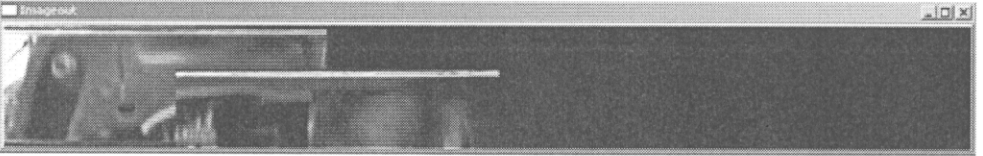
รูปที่ 5.12 ภาพที่เกิดจากการซ้อนกันของ 2 ภาพ

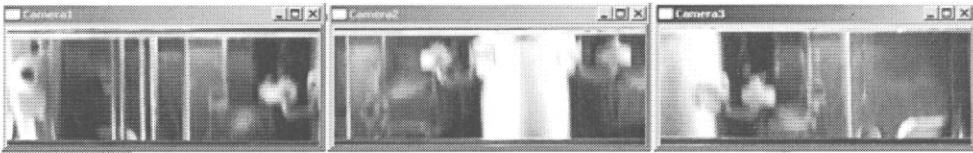
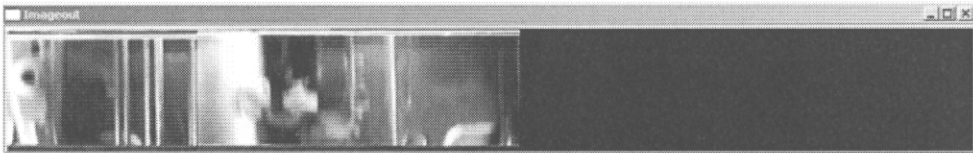
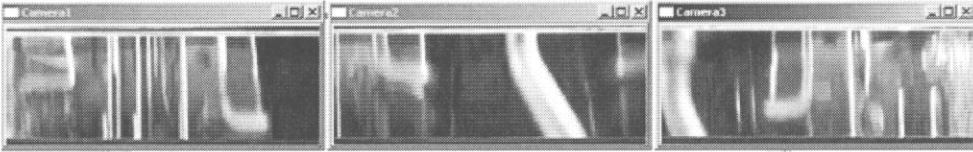
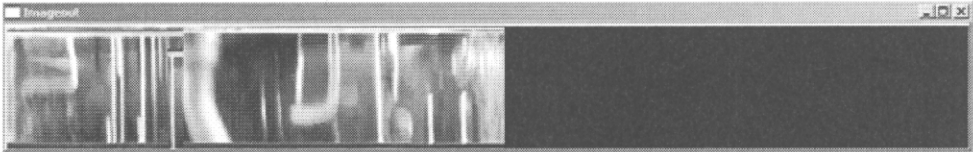
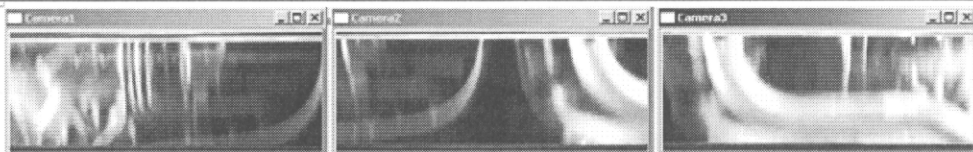

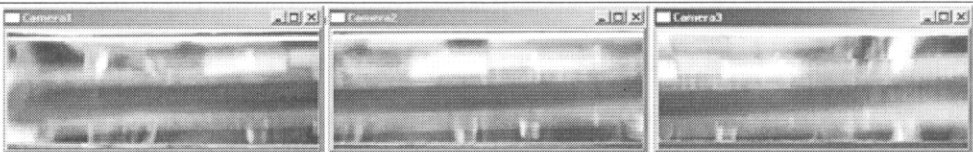
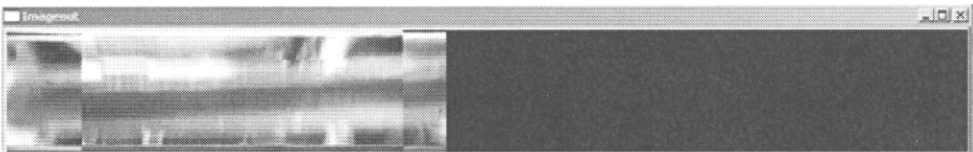


### การทดสอบกับรถยนต์จริง

ในการทดสอบระบบการตรวจสอบภาพได้ตามสภาวะแวดล้อมจริงนี้ ได้พัฒนากล่องรับข้อมูลภาพต้นแบบที่ใช้ฮาร์ดแวร์ของกล้อง Webcam จำนวนสามตัว และใช้เครื่องคอมพิวเตอร์แบบพกพาที่มีโปรเซสเซอร์ Pentium Duo Core เป็นตัวประมวลผลภาพ โดยการใช้รถยนต์ส่วนบุคคลในการทดสอบซึ่งมีความสูงของได้ท้องรถที่ประมาณ 25 เซนติเมตรจากระดับพื้นดิน การทดสอบประสิทธิภาพการทำงานได้ให้ทำการตรวจสอบภาพได้ท้องรถจำนวนสิบครั้ง ที่มีสภาวะแตกต่างกันคือ กลางวันมีแสงมาก กลางคืนมีแสงน้อย กลางวันมีแสงน้อยแต่ใช้แหล่งกำหนดแสงประกอบ และกลางคืนกับมีแหล่งกำหนดแสง

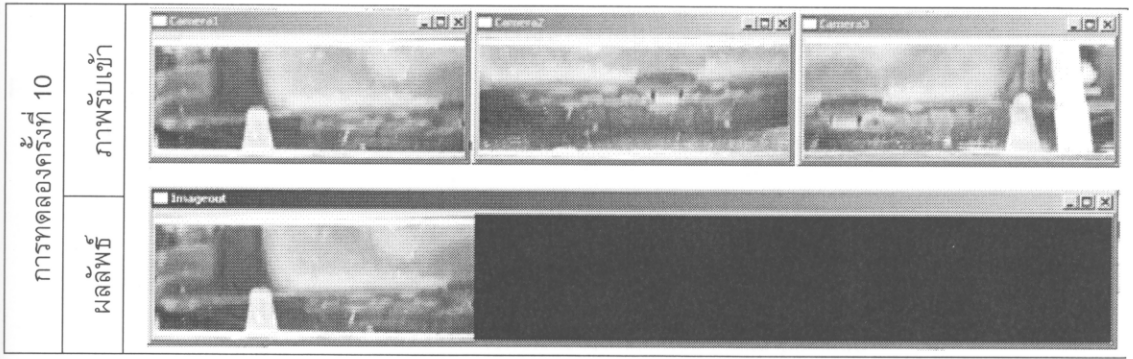
ตามตารางที่ 5.1 แสดงผลภาพที่ได้จากจากกล้องรับภาพและผลการต่อภาพ โดยได้แสดงภาพแค่หนึ่งเฟรมจากวิดีโอหนึ่งชุดของการทดลองแต่ละครั้ง พบว่าการต่อเชื่อมภาพที่ได้จากการทดลองพบว่ามีคุณภาพที่น้อยมาก เมื่อเทียบกับการทดลองภายในห้องซึ่ง สามารถสรุปสาเหตุการไม่สามารถหาผลลัพธ์ที่ดีได้จากการที่ภาพได้ห้องรถไม่มีคุณลักษณะที่เป็นจุดเด่นเพียงพอที่จะทำให้วิธีการหาจุดสำคัญโดยวิธีของ Harris มีความไม่ถูกต้อง ซึ่งตรงข้ามกับตอนทดสอบในห้องปฏิบัติการที่ใช้ภาพวิวที่มีลวดลายซับซ้อน

เพื่อการแก้ปัญหาดังกล่าวจึงได้มีการปรับวิธีการวิเคราะห์การต่อภาพและปรับปรุงกล้องรับภาพให้เหมาะสมเพื่อให้ประสิทธิภาพหรือผลลัพธ์มีความถูกต้องมากขึ้น

การทดลองครั้งที่ 1	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 2	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 3	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 4	ภาพรับเข้า	
	ผลลัพธ์	

การทดลองครั้งที่ 5	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 6	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 7	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 8	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 9	ภาพรับเข้า	
	ผลลัพธ์	





ตารางที่ 5.1 แสดงผลลัพธ์ของการต่อภาพใต้ห้องรถด้วยวิธีของลูคัส-คาเนส

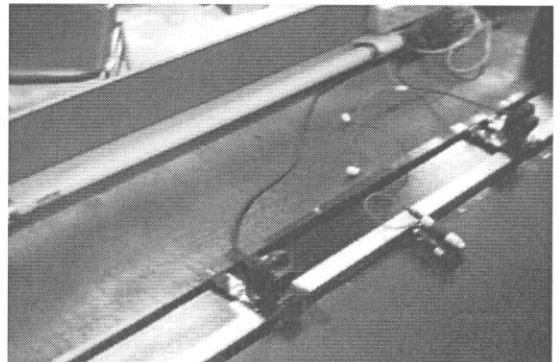
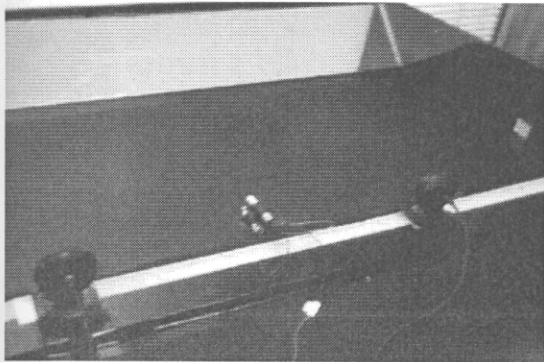
### 5.3.3 การเชื่อมต่อภาพโดยใช้เทคนิค Camshift

เนื่องด้วยความสามารถในการทำงานของเทคนิคการเชื่อมต่อภาพโดยใช้เทคนิคของ Harris และ เทคนิคลูคัส-คาเนส ไม่เป็นที่พอใจ ซึ่งมีสาเหตุสำคัญมาจากกรณีที่ภาพใต้ห้องรถไม่มีจุดอ้างอิงที่มาเพียงพอ จึงนำมาสู่การปรับปรุงเทคนิคดังนี้คือ ก) จำเป็นต้องสร้างจุดอ้างอิงขึ้นมาเอง และ ข) การตรวจจับจุดอ้างอิงดังกล่าวด้วยการวิเคราะห์ผลภาพ

#### การสร้างจุดอ้างอิงใต้ห้องรถ

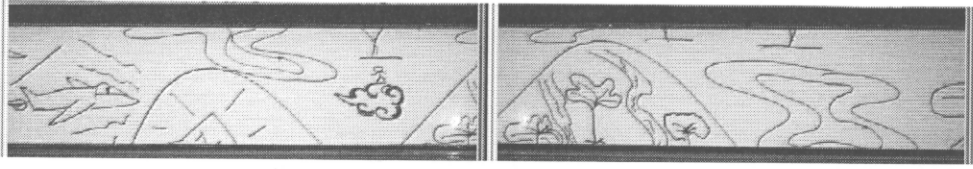
จากการศึกษาเพิ่มเติมการปรับแต่งกล่องรับภาพจากใต้ห้องรถให้มีความสามารถในการสร้างจุดอ้างอิงได้ด้วยนั้น แนวคิดที่ได้ทำการทดลองแล้วสามารถใช้งานได้ดีคือการสร้างจุดอ้างอิงด้วยแสงจาก Laser pointer ซึ่งจะทำให้เกิดจุดที่เกิดจากลำแสงบนใต้ห้องรถที่สามารถมองเห็นได้ด้วยตาเปล่า ซึ่งสามารถตรวจจับด้วยการวิเคราะห์ภาพใต้เช่นกัน ข้อดีของการนำเอา Laser pointer มาใช้อีกอย่างก็คืออุปกรณ์ดังกล่าวสามารถหาได้ง่ายและราคาถูก

จากรูปที่ 5.13 แสดงกล่องรับภาพที่ได้มีการปรับแต่งโดยการนำ laser pointer ไปติดตั้ง ณ ตำแหน่งกึ่งกลางระหว่างกล้องสองตัว ซึ่งจะให้มีจุดอ้างอิงที่ถูกสร้างขึ้นเกิดขึ้นในภาพจากทั้งสองกล้องในส่วนที่เหลื่อมกัน ซึ่งจะนำไปใช้ในขั้นตอนต่อไป จากการที่ได้กำหนดจุดอ้างอิงขึ้นเองจึงทำให้บริเวณที่เหลื่อมกันระหว่างสองกล้องนั้นไม่จำเป็นต้องมีพื้นที่มาก เหมือนที่ผ่านมา ซึ่งทำให้กล่องรับภาพลดจำนวนกล้องเหลือเพียงแค่สองตัวสำหรับความกว้างใต้ห้องรถที่เท่ากัน



รูปที่ 5.13 การติดตั้ง Laser pointer ในกล่องรับภาพใต้ห้องรถ

รูปที่ 5.14 แสดงผลลัพธ์ภาพจากทั้งสองกล้องที่มีจุดอ้างอิงที่สร้างขึ้นเองโดยใช้ laser pointer



รูปที่ 5.14 แสดงจุดอ้างอิงที่สร้างขึ้นเองโดยใช้ Laser pointers

### การตรวจจับและการจับคู่จุด Laser pointer

จากรูปที่ 5.14 จะเห็นว่าภาพจุดที่ได้จากการสร้างด้วย Laser pointer นั้น จะเกิดแค่จุดหนึ่งจุดที่มีรัศมีโดยรอบไม่ชัดเจน ซึ่ง ณ ตำแหน่งตรงกลางจะเป็นสีขาวและวงรอบเป็นสีแดงไม่คงที่ตามรัศมีนั้นๆ ด้วยคุณลักษณะเฉพาะนี้จึงไม่สามารถใช้วิธีการของ Harris ในการตรวจจับจุดสำคัญได้อีกต่อไป จากการศึกษาพบว่าวิธีการของ Camshift จะให้ผลลัพธ์ที่ดีกว่ามากซึ่งมีคุณสมบัติดังนี้คือ

วิธีการ Camshift [8] ย่อมาจาก Continuously Adaptive Mean Shift ซึ่งเป็นวิธีที่มีการปรับแต่งมาจากอัลกอริทึมที่เรียกว่า Mean Shift อันเป็นวิธีการที่สามารถตรวจจับคุณลักษณะของกลุ่มสีของพิกเซล (ที่สัมพันธ์กับพื้นที่ในภาพที่เปลี่ยนแปลงได้ตามเวลา) ตามความน่าจะเป็นที่กำหนดให้ได้ถูกต้องมาก ซึ่งเทคนิคนี้จำเป็นที่จะต้องกำหนดความน่าจะเป็นของกลุ่มสีที่ได้มาจากคุณลักษณะของภาพในบริเวณที่สนใจ ซึ่งสามารถคำนวณหาได้ด้วยวิธีที่เรียกว่า histogram back projection ซึ่งค่าของสีที่ใช้จะเป็นโมเดล HSV ทั้งนี้การคำนวณจะเป็นไปในลักษณะของการทำซ้ำเพื่อปรับปรุงหรือติดตามคุณสมบัติของความน่าจะเป็นของกลุ่มสีนั้นให้คงที่หรือดีขึ้นให้มากที่สุดเท่าที่จะทำได้ ซึ่งหลังจากปรับค่าที่เหมาะสมได้ในแต่ละครั้งก็จะสามารถหาพิกเซลที่เกี่ยวข้องกับกลุ่มสีนั้นได้ ซึ่งสุดท้ายจะทำให้สามารถคำนวณหาจุดศูนย์กลางและของเขตของบริเวณที่สนใจไว้ได้

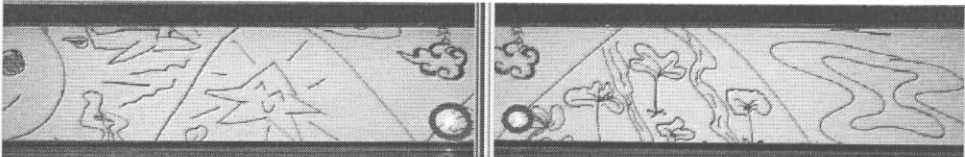
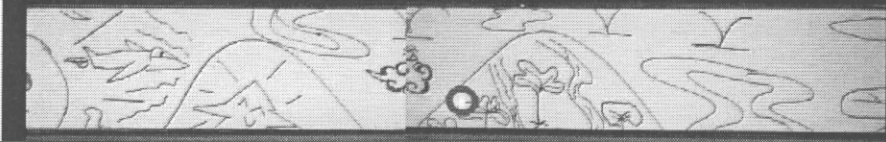
### ขั้นตอนการการเชื่อมต่อภาพด้วย Camshift

