สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

เมาส์ควบคุมด้วยจุดอ้างอิงและเสียง

Accessible Mouse

นาย ปรเมษฐ์ วรสิงห์
นาย ปริญญา กิจวัฒนาบูลย์
นาย ปิติพงษ์ บุญลือ

ปพ.
ปกาม
2549

เลขหมู่.....
เลขทะเบียน 72740
วัน,เดือน,ปี 22 มิ.ย. 2550

b 11772128
i.....

รายงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมาส์ควบคุมด้วยจุดอ้างอิงและเสียง

Accessible Mouse

โดย

ปรเมษฐ์ วรสิงห์

ปริญญา กิจวัฒนาบูลย์

ปิติพงษ์ บุญลือ

อาจารย์ที่ปรึกษา

อ.เกียรติกูล เจียรนัยธนะกิจ

รายงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2549

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง เมาส์ควบคุมด้วยจุดอ้างอิงและเสียง

Accessible Mouse

ผู้จัดทำ

1. นายปรเมษฐ์ วรสิงห์ รหัสประจำตัว 46010410
2. นายปริญญา กิจวัฒนาบูลย์ รหัสประจำตัว 46010425
3. นายปิติพงษ์ บุญลือ รหัสประจำตัว 46010452

อาจารย์ที่ปรึกษา

(อ. เกียรติกูล เจียรนัยธนะกิจ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมาส์ควบคุมด้วยจุดอ้างอิงและเสียง

นายปรเมษฐ์ วรสิงห์ 46010410
นายปริญญา กิจวัฒนาบูลย์ 46010425
นายปิติพงษ์ บุญลือ 46010452
อ.เกียรติกูล เจียรนัยธนะกิจ อาจารย์ที่ปรึกษา
ปีการศึกษา 2549

บทคัดย่อ

ปัจจุบันนี้ คอมพิวเตอร์มีบทบาทและความสำคัญต่อผู้คนมากขึ้น ซึ่งในส่วนการรับข้อมูลของคอมพิวเตอร์นั้น เมาส์(Mouse) เป็นอุปกรณ์ที่สำคัญอย่างยิ่งที่ใช้สำหรับรับข้อมูลจากผู้ใช้มาสู่คอมพิวเตอร์ โครงงานนี้จึงจัดทำขึ้นเพื่อเป็นทางเลือกหนึ่งสำหรับผู้พิการที่ไม่สามารถใช้มือในการควบคุมเมาส์ได้ โดยโครงงานเมาส์ควบคุมด้วยจุดอ้างอิงและเสียงนี้ มีความสามารถในการตรวจจับการเคลื่อนไหวของจุดอ้างอิงที่ติดอยู่บนใบหน้าได้ และตอบสนองโดยการควบคุมการเคลื่อนที่ของเมาส์อย่างถูกต้อง โดยในการทำงานจะต้องมีการประมวลผลภาพที่รวดเร็ว และในการสั่งงานนั้น สามารถสั่งงานให้เมาส์ click , double click , drag and drop ได้โดยการสั่งงานด้วยเสียงพูดผ่าน microphone ซึ่งผู้จัดทำโครงงานหวังว่าโครงงานนี้จะช่วยให้ผู้พิการสามารถใช้คอมพิวเตอร์ได้เหมือนกับบุคคลปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Accessible Mouse

Mr. Poramate Vorasingha
Mr. Parinya Kijwattanaboon
Mr. Pitipong Boonlue
Kietikul Jearanaitanakij Advisor
Academic Year 2549

ABSTRACT

Computer plays an important role over our life. In term of data receiving process, "Mouse" is one of the most important devices to transfer data from user to CPU. Accessible Mouse is an alternative for a computer user who cannot control mouse by hand. Accessible Mouse can be able to detect the color point movement and control the position of the mouse pointer. The software needs to perform an image processing within a rigid time period to determine a color point position and then controls the mouse to do corresponding tasks such as Click , Double Click , Drag and Drop. In addition, a human voice is detected through a microphone to accommodate the mouse control. The inspiration of this project is to equalize the working ability between normal and disable by inventing a device to reduce the gap of difference.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้คงไม่อาจเสร็จได้ด้วยดี หากไม่ได้รับความช่วยเหลือ และความร่วมมือจากหลายๆฝ่ายด้วยกัน บุคคลแรกที่ต้องกล่าวถึงเพราะเป็นส่วนสำคัญที่ทำให้วิทยานิพนธ์นี้เสร็จลงได้ ก็คือ อาจารย์เกียรติกูล เจียรนัยธนะกิจ อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ให้ความเอาใจใส่ แนะนำ และช่วยเหลือเสมอมา ซึ่งต้องขอขอบพระคุณเป็นอย่างมาก

และต้องขอขอบพระคุณบุคคลสำคัญที่สุดที่ทำให้ข้าพเจ้ามีวันนี้ ก็คือ บิดามารดา ผู้ที่คอยให้กำลังใจ เลี้ยงดู และมอบโอกาสที่ดีทางการศึกษามาโดยตลอด ทำให้ข้าพเจ้าได้ทราบถึงองค์ความรู้ต่างๆ ซึ่งหลายคนอาจมิได้โอกาสเช่นนี้ และต้องขอขอบพระคุณอาจารย์ท่านอื่นที่ไม่ได้ออกนาม ที่อยู่ ณ ภาควิศวกรรมคอมพิวเตอร์นี้ ทุกท่านไม่เคยที่จะบ่น คอยแต่ให้คำปรึกษาที่ดีและเป็นประโยชน์ สุดท้ายนี้ ข้าพเจ้าขอระลึกในพระคุณอันสุดประมาณและขอกราบขอบพระคุณมา ณ ที่นี้

ปรเมษฐ์ วรสิงห์
ปริญญา กิจวัฒนาบูลย์
ปิติพงษ์ บุญลือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้าที่
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญภาพ	VI
สารบัญตาราง	VIII
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มา	1
1.2 วัตถุประสงค์ของงานวิจัย	1
1.3 ผลที่คาดว่าจะได้รับ	1
1.4 ขอบเขตของงานวิจัย	2
1.5 คุณลักษณะของอุปกรณ์ที่ใช้กับเมาส์ควบคุมด้วยจุดอ้างอิงและเสียง	2
1.6 วิธีการดำเนินการ	2
บทที่ 2 การประมวลผลทางรูปภาพ	3
2.1 การทำ Thresholding	3
2.1.1 Global Thresholding	3
2.1.2 Semi Thresholding	5
2.1.3 Multilevel Thresholding	6
2.1.4 Variable Thresholding	7
2.2 การแยกข้อมูลภาพออกเป็นส่วนๆ (Image Segmentation)	7
2.2.1 Region segmentation methods	7
บทที่ 3 การเขียนโปรแกรมติดต่อกับกล้อง web cam	11
3.1 ActiveX Control (VideoOCX และ VideoOCXTools) คือ	11
3.2 Method ที่สำคัญ	11
บทที่ 4 การเขียนโปรแกรมติดต่อระหว่างแอปพลิเคชั่นกับวินโดวส์โดยใช้ API	14
4.1 API คือ	14
4.2 การเรียกใช้ Windows API	16
4.2.1 โดยการประกาศ Declare	16
4.2.2 การติดต่อ API โดยใช้ DllImport	18
4.3 กลุ่มฟังก์ชันของ Windows API	18
บทที่ 5 หลักการรู้จำเสียงโดย Hidden Markov Model	19
5.1 Hidden Markov Model คือ	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้าที่
5.2 หลักการรู้จำเสียง	21
5.3 การ Training	24
5.3.1 หลักการสร้างโมเดลเสียง (Acoustic model)	24
5.3.2 หลักการสร้างโมเดลภาษา (Language model)	29
5.3.3 การรู้จำเสียง	29
บทที่ 6 แนวคิดการออกแบบ	30
6.1 Functional requirements	30
6.2 Non-functional requirements	30
6.3 อัลกอริทึม 1	34
6.4 อัลกอริทึม 2	35
6.5 การใช้เสียงควบคุมโปรแกรม	36
6.6 คำสั่งควบคุมการทำงานโปรแกรม	37
บทที่ 7 ผลการทดลอง	38
7.1 การทดลองการควบคุมเมาส์	38
7.2 ทดสอบอัตราความเร็วของเมาส์	40
7.3 การทดสอบการลากและปล่อยของเมาส์	41
7.4 การทดสอบการควบคุมเมาส์	42
7.5 การออกแบบการทดสอบการควบคุมเมาส์ระดับ 2	44
7.6 สรุปผลการทดลอง	45
7.7 ข้อเสนอแนะ	45
บรรณานุกรม	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้าที่

รูปที่

รูปที่ 2-1 Flow chart แสดงการหา Optimum Threshold Estimation 4
รูปที่ 2-2 ภาพต้นฉบับก่อนการทำ Threshold 5
รูปที่ 2-3 ภาพที่ผ่านการทำ Threshold 6
รูปที่ 2-4 ภาพที่ผ่านการทำ Semi thresholding 6
รูปที่ 2-5 แสดงลักษณะของ Split region 10
รูปที่ 4-1 แสดงหลักการทำงานของ Windows API 15
รูปที่ 5-1 แสดง Automaton ความน่าจะเป็นของสภาพอากาศ 20
รูปที่ 5-2 การดึง Speech feature 22
รูปที่ 5-3 หน้าที่ของ Acoustic model 22
รูปที่ 5-4 หน้าที่ของ Language model 23
รูปที่ 5-5 Word network สำหรับการรู้จำ 24
รูปที่ 5-6 Left-to-right HMM 24
รูปที่ 5-7 แสดง Prototype model 25
รูปที่ 5-8 แสดง (a) Monophone model (b) Triphone model 27
รูปที่ 5-9 แสดง Tied-State model 28
รูปที่ 5-10 แสดง dgs.gram 29
รูปที่ 6-1 การออกแบบ High-level 31
รูปที่ 6-2 Use case diagram 31
รูปที่ 6-3 แสดง Sequence diagram ของการ click left 32
รูปที่ 6-4 แสดง Sequence diagram ของการ click right 32
รูปที่ 6-5 แสดง Sequence diagram ของการ double click 33
รูปที่ 6-6 แสดง Sequence diagram ของการ drag and drop mouse 33
รูปที่ 6-7 แสดง Sequence diagram ของ mouse move 34
รูปที่ 6-8 แสดงการค้นหาจุดอ้างอิง 35
รูปที่ 6-9 แสดงหน้าต่างแสดงกรอบจอยสติก 36
รูปที่ 6-10 แสดงหน้าต่าง Pop up ของกรอบจอยสติก 36
รูปที่ 7-1 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งซ้ายสุดของหน้าต่าง 38
รูปที่ 7-2 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งขวาสุดของหน้าต่าง 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ(ต่อ)

หน้าที่

รูปที่

รูปที่ 7-3 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งบนสุดของหน้าต่าง 39
รูปที่ 7-4 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งล่างสุดของหน้าต่าง 40
รูปที่ 7-5 หน้าต่างการทดสอบอัตราความเร็วของเมาส์ 41
รูปที่ 7-6 หน้าต่างการทดสอบการลากและปล่อยเมาส์ 42
รูปที่ 7-7 หน้าต่างการทดสอบการควบคุมเมาส์ระดับ 1 43
รูปที่ 7-8 หน้าต่างการทดสอบการควบคุมเมาส์ระดับ 2 44

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้าที่

ตารางที่

ตารางที่ 5-1 แสดงความน่าจะเป็นของสภาพอากาศวันพรุ่งนี้
อ้างอิงจากสภาพอากาศวันนี้ 20

ตารางที่ 5-2 แสดงค่าความน่าจะเป็นของสภาพอากาศ
เมื่อเห็นผู้ดูแลนำร่มมาด้วย 21

ตารางที่ 7-1 การเปรียบเทียบอัตราความเร็วของเมาส์ 41

ตารางที่ 7-2 การเปรียบเทียบอัตราการลากและปล่อยของเมาส์ 42

ตารางที่ 7-3 การเปรียบเทียบอัตราการควบคุมเมาส์ระดับ 1 43

ตารางที่ 7-4 การเปรียบเทียบอัตราการควบคุมเมาส์ระดับ 2 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มา

ในปัจจุบันการดำเนินชีวิตของผู้คนแทบทุกคนนั้น จะมีการนำเทคโนโลยีเข้ามาเกี่ยวข้องแทบทั้งสิ้น ไม่ว่าจะใช้ในการทำงาน การศึกษา การติดต่อสื่อสาร และอื่นๆอีกมากมาย ซึ่งในจำนวนผู้ใช้ที่มีอยู่มากมายหลายประเภทนั้น ก็รวมถึงผู้ที่พิการ ไม่สามารถใช้เมาส์ได้สะดวกและเกิดความลำบากอย่างยิ่งในการใช้งานคอมพิวเตอร์

ด้วยเหตุนี้จึงได้มีความคิด และเป็นแรงบันดาลใจในการพัฒนาโครงงานนี้ขึ้นมา ซึ่งมีชื่อว่า เมาส์ควบคุมด้วยจุดอ้างอิงและเสียง (Accessible Mouse) เพื่อช่วยเหลือผู้ที่พิการที่มีปัญหาเกี่ยวกับการใช้แขนในการบังคับเมาส์ เช่น ผู้ที่พิการแขนขาด เป็นต้น โดยในการทำงานนั้น จะมีจุดอ้างอิงติดอยู่บนใบหน้าผู้ใช้ และมีกล้อง Web Cam คอยจับการเคลื่อนไหวของจุดอ้างอิงที่ติดอยู่บนใบหน้าผู้ใช้ และนำภาพที่ได้มาประมวลผล เมื่อผู้ใช้ทำการขยับจุดอ้างอิงไปยังตำแหน่งต่างๆ ก็จะมีการคำนวณเพื่อขยับลูกศรบนหน้าจอให้ตรงกับตำแหน่งต่างๆ และในการสั่งงาน click , double click , drag and drop ก็จะใช้การสั่งงานด้วยเสียงพูดผ่านไมโครโฟน

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อสร้างเมาส์ควบคุมด้วยจุดอ้างอิงและเสียง สำหรับผู้พิการหรือบุคคลที่ไม่สามารถใช้มือในการควบคุมเมาส์ได้
2. เพื่อศึกษา API ที่ใช้ในการช่วยเขียนโปรแกรมในการติดต่อกับกล้อง web cam
3. เพื่อศึกษา API ที่ใช้ในการช่วยเขียนโปรแกรมจดจำเสียง
4. เพื่อศึกษาการเขียนโปรแกรมด้วย VB.net

1.3 ผลที่คาดว่าจะได้รับ

1. ได้รับความรู้เกี่ยวกับการเขียนโปรแกรมติดต่อกับกล้อง
2. ได้รับความรู้ความสามารถในการเขียนโปรแกรมจดจำเสียง
3. ได้รับความรู้ ความสามารถในการเขียนโปรแกรมประยุกต์บนระบบปฏิบัติการ window
4. โปรแกรมสามารถควบคุมเมาส์ด้วยจุดอ้างอิง และสั่งงานโดยการใช้เสียงพูดผ่านไมโครโฟนได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ขอบเขตของงานวิจัย

1. โปรแกรมเมาส์ควบคุมด้วยจุดอ้างอิงและเสียง สามารถตรวจจับการเคลื่อนไหวของจุดอ้างอิงที่ติดอยู่บนใบหน้าของผู้ใช้ และบังคับตำแหน่งของเมาส์ที่อยู่บนหน้าจอ ให้สอดคล้องกับการเคลื่อนไหวของจุดอ้างอิงได้

2. โปรแกรมสามารถสั่งงานให้เมาส์ click , double click , drag and drop ได้ โดยจะทำการสั่งงานด้วยเสียงพูดผ่านไมโครโฟน

3. โปรแกรมสามารถใช้กับจุดอ้างอิงได้ โดยไม่จำกัดขนาดจุดอ้างอิง

4. โปรแกรมมีการตอบสนองต่อการทำงานที่รวดเร็ว แม่นยำ

1.5 คุณลักษณะของอุปกรณ์ที่ใช้กับเมาส์ควบคุมด้วยจุดอ้างอิงและเสียง

1. กล้อง web cam
2. ไมโครโฟน

1.6 วิธีการดำเนินการ

งานวิจัยในโครงงานนี้จะเริ่มด้วยการศึกษาทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้อง และจำเป็นต้องใช้ในงานวิจัย ซึ่งหลักๆ ก็มีดังนี้ ศึกษาในเรื่องของการเขียนโปรแกรมติดต่อกับกล้อง web cam โดยใช้ภาษา VB.net และใช้ActiveX Control ที่มีชื่อว่า VideoOCX , การเขียนโปรแกรมติดต่อระหว่างแอปพลิเคชัน กับ Window โดยใช้ API

เมื่อได้ข้อมูลต่างๆ ที่ศึกษามาทั้งหมดแล้ว ก็จะนำมาทำการออกแบบ วางแผนการพัฒนาโครงงาน โดยเริ่มจาก แนวคิดและการออกแบบ อัลกอริทึมที่ใช้ในการตรวจจับหาจุดอ้างอิงที่ติดอยู่บนใบหน้า การใช้งานเมาส์ด้วยเสียงจากการพูดผ่านไมโครโฟน และเมื่อทำได้ทุกขั้นตอนแล้ว ก็จะเข้าสู่การทำสอบระบบ และสุดท้ายเป็นการสรุปผลการทดลองและข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
การประมวลผลทางรูปภาพ

2.1 การทำ Thresholding

Thresholdมีอยู่หลายแบบที่มีการใช้งานในเกี่ยวกับงานด้านคอมพิวเตอร์วิชั่นอย่างกว้างขวาง เช่น การทำเซกเมนเตชั่นกับภาพเพื่อหาส่วนที่ต้องการจากภาพหรือกำจัดส่วนที่ไม่ต้องการออกไปตัวอย่างการใช้งานที่ง่ายที่สุด คือ Thresholdแบบเดี่ยวที่ให้ผลลัพท์ที่น่าพอใจกับหลายๆรูปภาพนอกจากนี้ยังมีThresholdแบบที่เป็นแบบตัวแปรหรือ รูปแบบที่มีหลายระดับและยังใช้สถิติมาช่วยใช้ร่วมด้วย

2.1.1 Global Thresholding

เป็นกระบวนการที่ใช้ในการแยกวัตถุที่มีความแตกต่างจากพื้นหลังของภาพ การจำแนกแต่ละกลุ่มของพิกเซลว่าเป็นวัตถุเดียวกันหรือว่าเป็น ภาพพื้นหลัง สามารถกำหนดค่าให้กับภาพต้นฉบับได้ถ้าอยู่ในช่วงที่เรากำหนด

