

แฟรมเวิร์กของส่วนต่อประสานกับผู้ใช้ผ่านศึระษะมนุษย์เสมือนโดยใช้โมโนสโปกปีกพีชแห่งกีวาร์



นายอรรถกร พูนศิลป์

สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์


คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-17-3554-5

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

USER INTERFACE FRAMEWORK USING VIRTUAL HUMAN HEAD WITH MONOSCOPIC
FISH-TANK VR



Mr. Attakorn Poonsilp

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2005

ISBN 974-17-3554-5

อรรถกร พูนศิลป์ : เฟรมเวิร์กของส่วนต่อประสานกับผู้ใช้ผ่านศีรษะมนุษย์เสมือนโดยใช้
โมโนสโคปิกฟิชแทงก์วีอาร์. (USER INTERFACE FRAMEWORK USING VIRTUAL
HUMAN HEAD WITH MONOSCOPIC FISH-TANK VR) อ. ที่ปรึกษา :

อ.ดร.อรรถวิทย์ สุดแสง, 101 หน้า, ISBN 974-17-3554-5.

งานวิจัยนี้เริ่มต้นจากแนวคิดที่อยากให้การติดต่อกับคอมพิวเตอร์ง่ายเหมือนการติดต่อกับมนุษย์ด้วยกันเอง ผู้ใช้ไม่จำเป็นต้องเสียเวลาในการเรียนรู้การใช้งาน เพราะฉะนั้นจึงได้นำเสนอการออกแบบส่วนติดต่อกับผู้ใช้ ในลักษณะโมเดลสามมิติของศีรษะมนุษย์เสมือนจริงที่สามารถพูดคุยโต้ตอบโดยใช้เสียงกับผู้ใช้ได้อย่างเป็นธรรมชาติ การขยับปากสอดคล้องกับคำพูดสามารถแสดงอารมณ์ในแบบต่างๆ ผ่านระบบกล้ามเนื้อบนใบหน้า ดวงตาของโมเดลสามารถมองมายังผู้ใช้งาน รวมถึงการปรับเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในแบบเวลาจริงโดยใช้เทคโนโลยีฟิชแทงก์วีอาร์

เพื่อให้ระบบรับรู้ถึงตำแหน่งของผู้ใช้งาน เพราะฉะนั้นในการทดลองจะจับภาพผู้ใช้ผ่านกล้องวงจรปิดที่ให้ความละเอียดภาพ 320 x 240 จุด โดยระหว่างการใช้งาน ผู้ใช้ต้องสวมหมวกติดรูปแบบตรวจจับไว้ที่ศีรษะ ระบบสามารถที่จะตรวจหาตำแหน่งของศีรษะผู้ใช้โดยมีความเร็วในการประมวลผลไม่ต่ำกว่า 24.9 เฟรมต่อวินาที มีความคลาดเคลื่อนไม่เกิน 10 เซนติเมตร จำนวนเฟรมที่ตรวจจับไม่ได้มีไม่เกิน 5% ของจำนวนเฟรมทั้งหมด ส่งผลให้การจ้องมองไปยังผู้ใช้ และการปรับมุมมองให้สัมพันธ์กับตำแหน่งตาของผู้ใช้แบบฟิชแทงก์วีอาร์ สามารถทำได้อย่างถูกต้องและเสมือนจริง

สำหรับผลการพูดคุยโดยใช้เสียง โปรแกรมสามารถพูดคุยได้ตามรูปแบบประโยคสนทนาที่ได้กำหนดไว้ก่อน สามารถขยับปากสอดคล้องกับเสียงพูด และสามารถแสดงอารมณ์ได้ 7 แบบผ่านระบบกล้ามเนื้อ 18 เส้นบนใบหน้า

ระบบที่กล่าวมาสามารถนำไปประยุกต์ใช้งานได้หลากหลาย เช่น ระบบประชาสัมพันธ์อัตโนมัติ หรือระบบโฆษณาที่มีปฏิสัมพันธ์กับมนุษย์ เป็นต้น

ภาควิชา..... วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต..... ๐๕๖๐๕ พูนศิลป์
สาขาวิชา.....วิศวกรรมคอมพิวเตอร์.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา ...2548.....

4570640721 : MAJOR COMPUTER ENGINEERING

KEY WORD: 3-D VISUALIZATION / USER INTERFACE / IMAGE PROCESSING /
COMPUTER VISION / VIRTUAL REALITY

ATTAKORN POONSILP : USER INTERFACE FRAMEWORK USING VIRTUAL
HUMAN HEAD WITH MONOSCOPIC FISH-TANK VR. THESIS ADVISOR :
ATTAWITH SUDSANG, Ph.D., 101 pp. ISBN 974-17-3554-5.

This thesis proposes the design, the development and the architecture of user interface using 3D human head model. The proposed user interface is capable of interacting with the user in a natural way. The head model is featured with controllable facial expression and lip-synchronization to speech. These features, together with the integrated AI, allow human-like chat with the user. The realism is also enhanced with Fish-Tank VR technology, i.e., the head model's viewing perspective is automatically adjusted based on the observed position of the user from the attached camera.

While the system is running, the user has to wear a special cap attached with a chessboard pattern used in calculating the user's head position. The video sequences of the pattern would be taken by a webcam or a ccd camera which is generally mounted on the top of the monitor displaying the head model. In the experiment, the resolution of the video sequences is 320 x 240 pixels and effective frame rate is 24.9 fps. Preliminary results show that the error of the user's head computation is below 10 cm, and the chessboard pattern tracker run robustly. Consequently, impressive results could be acquired from Fish Tank VR and eye contact module.

In the chatting experiment, the system could chat to user with lip-synchronization to speech and make 7 types of facial expressions using 18 ways facial muscle system.

The proposed system could be used in various applications such as automatic public relation system, interactive advertisement etc.

Department..... Computer Engineering.....Student's signature.....

Field of study..... Computer Engineering.....Advisor's signature.....

Academic year ...2005.....

Attakorn Poonsilp
Attawith Sudsang

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี จากความช่วยเหลือของ อ.ดร.อรรถวิทย์ สุดแสง อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งอาจารย์ได้ให้ความช่วยเหลือเป็นอย่างมาก ช่วยประสิทธิ์ประสาทวิชาความรู้ในหลายๆ เรื่อง คอยหาเอกสารที่เกี่ยวข้องมาให้อ่าน ให้คำแนะนำต่างๆ ตลอดจนคอยดูแลกำกับการทำวิทยานิพนธ์เป็นอย่างดีและเอาใจใส่อย่างมาก คอยถามไถ่ความคืบหน้าเป็นระยะๆ ได้ให้คำปรึกษา พุดคุยในเรื่องต่างๆ ทั้งที่เกี่ยวข้องและไม่เกี่ยวกับวิทยานิพนธ์ ซึ่งเป็นประโยชน์ต่อทั้งการทำวิทยานิพนธ์และการใช้ชีวิตในสังคม

ขอขอบคุณ ผศ.ดร.วิวัฒน์ วัฒนาวุฒิ ประธานกรรมการสอบวิทยานิพนธ์ และกรรมการสอบวิทยานิพนธ์ ผศ.ประดนเดช นีละคุปต์ และ อ.ดร.เศรษฐา ปานงาม ที่ได้กรุณาสละเวลาอันมีค่าในการช่วยตรวจสอบวิทยานิพนธ์ฉบับนี้

ขอขอบคุณพี่แดงที่ให้คำแนะนำในเรื่องต่างๆ ขอขอบคุณน้องพิมพ์ที่ช่วยตรวจทานบทคัดย่อภาษาอังกฤษ พี่ที่คอยกระตุ้นเตือนให้ทำวิทยานิพนธ์ให้จบไวๆ รวมถึงคนอื่นๆ ทุกคนที่คอยถามความคืบหน้า และให้กำลังใจมาตลอด

สุดท้ายนี้ผู้วิจัยอยากจะกราบขอบพระคุณของบิดา มารดา และญาติพี่น้องที่คอยสนับสนุนผู้วิจัยด้านการศึกษาตั้งแต่วัยเยาว์ อีกทั้งยังเป็นกำลังใจและแนะนำสิ่งที่ดีให้กับผู้วิจัยเสมอมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญภาพ.....	ฌ
สารบัญตาราง.....	ฎ
บทที่	
1. บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	6
1.3 ขอบเขตของการวิจัย	6
1.4 วิธีดำเนินการวิจัย	6
1.5 ประโยชน์ที่คาดว่าจะได้รับ	7
2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง	8
2.1 งานวิจัยที่เกี่ยวข้อง	8
2.2 ทฤษฎีที่เกี่ยวข้อง.....	11
2.2.1 การประมวลผลภาพดิจิทัลเบื้องต้น.....	11
2.2.2 ระบบคอมพิวเตอร์กราฟฟิกสามมิติ	14
2.2.3 พฤติกรรมมนุษย์.....	23
2.2.4 การเทียบมาตรฐานกล้อง (Camera Calibration)	25
3. การออกแบบและรายละเอียดในการพัฒนา	27
3.1 การตรวจหาตำแหน่งของผู้ใช้เทียบกับจอภาพ	29
3.2 การปรับมุมมองแบบ Fish Tank VR.....	37
3.3 การทำให้โมเดลสบตาผู้ใช้	40
3.4 การพูดคุยโต้ตอบ	42
3.5 โมเดลสามมิติของศีรษะมนุษย์และระบบกล้องเน็อบนไบหน้า	44
3.6 การซัปปากให้สอดคล้องกับเสียงพูด (Lip-sync)	50

3.7 การแสดงอารมณ์	55
4. การทดลองและผลการทดลอง	60
4.1 ระบบที่ใช้ในการทดลอง	60
4.2 ผลการวัดความถูกต้องของการคำนวณตำแหน่งรูปแบบตรวจจับเทียบกับจอภาพ.....	61
4.3 ผลการวัดความถูกต้องของโมเดลในการมองไปยังจุดต่างๆ รอบตัวผู้ใช้	66
4.4 ผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้เมื่อไม่ใช้การแสดงผลแบบพีชแท่งกีวีอาร์.	67
4.5 ผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้เมื่อเปิดใช้การแสดงผลแบบพีชแท่งกีวีอาร์	69
4.6 ผลการทดสอบการพูดคุยสนทนา	74
4.7 สรุปผลการทดลอง.....	75
5. สรุปการวิจัย ข้อเสนอแนะ และงานวิจัยในอนาคต	79
5.1 สรุปการวิจัย	79
5.2 ข้อเสนอแนะ	80
5.3 งานวิจัยในอนาคต.....	81
รายการอ้างอิง.....	82
ภาคผนวก	85
ภาคผนวก ก โปรแกรมต้นแบบ	86
ภาคผนวก ข บทความที่นำเสนอในงานการประชุมวิชาการ	90
ประวัติผู้เขียนวิทยานิพนธ์	101

สารบัญภาพ

	หน้า
รูปที่ 1.1 อุปกรณ์ Head-mounted display	1
รูปที่ 1.2 ตัวอย่างระบบ Fish Tank VR	2
รูปที่ 1.3 แบบจำลองระบบ CAVE	3
รูปที่ 1.4 แนวคิดในการนำโมเดลศีรษะมนุษย์มาใช้ในการติดต่อกับผู้ใช้	5
รูปที่ 2.1 แสดงการทำงานของโปรแกรมพีชแทงก์วีอาร์ของ Jun Rekimoto	9
รูปที่ 2.2 โมเดลใบหน้าที่ใช้ในงานวิจัยของ Paula Salgado Lucena	10
รูปที่ 2.3 หุ่นยนต์ Kismet	10
รูปที่ 2.4 แสดงขั้นตอนพื้นฐานของการประมวลผลภาพ	12
รูปที่ 2.5 แสดงจุดภาพหรือพิกเซล	13
รูปที่ 2.6 แสดงการผสมแสงสีปฐมภูมิ	13
รูปที่ 2.7 ตัวอย่างภาพกราฟฟิคสองมิติ	14
รูปที่ 2.8 ตัวอย่างภาพกราฟฟิคสามมิติ	15
รูปที่ 2.9 แสดงการจัดเรียงของแกนในระบบมือขวาและระบบมือซ้าย	16
รูปที่ 2.10 พิกัด (x,y,z) ของสี่เหลี่ยมลูกบาศก์	16
รูปที่ 2.11 แสดงจุด เส้นตรง และโพลีกอน	17
รูปที่ 2.12 แสดงองค์ประกอบของโพลีกอนรูปสามเหลี่ยม	17
รูปที่ 2.13 แสดงการประกอบกันของโพลีกอนจนเป็นโมเดลชิ้นใหญ่	18
รูปที่ 2.14 แสดงรูปร่างวงแหวนที่สร้างจากโพลีกอนจำนวนต่างๆ กัน	18
รูปที่ 2.15 แสดงการเคลื่อนย้ายวัตถุ	19
รูปที่ 2.16 แสดงการหมุนวัตถุ	19
รูปที่ 2.17 แสดงการปรับขนาดวัตถุ	20
รูปที่ 2.18 การจัดองค์ประกอบของฉากโดยใช้การทราנסเฟอร์เมชัน	20
รูปที่ 2.19 การใช้ภาพพื้นผิวปะลงบนโมเดลสามมิติ	21
รูปที่ 2.20 แสดงผลจากการให้แสงและปะรูปพื้นผิว	21
รูปที่ 2.21 แสดงกรวยของการแสดงภาพ (Viewing Frustum)	22
รูปที่ 2.22 แสดงรูปแบบอารมณ์ต่างๆ ของใบหน้ามนุษย์	24
รูปที่ 3.1 แสดงโครงสร้างการทำงานทั้งหมดของงานวิจัย	28

รูปที่ 3.2 ตัวอย่างรูปแบบตรวจจับที่ใช้ในงานวิจัย	29
รูปที่ 3.3 แสดงการบิดเบี้ยวตามแนวรัศมี (ซ้าย) และภาพที่ปรับแก้การบิดเบี้ยวแล้ว (ขวา)	30
รูปที่ 3.4 ภาพจากกล้องที่ผ่านการปรับแก้การบิดเบี้ยวตามแนวรัศมีแล้ว	30
รูปที่ 3.5 ผลจากการแปลงเป็นภาพระดับเทา.....	31
รูปที่ 3.6 แสดงส่วนประกอบโครงสร้างที่ใช้.....	31
รูปที่ 3.7 ผลจากการขยายจุดสว่างในภาพ	32
รูปที่ 3.8 ผลจากการทำให้เป็นภาพลักษณะฐานสอง.....	32
รูปที่ 3.9 ผลจากการหาเส้นรอบรูป.....	33
รูปที่ 3.10 ผลจากการตัดเส้นรอบรูปที่มีพื้นที่น้อยกว่าที่กำหนด	33
รูปที่ 3.11 ผลจากการเลือกเฉพาะเส้นรอบรูปที่มีลักษณะนูนโค้ง	34
รูปที่ 3.12 แสดงรูปสี่เหลี่ยมที่ไม่ใช่รูปแบบตรวจจับ	34
รูปที่ 3.13 แสดงจุดมุมทั้งเจ็ดจุดที่ตรวจจับได้	35
รูปที่ 3.14 แสดงภาพผลลัพธ์จากการตรวจจับจุดมุม	36
รูปที่ 3.15 แสดงขั้นตอนทั้งหมดของกระบวนการประมวลผลภาพเพื่อระบุตำแหน่งตาของผู้ใช้ เทียบกับจอภาพ	36
รูปที่ 3.16 รูปแบบของทัศนมิติใน OpenGL	38
รูปที่ 3.17 การปรับทัศนมิติเมื่อศีรษะผู้ใช้ทำมุมต่างๆ กันกับจอภาพ	39
รูปที่ 3.18 แสดงการนำตำแหน่งศีรษะผู้ใช้อีกมากำหนดเป็นตำแหน่งของกล้อง.....	39
รูปที่ 3.19 แสดงส่วนประกอบต่างๆ ของโมเดลศีรษะมนุษย์	41
รูปที่ 3.20 แสดงมุมการหมุนของดวงตา.....	41
รูปที่ 3.21 แสดงหน้าจอการทำงานของ A.L.I.C.E Bot	43
รูปที่ 3.22 แสดงเส้นไวรฟ์เฟรมของโมเดลใบหน้าสามมิติ	44
รูปที่ 3.23 แสดงตำแหน่งกล้ามเนื้อบนใบหน้าคนและใบหน้าของโมเดล.....	45
รูปที่ 3.24 แสดงการย้ายจุดยอดของกล้ามเนื้อบนใบหน้า	46
รูปที่ 3.25 แสดงผลจากการประมาณค่าระหว่างกลางของรูปพื้นผิว	46
รูปที่ 3.26 แสดงการนำภาพพื้นผิวปะลงบนโมเดลใบหน้า (ซ้าย) และการลบเหลี่ยมโพลีกอน (ขวา)	47
รูปที่ 3.27 แสดงเวกเตอร์ปกติของผิวหน้าและจุดยอดโพลีกอน.....	47
รูปที่ 3.28 แสดงตำแหน่งทั้งหมดของกล้ามเนื้อที่ใช้ในงานวิจัยนี้	48
รูปที่ 3.29 แสดงการปรับกล้ามเนื้อเพื่อให้มีลักษณะการยิ้ม	49
รูปที่ 3.30 แสดงการปรับกล้ามเนื้อเพื่อให้มีลักษณะหน้าบึ้ง.....	50

รูปที่ 3.31 แสดงการปรับขากรรไกรเพื่อให้อ้าปาก	50
รูปที่ 3.32 ตัวอย่างปากที่สอดคล้องกับ Viseme แต่ละแบบ	52
รูปที่ 3.33 ตัวอย่างอารมณ์แบบต่างๆ ที่ใช้ในงานวิจัยนี้	56
รูปที่ 4.1 ส่วนประกอบต่างๆ ของระบบที่ใช้ในการทดลอง	60
รูปที่ 4.2 ตำแหน่ง (x,y) ที่ใช้ในการทดลอง	61
รูปที่ 4.3 แสดงผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้และไม่ใช้ระบบพีชแทงกีวัวร์	68
รูปที่ 4.4 แสดงผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้และเปิดใช้ระบบพีชแทงกีวัวร์	72
รูปที่ 4.5 แสดงภาพที่ได้จากมุมมองของผู้ใช้ เมื่อผู้ใช้เคลื่อนศีรษะไปทางซ้าย มองตรง และเคลื่อนศีรษะไปทางขวา	74
รูปที่ 4.6 แสดงการพูดคุยสนทนากับระบบ	75
รูปที่ ก.1 แสดงหน้าจอเมื่อสั่งให้ A.L.I.C.E Bot ทำงาน	87
รูปที่ ก.2 หน้าจอหลักของโปรแกรม	88

สารบัญตาราง

หน้า

ตารางที่ 4.1 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 80 เซนติเมตร.....	62
ตารางที่ 4.2 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 60 เซนติเมตร.....	63
ตารางที่ 4.3 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 40 เซนติเมตร.....	64
ตารางที่ 4.4 เปรียบเทียบพิกัดที่โมเดลมองและพิกัดที่สังเกตได้โดยคน.....	66



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ระบบคอมพิวเตอร์ตั้งแต่อดีตมาจนถึงปัจจุบันได้มีพัฒนาการอย่างมากมาย ทั้งความเร็วในการประมวลผล หน่วยความจำ รวมถึงระบบเพิ่มขยายต่างๆ เพื่อเพิ่มความสามารถให้ มีมากขึ้น แต่สำหรับส่วนต่อประสานกับผู้ใช้ (User Interface) นั้น ระบบส่วนใหญ่ยังคงกำหนดให้ ใช้แป้นพิมพ์ และเมาส์ในการป้อนข้อมูลเป็นหลัก ทำให้ในบางกรณีผู้ใช้ไม่สามารถใช้งานได้ อย่างสะดวกและเป็นธรรมชาติเท่าที่ควร เช่น กรณีของผู้ที่ไม่คุ้นเคยกับแป้นพิมพ์และเมาส์ หรือคน พิการ นอกจากนี้ เมื่อผ่านการประมวลผลแล้วระบบคอมพิวเตอร์ก็จะแสดงผลลัพธ์ออกมาใน รูปแบบตัวอักษร ภาพสองมิติ หรือภาพสามมิติที่ส่วนใหญ่ไม่มีปฏิสัมพันธ์ใดๆ กับผู้ใช้ จึงทำให้การ นำเสนอขาดความน่าสนใจ

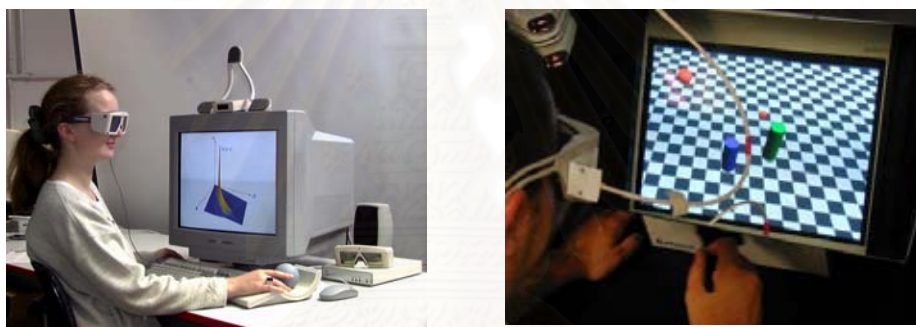
ในปี 1968 Ivan Sutherland [1] ได้พัฒนางานวิจัยต้นแบบของจอภาพแบบสวมศีรษะ (Head-mounted display) ดังรูปที่ 1.1 ซึ่งมีชื่อเสียงมาก เพราะสามารถแสดงผลโลก เสมือนจริงสามมิติที่มีปฏิสัมพันธ์กับผู้ใช้ได้อย่างเป็นธรรมชาติ ทำให้ศาสตร์ทางด้านความจริง เสมือน (Virtual Reality) มีความเป็นไปได้ในทางปฏิบัติมากขึ้น แต่อย่างไรก็ตาม อุปกรณ์นี้ยังมี อุปสรรคและข้อจำกัดในด้านการใช้งานค่อนข้างมาก เนื่องจากความละเอียดต่ำ อุปกรณ์มี ราคาแพงและหาซื้อได้ยาก นอกจากนี้ Sutherland ได้รับการยอมรับว่าเป็นคนแรกที่นำศัพท์คำว่า ความจริงเสมือน (Virtual Reality) มาใช้ ซึ่งมีความหมายว่าเป็นระบบการแสดงผลภาพสามมิติที่มี คุณภาพเพียงพอที่จะกระตุ้นการรับรู้ของผู้ใช้ให้มีความรู้สึกเสมือนว่าอยู่ในสิ่งแวดล้อมนั้นจริงๆ



รูปที่ 1.1 อุปกรณ์ Head-mounted display [2]

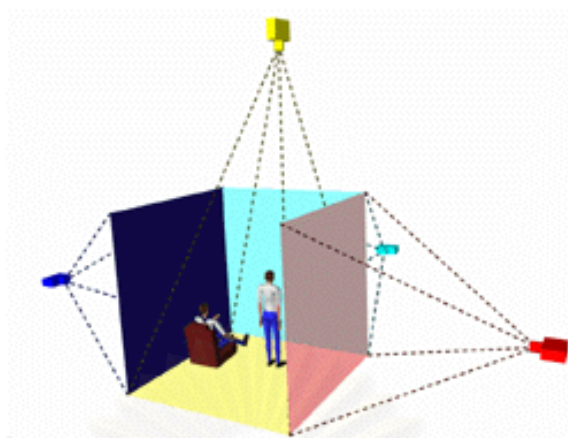
ต่อมาเมื่อต้นปี 1993 Colin Ware [3] ได้นำเสนอระบบความจริงเสมือนรูปแบบใหม่ que แสดงผลผ่านจอมอนิเตอร์ธรรมดาที่หาซื้อได้ทั่วไป โดยใช้ระบบการแสดงผลแบบมุมมองสามมิติ หรือที่เรียกว่า Stereoscopic (แสดงภาพสองภาพพร้อมๆ กัน สำหรับตาสองข้าง ใช้ร่วมกับแว่นสามมิติ) และใช้ระบบการติดตามตำแหน่งศีรษะของผู้ใช้ (Head Tracking) มาร่วมด้วย เพื่อปรับทัศนมิติ (Perspective) สำหรับการแสดงภาพสามมิติให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในขณะนั้น ระบบนี้ผู้พัฒนาได้ตั้งชื่อว่าฟิชแทงก์วีอาร์ (Fish Tank VR) เนื่องจากมีลักษณะคล้ายกับการมองเข้าไปในตู้ปลาซึ่งเปรียบได้กับโลกความจริงเสมือนนั่นเอง ดังรูปที่ 1.2

ด้วยการแสดงผลจากจอมอนิเตอร์ของคอมพิวเตอร์ตั้งโต๊ะธรรมดา ทำให้ระบบนี้มีข้อดีตรงที่มีความละเอียดและความคมชัดในการแสดงผลอย่างมาก แสดงภาพเคลื่อนไหวได้ดี ในขณะที่มีค่าใช้จ่ายน้อยมากหากเทียบกับระบบแบบเก่า อุปกรณ์ต่างๆ สามารถหาได้ทั่วไป และมีค่าใช้จ่ายในการบำรุงรักษาที่ต่ำ



รูปที่ 1.2 ตัวอย่างระบบ Fish Tank VR [4][5]

ต่อมาในเวลาไล่เลี่ยกัน เมื่อปลายปี 1993 มีนักวิจัยอีกกลุ่มได้เสนอระบบความจริงเสมือนในอีกรูปแบบหนึ่ง ตั้งชื่อว่า CAVE (CAVE Automatic Virtual Environment) [6] เป็นระบบที่ใช้เครื่องฉาย (Projector) ฉายภาพแวดล้อมสามมิติแบบสเตอริโอ (Stereoscopic) ลงบนฉากรับภาพ ที่ตั้งฉากกันเป็นรูปกล่องลูกบาศก์เปรียบเสมือนผนังของห้อง มีขนาดเท่ากับห้องๆ หนึ่ง สามารถรองรับได้หลายคน โดยแต่ละคนต้องสวมแว่นตาสามมิติ และสามารถเดินไปเดินมาได้อย่างอิสระภายในห้อง ซึ่งมีระบบตรวจจับตำแหน่งของศีรษะเพื่อปรับภาพสามมิติให้สอดคล้องกับมุมมองของผู้นำกลุ่มอีกด้วย แบบจำลองระบบสามารถแสดงได้ดังรูปที่ 1.3



รูปที่ 1.3 แบบจำลองระบบ CAVE [7]

ระบบ CAVE มีข้อดีเพิ่มขึ้นมาหลายอย่าง เช่น มีมุมมองที่กว้างขึ้น มีความละเอียดสูงกว่าจอภาพแบบสวมศีรษะหรือ Head-mounted display (ความละเอียดเทียบเคียงได้กับ HDTV หรือสองเท่าของ HDTV) รองรับผู้ใช้ได้หลายคน และเนื่องจากว่าฉากรับภาพแบนราบ รวมถึงมีระบบแก้ไขทางเรขาคณิตของภาพ (Geometric Correction) จึงทำให้ภาพที่ได้มีการบิดเบี้ยว (Distortion) ที่น้อยมากถ้าเทียบกับระบบอื่นๆ

จากลักษณะต่างๆ ของระบบความจริงเสมือนที่กล่าวมา Demiralp และ Laidlaw [8] ได้ทำการทดลองเปรียบเทียบข้อดีข้อเสีย ของระบบ CAVE เทียบกับ Fish Tank VR ในแง่ของประโยชน์ในการใช้งานจริง โดยมีผู้เข้าทำการทดสอบที่มาจากหลากหลายสาขาวิชา ทดลองใช้โปรแกรมสามมิติในสาขานั้นๆ แล้ววัดความพึงพอใจ ผลที่ได้คือผู้ใช้งานสามในสี่พึงพอใจกับระบบ Fish Tank VR มากกว่า ด้วยเหตุผลดังนี้