**ขั้นตอนที่ 1** กำหนดค่าเริ่มต้นของจุดที่ต้องการติดตาม ซึ่งสามารถทำได้โดยการตัดพื้นที่บริเวณจุดดังกล่าว และ ทำการหาค่า Histogram เก็บไว้ ซึ่งจำเป็นอย่างยิ่งที่จะต้องมียหลายค่า เช่น histogram ของจุดตอนแสงน้อย แสงมาก หรือ แสงปกติ เป็นต้น

**ขั้นตอนที่ 2** ใช้วิธีการ Camshift ในการติดตามค่า histogram ที่ได้กำหนดไว้ตามขั้นตอนที่ 1 จากภาพทั้งสองภาพ ณ บริเวณที่มีส่วนเหลื่อมล้ำกัน ซึ่งผลลัพธ์ที่ได้จะเป็นตำแหน่งจุดศูนย์กลางของจุดอ้างอิงในภาพทั้งสองภาพ

**ขั้นตอนที่ 3** นำตำแหน่งที่ได้มาคำนวณหาการ Translation ของภาพตามแกน x และแกน y ตามลำดับ แล้วนำค่าที่ได้ไปทำการเชื่อมต่อภาพต่อไป

จากรูปที่ 5.15 แสดงผลการจุดที่ได้จากการตรวจจับและติดตามโดยใช้วิธีการ Camshift รวมทั้งผลลัพธ์ที่ได้หลังจากเชื่อมต่อภาพได้เรียบร้อยแล้ว

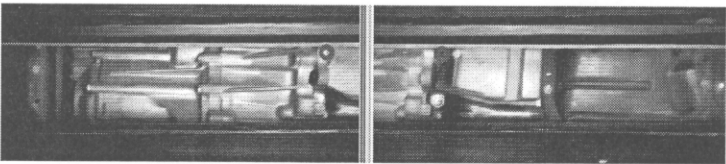
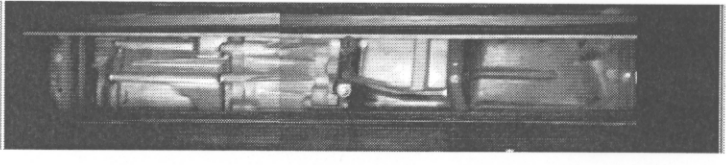
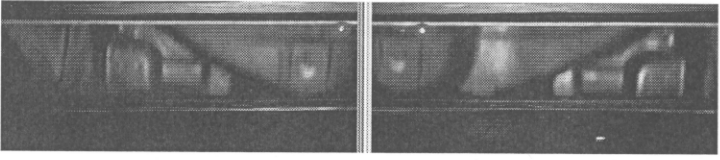
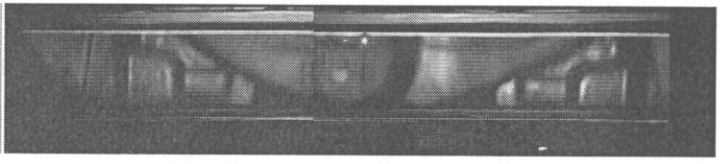
การติดตาม	
ผลลัพธ์	

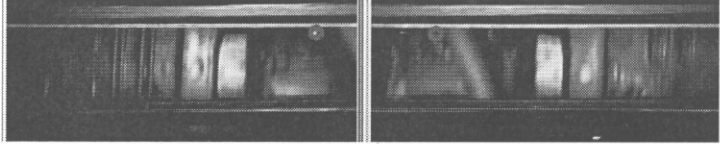
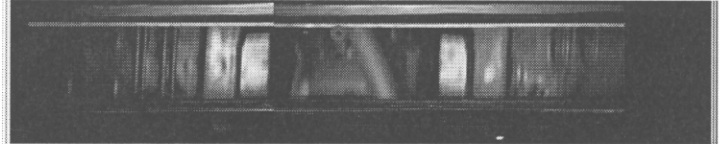
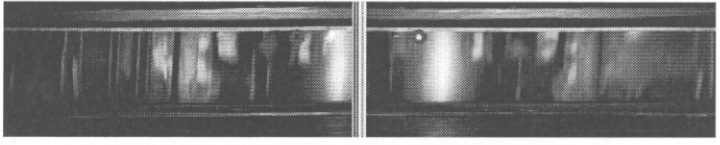
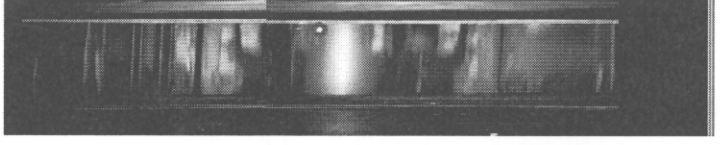
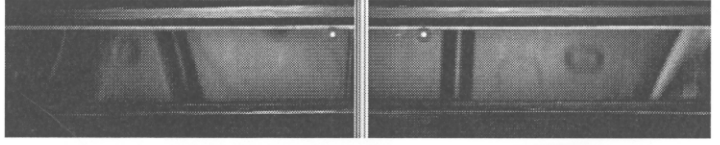
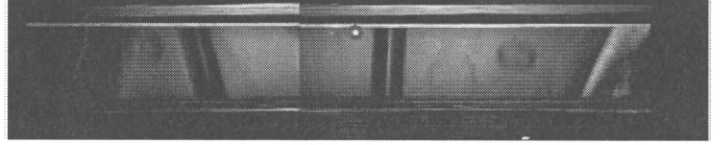
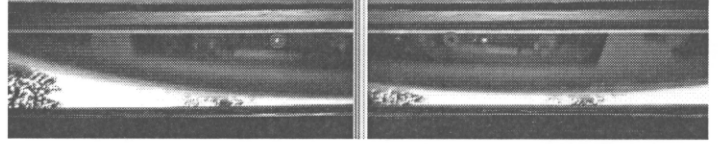
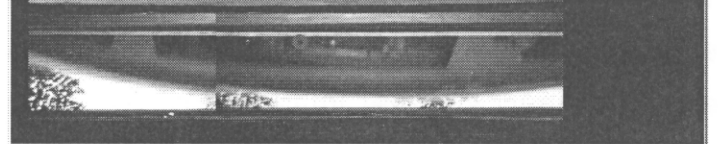
รูปที่ 5.15 ติดตามจุดที่สนใจโดยใช้อัลกอริทึม Camshift

#### 5.4 ผลการทดลองและวิเคราะห์

ในการทดสอบประสิทธิภาพของระบบการตรวจจับภาพได้ทั้งรถ ก็ได้ใช้ข้อกำหนดการตรวจสอบที่เป็นไปในลักษณะเดียวกันกับการทดสอบก่อนหน้านี้คือ การทดสอบตามสภาวะจริงที่มีการเปลี่ยนแปลงของแสงอย่างต่อเนื่อง ตารางที่ 5.2 แสดงการทดสอบจำนวน 6 ครั้ง ทั้งกลางวันและกลางคืน ทั้งที่มีแหล่งกำหนดแสงช่วยหรือไม่ก็ตาม ซึ่งได้ติดตั้งให้ระบบสามารถทำงานเองได้อย่างอัตโนมัติทั้งกระบวนการ

จากการทดสอบพบว่า วิธีการที่ได้นำเสนอไปข้างต้นนี้ สามารถทำการเชื่อมต่อภาพได้อย่างถูกต้องทั้งในสภาวะที่มีสัญญาณรบกวนในลักษณะต่างๆ โดยการทำงานของโปรแกรมมีความเสถียรและสามารถทำงานได้ตามความเร็วจริง 25 เฟรมต่อวินาที

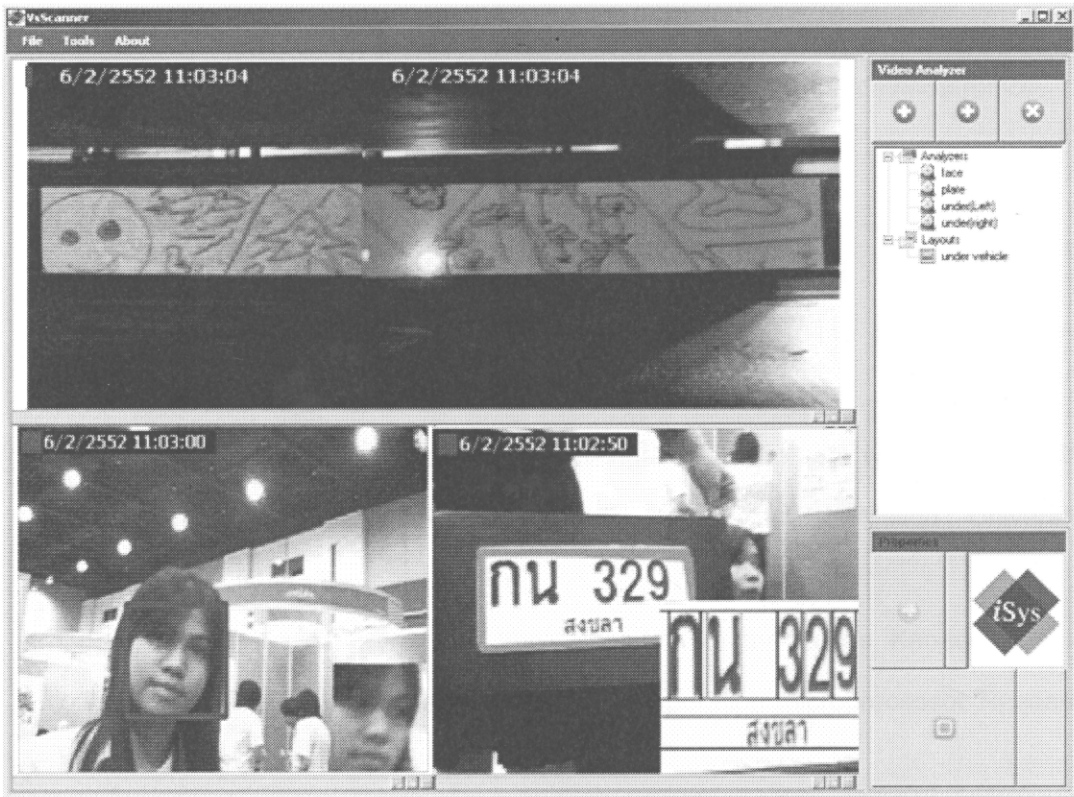
การทดลองครั้งที่ 1	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 2	ภาพรับเข้า	
	ผลลัพธ์	

การทดลองครั้งที่ 3	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 4	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 5	ภาพรับเข้า	
	ผลลัพธ์	
การทดลองครั้งที่ 6	ภาพรับเข้า	
	ผลลัพธ์	

ตารางที่ 5.2 แสดงผลลัพธ์ของการต่อภาพใต้ท้องรถด้วยวิธีของ Camshift

จากรูปต่อไปนี้จะแสดงเทคนิคการวิเคราะห์ภาพเพื่อการเชื่อมต่อภาพ โปรแกรมการเชื่อมต่อในส่วนนี้จะเป็นการนำโปรแกรมที่ได้เขียนขึ้นไว้ไปรวมเข้ากับโปรแกรมอื่นๆ คือ การตรวจจับใบหน้า และค้นหาป้ายทะเบียนรถยนต์ ทั้ง 3 โปรแกรมจะถูกเขียนขึ้นใหม่ให้กลายเป็นโมดูล ที่จะนำไปใส่ไว้ในโปรแกรม VsMonitor ซึ่งเป็นโปรแกรมที่สามารถเรียกดูภาพจากกล้องหลายๆ

กล้องได้ ดังนั้นโมดูลที่ใส่เข้าไปจะทำให้โปรแกรม VsMonitor สามารถนำภาพจากกล้องมาวิเคราะห์ได้ตามความสามารถของโมดูลแต่ละตัว



## 5.5 สรุปผล

งานวิจัยนี้เน้นที่การประยุกต์ใช้เครื่องมือทางด้านการวิเคราะห์ภาพที่มีอยู่แล้วมาใช้สำหรับทำระบบตรวจจับภาพได้ทั้งรถ ซึ่งสามารถแบ่งได้เป็นสองส่วนคือ การออกแบบและจัดทำอุปกรณ์กล้องรับภาพ โดยการใช้กล้องที่มีราคาถูกแต่หลายตัวเพื่อให้สามารถรับภาพได้ครบตามความกว้างและความสูงของรถ การรับภาพจากได้ทั้งรถจะได้จากภาพสะท้อนกระจกเงาที่ทำมุม 45 องศากับระบบ ทั้งนี้เพื่อลดจำนวนกล้องที่จะใช้ให้มากที่สุด โดยมีการนำ Laser pointer มาใช้สำหรับการต่อภาพและลดจำนวนกล้องที่ใช้ได้อีกระดับหนึ่ง ส่วนที่สอง จะเป็นการนำภาพที่ได้จากกล้องหลายตัวมาทำการเชื่อมต่อกัน โดยที่ภาพจากกล้องที่ติดกันจะมีส่วนที่เหลื่อมกันอยู่ ซึ่งในบริเวณดังกล่าวจะมีจุดอ้างอิงที่เกิดจากแสง laser ในการทำการ match กันระหว่างภาพตามจุดอ้างอิงนี้ ได้นำเอาเทคนิค Camshift มาใช้

อุปกรณ์ต้นแบบนี้ได้ทำการทดสอบการใช้งานในสภาพสิ่งแวดล้อมที่จะใช้งานจริง ซึ่งพบว่าผลลัพธ์ที่ได้มีความถูกต้องในขั้นที่ยอมรับได้ อย่างไรก็ตาม ยังจำเป็นต้องมีพัฒนาเพิ่มเติมในอีกหลายส่วน เช่น การต่อภาพของรถทั้งคันให้เป็นแค่ภาพเดียว การตรวจสอบภาพทางด้านข้างจากด้านในของรถ การให้แสงที่มีความสว่างอย่างเพียงพอและเหมาะสม หากให้มีแสงสว่างมากไปก็จะไม่สามารถวิเคราะห์จุดอ้างอิงได้อย่างถูกต้องอันนำไปสู่การเชื่อมต่อภาพที่ผิดพลาดได้

## 5.6 บรรณานุกรม

- [1] PPA Under Vehicle Surveillance System,  
<http://www.ppainnovation.com/uv/index.html>
- [2] Under Vehicle Surveillance System (UVSS) Integration, <http://davinoptronics-its.com/cms/license-plate-reading/applications/uvss.html>
- [3] Und-Aware™ Under-Vehicle-Surveillance-System,  
[http://www.aecinfo.com/1/pdcnewsitem/01/41/28/index\\_1.html](http://www.aecinfo.com/1/pdcnewsitem/01/41/28/index_1.html)
- [4] C. Harris and M.J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages 147–152, 1988.
- [5] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute, 1980
- [6] Jean-Yves Bouguet, Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm, Intel Corporation Microprocessor Research Labs.
- [7] Mitiche, A. Grisell, R. Aggarwal, J.K., Verdun, Que., On smoothness of a vector field-application to optical flow, IEEE Transactions on Pattern Analysis and Machine Intelligence, Nov 1988, Volume: 10, Issue: 6, page(s): 943-949, ISSN: 0162-8828
- [8] Comaniciu and D. Meer, P. , Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, May 2002, Volume: 24, Issue: 5, page(s): 603-619, ISSN: 0162-8828.