การทำThresholdทั่วไป จะเป็นแบบตรงไปตรงมา กำหนดให้ $a \in R^x$ ของภาพต้นฉบับและให้ช่วงที่ต้องการทำThresholdเป็น [h,k] Thresholdของภาพ b {0,1}x ให้

$$b(x) = \begin{cases} 1 & \text{ถ้า } h \le a \le k \\ 0 & \text{อื่น} \end{cases}$$

ทุกค่าของ x มีสองกรณีที่ต้องให้ความสนใจคือ ค่าของภาพ b มีค่าสูงเพียงอย่างเดียวหรือมีค่าต่ำเพียงอย่างเดียว ในกรณีแรกของ Threshold ภาพ b จะให้

$$b(x) = \begin{cases} 1 & \text{ถ้า } a \ge k \\ 0 & \text{อื่น} \end{cases}$$

กรณีที่สอง

$$b(x) = \begin{cases} 1 & \text{ถ้า } a(x) \le k \\ 0 & \text{อื่น} \end{cases}$$

ที่ค่า k แทนค่าThresholdที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมการพีชคณิตของภาพ

กำหนดให้ $a \in R^{X}$ เป็นภาพต้นฉบับ และ [h,k] เป็นช่วงของค่าThreshold ผลลัพท์ของ Thresholdคือภาพ b {0,1}X คำนวณได้จาก

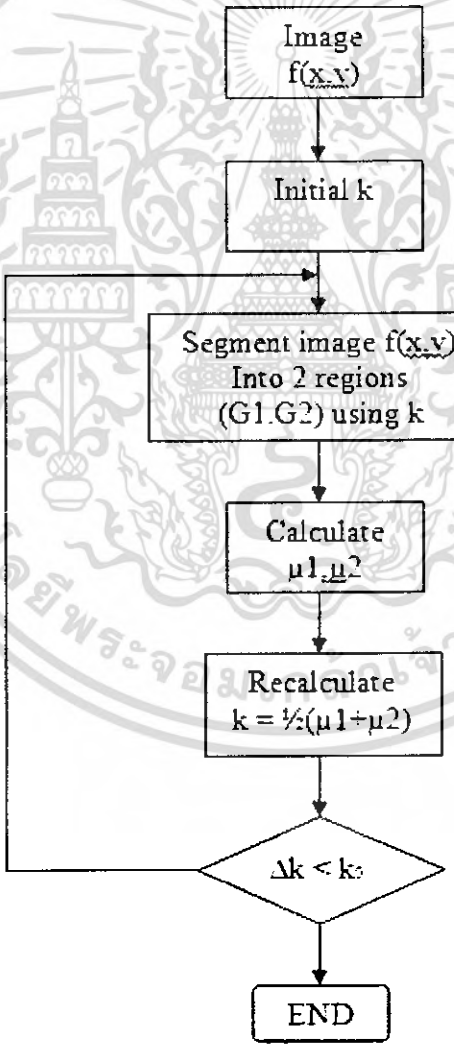
$$b := x[h,k]^{(\bullet)}$$

สมการที่มีลักษณะพิเศษ

$$b := k \leq x^{(\bullet)}$$

และ

$$b := k \geq x^{(\bullet)}$$



รูปที่ 2-1 Flow chart แสดงการหา Optimum Threshold Estimation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 Semi Thresholding

Semi Thresholding นำประโยชน์ของGlobal Thresholdingที่มีการเปลี่ยนแปลงมาใช้งาน ค่าของแต่ละพิกเซลจะอยู่ภายใน ช่วงของค่าThresholdจะเก็บค่าพิกเซลเดิมไว้ หรือถ้าค่าของพิเซลไม่อยู่ในช่วงของThresholdจะถูกเปลี่ยนค่าให้เป็น "0" ให้ภาพต้นฉบับเป็น $a \in R^X$ ช่วงของThresholdเป็น [h,k] ภาพที่เป็นเซมิเทรดโฮลดิ่ง คือ $b \in R^X$ โดยให้

$$b(x) = \begin{cases} a(x) & \text{ถ้า } h \le a(x) \le k \\ 0 & \text{อื่น} \end{cases}$$

สำหรับทุกค่าของ x
สำหรับพื้นที่ที่มีค่าพิกเซลค่าสูงแยกเป็น

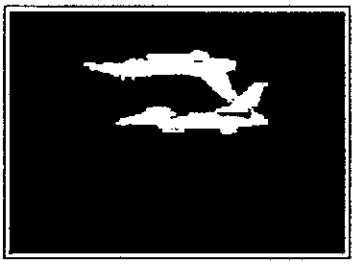
$$b(x) = \begin{cases} a(x) & \text{ถ้า } a \ge k \\ 0 & \text{อื่น} \end{cases}$$

สำหรับพื้นที่ที่มีค่าพิกเซลค่าต่ำแยกเป็น

$$b(x) = \begin{cases} a(x) & \text{ถ้า } a \le k \\ 0 & \text{อื่น} \end{cases}$$

รูปที่ 2-2 ภาพต้นฉบับก่อนการทำ Threshold

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2-3 ภาพที่ผ่านการทำThreshold

รูปที่ 2-4 ภาพที่ผ่านการทำ Semi thresholding

2.1.3 Multilevel Thresholding

จากวิธีการของThresholdingที่กล่าวมาทั้งสองวิธีนั้น การทำเซกเมนต์กับภาพจะทำให้ได้ข้อมูลที่มีขอบเขตหรือว่าพื้นที่นั้นจะถูกแบ่งออกเป็นสองส่วน การทำงานที่เกี่ยวกับภาพนั้นก็มีความจำเป็นที่จะต้องแบ่งเซกเมนต์ภาพออกมากกว่าสองส่วน มัลติลิเวลThresholdจะทำให้เราแบ่งภาพหรือจัดกลุ่มได้มากกว่า

ยกตัวอย่างเช่นhistogramของภาพที่มีจุดสูงสุดสามจุด สามารถแบ่งจุดเพื่อทำ Threshold ได้ 2 จุด โดยช่วงที่แบ่งไม่ทับซ้อนซึ่งกันและกัน กำหนดให้ $a \in R^x$ และ k1,..., kn เป็นค่าThreshold ที่ k1 > k2>k3>...>kn แบ่ง R ออกเป็น n+1 ที่สัมพันธ์กับค่าผลลัพธ์ของ v1,v2,v3,...,vn

โดยทั่วไปค่าที่ได้จะเป็น $1, \frac{n-1}{n}, \ldots, \frac{1}{n}, 0$ Thresholdของภาพ $b \in R^x$ กำหนดโดย

$$b(x) = \begin{cases} v_1 & \text{ถ้า } k < a(x) \\ v_i & \text{ถ้า } k < a(x) \le k_{i-1} \\ v_{n+1} & \text{ถ้า } a(x) \le k_n \end{cases}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4 Variable Thresholding

การทำThresholdที่เป็นแบบเดี่ยวยังทำให้ผลลัพท์ที่ทำกับภาพนั้นยังไม่ค่อยดี Variable Thresholding จะทำให้ ผลการทำThresholdกับภาพนั้นมีมีความแตกต่างกันหลายระดับขึ้น

ภาพที่มีระดับความเข้มของวัตถุ และแบล็คกราวด์สูงนั้น แต่มีความสว่างของภาพไม่สม่ำเสมอกันนั้น ภาพวัตถุและแบล็คกราววด์นั้นจะมีค่าต่ำที่อีกด้านหนึ่งของภาพมากกว่าอีกข้าง ตัวอย่างเช่น ภาพอาจจะถูกแบ่งเป็นพื้นที่เล็กๆย่อยลงไปอีก ความแน่นอนของการทำ Thresholdขณะนั้นทั้งขอบเขตและเป้าหมายของภาพที่ต้องการที่เป็นพื้นที่ย่อยนั้นรวมกันเป็นสิ่งเดียวกันได้

กำหนดให้ $a \in R^x$ เป็นภาพต้นฉบับ และภาพ $d \in R^x$ แสดงถึงการทำThresholdที่ค่า X ต่าง d(x) คือค่าที่ได้จากการทำThresholdที่ X ต่าง การทำThresholdภาพ b {0,1}X กำหนดให้

$$b(x) = \begin{cases} 1 & \text{ถ้า } a(X) \geq d(X) \\ 0 & \text{ถ้า } a(X) < d(X) \end{cases}$$

2.2 การแยกข้อมูลภาพออกเป็นส่วน ๆ (Image Segmentation)

Image segmentation ใช้ในการแยกองค์ประกอบต่างๆของรูปภาพออกจากกันตามลักษณะสำคัญที่เราพิจารณา ขั้นตอนนี้เป็นขั้นตอนสำคัญของการวิเคราะห์ภาพ
ประโยชน์

1) ลดจำนวนข้อมูลในรูปภาพที่ไม่จำเป็นในการวิเคราะห์ลง
2) จัดระเบียบข้อมูลในรูปภาพให้เป็นกลุ่มได้ดีขึ้น
3) แสดงข้อมูลในรูปที่เข้าใจง่ายขึ้น

2.2.1 Region segmentation methods

เป็นวิธีการแยกองค์ประกอบของภาพโดยดูจากตำแหน่งของพิกเซลและความเหมือนกันของคุณสมบัติของพิกเซลภายในพื้นที่ โดยถ้าพิกเซลที่อยู่ติดกันและมีคุณสมบัติเหมือนกันจะถูกจัดให้เข้ากลุ่มเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) Region Growing

วิธีการ Region Growing จะนำพิกเซลข้างเคียงมาพิจารณาซึ่งจะทำการจัดกลุ่ม (Region) ของพิกเซลเหล่านี้เข้าไว้ด้วยกันโดยการพิจารณาถึงความเข้มของพิกเซล ในการ Segmentation จะต้องมีการกำหนดกลุ่ม (Region) ที่ต้องการจะแบ่ง R_i โดยที่แต่ละกลุ่มจะต้องมีการกำหนดค่าความเข้มของพิกเซลเริ่มต้น $s_i, i = 1,\ldots, N$ (ค่าเหล่านี้จะถูกเลือกไว้โดยผู้ที่ต้องการ segmentation: supervised mode) ซึ่งมีไว้เพื่อใช้สำหรับการขยายตัวของกลุ่ม (growth)

ในการขยายตัวของกลุ่มนี้จำเป็นจะต้องมีกฎเพื่อใช้เป็นวิธีการขยายตัวของกลุ่มรวมทั้งกฎของการตรวจสอบความเป็นอันหนึ่งอันเดียวกันของกลุ่ม (homogeneity) ของทุกระยะของการขยายตัว

ในแต่ละระยะของการขยายตัวของกลุ่ม $R_i^{(K)}, i = 1,\ldots, N$ จะมีการตรวจสอบว่ามีพิกเซลที่ยังไม่ได้จัดกลุ่มหรือไม่ (เป็นพิกเซลทั้ง 8 ที่เชื่อมต่อ (8-neighbourbood) อยู่กับพิกเซลที่อยู่บริเวณขอบของกลุ่ม) และก่อนที่จะมีการกำหนดพิกเซลใด ๆ (x) เข้าสู่กลุ่มใดกลุ่มหนึ่งจะต้องมีการตรวจสอบอีกว่ากลุ่มที่จะขยายนั้นยังคงมีความเป็นอันหนึ่งอันเดียวกันอีกหรือไม่ (homogeneity)

$$P(R_i^{(k} \cup \{x\}) = \text{TRUE}$$

ประสิทธิภาพของการ Segmentation วิธีนี้จะขึ้นอยู่กับการเลือกค่าความเข้มของพิกเซลเริ่มต้นของในแต่ละกลุ่มซึ่งกำหนดให้โดยผู้ใช้ ซึ่งกระบวนการนี้สามารถหาได้อัตโนมัติจาก histogram ของภาพโดยการพิจารณาค่าสูงสุดของ histogram (ค่า peak) มาใช้เป็นค่าความเข้มของพิกเซลเริ่มต้นของกลุ่มได้ ซึ่งโดยปกติแล้วจะมีค่าความเข้มเริ่มต้นมากกว่าหนึ่งค่าต่อหนึ่งกลุ่ม

ดังนั้นจึงต้องมีวิธีการ merging เพื่อที่จะใช้ในการรวมกลุ่มที่มีลักษณะทาง statistical (พิจารณาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน) ใกล้เคียงกัน

$$m_i = \frac{1}{n} \sum_{(k,l) \in R_i} f(k,l)$$

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{(k,l) \in R_i} (f(k,l) - m_i)^2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ m_i เป็นค่าเฉลี่ยของค่าความเข้มในกลุ่มที่ i ที่มีจำนวนพิกเซลเท่ากับ n พิกเซลค่าเหล่านี้จะใช้ในการจะนำไปใช้ในการติดสินใจว่าจะสามารถรวมกลุ่ม R_1, R_2 เข้าไว้ด้วยกันได้หรือไม่ ถ้าหากค่าเฉลี่ยของกลุ่ม m_1, m_2 มีค่าใกล้เคียงกันก็สามารถรวมกันได้ดังนี้คือ

2) Split and Merge

2.1) Merge region เป็นวิธีการ segmentation วิธีหนึ่งซึ่งจะมีอัลกอลิทึมดังนี้คือถ้าสมมุติว่ามีการสแกนไปยังพิกเซลต่าง ๆภายในภาพจากพิกเซลแรกไปยังพิกเซลสุดท้ายของภาพไปตามแถวและหลักตามลำดับ ในช่วงระหว่างการสแกนก็จะมีการกำหนดพิกเซลนั้นไปยังกลุ่มต่าง ๆ

สมมุติปัจจุบันอยู่ที่พิกเซล (k,l) ดังนั้นพิกเซลแรกจนถึงพิกเซล $(k-1,l)$ ได้ถูกกำหนดให้อยู่ในกลุ่มต่าง ๆ หมดแล้ว ดังนั้นพิกเซลที่ (k,l) จึงเปรียบเสมือนเป็นอีกกลุ่มหนึ่งที่พยายามที่จะรวมเข้ากับกลุ่มที่มีอยู่ก่อนหน้านั้น (กลุ่มที่มีทั้งหมด R_i แต่จะเลือกเอาเฉพาะกลุ่มที่อยู่ข้างเคียงเท่านั้น ได้แก่กลุ่มที่มีพิกเซล ณ ตำแหน่ง $(k-1,l)$, $(k+1,l)$, $(k,l-1)$ และ $(k,l+1)$ เป็นสมาชิกอยู่) หากพบว่าไม่สามารถทำการรวมเข้ากับกลุ่มใดได้ก็ให้สร้างกลุ่มใหม่ขึ้นมา ประสิทธิภาพของวิธีนี้จะขึ้นอยู่กับกฎของการรวมกลุ่ม ($P(R_i \cup (k,l))$) ของพิกเซล (k,l) เข้ากับกลุ่ม R_i กฎของการรวมจะขึ้นอยู่กับค่าเฉลี่ยของกลุ่ม m_i และค่าเบี่ยงเบนมาตรฐาน s_i แต่สำหรับการรวมกันของ $R \cup (k,l)$ จะมีการเปลี่ยนแปลงการหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานเป็นดังนี้คือ

$$m_i' = \frac{1}{n+1}(f(k,l) + nm_i)$$

$$\sigma_i' = \sqrt{\frac{1}{n+1}\left(n\sigma_i^2 + \frac{n}{n+1}[f(k,l) - m_i]^2\right)}$$

การรวมกันสามารถทำได้เมื่อค่าความเข้มของพิกเซล $f(k,l)$ มีค่าใกล้เคียงกับค่าเฉลี่ย m_i' ดังต่อไปนี้คือ

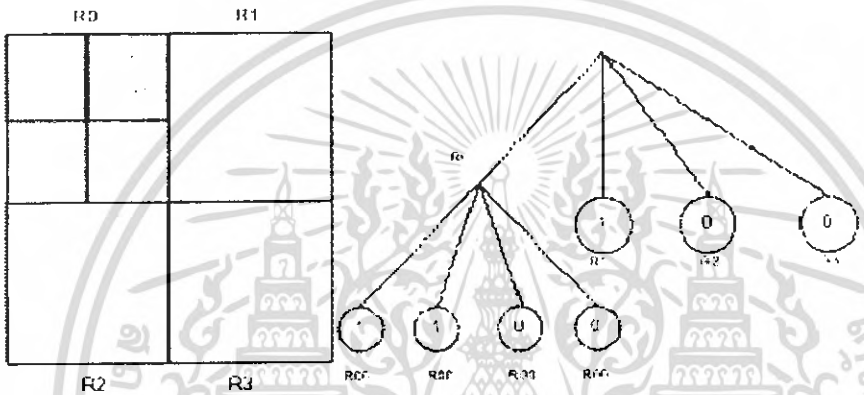
$$|f(k,l) - m_i| <= T_i(k,l)$$

เมื่อ T_i เป็นค่า Threshold ที่มีค่าเปลี่ยนแปลงอยู่ระหว่างกลุ่ม R_i กับความเข้มของพิกเซล $f(k,l)$ ซึ่งสามารถหาได้จาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$T_i(k,l) = \left(1 - \frac{\sigma_i}{m_i}\right)T$$

2.2) Split region เป็นลักษณะของการ segmentation อีกวิธีหนึ่งที่มีลักษณะตรงข้ามกับ Merge region (เป็นลักษณะ Top-down Approach) โดยเริ่มต้นจะมีการสมมุติว่าทั้งภาพจะมีเพียงหนึ่งกลุ่มเท่านั้น (Region) โดยถ้าหากว่าไม่เป็นความจริงก็ให้ทำการแยกกลุ่มนี้ออกเป็นสี่กลุ่มย่อย และจะมีการพิจารณาลักษณะนี้เรื่อย ๆ จนกระทั่งได้กลุ่มของภาพที่มีสมาชิกของกลุ่มที่มีค่าใกล้เคียงกันในระดับที่สามารถยอมรับได้ (homogeneous)



(ก) (ข)