1. Fish Tank VR ให้ภาพที่คมชัดและสว่างมากกว่าระบบ CAVE
2. Fish Tank VR มีพื้นที่การแสดงผลเล็กกว่า จึงมองเห็นองค์รวมและความสัมพันธ์ระหว่างวัตถุต่างๆ ได้ง่ายกว่า
3. Fish Tank VR ให้ความสะดวกสบายในการใช้งานมากกว่า ไม่จำเป็นต้องยืนและเดินไปเดินมาเหมือนระบบ CAVE

จากสิ่งที่ได้กล่าวมาทั้งหมด ทำให้ระบบความจริงเสมือนบนพีซีธรรมดาที่สามารถสร้างสภาพแวดล้อมที่น่าประทับใจและเป็นธรรมชาติแก่ผู้ใช้นั้นมีความเป็นไปได้ ผู้วิจัยจึงเกิด

แนวคิดที่ว่า น่าจะมีระบบติดต่อกับผู้ใช้ (User Interface) แนวใหม่ที่แสดงผลอยู่บนพื้นฐานของ Fish Tank VR เพราะจะทำให้ได้เปรียบระบบติดต่อกับผู้ใช้แบบเก่าซึ่งเป็นแบบ 2 มิติ หรือ 3 มิติ ทั่วๆ ไปอยู่ค่อนข้างมาก

เมื่อพิจารณาถึงระบบติดต่อกับผู้ใช้แบบดั้งเดิมที่ประกอบไปด้วยเมนูคำสั่ง ทูลบาร์ต่างๆ ผู้ใช้ส่วนใหญ่ที่ยังไม่คุ้นเคยกับคอมพิวเตอร์หรือไม่เคยใช้คอมพิวเตอร์มาก่อนมักจะประสบปัญหาในการใช้โปรแกรม ผู้ใช้บางคนอาจจะไม่เข้าใจว่าสิ่งต่างๆ ศัพท์เทคนิคต่างๆ ที่ปรากฏบนหน้าจอคืออะไร และมีหน้าที่อะไร บางคนก็อาจจะไม่คุ้นเคยกับการใช้คีย์บอร์ด ทำให้ล่าช้าในการป้อนข้อมูล สิ่งต่างๆ เหล่านี้เป็นอุปสรรคที่ส่งผลทำให้ต้องใช้เวลาพอสมควรในการเรียนรู้ว่าจะใช้งานได้

กรณีหนึ่งที่ได้เห็นได้ชัดคือโครงการต่างๆ ที่จัดซื้อคอมพิวเตอร์ให้กับโรงเรียนในชนบท มีวัตถุประสงค์ให้เด็กนักเรียนในชนบทได้มีโอกาสเรียนรู้คอมพิวเตอร์และสามารถเข้าถึงข้อมูลในอินเทอร์เน็ต แต่เมื่อระบบต่างๆ ติดตั้งพร้อมแล้ว ก็เกิดปัญหาครูอาจารย์ใช้งานไม่เป็นหรือใช้งานได้เพียงแค่อ้างอิง เด็กนักเรียนก็ไม่สามารถใช้งานได้เนื่องจากอ่านยังไม่คล่อง หรือเจอสัญลักษณ์แปลกๆ ที่ไม่ทราบความหมาย ไม่ทราบว่าต้องเอาเมาส์ไปกดตรงไหนจึงจะให้คอมพิวเตอร์ทำงานได้ตามต้องการ บางคนก็อาจจะเลิกล้มความตั้งใจและท้อแท้ และเป็นที่จะใช้คอมพิวเตอร์ไปเลย

จะดีสักเพียงใด หากเปิดคอมพิวเตอร์แล้วมีใบหน้าผู้หญิงปรากฏขึ้นมาบนจอ (รูปที่ 1.4) พร้อมกับทักทายว่า

“สวัสดีจ๊ะน้อง..... วันนี้ดูสีหน้าน้อง..... ไม่ดีเลย มีอะไรไม่สบายใจหรือเปล่า เดี่ยวพี่เปิดเพลงให้ฟังนะ” แล้วโปรแกรมฟังเพลงก็ถูกเรียกขึ้นมาทำงานอัตโนมัติ

หรือ

“เมื่อวานพิมพ์งานค้างไว้เนี่ย วันนี้จะทำต่อหรือเปล่า จะได้เรียกโปรแกรมพิมพ์งานให้”

ผู้ใช้อาจสั่งงานคอมพิวเตอร์ได้ เช่น

“คอมพิวเตอร์ คุณช่วยดูรูปถ่ายใบนี้ แล้วบันทึกไว้หน่อย” จากนั้นใบหน้าบนจอก็เหลือบมองไปยังรูปถ่าย และบันทึกภาพนั้นเก็บไว้โดยผ่านกล้องที่ติดอยู่บนจอภาพ

จากสิ่งต่างๆ ที่กล่าวมา ผู้วิจัยจึงได้เกิดแนวคิดที่จะออกแบบและวางรากฐานส่วนติดต่อกับผู้ใช้บนคอมพิวเตอร์พีซีเสียใหม่ให้มีความเสมือนจริง เป็นธรรมชาติ และมีความน่าสนใจ ผู้ใช้ไม่จำเป็นต้องใช้เวลาเรียนรู้การใช้งาน เพราะผู้ใช้สามารถสื่อสารกับคอมพิวเตอร์ได้โดยไม่ต่างอะไรกับการสื่อสารกับมนุษย์ด้วยตนเอง

แต่การที่จะทำให้ระบบรับรู้ เข้าใจในภาษาพูด สีหน้าและท่าทางของผู้ใช้นั้นยังมีอุปสรรคเนื่องด้วยปัจจัยต่างๆ เช่นความเร็วในการประมวลผล เทคโนโลยีปัญญาประดิษฐ์ การประมวลผลภาษาธรรมชาติ และการประมวลผลภาพดิจิทัลที่ในปัจจุบันยังมีข้อจำกัดอยู่มาก แต่ในอนาคตเมื่อเทคโนโลยีเหล่านี้มีความก้าวหน้าเพียงพอ ก็สามารถนำมารวมกับงานวิจัยชิ้นนี้เพื่อสร้างเป็นระบบการติดต่อกับผู้ใช้ที่สมบูรณ์แบบ

เพราะฉะนั้น ในงานวิจัยนี้จึงเน้นเฉพาะส่วนของการพัฒนาโมเดลสามมิติของศีรษะมนุษย์ให้สามารถมีปฏิริยาต่างๆ ต่อผู้ใช้ได้ เช่น รับรู้สภาพแวดล้อมรอบๆ ผู้ใช้ มองไปยังผู้ใช้ ชยับปากสนทนากับผู้ใช้ รวมถึงการเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในแบบเวลาจริงโดยนำ Fish Tank VR มาประยุกต์ใช้ ซึ่งจะทำให้ผู้ใช้เกิดความรู้สึกเสมือนว่ามีศีรษะมนุษย์อยู่ภายในจอคอมพิวเตอร์จริงๆ

ด้วยส่วนติดต่อกับผู้ใช้แบบนี้ จะช่วยเพิ่มความน่าสนใจในการใช้คอมพิวเตอร์ได้อีกมาก การใช้โมเดลศีรษะมนุษย์จะมีผลต่อจิตใจทำให้ผู้ใช้รู้สึกว่าคอมพิวเตอร์เป็นมิตรมากขึ้น และสามารถเป็นเพื่อนกับเราได้ (โดยเฉพาะเด็กๆ) ทำให้ชอบที่จะใช้งาน และไม่รู้สึกรำคาญแนวคิดของงานวิจัยนี้แสดงดังรูปที่ 1.4



รูปที่ 1.4 แนวคิดในการนำโมเดลศีรษะมนุษย์มาใช้ในการติดต่อกับผู้ใช้

1.2 วัตถุประสงค์

1. เพื่อสร้างกรอบการทำงาน (Framework) สำหรับระบบติดต่อกับผู้ใช้ในลักษณะโมเดลสามมิติของศีรษะมนุษย์ที่เน้นให้ผู้ใช้งานสามารถใช้งานได้อย่างเป็นธรรมชาติ
2. เพื่อให้คอมพิวเตอร์สามารถสื่อสารกับมนุษย์ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น โดยคอมพิวเตอร์สามารถแสดงออกโดยใช้ภาษาท่าทาง เช่น การใช้สายตา การยิ้ม เป็นต้น
3. เพื่ออำนวยความสะดวกให้กับผู้ที่ต้องการสร้างระบบติดต่อกับผู้ใช้ที่ผู้ใช้สามารถใช้งานได้ง่ายและไม่จำเป็นต้องใช้เวลาในการศึกษาเรียนรู้

1.3 ขอบเขตของการวิจัย

1. ระบบที่พัฒนาขึ้นสามารถแสดงโมเดลสามมิติศีรษะมนุษย์ให้สัมพันธ์กับมุมมองของผู้ใช้ โดยใช้เทคโนโลยีพีชแทงก์วีอาร์
2. โมเดลศีรษะสามารถมองไปยังผู้ใช้ได้
3. โมเดลศีรษะสามารถแสดงอารมณ์ ได้แก่ ปกติ, ยิ้ม, เศร้า, ประหลาดใจ, หวาดกลัว, โกรธ และ สะอิดสะเอียน
4. โมเดลศีรษะสามารถขยับปากเพื่อสนทนากับผู้ใช้ตามรูปแบบของประโยคที่ได้กำหนดไว้ล่วงหน้า
5. การควบคุมโมเดลศีรษะ เช่นรูปแบบการแสดงอารมณ์ หรือรูปแบบปากขณะพูดสามารถถูกกำหนดโดยผ่านสคริปต์
6. ผู้ใช้ต้องเคลื่อนไหวอย่างจำกัดให้อยู่ภายในมุมรับภาพของกล้อง
7. ผู้ใช้ต้องสวมหมวกที่ติดรูปแบบตรวจจับตลอดเวลาที่ใช้งาน

1.4 วิธีดำเนินการวิจัย

1. ศึกษาความรู้และงานวิจัยที่เกี่ยวข้องกับพีชแทงก์วีอาร์ (Fish Tank VR) , ลักษณะของศีรษะมนุษย์ และระบบกล้องเน็อบนไบหน้า
2. พัฒนาการแสดงผลโมเดลศีรษะมนุษย์ให้มีความสามารถปรับมุมมองแบบพีชแทงก์วีอาร์
3. เพิ่มความสามารถด้านการปฏิสัมพันธ์กับผู้ใช้ให้กับโมเดล ได้แก่ แสดงอารมณ์ ขยับปาก พูดคุย และสบตาผู้ใช้
4. ประเมินผล ตรวจสอบความเที่ยงตรงและความเสมือนจริง
5. สรุป และเรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

งานวิจัยชิ้นนี้พัฒนาขึ้นเพื่อให้ผู้ใช้ได้รับความสะดวกและความรู้สึกเสมือนจริงในการติดต่อสื่อสารกับคอมพิวเตอร์ ผู้ใช้จะรู้สึกว่าไม่มีเดสคท็อปและไบนามนุษย์สามมิติ อยู่ในจอมอนิเตอร์จริงๆ ซึ่งจะแปรเปลี่ยนมุมมองไปตามตำแหน่งของผู้ใช้ และมีปฏิสัมพันธ์กับผู้ใช้ในรูปแบบต่างๆ กัน ทำให้การติดต่อกับคอมพิวเตอร์มีความน่าสนใจมากยิ่งขึ้น



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

งานวิจัยและทฤษฎีที่เกี่ยวข้อง

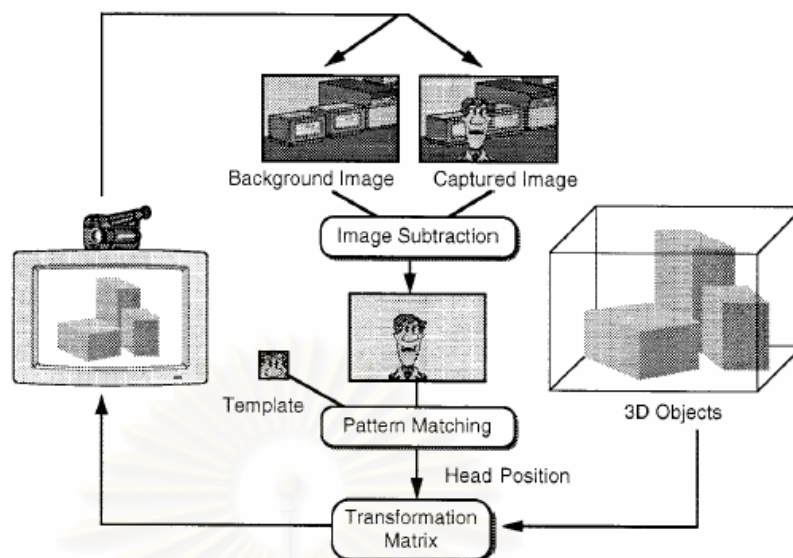
2.1 งานวิจัยที่เกี่ยวข้อง

Jun Rekimoto [9] ได้สร้างระบบความจริงเสมือนแบบฟิชแทงก์วัวร์ (Fish Tank Virtual Reality) ขึ้นมาในปี 1995 โดยใช้กล้องจับภาพผู้ใช้เพื่อติดตามตำแหน่งของศีรษะในแบบเวลาจริง แล้ววัดประสิทธิภาพในการทำงานและการแสดงผลแบบสามมิติบนจอภาพ สำหรับการประมวลผลเพื่อติดตามตำแหน่งของศีรษะ (Head Tracking) นั้น มีขั้นตอนดังนี้คือ

ค้นหาบริเวณที่เป็นใบหน้าผู้ใช้ โดยจะเก็บภาพฉากหลัง (background) ไว้ก่อน เมื่อเริ่มประมวลผลจึงลบภาพฉากหลังออกจากภาพผู้ใช้ที่ได้รับมาจากกล้องฟิชแทงก์วัวร์ เพื่อให้มีความแม่นยำที่ดีขึ้น ได้เลือกใช้ระบบสีแบบ YUV ในการประมวลผล ทำให้ได้ภาพบริเวณผู้ใช้ที่ชัดเจนถึงแม้จะมีฉากหลังที่ซับซ้อน จากนั้นระบบจะค้นหาตำแหน่งกึ่งกลางใบหน้าโดยใช้การจับคู่กับแม่แบบ (Template Matching) บริเวณบางส่วนของใบหน้าผู้ใช้จะถูกเก็บไว้ก่อนในรูปแบบของแม่แบบ เมื่อได้รับภาพบริเวณใบหน้าผู้ใช้เข้ามาในระบบจะคำนวณสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficients) ระหว่างแม่แบบกับทุกๆ บริเวณบนใบหน้า แล้วเลือกจุดที่ได้ค่าสูงที่สุดเป็นตำแหน่งกึ่งกลางใบหน้า

เมื่อได้ตำแหน่งกึ่งกลางของใบหน้าแล้ว จึงนำไปคำนวณหาตำแหน่งในสามมิติ (x, y, z) ของผู้ใช้ที่เทียบกับกล้อง โดยมีข้อกำหนดว่าผู้ใช้ต้องมีระยะห่างกับกล้องที่คงที่ หลังจากนั้นระบบจะนำพิกัดที่ได้ไปปรับทัศนมิติ (Perspective) และตำแหน่งกล้องเพื่อฉายโมเดลสามมิติให้สัมพันธ์กับตำแหน่งของผู้ใช้ การทำงานของโปรแกรมทั้งหมดแสดงได้ดังรูปที่ 2.1

ภาพโมเดลสามมิติที่ใช้ทดสอบความลึกนั้น เป็นภาพโมเดลต้นไม้ที่มีกิ่งก้านจำนวนสามต้นที่วางไว้ที่ความลึกแตกต่างกัน โดยให้ผู้ใช้ทดสอบทำการแยกแยะความลึกของต้นไม้ในส่วนต่างๆ ผลการทดลองปรากฏออกมาว่าเมื่อใช้ระบบการติดตามตำแหน่งของศีรษะเข้ามาช่วยด้วย สามารถเพิ่มความถูกต้องในการทำแบบทดสอบ โดยมีความผิดพลาดเพียงแค่ 5% เมื่อเทียบกับการไม่ใช้ตำแหน่งของศีรษะซึ่งมีความผิดพลาดสูงถึง 21.3%



รูปที่ 2.1 แสดงการทำงานของโปรแกรมพีชแทงกวีอาร์ของ Jun Rekimoto [9]

ทางผู้วิจัยที่ทำงานวิจัยชิ้นนี้มีความเชื่อว่ากล้องที่ติดอยู่บนจอภาพจะกลายเป็นอุปกรณ์รับข้อมูลเข้ามาตรฐานแบบที่สาม (นอกเหนือจากแป้นพิมพ์และเมาส์) ในอนาคตอันใกล้ เนื่องจากมีประโยชน์อย่างมหาศาลในด้านต่างๆ โดยเฉพาะศาสตร์ทางด้านการปฏิสัมพันธ์ระหว่างคอมพิวเตอร์และผู้ใช้ (Human-computer interaction)

มีคณะวิจัยอีกกลุ่มหนึ่ง ได้ทำวิจัยค้นคว้าเกี่ยวกับใบหน้าที่สามารถพูดคุยและแสดงอารมณ์ได้ นำโดย Paula Salgado Lucena [10] คณะวิจัยกลุ่มนี้ ได้พัฒนาส่วนใบหน้ามนุษย์ที่สามารถพูดได้ตามข้อความที่ป้อนเข้ามาจากคีย์บอร์ด (Talking Head) และสามารถที่จะกรอกลูกตาไปมา สายหัวไปมา และแสดงอารมณ์ สำหรับการทำงาน เริ่มแรกเมื่อผู้ใช้ป้อนข้อความตัวอักษร ตัวอักษรนั้นจะถูกส่งไปยังกลไกการสังเคราะห์เสียงพูด (Speech Synthesis Module) เพื่อแปลงข้อความตัวอักษรให้เป็นเสียง ในระหว่างนั้นก็สกัดเอาหน่วยเสียง (Phoneme) ออกมา แล้วนำไปจับคู่กับรูปแบบปากที่สัมพันธ์กับหน่วยเสียงนั้น (Viseme) ในงานวิจัยนี้ได้เตรียมรูปแบบปากไว้ถึง 16 แบบ การขยับปากให้สัมพันธ์กับหน่วยเสียงจะกระทำโดยโมดูลซิงโครไนเซชัน (Synchronization Mudule) และพฤติกรรมของใบหน้าอื่นๆ เช่น การกรอกลูกตาและสายหัวและแสดงอารมณ์จะกระทำโดยโมดูลสำหรับจัดการใบหน้า (Face Management Module) ใบหน้าที่ใช้เป็นตัวแบบรูปที่ 2.2



รูปที่ 2.2 โมเดลใบหน้าที่ใช้ในงานวิจัยของ Paula Salgado Lucena [10]

แต่อย่างไรก็ตามในงานวิจัยนี้เป็นโมเดลศีรษะมนุษย์ที่พูดตามสคริปต์เท่านั้น ไม่สามารถพูดคุยโต้ตอบ ไม่สามารถสบตาและรับรู้ความเป็นไปของผู้ใช้รวมถึงสภาวะแวดล้อมได้ ทำให้มีข้อจำกัดทางด้านความสมจริงอย่างมาก

นอกจากนี้ ยังมีงานวิจัยอีกชิ้นหนึ่งที่คล้ายคลึงกับงานวิจัยที่จะนำเสนอ งานวิจัยชิ้นนี้เป็นการสร้างหุ่นยนต์ (เฉพาะส่วนของศีรษะและใบหน้า) สามารถแสดงออกทางอารมณ์บนใบหน้า มีปฏิสัมพันธ์กับผู้ใช้ได้ในแบบต่างๆ กัน จุดมุ่งหมายหลักคือต้องการให้เป็นหุ่นยนต์ที่สามารถสื่อสารกับมนุษย์ได้อย่างเป็นธรรมชาติมากที่สุด และสามารถเป็นเพื่อนเล่นกับมนุษย์ได้ พัฒนาโดย Cynthia Breazeal [11][12] ซึ่งหุ่นยนต์มีชื่อว่า “Kismet” มีลักษณะดังรูปที่ 2.3



รูปที่ 2.3 หุ่นยนต์ Kismet [11][12]

มนุษย์สามารถสื่อสารกับ Kismet ได้โดยใช้เสียงพูดและท่าทางประกอบ เพราะฉะนั้นผู้ใช้งานไม่จำเป็นต้องเรียนรู้วิธีการใช้ใดๆ ก็สามารถสื่อสารกับ Kismet ได้อย่างมีประสิทธิภาพ

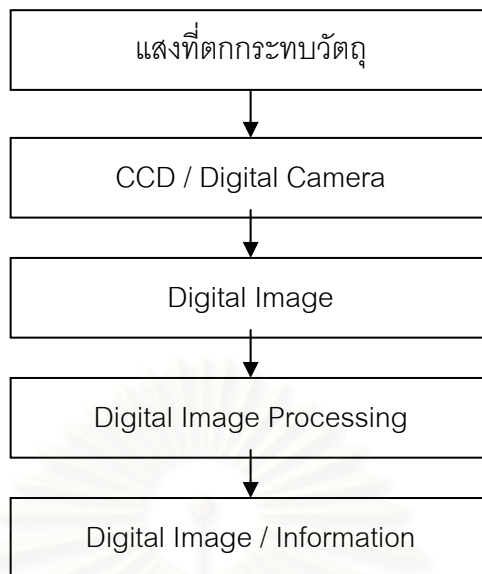
สำหรับการออกแบบ หุ่น Kismet มีกล้องและไมโครโฟนอย่างละสองตัว (Stereo) , มีระบบการสังเคราะห์เสียงพูด (Speech Synthesizer) , และอวัยวะบนใบหน้าเพื่อแสดงอารมณ์ ในแบบต่างๆ กัน ซึ่งจะมีลักษณะการแสดงออกทางอารมณ์และความสามารถเทียบได้กับเด็กทารกอายุหกเดือน แต่การสร้างเป็นหุ่นยนต์กลไกแบบนี้จะมีต้นทุนการสร้างที่สูง ไม่สามารถทำให้ดูเหมือนคนจริงได้เท่ากับระบบคอมพิวเตอร์กราฟฟิก และมีกลุ่มเป้าหมายที่จำกัด

2.2 ทฤษฎีที่เกี่ยวข้อง

2.2.1 การประมวลผลภาพดิจิทัลเบื้องต้น

ก่อนที่จะเข้าสู่การทำกระบวนการของภาพดิจิทัลเราจะเริ่มต้นที่ภาพธรรมดาก่อน โดยปกติแล้วสายตาของบุคคลทั่วไปจะมองเห็นภาพทิวทัศน์วิวต่างๆ เป็นลักษณะแบบอนาล็อก ซึ่งสามารถอธิบายได้ด้วยคณิตศาสตร์ที่มีตัวแปรแบบนับได้อย่างต่อเนื่อง แต่เครื่องคอมพิวเตอร์จะใช้เลขฐานสองเป็นหลักในการคำนวณ เมื่อนำภาพมาแปลงเข้าสู่เครื่องคอมพิวเตอร์ ภาพนั้นก็จะเป็นกลายเป็นภาพดิจิทัล (Digital Image) ซึ่งจะถูกเก็บไว้ในรูปแบบของเลขฐานสองทั้งหมด

การรับภาพเข้าสู่คอมพิวเตอร์สามารถทำได้หลายทาง เช่น จากกล้องดิจิทัล , กล้องวิดีโอ, กล้องวงจรปิด, สแกนเนอร์ เป็นต้น ซึ่งอุปกรณ์เหล่านี้จะมีหลักการทำงานที่เหมือนกันคือ มีเซนเซอร์อาร์เรย์ (Sensor Array) หรือ CCD (Charge-Coupled Device) ซึ่งมีหน้าที่ในการสุ่มตัวอย่าง (Sampling) เพื่อกำหนดจำนวนจุดของภาพผลลัพธ์ จากนั้นตรวจจับความเข้มแสงของแต่ละจุดแล้วกำหนดเป็นปริมาณตัวเลขดิจิทัลจากความเข้มแสงที่วัดได้ (Quantization) จากจุดนั้นๆ จึงได้ผลลัพธ์ออกมาเป็นภาพนิ่งดิจิทัลที่สามารถนำไปประมวลผลโดยคอมพิวเตอร์ได้ โดยขั้นตอนพื้นฐานของการประมวลผลภาพแสดงได้ดังรูปที่ 2.4



รูปที่ 2.4 แสดงขั้นตอนพื้นฐานของการประมวลผลภาพ

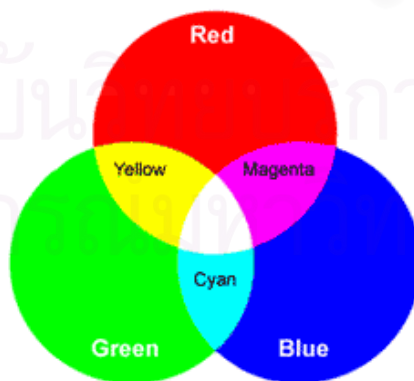
สำหรับภาพเคลื่อนไหวหรือภาพวิดีโอ เกิดจากภาพนิ่งหลายๆ ภาพเรียงต่อกันไป โดยทั่วไปแล้วความถี่ของการเปลี่ยนภาพที่มากพอจนตามนุษย์สังเกตเห็นไม่ออก คือ 24 ภาพต่อวินาที เพราะฉะนั้นการประมวลผลภาพวิดีโอ ดังเช่นที่ใช้ในงานวิจัยนี้ จะประกอบไปด้วยการรับภาพจากกล้อง จากนั้นนำภาพที่ได้จะเข้าสู่กระบวนการประมวลผลภาพดิจิทัล เมื่อประมวลผลจนได้ข้อมูลที่ต้องการในเฟรมนั้นแล้ว ก็วกกลับไปยังกระบวนการรับภาพจากกล้องดิจิทัลใหม่ซ้ำอีก โดยกระบวนการทั้งหมดนี้ ต้องประมวลผลให้เร็วเพื่อที่จะประมวลผลให้ได้มากกว่า 24 ภาพต่อวินาที ผู้ใช้จึงจะไม่รู้สึกสะดุดในการทำงาน

ภาพดิจิทัลเกิดจากหน่วยย่อยที่สุดที่เรียกว่าพิกเซล (Pixel) [13] หรือจุดภาพเรียงต่อกันทั้งแนวตั้งและแนวระนาบในลักษณะสองมิติจนได้เป็นภาพใหญ่ (จากรูปที่ 2.5 แสดงการขยายภาพจนเห็นกลุ่มของพิกเซล) ถ้าหากเป็นภาพแบบระดับเทา (Gray Scale) แต่ละพิกเซลจะแทนด้วยตัวเลขขนาด 8 บิตหนึ่งชุด ซึ่งระบุถึงค่าความสว่างของแต่ละจุดได้ทั้งหมด 256 ระดับ (0 คือระดับที่มืดที่สุดจนถึง 255 เป็นระดับที่สว่างที่สุด) เพราะฉะนั้นภาพดิจิทัลแบบระดับเทาสามารถแทนได้ด้วยเมตริกซ์สองมิติที่มีจำนวนแถวเท่ากับจำนวนพิกเซลในแนวตั้ง และมีจำนวนคอลัมน์เท่ากับจำนวนพิกเซลในแนวระนาบ ค่าในแต่ละช่องของเมตริกซ์ก็คือค่าความสว่างของแต่ละพิกเซลนั่นเอง การประมวลผลภาพดิจิทัล คือการนำค่าในเมตริกซ์นี้ไปผ่านฟังก์ชันหรืออัลกอริทึมที่ต้องการ จนได้เป็นเมตริกซ์ผลลัพธ์หรือภาพผลลัพธ์ออกมา