## บทที่ 6

### ระบบการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิด

#### A Generic Visual Surveillance

บทนี้จะนำเสนอระบบการเฝ้าระวังด้วยกล้องวงจรปิดที่จะให้การสนับสนุนการวิเคราะห์ผลภาพตามความเร็วจริงได้ เช่น การตรวจจับการเคลื่อนไหว การติดตาม และการบันทึกเหตุการณ์ผิดปกติ เป็นต้น ระบบสามารถรองรับการทำงานกับกล้องวงจรปิดจำนวนหลายกล้องพร้อมๆ กัน โดยที่จะเชื่อมต่อกับข้อมูลวิดีโอที่มีมาตรฐานแตกต่างกันได้ด้วย ซึ่งทั้งนี้จะขึ้นอยู่กับบริษัทผู้ผลิตที่มีความหลากหลาย สถาปัตยกรรมของระบบได้ถูกออกแบบเพื่อให้สามารถแสดงผลวิดีโอ จัดเก็บข้อมูลวิดีโอ ค้นหา และเปิดไฟล์วิดีโอแบบย้อนหลังได้อย่างมีประสิทธิภาพ ขั้นตอนวิธีการของการวิเคราะห์ภาพวิดีโอสามารถพัฒนาแยกได้อย่างอิสระจากระบบซึ่งสามารถเชื่อมต่อกันได้ในภายหลัง โดยได้มีการทดสอบประสิทธิภาพการทำงานของระบบสำหรับการวิเคราะห์ภาพเพื่อการตรวจจับการเคลื่อนไหวและการจัดเก็บภาพวิดีโอ

#### 6.1 ปัญหาและความสำคัญ

ในบทนี้ได้นำเสนอระบบการเฝ้าระวังด้วยกล้องวงจรปิดแบบใหม่ที่สามารถทำงานร่วมกับกล้องวิดีโอวงจรปิดแบบ IP ได้ เป้าหมายสำคัญคือการพัฒนาระบบให้มีประสิทธิภาพในการวิเคราะห์ข้อมูลภาพวิดีโอ จัดเก็บ ค้นหา และการแสดงผลแบบย้อนหลัง เป็นต้น ปัจจุบันระบบเฝ้าระวังด้วยกล้องวงจรปิดแบบ IP ได้รับความนิยมอย่างมาก เนื่องจากเป็นระบบที่สามารถตอบสนองความต้องการของผู้ใช้ที่จะนำระบบดังกล่าวมาเป็นเครื่องมือช่วยเหลือในการดูแลความปลอดภัยได้เป็นอย่างดี จากรูปที่ 1 แสดงถึงระบบการเฝ้าระวังด้วยกล้องวงจรปิดรุ่นที่สาม ซึ่งข้อมูลภาพวิดีโอจากกล้องแบบดิจิทัลสามารถที่จะถูกนำมาวิเคราะห์แบบอัตโนมัติและจัดเก็บอย่างชาญฉลาด ซึ่งจะทำการสืบค้นสามารถทำได้อย่างถูกต้องและรวดเร็ว อันเป็นลักษณะพื้นฐานของระบบการเฝ้าระวังด้วยกล้องวงจรปิดที่ทันสมัยนี้ ในช่วงเวลาที่ผ่านมาได้มีงานวิจัยหลายชิ้น เช่น [1][4][5] และ [6] ที่มีวัตถุประสงค์เพื่อที่จะออกแบบสถาปัตยกรรมของระบบการเฝ้าระวังด้วยกล้องวงจรปิดให้สามารถสนับสนุนเทคโนโลยีใหม่ๆ ได้ จากงานวิจัย [2][3] และ [7] ได้นำเสนอวิธีการติดตามของวัตถุที่ต้องสงสัยโดยใช้กล้องวงจรปิดหลายๆ ตัวพร้อมกันในลักษณะที่จะช่วยเหลือโดยการให้ข้อมูลซึ่งกันและกันได้แบบอัตโนมัติ

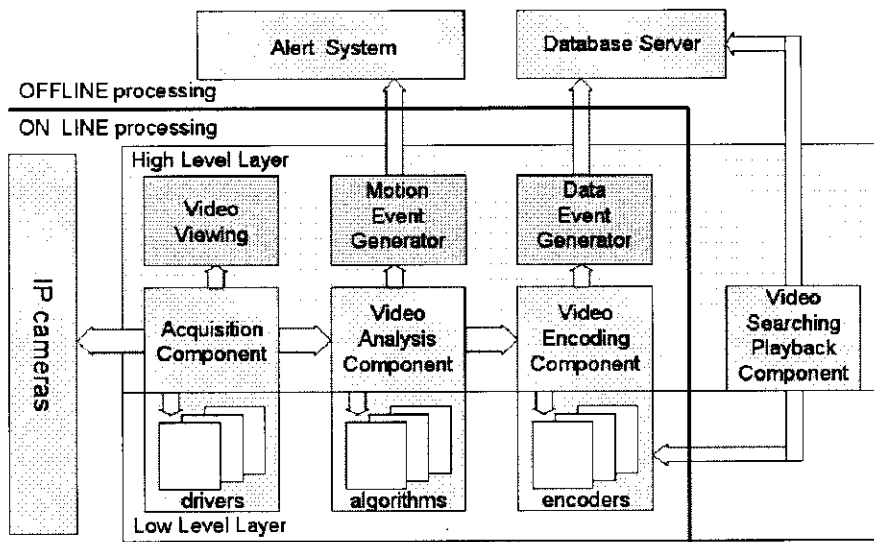
ระบบการเฝ้าระวังด้วยกล้องวงจรปิดที่มีขายอยู่ในท้องตลาดนั้นได้มีการนำไปใช้งานโดยบริษัทและองค์กรการปกครองส่วนต่างๆ อย่างกว้างขวาง ซึ่งระบบที่มีอยู่ค่อนข้างจะมีคุณภาพที่หลากหลายซึ่งโดยปกติแล้วจะขึ้นอยู่กับเงื่อนไขของราคา ระบบเฝ้าระวังที่พัฒนาขึ้นโดยส่วนใหญ่ก็จะให้การสนับสนุนเฉพาะผลิตภัณฑ์กล้องและระบบเครือข่ายบริษัทผู้ผลิต

เท่านั้น นั่นคือระบบจะถูกออกแบบเป็นระบบปิดเพื่อการทำงานที่มีความสัมพันธ์กับอุปกรณ์ กล้องวงจรปิดและเครือข่ายของบริษัทเดียว ซึ่งเป็นที่มาของงานวิจัยนี้ที่ต้องการพัฒนาระบบ การเฝ้าระวังด้วยกล้องวงจรปิดที่สามารถนำภาพจากกล้องของบริษัทต่าง ๆ มาทำการวิเคราะห์ และจัดเก็บเพื่อการค้นหาได้อย่างมีประสิทธิภาพในระบบเดียวกันได้

รายงานนี้จะนำเสนอระบบโดยแบ่งออกเป็นหัวข้อดังต่อไปนี้คือ ส่วนที่ 2 จะกล่าวถึง สถาปัตยกรรมการออกแบบของระบบทั้งหมด ส่วนที่ 3 จะนำเสนอการวิเคราะห์ข้อมูลภาพวิดีโอ ส่วนที่ 4 เป็นการนำเสนอระบบการจัดเก็บไฟล์วิดีโอ และใน 2 ส่วนสุดท้ายนำเสนอผลการ ประเมินประสิทธิภาพของระบบและสรุปผล ตามลำดับ

## 6.2 การออกแบบระบบเฝ้าระวังด้วยกล้องวิดีโอวงจรปิด

ฟังก์ชันการทำงานหลักของระบบสามารถแบ่งได้เป็น 6 ส่วนหลักๆ ได้แก่ ส่วนของรับ ภาพ การแสดงผล การวิเคราะห์วิดีโอ การจัดเก็บ การเตือน และการค้นหาและเล่นย้อนหลัง สถาปัตยกรรมของระบบได้ถูกออกแบบให้มี 2 เลเยอร์สำคัญคือ High Processing Layer (HPL) และ Low Processing Layer (LPL) แสดงดังรูปที่ 2



รูปที่ 6.1 สถาปัตยกรรมระบบ

จากสถาปัตยกรรมของระบบข้างต้นสามารถสังเกตได้ว่าจะมี High Processing Component (HPC) คือคอมโพเนนต์ที่ทำงานอยู่บนเยอร์ HPL และ LPC คือคอมโพเนนต์ที่ทำงานอยู่บนเลเยอร์ LPL ตามลำดับ โดยที่ HPL มีหน้าที่ให้การสนับสนุนการติดต่อกันระหว่าง HPC รวมถึงการเชื่อมต่อกับระบบภายนอก เช่น กล้องวงจรปิด ระบบการเตือน และระบบฐานข้อมูล เป็นต้น ซึ่งการทำงานจะให้ความสำคัญเกี่ยวกับการส่งผ่านข้อมูลภาพระหว่างคอมโพเนนต์ในส่วนต่างๆ โดยทั่วไปพบว่า คอมโพเนนต์ของการแสดงผลภาพจะสามารถทำงานสิ้นสุดลงได้ในที่เวลาไม่แตกต่างกันมากเมื่อเทียบกับคอมโพเนนต์ของรับภาพ ซึ่งในกรณีนี้จะไม่ทำให้เกิดปัญหาขึ้น อย่างไรก็ตาม ยังมีอีกหลายกรณีที่เวลาที่ใช้ในการประมวลผลในแต่ละคอม



โพเน็นท์ที่มีความแตกต่างกัน เช่น การวิเคราะห์ภาพโดยอัลกอริทึมที่มีความซับซ้อนหรือการบีบอัดภาพโดยการเข้ารหัสด้วยคุณภาพที่ต่างกัน โดยในกรณีดังกล่าวพบว่าเวลาที่ใช้ในการประมวลผลจะมากกว่าเวลาสำหรับการรับภาพ ซึ่งจะทำให้เกิดปัญหาที่จะไม่สามารถแลกเปลี่ยนข้อมูลภาพระหว่างคอมพิวเตอร์โพเน็นท์ที่อยู่ในการควบคุมของ HPL ได้อย่างถูกต้อง ซึ่งระบบที่นำเสนอได้แก้ไขด้วยการเพิ่มบัฟเฟอร์ระหว่างคอมพิวเตอร์โพเน็นท์ขึ้น โดยในการแลกเปลี่ยนข้อมูลภาพจะต้องผ่านบัฟเฟอร์ซึ่งเป็นกระบวนการการกำหนดสิทธิการเข้าถึงข้อมูลภาพทั้งเพื่อการอ่านและการเขียนข้อมูลที่จะทำให้เกิดการสูญเสียของข้อมูลภาพในระหว่างคอมพิวเตอร์โพเน็นท์ที่กำลังประมวลผลอยู่ได้

สำหรับคอมพิวเตอร์โพเน็นท์ LPC ซึ่งเป็นหน่วยประมวลผลภายในของ HPC นั้นจะไม่มี การติดต่อระหว่างกันบนบัสเลย์เออร์ LPL เลย ทั้งนี้ HPC จะมีความสามารถที่จะทำการโหลดหรือไม่ โหลด LPC เพื่อการสั่งการให้ทำการเอ็กซีคิวต์งานใดงานหนึ่งได้ ตัวอย่างของคอมพิวเตอร์โพเน็นท์ชนิด LPC ได้แก่ หน่วยรับข้อมูลภาพจากกล้องวงจรปิดแบบ IP ที่มีความแตกต่างกันของไดเรกทอรีของกล้องนั้นๆ หรือวิธีและอัลกอริทึมในการวิเคราะห์ภาพวิดีโอ หรือการบีบอัดข้อมูลเพื่อการจัดเก็บ เป็นต้น คอมพิวเตอร์โพเน็นท์ LPC สามารถที่จะพัฒนาแยกออกจากระบบหลักอย่างเป็นอิสระ โดยจะสามารถเชื่อมต่อเข้ากับระบบได้ในภายหลังด้วยตัวเชื่อมต่อ Plug-in โดยปกติแล้วเพื่อให้การประมวลผลข้อมูลภาพได้ตามเวลาจริง คอมพิวเตอร์โพเน็นท์ LPC จะถูกพัฒนาขึ้นโดยใช้ภาษาซีเป็นหลัก

จากระบบข้างต้นสามารถสังเกตได้ว่าจะประกอบด้วยสองส่วนหลักๆ ส่วนแรกเป็นการประมวลผลภาพตามความเร็วจริงหรือออนไลน์ที่มี 5 องค์กรประกอบสำคัญ ซึ่งส่วนของการรับภาพจะร้องขอข้อมูลภาพจากกล้องวงจรปิด ผ่านอินเทอร์เน็ตหรือเครือข่าย LAN แล้วส่งผ่านไปยังส่วนยังส่วนของการประมวลผลภาพวิดีโอ ผลลัพธ์ที่ได้จากการวิเคราะห์ที่มีลักษณะเป็นข้อมูลภาพหรือข้อมูลภาพวิดีโอหรือเป็นคุณลักษณะพิเศษของรูปภาพ ก็จะถูกนำไปจัดเก็บและใช้ในการแจ้งเตือนต่อไป ในส่วนที่สองของระบบคือการค้นหาและการเล่นย้อนกลับซึ่งเป็นองค์ประกอบสำคัญของระบบที่ทำงานในลักษณะออฟไลน์ โดยคุณลักษณะอย่างละเอียดของแต่ละคอมพิวเตอร์โพเน็นท์จะเป็นดังต่อไปนี้

### 6.2.1 การรับข้อมูลภาพ

การรับข้อมูลภาพจากกล้องสามารถทำได้โดยการเข้าถึงอุปกรณ์กล้องแบบ IP ได้โดยตรงผ่านระบบอินเทอร์เน็ตหรือระบบเครือข่าย ทั้งนี้ลักษณะของการเข้าถึงนั้นจะเป็นไปตามยี่ห้อและรุ่นของกล้องที่ผลิตโดยบริษัทต่าง ๆ การร้องขอภาพหนึ่งจากกล้องถือได้ว่าเป็นการรับข้อมูลภาพขั้นพื้นฐานที่สามารถทำได้โดยการระบุตำแหน่งของ URL ของกล้องนั้น (ซึ่งจะขึ้นอยู่กับยี่ห้อและรุ่น) ซึ่งจะทำให้ได้ข้อมูลภาพหนึ่งแบบ JPEG จากกล้อง หากต้องการภาพเวลาใด ก็สามารถร้องขอคำร้องผ่าน HTTP ไปยังเว็บเซิร์ฟเวอร์ของกล้องวงจรปิดนั้น ๆ ซึ่งทุกครั้งที่มีการร้องขอก็ต้องเสียเวลาขณะหนึ่งในกระบวนการการติดต่อร้องขอ ในกรณีที่ต้องการภาพเคลื่อนไหวก็มีความจำเป็นที่จะต้องเรียกหลาย ๆ ครั้ง ซึ่งในทางปฏิบัติจะทำให้การทำงาน

ช้าเพราะต้องสูญเสียความเร็วส่วนหนึ่งในการติดต่อร้องขอ อย่างไรก็ตาม การร้องภาพหนึ่งในลักษณะดังกล่าวจะทำให้ระบบสามารถควบคุมอัตราเฟรมต่อวินาทีที่ได้อย่างถูกต้องได้

รูปแบบการรับข้อมูลภาพที่มีความนิยมอีกชนิดหนึ่งสำหรับกล้องแบบไอพี ได้แก่ Motion JPEG (MJPEG) ซึ่งสามารถทำให้ได้ข้อมูลภาพเคลื่อนไหวจากกล้องแบบต่อเนื่องในลักษณะของภาพชนิด JPEG ซึ่งเมื่อต้องการข้อมูลภาพ ก็สามารถทำได้โดยการร้องขอผ่าน HTTP ติดต่อไปยังของกล้อง ณ URL ที่กำหนด (ขึ้นอยู่กับยี่ห้อและรุ่น) หลังจากนั้นกล้องจะตอบกลับการร้องขอ หากการติดต่อสำเร็จก็จะส่งภาพแบบ JPEG มาให้อย่างต่อเนื่อง การทำงานในลักษณะดังกล่าวจะประหยัดการใช้ช่องสัญญาณ (bandwidth) เนื่องการลดการติดต่อร้องขอภาพ กล้องวงจรปิดที่มีขายอยู่ในท้องตลาดส่วนใหญ่จะรองรับมาตรฐานนี้ เช่น Axis D-link Panasonic และ Bosch เป็นต้น