รูปที่ 2-5 แสดงลักษณะของ Split region

(ก) Original Image และ (ข) Quadtree representation

การตรวจสอบว่ากลุ่มใดสามารถยอมรับได้หรือไม่ทำได้โดยการคำนวณผลต่างของค่าความเข้มของพิกเซลที่ได้จากค่าความเข้มสูงสุดลบกับพิกเซลที่มีความเข้มน้อยสุดแล้วนำผลที่ได้ไปเปรียบเทียบกับค่า Threshold ว่ามีค่าน้อยกว่าหรือไม่คุณสมบัติของ Split region คือถ้าภาพเริ่มต้นมีขนาดเป็น NxN ก็จะทำให้ขนาดเท่ากับเป็นกำลังของ 2 ($N = 2^n$) ทุกกลุ่มที่ถูกสร้างด้วยอัลกอลิธึม Split จะเป็นสีเหลี่ยมเป็นขนาดเท่ากับ MxM เมื่อ M เป็นกำลังของ 2 ($M = 2^m$), m<=n และเนื่องจากอัลกอลิธึมนี้จะ เป็นลักษณะของการเรียกตัวเองดังนั้นจึงสามารถแสดงเป็นลักษณะของรูปต้นไม้ได้ซึ่งแต่ละโหนดจะมีการแตกออกมาเป็น 4 กลุ่มซึ่งเรียกว่า Quadtree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การเขียนโปรแกรมติดต่อกับกล้อง web cam

3.1 ActiveX Control (VideoOCX และ VideoOCXTools) คือ

ในปัจจุบันกล้อง web cam ถือได้ว่าเป็นอุปกรณ์ที่ได้รับความนิยมจากผู้ใช้เป็นจำนวนมาก จึงได้มีผู้พัฒนาเครื่องมือต่างๆที่ช่วยในการเขียนโปรแกรมที่ใช้ในการติดต่อกับกล้อง web cam ออกมามากมาย เพื่อตอบสนองต่อความต้องการของผู้ใช้ ซึ่งในโครงงานนี้ได้เลือกใช้เครื่องมือ VideoOCX ซึ่งเป็น ActiveX control ที่ช่วยให้ผู้เขียนโปรแกรมนั้น สามารถเขียนโปรแกรมติดต่อกับกล้อง web cam เพื่อใช้ในการจับภาพและทำ image processing ได้อย่างง่ายดาย โดยเพียงกำหนดค่าตัวแปรให้กับ VideoOCX และเรียกใช้งานผ่าน method ที่มีให้มา ตัวอย่าง method ที่สำคัญได้แก่

3.2 Method ที่สำคัญ

Syntax : BOOL *object*.**Start**()

Description : initiate the background capture process

Parameter : none

Return Type : true if successful

Syntax : BOOL *object*.**Stop**()

Description : end the internal capturing process

Parameter : none

Return Type : true if the internal capture process was stopped

Syntax : BOOL *object*.**ShowDriverDlg**()

Description : open dialog that enables the user to choose a video driver

Parameter : none

Return Type : true if successful

Syntax : void *object*.**SetErrorMessages**(BOOL *flag*)

Description : display every error that occurs

Parameter :true to turn error messages on

Return Type : none

Syntax : BOOL *object*.**Init**()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Description : initialize VideoOCX and connect to the specified image source

Parameter : none

Return Type : true if successful

Syntax : LPCTSTR *object*.**GetLastErrorString**()

Description : returns the message of the last occured error

Parameter : none

Return Type : error message

Syntax : long *object*.**GetGrayImageHandle**()

Description : receive an empty grayscale image

Parameter : none

Return Type : handle to the new allocated image

Syntax : long *object*.**GetColorImageHandle**()

Description : receive an empty color image

Parameter : none

Return Type : handle to the new allocated image

Syntax : BOOL *object*.**ToGray**(*long colorimagehandle, long grayimagehandle*)

Description : convert a color image to grayscale

Parameter : color image handle and grayscale image handle

Return Type : true if successful

Syntax : long *object*.**GetDataPointer**(*long imagehandle*)

Description : get a pointer to the image data

Parameter : image handle

Return Type : address of image data

Syntax : long *object*.**Capture**(*long imagehandle*)

Description : capture the current video image

Parameter : current video image is copied into the handle

Return Type : address to image data array

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Syntax : VOID *object*.**Show**(*long imagehandle*)

Description : display any image in the control

Parameter : handle to the image

Return Type : none

Syntax : VARIANT *object*.**GetMatrix**(*long imagehandle*)

Description : get dimensional array to the image pixels

Parameter : handle to an image

Return Type : 2 or 3 dimensional matrix containing the image pixels

Syntax : void *object*.**ReleaseMatrixToImageHandle**(*long imagehandle*)

Description : release memory for the last matrix

Parameter : handle to and image

Return Type : none

Syntax : void *object*.**ReleaseImageHandle**(*long imagehandle*)

Description : release the memory for an image

Parameter : handle to an image

Return Type : none

Syntax : BOOL *object*.**Close**()

Description : closes the current connection

Parameter : none

Return Type : true if successful

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การเขียนโปรแกรมติดต่อ
ระหว่างแอปพลิเคชั่น กับวินโดวส์โดยใช้ API

ระหว่างแอปพลิเคชั่น กับวินโดวส์โดยใช้ API แม้ว่า VB.net จะเตรียมฟังก์ชัน และความสามารถต่างๆ ไว้อย่างมากมาย แต่ก็ยังไม่สามารถตอบสนองความต้องการของผู้ใช้ได้ทั้งหมด ดังนั้นจึงได้มีการยอมให้ผู้พัฒนาแอปพลิเคชั่น สามารถที่จะเรียกใช้ความสามารถของระบบปฏิบัติการทั้ง Windows 95/98 , NT Workstation ได้อย่างเต็มที่ โดยการเรียกผ่านฟังก์ชัน Windows API

4.1 API คือ

API ย่อมาจาก (Application Programming Interface) ซึ่งเป็นส่วนเชื่อมต่อระหว่างแอปพลิเคชั่นกับ วินโดวส์โดยจะแสดงลักษณะการติดต่อสื่อสารกันระหว่างแอปพลิเคชั่น กับวินโดวส์ทำให้เข้าใจได้ว่าแอปพลิเคชั่น กับวินโดวส์มีการติดต่อสื่อสารกันอย่างไร โดยมีลักษณะการทำงานเป็นแบบ Event-Driven Programming ซึ่งมีลักษณะการทำงานที่ทำให้ผู้ใช้เป็นอิสระแตกต่างจากแบบเดิมที่เป็นSequence Programming และเป็นแบบ Graphic Oriented โดยจะมีฟังก์ชันอยู่มากมาย นับพันๆฟังก์ชัน ถูกแบ่งออกเป็นหมวดหมู่ตามประเภทของการใช้งาน เช่น API ด้านกราฟฟิก , API ด้านเน็ตเวิร์ค เป็นต้น

สำหรับโปรแกรมเมอร์ที่พัฒนาโปรแกรมด้วย .NET นั้น Windows API ดูจะถูกลดความสำคัญลงอย่างมาก ทั้งนี้เนื่องจากใน .NET นั้น Function ต่าง ๆ จำนวนมาก ใน แบบเดียวกับที่ Windows API มีให้นั้น ได้ถูกเตรียมมาให้ใน .NET Framework แล้ว การเรียกจาก Framework ให้ประสิทธิภาพ และเสถียรภาพที่ดีกว่า ปกติแล้วจึงไม่มีความจำเป็นอันใดที่จะต้องไปเรียกใช้ API อีก อย่างไรก็ตาม เผื่อไว้ว่า หากคุณต้องติดต่อกับ API แปลก ๆ จริงๆ หรือ DLL's ที่ไม่ธรรมดา เช่นที่พ่วงมากับ Application , Utilities อื่นๆ ก็ยังคงทำได้

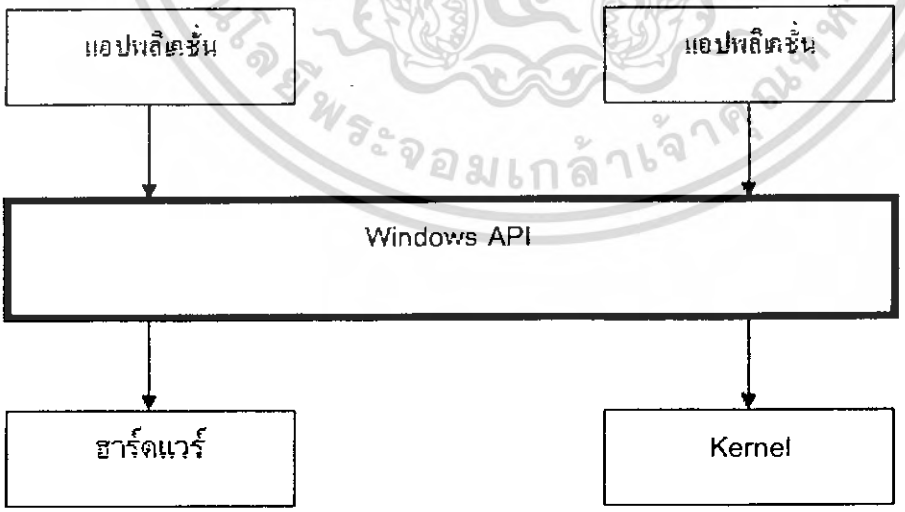
เหตุผลที่ต้องเข้าไปติดต่อกับ Windows API ในบางครั้ง ก็เนื่องจากว่า Windows API เป็น Function ระดับต่ำของ Windows ที่ช่วยให้เราเข้าถึง Software และ Hardware ได้ลึกกว่าที่คำสั่งที่มีมาให้กับภาษาโปรแกรมธรรมดา โดย Function นี้จะถูกเก็บอยู่ในไฟล์ Dll หลัก ๆ ของระบบปฏิบัติการ Windows ถึงแม้การติดต่อกับ Windows API จะยุ่งยาก แต่ก็ยังดีกว่าที่จะต้องเขียนภาษา Assembly แล้ว Link เข้ามาเพื่อจะติดต่อกับ Hardware หรือ Software ระดับต่ำสักกรณีหนึ่ง ซึ่งเป็นข้อดีอีกข้อหนึ่งของ Windows API คือ ช่วยลดความผิดพลาดจากการเข้าถึงฮาร์ดแอร์โดยตรง เพราะเราจะสั่งงานฮาร์ดแวร์ผ่าน Windows API เท่านั้น ซึ่งจะช่วยลดความสับสนในการใช้งานฮาร์ดแวร์ที่ผิดวิธีไปได้มาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการประกาศเรียกใช้ API Function ใน .NET นั้น มีอยู่สองวิธีในการประกาศ วิธีเดิม ๆ เช่น การใช้ Declare Statement แบบเดิมบน VB6.0 ยังคงใช้ได้ อีกวิธี ก็คือ การใช้ Dllimort ซึ่งเป็นวิธีประกาศที่ทำให้เราเข้าถึง และจัดการได้ดีกว่า อย่างไรก็ตาม ขนาดของ Data type ที่เปลี่ยนไป เป็นสิ่งที่ต้องคำนึงถึง เนื่องจาก ขนาดของ Datatype ใน .NET นั้น หลาย ๆ ค่ามีขนาดใหญ่ขึ้น เช่น Long Integer ใน VB6 มีขนาด 32 Bits ขณะที่ ใน .NET นั้น Long Integer เหมือนกัน แต่มีขนาด 64 Bits การใช้ Datatype ผิดขนาดก็จะทำให้เกิด Error ขึ้น ซึ่งแน่นอน การ Debug กับ Windows API ยากกว่าการ Debug กับ Code ธรรมดา เพราะไม่ได้มี Error message บอกชัดเจน และผลของ Error อาจทำให้เครื่อง Hang หรือ ปัญหาอื่นๆ ที่รุนแรงกว่า Code แบบธรรมดา

นอกจากนี้สาเหตุที่ต้องมี Windows API ก็เพราะว่า เราจะได้มีวิธีการมาตรฐานในการใช้งานวินโดวส์แบบเดียว ซึ่งจะช่วยลดเวลาของเราที่ต้องไปศึกษาอุปกรณ์ และองค์ประกอบต่างๆ ทำให้เราไม่จำเป็นต้องรู้ว่าจอภาพทำงานอย่างไร แป้นพิมพ์ทำงานอย่างไร เมาส์ทำงานอย่างไร แต่สามารถเรียกใช้งานผ่าน Windows API ได้ทันที

นอกจากนี้ยังมีข้อดีของ Windows API อื่นๆอีก ได้แก่ ความเป็นมาตรฐานที่ทุกคนรู้จัก จึงง่ายต่อการแก้ไขเปลี่ยนแปลง เช่น แม้ว่าเราจะเปลี่ยนไปใช้ฮาร์ดแวร์ใหม่ หรือแม้กระทั่งระบบปฏิบัติการเวอร์ชันใหม่ๆ แล้วก็ตาม ชื่อของฟังก์ชันเหล่านี้ก็ยังคงใช้ชื่อเดิมและทำงานได้เหมือนเดิม ทำให้เราไม่ต้องกังวลกับความเปลี่ยนแปลงที่จะเกิดขึ้นบ่อยเกินไปนัก ซึ่งมันก็คือชุดของฟังก์ชันที่พร้อมให้เราเรียกใช้งาน โดยที่ฟังก์ชันเหล่านั้นรับประกันได้ว่าทำงานอย่างมีประสิทธิภาพ เช่น เราไม่จำเป็นต้องเขียนโปรแกรมให้ยืดยาว เพื่อตรวจสอบว่าเนื้อที่ฮาร์ดดิสก์เหลือว่างอยู่เท่าไร แต่เรียกใช้ฟังก์ชัน GetDiskFreeSpace ฟังก์ชันเดียวก็รู้แล้ว เป็นต้น



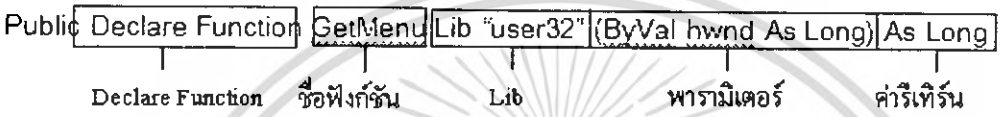
รูปที่ 4-1 แสดงหลักการทำงานของ Windows API

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การเรียกใช้ Windows API

4.2.1 โดยการประกาศ Declare

วิธีนี้เหมือนกับที่ใช้บน VB6 โดยการประกาศ Declare ตามด้วยชื่อ Function , Argument Type , Return value , ชื่อของ dll ที่ Function นั้นๆ ถูกเก็บอยู่ โดยการประกาศนั้นจะประกาศไว้ที่ส่วนของ General Declaration กล่าวคือ ต้องประกาศไว้ภายใน Class แต่ต้องไม่อยู่ในพื้นที่ของ Sub หรือ Function อื่นๆ นั่นคือต้องมีการประกาศค่าฟังก์ชันก่อนการใช้งาน สำหรับการใช้งานฟังก์ชันนั้น เราจะคัดลอกมาจาก API viewer ซึ่งฟังก์ชันที่คัดลอกมานั้น มีรูปแบบการประกาศดังนี้



จากตัวอย่าง เป็นการประกาศฟังก์ชัน GetMenu จะเห็นว่าฟังก์ชันที่ประกาศนั้นประกอบด้วยองค์ประกอบต่างๆ ดังนี้

- Declare Function เป็นคำสั่งแสดงว่านี้คือ การประกาศฟังก์ชัน
- ชื่อฟังก์ชัน เป็นชื่อของฟังก์ชัน จากตัวอย่างนี้ชื่อของฟังก์ชันคือ GetMenu
- Lib เป็นการบอกให้ทราบว่าให้ค้นหาจากไฟล์ .DLL ซึ่งจะต้องระบุชื่อไฟล์ .DLL นั้น ยกเว้นถ้าอยู่ในไฟล์ USER32.DLL , KERNEL32.DLL และ GDI32.DLL
- พารามิเตอร์ เป็นรายการของพารามิเตอร์ที่ต้องมีการผ่านค่าให้ฟังก์ชัน
- ค่าที่รีเทิร์น เป็นชนิดข้อมูลที่จะคืนค่าให้กับผู้เรียกใช้ จากตัวอย่างมีชนิดข้อมูลเป็น Long

ในการเขียนโปรแกรมติดต่อกับเมาส์นั้น เราจะนำค่าตำแหน่ง x,y มาตามโค้ด ดังนี้

```
'Object tracking จะอยู่ในฟอร์ม
Friend WithEvents ObjectTracking As AxObjectTrackingLib.AxObjectTracking
Me.ObjectTracking = New AxObjectTrackingLib.AxObjectTracking

'นำ ProcessFrame() Method ไปไว้ในฟังก์ชัน Timer_Tick() เพื่อไว้สำหรับการอัพเดต Object
Tracking
ObjectTracking.ProcessFrame()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
'2 method นี้ ไว้ใช้สำหรับ start และ stop tracking
ObjectTracking.StartCapture()
ObjectTracking.StopCapture()

'รับค่าข้อมูลของ object
Dim x As Integer
Dim y As Integer
Dim w As Integer
Dim h As Integer
Dim a As Double

x = ObjectTracking.ObjectX
y = ObjectTracking.ObjectY
w = ObjectTracking.ObjectWidth
h = ObjectTracking.ObjectHeight
a = ObjectTracking.ObjectAngle
```

ในการMoving หลังจากได้รับค่า attribute ของ object มาแล้ว ก็จะทำการแปลงตำแหน่งที่กล้อง web cam จับได้มาเป็นตำแหน่งบนหน้าจอ โดยจะทำการตั้งค่าตำแหน่งของเคอร์เซอร์โดย Windows API

```
Private Declare Function SetCursorPos Lib "user32" Alias "SetCursorPos" (ByVal x As
Integer, ByVal y As Integer) As Integer
```

และในส่วนของการกำหนดค่าให้กับ Mouse event

```
'Mouse event constant
Private Const MOUSEEVENTF_LEFTDOWN = &H2
Private Const MOUSEEVENTF_LEFTUP = &H4
Private Const MOUSEEVENTF_RIGHTDOWN = &H8
Private Const MOUSEEVENTF_RIGHTUP = &H10
Private Const MOUSEEVENTF_MIDDLEDOWN = &H20
Private Const MOUSEEVENTF_MIDDLEUP = &H40
```