รูปที่ 2.5 แสดงจุดภาพหรือพิกเซล

สำหรับภาพสี โดยทั่วไปแล้วระบบสีที่นิยมใช้กันมากที่สุดคือระบบสี RGB ภาพใดๆ ก็ตามที่ใช้ระบบสีนี้ ในแต่ละพิกเซลจะเกิดจากการผสมของแสงสีปฐมภูมิสามสี ได้แก่ สีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) ในอัตราส่วนต่างๆ กัน เพื่อสร้างเป็นสีที่ต้องการขึ้นมา (ตัวอย่างดังรูปที่ 2.6) ถ้าแต่ละแม่สีมีขนาด 8 บิต แม่สีแต่ละสีสามารถมีค่าความสว่างที่ต่างกันได้ 256 ระดับ ในระบบสี RGB จะมีแม่สีอยู่สามสี เพราะฉะนั้น ต้องใช้เนื้อที่เก็บเท่ากับ 8×3 เท่ากับ 24 บิต ถ้าหากนำสามสีมาผสมกัน ก็จะสามารถสร้างสีใหม่ได้ถึง $256 \times 256 \times 256$ เท่ากับ 16,777,216 สี มากเพียงพอสำหรับการแสดงภาพต่างๆ ไป



รูปที่ 2.6 แสดงการผสมแสงสีปฐมภูมิ [13]

2.2.2 ระบบคอมพิวเตอร์กราฟฟิกส์สามมิติ

2.2.2.1 ข้อแตกต่างระหว่างกราฟฟิกส์สองมิติและสามมิติ

ระบบคอมพิวเตอร์กราฟฟิกส์แบ่งได้เป็นสองแบบหลักๆ คือ สองมิติและสามมิติ กระบวนการในการสร้างกราฟฟิกส์สองมิติจะมีแกนที่เกี่ยวข้องเพียงแค่สองแกน คือแกน x และแกน y เท่านั้น ทำให้ภาพที่ได้มีลักษณะที่แบนราบ แสงเงาต่างๆ จะไม่สมจริง ถ้าจะทำเป็นแอนิเมชันหรือภาพเคลื่อนไหว จำเป็นต้องสร้างภาพใหม่ในทุกๆ เฟรม เช่น หนึ่งการ์ตูน รูปตัวอย่างที่ได้จากกระบวนการกราฟฟิกส์สองมิติ แสดงดังรูปที่ 2.7



รูปที่ 2.7 ตัวอย่างภาพกราฟฟิกส์สองมิติ

สำหรับระบบคอมพิวเตอร์กราฟฟิกส์แบบสามมิติ กระบวนการในการสร้าง จะมีความลึกเข้ามาเกี่ยวข้องด้วย โดยจะต้องทำงานบนแกนทั้งสามแกน คือ x , y และ z วัตถุต่างๆ ในฉากจะต้องทำเป็นโมเดลสามมิติโดยใช้โปรแกรมต่างๆ เช่น 3DS Max, Maya เป็นต้น ถึงแม้กระบวนการทำงานต่างๆ จะเป็นแบบสามมิติก็จริง แต่เนื่องจากจอภาพธรรมดาสามารถแสดงภาพแบบสองมิติเท่านั้น เพราะฉะนั้นสุดท้ายแล้วระบบก็จำเป็นต้องแปลงฉากและวัตถุสามมิติต่างๆ ให้เป็นภาพสองมิติเพื่อนำขึ้นแสดงผลบนจอภาพ กระบวนการนี้เรียกว่าการ เรนเดอร์ (Render)

ระบบคอมพิวเตอร์กราฟฟิกส์สามมิติมีข้อดีที่เหนือกว่าระบบสองมิติธรรมดาอยู่มาก อาทิเช่น สามารถเปลี่ยนมุมมอง หรือมุมมองได้ตามความต้องการ สามารถที่จะเรียกดูมุมไหนของโมเดลก็ได้ แสงเงาต่างๆ ก็จะถูกคำนวณใหม่ให้สัมพันธ์กับมุมมองและแหล่งกำเนิดแสงขณะนั้นโดยอัตโนมัติ ทำให้ภาพที่ได้มีความเสมือนจริงมาก สำหรับการทำแอนิเมชันก็สามารถทำ

ได้อย่างมีประสิทธิภาพมากกว่าระบบสองมิติ นักพัฒนาเพียงแต่เขียนโปรแกรมให้วัตถุเคลื่อนที่หรือจัดวางให้เป็นแบบที่ต้องการแล้วส่งเรนเดอร์ใหม่ ก็จะได้เฟรมใหม่มาทันที ไม่จำเป็นต้องวาดใหม่ทั้งฉากเหมือนระบบสองมิติ ด้วยข้อดีต่างๆ เหล่านี้ จึงได้เลือกระบบคอมพิวเตอร์กราฟฟิกส์สามมิติมาใช้ในงานวิจัยนี้ สำหรับตัวอย่างรูปที่ได้จากระบบกราฟฟิกส์สามมิติ แสดงดังรูปที่ 2.8

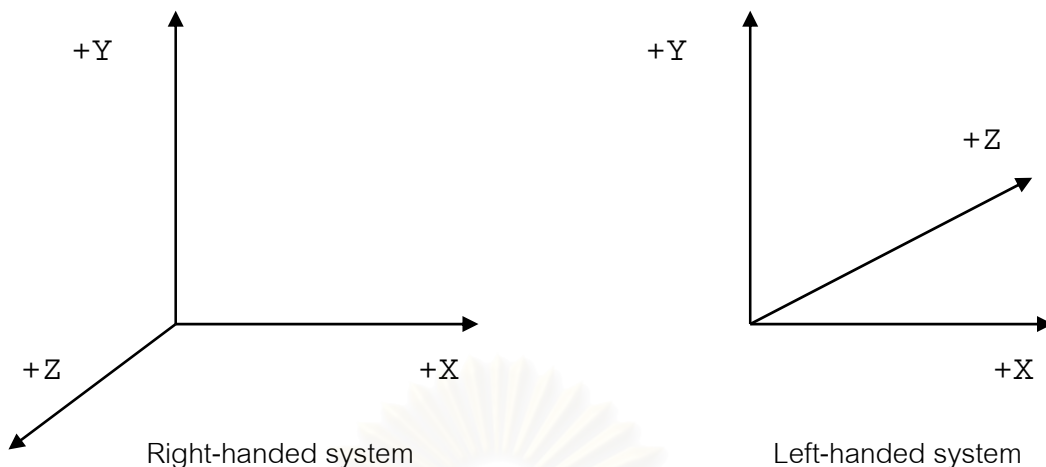


รูปที่ 2.8 ตัวอย่างภาพกราฟฟิกส์สามมิติ

ปัจจุบันนี้ได้มีผู้ผลิตกราฟฟิกส์การ์ดที่ทำหน้าที่ช่วยประมวลผลด้านกราฟฟิกส์สามมิติโดยเฉพาะ เช่นบริษัท nVidia และ ATI ที่ผลิตการ์ดในตระกูล GeForce และ Radeon ตามลำดับ ทำให้การทำงานด้านนี้มีประสิทธิภาพเพิ่มขึ้นอีกมาก สำหรับเครื่องมือที่ช่วยอำนวยความสะดวกในการเขียนโปรแกรมกราฟฟิกส์สามมิติให้กับนักพัฒนาโปรแกรมมีอยู่สองตัวหลักๆ คือ DirectX ที่พัฒนาโดยบริษัทไมโครซอฟต์ และ OpenGL พัฒนาโดยบริษัทซิลิกอนกราฟฟิกส์ สำหรับในงานวิจัยนี้ได้เลือกใช้ OpenGL [14] ในการพัฒนา

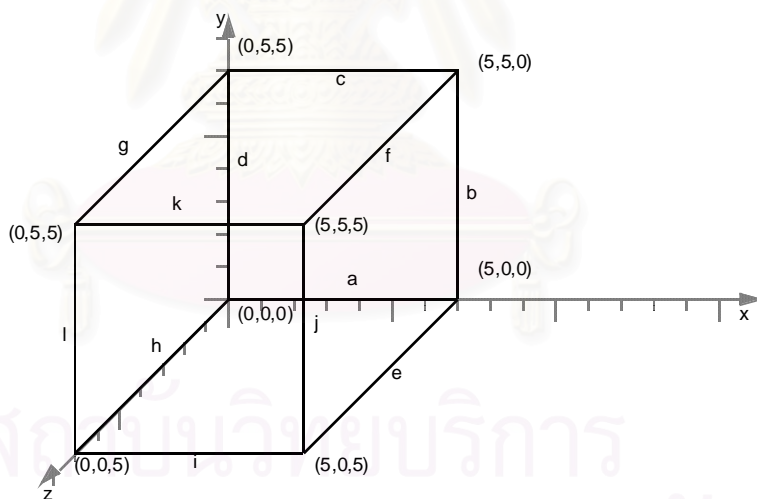
2.2.2.2 หลักการเบื้องต้นของกราฟฟิกส์สามมิติ

คอมพิวเตอร์กราฟฟิกส์สามมิติมีการทำงานบนแกนทั้งสามแกน คือ x, y, z การจัดเรียงของแกนทั้งสามแบ่งได้เป็นสองแบบหลักๆ คือ ระบบมือซ้าย (Left-handed system) และระบบมือขวา (Right-handed system) แสดงดังรูปที่ 2.9



รูปที่ 2.9 แสดงการจัดเรียงของแกนในระบบมือขวาและระบบมือซ้าย

วัตถุทุกชิ้นที่อยู่ภายในโลกสามมิติ จะต้องมิตัวเลขสำหรับอ้างอิงตำแหน่งที่อยู่แน่นอน จำนวนสามตัว ได้แก่ตำแหน่งบนแกน x, y และ z หรือที่เรียกว่าพิกัด (x, y, z) ยกตัวอย่างพิกัดของมุมทุกมุมบนสี่เหลี่ยมลูกบาศก์แสดงได้ดังรูปที่ 2.10

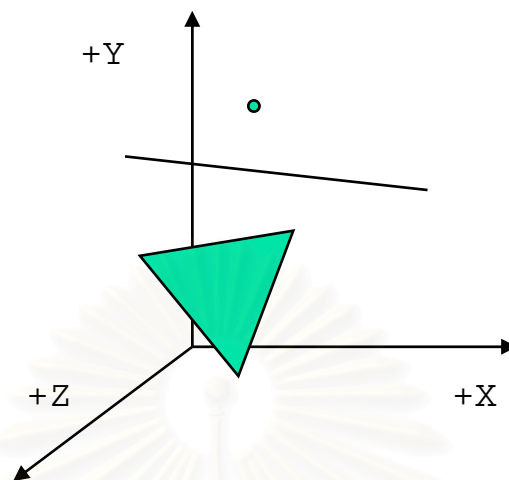


รูปที่ 2.10 พิกัด (x,y,z) ของสี่เหลี่ยมลูกบาศก์

พิกัด (5,5,5) คือจุดที่เทียบกับแกน x เท่ากับ 5 หน่วย เทียบกับแกน y เท่ากับ 5 หน่วย และเทียบกับแกน z เท่ากับ 5 หน่วยตามลำดับ

พิกัด (x,y,z) หนึ่งชุดจะแทนจุด (Point) หนึ่งจุด หากมีสองจุดแล้วลากเส้นเชื่อมเข้าด้วยกันจะได้เป็นเส้นตรง (Line) และหากมีสามจุด (หรือมากกว่า) บริเวณพื้นผิวที่ถูกระบุ

โดยจุดทั้งหมดนั้นจะเรียกว่าโพลีกอน (Polygon) และแต่ละจุดจะเรียกว่าจุดยอด (Vertex) ดังรูปที่ 2.11 และรูปที่ 2.12

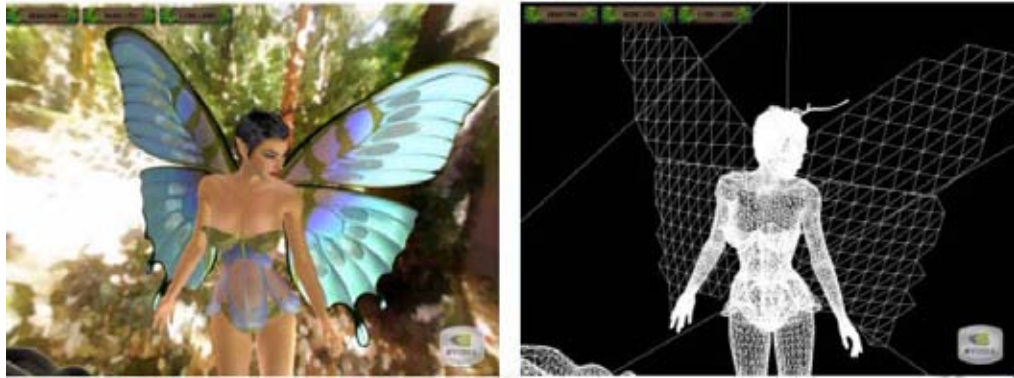


รูปที่ 2.11 แสดงจุด เส้นตรง และโพลีกอน



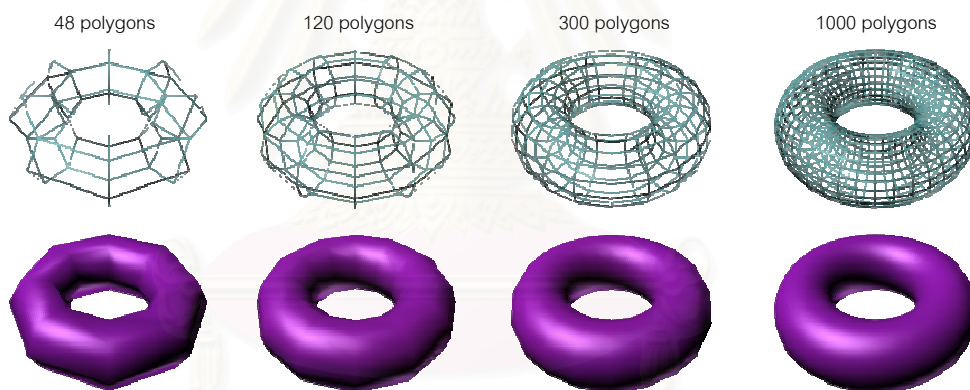
รูปที่ 2.12 แสดงองค์ประกอบของโพลีกอนรูปสามเหลี่ยม

โพลีกอนคือส่วนประกอบพื้นฐานที่สุดสำหรับแสดงพื้นผิวของวัตถุ ส่วนใหญ่แล้วจะนิยมใช้โพลีกอนรูปสามเหลี่ยมเนื่องจากจะให้ประสิทธิภาพในการคำนวณสูงที่สุด โมเดลสามมิติแต่ละชิ้นจะเกิดจากโพลีกอนหลายๆ ชิ้นมาประกอบกันจนได้เป็นรูปร่างที่ต้องการ (เรียกว่า Mesh) มีตัวอย่างเช่นรูปที่ 2.13) โพลีกอนทุกชิ้นจะมีเวกเตอร์ปกติกำกับไว้ (Normal Vector) สำหรับใช้เพื่อคำนวณการตกกระทบของแสงและการสะท้อนแสงของโพลีกอนชิ้นนั้น



รูปที่ 2.13 แสดงการประกอบกันของโพลีกอนจนเป็นโมเดลชิ้นใหญ่

จำนวนของโพลีกอนก็มีผลต่อความสมจริง ยิ่งมีจำนวนโพลีกอนมาก ก็ยิ่งสามารถแสดงรายละเอียดของโมเดลได้มากขึ้นและมีความสมจริงมากยิ่งขึ้น แต่ก็ทำให้ต้องใช้เวลาในการประมวลผลมากขึ้น จากรูปที่ 2.14 แสดงรูปทรงวงแหวนที่สร้างจากโพลีกอนจำนวนต่างๆ กัน และผลได้

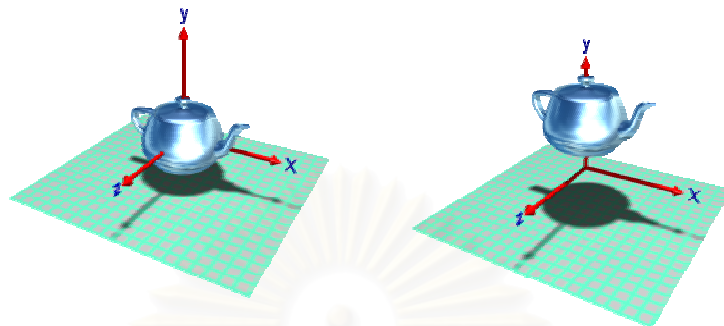


รูปที่ 2.14 แสดงรูปทรงวงแหวนที่สร้างจากโพลีกอนจำนวนต่างๆ กัน

2.2.2.3 การจัดตำแหน่งและรูปแบบวัตถุ

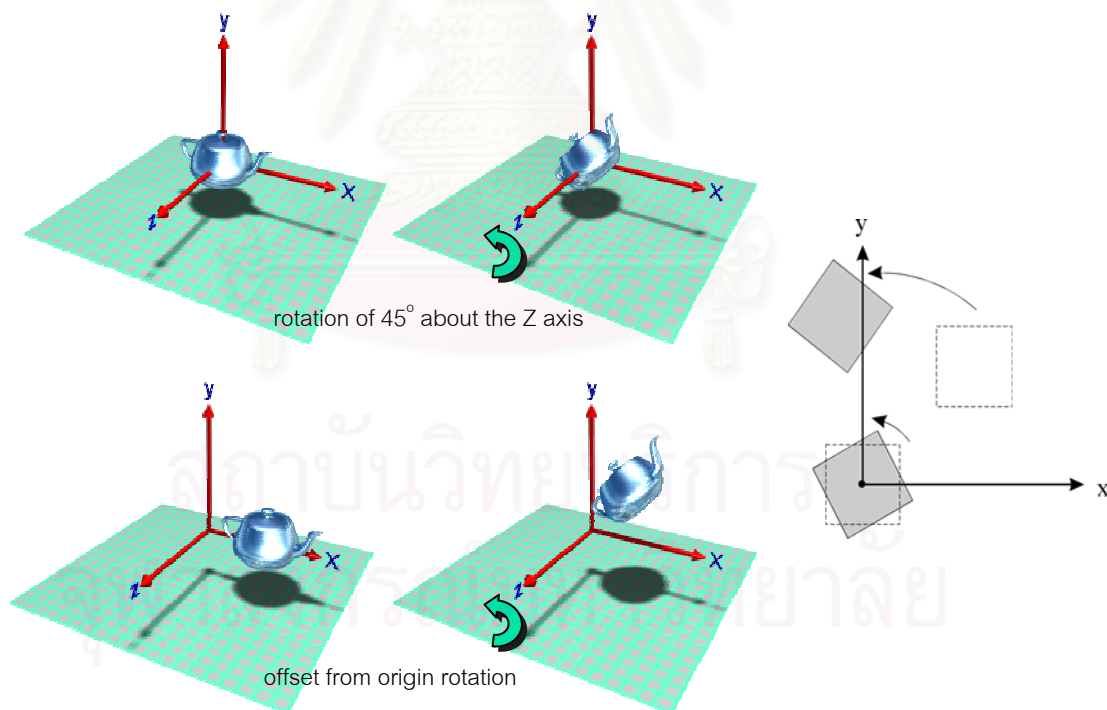
เมื่อมีโมเดลสามมิติหรือวัตถุที่ต้องการแล้ว สิ่งที่ต้องทำต่อไปคือการนำวัตถุชิ้นนั้นไปจัดรูปแบบและวางลงบนตำแหน่งที่ต้องการในฉาก เช่น โมเดลรูปหนังสือ อาจจะถูกวางลงบนโมเดลรูปโต๊ะเป็นต้น สิ่งเหล่านี้เรียกว่าการทำทรานส์ฟอร์มเมชัน (Transformation) ซึ่งจะมีอยู่สามแบบคือ การเคลื่อนย้าย (Translation) , การหมุน (Rotation) , และการปรับขนาด (Scaling)

การเคลื่อนย้าย (Translation) ก็คือการเปลี่ยนตำแหน่งของวัตถุ ให้อยู่ในตำแหน่งพิกัดที่ต้องการ เช่น จากเดิมวัตถุอยู่ที่พิกัด $(0,0,0)$ ต้องการเคลื่อนย้ายให้ไปอยู่ในพิกัด $(0,1,0)$ ก็จะเป็นดังรูปที่ 2.15



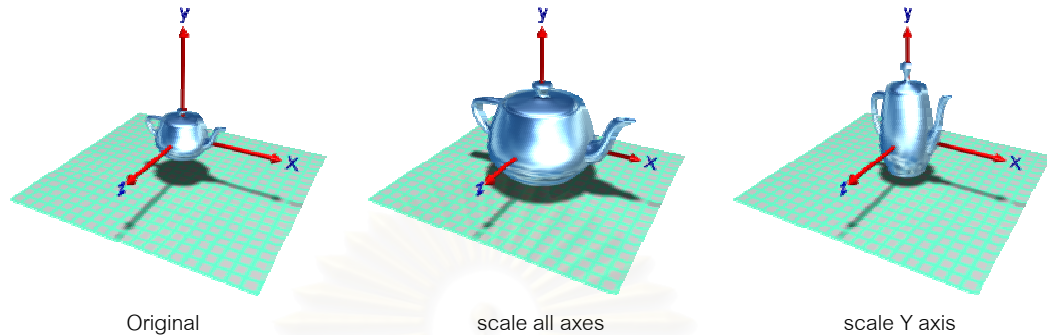
รูปที่ 2.15 แสดงการเคลื่อนย้ายวัตถุ

การหมุน (Rotation) ก็คือการหมุนวัตถุรอบแกนใดๆ จนได้ทิศทางของวัตถุตามต้องการ เช่น ถ้าต้องการหมุนวัตถุรอบแกน z ไป 45 องศา จะได้เป็นดังรูปที่ 2.16



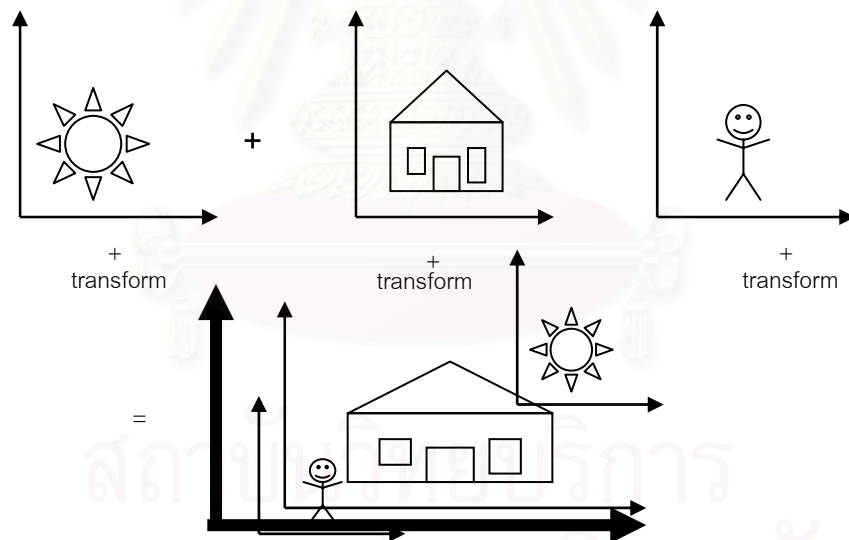
รูปที่ 2.16 แสดงการหมุนวัตถุ

การปรับขนาด (Scaling) ก็คือการปรับเปลี่ยนขนาดของวัตถุ (อาจจะขยายขนาดให้ใหญ่ขึ้น หรือลดขนาดให้เล็กลง) โดยอิงจากแกนหนึ่งๆ หรือหลายแกนพร้อมกัน รูปที่ 2.17 แสดงการปรับขนาดของวัตถุเทียบกับแกนทั้งสามแกน และปรับขนาดของวัตถุเทียบกับแกน y



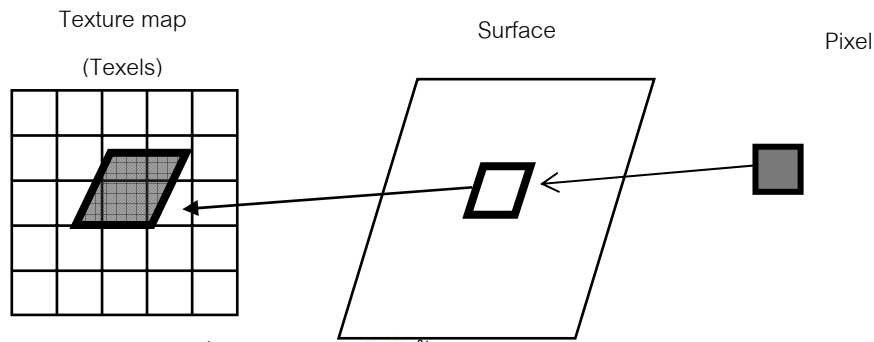
รูปที่ 2.17 แสดงการปรับขนาดวัตถุ

ถ้าหากในฉากประกอบไปด้วยวัตถุหลายชิ้น ก็จำเป็นที่จะต้องทำทรานส์ฟอร์มเมชันให้กับวัตถุทุกชิ้น เพื่อให้ทุกชิ้นอยู่ในตำแหน่งและมีรูปแบบที่ต้องการ ดังรูปที่ 2.18



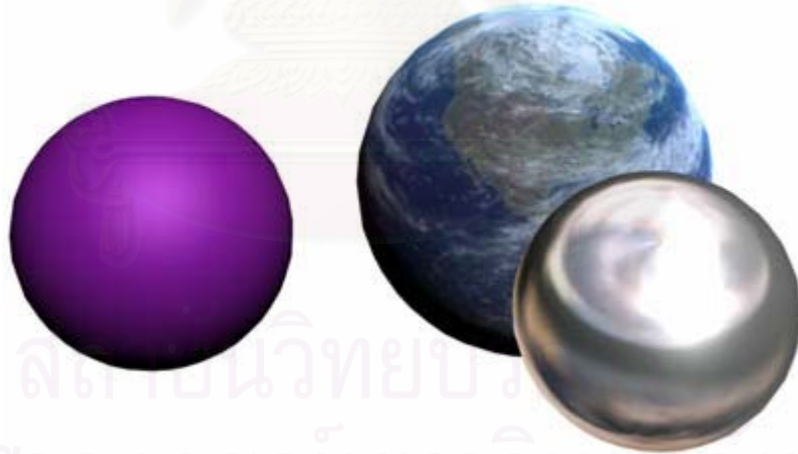
รูปที่ 2.18 การจัดองค์ประกอบของฉากโดยใช้การทรานส์ฟอร์มเมชัน

เพื่อเพิ่มความสวยงามและความสมจริงให้กับวัตถุและฉาก สามารถทำได้หลายวิธี เช่น การใช้ภาพพื้นผิวปะลงบนโมเดลสามมิติ (Texture Mapping) รวมถึงการให้แสง (Lighting) และเงา (Shadow) เป็นต้น สำหรับการให้ภาพพื้นผิวปะลงบนโมเดลสามมิตินั้น จะต้องใช้ภาพนิ่ง เช่น ภาพตระกูล JPG (JPEG) หรือ BMP (Bitmap) นำจุดภาพหรือพิกเซลปะลงบนพื้นผิวโพลีกอนตามตำแหน่งที่กำหนดจนครบทุกโพลีกอนดังรูปที่ 2.19



รูปที่ 2.19 การใช้ภาพพื้นผิวปะลงบนโมเดลสามมิติ

สำหรับการให้แสงและเงา นั้นจะต้องสร้างแหล่งกำเนิดแสงขึ้นมาก่อน ซึ่งแหล่งกำเนิดแสงแต่ละอันสามารถที่จะจัดตำแหน่งได้เช่นเดียวกับวัตถุทั่วไป เมื่อจัดได้ตำแหน่งที่ต้องการแล้ว ก็ต้องกำหนดคุณสมบัติของแหล่งกำเนิดแสงด้วย ใน OpenGL สามารถกำหนดคุณสมบัติแหล่งกำเนิดแสงได้สามแบบคือ แหล่งกำเนิดแสงแบบจุด แหล่งกำเนิดแสงแบบมีทิศทาง และแหล่งกำเนิดแสงแบบสปอตไลท์ เมื่อได้มีการกำหนดแสงให้โมเดลเรียบร้อยแล้วจะทำให้โมเดลมีความเหมือนจริงมากยิ่งขึ้นดังรูปที่ 2.20