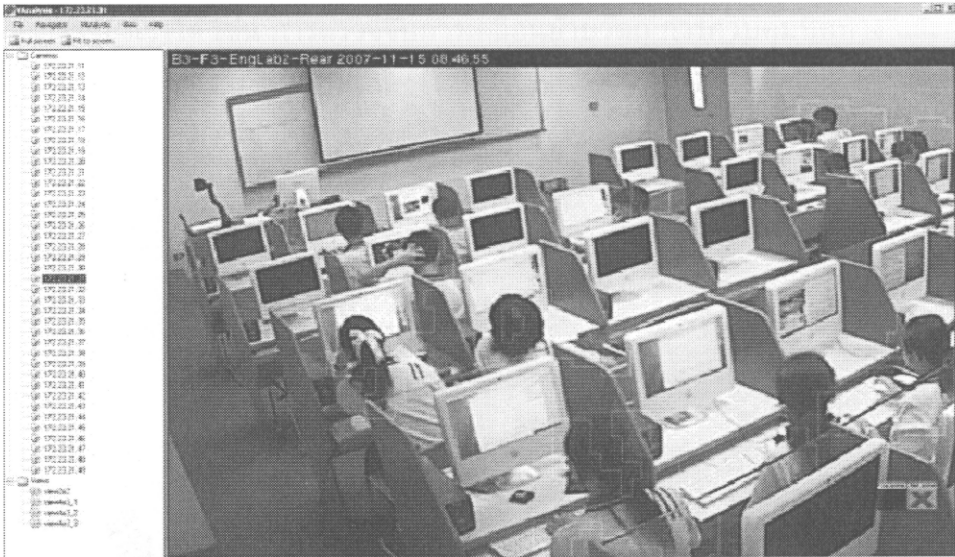
การส่งข้อมูลภาพวิดีโอแบบ MPEG-2 และ MPEG-4 เป็นรูปแบบที่กำลังได้รับความนิยมที่สูงสุดในขณะนี้ เนื่องจากสามารถส่งข้อมูลภาพวิดีโอคุณภาพสูงได้โดยใช้ช่องสัญญาณต่ำ ซึ่งในกรณีนี้จำเป็นจะต้องมีตัวถอดรหัสของ MPEG-2 และ MPEG-4 ด้วยถึงจะสามารถนำภาพจากกล้องมาใช้ได้ อย่างไรก็ตามสำหรับกล้องในบางยี่ห้อจะมีการกำหนดรูปแบบการเข้ารหัสและถอดรหัสเพิ่มเติมเฉพาะสำหรับผลิตภัณฑ์นั้นๆ หรือที่เรียกว่าเป็นรูปแบบ Native ทั้งนี้เพื่อการเพิ่มเติมข้อมูลสำคัญต่าง ๆ เช่น การเพิ่มลายน้ำแบบพิเศษลงไปเพื่อการจัดการกับความผิดปกติของวิดีโอที่จะสามารถตรวจพบได้หากมีการปรับปรุงหรือตัดต่อวิดีโอดังกล่าว โดยปกติแล้วบริษัทผู้ผลิตจะพัฒนาคอมโพเนนต์ ActiveX เฉพาะเพื่อให้สามารถการเข้าถึงรูปแบบวิดีโอที่เป็นแบบ Native เหล่านี้ได้

## 6.2.2 ส่วนของการติดต่อกับผู้ใช้

วัตถุประสงค์หลักของการพัฒนาระบบเฝ้าระวังด้วยกล้องวงจรปิด คือการที่จะสามารถนำภาพจากกล้องที่มีหลากหลายแบบมาทำการวิเคราะห์และแสดงผลได้บนระบบเดียวกัน แม้ว่ากล้องเหล่านั้นจะมีความแตกต่างกันในเรื่องยี่ห้อหรือรุ่นก็ตาม โดยระบบที่ออกแบบจะต้องสามารถทำงานได้ง่ายและสะดวกต่อการใช้งานซึ่งได้พัฒนาในส่วนของการติดต่อกับผู้ใช้นี้คือ

ผู้ใช้สามารถทำการเพิ่มกล้องใหม่เข้าไปในระบบได้ โดยจะต้องกำหนดค่าตัวแปรต่าง ๆ ของกล้องวงจรปิดที่จะทำให้ระบบสามารถร้องขอข้อมูลภาพได้ และต้องกำหนดค่าตัวแปรของการวิเคราะห์ภาพวิดีโอแบบอัตโนมัติ รวมถึงต้องระบุชนิดของตัวเข้ารหัสเพื่อการบีบอัดข้อมูลในขั้นตอนของการจัดเก็บด้วย โดยสำหรับเวอร์ชันปัจจุบันนี้ จะมีตัวเข้ารหัสเพียงตัวเดียวเท่านั้น

ในการแสดงผลภาพวิดีโอจากกล้องต่าง ๆ ผู้ใช้สามารถจัดกลุ่มกล้องวงจรปิดหลาย ๆ ตัวที่สนใจไว้ด้วยกันได้โดยการใช้ไอคอนสัญลักษณ์ที่จะผู้ใช้ระบุจำนวนและระบุตำแหน่งของกล้องในการแสดงผลของแต่ละกลุ่มได้ อันจะทำให้ผู้ใช้สามารถแสดงผลภาพวิดีโอจากกล้องต่าง ๆ ในกลุ่มเดียวกันได้พร้อม ๆ กัน ซึ่งการแสดงผลสามารถทำได้ง่าย ๆ ด้วยการลากแล้ววางในตำแหน่งของหน่วยแสดงผลภาพวิดีโอเท่านั้น



รูปที่ 6.2 ตัวอย่างการวิเคราะห์ภาพที่ได้จากกล้องวงจรปิดของ ร.ร. มอ วิทยานุสรณ์

การกำหนดค่าในการติดตั้งของระบบต่าง ๆ เช่น ค่าเริ่มต้นการทำงานของแอปพลิเคชัน ข้อมูลรายละเอียดกล้องวงจรปิด รายละเอียดของตัววิเคราะห์ ตัวเข้ารหัสและการแสดงผล จะถูกบันทึกโดยอัตโนมัติในไฟล์แบบ XML ในรูปที่ 6.2-6.4 แสดงตัวอย่างของหน้าจอของระบบที่จะใช้ติดต่อการทำงานกับผู้ใช้

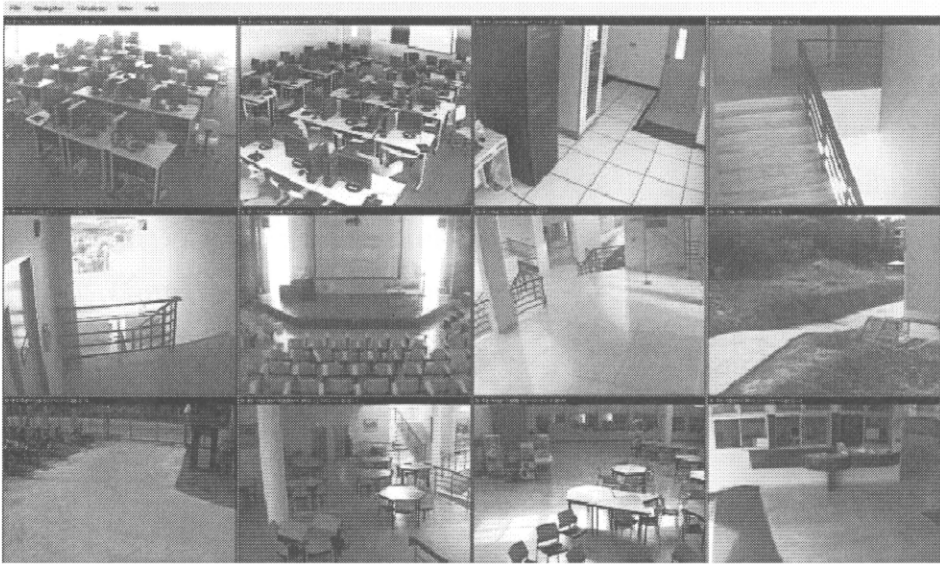


รูปที่ 6.3 แสดงหน้าต่างการจัดกลุ่มกล้องขนาด 3x4

### 6.3 ระเบียบวิธีการวิเคราะห์ภาพวิดีโอ

โดยทั่วไปแล้ว ระบบการเฝ้าระวังด้วยกล้องวงจรปิดจะมีการใช้เทคนิคการวิเคราะห์ภาพวิดีโอแบบอัตโนมัติ เช่น การตรวจจับการเคลื่อนไหว การติดตามวัตถุและการรู้จำ เป็นต้น จากอัลกอริทึมพื้นฐานเหล่านี้ ก็สามารถที่จะพัฒนาแอปพลิเคชันที่มีความซับซ้อนเพื่อให้เหมาะกับ

ความต้องการของเราได้ เช่น การตรวจจับการเคลื่อนย้ายวัตถุ การควบคุมทิศทาง และการแจ้งเตือนพื้นที่ที่ต้องการการป้องกัน เป็นต้น



รูปที่ 6.4 แสดงหน้าต่างการจัดกลุ่มกล้องขนาด 3x4 เต็มหน้าจอ

สำหรับการทดสอบระบบที่พัฒนาขึ้นนี้ อัลกอริทึมที่นำมาใช้ในการทดสอบคือ การตรวจจับการเคลื่อนไหว โดยอัลกอริทึมนี้ได้มีการพัฒนาและทดสอบแยกจากระบบข้างต้น แล้วค่อยมารวมกันภายหลังในลักษณะของโมดูล LPC ที่จะทำให้ระบบสามารถการเชื่อมต่อการวิเคราะห์นี้ได้

### 6.3.1 การตรวจจับการเคลื่อนไหวของวัตถุ

การตรวจจับการเคลื่อนไหวมีวัตถุประสงค์เพื่อการค้นหาวัตถุที่สนใจที่กำลังเคลื่อนไหวน ซึ่งอาจจะเป็น การเดินของมนุษย์หรือยานพาหนะที่กำลังเคลื่อนที่อยู่มากได้ วิธีการที่เรียกว่า Background subtraction เป็นวิธีการที่ถูกนำมาใช้ในงานวิจัยนี้ เนื่องจากมีความเหมาะสมสำหรับกรณีที่กล้องไม่เคลื่อนที่ ทั้งนี้วิธีการดังกล่าวถือได้ว่าเป็นเทคนิคที่ใช้กันโดยทั่วไปเพื่อการตรวจจับการเคลื่อนไหวของวัตถุในภาพวิดีโอ แนวคิดพื้นฐานสำหรับตรวจจับวัตถุเคลื่อนไหวน  $O$  คือการนำเอาภาพพื้นหลัง  $B_t$  มาลบออกจากภาพปัจจุบัน  $I_t$

$$O = \min(I_t - B_t, |I_{max})$$

ความแม่นยำของวิธีการนี้จะขึ้นอยู่กับคุณภาพของพื้นหลัง ซึ่งโดยปกติแล้วคุณภาพจะไม่ค่อยดีนัก ทั้งนี้เนื่องมาจากปัจจัยต่าง ๆ เช่น การกำหนด Threshold เพื่อการกำหนดว่าพิกเซลใดเป็นพื้นหลังหรือวัตถุอย่างไม่ถูกต้อง การเปลี่ยนแปลงของค่า contrast ค่าความสว่างภายในภาพอย่างรวดเร็ว และการเกิดเงา เป็นต้น ได้มีงานวิจัยจำนวนมากที่พยายามจะปรับปรุงเพื่อให้ได้มาซึ่งพื้นหลังที่ถูกต้อง ตัวอย่างเช่น วิธีการของ Gaussian mixture และวิธีการพิจารณาการกระจายแบบ Pixel-Wise Gaussian เป็นต้น อย่างไรก็ตามวิธีการต่าง ๆ

เหล่านี้มีผลทำให้อัลกอริทึมที่ได้มีความซับซ้อนยิ่งขึ้น ซึ่งจะทำให้ต้องใช้เวลาในการประมวลเพิ่มขึ้นอย่างหลีกเลี่ยงไม่ได้ ด้วยเหตุผลนี้ เราจึงเลือกโมเดลพื้นหลังอย่างง่ายโดยจะคำนวณค่าพื้นหลังใหม่จากการปรับค่าพื้นหลังเก่าในลักษณะของการหาค่าเฉลี่ย ดังสมการต่อไปนี้คือ

$$B_t = \alpha I_t + (1 - \alpha) B_{t-1}$$

อย่างไรก็ตาม การใช้เฉพาะค่าความแตกต่างระหว่างเฟรมปัจจุบันและพื้นหลังจะทำให้ได้วัตถุเคลื่อนไหวไม่ดีเท่าที่ควร จึงจำเป็นอย่างยิ่งที่ต้องพิจารณาถึงประวัติของการเคลื่อนไหวของวัตถุนั้น ๆ ด้วย ซึ่งการวิเคราะห์ประวัติการเคลื่อนที่นี้มีวัตถุประสงค์เพื่อการเพิ่มความเสถียรในการตรวจจับการเคลื่อนไหวของวัตถุให้ได้อย่างถูกต้องมากยิ่งขึ้น โดยมีแนวคิดดังนี้คือ การเคลื่อนที่ของวัตถุ ณ เวลา  $t$  จะกลายเป็นประวัติการเคลื่อนที่ของวัตถุ ณ เวลาที่  $t+1$  โดยที่ประวัติของการเคลื่อนที่ของวัตถุ  $O_t$  จะหายไปจากเวลาหนึ่งไปอีกเวลาหนึ่งตามคงที่ค่าหนึ่ง  $h$  โดยที่  $h < 1$  ดังสมการต่อไปนี้

$$O_t = \max(h \times O_{t-h}, O_t)$$

หากนำวิธีการของ Gradient มาใช้เพื่อวิเคราะห์ภาพประวัติการเคลื่อนที่ก็จะทำให้สามารถระบุทิศทางของการเคลื่อนที่ของวัตถุได้อย่างถูกต้อง ด้วยอัลกอริทึมที่อธิบายไปแล้วข้างต้นนี้ จะทำให้สามารถระบุขนาด ตำแหน่งของวัตถุ และรวมถึงทิศทางของการเคลื่อนที่ของวัตถุได้อย่างไรก็ตาม การกำหนดตัวแปรตั้งต้นต่าง ๆ ที่ไม่เหมาะสมนั้น อาจจะทำให้อัลกอริทึมทำงานผิดพลาดได้ หรือทำให้ได้ผลลัพธ์ที่ไม่ดีเท่าที่ควร ค่าที่ต้องพิจารณาคือค่า  $h$  ซึ่งหากมีค่าเข้าใกล้ 1 ก็จะทำให้ได้ประวัติการเคลื่อนที่ของวัตถุที่ยาวนาน อันหมายถึงการที่จะสามารถวิเคราะห์ทิศทางของวัตถุได้อย่างถูกต้อง แต่จะทำให้การคำนวณขนาดและตำแหน่งของวัตถุผิดพลาดแต่ในทางตรงข้าม หากกำหนดให้มีค่าน้อยก็จะทำให้ได้ทิศทางของการเคลื่อนที่ของวัตถุที่ผิดพลาดเป็นต้น สำหรับการวิเคราะห์ในงานวิจัยนี้จะพิจารณาถึงการเดินของคนหรือตรวจจับการเคลื่อนที่ของยานพาหนะ ซึ่งอัลกอริทึมดังกล่าวได้ให้ผลลัพธ์อย่างเป็นที่น่าพอใจโดยเฉพาะอย่างยิ่งเมื่อวัตถุนั้น ๆ เคลื่อนไหวอยู่ไกลจากกล้องที่กำลังพิจารณา

### 6.3.2 การเตือนเหตุการณ์เมื่อเริ่มต้นเคลื่อนไหว

การเตือนในที่นี้หมายถึงการเตือนเมื่อมีเหตุการณ์การเคลื่อนไหวเกิดขึ้นหรือหายไป ณ บริเวณที่เฝ้าระวังอยู่ ข้อมูลเหตุการณ์นี้จะถูกนำไปใช้ในระบบด้วยคอมพิวเตอร์อื่น ๆ เช่น การบีบอัด การจัดเก็บภาพและระบบแจ้งเตือน เป็นต้น ปัญหาที่ต้องคำนึงถึงอย่างมากเกี่ยวกับเทคนิคการตรวจจับการเคลื่อนที่คือ เมื่อวัตถุหยุดการเคลื่อนที่ชั่วขณะหนึ่ง (ซึ่งเกิดขึ้นบ่อยในสถานการณ์จริง) วัตถุนั้น ๆ จะไม่สามารถถูกตรวจจับได้ แล้วเมื่อเกิดการเคลื่อนไหวอีกครั้ง ก็จะทำให้ระบบเราคิดว่าเป็นวัตถุใหม่เกิดขึ้นทั้ง ๆ ที่เป็นวัตถุเดิม ระบบที่พัฒนาขึ้นนี้ได้แก้ปัญหาด้วยการนำการวิเคราะห์ Low pass filter มาใช้ ซึ่งจะทำให้ได้วัตถุการเคลื่อนไหวเพียงวัตถุเดียวหากเกิดการหยุดชั่วขณะของวัตถุนั้น ๆ ภายในบริเวณที่เฝ้าระวัง

## 6.4 ระเบียบวิธีการอัดและการจัดเก็บภาพวิดีโอ

การจัดเก็บวิดีโอของระบบเฟิร์มแวร์ด้วยกล้องวงจรปิดนี้ได้ถูกออกแบบโดยกำหนดให้มีการจัดเก็บลำดับภาพเฉพาะเวลาที่มีการเคลื่อนไหวในบริเวณที่เฟิร์มแวร์เท่านั้น ตัวเข้ารหัสจะเริ่มทำงานเมื่อเหตุการณ์การเคลื่อนไหวเริ่มขึ้น และหยุดเมื่อเหตุการณ์การเคลื่อนไหวหยุด