72740
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
'Mouse event raising
Private Declare Sub mouse_event Lib "user32" (ByVal dwFlags As Integer, ByVal dx As
Integer, ByVal dy As Integer, ByVal cButtons As Integer, ByVal dwExtraInfo As Integer)
```

4.2.2 การติดต่อ API โดยใช้ DllImport

DllImport เป็น Attribute ที่สามารถใช้เป็นอีกวิธีหนึ่งในการติดต่อกับ Windows API ได้ แต่ให้สามารถที่มากกว่า ในการควบคุมจัดการกับ Function

4.3 กลุ่มฟังก์ชันของ Windows API

Windows 95/98 , NY Workstations จะมีฟังก์ชันในรูปของ Windows API ให้ใช้มากกว่า 1000 ฟังก์ชัน ซึ่งฟังก์ชันต่างๆจะถูกเก็บในไฟล์ .DLL ซึ่งอยู่ในโฟลเดอร์ c:\Windows\Systems ซึ่งมีไฟล์สำคัญๆ ได้แก่

ADVAPI32.DLL มาจาก Advanced API ซึ่งจะเก็บฟังก์ชันเกี่ยวกับความปลอดภัย

COMDLG.DLL เก็บฟังก์ชันของ Common Dialog

GDI32.DLL GDI ซึ่งจะเก็บฟังก์ชันเกี่ยวกับการแสดงผล และฟังก์ชันกราฟฟิก

KERNEL32.DLL เก็บฟังก์ชันที่ใช้จัดการทรัพยากรของระบบปฏิบัติการ

LZ32.DLL เก็บฟังก์ชันที่ใช้เกี่ยวกับการบีบอัดข้อมูล

MPR.DLL เก็บฟังก์ชันที่เกี่ยวกับ Multiple Provide Router

NETAPI32.DLL เก็บฟังก์ชันที่เกี่ยวกับการจัดการเครือข่าย (Network)

SHELL32.DLL เก็บฟังก์ชันที่เกี่ยวกับเชลล์ (Windows Explorer)

USER32.DLL เก็บฟังก์ชันที่เกี่ยวกับการติดต่อกับผู้ใช้งาน (User Interface)

VERSION32.DLL เก็บฟังก์ชันที่เกี่ยวกับการจัดการรุ่นของซอฟต์แวร์

WINMM.DLL เก็บฟังก์ชันที่เกี่ยวกับมัลติมีเดีย

WINSPOLL.DRV เก็บฟังก์ชันที่เกี่ยวกับการพิมพ์ผ่านเครื่องพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

หลักการรู้จำเสียงโดย Hidden Markov Model

ในปัจจุบันระบบรู้จำเสียงภาษาไทยยังไม่ได้มีการพัฒนาขึ้นมาเป็นผลิตภัณฑ์ให้ใช้งานได้ ระบบรู้จำเสียงที่ใช้งานได้จริงในต่างประเทศนั้น สามารถประยุกต์ใช้กับภาษาไทยได้แต่ยังเกิดความผิดพลาดอยู่มาก ในส่วนของโครงงานนี้จะเน้นไปในส่วนของการพัฒนาโปรแกรมแบบ Speech Recognitionหรือการรู้จำเสียง ซึ่งในการรู้จำเสียงนี้ใช้หลักการของ Hidden Markov Model โดยใช้อุปกรณ์ที่ใช้ในงานวิจัยนี้ คือ Hidden Markov toolkit (HTK)

5.1 Hidden Markov Model คือ

ก่อนที่จะไปกล่าวถึง Hidden Markov Model เราจะมากล่าวถึงแนวคิดของ Markov Model กันเสียก่อน ซึ่งมีใจความดังนี้ หากเราพูดถึงเรื่องสภาพอากาศที่มีอยู่สามแบบได้แก่ sunny rainy และ foggy และสมมติให้ว่าแต่ละวันนั้นสภาพอากาศจะเป็นแบบเดียวกันทั้งวันโดยไม่มีการเปลี่ยนแปลง

Weather prediction (การทำนายสภาพอากาศ): นั้นเป็นการพยายามที่จะทายว่าอากาศในวันพรุ่งนี้นั้นจะเป็นอย่างไรโดยใช้การเก็บรวบรวมข้อมูลสภาพอากาศจากที่ผ่านๆมา โดยข้อมูลจะเก็บดังนี้ ให้อากาศในวันนี้มีค่าเป็น q_n และวันก่อนหน้านี้เป็น q_{n-1}และวันก่อนๆ เป็น q_{n-2}, q_{n-3}, q_{n-4} ตามลำดับเราจะได้เงื่อนไขความน่าจะเป็นในการทำนายสภาพอากาศดังนี้

$$P(q_n | q_{n-1}, q_{n-2}, \ldots, q_1) \qquad \text{--------(1)}$$

หมายความว่า ความน่าจะเป็นของสภาพอากาศที่เราไม่ทราบในวันที่ n นั้นขึ้นอยู่กับสภาพอากาศที่เราทราบมาแล้วจะวันที่ผ่านๆมาโดยเป็นไปได้สามแบบคือ sunny rainy และ foggy

จากสมการความน่าจะเป็นที่ 1 เราก็สามารถจะทำนายอากาศวันพรุ่งนี้และวันต่อไปได้โดยใช้ข้อมูลที่เก็บเป็นสถิติไว้จากวันที่ผ่านๆมา ยกตัวอย่างเช่น หากเราทราบสภาพอากาศของสามวันที่ผ่านมาเป็นดังนี้ {Sunny,Sunny,Foggy} ตามลำดับ ความน่าจะเป็นที่พรุ่งนี้อากาศจะเป็น rainy จะเป็นดังนี้

$$P(q_4=\text{Rainy} | q_3=\text{Foggy}, q_2=\text{Sunny}, q_1=\text{Sunny}) \qquad \text{--------(2)}$$

ปัญหาอย่างหนึ่งที่ตามมานั่งก็คือ ยิ่ง n มีค่ามากเราก็ต้องเก็บความน่าจะเป็นของสภาพอากาศในอดีตที่ผ่านมามากขึ้นตาม ยกตัวอย่าง หากให้ n=6 แล้วเราต้องทำการเก็บสถิติไว้ถึง $3^{(6-1)}$ = 243 วัน ฉะนั้นจึงได้มีการสร้างสมมติฐานอย่างหนึ่งขึ้นมาซึ่งเรียกว่า Markov assumption

สำหรับลำดับ $\{ q_1, q_2, \ldots, q_n \}$; เราจะได้

$$P(q_n | q_{n-1}, q_{n-2}, \ldots, q_1) = P(q_n | q_{n-1}) \qquad \text{---------(3)}$$

ซึ่งสมการที่ (3) นี้เรียกว่า first-order Markov assumption: โดยเราสามารถกล่าวได้ว่าความน่าจะเป็นของสภาพต่างๆ ณ เวลา n ขึ้นอยู่กับ q_{n-1} ณ เวลา n-1 เท่านั้น (second-order

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

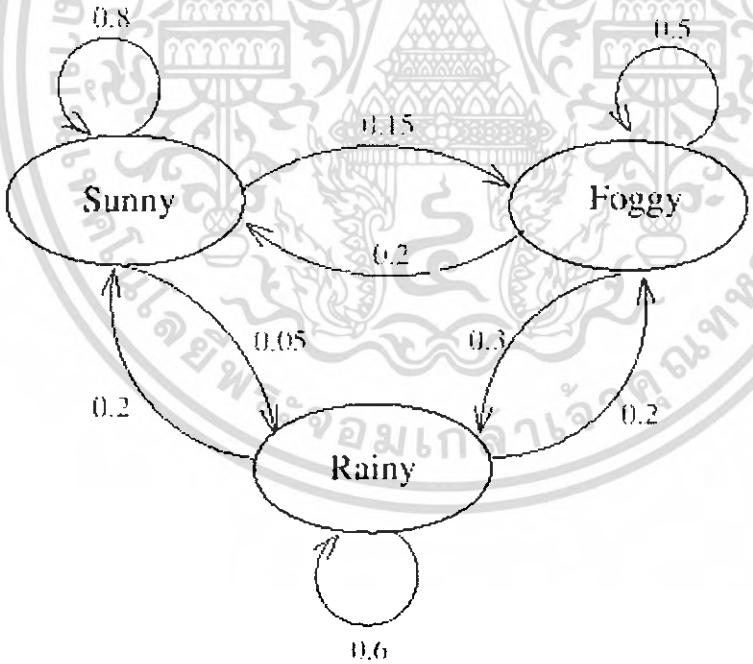
Markov assumption: ก็จะหมายถึง ความน่าจะเป็นของสภาพต่างๆ ณ เวลา n ขึ้นอยู่กับ q_{n-1} และ q_{n-2} แต่โดยปรกติแล้วเวลาผู้คนพูดกล่าวถึง Markov assumption ก็มักจะหมายถึง first-order Markov assumption)

ตัวอย่างการใช้งานสำหรับ Markov Model นั้น สมมติเรามีตารางที่บอกถึงค่าความน่าจะเป็นของอากาศสำหรับวันพรุ่งนี้โดยอิงจากสภาพอากาศในวันนี้เป็นดังนี้

Today's weather	Tomorrow's weather		
	Sunny	Rainy	Foggy
Sunny	0.8	0.05	0.15
Rainy	0.2	0.6	0.2
Foggy	0.2	0.3	0.5

ตารางที่ 5-1 แสดงความน่าจะเป็นของสภาพอากาศวันพรุ่งนี้อ้างอิงจากสภาพอากาศวันนี้

สำหรับ First-order Markov models นั้นเราสามารถจะนำเอาข้อมูลความน่าจะเป็นนี้ไปเขียนให้อยู่ในรูปของ finite state automaton ได้โดย weather domain ก็จะมีอยู่ด้วยกัน 3 state ได้แก่ sunny rainy และ foggy (S={Sunny,Rainy,Foggy}) และค่าความน่าจะเป็นก็จะเป็นไปตามตารางที่ให้มาซึ่งก็จะได้ automatonออกมาดังนี้



รูปที่ 5-1 แสดง Automaton ความน่าจะเป็นของสภาพอากาศ

จากที่กล่าวมานั้นเป็นเรื่องของ Markov Model ต่อไปเราจะทำการกล่าวถึงเรื่อง Hidden Markov Model กันต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หากลองสมมติว่าเราถูกบังคับให้อยู่ในห้องที่ไม่เห็นสภาพอากาศภายนอกเป็นเวลาหลายๆวันแล้ว เราจะไม่สามารถรู้ได้อย่างแน่นอนว่า สภาพอากาศในแต่วันนั้นเป็นอย่างไรกันบ้าง เราจะคาดเดาได้ว่าสภาพอากาศเป็นอย่างไรในแต่ละวันก็ได้จาก การที่คนที่คอยนำอาหารมาส่งให้เรานั้น ถือร่มมารึป่าวและหากเขาถือร่มมาด้วยเราก็จะได้ตารางความน่าจะเป็นของสภาพอากาศเป็นดังนี้

Weather	Probability of umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

ตารางที่ 5-2 แสดงค่าความน่าจะเป็นของสภาพอากาศเมื่อเห็นผู้ดูแลนำร่มมาด้วย

อย่างไรก็ตามสภาพอากาศที่แท้จริงนั้นเราก็ไม่สามารถทราบได้ โอกาสที่จะเป็นไปได้ของสภาพอากาศก็คือ $q_i \in$ {Sunny,Rainy,Foggy} สามารถจะบ่งบอกได้จากการที่ผู้ดูแลถือร่มมาด้วยหรือไม่เท่านั้น (x_i,Umbrella or NoUmbrella) จากเงื่อนไขเหล่านี้เราสามารถที่รวบรวมเป็น condition probability ตาม Bayes' rule ได้ดังนี้

$$P(q_i|x_i)=P(x_i|q_i)P(q_i)/P(x_i)$$

สำหรับวัน n วันและมีลำดับดังนี้ Q= $\{q_1, \ldots, q_n\}$เช่นเดียวกับลำดับของการนำเอาร่มมาเป็น X=$\{x_1,\ldots,x_n\}$ จะได้สมการเป็น

$$P(q_1, \ldots, q_n|x_1,\ldots,x_n)=P(x_1,\ldots,x_n|q_1, \ldots, q_n)P(q_1, \ldots, q_n)/P(x_1,\ldots,x_n)$$

ในการสรุปความน่าจะเป็นนั้นถึงแม้ว่าเราจะตัดเอา $P(x_1,\ldots,x_n)$ ออก ก็ไม่ได้ทำให้ผลสรุปมีการเปลี่ยนแปลงไปแต่อย่างใดเพราะเป็นแค่การเปรียบเทียบด้วยค่าที่เท่ากันดังนั้นเราจึงสามารถละออกจากสมการได้แล้วเราจะได้ค่าของผลสรุปที่เราจะให้เป็นค่า L ดังนี้

$$P(q_1, \ldots, q_n|x_1,\ldots,x_n) \propto L(q_1, \ldots, q_n|x_1,\ldots,x_n) = P(x_1,\ldots,x_n|q_1, \ldots, q_n)P(q_1, \ldots, q_n)$$

5.2 หลักการรู้จำเสียง

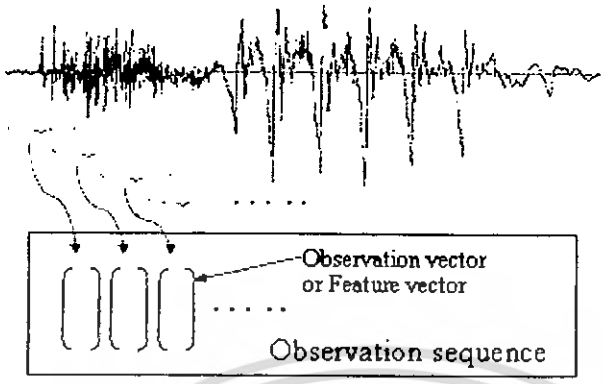
ในการรู้จำเสียงนั้น ระบบจะไม่ได้เอาสัญญาณเสียงที่ได้มาโดยตรงไปเทียบ แต่จะดึงเอาเฉพาะค่าสำคัญของเสียงออกมา ซึ่งเรียกค่าสำคัญนี้ว่า Speech feature ค่าสำคัญที่เป็นที่นิยมมากที่สุดในปัจจุบันคือ Mel-frequency cepstral coefficient (MFCC) ซึ่งถือว่าเป็นตัวแทนของเสียงในการรู้จำได้เป็นอย่างดี โดยขั้นตอนการแปลงนี้ เราใช้คำสั่ง