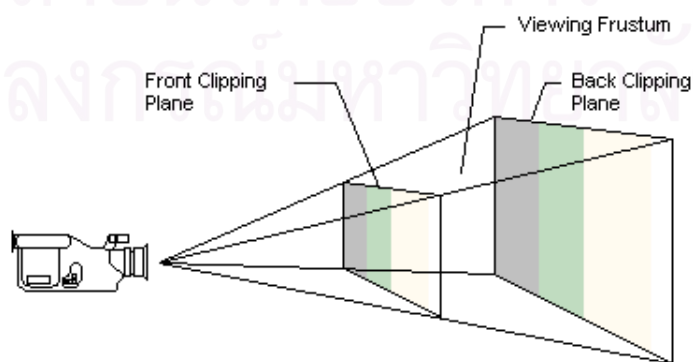
รูปที่ 2.20 แสดงผลจากการให้แสงและปะรูปพื้นผิว

2.2.2.4 กล้องในโลกสามมิติ (Camera)

จากหัวข้อที่ผ่านมา เมื่อกำหนดวัตถุสามมิติให้อยู่ในตำแหน่งที่ต้องการเรียบร้อยแล้ว และจัดฉากต่างๆ พร้อมหมดแล้ว ขั้นตอนต่อไปคือต้องเลือกว่าจะให้ส่วนไหนของฉากขึ้นแสดงบนจอภาพ เปรียบเสมือนการถ่ายรูปหรือถ่ายภาพยนต์ เมื่อฉากต่างๆ พร้อม ก็ต้องถ่ายฉากนั้นๆ ออกมาโดยใช้กล้องเลือกมุมที่ต้องการสื่อออกไปให้คนได้เห็นและรับชมแล้วถ่ายเก็บไว้ กล้องในโลกสามมิติก็มีหลักการทำงานเช่นเดียวกัน สิ่งที่กล้องจับอยู่ คือสิ่งที่จะถูกแสดงบนจอภาพ กล้องสามารถที่จะถูกจัดวางตำแหน่ง (Translate) และหมุน (Rotate) ไปยังทิศที่ต้องการได้ เช่นเดียวกับวัตถุสามมิติ การทำเช่นนี้เรียกว่าการจัดมุมมองหรือทัศนมิติ (Perspective)

กล้องในโลกแห่งความเป็นจริง สามารถที่จะถอดเปลี่ยนเลนส์ได้ อาจจะใช้เลนส์เทเล (Tele) หรือเลนส์ไวด์ (Wide) เพื่อกำหนดมุมรับภาพให้เหมาะสมกับภาพที่จะถ่าย หรือกำหนดการซูมเข้าซูมออกจนได้มุมมองที่ต้องการ กล้องในโลกสามมิติก็สามารถทำสิ่งที่กล่าวมาได้เช่นเดียวกัน โดยการกำหนดสิ่งที่เรียกว่ามุมรับภาพ หรือ FOV (Field of View) โดยทั่วไปแล้วถ้าจะกำหนดให้กล้องมีมุมรับภาพใกล้เคียงกับสายตามนุษย์มากที่สุด ควรจะใช้มุมรับภาพในแนวระนาบ 110 ถึง 120 องศา และมุมรับภาพในแนวดิ่งประมาณ 90 องศา

สิ่งที่กล้องในโลกสามมิติมีเพิ่มเติมจากกล้องในโลกแห่งความเป็นจริงคือการกำหนดระนาบตัดภาพ (Clipping Plane) และกรวยของการแสดงภาพ (Viewing Frustum) สิ่งที่ต้องกำหนดคือระนาบตัดภาพด้านหน้า (Front Clipping Plane) และระนาบตัดภาพด้านหลัง (Back Clipping Plane) วัตถุที่จะถูกกล้องถ่ายออกมาจะต้องอยู่ระหว่างระนาบตัดภาพด้านหน้า ระนาบตัดภาพด้านหลัง และมุมรับภาพเท่านั้น ปริมาตรส่วนนี้เรียกว่ากรวยของการแสดงภาพ (Viewing Frustum) วัตถุใดๆ ที่อยู่นอกเหนือบริเวณนี้จะถูกตัดทิ้งไป ไม่นำมาร่วมคำนวณ และไม่ปรากฏในภาพผลลัพธ์ ดังรูปที่ 2.21



รูปที่ 2.21 แสดงกรวยของการแสดงภาพ (Viewing Frustum)

เมื่อกำหนดทุกอย่างเรียบร้อยแล้ว ทั้งฉาก ทั้งกล้อง ในโลกแห่งความเป็นจริงเราจะต้องกดปุ่มชัตเตอร์ของกล้องเพื่อถ่ายภาพ แต่ในโลกสามมิติเราต้องสั่งเรนเดอร์ (Render) เพื่อให้โปรแกรมคำนวณและสร้างภาพสองมิติจากฉากสามมิติเพื่อที่จะแสดงผลบนจอภาพ ถ้าหากต้องการแสดงเป็นภาพเคลื่อนไหว โปรแกรมควรจะเรนเดอร์เป็นจำนวน 24 ครั้งต่อวินาทีหรือมากกว่า จึงจะได้ภาพเคลื่อนไหวที่ไม่กระตุก

2.2.3 พฤติกรรมมนุษย์

2.2.3.1 ใบหน้า

ใบหน้านั้นเป็นส่วนหนึ่งของร่างกายมนุษย์ เป็นส่วนที่มีความจำเป็นและสำคัญอย่างมากในการใช้เพื่อติดต่อสื่อสารกับคนอื่น ๆ ด้วยภาษาท่าทาง (non-verbal language) ที่นอกเหนือไปจากภาษาคำพูด (verbal language) และยังเป็นส่วนที่แสดงความรู้สึกต่างๆ ให้คนอื่นรับรู้ การแสดงออกทางสีหน้านั้นบางทีก็เปรียบเสมือนกับเป็นแหล่งข้อมูลด้านอารมณ์เพื่อเป็นส่วนประกอบให้คำพูดสมบูรณ์ นอกจากนี้ยังรวมไปถึงความสามารถในการบ่งบอกลักษณะของบุคคลภายในครอบครัวของคนๆ นั้นได้อีกด้วย คือจะต้องมีลักษณะใบหน้าที่คล้ายๆ กัน

การเคลื่อนไหวของกล้ามเนื้อบนใบหน้าของมนุษย์นั้นจะบ่งบอกถึงพฤติกรรมการแสดงออกของอารมณ์มนุษย์ในสภาวะต่างๆ กันไป อาทิเช่น ใบหน้าและมุมปากมีลักษณะโค้งขึ้น และมีรอยหยักหรือรอยย่นบ่งบอกถึงสภาวะที่มีความสุข ใบหน้าที่มีลักษณะของตาที่เขม่นแสดงว่าโกรธ ใบหน้าที่มีลักษณะดวงตาที่โตแสดงว่ามีอาการประหลาดใจ เป็นต้น

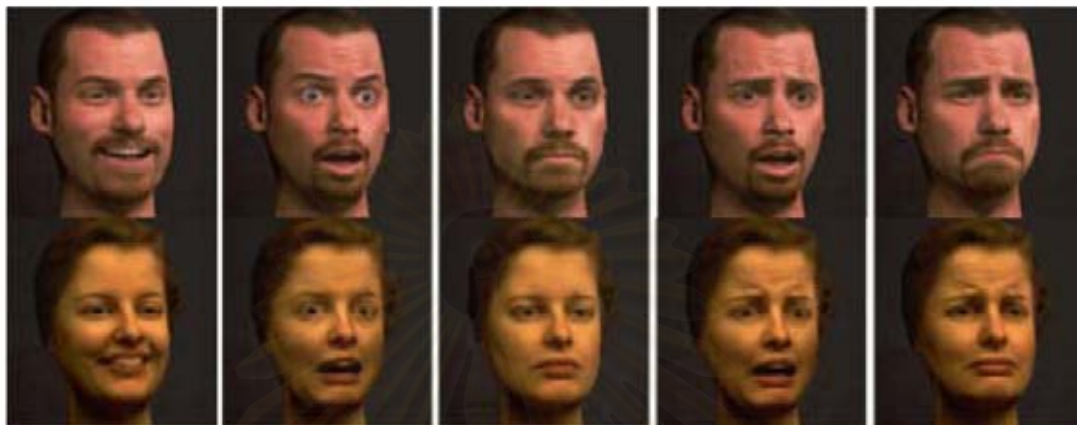
พฤติกรรมการแสดงออกของใบหน้าที่บ่งบอกถึงลักษณะของบุคคลนั้นๆ ว่าเป็นมิตรมากน้อยแค่ไหน ซึ่งมนุษย์นั้นจะรู้สึกว่าการแสดงออกทางสีหน้าจะมีอิทธิพลอย่างมากในการติดต่อสื่อสารระหว่างกัน

การแสดงออกทางสีหน้านั้นแบ่งได้เป็น 6 กลุ่มใหญ่ๆ คือ

1. Happiness (มีความสุข)
2. Anger (โกรธ)
3. Surprise (ประหลาดใจ)
4. Sadness (เศร้า)
5. Fear (กลัว)

6. Disgust (ขยะแขยง หรือสะอิดสะเอียน)

ตัวอย่างใบหน้าแสดงดังรูปที่ 2.22 เรียงจากซ้ายไปขวา ได้แก่ มีความสุข, ประหลาดใจ, เศร้า, กลัว, และขยะแขยง



รูปที่ 2.22 แสดงรูปแบบอารมณ์ต่างๆ ของใบหน้ามนุษย์ [15]

2.2.3.2 ดวงตา

ดวงตานั้นมีความสำคัญเช่นเดียวกับใบหน้า สำหรับการติดต่อสื่อสารด้วย ดวงตาจะเป็นไปเพื่อจุดประสงค์ดังนี้คือ

1. เพื่อที่จะได้รับรู้ถึงการตอบสนองของผู้ฟัง
2. ควบคุมการติดต่อสื่อสาร
3. แสดงถึงลักษณะของความสัมพันธ์ระหว่างผู้พูด ผู้ฟัง สถานภาพ ความคิดเห็นและบุคลิกภาพ
4. ถ้าหากในกลุ่มสนทนามีคนหลายคน การใช้สายตาจะบ่งบอกว่ากำลังสื่อสารอยู่กับใคร

ผู้พูดมักจะใช้การสื่อสารด้วยดวงตาประกอบไปด้วยเมื่อเขาต้องการที่จะได้รับการตอบกลับหรือได้รับความสนใจจากผู้ฟัง หรือสามารถใช้ในการวิเคราะห์ทัศนคติของผู้ฟัง เช่นเมื่อผู้ฟังไม่สนใจก็อาจจะมองไปทางอื่น การสื่อสารโดยใช้ดวงตาประกอบสามารถที่จะควบคุมให้การสนทนานั้นๆ เป็นไปด้วยดีและมีความเข้าใจกันมากยิ่งขึ้น

2.2.4 การเทียบมาตรฐานกล้อง (Camera Calibration)

การเทียบมาตรฐานกล้อง (Camera Calibration) [16] เป็นเทคนิคที่จะนำไปใช้ในการหาความสัมพันธ์ระหว่างตำแหน่งจุดต่างๆ ในโลกจริง และจุดพิกัด (x, y) ที่ปรากฏบนภาพดิจิทัล เพราะฉะนั้นเมื่อทราบตำแหน่ง (x, y) ของจุดบนภาพดิจิทัล ก็สามารถใช้กระบวนการนี้ในการหาตำแหน่ง (x, y, z) ของจุดในโลกจริง (เทียบกับกล้อง) ได้

กระบวนการนี้จะเกี่ยวข้องกับพารามิเตอร์ของกล้องแบบภายใน (Intrinsic) และภายนอก (Extrinsic) ดังนี้

1. พารามิเตอร์ของกล้องแบบภายใน (Intrinsic camera parameters) คือค่าที่บ่งบอกลักษณะเฉพาะของกล้องนั้นๆ ซึ่งได้แก่
 - ความยาวโฟกัส (Focal length) ซึ่งก็คือระยะระหว่างเลนส์ของกล้องและฉากรับภาพ
 - ตำแหน่งของจุดกึ่งกลางภาพในระบบพิกัดจุด (Principal point)
 - ขนาดของแต่ละจุดภาพ (Effective pixel size)
 - สัมประสิทธิ์การบิดเบี้ยวในแนวรัศมีของเลนส์ (Radial distortion coefficient)
2. พารามิเตอร์ของกล้องแบบภายนอก (Extrinsic camera parameters) คือค่าที่บ่งบอกถึงความสัมพันธ์เชิงระยะทาง (Spatial Relationship) ระหว่างกล้อง (Camera) กับโลกภายนอก (World) ได้แก่
 - เมตริกซ์การหมุน (Rotational Matrix)
 - เวกเตอร์การเคลื่อนที่ (Translation Vector)

ถ้าหากใช้กล้องแบบบูรณาการทั้งหมดแล้ว จุดพิกัดสามมิติในโลกจริง M และจุดที่ปรากฏบนภาพ m เป็นไปตามสมการที่ 2.1

$$m = A[Rt]M \quad \dots(2.1)$$

โดยที่ A คือเมตริกซ์ของพารามิเตอร์แบบภายใน:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \text{ เมื่อ}$$

(c_x, c_y) คือพิกัดของจุดกึ่งกลางภาพ (Principal point)

(f_x, f_y) คือระยะโฟกัสบนแกน x และ y

(R, t) คือพารามิเตอร์แบบภายนอก: เมตริกซ์การหมุน R และเวกเตอร์การเคลื่อนที่ t

เมื่อทราบถึงความสัมพันธ์ระหว่างจุดในโลกจริงของผู้ใช้กับจุดที่ปรากฏบนภาพก็สามารถแปลงตำแหน่งของจุดสองมิติบนภาพไปเป็นจุดบนพิกัดสามมิติในโลกของผู้ใช้ได้ แต่ในการแปลงจุดในพิกัดสองมิติให้อยู่ในพิกัดสามมิตินั้นใช้จุดเดียวไม่เพียงพอ (เพราะการแปลงโดยใช้จุดสองมิติเพียงจุดเดียวจะได้ผลลัพธ์เพียงแค่แนวการฉายของจุดในระนาบสามมิติ) จำเป็นต้องใช้หลายๆ จุดที่มีรูปแบบการเรียงตัวที่แน่นอน (Pattern) จึงจะเพียงพอสำหรับการคำนวณ



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

การออกแบบและรายละเอียดในการพัฒนา

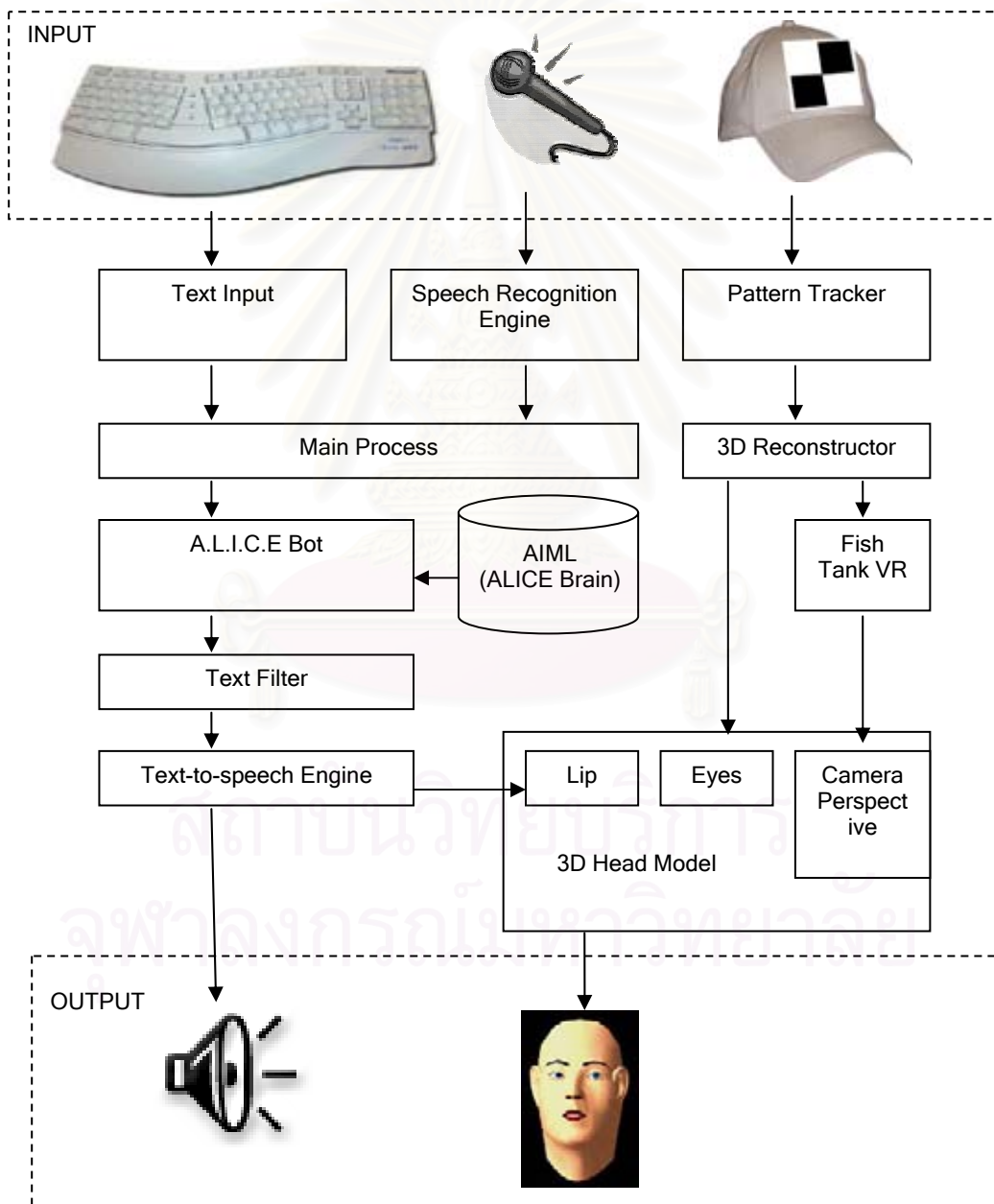
จากแนวคิดในงานวิจัยที่กล่าวมา สามารถนำมากำหนดเป็นรายละเอียดและ
ความสามารถของระบบได้ดังนี้

1. ระบบที่พัฒนาขึ้นสามารถแสดงโมเดลสามมิติของศีรษะมนุษย์เสมือนจริง
2. โมเดลศีรษะสามารถมองไปยังผู้ใช้หรือสบตาผู้ใช้ได้
3. สามารถสนทนาพูดคุยกับผู้ใช้ได้
4. สามารถขยับปากให้สอดคล้องกับคำพูดได้
5. สามารถแสดงอารมณ์บนใบหน้า ได้แก่ ปกติ, ยิ้ม, เศร้า, ประหลาดใจ, หวาดกลัว, โกรธ, สะอิดสะเอียน เป็นต้น
6. สามารถปรับมุมมองของโมเดลให้สอดคล้องกับตำแหน่งดวงตาของผู้ใช้ได้แบบ
เวลาจริง โดยผู้ใช้ต้องสวมหมวกติดรูปแบบตรวจจับ (Pattern) เอาไว้ และต้อง
เคลื่อนไหวให้อยู่ในมุมรับภาพของกล้อง
7. การควบคุมโมเดลศีรษะสามารถกระทำโดยผ่านสคริปต์

ก่อนอื่นขออธิบายแนวคิดการทำงานในภาพรวมเสียก่อน เมื่อจะคุยกับระบบ ผู้ใช้
สามารถที่จะพิมพ์ หรือพูดโดยใช้เสียงก็ได้ กรณีที่ผู้ใช้พูดใส่ไมโครโฟน เสียงจะถูกแปลงเป็น
ข้อความตัวอักษรเสียก่อนโดยใช้กลไกการรู้จำเสียงพูด (Speech Recognition Engine) [17] เมื่อ
ได้เป็นข้อความตัวอักษร ข้อความนี้จะถูกนำไปเป็นอินพุตให้กับ A.L.I.C.E Bot [18] ซึ่งเป็น A.I ที่
เป็นมาตรฐานสำหรับพูดคุยกับผู้ใช้ จากนั้น A.L.I.C.E Bot ก็จะให้คำตอบรับออกมาเป็นข้อความ
ตัวอักษรเช่นเดียวกัน จึงต้องนำข้อความนี้เข้าสู่กลไกสังเคราะห์เสียงพูด (Text-to-speech
Engine) เพื่อสร้างเป็นเสียงให้ผู้ใช้ได้ยิน ในขณะที่กลไกสังเคราะห์เสียงพูดทำงาน จะส่งสัญญาณ
ไปยังส่วนการทำงานของโมเดลใบหน้าสามมิติ เพื่อกำหนดรูปแบบการขยับปากให้สอดคล้องกับ
เสียงที่เปล่งออกมา สำหรับการขยับปาก รวมถึงการแสดงอารมณ์ต่างๆ บนใบหน้า นั้น จะใช้
หลักการของระบบกล้ามเนื้อ (Muscle) ที่อยู่ตามจุดต่างๆ บนใบหน้า ซึ่งจะกล่าวโดยละเอียด
ต่อไปในภายหลัง

ส่วนสำคัญอีกส่วนหนึ่งคือส่วนของการรับรู้ตำแหน่งผู้ใช้ ระบบสามารถรับรู้
ตำแหน่งของผู้ใช้ได้จากการตรวจหาตำแหน่งของรูปแบบตรวจจับ (Pattern) ที่ติดไว้บนหมวกที่ผู้ใช้

สวมอยู่โดยผ่านกระบวนการประมวลผลภาพดิจิทัล (Image processing) ที่ได้จากกล้องวงจรมืด เมื่อได้ตำแหน่งของรูปแบบตรวจจับ ก็สามารถนำไปคำนวณหาตำแหน่งพิกัดของผู้ใช้และมุมมองของผู้ใช้ที่กระทำกับจอภาพโดยผ่านโมดูล 3D Reconstructor จากนั้นจะนำข้อมูลที่คำนวณได้ไปปรับดวงตาของโมเดล และส่งต่อข้อมูลไปให้โมดูล Fish Tank VR เพื่อปรับมุมมองการแสดงผลโมเดลสามมิติให้สอดคล้องกับตำแหน่งผู้ใช้ สำหรับภาพแสดงการทำงานของส่วนประกอบต่างๆ แสดงดังรูปที่ 3.1

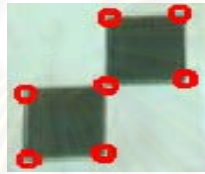


รูปที่ 3.1 แสดงโครงสร้างการทำงานทั้งหมดของงานวิจัย

รายละเอียดการทำงานในแต่ละส่วน มีดังนี้

3.1 การตรวจหาตำแหน่งของผู้ใช้เทียบกับจอภาพ

การตรวจหาตำแหน่งของผู้ใช้เทียบกับจอภาพ จะเริ่มจากการตรวจจับตำแหน่งจุดมุมทั้ง 7 จุดของรูปแบบตรวจจับดังรูปที่ 3.2 จากนั้นจึงใช้เทคนิคการเทียบมาตรฐานกล้อง (Camera Calibration) เพื่อคำนวณตำแหน่งพิกัดสามมิติและการวางตัวของรูปแบบตรวจจับเทียบกับกล้อง จากนั้นจึงคำนวณการย้ายแกนเพื่อหาตำแหน่งพิกัดสามมิติของศีรษะผู้ใช้เทียบกับกึ่งกลางจอภาพตามที่ต้องการ



รูปที่ 3.2 ตัวอย่างรูปแบบตรวจจับที่ใช้ในงานวิจัย

สำหรับในงานวิจัยนี้ กำหนดให้คอมพิวเตอร์รับภาพผู้ใช้โดยกล้องวงจรปิดที่ต่อกับการ์ดจับภาพวีดีโอ (Video capture card) กล้องจะถูกติดตั้งอยู่บนจอมอนิเตอร์ (เป็นตำแหน่งที่กล้องสามารถจับภาพศีรษะได้ดีที่สุด) ก่อนที่จะนำกล้องไปใช้งาน จะต้องผ่านการเทียบมาตรฐาน (Calibration) เพื่อหาค่าพารามิเตอร์แบบภายในเสียก่อน โดยการนำเซตของจุด (x, y, z) บนกรอบอ้างอิง (Reference frame) เช่นจุดมุมตารางหมากรุกดังรูปที่ 3.2 และเซตของจุด (x, y) บนภาพที่ได้จากกล้อง นำมาคำนวณหาพารามิเตอร์แบบภายใน

เมื่อได้พารามิเตอร์แบบภายในแล้วจึงเริ่มรับภาพผู้ใช้เข้ามาจากกล้อง แต่ภาพที่ได้จากกล้องส่วนใหญ่จะมีการบิดเบี้ยวตามแนวรัศมี หากนำไปใช้คำนวณหาตำแหน่งของศีรษะภายหลังอาจให้ผลที่ผิดพลาดได้ เพราะฉะนั้นจึงมีความจำเป็นที่จะต้องปรับแก้ภาพให้มีความเที่ยงตรงเสียก่อน โดยการนำสัมประสิทธิ์การบิดเบี้ยวตามแนวรัศมี (Radial distortion coefficient) k_1, k_2, p_1, p_2 ที่ได้จากขั้นตอนการหาพารามิเตอร์แบบภายใน มาร่วมคำนวณเพื่อทำการปรับแก้การบิดเบี้ยว (Undistortion) สำหรับความสัมพันธ์ระหว่างภาพที่เกิดการบิดเบี้ยวและภาพในอุดมคติที่ไม่มีการบิดเบี้ยว เป็นไปตามสมการที่ 3.1 และสมการที่ 3.2

$$\tilde{x} = x + x[k_1 r^2 + k_2 r^4] + [2p_1 xy + p_2(r^2 + 2x^2)] \quad \dots(3.1)$$

$$\tilde{y} = y + y[k_1 r^2 + k_2 r^4] + [2p_2 xy + p_2(r^2 + 2y^2)] \quad \dots(3.2)$$

โดยที่ x, y คือตำแหน่งของจุดบนภาพในอุดมคติ (ภาพที่ไม่มีการบิดเบี้ยว) และ \tilde{x}, \tilde{y} คือตำแหน่งของจุดบนภาพที่เกิดการบิดเบี้ยวตามแนวรัศมี และ $r^2 = x^2 + y^2$ ซึ่งเมื่อปรับแก้การบิดเบี้ยวตามแนวรัศมีแล้ว จะได้ผลเป็นดังรูปที่ 3.3



รูปที่ 3.3 แสดงการบิดเบี้ยวตามแนวรัศมี (ซ้าย) และภาพที่ปรับแก้การบิดเบี้ยวแล้ว (ขวา)

จากนั้นจึงเข้าสู่กระบวนการตรวจหาตำแหน่งของศีรษะ โดยระหว่างการใช้งาน ผู้ใช้จำเป็นต้องสวมหมวกติดรูปแบบตรวจจับที่มีลักษณะเป็นลายตารางหมากรุกขนาด 2×2 ช่อง เพื่อตรวจจับจุดมุมทั้ง 7 จุด ซึ่งจำเป็นสำหรับการคำนวณหาพารามิเตอร์แบบภายนอก ข้อดีของการใช้การตรวจจุดมุมของตารางหมากรุกแทนการตรวจอวัยวะบนใบหน้าคือ วิธีการนี้มีความรวดเร็วในการประมวลผล และมีความแม่นยำสูงกว่า และไม่ขึ้นกับลักษณะหน้าตาของผู้ใช้ ไม่ว่าผู้ใดคนใด เพียงแค่สวมอุปกรณ์สวมศีรษะก็สามารถใช้งานในระบบได้ทันที สำหรับการตรวจจับจุดมุมทั้ง 7 จุด มีขั้นตอนดังต่อไปนี้