### 6.4.1 ตัวเข้ารหัส

ตัวเข้ารหัสจะมีลักษณะเหมือนกับคอมพิวเตอร์ LPC อื่น ๆ ที่สามารถพัฒนาแยกจากระบบรวมได้ (หรือใช้ตัวเข้ารหัสที่มีอยู่แล้ว) และสามารถเชื่อมต่อเข้ากับระบบการทำงานในภายหลังได้ ในเวอร์ชันปัจจุบัน ระบบจะใช้ Window Media Encoder SDK (ลิขสิทธิ์ของตัวเข้ารหัสจะเป็นไปตามลิขสิทธิ์ของ Windows ที่ใช้ซึ่งสามารถดาวน์โหลดจากเว็บไซต์ของไมโครซอฟต์ได้) สิ่งที่สำคัญมากในการใช้ WME คือการอธิบายคุณสมบัติของการเข้ารหัสให้ถูกต้อง โดยทั่วไปแล้วจะมีอยู่ด้วยกันสองลักษณะคือ การพิจารณาถึงคุณภาพของภาพวิดีโอที่ต้องการ หรือการพิจารณาถึงค่าบิตเรตของวิดีโอที่เหมาะสมในการส่งผ่านในระบบเครือข่ายในลักษณะต่าง ๆ ในกรณีของระบบที่พัฒนาขึ้นนั้นสามารถใช้ Windows Media Audio และ Video codecs version 9.0 ที่เข้ารหัสแบบพิจารณาคุณภาพที่ 100% เพื่อให้ใช้เวลาสำหรับกระบวนการเข้ารหัสที่น้อยที่สุด อย่างไรก็ตาม การกำหนดดังกล่าวจะทำให้ต้องการพื้นที่ของดิสก์ในการจัดเก็บข้อมูลภาพวิดีโอมากขึ้น

### 6.4.2 การเตือนเหตุการณ์เมื่อสิ้นสุดการเคลื่อนไหว

การเตือนจะเกิดขึ้นเมื่อเหตุการณ์การเคลื่อนไหวสิ้นสุดลง ซึ่งจะมีประโยชน์เป็นอย่างมากที่จะทำให้ระบบสามารถติดต่อกับส่วนต่างที่เป็นออฟไลน์ได้ อันได้แก่ แอปพลิเคชันการค้นหาและเล่นย้อนหลัง เป็นต้น ซึ่งการเตือนนี้จะเกิดขึ้นเมื่อเสร็จสิ้นกระบวนการเข้ารหัสของภาพวิดีโอของเหตุการณ์การเคลื่อนไหวที่ใด ๆ แล้ว โดยระบบจะส่งข้อมูล Metadata ที่อธิบายลักษณะของวิดีโอที่เกิดขึ้นด้วย

ข้อมูล Metadata ต่อไปนี้จะถูกส่งไปยังส่วนออฟไลน์ ได้แก่ ชื่อ ขนาด ตำแหน่งของไฟล์วิดีโอที่ถูกจัดเก็บ เวลาที่เริ่มต้น เวลาสิ้นสุดการเคลื่อนไหว ชื่อของอัลกอริทึมที่ใช้ และประเภทของการเข้ารหัส เป็นต้น ข้อมูลเหล่านี้จะถูกนำไปใช้ในแอปพลิเคชันของการค้นหา การเล่นย้อนหลัง และระบบการแจ้งเตือน ตามลำดับ

ระบบสำคัญสองระบบที่ไม่ได้อธิบายในรายละเอียดในที่นี้คือระบบการแจ้งเตือนและแอปพลิเคชันการค้นหาและเล่นย้อนหลัง ซึ่งทั้งสองระบบนี้มีความจำเป็นสำหรับระบบเฟิร์มแวร์ด้วยกล้องวงจรปิดเป็นอย่างมาก

## 6.5 ผลการทดลองและวิเคราะห์

การทดสอบระบบจะพิจารณาออกเป็นสองส่วนคือ การทดสอบการทำงานของแต่ละคอมพิวเตอร์หนึ่งเครื่อง และการทดสอบการทำงานโดยรวมทั้งระบบ ซึ่งการทดสอบได้ใช้เครื่องคอมพิวเตอร์แบบตั้งโต๊ะ Intel Core 2 Duo processor 1.7 GHz ขนาดแรม 1 Gbyte และการ์ด Ethernet ที่ความเร็ว 100 Mb/s

### 6.5.1 การรับภาพและแสดงผล

การทดสอบความสามารถในการรับภาพและแสดงผลภาพได้จัดทำโดยใช้กล้องวงจรปิดของบริษัท Axis รุ่น 206 ที่ติดตั้งในโรงเรียน มอ.วิทยานุสรณ์จำนวน 47 ตัว กล้องวงจรปิดของบริษัท Bosch ที่ติดตั้งบริเวณทางเข้ามหาวิทยาลัยจำนวน 7 ตัว กล้องวงจรปิดของ Panasonic ที่ติดตั้งในภาควิศวกรรมคอมพิวเตอร์ 1 ตัว กล้อง D-link 1 ตัว และมีเว็บแคมแบบ USB 2 ตัว

การทดสอบมีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพการทำงานพื้นฐานของระบบ โดยใช้กล้องวงจรปิด 16 ตัวจากบริษัทผู้ผลิตที่แตกต่างกัน ด้วยอัตรา 25 ถึง 30 เฟรมต่อวินาที และมีขนาดภาพ 640x480 พิกเซล โดยทำการจัดกล้องทั้งหมดไว้ในกลุ่มเดียวกันเพื่อที่จะทำให้สามารถเชื่อมต่อและแสดงผลแบบ real-time พร้อมๆ กันได้ จากการทดสอบพบว่าหลังจากที่ได้มีการใช้งานไปเป็นเวลา 48 ชั่วโมง ก็ไม่เจอข้อผิดพลาดใดๆ เลย กระบวนการทดสอบนี้ทำได้ช้าๆ เป็นเวลาหลายๆ สัปดาห์ ซึ่งพบว่าจำนวนกล้องสูงสุดที่สามารถรับมาเพื่อแสดงผลได้อย่างถูกต้องอยู่ที่ 16 ตัว ซึ่งอาจจะอธิบายได้ด้วยคุณภาพของการเชื่อมต่อกับเครือข่ายผ่านทางอินเทอร์เน็ต ที่จำกัดขนาดของช่องสัญญาณ (bandwidth) ซึ่งในการประมวลผลรูปภาพ JPEG หนึ่งรูปจะมีการใช้ช่องสัญญาณอยู่ที่ค่าเฉลี่ย 74 Kbyte หากทดสอบด้วยด้วยอัตรา 25-30 เฟรมต่อวินาทีและใช้กล้องวงจรปิดทั้งหมด 16 ตัว ช่องสัญญาณที่ใช้จะอยู่ในช่วง 29.6 ถึง 35.5 Mbyte

### 6.5.2 การตรวจจับการเคลื่อนที่ของวัตถุ

อัลกอริทึมของการตรวจจับการเคลื่อนที่ไหวได้ถูกพัฒนาขึ้นด้วยภาษา C โดยใช้ไลบรารี OpenCV ที่ถูกติดตั้งไว้ภายในระบบที่พัฒนาขึ้น รูปที่ 6.5 แสดงภาพที่ได้จากกล้องวงปิดแบบ IP ของ Panasonic ที่ติดตั้งไว้ที่ชั้นที่หนึ่งของตึกหุ่นยนต์ รูปที่ 6.6 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์โดยใช้อัลกอริทึมการตรวจจับการเคลื่อนที่ไหวดังกล่าว จากการทดสอบพบว่าความแม่นยำของวิธีการนี้ (โดยพิจารณา การระบุตำแหน่ง ขนาด และการกำหนดทิศทาง) จะขึ้นอยู่กับค่าพารามิเตอร์  $h$  โดยที่จะได้ผลลัพธ์ที่ผิดพลาดมากหากมีการเคลื่อนที่ของวัตถุใกล้ชิดกับกล้องมากๆ อย่างไรก็ตามผลลัพธ์ที่ได้โดยรวมเป็นที่น่าพอใจ เนื่องจากโดยส่วนใหญ่แล้วการเคลื่อนที่ไหวของบุคคลจะอยู่ห่างจากกล้องพอสมควร

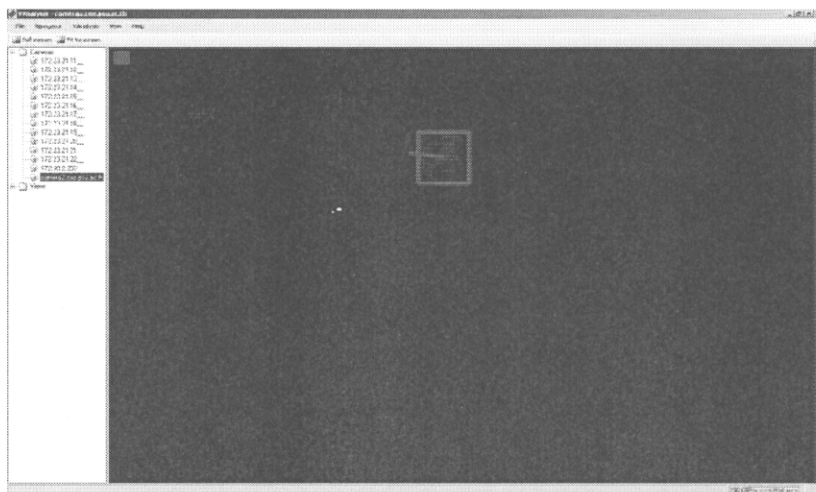
### 6.5.3 การทดสอบระบบโดยรวม

การทดสอบการทำงานของระบบโดยรวมแบ่งออกเป็นสองตอนคือ

**ขั้นตอนที่ 1** จะพิจารณาการทำงานของกล้องเพียงตัวเดียวโดยใช้กล้องวงจรปิดแบบ IP ของ Panasonic (ในบริเวณชั้นล่างของตึกหุ่นยนต์) โดยตั้งค่าไว้ที่ 10 เฟรมต่อวินาทีและใช้ภาพที่มีขนาดภาพ 640x480 พิกเซล ซึ่งระบบจะทำงานตลอด 24 ชั่วโมง จากการสังเกตผลพบว่าการตรวจจับการเคลื่อนไหวของบุคลากรสามารถทำงานได้โดยส่วนใหญ่ถูกต้อง อย่างไรก็ตาม ในบางกรณีก็ยังคงมีความผิดพลาดซึ่งสามารถแยกออกได้เป็นสองลักษณะคือ การที่แสงสว่างมีการเปลี่ยนแปลงอย่างรวดเร็วอันอาจเนื่องมาจากแสงอาทิตย์ และปัญหาของพื้นหลังที่มีการเปลี่ยนแปลงตลอด เช่น การเคลื่อนไหวของต้นไม้เนื่องจากลมพัด เป็นต้น



รูปที่ 6.5 ภาพวิดีโอจากกล้องใต้ตึกภาควิศวกรรมคอมพิวเตอร์



รูปที่ 6.6 ผลการค้นหาวัดจุดเคลื่อนไหว



**ขั้นตอนที่ 2** การทดสอบจะถูกกำหนดให้เหมือนกับตอนที่ 1 แต่เพิ่มจำนวนของกล้องให้มากขึ้นเป็น 7 ตัว ได้แก่ กล้องวงจรปิด Panasonic จำนวน 1 ตัวหน้าภาควิชา และกล้องวงจรปิดชนิด Bosch จำนวน 6 ตัวจากมหาวิทยาลัยสงขลานครินทร์ และได้ทำการรันระบบเป็นเวลาหนึ่งวัน ซึ่งพบว่าการทำงานของระบบทั้งการรับข้อมูลภาพและการวิเคราะห์ภาพวิดีโอสามารถทำงานบรรลุได้ตามวัตถุประสงค์ อย่างไรก็ตาม เมื่อการทดสอบการรันโปรแกรมไปเป็นระยะเวลาหนึ่งประมาณ 6 ชั่วโมงก็พบปัญหาที่สำคัญคือ ตัวเข้ารหัสจะทำงานผิดพลาด โดยมีข้อความแจ้งว่า object instance can't be created ซึ่งจากการสังเกตพบว่าเมื่อใช้กล้องในวิเคราะห์เป็นจำนวนมากขึ้นก็จะพบข้อผิดพลาดเร็วขึ้นด้วย ซึ่งสันนิษฐานว่าการบีบอัดข้อมูลจำนวนมากเพื่อจัดเก็บในฮาร์ดดิสก์ในเวลาเดียวกันจะมีผลโดยตรงต่อความเสถียรของระบบ

## 6.6 สรุปผล

วัตถุประสงค์ของงานวิจัยนี้คือการพัฒนาระบบเฝ้าระวังด้วยกล้องวงจรปิด ซึ่งได้นำเสนอโครงสร้างสถาปัตยกรรมของระบบโดยละเอียด ทั้งนี้การทดสอบเบื้องต้นของระบบที่แสดงให้เห็นผลลัพธ์ที่สามารถทำงานได้ถูกต้อง อย่างไรก็ตาม ยังจำเป็นอย่างยิ่งที่จะต้องมีการแก้ไขข้อผิดพลาดและทดสอบเพื่อปรับปรุงให้การทำงานมีความถูกต้องและดียิ่งขึ้นซึ่งจำเป็นอย่างมากหากต้องนำระบบดังกล่าวไปใช้งานจริง

สำหรับงานที่ต้องทำเพิ่มเติมในอนาคตนั้น ได้แก่ การพัฒนาอัลกอริทึมในการวิเคราะห์ข้อมูลวิดีโอเพื่อนำไปใช้ในงานเฉพาะทาง การพัฒนาระบบการแจ้งเตือนที่สามารถทำงานร่วมกับระบบที่พัฒนาขึ้นได้อย่างสมบูรณ์ การพัฒนาการค้นหาและเล่นย้อนหลังที่จะต้องนำผลลัพธ์ที่ได้จากการวิเคราะห์ของระบบนี้ไปใช้เพื่อการทดสอบต่อไป ซึ่งหากสามารถปรับให้ได้ดังกล่าวแล้วก็จะทำให้ระบบที่นำเสนอนี้สามารถนำไปใช้งานได้จริง

## 6.7 บรรณานุกรม

- [1] R.G.J. Wijnhoven, E.G.T. Jaspers, P.H.N. de With, Flexible Surveillance System Architecture for Prototyping Video Content Analysis Algorithms in Conference on Real-Time Imaging IX, Proceedings of the SPIE, January 2006, San Jose, CA, USA
- [2] B. Abreu, L. Botelho A. and all. Video-based multi-agent traffic surveillance system, Proceedings of the IEEE Intelligent Vehicles Symposium, 2000. Page(s):457 – 462.
- [3] T.Matsuyama and N.Ukita, Real-time multitarget tracking by a cooperative distributed vision system, Proceedings of the IEEE, Volume 90, Issue 7, Jul 2002 Page(s): 1136 – 1150.
- [4] Xiaojing Yuan, Zehang Sun, and Y. Varol, and G. Bebis, A distributed visual surveillance system, Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 21-22 July 2003, page(s): 199- 204.
- [5] B.Georis, X.Desurmont, and all, IP-distributed computer-aided video-surveillance system, Intelligence Distributed Surveillance Systems, IEEE Symposium on (Ref. No. 2003/10062) Volume, Issue , 26 Feb. 2003 Page(s): 18/1 - 18/5
- [6] C. Jaynes and S. Webb and R. Steele and Q. Xiong, An Open Development Environment for Evaluation of Video Surveillance Systems, Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance, 2002.
- [7] J. Piater and J. Crowley, Multi-modal tracking of interacting targets using Gaussian approximations, Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. IEEE Computer Society, Dec. 2001
- [8] Andrew Kirillov, Camera Vision - video surveillance on C#, <http://www.codeproject.com/KB/audio-video/cameraviewer.aspx>.