```
HCopy -T 1 -C config/code.config -S code.scp
```

ซึ่งเป็นการแปลงจาก .wav เป็น .mfc ค่า MFCC นี้จะเก็บเป็นเวกเตอร์ โดยเวกเตอร์นึงแทนสัญญาณเสียงยาวประมาณ 20 มิลลิวินาที แต่ละเวกเตอร์ก็แทนสัญญาณเสียง ที่ค่อยๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เลื่อนไปแบบ overlap เช่นเลื่อนไปทีละ 10 มิลลิวินาที ดังรูปที่ 5-2 ดังนั้นหากมีเสียงที่ยาวซัก 1 วินาทีเข้ามา ก็จะแทนด้วยเวกเตอร์จำนวน 100 อัน



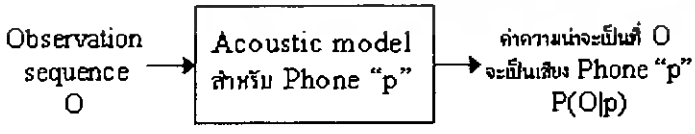
รูปที่ 5-2 การดึง Speech feature

Sequence ของเวกเตอร์ที่แทนสัญญาณเสียงนี้ เรียกกันว่า Observation sequence และ เรียกเวกเตอร์ใดๆ ใน Observation sequence ว่า Observation vector

Phone คือ หน่วยย่อยสุดทางเสียง ตัวอย่างเช่น คำว่า "การ" จะอ่านออกเสียงด้วยเสียง "ก" ตามด้วยสระ "า" และลงท้ายด้วย เสียงตัวสะกด "น" โดยจะมีสัญลักษณ์มาตรฐานแทนเสียง Phone แต่ละเสียง ตัวอย่างเช่น "k" แทน เสียง "ก" "aa" แทนสระ "า"

โมเดลการออกเสียง (Pronunciation model) จะมีหน้าที่ในการบอกว่า คำ(word) ใดๆ ออกเสียงอย่างไรว่า โดยจะบอกเป็น Sequence ของ Phone เช่น "การ" ออกเสียงว่า "k aa n^" "ขวา" ออกเสียงว่า "khw aa" โดยวิธีการสร้างโมเดลการออกเสียงต้องสร้าง Pronunciation dictionary ที่เก็บ list ของคำคู่กับเสียงอ่านของคำๆนั้น

โมเดลเสียง (Acoustic model) จากรูปที่ 5-3 โดยปกติเราจะมีโมเดลเสียง 1 โมเดลต่อ 1 Phone เมื่อเราป้อน Observation sequence เข้าไปยังโมเดลเสียงใดๆ มันจะคำนวณค่าความน่าจะเป็นที่ Observation sequence นั้นจะเป็นเสียงของ Phone นั้นๆ ความน่าจะเป็นนี้คือ P(O|p) โดยที่ p คือโมเดลเสียงของ Phone ใดๆ

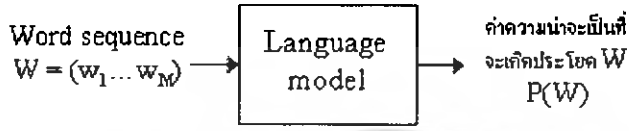


รูปที่ 5-3 หน้าที่ของ Acoustic model

โมเดลภาษา (Language model) นั้นจะบอกเราว่าคำ (Word) นี้ ตามด้วยคำนี้ได้หรือไม่ หรือในบางโมเดลจะบอกค่าความน่าจะเป็นที่คำใดๆ จะพูดต่อกัน อาทิเช่น โมเดลภาษาอาจจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

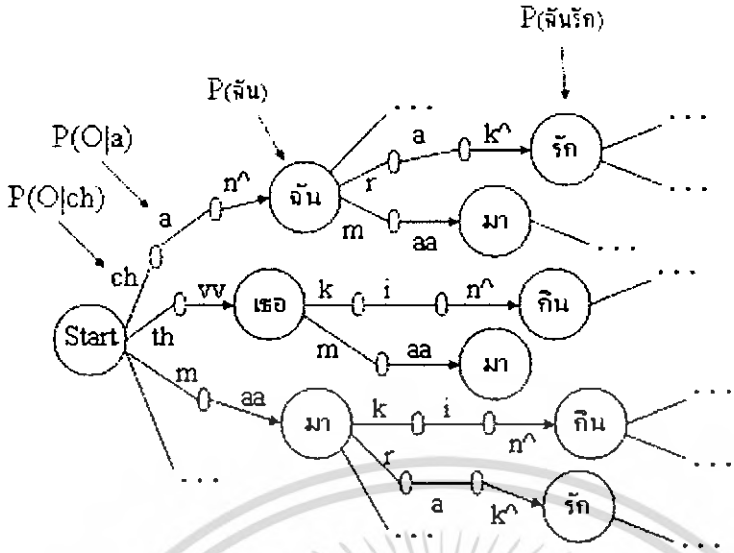
บอกว่า "จะ ไป" เกิดได้ แต่ "ไป จะ" ไม่ได้ หรืออาจจะบอกเป็นค่าความน่าจะเป็นว่า "จะ ไป" มีโอกาสเกิดได้ 0.8 แต่ "ไป จะ" มีโอกาสเกิดได้แค่ 0.01 เป็นต้น นอกจากนี้โมเดลภาษายังสามารถบอกด้วยว่า ทั้งประโยคมีโอกาสเกิด ได้เท่าไหร่ สมมติว่าเรามีประโยค ซึ่งประกอบด้วยคำต่อๆ กันหลายๆ คำ แทนด้วย W = (w1...wM) โดยที่ w แทนคำแต่ละคำ โมเดลภาษาจะบอกว่า W สามารถเกิดได้หรือไม่ หรือบอกเป็นค่าความน่าจะเป็นว่ามีโอกาสเกิดมากน้อยแค่ไหน โดยแทนค่าความน่าจะเป็นด้วย P(W) ดังรูปที่ 5-4 แสดงให้เห็นหน้าที่ของโมเดลภาษา



รูปที่ 5-4 หน้าที่ของ Language model

โดยในการรู้จำนั้นระบบจะรับ Observation sequence เข้ามา และทำการเดาว่าเป็นคำใดๆต่อกัน จากนั้นจะนำประโยคที่ได้เข้าไปยัง Language model ได้ค่าความน่าจะเป็น P(W) ที่จะเกิดประโยคดังกล่าว แล้วนำประโยคนั้นมาทำการแปลงเป็นเสียงอ่านโดยอาศัย Pronunciation model เมื่อได้ Sequence ของ Phone แล้วก็จะเอาAcoustic model ของแต่ละ Phone มาต่อกัน แล้วทำการป้อน Observation sequence เข้าไปยัง Acoustic model ของทั้งประโยค ก็ได้จะค่าความน่าจะเป็น P(O|W) ซึ่งเกิดจาก P(O|p) ของแต่ละ Phone คูณกัน จากนั้นก็จะได้เป็นโอกาสที่สัญญาณเสียงดังกล่าวจะเป็นเสียงประโยค W แต่เนื่องจากประโยคที่เป็นไปได้มีได้หลากหลายรูปแบบมาก จึงใช้วิธีการแก้ไขคือ Beam search เป็นการสร้าง Word network ซึ่งเป็นการเป็นการนำเอาคำมาต่อๆกันในลักษณะของ Network ดังรูปที่ 5-5 ระหว่างคำก็กำกับด้วยโอกาสที่แต่ละคำจะต่อกัน หรือ P(wi|wi-1) และในแต่ละคำก็ประกอบด้วย Acoustic model ของ Phone ที่ต่อกันเป็นเสียงอ่านของคำนั้นๆ แล้วเวลาทำงานก็จะผ่านสัญญาณเสียงเข้าไป ในขณะที่ผ่าน Node ของ Network แต่ละ Node ก็จะมีการคูณค่าความน่าจะเป็น P(O,W) ต่อๆ ไปเรื่อยๆ หากในเส้นทางใดที่ค่าความน่าจะเป็นรวมขณะนั้น ตกต่ำกว่าค่า Threshold ที่กำหนด ก็จะเลิกวิ่งไปเส้นทางนั้น ซึ่งจะช่วยลดจำนวนประโยคที่จะต้องคำนวณลงได้มากมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

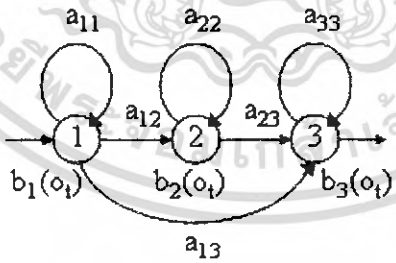


รูปที่ 5-5 Word network สำหรับการรู้จำ

5.3 การ Training

5.3.1 หลักการสร้างโมเดลเสียง (Acoustic model)

ในปัจจุบันนี้ Acoustic model ที่เป็นที่นิยมในการใช้ ใช้วิธีที่เรียกว่า Hidden Markov Model (HMM) ซึ่ง HMM ในรูปแบบมาตรฐานสำหรับแทนเสียง Phone 1 Phone จะมีลักษณะดังรูปที่ 5-6 ซึ่ง Observation sequence จะถูกป้อนเข้าทางซ้ายหรือ เข้าทาง Node ที่ 1 และออกทาง Node สุดท้าย ในระหว่างที่ Observation vector แต่ละตัว วิ่งอยู่ใน HMM นี้ จะวิ่งไปข้างหน้า หรือย่ำอยู่ที่ Node เดิม (ตาม Link ที่มีอยู่) เท่านั้น ไม่มีการย้อนกลับ เราจึงเรียก HMM แบบนี้ว่า Left-to-right model

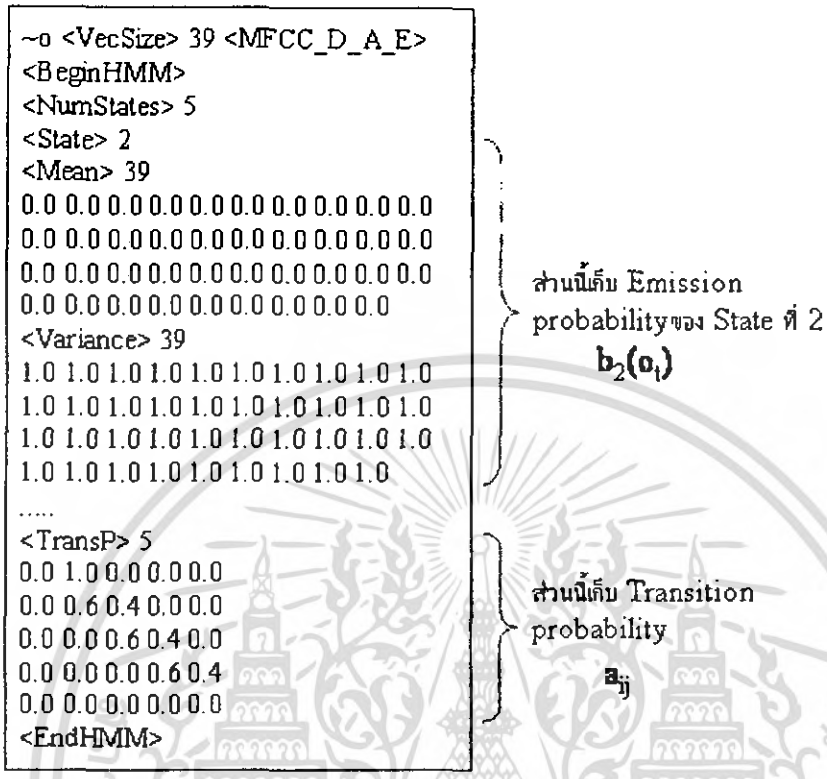


รูปที่ 5-6 Left-to-right HMM

Link ที่เชื่อมระหว่าง Node เรียกว่า Transition และ Node แต่ละอันเรียกว่า State ในขณะที่ Observation vector หนึ่งๆ กระโดดจาก State i ไป State j จะมีความน่าจะเป็นของการกระโดดคือ aij เราเรียกความน่าจะเป็นของการกระโดดนี้ว่า Transition probability และเมื่อ Observation vector เช่น ot ไปตกที่ State i จะมีความน่าจะเป็นของการตกคือ bi(ot) ซึ่งเรา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียกว่า Emission probability ซึ่งไฟล์เริ่มต้นของ HMM 1 ตัวที่ยังไม่ได้รับการ Train จะมีลักษณะดังรูปที่ 5-7



รูปที่ 5-7 แสดง Prototype model

ในบรรทัดแรกความยาวของ vector ได้ถูกกำหนดให้เป็น 39 และได้ประกาศตัวแปรเป็นแบบ MFC-coefficients(MFCC) ในส่วนถัดมาจะเป็นการกำหนดค่าให้กับตัวแปรต่างๆเช่น จำนวนของ state แต่ละ state มี mixture mean และ variance เป็นเท่าไหร่ๆและในส่วนสุดท้ายของ model ก็จะเป็น <TransP> ซึ่งจะบ่งบอกถึง transition matrix ดังในตัวอย่างที่ให้มานี้จะเห็นได้ว่า เราได้กำหนดความน่าจะเป็นของการที่จะ คงอยู่ ณ state เดิมไว้ที่ 60% และความน่าจะเป็นที่จะไปต่อยัง state ถัดไปได้ถูกกำหนดไว้เป็น 40% และความน่าจะเป็นที่จะไปยัง state อื่นๆ เป็น 0 โดยค่าต่างๆในตัว prototype นี้ไม่มีตัว tool ที่จะช่วยในการสร้างฉะนั้นเราจึงต้องเขียนมันขึ้นมาเองโดยผู้ใช้งาน โดยขั้นตอนในการ Train จะมีขั้นตอนดังนี้

1. แปลงไฟล์เสียง .wav ที่จะใช้ Train Acoustic model เป็นไฟล์ค่าลักษณะเด่น .mfc ก่อนด้วยคำสั่ง HCopy

2. เตรียม Label file หรือไฟล์ที่กำกับว่าไฟล์เสียงแต่ละไฟล์ประกอบด้วย Phone อะไรต่อกันบ้าง

3. เริ่มต้นการ Train HMM ด้วยวิธีที่เรียกว่า Flat start ซึ่งHTK จะมีตัว tool ที่ชื่อว่า HCompV ซึ่งจะทำหน้าที่สแกนตัว prototype ไฟล์ทั้งหมดแล้วทำการคำนวณค่า mean และ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

variance ให้มีค่าเท่ากัน หากเราสมมุติว่า ตัว training ไฟล์ทั้งหมดได้ถูกเก็บไว้ใน train.scp เราจะใช้คำสั่ง

HCompV –C config –f 0.01 –m –S train.scp –M hmm0 proto

โดย HCompV จะทำการแบ่ง Observation sequence ใน .mfc ทุกไฟล์เป็นส่วนๆ เท่ากับจำนวน Phone ที่กำกับทั้งหมดใน Label file แล้วทำการ คำนวณค่า Emission และ Transition probability ใส่ใน proto โดยมองว่าทุก Phone เป็น Phone เดียวกัน ผลก็จะได้ proto ตัวใหม่ ในไดเรกทอรี่ hmm0 เป็นเหมือน HMM ที่เฉลี่ยค่า Probability ให้เท่าๆ กัน สำหรับทุก Phone

4. หลักจากที่ได้ proto ตัวใหม่มาแล้ว ก็ทำการ copy ไปเป็น Phone แต่ละตัวที่ต้องใช้ในระบบด้วยคำสั่ง

```
sed 's/'$Proto'/'$phn'/' $ModelDir/$TargetDir/$Proto > $ModelDir/$TargetDir/$phn
```

คำสั่งนี้จะทำการ copy proto ไปเป็นไฟล์ของแต่ละ Phone ที่เราได้กำหนดไว้

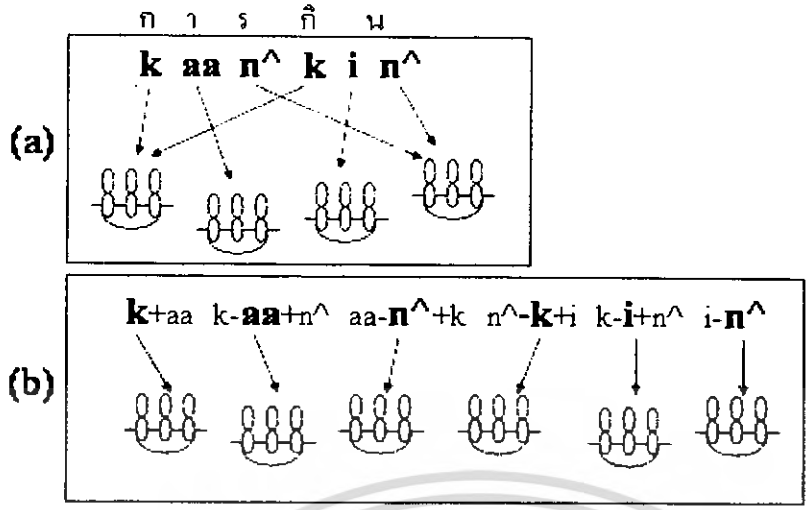
5. ทำการ Re-estimate แต่ละ Phone HMM ด้วย .mfc ทั้งหมดใหม่ ด้วยคำสั่ง

```
HERest -T 1 -A -C $ConfigDir/$TrainCfg \
        -I $ConfigDir/$MonophnMlf \
        -S $TrainScp -d $ModelDir/$SourceDir \
        -M $ModelDir/hmm_1 $ConfigDir/$MonophnLst
```

เป็นการปรับค่า Emission และ Transition probability ในแต่ละ Phone หลายๆรอบ (โดยในงานวิจัยนี้ทำการ Re-estimate 3 รอบ) จะได้ HMM ของ Phone ใหม่ที่มีค่าแตกต่างกันในแต่ละPhone ซึ่งไฟล์ Phone HMM ที่ได้นี้จะทำการเก็บเป็นไฟล์เดียว ชื่อไฟล์ว่า newMacros

6. Monophone model คือ HMM 1 ตัวจะแทน Phone 1 Phone โดยไม่สนใจว่าPhone ข้างๆ เป็น Phone อะไร ดังนั้นสัญญาณเสียงของ Phone เดียวกันจะถูกเอามา Train HMM ของ Phone นั้นๆ แต่ความจริงแล้วเสียงของ Phone นั้นขึ้นกับ Phone ข้างๆด้วยว่าเป็น Phone อะไร เสียงของ Phone เดียวกันอาจจะแตกต่างกันได้หากอยู่ในตำแหน่งต่างๆกันไป จึงต้องทำการสร้าง Phone model แยกกันในทุกๆกรณีที่ Phone รอบข้างต่างกัน ซึ่งก็คือ Triphone model ดังรูป 5-8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-8 แสดง (a) Monophone model (b) Triphone model

ขั้นตอนต่อไปจึงนำ Monophone model ที่สร้างไว้ใน newMacros มาแปลงเป็น Triphone model โดยใช้คำสั่ง

```
HLEd -T 1 -A -D -n $ConfigDir/$TriphnLst \
        -l '*' -i $ConfigDir/$TriphnMlf \
            $ConfigDir/$MktriLed $ConfigDir/$MonophnMlf
```

ซึ่งเป็นการนำ Label file $ConfigDir/$MonophnMlf มาแปลงเป็น Triphone ใน $ConfigDir/$TriphnMlf และทำการสร้างลิสต์ของ Triphone ที่มีทั้งหมดไว้ที่ $ConfigDir/$TriphnLst ซึ่งการทำงานในส่วนนี้จะเป็นไปตามคำสั่งที่ได้ระบุไว้ในไฟล์ชื่อ $ConfigDir/$MktriLed ซึ่งมีคำสั่งดังนี้

```
WB sil
TC
```

โดย WB ย่อมาจาก word boundary symbols โดยคำสั่งนี้จะทำการ block ส่วนเพิ่มเติมของ context จากคำสั่ง TI

7. หลังจากนั้น ใช้คำสั่ง

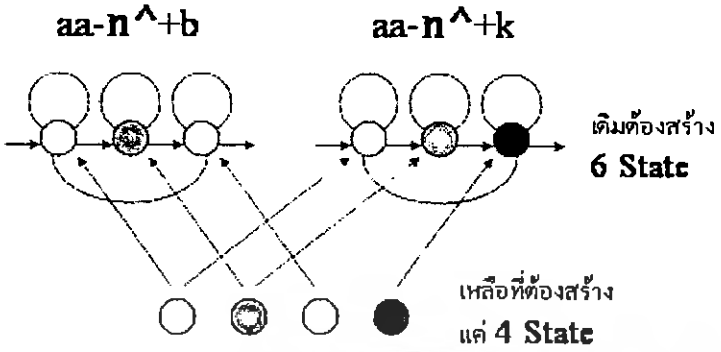
```
HHEd -T 1 -A -D -H $ModelDir/$SourceDir/newMacros \
        -M $ModelDir/$TargetDir \
        $ConfigDir/$MktriHed $ConfigDir/$MonophnLst
```

เป็นการโหลด HMM macro file จาก $ModelDir/$SourceDir/newMacros เพื่อสร้าง Triphone HMM โดยนำไปไว้ที่ $ModelDir/$TargetDir ซึ่งในที่นี้จะต้องใช้ไฟล์ $ConfigDir/$MktriHed ด้วย

8. ทำการ Re-estimate Phone model ที่เป็น Triphone นี้ใหม่ โดยทำเหมือนกับในขั้นตอนที่ผ่านมา ผลลัพธ์สุดท้ายจะได้ newMacros มา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. นำมาสร้าง Tied-state model ซึ่งคือการนำ Triphone บางตัวที่น่าจะมีองค์ประกอบคล้ายๆกันมา Share State ร่วมกันดังรูปที่ 5-9



รูปที่ 5-9 แสดง Tied-State model

โดยทำการสร้างไฟล์ใหม่อันนึงที่ใช้เก็บลิสต์ของ Triphone ทั้งหมด และใช้คำสั่ง

```
perl $ScriptDir/$Fulltripl $ConfigDir/$Dict > temp
cat $ConfigDir/$TriLst >> temp
echo "sil" >> temp
sort -u temp > $ConfigDir/$FullLst
rm -rf temp
```

ผลของคำสั่งนี้จะได้ไฟล์ $ConfigDir/$FullLst

10. ต่อมาทำการ copy Triphone HMM ไปเป็น Tied-state HMM ด้วยคำสั่ง

```
HHEd -T 1 -H $ModelDir/$SourceDir/newMacros -M $ModelDir/$TargetDir \
        $ConfigDir/$TieHed $ConfigDir/$TriphnLst
```

ซึ่งจะต้องใช้ไฟล์ $ConfigDir/$TieHed นี้ด้วย ซึ่งเป็นไฟล์ที่กำหนดว่า เสียง Phone ใดจัดอยู่ในกลุ่มเดียวกัน ผลจากการ Tie จะได้เป็น HMM ใหม่ในไฟล์ $ModelDir/$SourceDir/newMacros

11. เมื่อได้ Tied-state HMM มาแล้วก็ทำการ Re-estimate ใหม่ 3 รอบด้วยคำสั่ง