- 1.) รับภาพต้นฉบับที่ปรับแก้การบิดเบี้ยวตามแนวรัศมีเรียบร้อยแล้วเข้ามาเพื่อเริ่มทำการประมวลผล ดังรูปที่ 3.4



รูปที่ 3.4 ภาพจากกล้องที่ผ่านการปรับแก้การบิดเบี้ยวตามแนวรัศมีแล้ว

- 2.) แปลงระบบสีจากภาพสี (RGB) ให้เป็นภาพระดับเทา (Gray-scale) ดังรูปที่ 3.5 โดยสัดส่วนของสีแดง (R) เขียว (G) และน้ำเงิน (B) ในการคำนวณค่าความสว่าง (Y) เป็นดังสมการที่ 3.3

$$Y = 0.212671R + 0.715160G + 0.072169B \quad \dots(3.3)$$

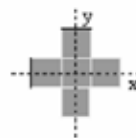


รูปที่ 3.5 ผลจากการแปลงเป็นภาพระดับเทา

- 3.) ขยายขนาดของจุดสว่างในภาพ (Dilation) เพื่อให้จุดภาพที่มีค่าสว่าง เชื่อมเข้าหากัน ส่งผลให้รูปสี่เหลี่ยมสีดำสองอันแยกตัวออกจากกันอย่างชัดเจน มีประโยชน์ในการหาเส้นรอบรูป (Contour) ของทรงสี่เหลี่ยมในขั้นตอนถัดๆ ไป เพราะถ้าหากสี่เหลี่ยมสีดำสองอันนี้ติดกันสนิทอาจจะเกิดปัญหาได้เส้นรอบรูปที่ผิดพลาด สำหรับการขยายจุดสว่างในภาพใช้สมการที่ 3.4

$$D(A, B) = A \oplus B = \bigcup_{\beta \in B} (A + \beta) \quad \dots(3.4)$$

โดยที่ A คือเซตของเวกเตอร์ที่แทนจุดต่างๆ บนภาพ และ B คือเซตของเวกเตอร์ที่แทนส่วนประกอบโครงสร้าง (Structuring Element) ดังรูปที่ 3.6



รูปที่ 3.6 แสดงส่วนประกอบโครงสร้างที่ใช้

หลังจากการทำกรขยายขนาดของจุดสว่างเรียบร้อยแล้ว ได้ภาพผลลัพธ์เป็นดังรูปที่ 3.7



รูปที่ 3.7 ผลจากการขยายจุดสว่างในภาพ

- 4.) กำหนดค่าขีดแบ่งและทำให้เป็นภาพลักษณะฐานสอง (Thresholding) ภาพที่ได้ จะมีแค่ค่า 0 (มืด) และ 1 (สว่าง) เท่านั้น ตามสมการที่ 3.5

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) > T \\ 0 & \text{if } f(x, y) \leq T \end{cases} \quad \dots(3.5)$$

โดยที่ค่าขีดแบ่ง T คือค่าเฉลี่ยความสว่างทั่วทั้งภาพ ได้ผลดังรูป 3.8



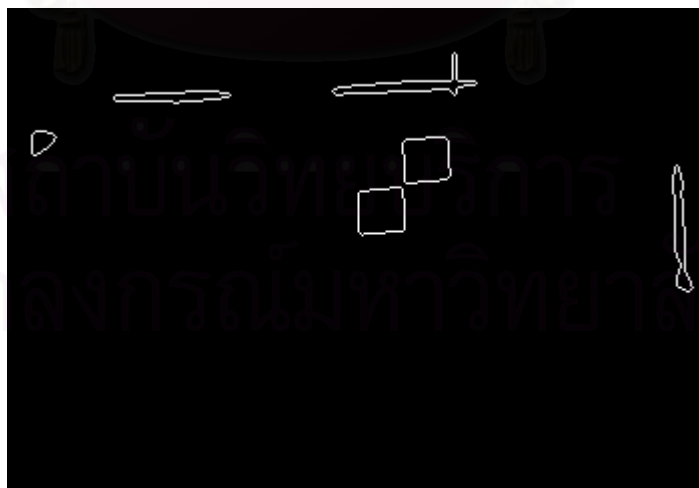
รูปที่ 3.8 ผลจากการทำให้เป็นภาพลักษณะฐานสอง

5.) หาเส้นรอบรูป (Contour) ได้ผลเป็นดังรูปที่ 3.9



รูปที่ 3.9 ผลจากการหาเส้นรอบรูป

- 6.) เส้นรอบรูปที่ได้จะมีสองแบบคือ เส้นที่ล้อมรอบกลุ่มของจุดสว่าง และเส้นที่ล้อมรอบกลุ่มของจุดมืด เราจะคัดเส้นที่ล้อมรอบกลุ่มของจุดสว่างทิ้งไป ให้เหลือแต่เส้นที่ล้อมรอบกลุ่มของจุดมืดเอาไว้
- 7.) ตัดเส้นรอบรูป ที่มีขนาดพื้นที่น้อยกว่าที่กำหนดออกไป โดยจะถือว่าเป็นเส้นรอบรูปที่เกิดจากจุดรบกวน พบว่าผลที่ได้ช่วยตัดเส้นที่ไม่จำเป็นออกไปได้อย่างมาก ดังรูปที่ 3.10



รูปที่ 3.10 ผลจากการตัดเส้นรอบรูปที่มีพื้นที่น้อยกว่าที่กำหนด

- 8.) เลือกเฉพาะเส้นรอบรูปที่คล้ายกับรูปสี่เหลี่ยม คือมีมุมอยู่สี่มุม นอกนั้นตัดทิ้งไป
- 9.) เลือกเฉพาะเส้นรอบรูปที่มีลักษณะนูนโค้ง (Convex) สำหรับเส้นรอบรูปที่ส่วนใดส่วนหนึ่งมีลักษณะเว้าเข้า (Concave) ตัดทิ้งไป จะได้ผลออกมาเหลือแต่เส้นรอบรูปที่มีลักษณะของสี่เหลี่ยมที่คล้ายกับรูปแบบตรรกจจับเท่านั้นดังรูปที่ 3.11



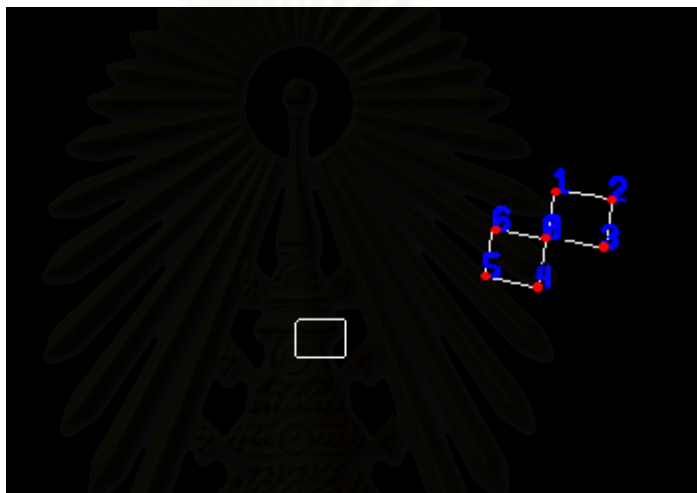
รูปที่ 3.11 ผลจากการเลือกเฉพาะเส้นรอบรูปที่มีลักษณะนูนโค้ง

ในบางกรณีอาจจะมีเส้นรอบรูปที่เป็นสี่เหลี่ยมอื่นๆ ที่ไม่ใช่รูปแบบตรรกจจับก็ได้ ดังรูปที่ 3.12 จะมีเส้นรอบรูปของหน้าจอคอมพิวเตอร์เกินเข้ามา ซึ่งจะต้องทำการตัดทิ้งออกไปภายหลัง



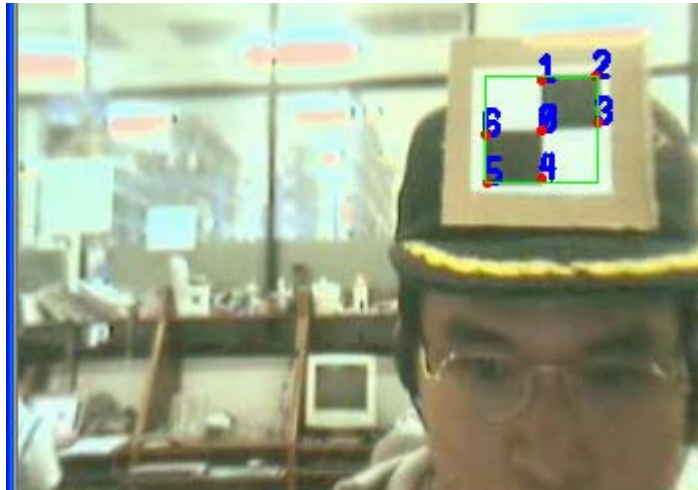
รูปที่ 3.12 แสดงรูปสี่เหลี่ยมที่ไม่ใช่รูปแบบตรรกจจับ

- 10.) เนื่องจากรูปแบบตรวจจับ จะเป็นสี่เหลี่ยมสีดำสองอันอยู่ตำแหน่งที่ติดกัน เพราะฉะนั้นการตรวจสอบจะนำเส้นรอบรูปสี่เหลี่ยมทั้งหมดมาคำนวณดูว่ามีสี่เหลี่ยมคู่ใดที่มีจุดมุมอยู่ใกล้กันที่สุด โดยการตรวจสอบไปที่ละคู่ จนครบทุกคู่ที่เป็นไปได้ เมื่อได้คู่ที่มีจุดมุมใกล้กันที่สุดออกมา จึงเลือกสี่เหลี่ยมคู่นั้นเป็นคู่ที่ต้องการ จากนั้นหาจุดมุมทั้ง 4 ของสี่เหลี่ยมทั้งคู่ จะได้เป็นจุดมุม 8 จุด จุดที่อยู่ใกล้กันที่สุดสองจุด จะถูกแทนที่ด้วย 1 จุดที่อยู่ระหว่างกลางสองจุดนั้น ได้เป็นจุดมุมทั้ง 7 จุด ดังรูปที่ 3.13

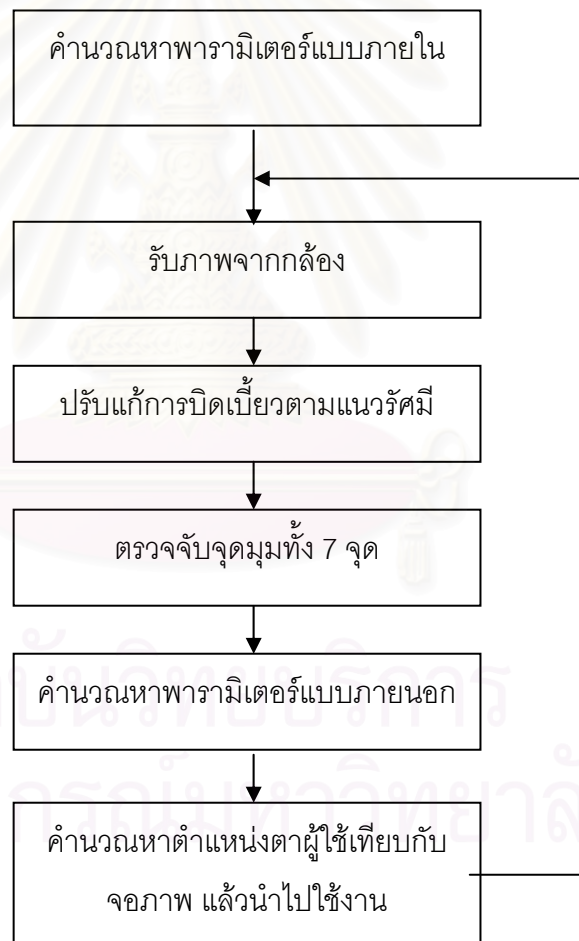


รูปที่ 3.13 แสดงจุดมุมทั้งเจ็ดจุดที่ตรวจจับได้

- 11.) เนื่องจากว่าความแม่นยำของการตรวจจับจุดมุมทั้ง 7 จุดจากภาพดิจิทัล ขึ้นอยู่กับปัจจัยภายนอกค่อนข้างมาก เช่น สภาพแสง หรือ ระบบปรับแสงของกล้อง สิ่งเหล่านี้ทำให้จุดมุมที่ตรวจจับได้อาจเกิดความคลาดเคลื่อน เช่น จุดอาจจะกระโดดไปมา ไม่นิ่ง เพราะฉะนั้นในขั้นตอนนี้จึงแก้ไขโดยการเฉลี่ยค่าตำแหน่งจุดมุมทุกจุดจาก 4 เฟรมล่าสุด ได้เป็นจุดมุม 7 จุดที่จะนำไปใช้งานจริงๆ



รูปที่ 3.14 แสดงภาพผลลัพธ์จากการตรวจจับจุดมุ่ม



รูปที่ 3.15 แสดงขั้นตอนทั้งหมดของกระบวนการประมวลผลภาพเพื่อระบุตำแหน่งตาของผู้ใช้เทียบกับจอภาพ

เมื่อได้จุดมุมทั้ง 7 จุดแล้ว (ดังรูปที่ 3.14) จึงนำไปคำนวณหาพารามิเตอร์แบบภายนอก จะได้ เวกเตอร์การเคลื่อนที่ (ตำแหน่งของรูปแบบตรวจจับเทียบกับกล้อง) และเมตริกซ์การหมุน (มุมที่ รูปแบบตรวจจับกระทำกับกล้อง) เมื่อรู้ตำแหน่งและมุมของรูปแบบตรวจจับก็สามารถประมาณตำแหน่งของตาและมุมมองที่ผู้ใช้กระทำกับกล้องได้ (การใช้หมวกทำให้รูปแบบตรวจจับมีตำแหน่งคงที่เทียบกับตา) จากนั้นระบบจะต้องคำนวณการย้ายแกนเพื่อให้จุดอ้างอิงเปลี่ยนจากกล้องไปอยู่ที่ตำแหน่งกึ่งกลางจอภาพ (เนื่องจากผู้ใช้มองไปยังจอภาพ) จึงจะได้ตำแหน่งตาของผู้ใช้เทียบกับจอภาพตามต้องการ ในงานวิจัยนี้ ตำแหน่งของผู้ใช้จะคำนวณในหน่วยของเซนติเมตร (cm) ทั้งหมด

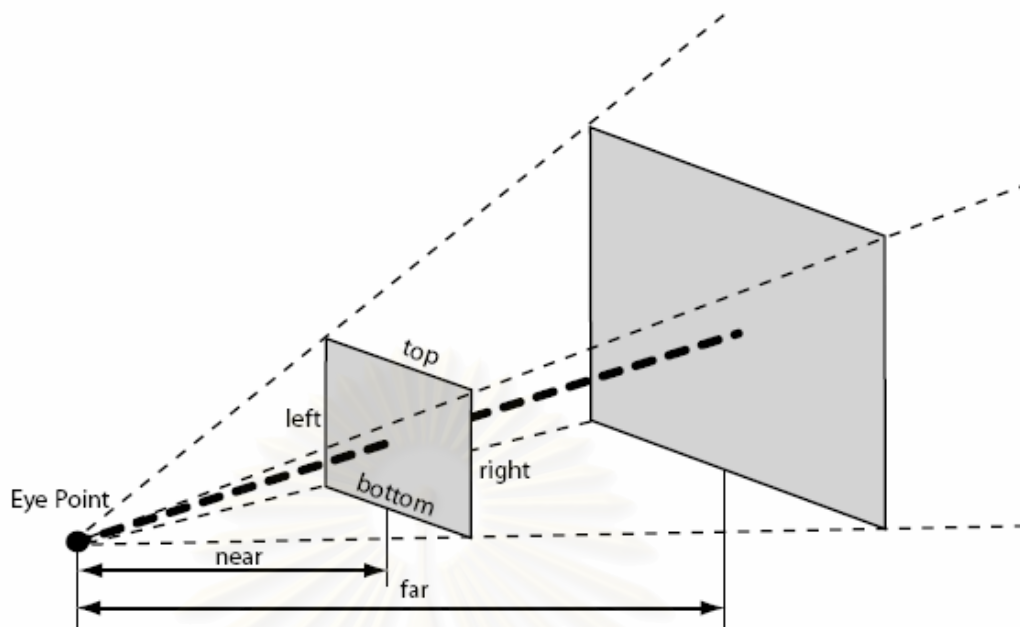
ขั้นตอนทั้งหมดสามารถสรุปการทำงานได้ดังรูปที่ 3.15

3.2 การปรับมุมมองแบบ Fish Tank VR

Colin Ware ได้นำเสนอระบบความจริงเสมือนรูปแบบใหม่ที่ชื่อว่า Fish Tank VR [3][19] ระบบ VR แบบนี้จะแสดงผลผ่านจอมอนิเตอร์ธรรมดาที่หาซื้อได้ทั่วไป โดยใช้ระบบการติดตามตำแหน่งศีรษะของผู้ใช้ (Head Tracking) มาร่วมด้วย เพื่อปรับมุมมองหรือทัศนมิติ (Perspective) สำหรับการแสดงผลภาพสามมิติให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในขณะนั้น สำหรับในงานวิจัยชิ้นนี้จะสร้างภาพกราฟิกโดยใช้ OpenGL เป็นหลัก เพราะฉะนั้นการปรับทัศนมิติจะกระทำผ่านฟังก์ชันของ OpenGL ตำแหน่งศีรษะของผู้ใช้เทียบกับจอภาพที่ได้ในหัวข้อที่แล้ว จะถูกนำมาเซตเป็นตำแหน่งให้กับกล้องของ OpenGL (ในงานวิจัยนี้ ระยะห่างหนึ่งเซนติเมตรในโลกจริง จะถูกกำหนดให้เท่ากับหนึ่งหน่วยใน OpenGL)

กล้องใน OpenGL มีคุณสมบัติหลายอย่างที่เหมือนกับกล้องที่เราใช้กันจริงๆ อาทิเช่น กล้องมือถือมักจะถูกเปลี่ยนเลนส์ เช่น เปลี่ยนเป็นเลนส์มุมกว้าง (Wide) หรือเลนส์เทเล (Tele) ซึ่งจะมีมุมมองที่แตกต่างกันออกไป กล้องใน OpenGL ก็ยังสามารถทำแบบนี้ได้เช่นกัน โดยการปรับทัศนมิติ (Perspective) เมื่อนำมาใช้ในระบบ Fish Tank VR จะเป็นดังรูปที่

3.16

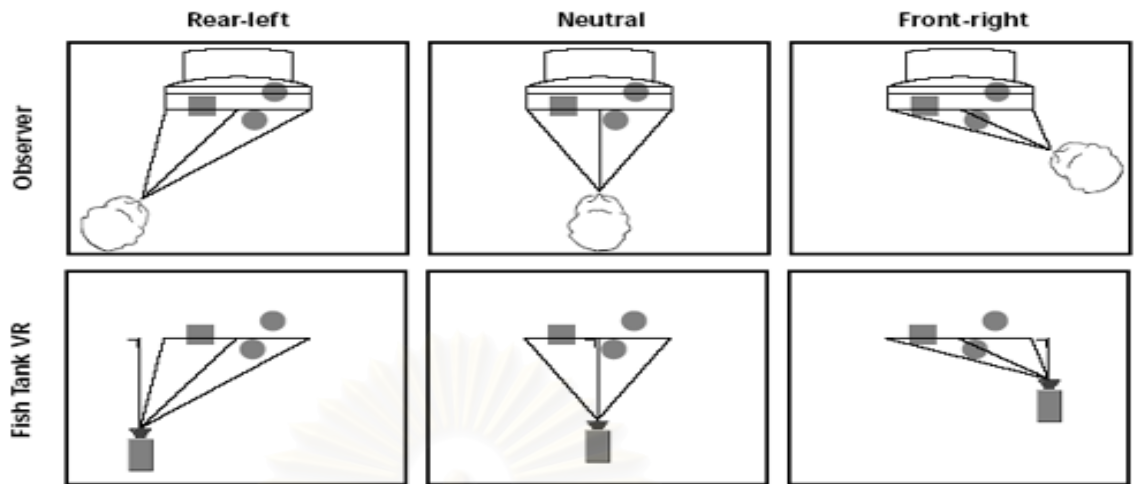


รูปที่ 3.16 รูปแบบของทัศนมิติใน OpenGL [14]

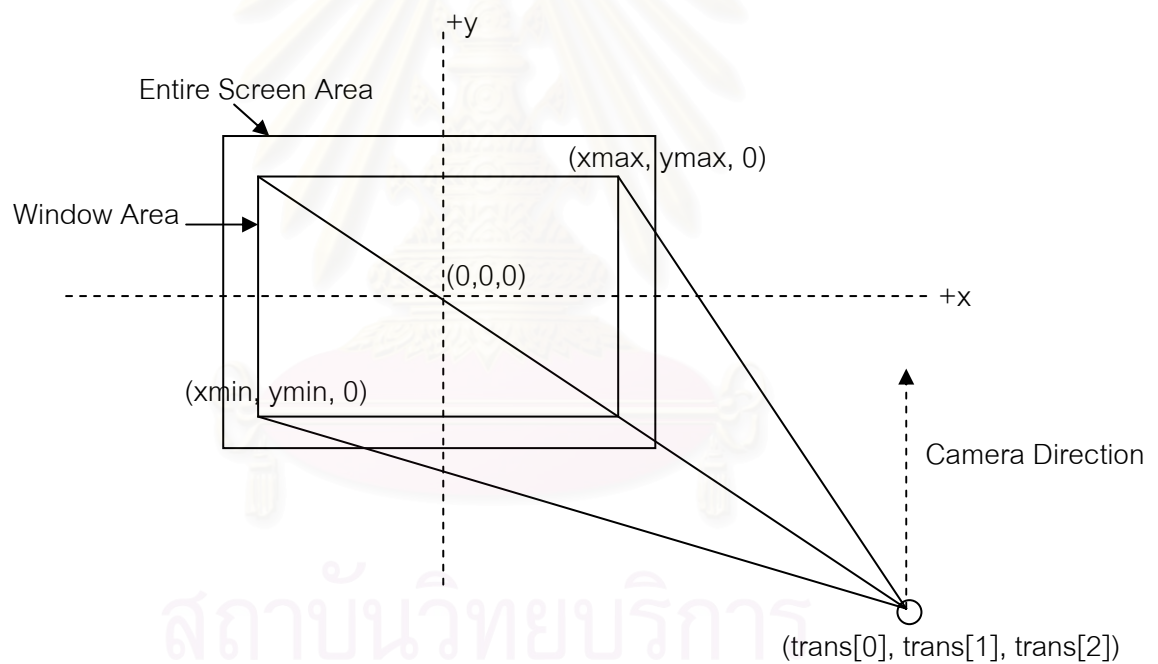
ตำแหน่งศีรษะของผู้ใช้เทียบกับจอภาพที่ได้ในหัวข้อที่แล้ว จะถูกนำมาเซตเป็นตำแหน่งให้จุด “Eye Point” เทียบกับ “Near plane” ซึ่งวัตถุสามมิติที่อยู่ระหว่าง “Near plane” และ “Far plane” จะถูกฉาย (หรือโปรเจก) ไปยัง “Near plane” โดยมีศูนย์กลางของการโปรเจกอยู่จุด “Eye point” ก่อให้เกิดเป็นภาพสองมิติเพื่อแสดงออกทางจอภาพ

รูปที่ 3.16 แสดงการตั้งทัศนมิติเมื่อศีรษะผู้ใช้อยู่ในแนวเดียวกับจอภาพและมองตั้งฉากกับจอภาพ ถ้าหากผู้ใช้เคลื่อนศีรษะไปทางซ้ายหรือขวาและมองทำมุมเอียงกับจอภาพ จะต้องเคลื่อนตำแหน่งกล้อง (Eye Point) ใน OpenGL ให้สัมพันธ์กัน (แต่ฉากรับภาพจะยังคงอยู่ที่เดิม ไม่เปลี่ยนแปลง) และกล้องต้องมีทิศทางตั้งฉากกับฉากรับภาพ “Near plane” เสมอ เพราะฉะนั้นการปรับทัศนมิติจะเป็นดังรูปที่ 3.17

จากรูปที่ 3.17 จะเห็นว่าฉากรับภาพ “Near plane” ไม่จำเป็นต้องสมมาตรกับกล้อง ซึ่ง OpenGL ได้เตรียมคำสั่งสำหรับการตั้งทัศนมิติเช่นนี้ไว้แล้ว คือ `glFrustum()`



รูปที่ 3.17 การปรับทัศนมิติเมื่อศีรษะผู้ใช้ทำมุมต่างๆ กันกับจอภาพ [3]



รูปที่ 3.18 แสดงการนำตำแหน่งศีรษะผู้ใช้มากำหนดเป็นตำแหน่งของกล้อง

จากรูปที่ 3.18 ตำแหน่งตาของผู้ใช้เทียบกับจอภาพที่ได้จากหัวข้อที่แล้ว จะเก็บไว้ในตัวแปรอาร์เรย์ ($trans[0]$, $trans[1]$, $trans[2]$) ซึ่งแทนพิกัด (x , y , z) ตามลำดับ สำหรับตัวแปร $xmin$, $ymin$, $xmax$, $ymax$ ก็คือขอบเขตของหน้าต่างที่จะแสดงภาพที่ได้จาก OpenGL (มีหน่วยเป็นเซนติเมตร) ซึ่งทัศนมิติแบบที่กล่าวมาสามารถกำหนดได้โดยใช้คำสั่งของ OpenGL ดังนี้

```

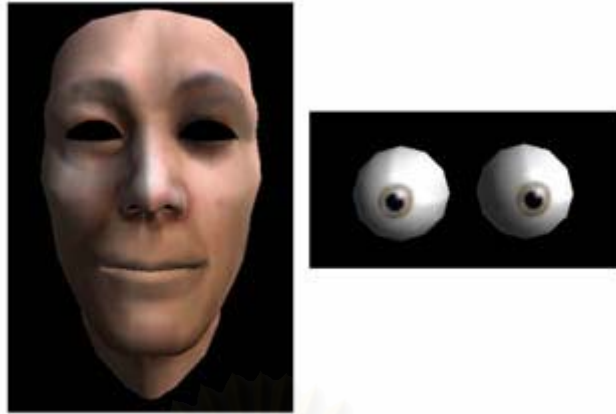
1:   glFrustum(
2:       xmin - trans[0],
3:       xmax - trans[0],
4:       ymin - trans[1],
5:       ymax - trans[1],
6:       trans[2],
7:       500.0f
8:   );
9:
10:  // Locate and align the camera.
11:  gluLookAt(
12:      trans[0],
13:      trans[1],
14:      trans[2],
15:      trans[0],
16:      trans[1],
17:      trans[2] - 1,
18:      0,1,0
19:  );

```

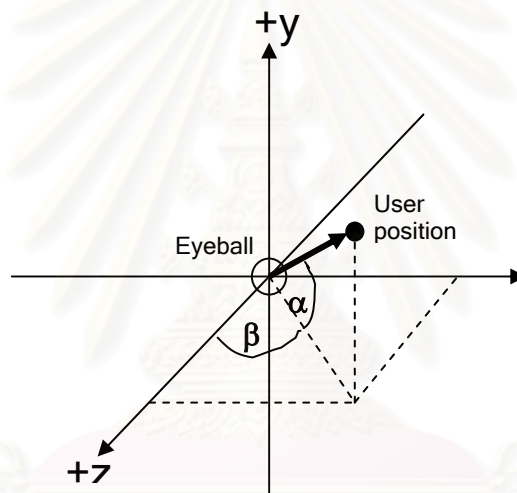
จากโค้ดตัวอย่าง ใช้คำสั่ง `glFrustum()` เพื่อกำหนดทัศนมิติของกล้องให้เป็นไปตามต้องการ จากนั้นจึงใช้คำสั่ง `gluLookAt()` เพื่อกำหนดตำแหน่งของกล้อง และทิศทางที่กล้องหันไป