## บทที่ 7

### สรุปผลและข้อเสนอแนะ

รายงานวิจัยฉบับนี้ได้นำเสนอผลการวิจัยในสองหัวข้อสำคัญคือ การวิเคราะห์ข้อมูลภาพวิดีโอเพื่อการประยุกต์ใช้กับระบบเฝ้าระวังด้วยกล้องวงจรปิด ซึ่งได้นำเสนอในบทที่สองถึงบทที่ห้า และการพัฒนาระบบจัดการข้อมูลภาพจากกล้องวิดีโอวงจรปิด ซึ่งได้นำเสนอรายละเอียดในบทที่หก

บทที่สอง ได้นำเสนอกระบวนการวิธีการตรวจจับการเคลื่อนไหวของวัตถุ ซึ่งมีความสำคัญเป็นอันดับแรกที่ต้องใช้ในระบบการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิด วัตถุเคลื่อนไหวสามารถระบุได้ด้วยวิธีการลบพื้นหลังซึ่งมีสองลักษณะคือการพื้นหลังโดยการระบุบริเวณวัตถุและไม่ระบุบริเวณวัตถุ คุณสมบัติของวัตถุที่เคลื่อนไหวสามารถคำนวณได้ด้วยการวิเคราะห์ประวัติการเคลื่อนที่ก่อนหน้านั้น เช่น ตำแหน่ง ขนาด และ ทิศทาง ในงานวิจัยนี้ได้นำเสนอระบบที่สามารถทำงานได้ถูกต้องมากยิ่งขึ้น โดยการนำแนวทางการวิเคราะห์พื้นหลังทั้งสองแบบมาผสมผสานกันอย่างเหมาะสม ซึ่งจากการทดสอบพบว่าผลลัพธ์ที่ได้มีความถูกต้องมากขึ้น

บทที่สาม ได้นำเสนอกระบวนการวิธีการวัดคุณภาพของภาพใบหน้าเพื่อนำไปใช้ในการวิเคราะห์ภาพใบหน้าที่ดีที่สุด ขั้นตอนเริ่มแรก ได้มีการปรับปรุงกระบวนการตรวจจับใบหน้าให้ดียิ่งขึ้น โดยการนำวิธีการของ Harr มาผสมกับวิธีของ Camshift กระบวนการนี้จะทำให้ได้ข้อมูลภาพหน้าในแต่ละเฟรมของภาพวิดีโอ ซึ่งนำไปสู่การวิเคราะห์คุณภาพใบหน้าที่ดีที่สุด ซึ่งได้นำเสนอการคำนวณความสัมพันธ์ของข้อมูล Covariance ระหว่างภาพของใบหน้าใดๆ นำมาใช้เป็นเกณฑ์วัดเพื่อการตรวจสอบคุณภาพของใบหน้าที่ดีที่สุดได้อย่างเหมาะสม ซึ่งจากการทดสอบพบว่าสามารถวิเคราะห์หาภาพที่มีคุณภาพได้อย่างถูกต้อง

บทที่สี่ ได้นำเสนอกระบวนการวิธีการตรวจจับป้ายทะเบียน โดยได้เสนอวิธีสำหรับสองขั้นตอนสำคัญคือ การค้นหาป้ายว่าอยู่ที่ตำแหน่งใดภายในภาพและการตัดแยกตัวอักษรออกจากกัน เมื่อสำเร็จแล้วก็จะนำไปสู่กระบวนการรู้จำอักษรนั้นๆ ต่อไป อย่างไรก็ตามวิธีการรู้จำตัวอักษรไม่ได้นำเสนอไว้ในที่นี้ จากผลการทดสอบพบว่าขั้นตอน ที่มีผลต่อความถูกต้องในการตัดแยกตัวอักษรคือขั้นตอนที่หนึ่ง ซึ่งจะต้องมีการปรับค่าตัวแปรบางตัวให้เหมาะสมกับสภาพแวดล้อมถึงจะทำให้ได้ผลลัพธ์ที่ดีขึ้นได้ ซึ่งจำเป็นอย่างยิ่งที่จะต้องมีการปรับปรุงขั้นตอนนี้ต่อไป

บทที่ห้า ได้นำเสนอกระบวนการวิธีการตรวจจับภาพใต้ท้องรถ ซึ่งประกอบด้วยสองส่วนหลักคือ การจัดทำอุปกรณ์การรับภาพจากใต้ท้องรถ และการนำภาพที่ได้มาประมวลผล โดยเน้นไปที่การพัฒนาระบบที่มีราคาไม่แพง การออกแบบอุปกรณ์จึงเน้นการใช้กล้องที่มีราคาถูกและมีจำนวนมากพอที่จะสามารถทำงานได้ตามต้องการ ในงานวิจัยนี้ได้ออกแบบโดยให้ใช้อาร์เรย์ของกล้องจำนวนมากว่าหนึ่งตัวเพื่อที่จะสามารถรับภาพทั้งหมดตามความกว้างของรถ

ภาพที่ได้จากหลายๆ กล้องนี้จะถูกนำมาต่อเข้าด้วยกัน ซึ่งพบว่าผลลัพธ์ที่ได้มีความถูกต้องและความเร็วมากพอที่สามารถทำงานได้ตามความเร็วจริง อย่างไรก็ตามหากมองถึงระบบโดยรวมแล้วยังพบว่า ระบบที่นำเสนอยังขาดการความสมบูรณ์ในหลายๆ ส่วน เช่น การกำหนดการเริ่มและสิ้นสุดการทำงานอัตโนมัติของระบบ และการต่อภาพได้ทั้งหมดให้เป็นภาพแค่ภาพเดียว เป็นต้น

บทที่หก ได้นำเสนอระบบซอฟต์แวร์การเฝ้าระวังด้วยกล้องวิดีโอวงจรปิดที่มีความสามารถที่จะติดต่อไปยังกล้องแบบ IP ที่มีอยู่ในท้องตลาดได้ โดยระบบดังกล่าวสามารถนำกระบวนการวิธีการประมวลผลภาพที่ได้กล่าวมาแล้วทั้งหมดก่อนหน้านี้มาใช้ร่วมกันได้ ระบบที่ได้ออกแบบมีความสามารถที่จะทำงานในเครื่องคอมพิวเตอร์แบบที่มีหลายคอร์หรือมีหลายโปรเซสเซอร์ได้อย่างมีประสิทธิภาพ รวมถึงมีการจัดการหน่วยความจำและการใช้งานแบนวิธซ์ของเครือข่ายคอมพิวเตอร์ได้เป็นอย่างดี ระบบนี้ได้ทดสอบการใช้งานจริงกับระบบกล้องของมหาวิทยาลัยสงขลานครินทร์ ระบบกล้องของโรงเรียน ม.อ.วิทยานุสรณ์ และระบบกล้องของภาควิชาวิศวกรรมคอมพิวเตอร์ ซึ่งพบว่าซอฟต์แวร์ดังกล่าวสามารถนำไปใช้งานได้จริง ปัจจุบันซอฟต์แวร์นี้ได้นำมาใช้เป็นระบบการทำงานหลักของระบบการเฝ้าระวังด้วยกล้องวิดีโอวงจรปิดของภาควิชาวิศวกรรมคอมพิวเตอร์

## ภาคผนวก

### บทความตีพิมพ์ในประชุมวิชาการ

1. Sofina Yakhu and Nikom Suvonvorn, (2009), "Key Frame Extraction from Video Sequence Using a Face Quality Index Applied to Surveillance System", In Proceedings of ECTI Conference on Application Research and Development (ECTI-CARD), May 4-6, 2009, Thailand, p.409-413. ISBN: 978-974-8285-62-7.
2. N. Suvonvorn, (2008), "A Generic Surveillance System", In Proceedings of 6th PSU Engineering Conference, May 8-9, 2008, Hat Yai, Songkhla, Thailand, p.204-208.

## A Generic Visual Surveillance System

Nikom SUVONVORN<sup>1\*</sup>

<sup>1</sup>Department of Computer Engineering, Prince of Songkla University,  
Hat Yai, Songkhla 90112  
E-mail: [korn@coe.psu.ac.th](mailto:korn@coe.psu.ac.th)\*

### Abstract

We introduce in this paper a generic software for real-time video surveillance system to support ongoing research in motion detection, tracking and abnormal event recognition. The proposed system is a multi-camera platform that able to connect with different standard of video inputs for many manufacturers. The architecture of our platform is designed in an efficient way for video display, storage, rapidly searching and playback. The algorithms for video analysis can be developed separately from platform as plug-in. In the current version, some classic algorithms are integrated such as motion detection, and tracking method. The storage system for video encoding and management is functionally based on motion events that invoking by motion detector for starting/stopping video encoding. Some preliminary tests of system are performed in real outdoor/indoor situations.

**Keywords:** surveillance system, video content analysis, motion detection, object tracking.

### 1. Introduction

In this article, we introduce a novel platform of video surveillance system working with IP-based cameras. Our goal is first to obtain an efficient complete system for video viewing, video analysis, storage and searching and playback.

For many reasons including especially security, it can be noticed that popularity of IP-based cameras for video surveillance is growing really fast. In the third-generation of video surveillance system, illustrated in figure 1, video information as a stream or sequence of digital images provides an intelligent and efficient approach for automatic system analysis. Motion detection, object tracking, and abnormal event detection are some of video analysis techniques needed as basic functions for modern surveillance system. Many researches [1][4][5][6] proposed novel surveillance system architecture for supporting the technologies. [2][3][7] presented the cooperative target tracking methods using multiple cameras of surveillance system.

The surveillance systems existing in the market using by private companies and government organizations provide a large range in terms of prices and qualities. However, most of them support only its own products, system infrastructure for their cameras. We proposed here a system which is designed for

working independently with camera devices and network infrastructure.

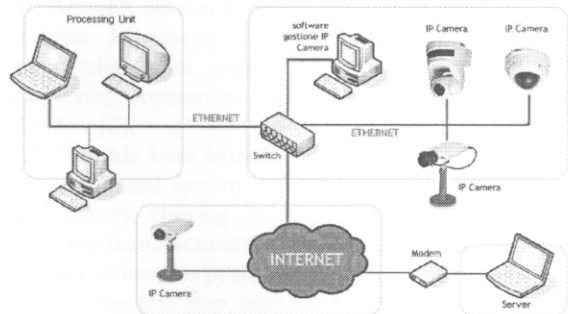


Figure 1. Third generation of video surveillance System.

The paper is organized as follows: section 2 describes architecture design of overall system, section 3 is devoted to the videos content analysis, section 4 presents video storage system, and the last two sections show system performance evaluation and conclusion respectively.

### 2. Overall system overview

The principle functions of our system can be divided into six important components: acquisition, display, video analysis, storage, alert, searching and playback. These components lay on two processing layer architecture: High Processing Layer (HPL) and Low Processing Layer (LPL), shown in figure 2. Note that High Processing Component (HPC) means all component located on HPL, and respectively for Low Processing Component (LPC). The HPL facilitates the communication between HPC components, and also with external system such as cameras, alert system and database server. It emphasizes how images would be transferred form one to another component. Let's notice that the processing time of each component is different. Normally, displaying images from acquisitions module can be terminated at the same time with no problem. However, in some situation such as images analysis by complex algorithms or image compression by encoders with different qualities, these naturally spend more processing time than the acquisitions rate. This problem of images exchange between components is take care by HPL mechanism. Each communication between any two

HPCs, a buffer is used as resource exchange zone. Only one component can access to buffer at a time for reading or writing.

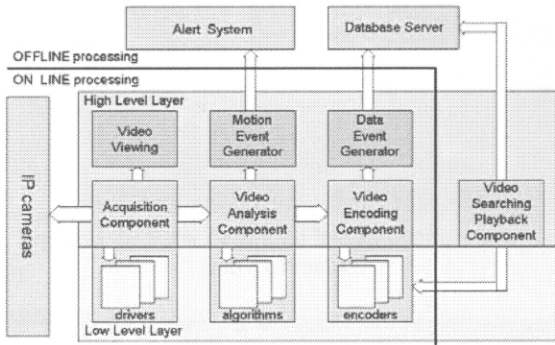


Figure 2. System Architecture.

In contrary, the LPC is considered as an internal unit of HPC component. There is no communication between LPC components on LPL layer. HPC can load/unload LPC in order to execute a specific task. For example, LPC of acquisition component corresponds to different drivers of IP-camera. In the case of video analysis and storage components, LPC represents different kind of image filters and encoders respectively. Actually, there is no LPC for displaying component. All LPC components are developed separately from the main system and acts as plug-ins. In order to decrease the computation time for video analysis, the algorithms of LPC components are mostly implemented in C.

It is important to be noticed that the first five components are some kind of online processing. The acquisition component requests image streams from IP-camera localized on internet or on local area network. Images stream are decoded and then transferred to video processing component. It produces the results as image or sequence of images or image's features that will be used by storage and alert modules.

The last component of system corresponds to the searching and playback which is an offline component. It uses the same drivers of encoder for decoding the images sequence and display. Some facilities for searching are actually developed in order to select the events of interest.

### 2.1 Acquisition

A lot of manufacturers provide a great range of IP video cameras. The simplest video format supported by almost all IP cameras is just JPEG. Cameras allow retrieving a single image by accessing a special URL. The disadvantage is that it is required to send a new HTTP request to the camera's web server each time you need to retrieve a new image, which makes loss of speed because of extra HTTP headers being sent/received. The advantage is that the acquisition component can easily control the frame rate.

The second popular format is Motion JPEG (MJPEG). This format allows downloading a stream

of JPEGs. In this case, the acquisition component does an HTTP request to a special camera's URL. Then, the camera replies to request with a stream of JPEGs delimited with a special delimiter specified in HTTP headers. MJPEG approach seems to be much better than JPEG, because it speed up the acquisition rate with lower bandwidth. Cameras from most of manufacturers support these two basic formats such as Axis, D-link, Panasonic, Bosch and etc.

MPEG-2 and MPEG-4 video format are actually the most popular. Normally, the encoders/decoders of MPEG-2 and MPEG-4 are specifically defined by the provider, called native format. By adding some special watermark filters, unusual manipulation of video can be detected. Most of manufacturers provide the ActiveX control for accessing to their native video formats. In general, the native video format is the most suitable for such system. Our acquisition and displaying component inspired the work of Andrew Kirillov [8].

### 2.2 Graphic User Interface

Proposed system is a multiple-camera platform that allows viewing many different cameras from different manufacturers at a time. User design should be easy as most as possible for using.

At the first step, user has to add new cameras by configuring the camera's parameters and video analyzers. For the current version, this is only one encoder, so there is no need to define. The available camera's drivers and image analyzers are automatically listed in the configuration dialog.

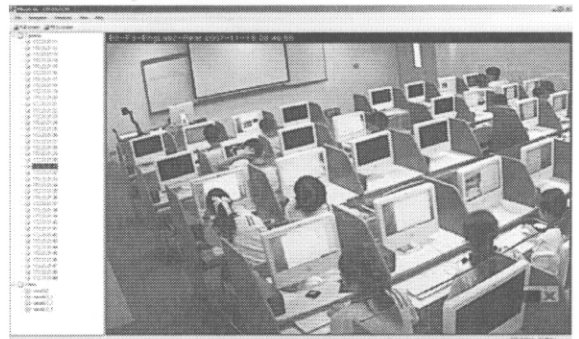


Figure 3.1. Connect to a signal camera at PSU Wittayanusorn School.

A view for grouping the interest cameras can be created by using a dialog wizard in order to simultaneously display. User can define number of cameras and position in the view that can be complete simply by drag and drop style: selecting available pre-defined cameras into view.

All configurations about application, cameras, analyzer, encoders and views are saved automatically in XML configuration files. The figure 3.1-3.3 shows examples of user interface.

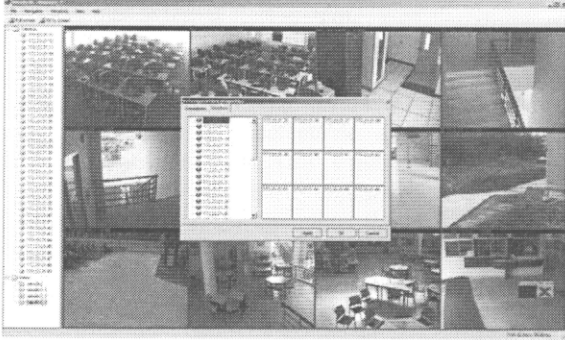


Figure 3.2. View definition by regrouping cameras into 3x4 matrix.

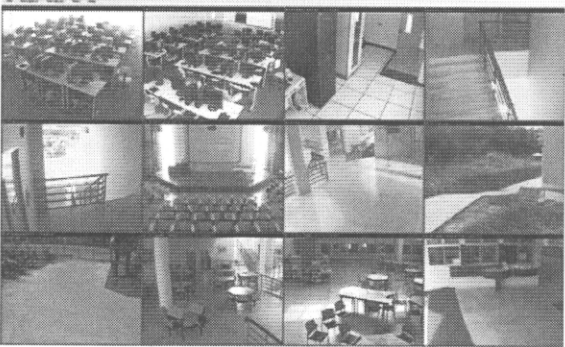


Figure 3.3. View in full screen.

### 3. Videos Analysis

For the video surveillance system, techniques for motion detection, object tracking and recognition are mostly required. From these basic algorithms, the more complex applications can be developed, for example, object removal detection, direction control, preventive area alert and etc.

In order to test our system, an algorithm for motion detection is implemented. Note that this algorithm has to develop and test separately from the principle system. Then, by adding some interface codes (LPC), the system can recognize them as its plug-ins.