```
HERest -T 1 -A -D -C $ConfigDir/$TrainCfg $Flags \
        -I $ConfigDir/$TriphnMlf -S $TrainScp \
        -H $ModelDir/$TargetDir$j/newMacros \
        -M $ModelDir/$TargetDir$i $ConfigDir/$TieLst
```

สุดท้ายแล้วจะได้ $ModelDir/$TargetDir$j/newMacros ซึ่ง HMM model ที่ได้นี้สามารถนำไปใช้งานได้เลย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.2 หลักการสร้างโมเดลภาษา (Language model)

การสร้าง Language model มีวิธีหลักๆ 2 วิธี วิธีแรกคือการเขียนกฎหรือ Rule-based model และอีกวิธีคือการเรียนรู้จากฐานข้อมูลหรือ Statistical model สำหรับงานวิจัยนี้เป็นการรู้จำเสียงพูดคำสั่งของเมาส์ ที่มีเพียงไม่กี่คำ และแต่ละคำสามารถพูดขึ้นก่อนหลังใครก็ได้ จึงสามารถเขียนด้วยกฎง่ายๆได้ ไม่จำเป็นต้องเรียนรู้จากฐานข้อมูล ซึ่งใน HTK toolkit นั้นเราสามารถเขียนกฎของ Language model ได้โดยสร้างไฟล์ /config/dgs.gram โดยตัวอย่างเช่น การรู้จำเสียงพูดตัวเลขต่อเนื่องศูนย์ถึงเก้านั้น เราสามารถทำได้ดังรูปที่ 5-10

```
$digit = nvng1|s@@ng4|saam4|siil|haa2|
         hok1|cet1|pxxt1|kaaw2|suun4;
( SENT-START (<$digit>) SENT-END )
```

รูปที่ 5-10 แสดง dgs.gram

โดยในส่วนของบรรทัดแรกจะบอกว่ามีคำอะไรเกิดขึ้นได้บ้าง ในตัวอย่างนี้มีสิบคำคือ ศูนย์ถึงเก้า (suun4 ถึง kaw2) และส่วนต่อมาเป็นการกำหนดไวยากรณ์ของประโยคที่เกิดขึ้นได้ โดยเริ่มที่ SENT-START คือเริ่มประโยค และจบด้วย SENT-END คือสิ้นสุดประโยค ภายในประโยคประกอบด้วย &digit ก็คือตัวเลข ที่กำหนดในบรรทัดแรก คร่อมด้วย <> หมายถึงต่อกันกี่ตัวก็ได้

หลักจากสร้างไฟล์ dgs.gram แล้ว จะต้องทำการแปลงไฟล์นี้ให้อยู่ในรูปแบบของ Word network ที่ใช้ในการรู้จำใน HTK ซึ่งการแปลงนี้ทำได้โดยใช้คำสั่ง

```
HParse $ConfigDir/$Gram $LmDir/$Wdnet
```

ผลของคำสั่งจะทำการแปลง dgs.gram เป็น dgs.wdnet ซึ่งใช้ในการรู้จำเสียงต่อไป

5.3.3 การรู้จำเสียง

สามารถทำได้โดยใช้คำสั่ง

```
HVite -H am/tiehmm2m_3/newMacros \
      -S test.scp -l '*' -w lm/dgs.wdnet \
      -i output.mlf config/dgs.dict config/tie.list
```

คำสั่งนี้ระบบจะทำการโหลด HMM macro file มาจาก am/tiehmm2m_3/newMacros และทำการรับเอาไฟล์ .mfc ที่ลิสต์ไว้ใน test.scp แล้วทำการ recognize จาก Word network ใน lm/dgs.wdnet ทีละไฟล์และเก็บผลการรู้จำไว้ที่ไฟล์ output.mlf ที่สร้างขึ้นมาใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
แนวคิดการออกแบบ

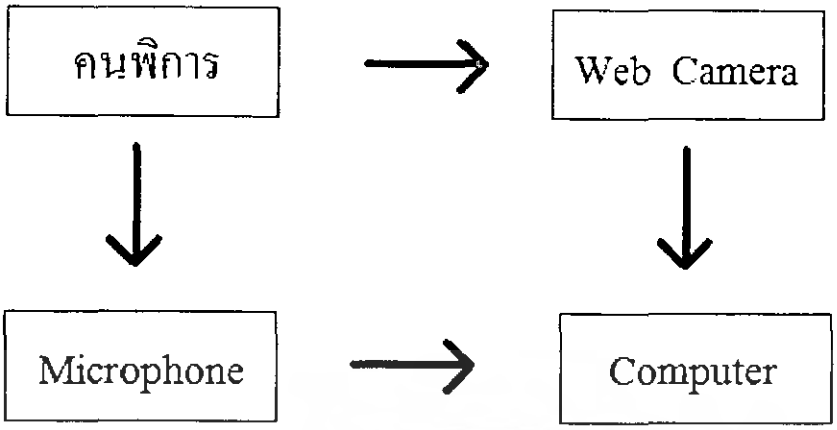
6.1 Funtional requirements

- การเคลื่อนที่ของเคอร์เซอร์เมาส์ (Cursor Mouse Move)
- คลิกซ้าย (Mouse Left Click)
- คลิกขวา (Mouse Right Click)
- ดับเบิ้ลคลิก (Mouse Double Click)
- ลากและปล่อย (Drag and Drop)

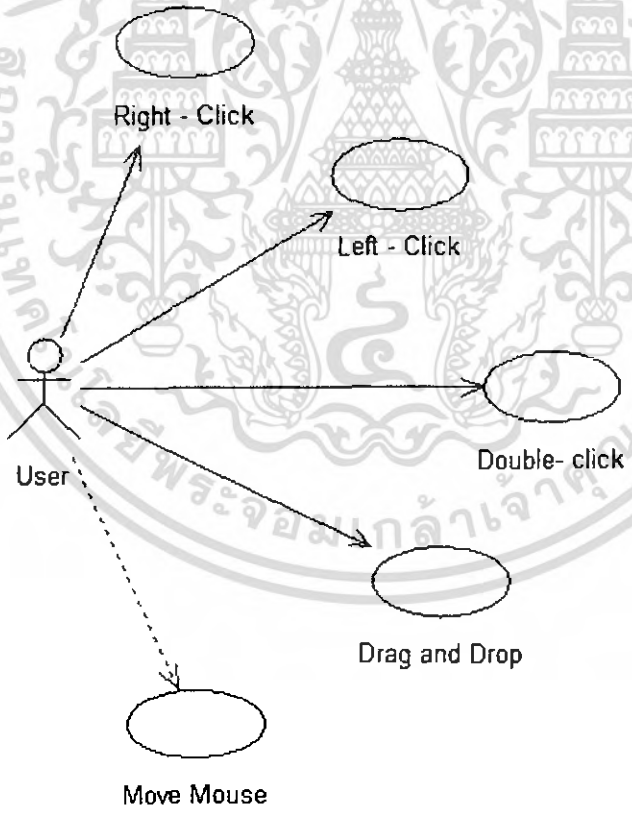
6.2 Non-funtional requirements

- ความเร็ว
- ความแม่นยำ
- ใช้ง่าย
- รองรับผู้ใช้ให้ได้มากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

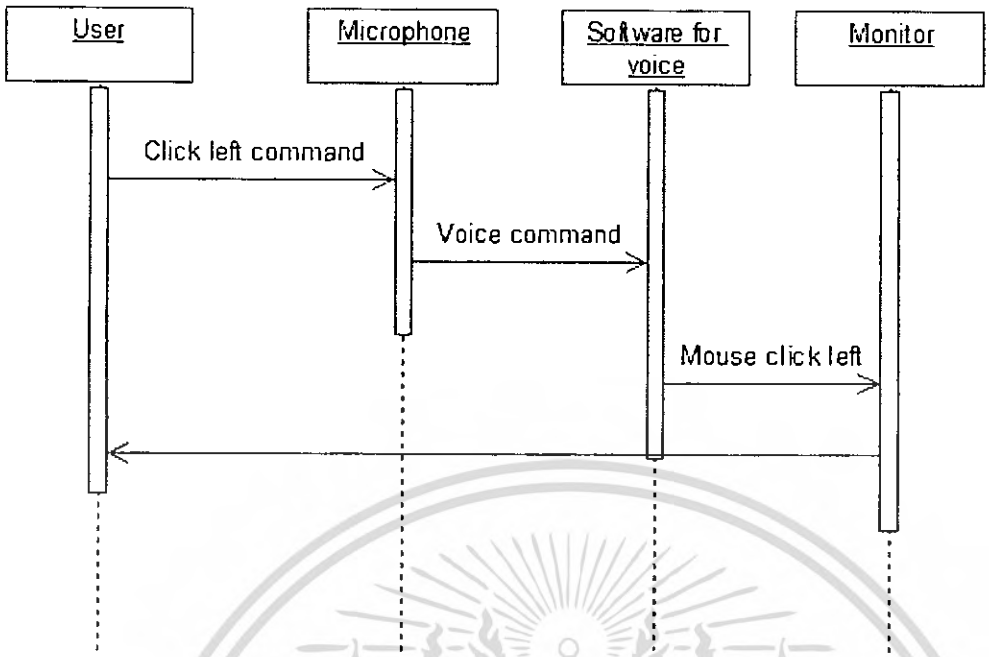


รูปที่ 6-1 การออกแบบ High-level

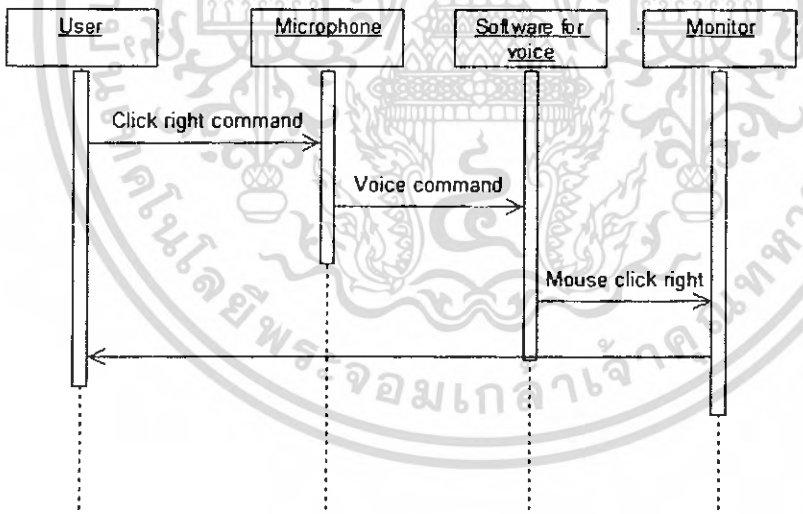


รูปที่ 6-2 Use case diagram

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

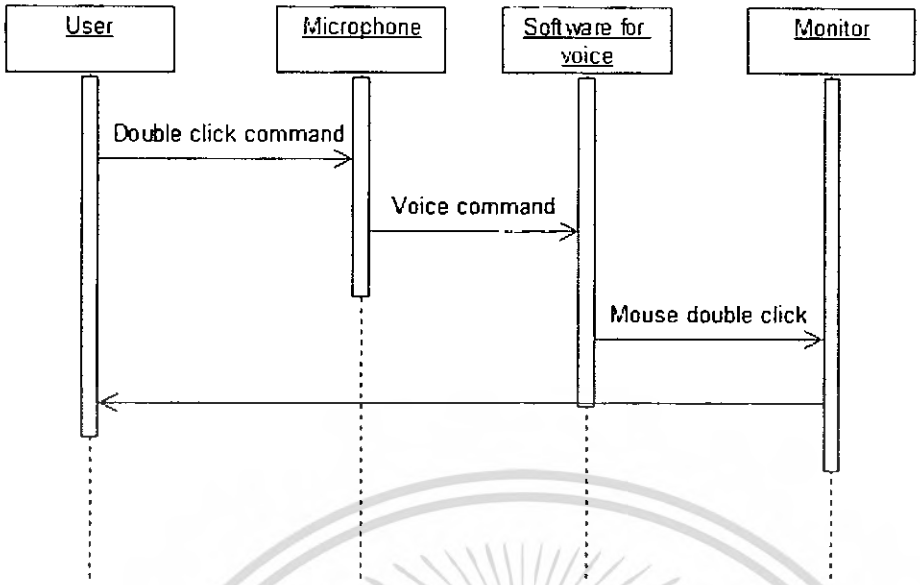


รูปที่ 6-3 แสดง Sequence diagram ของการ click left

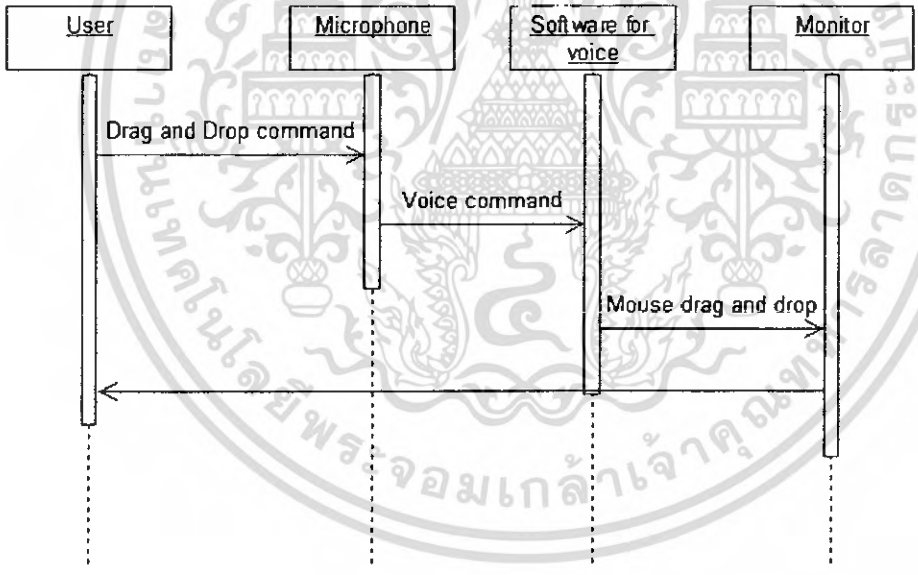


รูปที่ 6-4 แสดง Sequence diagram ของการ click right

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

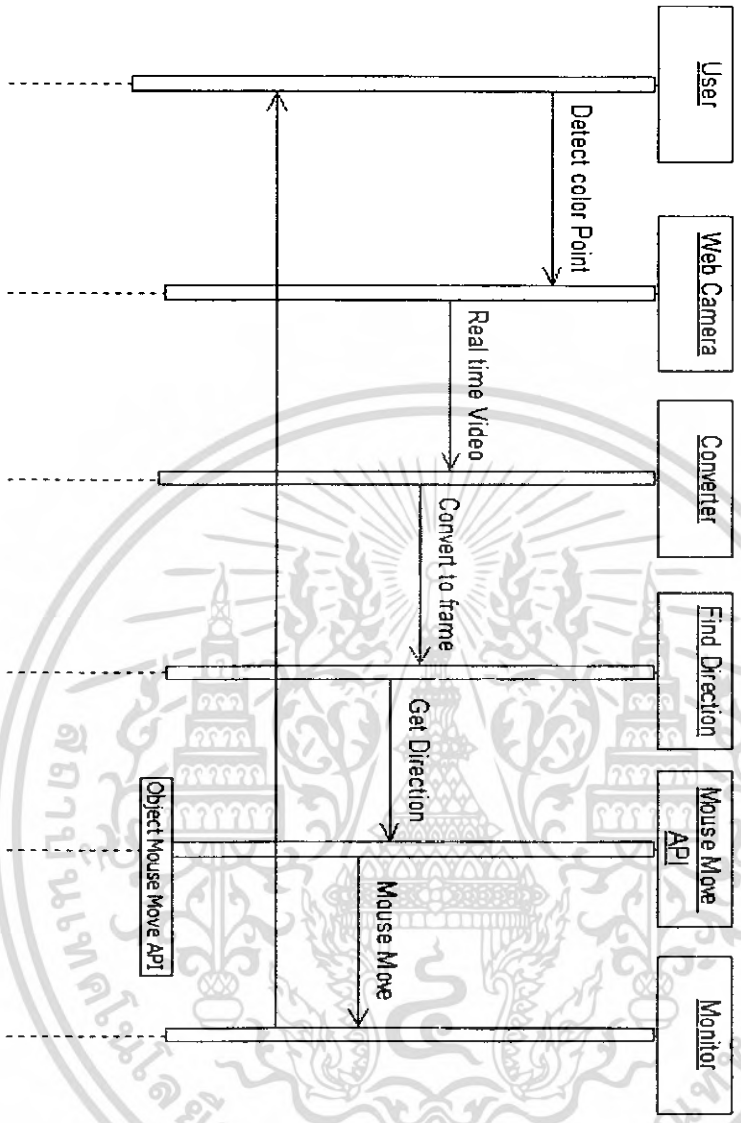


รูปที่ 6-5 แสดง Sequence diagram ของการ double click



รูปที่ 6-6 แสดง Sequence diagram ของการ drag and drop mouse

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 6-7 แสดง Sequence diagram ของ mouse move

6.3 อัลกอริทึม 1

ในส่วนของอัลกอริทึมนี้จะใช้ในการค้นหาจุดอ้างอิงที่ติดอยู่บนใบหน้าโดยขั้นตอนในการทำงานมีดังนี้

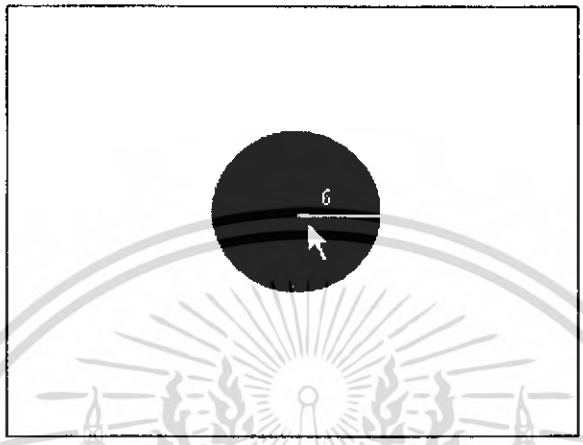
1) เริ่มแรกเราจะใช้เมาส์คลิกเพื่อกำหนดตำแหน่งของจุดอ้างอิง

2) จากนั้นจะทำการสแกนออกจากจุดศูนย์กลาง เพื่อหาขอบบน ขอบล่าง ขอบซ้าย ขอบขวาของจุดอ้างอิง และจะทำการตีกรอบวงกลมรัศมีไม่เกิน 6 พิกเซลล้อมรอบจุดอ้างอิงไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) ทำการวิเคราะห์หาตำแหน่งและทิศทางการเคลื่อนที่ของจุดอ้างอิง โดยจะทำในลักษณะที่ว่า เมื่อจุดอ้างอิงเคลื่อนที่ไปเท่าไรในขนาดหน้าต่าง 320 * 240 ก็จะนำไประบุเป็นตำแหน่งของเมาส์บนหน้าจอภาพ ว่าเมาส์จะเคลื่อนที่ไปได้เท่าไรบนหน้าจอภาพ

4) นำผลที่ได้มาระบุเป็นตำแหน่งของเมาส์ที่บนหน้าจอภาพ



รูปที่ 6-8 แสดงการค้นหาจุดอ้างอิง

จากการทำดังกล่าวพบว่า ในการหาความสัมพันธ์ระหว่างตำแหน่งของจุดอ้างอิงและเมาส์นั้น หากต้องการเลื่อนเมาส์จากหน้าจอซ้ายสุดมายังหน้าจอขวาสุด จะต้องทำการเลื่อนตำแหน่งของจุดอ้างอิงจากซ้ายสุดมาขวาสุดเช่นกัน ทำให้อาจเกิดความยากลำบากแก่ผู้ใช้ได้ จึงได้ทำการปรับสเกลให้มากขึ้น ดังนี้

จากขั้นตอนในการวิเคราะห์หาตำแหน่งและทิศทางการเคลื่อนที่ของจุดอ้างอิงนั้น จะเปลี่ยนขนาดหน้าต่างเป็น 100 * 60 เพื่อที่ในการขยับเมาส์จากตำแหน่งหน้าจอซ้ายสุด มาตำแหน่งขวาสุด ได้โดยไม่ต้องทำการขยับจุดอ้างอิงมากเหมือนตอนแรก

6.4 อัลกอริทึม 2

เนื่องจากอัลกอริทึม 1 นั้น ในการความยากลำบากแก่ผู้ใช้ในการเลื่อนตำแหน่งของจุดอ้างอิงจากซ้ายสุดมาขวาสุด และเมาส์ยังไม่มีความนิ่งพอ อัลกอริทึม 2 นี้จึงได้ใช้วิธีควบคุมเมาส์แบบจอยสติกแทน โดยขั้นตอนในการทำงานมีดังนี้

1) เริ่มแรกเราจะใช้เมาส์คลิกเพื่อกำหนดตำแหน่งของจุดอ้างอิง

2) จากนั้นจะทำการสแกนออกจากจุดศูนย์กลาง เพื่อหาขอบบน ขอบล่าง ขอบซ้าย ขอบขวาของจุดอ้างอิง และจะทำการตีกรอบสี่เหลี่ยมขนาดเพิ่มจากขอบของจุดอ้างอิงออกไปข้างละ 10 พิกเซล เพื่อไว้เป็นกรอบสำหรับจอยสติก

3) ทำการวิเคราะห์หาตำแหน่งและทิศทางการเคลื่อนที่ของจุดอ้างอิง โดยกำหนดลักษณะที่ว่า เมื่อขอบของจุดอ้างอิงเคลื่อนที่ไปติดกับกรอบในตำแหน่งซ้าย ขวา บน ล่าง ตำแหน่งเมาส์ก็จะเคลื่อนที่ไปในทิศทางซ้าย ขวา บน ล่างด้วยตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทำดังกล่าวพบว่าในการเลื่อนเมาส์จากหน้าจอซ้ายสุดมายังหน้าจอขวาสุดนั้น ทำได้ง่ายขึ้น และเมาส์มีความนิ่งมากขึ้น แต่ปัญหาที่พบต่อมาพบว่า ในขณะที่ควบคุมเมาส์นั้น ผู้ใช้จะมองไม่เห็นกรอบจอยสติก จึงได้ทำหน้าต่างPop Upของกรอบจอยสติกขึ้นมาที่มุมขวาบนของหน้าจอ

รูปที่ 6-9 แสดงหน้าต่างแสดงกรอบจอยสติก

รูปที่ 6-10 แสดงหน้าต่าง Pop up ของกรอบจอยสติก

เมื่อผู้ใช้ทำการเลื่อนเมาส์ไปยังตำแหน่งของหน้าต่าง Pop up ของกรอบจอยสติกที่อยู่ที่ตำแหน่งมุมขวาบนของหน้าจอนั้น หน้าต่าง Pop up ก็จะย้ายไปอยู่ที่ตำแหน่งมุมซ้ายบนของหน้าจอแทน เพื่อไม่ให้เกิดการบังการใช้งานที่ตำแหน่งมุมขวาบนของหน้าจอของเมาส์

6.5 การใช้เสียงควบคุมโปรแกรม

ในการสั่งงานโปรแกรมจะเป็นการสั่งงานด้วยเสียง โดยการรับเสียงจะมีตัว Timer 2 ตัว ผลัดกันทำงาน โดยเมื่อตัวแรกทำการบันทึกไฟล์เสียงเข้ามาเพื่อ recognition ตามเวลาที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดแล้ว ตัวที่สองก็จะทำการบันทึกไฟล์เสียงแทนก่อนที่จะนำมา recognition โดยผลัดกันทำเช่นนี้ไปเรื่อยๆ

โปรแกรมที่ใช้ในการบันทึกไฟล์เสียง คือ WaveEx ActiveX Control โดยใช้คำสั่ง

WaveExCtrl1.Record(filename)

เมื่อใช้คำสั่งนี้ โปรแกรมจะเริ่มทำการบันทึกไฟล์เสียง โดยในโครงงานนี้กำหนดให้การบันทึกเสียงเป็นไฟล์ .wav แบบ 16 kHz,16 bit,Mono และเก็บไฟล์ในรูปแบบของ PCM wav เมื่อบันทึกแล้วจะเก็บไฟล์โดยใช้ชื่อว่า filename และทำการหยุดบันทึกเสียงโดยใช้คำสั่ง

WaveExCtrl1.RecordStop()

โปรแกรมก็จะทำการหยุดบันทึกเสียงเพื่อนำไป recognition ต่อไป

เมื่อรู้ว่าคำที่พูดเข้ามาเป็นคำอะไร ก็จะนำมาเรียกใช้ Windows API เพื่อควบคุมการทำงานของเมาส์ และไปควบคุมโปรแกรม โดยในโครงงานนี้ได้มีการนำคำมาใช้แทนคำสั่งปกติที่เคยใช้ด้วย คือ

"กด" ใช้แทนคำสั่งการคลิก "click"

"ดับ" ใช้แทนคำสั่งการดับเบิลคลิก "doubleclick"

"ลาก" ใช้แทนคำสั่งการแดรก "drag"

"ปล่อย" ใช้แทนคำสั่งการดร็อป "drop"

"ขวา" ใช้แทนคำสั่งการคลิกขวา "rightclick"

6.6 คำสั่งควบคุมการทำงานโปรแกรม

1. การคลิก click