3.3 การทำให้โมเดลสบตาผู้ใช้

เพื่อให้โมเดลสามารถสบตากับผู้ใช้ได้ จึงได้ออกแบบโมเดลสี่ระฆัง โดยใช้โมเดลสามชั้นหลักๆ คือ ส่วนศีรษะ ดวงตาซ้าย และดวงตาขวา มาประกอบกัน เพื่อให้สามารถแยกควบคุมโมเดลแต่ละชั้นได้อย่างอิสระ (ดังรูปที่ 3.19) การทำให้โมเดลมองไปยังจุดต่างๆ สามารถทำได้โดยการหมุนโมเดลดวงตาแต่ละข้าง รอบแกน x และ y เพื่อให้ลูกตาดำจ้องไปยังทิศทางที่ต้องการ



รูปที่ 3.19 แสดงส่วนประกอบต่างๆ ของโมเดลศีรษะมนุษย์



รูปที่ 3.20 แสดงมุมการหมุนของดวงตา

จากรูปที่ 3.20 จุดกำเนิดของแกนอยู่ตรงจุดศูนย์กลางของดวงตาที่ต้องการจะหมุน, User position คือตำแหน่งของผู้ใช้ที่คำนวณได้ เพราะฉะนั้น ดวงตาถูกนี้ต้องหมุนรอบแกน x เป็นมุม α และหมุนรอบแกน y เป็นมุมเท่ากับ β องศา โดยที่

$$\alpha = \tan^{-1}\left(\frac{y}{\sqrt{x^2 + z^2}}\right) \quad \dots(3.6)$$

$$\beta = \tan^{-1}\left(\frac{x}{z}\right) \quad \dots(3.7)$$

เมื่อได้มุมแล้ว ก็สามารถสั่งดวงตาให้หมุนโดยใช้คำสั่ง `glRotatef()` ดวงตาก็มจะ
 จ้องไปยังตำแหน่งของผู้ใช้งาน

3.4 การพูดคุยโต้ตอบ

งานวิจัยนี้ได้เลือกใช้ A.L.I.C.E Bot [20][21][22][23] เป็นปัญญาประดิษฐ์ (Artificial Intelligence) ที่ทำหน้าที่สำหรับพูดคุยโต้ตอบกับผู้ใช้ เนื่องจาก A.L.I.C.E Bot เป็น A.I ที่มีมาตรฐานดี และได้รับความนิยมอย่างสูง มีการพัฒนาอย่างต่อเนื่อง ความสามารถในการโต้ตอบจะขึ้นอยู่กับข้อมูล AIML (Artificial Intelligence Markup Language) [17] ซึ่งเป็นที่เก็บความรู้ และประโยคพูดคุยที่ซับซ้อนๆ เปรียบเสมือนเป็นมันสมองของ A.L.I.C.E Bot สำหรับไฟล์ AIML มาตรฐานที่สามารถดาวน์โหลดได้ฟรีจากอินเทอร์เน็ต จะเก็บรูปแบบประโยคสนทนาไว้กว่า 41,000 รูปแบบ ซึ่งเพียงพอต่อการพูดคุยทั่วๆ ไป หากผู้ใช้ต้องการเพิ่มรูปแบบประโยคใหม่ๆ ก็สามารถทำได้โดยไม่ยาก สำหรับข้อมูล AIML จะเก็บในรูปแบบของ XML ซึ่งมีตัวอย่างแบบง่ายเป็นดังนี้

```

1: <category>
2:   <pattern>WHAT IS YOUR NAME</pattern>
3:   <template>My name is Anna.</template>
4: </category>
5: <category>
6:   <pattern>WHAT IS YOUR NAME AGAIN</pattern>
7:   <template>My name is Anna. Whats your name?</template>
8: </category>
9: <category>
10:  <pattern>WHAT IS YOUR NATIONALITY</pattern>
11:  <template>American.</template>
12: </category>

```

บทสนทนาแต่ละบทจะถูกกำหนดด้วย <category> เมื่อ A.L.I.C.E Bot รับประโยคพูดคุยเข้ามา จะนำประโยคนั้นมาเปรียบเทียบ (Matching) กับฐานข้อมูล AIML หากพบว่าตรงกับบทสนทนาอันไหน ก็จะเอาคำตอบของบทสนทนานั้นส่งออกไป ยกตัวอย่างเช่น ถ้าหากข้อความรับเข้าเป็น “What is your name” ก็จะถูกจับคู่กับคำตอบ “My name is Anna” ส่งเป็นผลลัพธ์ออกไป

A.L.I.C.E Bot ที่มีให้ดาวน์โหลดได้ทางอินเทอร์เน็ตนั้นมีอยู่ด้วยกันหลายเวอร์ชัน ใช้ได้กับหลายๆ ภาษา แต่เวอร์ชันที่ใช้ในงานวิจัยชิ้นนี้ ได้เลือกใช้ A.L.I.C.E Bot - Program D เนื่องจากเป็นเวอร์ชันที่เพิ่งได้รับการพัฒนาขึ้นมาใหม่ที่สุด และมีความสามารถในการทำให้เป็นเซิร์ฟเวอร์ได้ ซึ่งทั้งหมดถูกเขียนขึ้นโดยใช้ภาษาจาวา เพราะฉะนั้นเครื่องที่จะรัน A.L.I.C.E Bot เวอร์ชันนี้ได้จึงจำเป็นที่จะต้องลงจาวาเวอร์ชวลแมชชีน (Java Virtual Machine) เสียก่อน จากนั้นเมื่อเริ่มการทำงานของ A.L.I.C.E Bot หน้าจอจะมีข้อความดังรูปที่ 3.21

```

[18:02:18] Starting Alicebot Program D version 4.1.5
[18:02:19] Using Java VM 1.5.0_02-b09 from Sun Microsystems Inc.
[18:02:19] On Windows XP version 5.1 (x86)
[18:02:19] Predicates with no values defined will return:
"undefined".
[18:02:19] Initializing Multiplexor.
[18:02:20] Loading Graphmaster.
[18:02:20] Starting up with
"D:\Work\Thesis\ALICE_Project\conf\startup.xml".
[18:02:21] Configuring bot "TestBot-1".
[18:02:22] Loaded 287 input substitutions.
[18:02:22] Loaded 19 gender substitutions.
[18:02:22] Loaded 9 person substitutions.
[18:02:22] Loaded 60 person2 substitutions.
[18:02:22] Loaded 4 sentence-splitters.
[18:02:23] There is no "sr" element in AIML.
[18:02:23] 6000 categories loaded so far.
[18:02:25] 12000 categories loaded so far.
[18:02:26] 18000 categories loaded so far.
[18:02:26] 1 bots thinking with 23358 categories.
[18:02:26] Alicebot Program D (c) 1995-2002 A.L.I.C.E. AI Foundation
[18:02:26] All Rights Reserved.
[18:02:26] This program is free software; you can redistribute it
and/or
[18:02:26] modify it under the terms of the GNU General Public
License
[18:02:26] as published by the Free Software Foundation; either
version 2
[18:02:26] of the License, or (at your option) any later version.
[18:02:26] Alicebot Program D version 4.1.5 Build [00]
[18:02:26] 23358 categories loaded in 6.54 seconds.
[18:02:26] The AIML Watcher is not active.
[18:02:26] HTTP server listening at http://virtuallife:2001

TEST : [Ljava.lang.String;@183f74d
[18:02:28] Socket Agent started.
[18:02:28] Interactive shell: type "/exit" to shut down; "/help" for
help.
[18:02:28] Testy 1> My favorite movie is Memento or The Matrix. Have
you seen either?

[18:02:28] [Testy 1] user>

```

รูปที่ 3.21 แสดงหน้าจอการทำงานของ A.L.I.C.E Bot

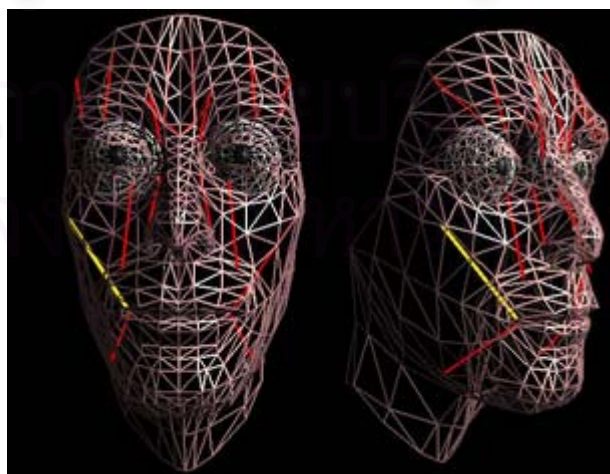
เมื่อขึ้นหน้าจอเช่นนี้ แสดงว่า A.L.I.C.E Bot ทำงานเป็นเซิร์ฟเวอร์พร้อมแล้ว สำหรับการติดต่อที่พอร์ตหมายเลข 2001 การติดต่อกับ A.L.I.C.E Bot เวอร์ชันนี้ก็เพียงแค่เขียนโปรแกรมให้เปิดซ็อกเกต (socket) ให้ติดต่อไปยังหมายเลขพอร์ต (port) ที่ A.L.I.C.E Bot ทำงานอยู่ แล้วส่งข้อความไปในรูปแบบนี้ "text=ข้อความที่อยากจะพูดคุย" A.L.I.C.E Bot ก็จะตอบรับเป็นข้อความตอบกลับมาจากที่ซ็อกเกตเดียวกัน

เนื่องจาก A.L.I.C.E Bot รับอินพุตและส่งออกเป็นข้อความตัวอักษร (Text) เพราะฉะนั้น ระบบจำเป็นต้องแปลงเสียงของผู้ใช้งานให้เป็นตัวอักษรเสียก่อนโดยใช้กลไกการรู้จำเสียงพูด (Speech Recognition Engine) จากนั้นส่งข้อความที่ได้ให้เป็นอินพุตของ A.L.I.C.E Bot จะได้อาตัพูตคำตอบออกมา คำตอบที่ได้ต้องเข้าสู่โมดูลกรองข้อความ (Text Filter) เพื่อปรับข้อความให้เหมาะสมกับการพูดออกเสียง (บางครั้ง A.L.I.C.E Bot อาจจะส่งประโยคออกมาในลักษณะนี้ “Yes..... : -)” จึงต้องปรับให้อยู่ในรูป “Yes.” มิฉะนั้นระบบจะออกเสียงผิดพลาด) จากนั้นจึงส่งต่อไปกับกลไกการสังเคราะห์เสียง (Text-to-speech engine) เพื่อแปลงข้อความเป็นเสียงพูดให้ผู้ใช้ได้ยิน ในระหว่างที่กลไกการสังเคราะห์เสียงทำงาน มันจะส่งข้อมูลรูปแบบการขยับปาก (Viseme) ออกมาด้วย สามารถนำข้อมูลนี้ไปใช้ประโยชน์ในการขยับปากโมเดลสามมิติให้สอดคล้องกับเสียงพูด ซึ่งจะกล่าวถึงโดยละเอียดต่อไปในภายหลัง

เพื่อให้ Speech Recognition Engine ทำงานได้ดี ผู้ใช้จำเป็นต้องเทรนระบบให้รู้จำเสียงพูดก่อน ขั้นตอนนี้อาจจะใช้เวลานาน งานวิจัยนี้จึงได้เพิ่มทางเลือกให้ผู้ใช้สามารถพิมพ์ข้อมูลเข้าโดยตรงได้ด้วย

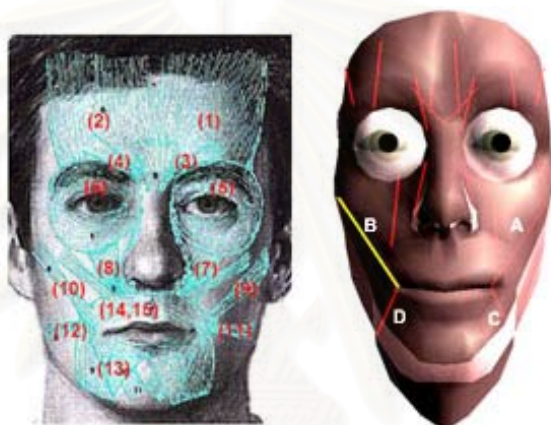
3.5 โมเดลสามมิติของศีรษะมนุษย์และระบบกล้ามเนื้อบนใบหน้า

วัตถุทุกชิ้นในโลกสามมิติจะประกอบขึ้นจากหน่วยย่อยที่สุดที่เรียกว่าโพลีกอน (Polygon) โมเดลใบหน้าเองก็เช่นกัน ประกอบจากโพลีกอนสามเหลี่ยมจำนวนมากแสดงเป็นเส้นไวร์เฟรม (Wireframe) ได้ดังรูปที่ 3.22



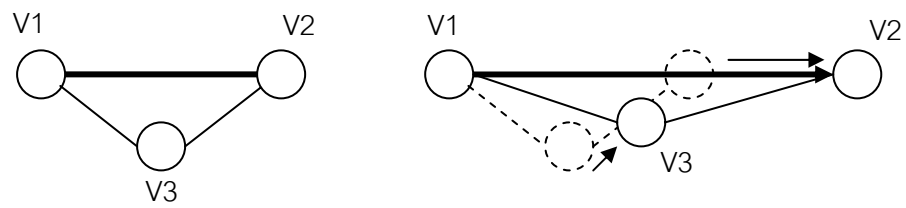
รูปที่ 3.22 แสดงเส้นไวร์เฟรมของโมเดลใบหน้าสามมิติ

เพื่อที่จะจำลองโมเดลศีรษะมนุษย์ให้เหมือนคนจริงมากที่สุด จึงได้จำลองระบบกล้ามเนื้อบนใบหน้าคน (Muscle System) [24][25][26][27] มาประยุกต์ใช้กับโมเดล เพื่อให้การขยับปากและแสดงอารมณ์เป็นไปอย่างธรรมชาติ กล่าวคือในขณะที่คนเราขยับปากหรือแสดงอารมณ์นั้น กล้ามเนื้อตามจุดต่างๆ จะยืดออก (Relax) หรือหดเข้า (Contract) ส่งผลให้ผิวหนังบนใบหน้าเกิดการขยับยืดหรือหดตามกล้ามเนื้อไปด้วยเกิดเป็นรูปทรงของปากและสีหน้าในแบบต่างๆ กัน เมื่อนำมาใช้กับโมเดล กล้ามเนื้อแต่ละเส้นจะถูกกำหนดโดยจุดยอด (Vertex) บนโพลีกอนจำนวนสองจุดเชื่อมเข้าหากัน ยกตัวอย่างเช่น จุดยอดของโพลีกอนบริเวณมุมปากด้านขวาของโมเดล เชื่อมกับจุดยอดของโพลีกอนบริเวณด้านบนของแก้มขวา จะได้เป็นกล้ามเนื้อสำหรับควบคุมการยกมุมปากด้านขวาขึ้น (แสดงดังเส้นสว่าง B ในรูปที่ 3.23)



รูปที่ 3.23 แสดงตำแหน่งกล้ามเนื้อบนใบหน้าคนและใบหน้าของโมเดล

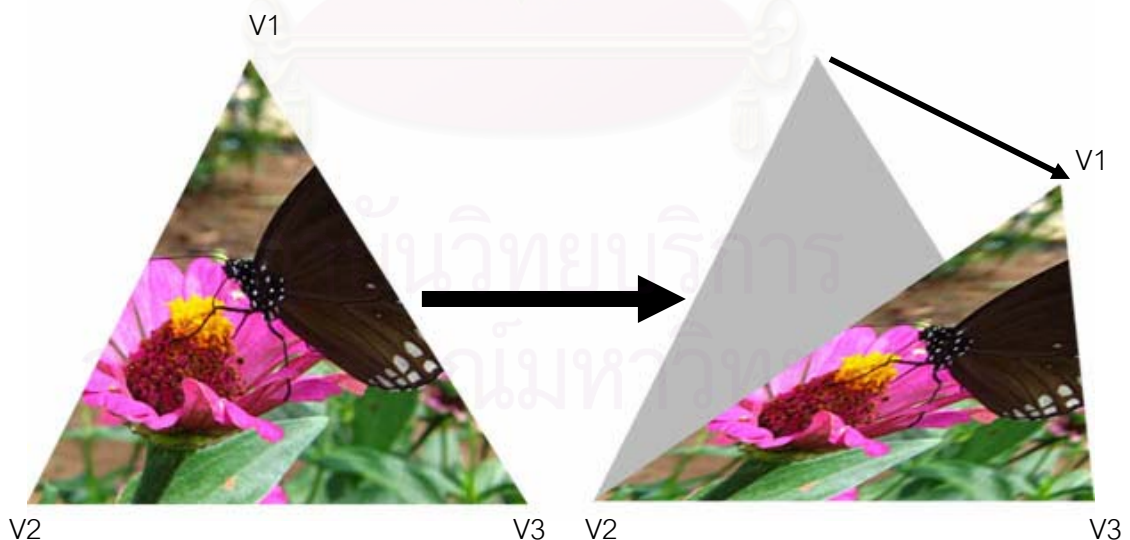
กล้ามเนื้อแต่ละเส้นสามารถยืดหดได้ การทำให้กล้ามเนื้อยืดออก คือการควบคุมจุดยอดสองจุดของกล้ามเนื้อนั้นให้ออกห่างจากกัน สำหรับการทำให้กล้ามเนื้อหดเข้า คือการควบคุมจุดยอดสองจุดของกล้ามเนื้อนั้นให้ชิดกันมากกว่าเดิม สำหรับจุดยอดอื่นๆ ที่ใกล้เคียงจะต้องนำมาคำนวณทางคณิตศาสตร์เพื่อหาผลกระทบด้วย จุดอื่นๆ ควรจะเคลื่อนที่ให้สอดคล้องไปในทางเดียวกันกับจุดยอดทั้งสองจุดนั้น (เปรียบเสมือนทุกจุดติดสปริงต่อถึงกันไว้ในแนวเส้นไคร์เฟรม ทำให้การเคลื่อนที่ของจุดใดๆ สามารถส่งอิทธิพลถึงจุดรอบข้างได้)



รูปที่ 3.24 แสดงการย้ายจุดยอดของกลัมนเนื้อบนใบหน้า

รูป 3.24 แสดงการเคลื่อนไหวของจุดยอดบนแนวกลัมนเนื้อแบบง่าย เส้นสีดำที่บ คือเส้นแสดงแนวกลัมนเนื้อซึ่งมีจุด V1 และ V2 เป็นจุดยอด เมื่อกลัมนเนื้อเส้นนี้คลายตัว จุด V1 และ V2 จะเคลื่อนออกห่างจากกัน สำหรับจุดรอบข้างอื่นๆ เช่นจุด V3 ก็จะเคลื่อนในแนวที่สัมพันธ์กัน

เมื่อจุดยอดเกิดการเปลี่ยนแปลงตำแหน่ง รูปพื้นผิวโมเดล (Texture) ที่นำมาปะไว้ ก็ต้องปรับรูปให้สอดคล้องกับจุดยอดด้วยเช่นกัน รูปที่ 3.25 แสดงโพลีกอนที่นำพื้นผิวมาติดไว้ ประกอบด้วยจุดยอด V1, V2 และ V3 เมื่อจุดยอด V1 เคลื่อนที่ไปตำแหน่งอื่นก็จำเป็นต้องคำนวณการบิดเบี้ยวของภาพพื้นผิวให้สัมพันธ์กับตำแหน่งจุดโดยใช้เทคนิคการประมาณค่าระหว่างกลาง (Interpolation) เพื่อหาค่าสีของจุดภาพในรูปเป้าหมายที่ต้องการ



รูปที่ 3.25 แสดงผลจากการประมาณค่าระหว่างกลางของรูปพื้นผิว

เมื่อนำภาพพื้นผิวของใบหน้า ประกอบติดเข้ากับโครงโมเดลใบหน้าสามมิติ จะ
ได้ผลเป็นดังรูป 3.26



รูปที่ 3.26 แสดงการนำภาพพื้นผิวปะลงบนโมเดลใบหน้า (ซ้าย) และการลบเหลี่ยมโพลีกอน (ขวา)

จากรูปซ้ายจะเห็นว่าโมเดลใบหน้าที่มีลักษณะเป็นเหลี่ยม อย่างชัดเจน ทั้งนี้
เนื่องมาจากว่า โดยค่ามาตรฐานของ OpenGL นั้น จะคำนวณการตกกระทบของแสงโดยใช้
เวกเตอร์ปกติของผิวหน้าโพลีกอน (Face Normal) ทำให้แต่ละโพลีกอนมีการตกกระทบของแสง
และสะท้อนแสงที่ต่างกันออกไป จึงเห็นเป็นเหลี่ยมโพลีกอนได้อย่างชัดเจน สามารถแก้ไขได้โดย
การเปลี่ยนข้อมูลการคำนวณการตกกระทบของแสงจากเดิมที่ใช้เวกเตอร์ปกติของผิวหน้าโพลีกอน
ให้เป็น เวกเตอร์ปกติของจุดยอด (Vertex Normal) (รูปที่ 3.27) จะทำให้ผลที่ได้ดูนุ่มนวล
(Smooth shading) และเหลี่ยมคมของโพลีกอนหายไป

- The **normal** of a triangle face is the direction perpendicular to the face



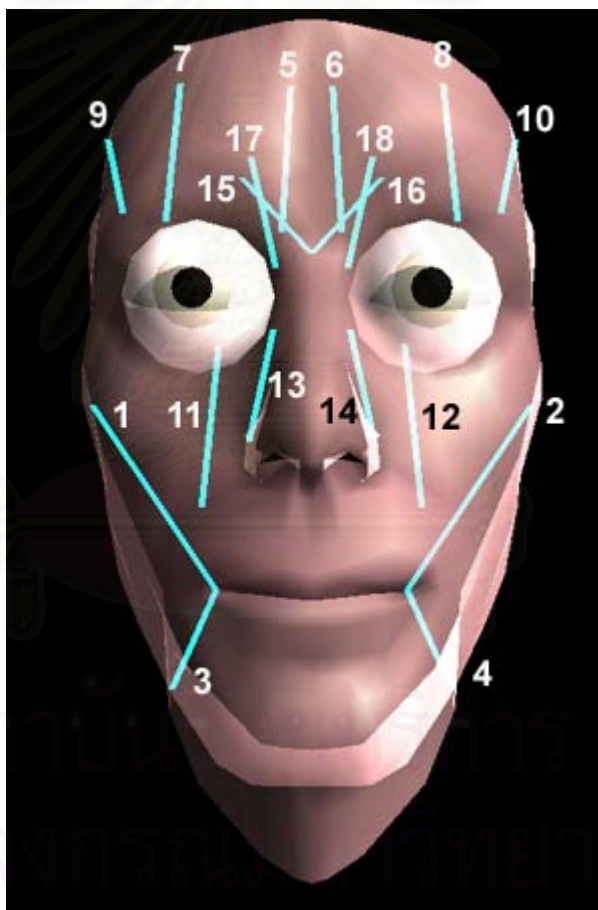
- The **normal** of a triangle vertex in a mesh is (usually) computed by averaging the normals of all faces on which it lies



รูปที่ 3.27 แสดงเวกเตอร์ปกติของผิวหน้าและจุดยอดโพลีกอน

ข้อดีที่โดดเด่นอีกอย่างหนึ่งสำหรับระบบกล้ามเนื้อคือ ข้อมูลการยึดหดกล้ามเนื้อจะเป็นอิสระกับโมเดลใบหน้า ด้วยเหตุนี้จึงทำให้ข้อมูลการยึดหดกล้ามเนื้อเพียงชุดเดียวสามารถนำไปประยุกต์ใช้ได้กับโมเดลใบหน้าหลายๆ แบบ ผู้พัฒนาเพียงแค่สร้างโมเดลใบหน้าและกำหนดจุดยอดของกล้ามเนื้อทั้งหมดเอาไว้ ก็สามารถนำมาใช้แสดงอารมณ์และขยับปากพูดคุยได้ทันที ไม่จำเป็นต้องทำข้อมูลอารมณ์สำหรับโมเดลทุกอันทำให้ประหยัดเวลาไปได้อย่างมาก

สำหรับกล้ามเนื้อที่ใช้ในงานวิจัยนี้ มีทั้งหมด 18 แบบกระจายตามจุดต่างๆ บนใบหน้า แบ่งเป็นสองชุด ชุดละ 9 แบบสำหรับใบหน้าสี่ข้างและ 9 แบบสำหรับใบหน้าสี่ขวา กล้ามเนื้อทั้งหมดนี้ออกแบบเพื่อให้ครอบคลุมการขยับปากเพื่อพูดและแสดงอารมณ์ในแบบต่างๆ กัน สำหรับตำแหน่งกล้ามเนื้อตามจุดต่างๆ แสดงได้ดังรูปที่ 3.28



รูปที่ 3.28 แสดงตำแหน่งทั้งหมดของกล้ามเนื้อที่ใช้ในงานวิจัยนี้

1. Left Zygomatic Major
2. Right Zygomatic Major
3. Left Angular Depressor
4. Right Angular Depressor

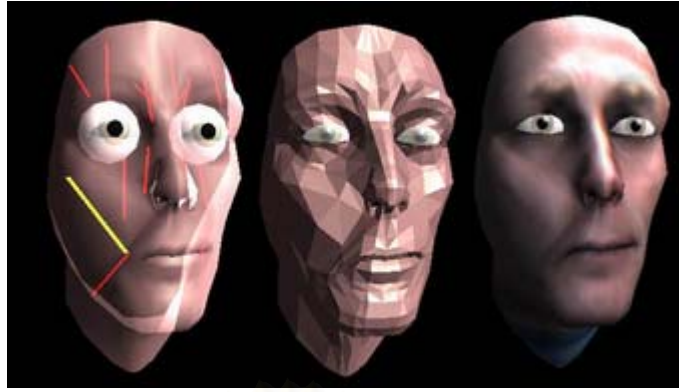
5. Left Frontalis Inner
6. Right Frontalis Inner
7. Left Frontalis Major
8. Right Frontalis Major
9. Left Frontalis Outer
10. Right Frontalis Outer
11. Left Labi Nasi
12. Right Labi Nasi
13. Left Inner Labi Nasi
14. Right Inner Labi Nasi
15. Left Lateral Corrigator
16. Right Lateral Corrigator
17. Left Secondary Frontalis
18. Right Secondary Frontalis

เพื่อให้เห็นภาพการทำงานของระบบกล้ามเนื้อ ขอยกตัวอย่างด้วยการหด (Contract) เส้นกล้ามเนื้อที่ชื่อว่า Left Zygomatic Major (กล้ามเนื้อที่เชื่อมระหว่างมุมปาก ด้านซ้ายและบริเวณบนของแก้มซ้าย) จะส่งผลให้มุมปากด้านซ้ายถูกยกขึ้น ลักษณะคล้ายกับการยิ้ม ได้ผลเป็นดังรูปที่ 3.29



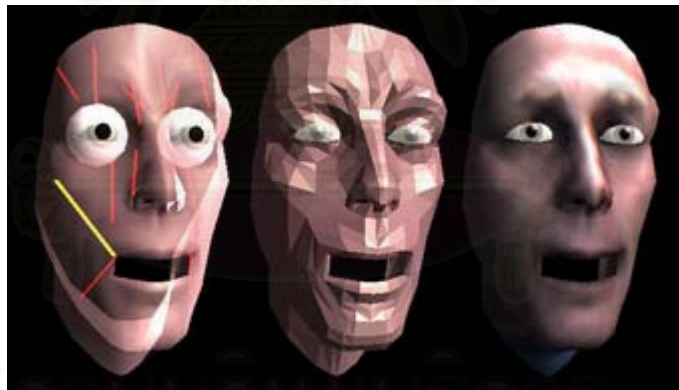
รูปที่ 3.29 แสดงการปรับกล้ามเนื้อเพื่อให้มีลักษณะการยิ้ม

ในทางกลับกัน เมื่อยืด (Relax) เส้นกล้ามเนื้อ Left Zygomatic Major จะส่งผลให้มุมปากซ้ายตกลง คล้ายกับคนหน้าบึ้ง ดังรูปที่ 3.30