#### 3.1 Motion History Detection/Segmentation

Mainly purpose of motion detection is to search the pedestrian or vehicle presences in images sequence. The "Background subtraction" method is employed in this paper which is suitable for fixed-cameras. It is a classic technique for detecting moving object in images sequence. The basic idea is that moving object  $O$  can be detected by subtracting a stationary background  $B_t$  from the current image  $I_t$ .

$$O = \min(|I_t - B_t|, I_{\max})$$

The accuracy of method depends crucially on the quality of background normally degraded by the following factors: predefined threshold setting for assigning pixels to background/foreground, quick contrast/illumination changes in scene, shadows and etc. Lots of researches have been carried out in order

to improve the background subtraction method such as mixture of Gaussian, Pixel-Wise Gaussian distribution. However, these lead to more complex algorithm which increases the inevitable computation time. By this reason, we chose a simplistic background model that updating its current background to the new one by average.

$$B_t = \alpha I_t + (1 - \alpha) B_{t-\Delta t}$$

Motion history detection has objective for increasing the robustness of moving object detection method. Instead of using just only the difference values between the current frame and background, the motion history of moving object over time is considered. Moving object detected at time  $t$  becomes history of its own at time  $t+1$ . Motion history image  $O_t$  of moving object will disappear from time-to-time which is represented by the multiplication by  $h$  ( $h < 1$ ) illustrated by the following equation.

$$O_t = \max(h \times O_{t-\Delta t}, O)$$

By applying the gradient method to motion history image, we can determine the direction of moving object. Then, motion segmentation can be done by classifying its direction. Figure 4 shows the summary diagram for motion history detection and segmentation.

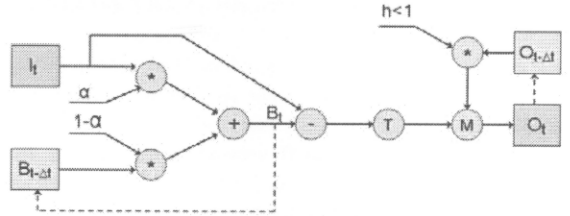


Figure 4. Motion history detection and segmentation.

With the algorithm described above, we obtain position size and direction of moving objects which is necessary for testing our system. However, by defining inappropriate parameters, algorithm will give undesirable results. First remark, with  $h$  close to 1, we got a long history of moving object that finally leads to wrong detection for its size and position, but inversely improve the direction. And when  $h$  close to 0, we got opposite results. Second remark, the method will give result incorrect when moving object is not rigid. In this case, object can be perceived as multiple parts with different directions. For our specific case about pedestrian/vehicle detection, algorithm gives comfortable results, especially when it stays far from camera.

#### 3.2 Motion Event Generator

Event Generator in this circumstance means that events will be raised systematically when motion occurs or disappears. These events are necessary for driving the other components of system such as



encoder and external alert system.

A serious problem about motion detection technique described above is that when object stop moving for a while, which occurs very often in real situation, object is undetectable and then will be immediately recovered. This situation will provide eventually unforeseen event separating two motions of the same object. We encounter the problem simply by introducing an event smoother which is actually implemented by a self-decreasing timer. Motion detection event invoked by detector for indicating stopping motion will not be raised immediately; the generator will wait for some seconds.

#### 4. Video encoding and storage

Video storage component of our surveillance system is designed by default to store images sequence when only having motions on the scene. Encoder starts encoding images after capturing a motion event and stop when new event is raised.

##### 4.1 Encoders

Like the other LPC components, the encoders can be developed (or using an existing one) and interfaced to the principal platform by adding some codes that act as plug-ins.

In our current version, Window Media Encoder SDK is used by default (a license windows can get it directly from Microsoft's website). The most important step for using WME is the description of encoding profiles. In general, the codecs for quality-based encoding or bit rate-based encoding needs to be explicitly specified in the profile. In our case, the Windows Media Audio and Video codecs version 9.0 is used for 100% quality-based encoding. This configuration gives the best video quality with less time consumption of encoding process, but needs more disk space for storage. The choice is made based on the fact that main CPU should be used further in the video analysis process.

##### 4.2 Data Event Generator

Data Event Generator is a HPC utility of our platform to communicate with the external offline searching and playback application. After finishing the encoding process of images sequence of current motion, an event is immediately raised by encoder component that invoke the Data Event Generator. This HPC will send a metadata file containing video file description to the external database server.

The following items are indicated in metadata: name, size and location of video file, starting and ending time of motion, algorithm's name and encoder type. This information will be used by searching and playback application, alert system.

There are two external systems that do not mention here in details are alerts system and searching/playback application. Both of them are very necessary for completing the global need of video surveillance system.

## 6. Performance and discussion

The test of system is performed into two parts: component testing and global system testing. A desktop PC with Intel Core 2 Duo processor cadenced at 1.7 Ghz, 1 Gbyte memory and 100 Mb/s Ethernet card is used for testing. The PC is located at COE department.

### 6.1 Acquisition and viewing

The test is done with some 47 cameras from Axis manufacturer (model 206) locating at PSU Wittayanusorn School, 7 cameras of Bosch Company at PSU entrance gates, 1 Panasonic camera at COE department, 1 D-Link IP-camera (install locally) and 2 USB webcam.

Firstly, the test is performed according to basic functional design, 16 cameras from different manufacturers are installed with 25-30 fps at 640x460 pixels, and then regrouping as view for connecting and displaying in real-time in the same view simultaneously. The system is executed during 48 hours continuously, no error found. The processes are repeated for a couple of works. By considering the maximum number of cameras viewing at the same time, we found that this is limited to around 16 cameras that may implicitly explain by quality of network connection (though internet) and bandwidth limit of testing PC. In fact, one image in jpeg format corresponds by average 74Kbyte, with 25-30 fps of 16 cameras; bandwidth is then bounded from 29.6 to 35.5 Mbyte.

### 6.2 Motion history detection/segmentation

The algorithm of motion detection is implemented in C and used OpenCV library. The test is performed under our system.

Figure 5.1 illustrates that our system is viewing the sequence of images from Panasonic IP-camera located at the first floor under robot building. Figure 5.2 shows some result of motion detection algorithm.

Like the remarks that I did on the section 3.1, the precision of method (for location, size and direction) is very sensitive to the parameter  $h$ . Moreover, method will give wrong result when the moving of non-rigid object is too close to the camera. However, it gives satisfied results if this configuration is determined appropriately.

### 6.2 Overall system testing

Overall system testing is done at first step with only one camera. Image acquisition from Panasonic IP-camera (at the first floor of robot building) is defined at 10 fps of image 640x480 pixels. System is executed continuously during 24 hours. By observing the results, we found that motions of people working in/out from robot building are quite well detected. However, some noises are also registered. We can recognize for two types: contrast/illumination change rapidly due to sunshine and non-static background problem (tree moving because of the winds).



Figure 5.1 Video viewing from camera at robot building

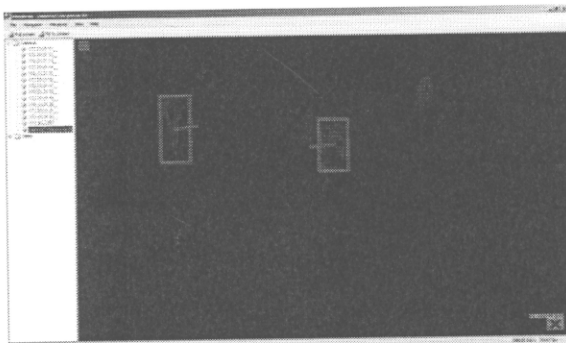
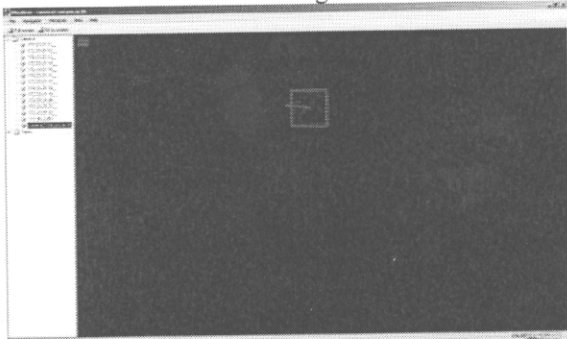


Figure 5.2 motion detection results, red rectangles indicate motions occurring in the area with directions, blue color represents motion history.

Second step, the configuration of testing is setup like before but increasing number of cameras up to 7 (1 Panasonic + 6 Bosch at PSU) and executing for one day. We found that, the acquisition component and video analysis are very well done. However, this test recovers another important problem is that the encoder is mal-functional after a triple hours later. It generates the error "object instance can't be created", and found that more cameras used, faster get the error. We have assumption that maybe there are too much encoder instances that try to write video images onto disk at the same time.

## 7. Conclusion

We proposed in this article a generic version of video surveillance system. System architecture with details is presented. Some preliminary tests of system show encourage results. Nevertheless, more extra debugging and testing has to be performed in order to

improve the robustness that is very important for such system.

Future works, firstly, more complex algorithm of video content analysis for the specific application have to be explored. Secondly, the external alert system must be completed and tested together with this platform, and the same for searching/playback application that use results produced by system. All together the system may become a good choice as SME solution of video surveillance system

## References

### Conference

- [1] R.G.J. Wijnhoven, E.G.T. Jaspers, P.H.N. de With, Flexible Surveillance System Architecture for Prototyping Video Content Analysis Algorithms in Conference on Real-Time Imaging IX, Proceedings of the SPIE, January 2006, San Jose, CA, USA
- [2] B. Abreu, L. Botelho A. and all. Video-based multi-agent traffic surveillance system, Proceedings of the IEEE Intelligent Vehicles Symposium, 2000. Page(s):457 – 462.
- [3] T.Matsuyama and N.Ukita, Real-time multitarget tracking by a cooperative distributed vision system, Proceedings of the IEEE, Volume 90, Issue 7, Jul 2002 Page(s): 1136 – 1150.
- [4] Xiaojing Yuan, Zehang Sun, and Y. Varol, and G. Bebis, A distributed visual surveillance system, Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 21-22 July 2003, page(s): 199- 204.
- [5] B.Georis, X.Desurmont, and all, IP-distributed computer-aided video-surveillance system, Intelligence Distributed Surveillance Systems, IEEE Symposium on (Ref. No. 2003/10062) Volume, Issue , 26 Feb. 2003 Page(s): 18/1 - 18/5
- [6] C. Jaynes and S. Webb and R. Steele and Q. Xiong, An Open Development Environment for Evaluation of Video Surveillance Systems, Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance, 2002.
- [7] J. Piater and J. Crowley, Multi-modal tracking of interacting targets using Gaussian approximations, Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. IEEE Computer Society, Dec. 2001

### Web

- [8] Andrew Kirillov, Camera Vision - video surveillance on C#, <http://www.codeproject.com/KB/audio-video/cameraviewer.aspx>.

# Key Frame Extraction from Video Sequence Using a Face Quality Index Applied to Surveillance System

Sofina Yakhu and Nikom Suvonvorn\*

Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University,  
Hat Yai, Songkhla 90112

E-mail: s4810167@psu.ac.th, kom@coe.psu.ac.th\*

## Abstract

We propose in this paper a method for measuring the quality of face images, detected from video sequence, which is used for key frame extraction. The method is aimed for applying to the domain of video surveillance application. The inspection of covariance matrix values of face images within periods of time is the fundamental concept of our face quality index measurement. The system performance evaluation is done on testing dataset with discussion.

**Keywords:** Key Frame Extraction, Face Detection, Face Tracking, Face Quality Index, Video Surveillance

## 1. Introduction

In the domain of video surveillance system, the key frame extraction is very important issues. Many works have been published [1] [2]. Our goal on this research is inspired by the real problem. It's quite very often that the video obtained from CCTV system cannot identify the person because of the low resolution or low image quality degraded by the compression. Although the current IP-camera using in the market is really advance with very high resolution (1-2M pixels) but in the real world application with a lot of cameras it is limited to the network bandwidth that cannot stream the full quality to the video network recorder. The usual resolution in general is VGA or CIE with MPEG4 or H.264 encoding.

The idea of our research is firstly try to identify the key frame from the low resolution and low quality video sequence streaming from the IP-camera. If there is an interesting event indicating by the key frame, the video network server will ask the camera to send the highest resolution of the current image frame. With this solution, we will obtain the best quality of image at a specific event while reducing network bandwidth. In this paper, we focalize only on the video analysis technique for key frame selection but not the overall system implementation. Our proposed method is emphasized on the face quality index measurement based on the covariance matrix values of detected face images.

The paper is organized as follows: section 2 reviews the face detection and tracking method, section 3 is devoted to our proposed method, section 4 presents the experimentation results with discussion, and the last sections summarized by conclusion.

## 2. Reviewed Techniques

In this section two algorithms for object detection and tracking are presented: the Harr-Like and Camshift algorithm respectively. We apply these methods as the initialization step of our method in the section 3.

### 2.1 Haar-Like Object Detection

The object detection techniques based on Harr-like features are proposed by several authors. In this paper we emphasize on the method of Viola and all [3][5]. The detection framework can be resumed into two important steps. Firstly, with a given over-complete set of Haar-like feature (180,000 rectangle features associated with each images sub-window) and a training set of positive and negative images, the weak AdaBoost learning algorithm is used both for selecting a small set of features and training the classifier. The algorithm is designed to select a single feature which best separates the positive from the negative examples. The classification function  $h_j(x)$  consists of a feature  $f_j(x)$ , a threshold  $\theta_j$  and  $p_j$  indicating the direction of the inequality sign. Here,  $x$  is a 24x24 pixel of an image.

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

Secondly, the classifiers (stages) are constructed into a cascade which is again trained using AdaBoost algorithm by adjusting the threshold to minimize false negatives. With the design, a positive result from the first stage triggers the better evaluation of second one while a negative result will be rejected immediately from the decision. This kind of decision tree increases the detection rate and inversely reduces the computation time.

### 2.2 Camshift Object Tracking

Camshift [4] is an object tracking algorithm derived from mean shift algorithm which is a non-parametric technique. The colored object represented in a probability distribution is tracked in the video sequence by climbing the gradient values in order to find the nearest dominant peak. Camshift tracks the object by updating its probability distribution dynamically which allows better tracking the moving

object having size changing. The algorithm can resume by the following. For each video frame, the raw image is converted to a color probability distribution image via a color histogram model such as HSV space. The center and size of the tracking object are computed from the color probability image as its centroid and area. The current size and location of the tracked object are reported and used to set the size and location of the search window in the next video frame. The process is then repeated for continuous tracking.

### 3. Proposed Method

In this section we describe our proposed method in details how to select the key frame from video sequence using a face quality index. Our technique consist of two steps: face detection and tracking, and face quality index measurement.

#### 3.1 Face Detection and Tracking

The aim of this step is to find the face area in each frame of video sequence that will be used further in the face quality measurement step. We combine two methods based on the work of Viola [3] for face detection and Bradski [4] for face tracking respectively. The face detection technique using Harr-like method achieves high detection accuracy while minimizing the computation time, tested with MIT+CMU frontal face test set. However, in the real situation for the video sequence from CCTV system where the resolution of images are large (normally VGA or CIF) and degraded by the compression, the detection rate decrease and false positive increase. Concerning the Camshift algorithm, the concept of tracking method operates by climbing the gradient of color probability distribution. In real situation, it is often that the background color or its texture is quite similar to the tracking region. As Camshift updates dynamically its probability distribution, this let to miss the tracking area with less accuracy in term of face localization.

We then join these methods in order to combine their main advantages that will increase relatively the detection rate and the robustness. The overall process (figure 1) is performed by the following: for each video frame, face detection by Haar-Like algorithm is applied. If a face is detected, its histogram in HSV model is computed and used as color probability distribution for Camshift initialization or correction. If no faces are detected in a frame, the tracked area obtained from Camshift is automatically assumed to be the missing face. In the correcting phase where the face is detected and the tracked face is found, the correction is done simply by a low-pass filter (average) on its size and position.

Figure 2 show some results from our detection and tracking approach. At the end of this step we obtained a sequence of face images extracted from every frame in the video sequence.

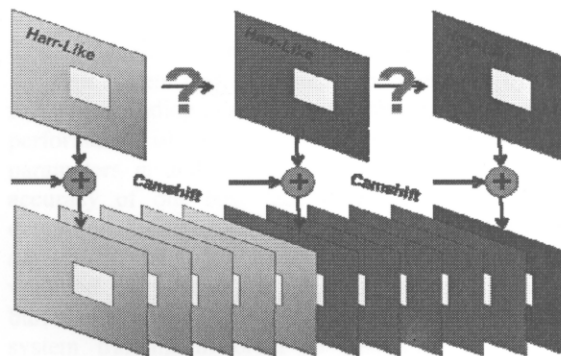


Figure 1 : Harr-Like and Camshift Mixture .



Figure 2. Results of Harr-like and Camshift Mixture

#### 3.2 Face Quality Index for Key Frame Extraction

The face quality measurement is aim for determining the key frame from the video sequence which relate to the best appearance feature of face characteristics. The procedure can be divided into two stages: post-processing and face quality measurement. Figure 3 illustrate the overall process.