```
Call mouse_event(MOUSEEVENTF_LEFTDOWN,0,0,0,0)
Call mouse_event(MOUSEEVENTF_LEFTUP,0,0,0,0)
```

2. การดับเบิลคลิก doubleclick

```
Call mouse_event(MOUSEEVENTF_LEFTDOWN,0,0,0,0)
Call mouse_event(MOUSEEVENTF_LEFTUP,0,0,0,0)
Call mouse_event(MOUSEEVENTF_LEFTDOWN,0,0,0,0)
Call mouse_event(MOUSEEVENTF_LEFTUP,0,0,0,0)
```

3. การแดรก drag

```
Call mouse_event(MOUSEEVENTF_LEFTDOWN,0,0,0,0)
```

4. การดร็อป drop

```
Call mouse_event(MOUSEEVENTF_LEFTUP,0,0,0,0)
```

5. การคลิกขวา rightclick

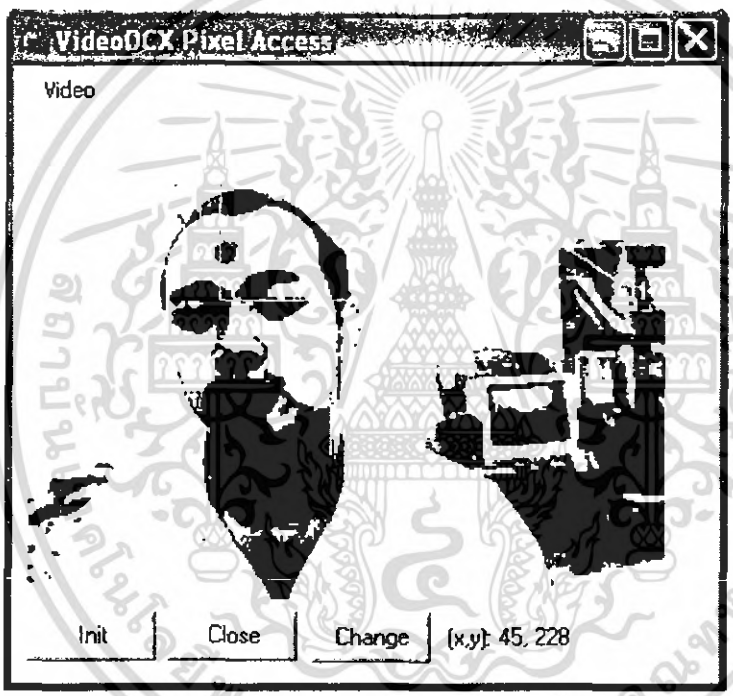
```
Call mouse_event(MOUSEEVENTF_RIGHTDOWN,0,0,0,0)
Call mouse_event(MOUSEEVENTF_RIGHTUP,0,0,0,0)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7
การทดสอบระบบ

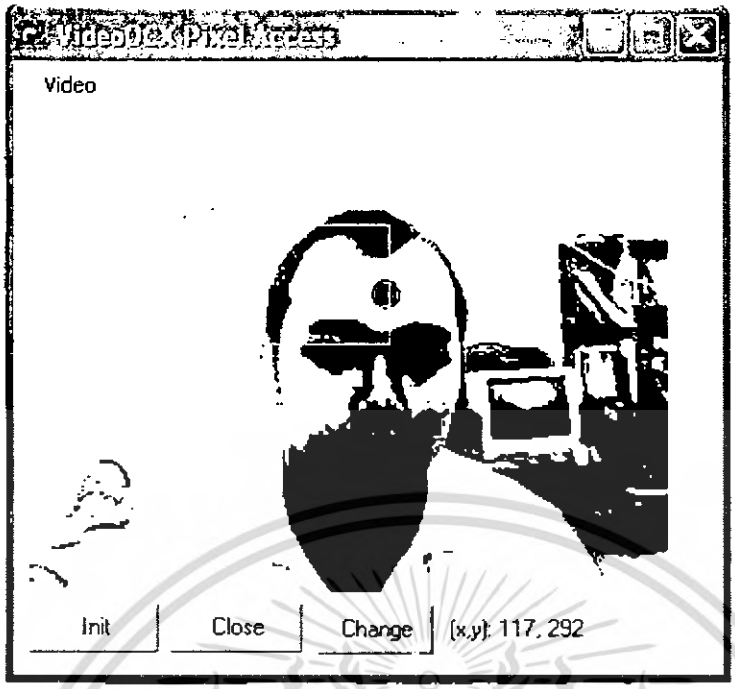
7.1 การทดลองการควบคุมเมาส์

ในส่วนของการทดลองการควบคุมเมาส์นั้น ได้มีการทดลองนำจุดอ้างอิงติดบนใบหน้า และทำการทดสอบการเลื่อนจุดอ้างอิงไปยังตำแหน่งซ้ายสุด ขวาสุด บนสุดและล่างสุด ของหน้าต่าง เพื่อจะทำการดูว่า ตำแหน่งเมาส์บนหน้าจอภาพนั้นจะเคลื่อนที่ได้ตามที่ต้องการหรือไม่

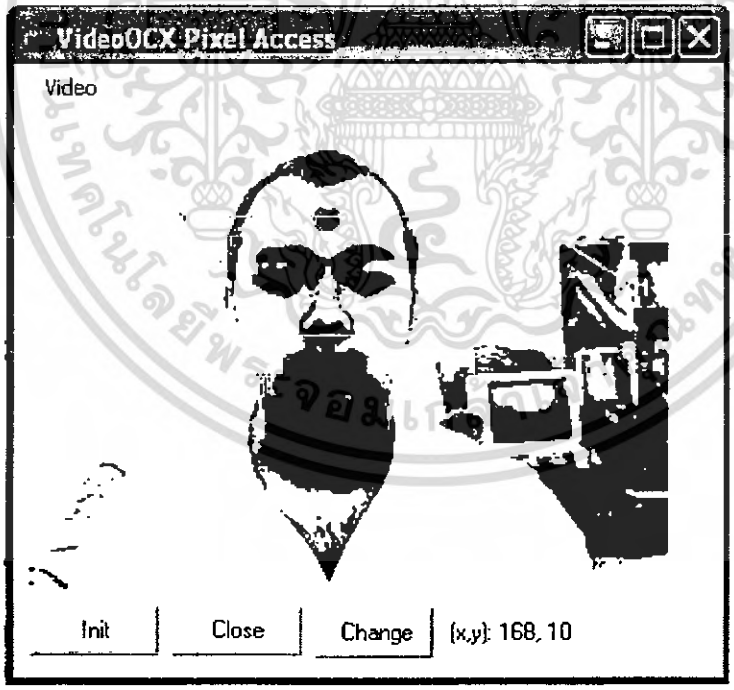


รูปที่ 7-1 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งซ้ายสุดของหน้าต่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

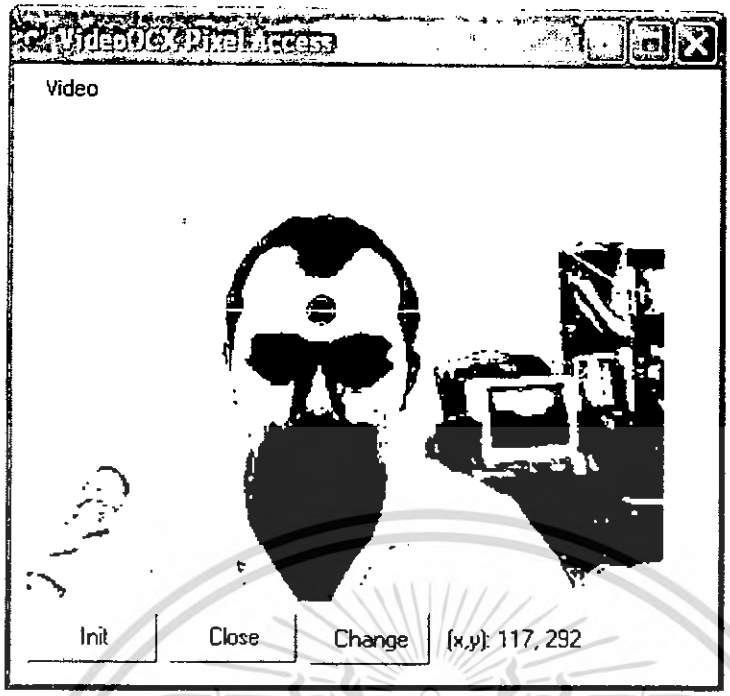


รูปที่ 7-2 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งขวาสุดของหน้าต่าง



รูปที่ 7-3 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งบนสุดของหน้าต่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7-4 แสดงการเลื่อนจุดอ้างอิงไปยังตำแหน่งล่างสุดของหน้าต่าง

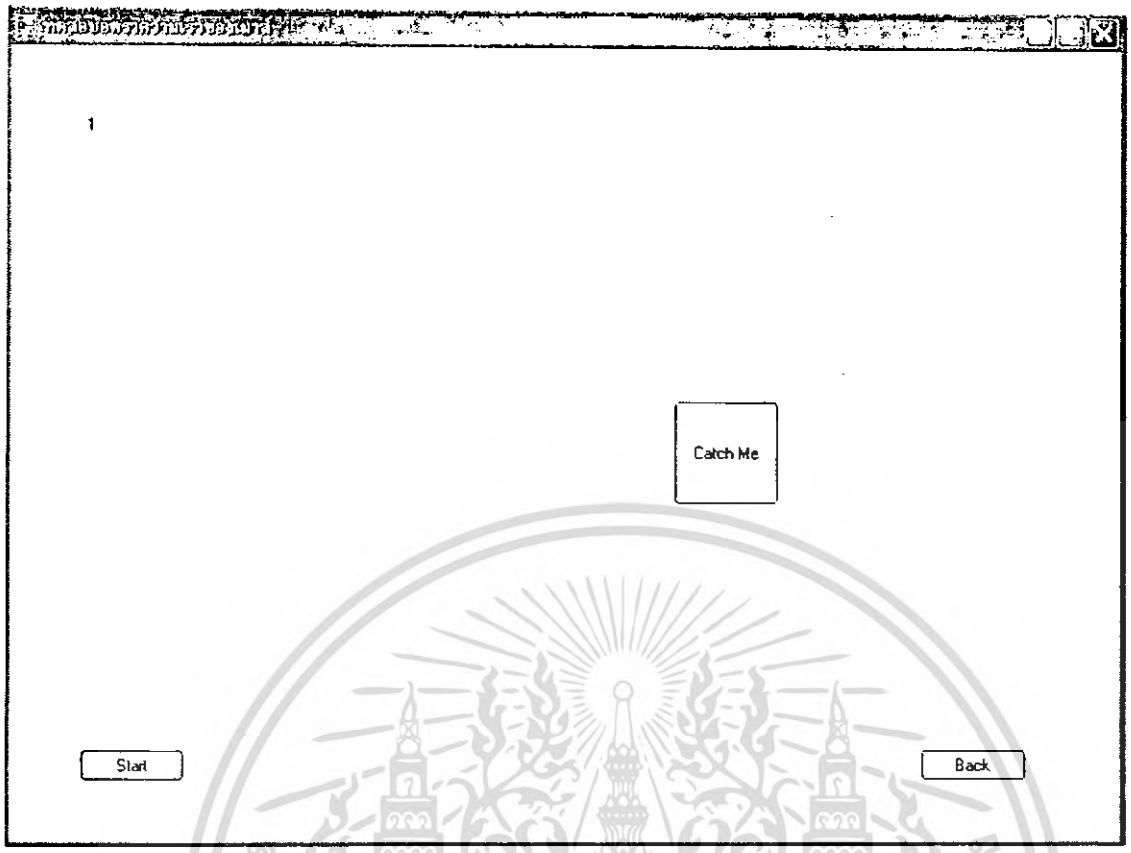
7.2 ทดสอบอัตราความเร็วของเมาส์

- สูตรคำนวณหาอัตราความเร็วของเมาส์
  อัตราความเร็วของเมาส์ = ระยะทาง / เวลาในการคลิก หน่วย pixel/Sec.
- สูตรคำนวณหาระยะทาง
  ระยะทาง = $sqt(x2-x1)^2 + (y2-y1)^2$