รูปที่ 3.30 แสดงการปรับกล้ามเนื้อเพื่อให้มีลักษณะหน้าบึ้ง

การทำให้ปากเป็นรูปทรงต่างๆ ตามที่ต้องการนั้น การใช้ระบบกล้ามเนื้อเพียงอย่างเดียวไม่เพียงพอ เพราะกล้ามเนื้อทำได้เพียงแค่การยืดและหดพื้นผิวให้เป็นรูปต่างๆ แต่ไม่สามารถทำให้ปากอ้าได้ เพราะฉะนั้น จึงจำเป็นที่จะต้องจำลองระบบขากรรไกรมาประกอบด้วย จึงจะทำให้การจำลองปากเหมือนจริงที่สุด ขากรรไกรจะควบคุมองศาการอ้าปาก ถ้าหากค่าองศาเท่ากับศูนย์หมายถึงปากเม้มปิดสนิท แต่ถ้าองศามีค่าบวกเพิ่มขึ้น โพลีกอนบริเวณริมฝีปากบนและล่างก็จะถูกเคลื่อนย้ายให้ออกห่างจากกัน ปากก็จะอ้าตามค่าองศานั้น ดังรูปที่ 3.31



รูปที่ 3.31 แสดงการปรับขากรรไกรเพื่อให้อ้าปาก

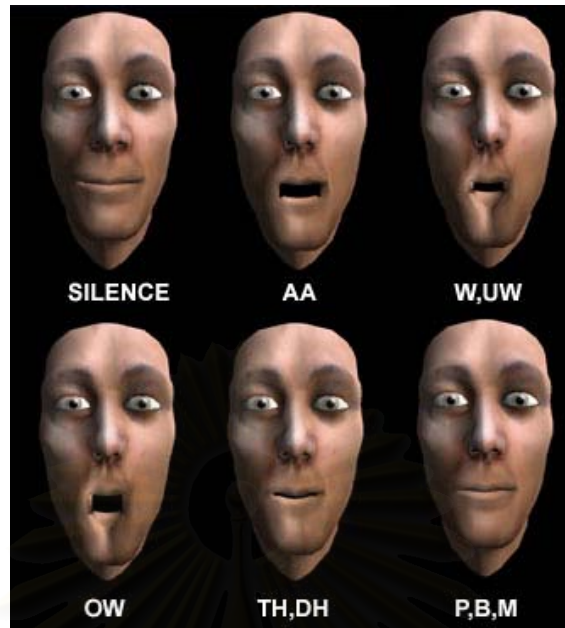
3.6 การขยับปากให้สอดคล้องกับเสียงพูด (Lip-sync)

ในระหว่างที่กลไกการสังเคราะห์เสียงทำงาน จะส่งสัญญาณที่เรียกว่า “Viseme” (รูปแบบของปากที่สัมพันธ์กับหน่วยเสียงที่เปล่งออกมาในขณะนั้น) ไปให้กับโมดูลศีรษะมนุษย์สามมิติเพื่อปรับรูปแบบปากให้สัมพันธ์กัน สำหรับ Viseme ทั้งหมดที่กลไกการสังเคราะห์เสียงสามารถส่งออกมาได้ มีอยู่ทั้งหมด 22 รูปแบบ คือ

1. SILENCE

2. AE, AX, AH
3. AA
4. AO
5. EY, EH, UH
6. ER
7. Y, IY, IH, IX
8. W, UW
9. OW
10. AW
11. OY
12. AY
13. H
14. R
15. L
16. S, Z
17. SH, CH, JH, ZH
18. TH, DH
19. F, V
20. D, T, N
21. K, G, NG
22. P, B, M

เมื่อโมเดลใบหน้าได้รับสัญญาณ Viseme เข้ามาก็จะนำมาจับคู่ (Match) เข้ากับรูปแบบปากที่กำหนดไว้ก่อนแล้ว เนื่องจากโมเดลที่ใช้ในงานวิจัยนี้ไม่มีส่วนของฟันและลิ้น จึงทำให้ไม่สามารถแสดงปากเพื่อให้สอดคล้องกับ Viseme ทั้งหมด 22 แบบได้ (หากให้มีฟันและลิ้นด้วยระบบใบหน้าจะซับซ้อนมาก อาจจะเกินขอบเขตของงานวิจัยไป และทำให้ใช้เวลาในการประมวลผลมาก) เพราะฉะนั้นในงานวิจัยนี้จะใช้รูปแบบปากหลักๆ 7 แบบ เพื่อรอการถูกจับคู่กับ Viseme ทั้ง 22 แบบ โดยเลือกจับคู่แบบที่คล้ายกันมากที่สุดหรือทดแทนกันได้ ผลการทดสอบที่ออกมาก็สมจริงเป็นที่น่าพอใจ สำหรับตัวอย่างปากแสดงดังรูปที่ 3.32



รูปที่ 3.32 ตัวอย่างปากที่สอดคล้องกับ Viseme แต่ละแบบ

เมื่อต้องการขยับปาก ชากรรไกรจะทำหน้าที่ขยับปากขึ้นลง ส่วนรูปทรงของปากจะเกี่ยวข้องกับกล้ามเนื้อทั้งสิ้นี่จุดได้แก่

1. Left Zygomatic Major (ควบคุมมุมปากด้านซ้ายในแนวเฉียงขึ้น)
2. Right Zygomatic Major (ควบคุมมุมปากด้านขวาในแนวเฉียงขึ้น)
3. Left Angular Depressor (ควบคุมมุมปากด้านซ้ายในแนวเฉียงลง)
4. Right Angular Depressor (ควบคุมมุมปากด้านขวาในแนวเฉียงลง)

กล้ามเนื้อทั้งสิ้นี่นี้จะบีบรัดหรือคลายตัวเพื่อให้รูปปากเป็นไปตามต้องการ เพราะฉะนั้น การเก็บลักษณะปากที่สอดคล้องกับ Viseme แต่ละแบบ จะประกอบไปด้วยข้อมูลการยืดหดกล้ามเนื้อทั้ง 4 เส้น และองศาของการอ้าชากรรไกร เมื่อมีสัญญาณ Viseme เข้ามา ก็เรียกใช้ข้อมูลเหล่านี้มาปรับกล้ามเนื้อและชากรรไกรให้เป็นไปตามต้องการ สำหรับข้อมูลทั้งหมดเป็นดังนี้ (ค่า + คือการหดกล้ามเนื้อ , ค่า - คือการยืดกล้ามเนื้อ)

1. Mouthshape_SILENCE

Jaw_Angle = 0.00

Left_Zygomatic_Major = 0.00

Left_Angular_Depressor = 0.00

Right_Zygomatic_Major = 0.00

Right_Angular_Depressor = 0.00

Left_Frontalis_Inner = 0.00	Right_Frontalis_Inner = 0.00
Left_Frontalis_Major = 0.00	Right_Frontalis_Major = 0.00
Left_Frontalis_Outer = 0.00	Right_Frontalis_Outer = 0.00
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

2. Mouthshape_TH_DH

Jaw_Angle = 0.30	
Left_Zygomatic_Major = -0.90	Right_Zygomatic_Major = -0.90
Left_Angular_Depressor = -0.40	Right_Angular_Depressor = -0.40
Left_Frontalis_Inner = 0.00	Right_Frontalis_Inner = 0.00
Left_Frontalis_Major = 0.00	Right_Frontalis_Major = 0.00
Left_Frontalis_Outer = 0.00	Right_Frontalis_Outer = 0.00
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

3. Mouthshape_P_B_M

Jaw_Angle = 0.00	
Left_Zygomatic_Major = -0.70	Right_Zygomatic_Major = -0.60
Left_Angular_Depressor = -0.20	Right_Angular_Depressor = -0.20
Left_Frontalis_Inner = 0.00	Right_Frontalis_Inner = 0.00
Left_Frontalis_Major = 0.00	Right_Frontalis_Major = 0.00
Left_Frontalis_Outer = 0.00	Right_Frontalis_Outer = 0.00
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

4. Mouthshape_E

Jaw_Angle = 1.00

Left_Zygomatic_Major = -1.30

Right_Zygomatic_Major = -1.00

Left_Angular_Depressor = -0.40

Right_Angular_Depressor = -0.30

Left_Frontalis_Inner = 0.00

Right_Frontalis_Inner = 0.00

Left_Frontalis_Major = 0.00

Right_Frontalis_Major = 0.00

Left_Frontalis_Outer = 0.00

Right_Frontalis_Outer = 0.00

Left_Labi_Nasi = 0.00

Right_Labi_Nasi = 0.00

Left_Inner_Labi_Nasi = 0.00

Right_Inner_Labi_Nasi = 0.00

Left_Lateral_Corigator = 0.00

Right_Lateral_Corigator = 0.00

Left_Secondary_Frontalis = 0.00

Right_Secondary_Frontalis = 0.00

5. Mouthshape_AA

Jaw_Angle = 1.50

Left_Zygomatic_Major = -1.30

Right_Zygomatic_Major = -1.00

Left_Angular_Depressor = -0.40

Right_Angular_Depressor = -0.30

Left_Frontalis_Inner = 0.00

Right_Frontalis_Inner = 0.00

Left_Frontalis_Major = 0.00

Right_Frontalis_Major = 0.00

Left_Frontalis_Outer = 0.00

Right_Frontalis_Outer = 0.00

Left_Labi_Nasi = 0.00

Right_Labi_Nasi = 0.00

Left_Inner_Labi_Nasi = 0.00

Right_Inner_Labi_Nasi = 0.00

Left_Lateral_Corigator = 0.00

Right_Lateral_Corigator = 0.00

Left_Secondary_Frontalis = 0.00

Right_Secondary_Frontalis = 0.00

6. Mouthshape_OW

Jaw_Angle = 2.00

Left_Zygomatic_Major = -1.60

Right_Zygomatic_Major = -1.60

Left_Angular_Depressor = -0.70

Right_Angular_Depressor = -0.70

Left_Frontalis_Inner = 0.00

Right_Frontalis_Inner = 0.00

Left_Frontalis_Major = 0.00

Right_Frontalis_Major = 0.00

Left_Frontalis_Outer = 0.00	Right_Frontalis_Outer = 0.00
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

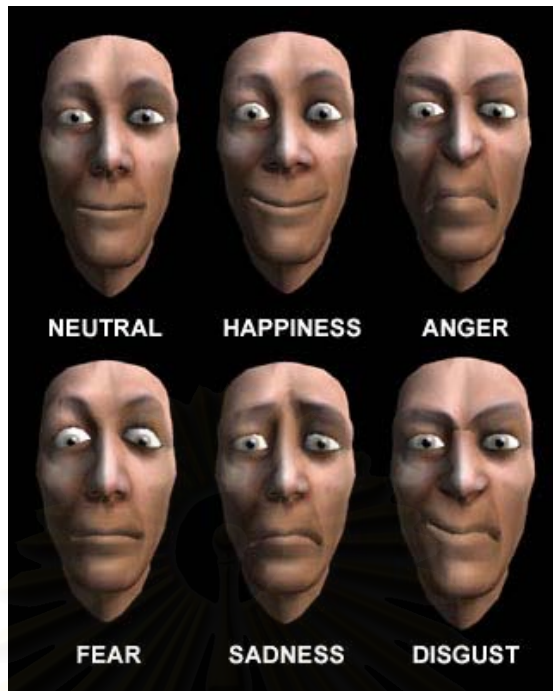
7. Mouthshape_W_UW

Jaw_Angle = 1.00	
Left_Zygomatic_Major = -1.60	Right_Zygomatic_Major = -1.60
Left_Angular_Depressor = -0.70	Right_Angular_Depressor = -0.70
Left_Frontalis_Inner = 0.00	Right_Frontalis_Inner = 0.00
Left_Frontalis_Major = 0.00	Right_Frontalis_Major = 0.00
Left_Frontalis_Outer = 0.00	Right_Frontalis_Outer = 0.00
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

3.7 การแสดงอารมณ์

การแสดงอารมณ์ จะใช้หลักการของกล้ามเนื้อควบคุม เช่นเดียวกับการขยับปาก ให้สอดคล้องกับเสียงพูด เพียงแต่การแสดงอารมณ์ต้องเกี่ยวข้องกับกล้ามเนื้อที่จุดอื่นๆ ด้วย นอกเหนือจากปาก เช่น กล้ามเนื้อที่แก้ม ตา หน้าผาก เป็นต้น เมื่อปรับค่าการยืดหดกล้ามเนื้อเหล่านี้ ก็จะได้เป็นลักษณะอารมณ์ที่แตกต่างกัน

อารมณ์ที่สามารถแสดงได้ทั้งหมดมีเจ็ดแบบ คือ ปกติ (Normal), ดีใจ (Happiness), โกรธ (Anger), ประหลาดใจ (Surprise), เศร้า (Sadness), หวาดกลัว (Fear), และขยะแขยง (Disgust) ยกตัวอย่างดังรูปที่ 3.33



รูปที่ 3.33 ตัวอย่างอารมณ์แบบต่างๆ ที่ใช้ในงานวิจัยนี้

สำหรับข้อมูลการขยับกล้ามเนื้อทั้งหมดเป็นดังนี้ (ค่า + คือการหดกล้ามเนื้อ , ค่า - คือการยืดกล้ามเนื้อ)

1. Neutral

Jaw_Angle = 0.00

Left_Zygomatic_Major = 0.00

Right_Zygomatic_Major = 0.00

Left_Angular_Depressor = 0.00

Right_Angular_Depressor = 0.00

Left_Frontalis_Inner = 0.00

Right_Frontalis_Inner = 0.00

Left_Frontalis_Major = 0.00

Right_Frontalis_Major = 0.00

Left_Frontalis_Outer = 0.00

Right_Frontalis_Outer = 0.00

Left_Labi_Nasi = 0.00

Right_Labi_Nasi = 0.00

Left_Inner_Labi_Nasi = 0.00

Right_Inner_Labi_Nasi = 0.00

Left_Lateral_Corigator = 0.00

Right_Lateral_Corigator = 0.00

Left_Secondary_Frontalis = 0.00

Right_Secondary_Frontalis = 0.00

2. Happiness

Jaw_Angle = 0.00

Left_Zygomatic_Major = 1.10	Right_Zygomatic_Major = 1.00
Left_Angular_Depressor = 0.00	Right_Angular_Depressor = 0.00
Left_Frontalis_Inner = 0.80	Right_Frontalis_Inner = 0.90
Left_Frontalis_Major = 0.20	Right_Frontalis_Major = 0.20
Left_Frontalis_Outer = 0.10	Right_Frontalis_Outer = 0.30
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.00
Left_Lateral_Corigator = 0.00	Right_Lateral_Corigator = 0.00
Left_Secondary_Frontalis = 0.00	Right_Secondary_Frontalis = 0.00

3. Anger

Jaw_Angle = 0.00	
Left_Zygomatic_Major = 0.00	Right_Zygomatic_Major = 0.00
Left_Angular_Depressor = 0.70	Right_Angular_Depressor = 0.60
Left_Frontalis_Inner = 0.20	Right_Frontalis_Inner = 0.00
Left_Frontalis_Major = 0.10	Right_Frontalis_Major = 0.00
Left_Frontalis_Outer = 2.00	Right_Frontalis_Outer = 2.00
Left_Labi_Nasi = 1.80	Right_Labi_Nasi = 1.90
Left_Inner_Labi_Nasi = 1.20	Right_Inner_Labi_Nasi = 1.10
Left_Lateral_Corigator = 1.40	Right_Lateral_Corigator = 1.30
Left_Secondary_Frontalis = 0.20	Right_Secondary_Frontalis = 0.30

4. Surprise

Jaw_Angle = 0.00	
Left_Zygomatic_Major = 0.20	Right_Zygomatic_Major = 0.20
Left_Angular_Depressor = 0.40	Right_Angular_Depressor = 0.40
Left_Frontalis_Inner = 1.40	Right_Frontalis_Inner = 0.90
Left_Frontalis_Major = 0.50	Right_Frontalis_Major = 0.40
Left_Frontalis_Outer = 0.20	Right_Frontalis_Outer = 0.70
Left_Labi_Nasi = 0.00	Right_Labi_Nasi = 0.00
Left_Inner_Labi_Nasi = 0.00	Right_Inner_Labi_Nasi = 0.10

Left_Lateral_Corigator = 0.00 Right_Lateral_Corigator = 0.00
 Left_Secondary_Frontalis = 0.50 Right_Secondary_Frontalis = 0.60

5. Sadness

Jaw_Angle = 0.00
 Left_Zygomatic_Major = 0.00 Right_Zygomatic_Major = 0.00
 Left_Angular_Depressor = 0.70 Right_Angular_Depressor = 0.70
 Left_Frontalis_Inner = 3.10 Right_Frontalis_Inner = 3.90
 Left_Frontalis_Major = 0.00 Right_Frontalis_Major = 0.00
 Left_Frontalis_Outer = 0.00 Right_Frontalis_Outer = 0.00
 Left_Labi_Nasi = 0.20 Right_Labi_Nasi = 0.70
 Left_Inner_Labi_Nasi = 0.20 Right_Inner_Labi_Nasi = 0.20
 Left_Lateral_Corigator = 0.00 Right_Lateral_Corigator = 0.00
 Left_Secondary_Frontalis = 0.00 Right_Secondary_Frontalis = 0.00

6. Fear

Jaw_Angle = 0.00
 Left_Zygomatic_Major = 0.60 Right_Zygomatic_Major = 0.50
 Left_Angular_Depressor = 0.50 Right_Angular_Depressor = 0.50
 Left_Frontalis_Inner = 0.20 Right_Frontalis_Inner = 0.20
 Left_Frontalis_Major = 0.60 Right_Frontalis_Major = 0.50
 Left_Frontalis_Outer = 0.20 Right_Frontalis_Outer = 0.20
 Left_Labi_Nasi = 0.20 Right_Labi_Nasi = 0.20
 Left_Inner_Labi_Nasi = 0.10 Right_Inner_Labi_Nasi = 0.10
 Left_Lateral_Corigator = 0.20 Right_Lateral_Corigator = 0.20
 Left_Secondary_Frontalis = 0.40 Right_Secondary_Frontalis = 0.30

7. Disgust

Jaw_Angle = 0.00
 Left_Zygomatic_Major = 0.80 Right_Zygomatic_Major = 0.10
 Left_Angular_Depressor = 0.00 Right_Angular_Depressor = 0.50

Left_Frontalis_Inner = 0.20

Right_Frontalis_Inner = 0.00

Left_Frontalis_Major = 0.00

Right_Frontalis_Major = 0.00

Left_Frontalis_Outer = 0.40

Right_Frontalis_Outer = 1.70

Left_Labi_Nasi = 1.10

Right_Labi_Nasi = 2.00

Left_Inner_Labi_Nasi = 0.70

Right_Inner_Labi_Nasi = 0.20

Left_Lateral_Corigator = 2.30

Right_Lateral_Corigator = 2.40

Left_Secondary_Frontalis = 0.00

Right_Secondary_Frontalis = 0.00



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลอง และผลที่ได้รับจากการทดลอง โดยการทดลองจะอิงกับทฤษฎีและแนวคิดที่กล่าวมาทั้งหมดจากบทที่แล้ว ซึ่งจะทดลองในหลายๆ ด้านเพื่อทดสอบความสามารถทั้งหมดของระบบที่พัฒนาขึ้น สำหรับการทดลองในทุกหัวข้อจะวิเคราะห์ผลการทดลองกำกับไว้ด้วย สุดท้ายก็จะเป็นการสรุปผลการทดลองทั้งหมด

4.1 ระบบที่ใช้ในการทดลอง

ระบบที่ใช้ในการทดลองมีคุณสมบัติดังนี้

1. ใช้ซีพียู Intel Celeron 2.40 GHz มีหน่วยความจำ 256 เมกะไบต์
2. ใช้การ์ดจอ nVidia GeForce FX5200, 128 เมกะไบต์
3. ติดตั้งกล่องวงจรปิดบริเวณด้านบนของจอมอนิเตอร์ ต่อสายคอมโพสิทวิดีโอ (Composite Video) ไปยังการ์ดจับภาพวิดีโอของ LifeView FlyVideo 2000 ในระบบ PAL มีจำนวนเฟรมเรต 25 เฟรมต่อวินาที
4. ติดตั้งหูฟังสเตอริโอและไมโครโฟน

ระบบที่ใช้ในการทดลองแสดงดังรูปที่ 4.1

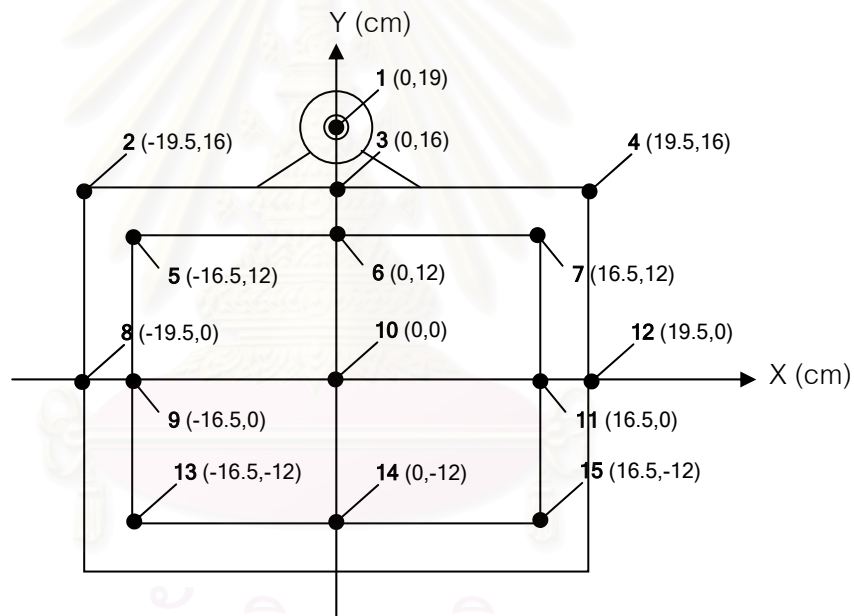


รูปที่ 4.1 ส่วนประกอบต่างๆ ของระบบที่ใช้ในการทดลอง

4.2 ผลการวัดความถูกต้องของการคำนวณตำแหน่งรูปแบบตรวจจับเทียบกับจอภาพ

หัวข้อนี้จะกล่าวถึงความแม่นยำของการคำนวณหาตำแหน่งของรูปแบบตรวจจับเทียบกับจอภาพ โดยการทดลองจะเปรียบเทียบระหว่างผลที่วัดจริงโดยคนและผลที่โปรแกรมคำนวณได้ สำหรับกระบวนการทดลองมีขั้นตอนดังนี้

1. กำหนดตำแหน่งทั้งหมดที่จะทดลอง โดยตำแหน่งทั้งหมดจะแบ่งออกเป็นสามระยะ คือ ระยะที่ตั้งฉากและห่างจากจอภาพเท่ากับ 80 , 60 และ 40 เซนติเมตร (ค่าในแกน z) แต่ระยะจะใช้ตำแหน่งในการทดลองเท่ากับ 15 ตำแหน่ง รอบๆ จอภาพ แสดงได้ดังรูปที่ 4.2 จากรูปจะแสดงหมายเลขจุดและพิกัดของจุดนั้นๆ จุดกึ่งกลางจอภาพจะมีพิกัด $(x,y) = (0,0)$ ตำแหน่งเลนส์ของกล้องจะมีพิกัด $(0,19)$ มีหน่วยเป็นเซนติเมตร เป็นต้น



รูปที่ 4.2 ตำแหน่ง (x,y) ที่ใช้ในการทดลอง

2. วางตำแหน่งรูปแบบตรวจจับไว้ที่ตำแหน่งต่างๆ เรียงลำดับจากจุดหมายเลข 1 จนถึงหมายเลขที่ 15 และมีระยะห่างที่ตั้งฉากกับจอภาพเท่ากับ 80 เซนติเมตร บันทึกผลการทดลองที่วัดโดยคนและที่คำนวณได้โดยโปรแกรม
3. ทำซ้ำข้อที่สอง สำหรับระยะที่ 60 และ 40 เซนติเมตร ตามลำดับ จะได้ผลการทดลองดังตาราง ที่ 4.1, 4.2 และ 4.3 (ค่าที่เป็นกากบาท X แสดงว่าไม่สามารถคำนวณได้เนื่องจากรูปแบบตรวจจับไม่ได้ อยู่ภายในมุมรับภาพของกล้องที่ใช้)

ตารางที่ 4.1 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 80 เซนติเมตร

No.	ผลการวัดโดยคน			ผลที่โปรแกรมคำนวณ ได้			ค่าความคลาดเคลื่อน		
	X	Y	Z	X	Y	Z	X	Y	Z
1	0	19	80	0.31	18.26	80.51	0.31	0.74	0.51
2	-19.5	16	80	-19.86	16.69	75.34	0.36	0.69	4.66
3	0	16	80	0.95	16.59	80.93	0.95	0.59	0.93
4	19.5	16	80	18.63	16.38	80.34	0.87	0.38	0.34
5	-16.5	12	80	-16.01	12.85	74.23	0.49	0.85	5.77
6	0	12	80	-0.31	12.72	80.34	0.31	0.72	0.34
7	16.5	12	80	16.25	12.67	78.47	0.25	0.67	1.53
8	-19.5	0	80	-18.99	1.71	65.83	0.51	1.71	14.17
9	-16.5	0	80	-16.12	1.99	67.99	0.38	1.99	12.01
10	0	0	80	1.47	1.77	81.03	1.47	1.77	1.03
11	16.5	0	80	16.5	1.29	82.69	0	1.29	2.69
12	19.5	0	80	20.96	1.88	81.31	1.46	1.88	1.31
13	-16.5	-12	80	X	X	X	X	X	X
14	0	-12	80	X	X	X	X	X	X
15	16.5	-12	80	X	X	X	X	X	X
ค่าความคลาดเคลื่อนเฉลี่ยของแกน X = 0.6 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Y = 1.1 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Z = 3.8 เซนติเมตร									

ตารางที่ 4.2 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 60 เซนติเมตร

No.	ผลการวัดโดยคน			ผลที่โปรแกรมคำนวณ ได้			ค่าความคลาดเคลื่อน		
	X	Y	Z	X	Y	Z	X	Y	Z
1	0	19	60	0.07	20.03	63.24	0.07	1.03	3.24
2	-19.5	16	60	-18.46	16.62	51	1.04	0.62	9
3	0	16	60	-0.91	16.82	67.54	0.91	0.82	7.54
4	19.5	16	60	19.83	17.15	60.35	0.33	1.15	0.35
5	-16.5	12	60	-16.21	12.83	52.98	0.29	0.83	7.02
6	0	12	60	0.84	12.13	67.83	0.84	0.13	7.83
7	16.5	12	60	16.28	11.26	60.54	0.22	0.74	0.54
8	-19.5	0	60	-14.71	3.94	37.56	4.79	3.94	22.44
9	-16.5	0	60	-14.66	5.57	48.05	1.84	5.57	11.95
10	0	0	60	0.76	0.65	60.33	0.76	0.65	0.33
11	16.5	0	60	16.49	1.01	59.39	0.01	1.01	0.61
12	19.5	0	60	19.38	1.32	56.19	0.12	1.32	3.81
13	-16.5	-12	60	X	X	X	X	X	X
14	0	-12	60	X	X	X	X	X	X
15	16.5	-12	60	X	X	X	X	X	X
ค่าความคลาดเคลื่อนเฉลี่ยของแกน X = 0.9 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Y = 1.5 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Z = 6.2 เซนติเมตร									