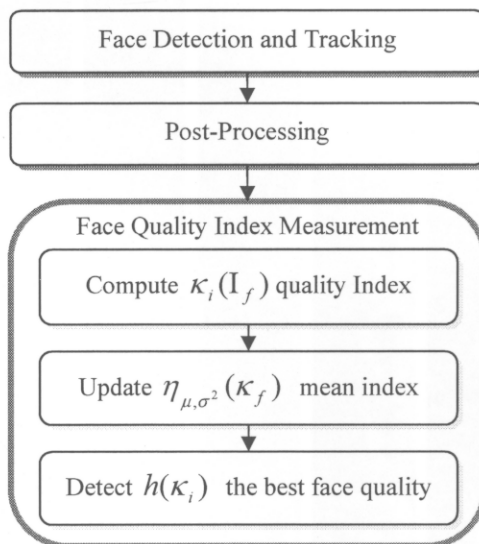


Figure 3. Face Quality Index Measurement Process

In the post-processing, the detected face images from every frame are transformed into the same size (150x150) with bilinear interpolation, and then the basic histogram equalization is applied. This aim to reduce the two noise factors, the number of pixels and the intensity variation effected from light changing in the environment, which will interfere to our face quality measurement method.

After the post-processing, the measurement of face quality can then be computed. Let  $\kappa_i(\mathbf{I}_f)$  is the quality index of face image at frame $_i$ . It is defined by the following equation:

$$\kappa_i(\mathbf{I}_f) = \text{Trace}(\Sigma(\mathbf{I}_f)) + \text{Trace}(\Sigma(\mathbf{I}_f^T))$$

Here  $\Sigma$  is the covariance matrix of face image  $\mathbf{I}_f$ . This function gives a big value when the image is burred and a small one for very noisy image. However, the function indicates us only the tendency of values with respect to the noisy, good, and burred images. There is no the exact value that could be used to evaluate a best quality face. Actually, this value depend on the physical appearance at that time, such as the characteristics of moving object, the contrast changing in the environment, and etc. Intuitively, our method will takes the consideration on these parameters by looking globally the best face image in a short period of time. The procedure of our method is the following: we first compute a reference value and then determine the best quality face to those which have  $\kappa_i(\mathbf{I}_f)$  value around that reference. We introduce the Gaussian model as low-pass filter  $\eta$  for determining the reference value  $\mu$  which needs to be updated continuously.

$$\eta_{\mu,\sigma^2}(\kappa_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\kappa_i-\mu)^2}{2\sigma^2}}$$

Then, the best quality face image can be detected by the following index function:

$$h(\kappa_i) = \begin{cases} 1 & \text{if } \mu_{\kappa_i} + nm\sigma_{\kappa_i} < \kappa_i < \mu_{\kappa_i} + (n+1)m\sigma_{\kappa_i} \\ 0 & \text{otherwise} \end{cases}$$

So, the  $n$  and  $m$  are our model parameter representing the sub-window location referred to the mean value and the sub-window size respectively. These parameters are determined experimentally. We will discuss in the next section. Figure 4 shows the best face quality index  $h(\kappa_i)$ .

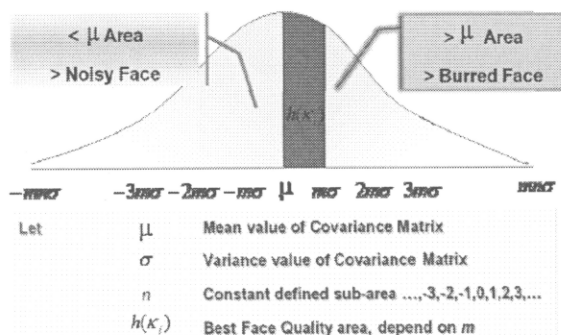


Figure 4. Best Face Quality Index

#### 4. Experimentation and Discussion

Our system was evaluated experimentally by the following: finding  $m$  and  $n$  parameters, and testing the performance of system in the real situation. The parameters  $m$  and  $n$  play an important role to the accuracy of our best face detection method. We determine these two parameters by: firstly a learning dataset of 10 faces (male and female) with its variation (very noisy, noisy, clear, burred and very burred) are created (figure 5), and then it is used for system training in order to determine  $m$  and  $n$  parameters. We found that the optimal parameter  $n$  for the clear face image is 0 and  $m$  corresponds to  $1/15$ . This corresponds to the first sub-window next to the mean value with the larger size equal to  $1/15\sigma$ , meaning that all face images having  $\sigma_i$  greater than  $\mu_{\kappa_i}$  and less than  $\mu_{\kappa_i} + \sigma_{\kappa_i}/15$  will be considered as a face having the best quality.

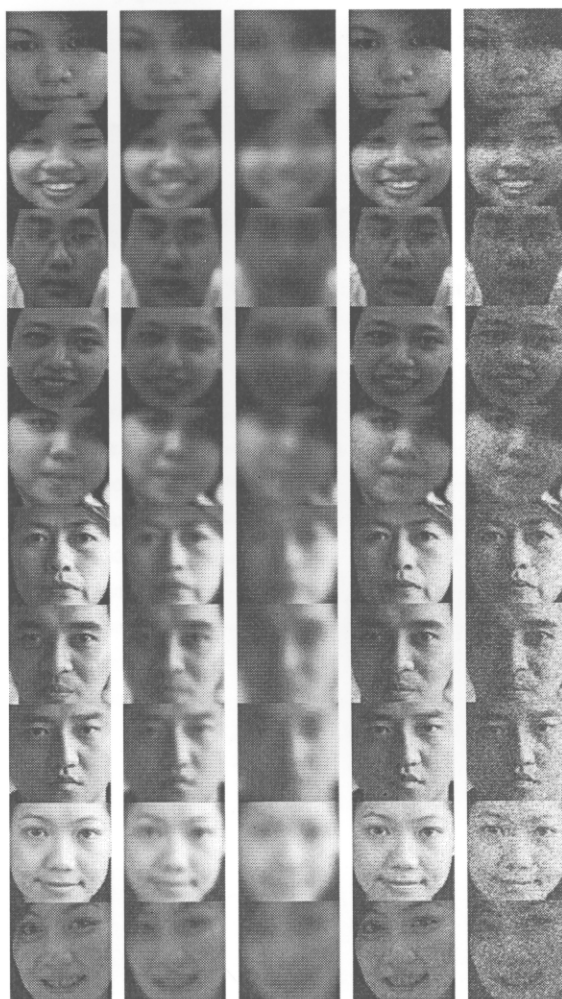


Figure 5. Learning dataset

For testing the performance of our method, we created a testing dataset of four video sequences with one minute each at 30 fps. Each sequence corresponds to a moving person in an indoor environment which

contains the testing characteristics: (1) burred face images from fast moving, or far distance from camera, (2) noisy images with poor light condition, and (3) clear face images when no motion with a good light condition and close to the camera. Figures 6, 7, 8 and 9 illustrate graphs of the face quality index  $\kappa_i(\mathbf{I}_f)$ , its mean  $\eta_{\mu,\sigma^2}(\kappa_i)$ , and the best quality image detected by function  $h(\kappa_i)$ . Visually, we can notice that the quantity of  $\kappa_i(\mathbf{I}_f)$  value for each sequence is difference which relatively depends on the face skin texture and environment. So, the best quality face determined by our method doesn't mean the most clear face image of that person, but it represents the best one at that period of time. We observed experimentally that our method give quite good results of selected face. So, the much burred and very noisy images are rejected completely.



Figure 7. Sequence#2 (a) graph (b,c) simple frames from sequence; and (d) an example of selected face.

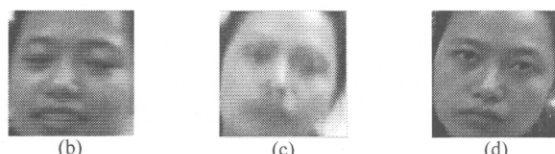
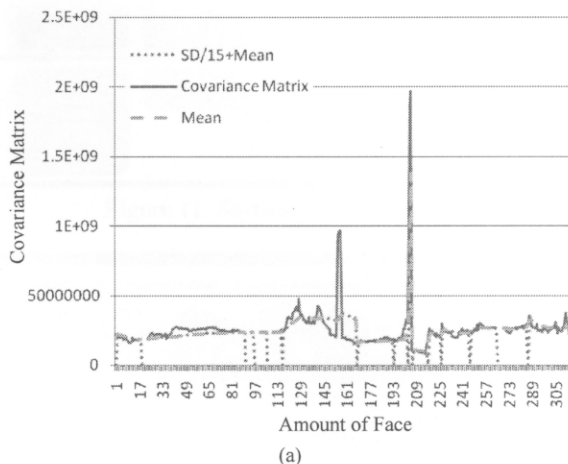


Figure 8. Sequence#3 (a) graph (b,c) simple frames from sequence; and (d) an example of selected face.

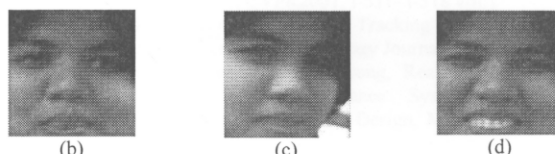
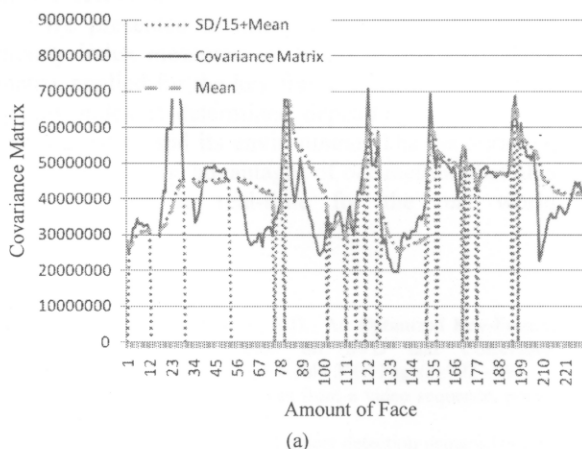


Figure 9. Sequence#4 (a) graph (b,c) simple frames from sequence; and (d) an example of selected face.

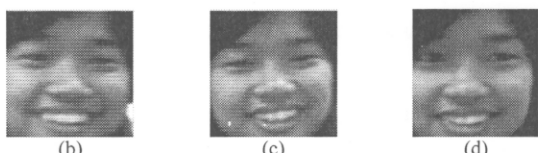
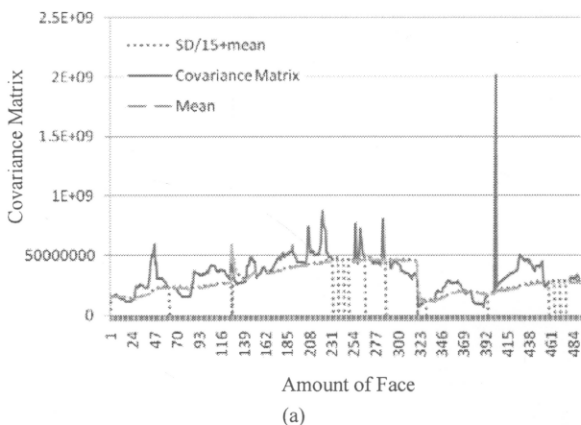
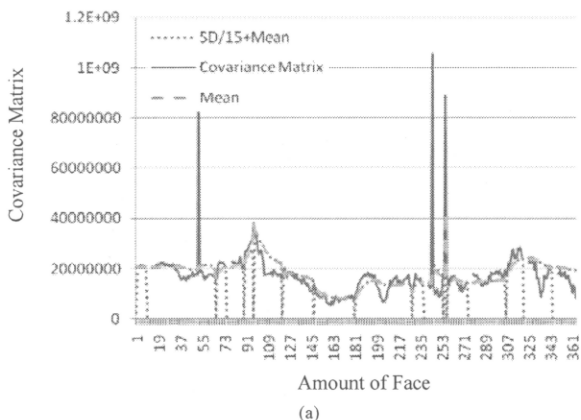


Figure 6. Sequence#1 (a) graph of  $\kappa_i(\mathbf{I}_f)$ ,  $\eta_{\mu,\sigma^2}(\kappa_i)$  and  $h(\kappa_i)$ ; (b,c) simple frames from sequence; and (d) an example of selected face.



### 5. System Integration

Our algorithm is integrated with a video analysis framework proposed in [6][7]. This framework allows our algorithm to analysis image sequences from IP-cameras which is executed on the multi-cores and multi-processors architecture. The framework architecture is divided into five important components: video acquisition, video analysis, video encoding, event alert, and playback. These components lay on three layer architectures: Application Layer (AL), Communication Layer (CL) and Processing Layer (PL), shown in figure 10. Note that Processing Component (PC) means all component located on PL, and respectively for Communication Component (CC). The CL facilitates the communication between components, and also with external sources such as cameras, alert system and database server. It emphasizes how images would be transferred from one to another component. All PC components are developed separately from the framework and acts as plug-ins. We implement our algorithm as a PC component of this framework.

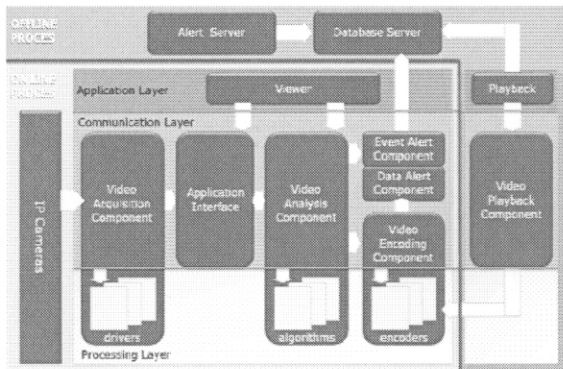


Figure 10. Video Analysis Framework using for our algorithm

The software architecture of framework is entirely implemented under the .NET framework version 2.0. It consists four layers (figure 11): application layer, application core layer, system core layer, and external module layer. The *VsCore* in the system core layer is the most important library that implements how to synchronize between communication components for supporting the three possible types of analysis: local (*VsCamera*), cooperative (*VsChannel*) and global analysis (*VsPage*). The system core also defines how external processing modules can be interfaced and executed during the runtime service. Four types of interfaces are predefined: *VsProvider* for image acquisition, *VsAnalyzer* for any kind of image filters, *VsEncoder* for image encoding into files, and *VsData* for database connection.

For the application core, its role is to define the functionality of a specific application that implements under the system core. The application layer interfaces the application core in different ways to the users which allows, the same function can be called from either standalone or web application. Figure 12 shows the standalone *VsMonitor* app interface that

allows viewing analysing (using our algorithm) and encoding the image sequence from IP-cameras.

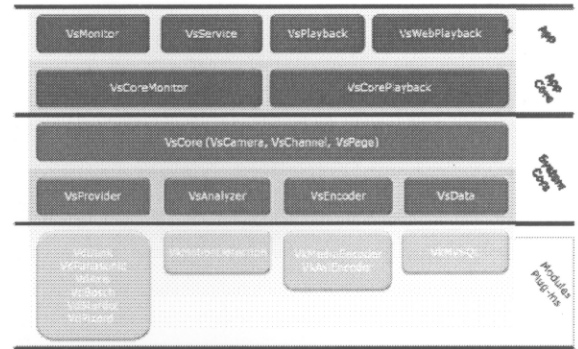


Figure 11. Software architecture



Figure 12. VsMonitor with our algorithm implemented as a PC module

### 6. Conclusion

We presented in this paper a method for face quality index measurement based on the covariance matrix applied for the key frame extraction. The face quality index is determined depending on face skin characteristics and its environment. The encouraging results show some advantages of our method that also could be run in real-time for the real world application.

### References

- [1] Ntalianis, K.S. and Kollias, S.D., An Optimized Key-Frames Extraction Scheme Based on SVD and Correlation Minimization, ICME 2005, 6-6 July 2005 Page(s):792 - 795
- [2] Method of selecting key-frames from a video sequence, patent United States Patent 7184100
- [3] Viola, P. and Jones, M, Rapid object detection using a boosted cascade of simple features, CVPR2001, 1-511- 1-518 vol.1
- [4] Gary R Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal, No.Q2. 98
- [5] N. Suvonvorn and Anant Chocksurivong, Real-Time Face Detection/Identification for Surveillance System, IEEE International Conference on Electronic Design, Malaysia, 1-3 December 2008
- [6] N. Suvonvorn, 2008, A Video Analysis Framework for Surveillance System, IEEE International Workshop on Multimedia Signal Processing, Cairns, Queensland, Australia, 8-10 October 2008.
- [7] N. Suvonvorn, 2008, "A Generic Surveillance System", The Sixth PSU-Engineering Conference, Songkhla, Thailand, 8-9 May 2008