การออกแบบการทดลองอัตราความเร็วของเมาส์ ตามรูปที่ 7-5

1. เริ่มโดยการกดปุ่ม Start
2. จากนั้นภาพ "Catch me" จะปรากฏขึ้น
3. ให้ใช้เมาส์คลิกซ้ายที่ภาพ "Catch me"
4. ภาพ "Catch me" จะสุ่มตำแหน่งขึ้นมาใหม่เรื่อยๆ
5. เมื่อคลิกครบ 10 ครั้ง หรือ กดปุ่ม Exit จะเป็นการออกจากโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7-5 หน้าต่างการทดสอบอัตราความเร็วของเมาส์

ผลลัพธ์จากการทดสอบอัตราความเร็วเมาส์

	ค่าเฉลี่ยอัตราความเร็วของเมาส์ (Pixel/Sec)
เมาส์ปกติ	93.98
ค่าจากโครงเดิม	10.72
จุดอ้างอิง	11.73

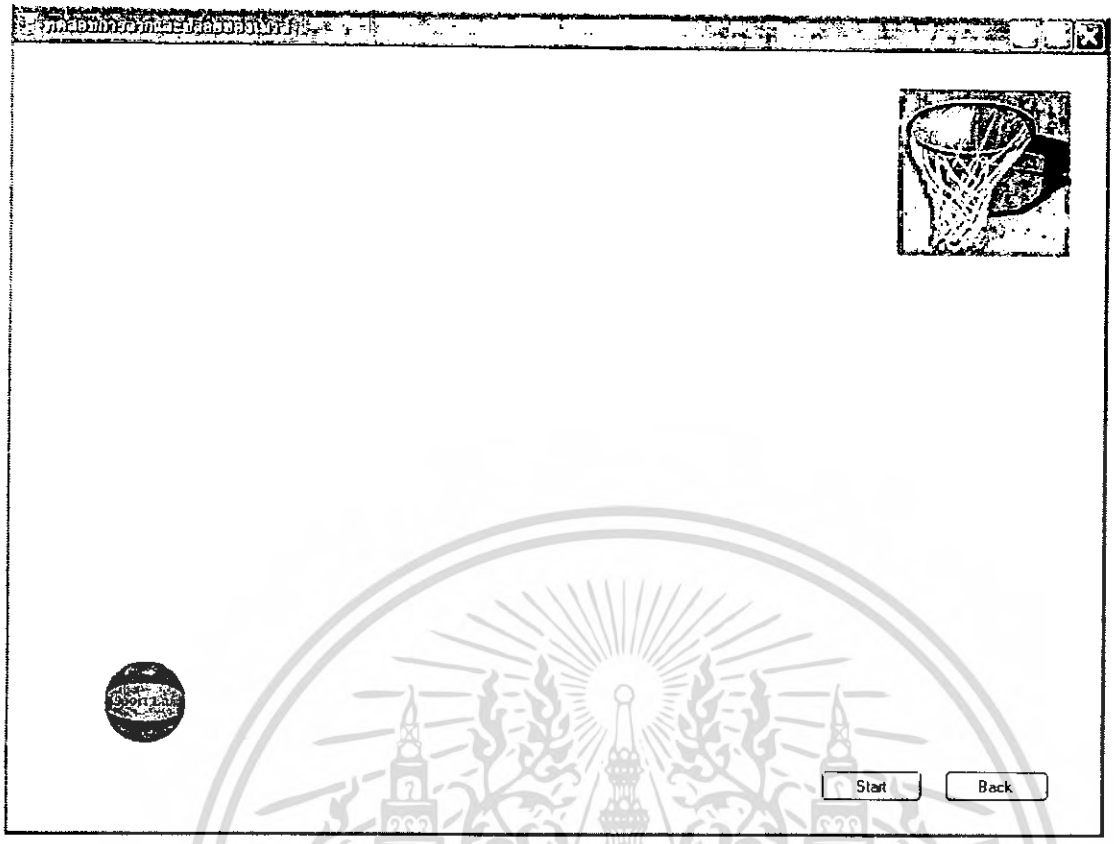
ตารางที่ 7-1 การเปรียบเทียบอัตราความเร็วของเมาส์

7.3 การทดสอบการลากและปล่อยของเมาส์

การออกแบบการทดสอบการลากและปล่อยของเมาส์ ตามรูปที่ 7-6

1. เริ่มโดยการกดปุ่ม Start
2. จากนั้นให้ใช้เมาส์ลากลูกบาสเกตบอลลงห่วง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7-6 หน้าต่างการทดสอบการลากและปล่อยเมาส์

ผลลัพธ์จากการทดสอบการลากและปล่อยของเมาส์

	เวลาที่ใช้ในการลากลูกบาสเกตบอลลงห่วง(Sec)
เมาส์ปกติ	1
ค่าจากโครงเดิม	128
จุดอ้างอิง	25

ตารางที่ 7-2 การเปรียบเทียบอัตราการลากและปล่อยของเมาส์

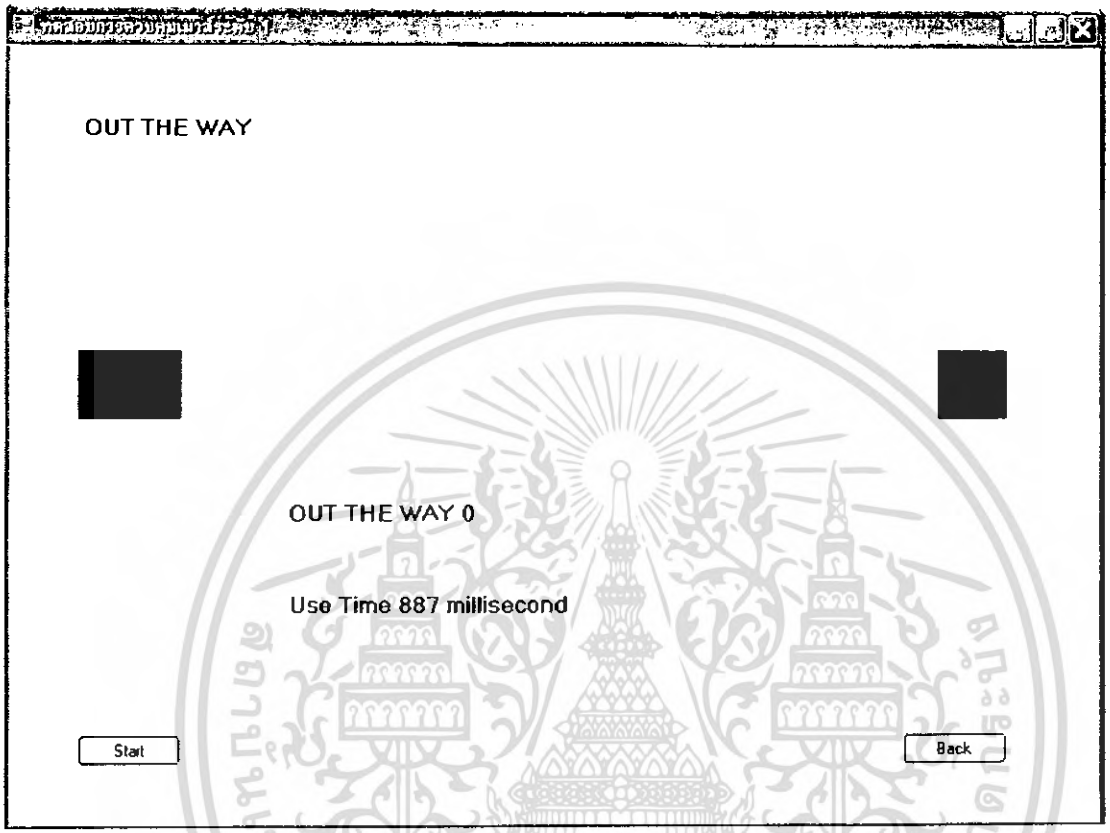
7.4 การทดสอบการควบคุมเมาส์

การออกแบบการทดสอบการควบคุมเมาส์ระดับ 1 ดามรูปที่ 7-7

1. เริ่มโดนการกดปุ่ม Start
2. จากนั้นเมาส์จะไปอยู่ที่ตำแหน่ง Start (บริเวณพื้นที่สีแดง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ให้เลื่อนเมาส์ไปตำแหน่ง END (บริเวณพื้นที่สีขาว) โดยเลื่อนให้อยู่ตามแนวที่กำหนดไว้ เมื่อเมาส์อยู่ในแนวที่กำหนด ด้านบนซ้ายของหน้าต่างจะขึ้นว่า "IN THE WAY" เมื่อเมาส์ออกจากแนวที่กำหนด ด้านบนซ้ายของหน้าต่างจะขึ้นว่า "OUT THE WAY"



รูปที่ 7-7 หน้าต่างการทดสอบการควบคุมเมาส์ระดับ 1

เมื่อเลื่อนเมาส์จากตำแหน่ง START ถึง END แอปพลิเคชันจะทำการบันทึกเวลาที่ใช้ในการเลื่อนเมาส์ และจำนวนครั้งที่เมาส์ออกนอกเส้นทาง

ผลลัพธ์จากการทดสอบการควบคุมเมาส์ระดับที่ 1

	การหลุดออกนอกเส้นทาง(ครั้ง)/เวลาที่ใช้(Sec)
เมาส์ปกติ	0/2
ค่าจากโครงเดิม	13/84
จุดอ้างอิง	0/11

ตารางที่ 7-3 การเปรียบเทียบอัตราการควบคุมเมาส์ระดับ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

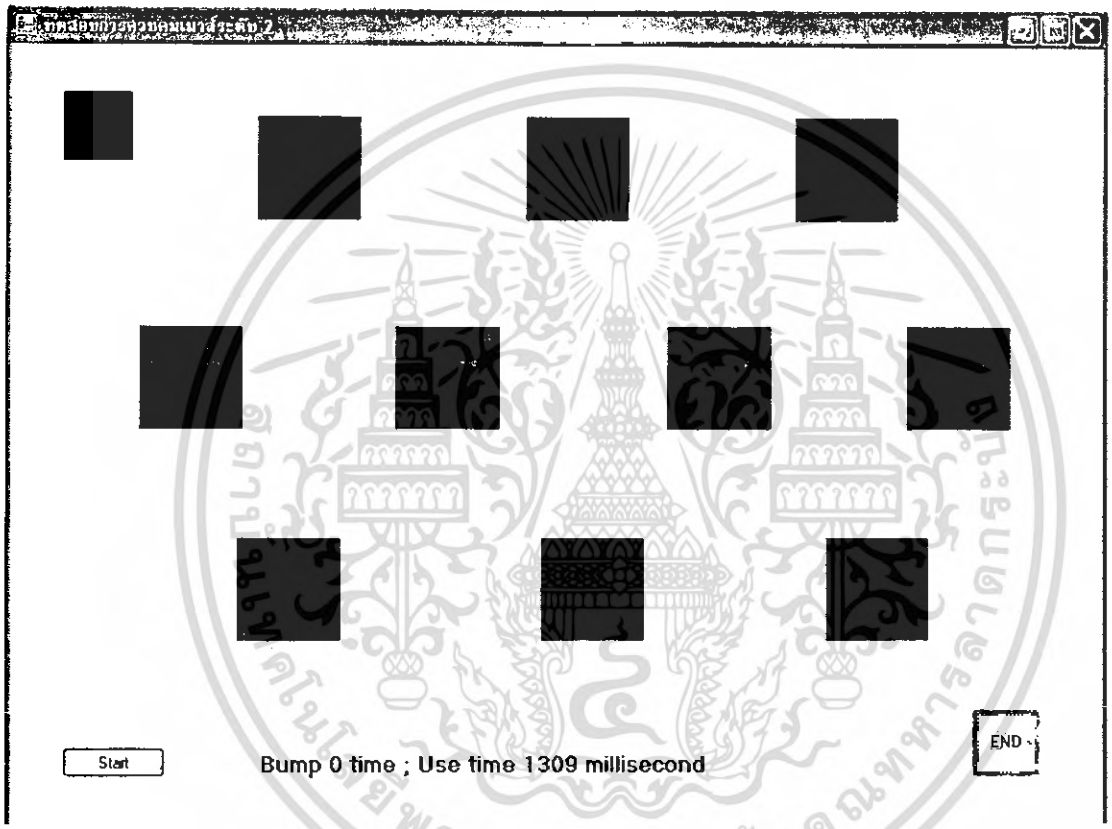
7.5 การออกแบบการทดสอบการควบคุมเมาส์ระดับ 2

ตามรูปที่ 7-8

1. เริ่มโดยกดปุ่ม Start

2. จากนั้นเมาส์จะไปอยู่ที่ตำแหน่ง Start (บริเวณพื้นที่สีขาว)

3. ให้เลื่อนเมาส์ไปตำแหน่ง END (บริเวณพื้นที่สีเหลือง) โดยเลื่อนให้อยู่ตามแนวที่กำหนดไว้ เมื่อเมาส์อยู่ในแนวที่กำหนด ด้านล่างของหน้าต่างจะขึ้นว่า "IN THE WAY" เมื่อเมาส์ชนผนังอิฐ ด้านล่างของหน้าต่างจะขึ้นว่า "BUMP"



รูปที่ 7-8 หน้าต่างการทดสอบการควบคุมเมาส์ระดับ 2

เมื่อเลื่อนเมาส์จากตำแหน่ง START ถึง END แอปพลิเคชันจะทำการบันทึกเวลาที่ใช้ในการเลื่อนเมาส์ และจำนวนครั้งที่เมาส์ชนผนังอิฐ

ผลลัพธ์จากการทดสอบการควบคุมเมาส์ระดับที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	ชนกำแพง(Bump)/เวลาที่ใช้(Sec)
เมาส์ปกติ	0/4
ค่าจากโครงเดิม	47/136
จุดอ้างอิง	2/24

ตารางที่ 7-4 การเปรียบเทียบอัตราการควบคุมเมาส์ระดับ 2

7.6 สรุปผลการทดลอง

จากการทดลองพบว่า ในอัลกอริทึม 1 อัตราการเคลื่อนที่ของเมาส์จากตำแหน่งซ้ายสุดไปขวาสุด จากตำแหน่งบนสุดมาล่างสุดนั้น เคลื่อนที่ได้ใกล้เคียงกับอัตราการเคลื่อนที่ของเมาส์ที่บังคับด้วยมือ แต่การควบคุมการเคลื่อนที่ของเมาส์ยังไม่สามารถควบคุมได้ดี เหมือนกับการควบคุมเมาส์ด้วยมือ และเมาส์ยังไม่มีความนิ่งเพียงพอ

จึงเกิดแนวคิดเปลี่ยนมาใช้อัลกอริทึม 2 โดยใช้การเคลื่อนที่ของเมาส์แบบจอยสติก พบว่าเมาส์มีความนิ่งมากขึ้น แต่ความเร็วในการเคลื่อนที่ของเมาส์จากตำแหน่งซ้ายสุดไปขวาสุดจะมีความเร็วลดลง และการตอบสนองการสั่งงานด้วยเสียงนั้นยังมีความล่าช้าอยู่บ้าง เนื่องจากต้องนำเสียงที่ได้ทำการบันทึกไปทำการ recognition ซึ่งในโครงงานนี้จะทำการบันทึกทุกๆประมาณ 1 วินาที เพราะหากใช้เวลาน้อยกว่านี้ ในบางครั้งอาจทำการบันทึกเสียงไม่ทัน โดยกรณีที่ทำให้การทำงานของเมาส์ไม่ตรงกับการสั่งงานด้วยเสียง อาจเนื่องมาจากสาเหตุการบันทึกเสียงขาด และการ recognition เกิดความผิดพลาด แสดงผลลัพธ์ออกมาผิดคำสั่ง

และในการใช้งานเมาส์ควบคุมด้วยจุดอ้างอิงและเสียงนั้น พบว่ามีเปอร์เซ็นต์ความถูกต้องในการทำงานอยู่ที่ 70 เปอร์เซ็นต์

7.7 ข้อเสนอแนะ

ในการใช้งานของจุดอ้างอิง ผู้ใช้ควรมีการฝึกใช้งานบ่อยๆ เพื่อให้เกิดความชำนาญในการใช้มากขึ้น และการเคลื่อนที่ของเมาส์ ยังมีความล่าช้าอยู่ ควรมีการเขียนโปรแกรมให้เพิ่มความเร็วในการขยับของเมาส์ได้ และในการขยับเมาส์ให้มาอยู่ตามจุดที่กำหนด ถ้ามีจุดที่ขนาดเล็กนั้น ทำได้ค่อนข้างยากในการควบคุมเมาส์ไปที่จุดนั้น จึงควรออกแบบโปรแกรมให้เมาส์เลื่อนได้ด้วยความละเอียดตามที่ผู้ใช้ต้องการ

ในส่วนของการสั่งงานด้วยเสียง อาจมีปัญหาตรงที่ว่า เมื่อสั่งงานแล้วมันไม่ตอบสนองทันที เนื่องจากต้องทำการบันทึกแล้วมาประมวลผลอีกที ดังนั้นควรจะหาเครื่องมือในการเขียนโปรแกรมที่เป็นการทำงานแบบ real time เพื่อการตอบสนองที่เร็วกว่าเดิม

ในส่วนของไมโครโฟนที่มีความไวสูงนั้น บางครั้งจะรับเอาเสียงจากภายนอกที่เป็นสัญญาณรบกวนเข้าไปด้วย ทำให้บางครั้งการตอบสนองจากการทำงานอาจเกิดความผิดพลาดได้ ในการใช้จึงควรหาสถานที่ที่มีความเงียบสงบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] สุรสิทธิ์ คิวประสพศักดิ์ - นันทนี แขวงโสภา : *"อินไซท์ Visual Basic .NET ฉบับสมบูรณ์"* , บ. โปรวิชั่น จำกัด , 2546

[2] สุทธิศักดิ์ พงศ์ธนาพานิช : *"การเขียนโปรแกรมด้วย Visual Basic 6.0 ระดับสูง(การใช้งานฟังก์ชั่นวินโดวส์ API-32 บิต)"* , บ. ไพรเมด จำกัด , 2542

[3] Rafael C. Gonzalez, Richard E. Woods : *"Digital Image Processing"*, Second Edition , Prentice Hall 2002

[4] http://www.videoocx.de/

[5] http://www.microsoft.com/speech/download/sdk51/

[6] http://web.yl.is.s.u-tokyo.ac.jp/~dunnan/virtualmouse/

[7] http://thaispeech.longdo.org/speechy/thaispeech1/index_html

[8] http://htk.eng.cam.ac.uk/

[9] http://www.cygwin.com/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้