ตารางที่ 4.3 ผลการทดลองการคำนวณตำแหน่งรูปแบบตรวจจับที่ระยะ 40 เซนติเมตร

No.	ผลการวัดโดยคน			ผลที่โปรแกรมคำนวณ ได้			ค่าความคลาดเคลื่อน		
	X	Y	Z	X	Y	Z	X	Y	Z
1	0	19	40	-0.52	19.28	40.8	0.52	0.28	0.8
2	-19.5	16	40	X	X	X	X	X	X
3	0	16	40	0.56	16.51	42.8	0.56	0.51	2.8
4	19.5	16	40	X	X	X	X	X	X
5	-16.5	12	40	X	X	X	X	X	X
6	0	12	40	0.6	12.48	41.51	0.6	0.48	1.51
7	16.5	12	40	17.17	12.64	36.22	0.67	0.64	3.78
8	-19.5	0	40	X	X	X	X	X	X
9	-16.5	0	40	X	X	X	X	X	X
10	0	0	40	X	X	X	X	X	X
11	16.5	0	40	X	X	X	X	X	X
12	19.5	0	40	X	X	X	X	X	X
13	-16.5	-12	40	X	X	X	X	X	X
14	0	-12	40	X	X	X	X	X	X
15	16.5	-12	40	X	X	X	X	X	X
ค่าความคลาดเคลื่อนเฉลี่ยของแกน X = 0.6 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Y = 0.5 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Z = 2.2 เซนติเมตร									

จากการทดลอง มีอุปสรรคสำคัญอย่างหนึ่งคือมุมรับภาพของกล้องค่อนข้างที่จะแคบ ทำให้ตำแหน่งของรูปแบบตรวจจ็บบ้างเป็นต้องเคลื่อนไหวอยู่ในมุมที่จำกัดพอสมควร จากผลการทดลอง ค่าที่เป็นกากบาท (X) คือค่าที่ไม่สามารถคำนวณได้เนื่องจากรูปแบบตรวจจ็บบอกนอกมุมรับภาพของกล้อง ดังเช่นค่าของจุดทดสอบที่ 13, 14 และ 15 ของการทดลองทั้งสามตาราง ที่ไม่สามารถตรวจสอบได้เลย เนื่องจากอยู่ส่วนล่างของจอภาพ ทำให้อยู่นอกมุมรับภาพของกล้องที่ติดอยู่ด้านบนจอภาพ ถึงแม้ว่าจะถอยห่างไปถึงระยะ 80 เซนติเมตรแล้วก็ตาม (หากถอยห่างกว่านี้ เช่นถอยไปถึงที่ระยะ 90 ถึง 100 เซนติเมตร โปรแกรมจะไม่สามารถตรวจจ็บบ้างได้เลย เนื่องจากรูปแบบตรวจจ็บบ้างจะมีขนาดเล็กเกินไป)

จากผลการทดลองโดยรวมที่ออกมา ถือว่าความแม่นยำอยู่ในระดับที่ดี นั่นคือค่าความคลาดเคลื่อนเฉลี่ยของแกน X และ Y อยู่ในระดับที่ไม่เกิน 2 เซนติเมตร แต่ค่าความคลาดเคลื่อนของแกน Z อาจสูงเกินกว่า เช่นที่ระยะ 60 เซนติเมตร ค่า Z มีความคลาดเคลื่อนเฉลี่ยสูงถึง 6.2 เซนติเมตร สำหรับจุดที่มีความคลาดเคลื่อนสูงที่สุดคือจุดหมายเลข 8 และ 9 จุดสองจุดนี้มีผลการคำนวณผิดพลาดที่สูงที่สุดทั้งตารางที่ 1 และตารางที่ 2 (ตารางที่ 3 ไม่สามารถตรวจจ็บบ้างจุดนี้ได้) หากดูในแผนผังจะพบว่าจุดที่ 8 และ 9 จะอยู่บริเวณขอบจอด้านซ้ายกลาง (จะอยู่บริเวณมุมซ้ายล่างเมื่อเทียบกับกล้อง) ทั้งนี้เนื่องจากยังเกิดการบิดเบี้ยวในแนวรัศมี (Radial Distortion) ที่บริเวณมุมด้านซ้ายอยู่ ยังไม่สามารถที่จะกำจัดการบิดเบี้ยวออกไปได้ทั้งหมด ขั้นตอนในการเทียบมาตรฐานกล้อง (Calibration) ให้สัมพันธ์กับการบิดเบี้ยวในแนวรัศมีที่ไม่แม่นยำร้อยเปอร์เซ็นต์ ส่งผลให้ผลการตรวจจ็บบ้างตำแหน่งเฉพาะในบริเวณนั้นเกิดการผิดพลาด

สำหรับตารางที่ 4.3 ที่ตรวจจ็บบ้างตำแหน่งในระยะ 40 เซนติเมตร พบว่ามีหลายจุดที่ไม่สามารถทดสอบได้เนื่องระยะนี้เป็นระยะที่ใกล้กับกล้องมาก ทำให้พื้นที่ในการรับภาพยิ่งแคบ จุดส่วนใหญ่จึงออกนอกมุมรับภาพของกล้องไป ทำให้ตรวจสอบได้เพียงสี่จุดทดสอบเท่านั้น ถ้าสังเกตจากจุดทดสอบ ที่ 5 และ 7 จะพบว่าจุดสองจุดนี้สมมาตรกันเมื่อเทียบกับกล้อง โปรแกรมสามารถตรวจจ็บบ้างจุดที่ 7 ได้ เพราะฉะนั้นในทางทฤษฎีควรตรวจจ็บบ้างจุดที่ 5 ได้ด้วย (เพราะน่าจะอยู่ในมุมรับภาพเหมือนกัน) แต่ในความเป็นจริงไม่สามารถตรวจจ็บบ้างจุดที่ 5 ได้ เนื่องจากการบิดเบี้ยวในแนวรัศมีด้านซ้ายที่ยังคงอยู่เองที่ทำให้ภาพเกิดการบิดเบี้ยวจนตกขอบไป

ในการทดลองนี้ โปรแกรมสามารถตรวจจ็บบ้างตำแหน่งได้อย่างรวดเร็ว โดยมีความเร็วถึง 24.9 เฟรมต่อวินาที ซึ่งน่าพอใจ แต่อย่างไรก็ตาม มีบางเฟรมที่มีข้อผิดพลาดในการตรวจจ็บบ้าง ถึงแม้จะอยู่ในมุมรับภาพของกล้องก็ตาม โดยเฟรมที่ตรวจจ็บบ้างไม่ได้นั้นอยู่ที่ประมาณ 3%-

5% ของเฟรมทั้งหมด ข้อผิดพลาดจะพบบ่อยขึ้นหากตำแหน่งของรูปแบบตรวจจับอยู่ไกลจากกล้อง โดยเฉพาะในระยะ 80 เซนติเมตรขึ้นไป

4.3 ผลการวัดความถูกต้องของโมเดลในการมองไปยังจุดต่างๆ รอบตัวผู้ใช้

ในหัวข้อนี้จะทำการทดลองการวัดความถูกต้องของโมเดลในการมองไปยังจุดต่างๆ รอบตัวผู้ใช้ โดยจะกำหนดให้โมเดลมองไปยังพิกัดหนึ่งๆ ก่อน จากนั้นจึงให้ผู้ใช้ประมาณเอาว่าโมเดลมองไปยังจุดไหน แล้ววัดเพื่อหาค่าพิกัดจุดนั้นออกมา จากนั้นจึงนำค่าพิกัดทั้งสองมาเปรียบเทียบกันเพื่อหาค่าความคลาดเคลื่อน สำหรับในการทดลองนี้จะทดลองที่ระยะที่ห่างจากจอภาพในแนวตั้งฉากเพียงระยะเดียวคือ 60 เซนติเมตร เนื่องจากว่าผู้ใช้ไม่มีทางรู้ได้โดยที่โมเดลมองไปยังจุดใกล้หรือไกลแค่ไหน หรือโฟกัสไปยังระยะไหน เพราะฉะนั้นพิกัดที่สังเกตโดยคนจะให้ค่า Z เท่ากันทั้งหมด สำหรับผลการทดลองเป็นดังตารางที่ 4.4

ตารางที่ 4.4 เปรียบเทียบพิกัดที่โมเดลมองและพิกัดที่สังเกตได้โดยคน

No.	พิกัดที่กำหนดให้โมเดล			พิกัดที่สังเกตโดยคน			ค่าความคลาดเคลื่อน		
	X	Y	Z	X	Y	Z	X	Y	Z
1	0	19	60	0	29	60	0	10	0
2	-16.5	12	60	-20	14	60	3.5	2	0
3	0	12	60	0	14	60	0	2	0
4	16.5	12	60	19	12	60	2.5	0	0
5	-16.5	0	60	-21	0	60	4.5	0	0
6	0	0	60	0	0	60	0	0	0
7	16.5	0	60	18	0	60	1.5	0	0
ค่าความคลาดเคลื่อนเฉลี่ยของแกน X = 1.7 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Y = 2.0 เซนติเมตร									
ค่าความคลาดเคลื่อนเฉลี่ยของแกน Z = 0.0 เซนติเมตร									

จากตารางที่ 4.4 ค่าความคลาดเคลื่อนเฉลี่ยถือว่าอยู่ในเกณฑ์ที่ดี แต่หากดูจุดทดสอบรายตัว จะเห็นว่าจุดที่ไกลจากจุดศูนย์กลางจอภาพมากจะมีความคลาดเคลื่อนมาก โดยเฉพาะจุดที่ 1 มีความคลาดเคลื่อนทางแกน Y ถึง 10 เซนติเมตร ซึ่งเกิดจากความผิดพลาดในการประเมินตำแหน่งของคน และรวมกับความผิดพลาดของการคำนวณตำแหน่งโดยโปรแกรม สำหรับจุดที่มีค่า X หรือ Y เป็น 0 คนสามารถสังเกตได้ง่าย จึงไม่ค่อยพบความคลาดเคลื่อน โดยเฉพาะจุดที่ 6 ซึ่งจ้องตรงมายังผู้ใช้พอดี ทำให้ไม่มีความคลาดเคลื่อนเกิดขึ้นเลย

4.4 ผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้ เมื่อไม่ใช้การแสดงผลแบบพีชแท่งกีวาร์

การทดสอบในหัวข้อนี้ จะเป็นการให้ผู้ใช้สวมหมวกติดรูปแบบตรวจจับแล้วทดลองเคลื่อนที่ศีรษะไปยังจุดต่างๆ กัน แล้วตรวจผลที่ได้ว่าโมเดลมีปฏิกิริยาอย่างไร โดยกำหนดให้โมเดลต้องจ้องตาผู้ใช้ตลอดและยังไม่เปิดใช้ระบบพีชแท่งกีวาร์ ผลที่ได้เป็นดังรูปที่ 4.3



(a)



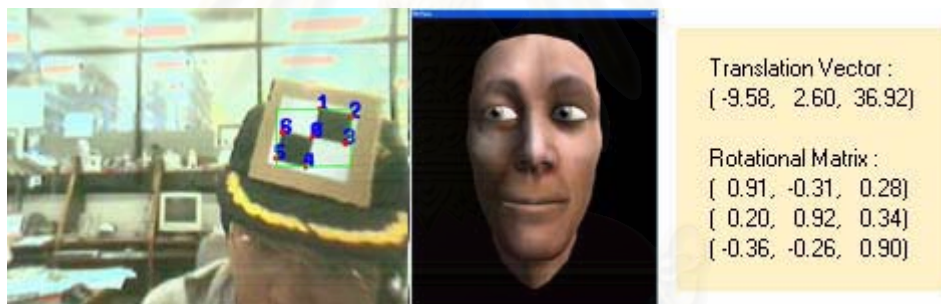
(b)



(c)



(d)



(e)



(f)

รูปที่ 4.3 แสดงผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้และไม่ใช้ระบบพิกซาทงกวีอาร์

จากรูปที่ 4.3 จะเห็นได้ว่า เมื่อผู้ใช้เคลื่อนศีรษะไปทางขวา (รูปที่ 4.3 (b)) โมเดล จะกรอกลูกตามองไปยังผู้ใช้ซึ่งอยู่ด้านซ้ายมือของโมเดล เพราะหันหน้าเข้าหากัน สำหรับเวกเตอร์ การเคลื่อนที่ (Translation Vector) จะแสดงตำแหน่งศีรษะของผู้ใช้เทียบกับจอภาพ และ เมตริกซ์ การหมุน (Rotational Matrix) จะแสดงมุมศีรษะของผู้ใช้ในขณะนั้น

4.5 ผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้ เมื่อเปิดใช้การแสดงผลแบบพีชแทงก์วี่อาร์

สำหรับหัวข้อนี้ การทดสอบจะคล้ายคลึงกับหัวข้อที่ 4.4 แต่ในหัวข้อนี้จะเปิดใช้ระบบพีชแทงก์วี่อาร์ด้วย ผลที่ได้แสดงดังรูปที่ 4.4



(a)



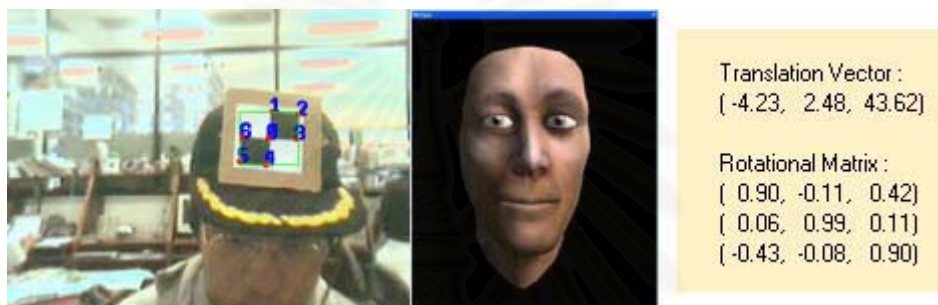
(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)



(j)



(k)



(l)



(m)



(n)

รูปที่ 4.4 แสดงผลที่ได้จากระบบเมื่อให้โมเดลมองไปยังผู้ใช้และเปิดใช้ระบบฟิชแทงก์วัวร์

จากรูปที่ 4.4 จะเห็นได้ว่าเมื่อเปิดใช้ระบบฟิชแทงก์วัวร์ มุมมองของโมเดลจะถูกปรับให้สัมพันธ์กับตำแหน่งของใบหน้าผู้ใช้โดยอัตโนมัติ เพราะฉะนั้นเมื่อผู้ใช้เคลื่อนศีรษะไปทางขวามือ ก็ควรที่จะเห็นแก้มซ้ายและใบหน้าด้านซ้ายของโมเดลมากขึ้น ดังรูปที่ 4.4 (b) ในขณะเดียวกันโมเดลก็จะจ้องหน้าผู้ใช้ไปด้วย

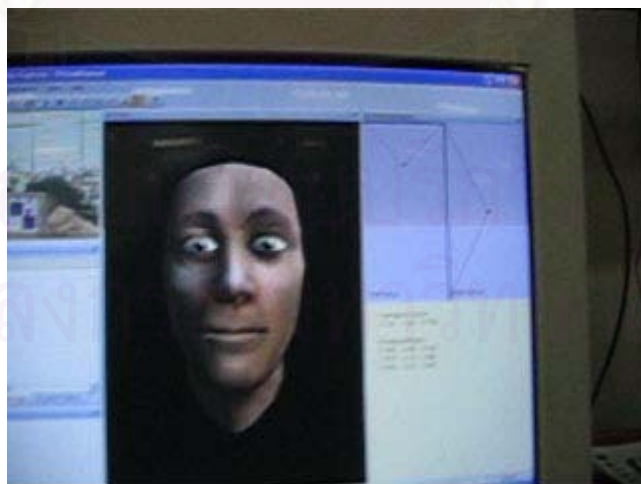
ถึงแม้ว่าผู้ใช้จะเอียงหรือหันศีรษะมองไปทางอื่น โมเดลก็ยังคงมองไปยังผู้ใช้ และโปรแกรมก็ยังสามารถที่จะคำนวณหาตำแหน่งของผู้ใช้ได้อย่างถูกต้อง แม้รูปแบบตรวจจับจะไม่ตั้งฉากกับกล้อง ดังรูปที่ 4.4 (d)

สำหรับรูปที่ 4.4 (l), 4.4 (m), 4.4 (n) แสดงเฟรมที่โปรแกรมไม่สามารถตรวจหาตำแหน่งของรูปแบบตรวจจับได้ เนื่องด้วยมาจากมุมบางมุมที่แสงไฟฟลูออเรสเซนต์ส่องกระทบรูปแบบตรวจจับมากเกินไป ทำให้สภาพแสงเปลี่ยนแปลงและเกิดข้อผิดพลาดในขั้นตอนการทำเทรชโฮล (Threshold) ภาพ หรือเกิดจากผู้ใช้งานเงาหน้ามากเกินไปจนทำให้ส่วนหนึ่งของหมวกบังรูปแบบตรวจจับ ในกรณีที่เฟรมปัจจุบันไม่สามารถตรวจจับตำแหน่งได้ โมเดลใบหน้าจะยังคงอยู่ในสถานะเดิมจากเฟรมก่อนหน้าล่าสุดที่คำนวณตำแหน่งได้สำเร็จ

สำหรับรูปที่ถ่ายมาจากมุมมองของผู้ใช้ เมื่อผู้ใช้เคลื่อนศีรษะไปทางซ้าย, มองตรง และ เคลื่อนศีรษะไปทางขวา แสดงดังรูปที่ 4.5 (a), 4.5 (b), และ 4.5 (c) ตามลำดับ



(a)



(b)



(c)

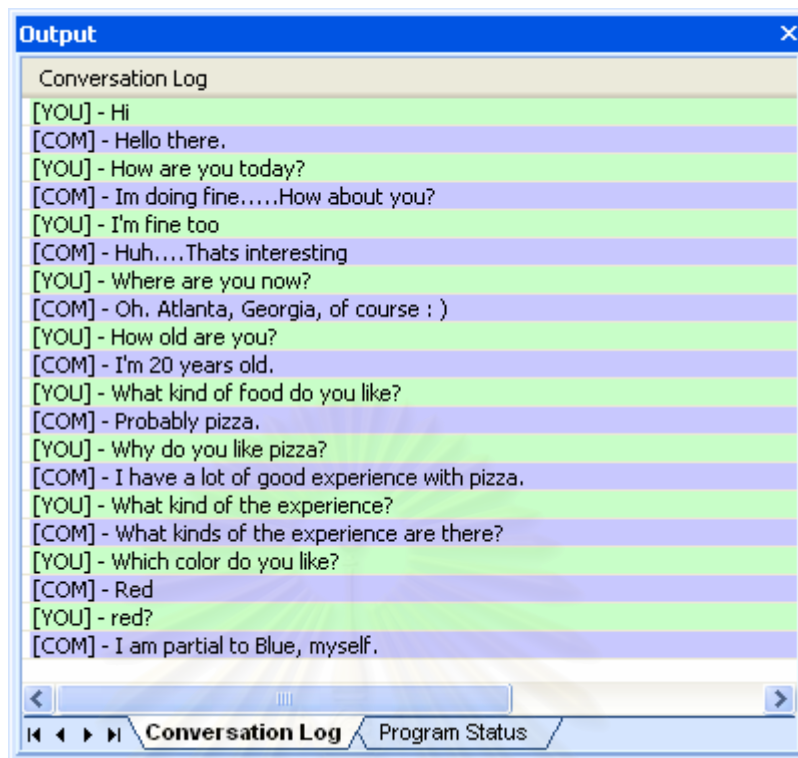
รูปที่ 4.5 แสดงภาพที่ได้จากมุมมองของผู้ใช้ เมื่อผู้ใช้เคลื่อนศีรษะไปทางซ้าย มองตรง และเคลื่อนศีรษะไปทางขวา

4.6 ผลการทดสอบการพูดคุยสนทนา

เมื่อได้ทำการทดสอบด้านการพูดคุยสนทนา ปัญหาอย่างหนึ่งที่พบคือระบบรู้จำเสียงพูดของผู้ใช้ไม่ค่อยแม่นยำ มักจะให้เอาต์พุตที่ผิดพลาด อาจเป็นเพราะยังไม่ได้เทรนให้ระบบรู้จำเสียงได้มากเพียงพอ และเวลาที่ทดสอบเป็นเวลากลางวัน มีเสียงรบกวนจากสภาพแวดล้อมค่อนข้างมาก ประกอบกับไมโครโฟนที่ใช้ยังไม่ค่อยมีความไวต่อเสียงพูด และยังคงตอบสนองต่อความถี่ต่างๆ ไม่ดีพอ ทำให้การใช้เสียงพูดทำได้ลำบาก แต่อย่างไรก็ตามในอนาคตหากมีอุปกรณ์ต่างๆ ที่ดี เช่น ไมโครโฟนที่ใช้สำหรับงานนี้โดยเฉพาะ หรือโปรแกรมรู้จำเสียงพูดที่มีประสิทธิภาพมากกว่านี้ ก็สามารถที่จะนำมาใช้ร่วมกับงานวิจัยนี้ได้ทันทีเพื่อผลลัพธ์ที่ดียิ่งขึ้น

สำหรับส่วนอื่นๆ เช่น ส่วนของการสังเคราะห์เสียง สามารถที่จะสังเคราะห์เสียงออกมาได้อย่างถูกต้อง และมีจังหวะการขยับปากตามรูปแบบที่ได้กำหนดไว้ได้ตรงกับเสียงพูด แต่เนื่องจากไม่ได้กำหนดรูปแบบปากไว้ค่อนข้างน้อย เพราะโมเดลไม่มีฟันและลิ้น จึงทำให้การพูดในบางคำอาจจะดูไม่เหมือนคนจริงนัก

การทดสอบการพูดคุยสนทนา จะทดสอบด้วยการพูดคุยประโยคง่ายๆ ดังรูปที่



รูปที่ 4.6 แสดงการพูดคุยสนทนากับระบบ

จากรูปที่ 4.6 ผลการพูดคุยที่แสดงจะเป็นผลที่ได้จากการติดต่อปัญหาประดิษฐ์ที่ชื่อ A.L.I.C.E Bot ซึ่งสามารถพูดคุยประโยคพื้นฐานได้เป็นอย่างดี แต่หากเป็นประโยคซับซ้อนหรือประโยคแปลกๆ ที่ไม่ได้กำหนดไว้ในฐานข้อมูล โปรแกรมก็อาจจะให้คำตอบออกมาผิดพลาดหรือไม่เกี่ยวกับเรื่องก็ตามไป ความสามารถในการพูดคุยโต้ตอบนั้น จะขึ้นอยู่กับข้อมูลในไฟล์ AIML เป็นหลัก ซึ่งผู้ใช้สามารถที่จะเพิ่มข้อมูลการพูดคุย และรูปแบบประโยคสนทนาใหม่ๆ เข้าไปในฐานข้อมูล AIML ได้ตามต้องการ

4.7 สรุปผลการทดลอง

จากการทดลองที่ผ่านมาทั้งหมดโดยรวมแล้วพบว่าโปรแกรมให้ผลลัพธ์ออกมาได้อย่างถูกต้อง และน่าพอใจ ระบบสามารถเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาผู้ใช้ได้ในแบบเวลาจริง สามารถพูดคุยโต้ตอบกับโมเดลโดยใช้เสียงพูดได้ และโมเดลสามารถแสดงอารมณ์แบบต่างๆ ได้ตามที่กำหนดไว้ สำหรับความแม่นยำและปัจจัยต่างๆ ที่มีผลกระทบต่อความแม่นยำและความเสมือนจริงของระบบ จะแยกอธิบายไปที่ละส่วน ดังต่อไปนี้

การคำนวณหาตำแหน่งของรูปแบบตรวจจับมีค่าความคลาดเคลื่อนเฉลี่ยไม่เกิน 10 เซนติเมตร โดยจะมีค่าความคลาดเคลื่อนน้อยหากรูปแบบตรวจจับอยู่กลางภาพ และจะมีความคลาดเคลื่อนมากขึ้นหากรูปแบบตรวจจับอยู่ตามขอบภาพโดยเฉพาะขอบภาพทางด้านซ้าย เนื่องมาจากบริเวณนั้นเป็นบริเวณที่ยังมีการบิดเบี้ยวในแนวรัศมีหลงเหลืออยู่ ไม่สามารถกำจัดออกไปได้ทั้งหมดร้อยเปอร์เซ็นต์ การตรวจจับรูปแบบตรวจจับสามารถทำได้อย่างรวดเร็วโดยได้เฟรมเรตสูงถึง 24.9 เฟรมต่อวินาทีบนเครื่องที่ใช้ซีพียูความเร็ว 2.40GHz ความแม่นยำที่ได้จากระดับตอนนี้จะส่งผลถึงความแม่นยำและประสิทธิภาพของการทำงานส่วนอื่นๆ ที่เหลือ เพราะฉะนั้นเมื่อนำระบบไปใช้งาน ควรคำนึงถึงปัจจัยต่างๆ ที่จะมีผลต่อความแม่นยำและประสิทธิภาพของการตรวจจับตำแหน่งซึ่งได้แก่

1. มุมของรูปแบบตรวจจับที่กระทำกับกล้อง ระบบสามารถตรวจจับตำแหน่งได้ แม้ว่ารูปแบบตรวจจับจะเอียง และไม่ตั้งฉากกับกล้อง แต่อย่างไรก็ตาม หากผู้ใช้เอียงศีรษะมากเกินไป (มากกว่า 60 องศาเทียบกับกล้อง) หรือเอียงเข้าหาแสงไฟโดยตรง จะทำให้เกิดแสงสะท้อนจากรูปแบบตรวจจับมาก และส่งผลให้การตรวจจับผิดพลาด
2. ความเร็วในการเคลื่อนไหวศีรษะของผู้ใช้ หากผู้ใช้เคลื่อนไหวศีรษะเร็ว จะส่งผลให้ภาพในส่วนที่เคลื่อนไหวเกิด “Motion Blur” ทำให้ตรวจจับตำแหน่งจุดมุมผิดพลาด ซึ่งอัตราความเร็วที่ทำให้เกิด “Motion Blur” จะไม่เท่ากันในกลุ่มแต่ละแบบ
3. มุมรับภาพของกล้อง หากผู้ใช้เคลื่อนไหวศีรษะออกนอกมุมรับภาพ ระบบก็จะไม่สามารถตรวจจับตำแหน่งได้ และมุมรับภาพก็จะไม่เท่ากันในกลุ่มแต่ละแบบ
4. การบิดเบี้ยวในแนวรัศมีของภาพ หากไม่สามารถกำจัดการบิดเบี้ยวในแนวรัศมีออกไปได้ทั้งหมด จะส่งผลให้การคำนวณตำแหน่งมีค่าความคลาดเคลื่อนมากขึ้นในบริเวณขอบของภาพ
5. ระยะของรูปแบบตรวจจับเทียบกับกล้อง หากอยู่ห่างเกินกว่า 80 เซนติเมตร จะไม่สามารถตรวจจับได้ เนื่องจากรูปแบบตรวจจับมีขนาดเล็กเกินไป และหากเข้ามาใกล้กล้องมากกว่า 30 เซนติเมตร จะอยู่นอกระยะโฟกัสของกล้อง
6. การเทียบมาตรฐานกล้อง เช่น การหาค่าพารามิเตอร์ทั้งภายนอกและภายในของกล้อง การปรับโฟกัสของกล้อง การหาสัมประสิทธิ์การบิดเบี้ยวในแนวรัศมี

การวัดระยะของกล้องเทียบกับกึ่งกลางจอภาพ การวัดความกว้างและความยาวของจอภาพ หากขั้นตอนการเทียบมาตรฐานเหล่านี้ให้ผลออกมาไม่ดีแล้ว ขั้นตอนการตรวจจับตำแหน่งสีระของผู้ใช้ก็จะออกมาไม่ดีและมีความคลาดเคลื่อนมากตามไปด้วย แต่ถ้าหากขั้นตอนเหล่านี้กระทำด้วยความแม่นยำ ผลการตรวจหาตำแหน่งสีระก็จะมีความแม่นยำเช่นกัน

สำหรับการมองไปยังจุดต่างๆ รอบตัวผู้ใช้มีความคลาดเคลื่อนไม่เกิน 2 เซนติเมตร ซึ่งอยู่ในเกณฑ์ในยอมรับได้ และเมื่อวัดด้วยความรู้สึกแล้ว ก็ให้ผลที่ตรงกับการสังเกตของผู้ใช้ เมื่อนำไปประยุกต์ให้จ้องมองไปยังตำแหน่งของผู้ใช้ ความแม่นยำจะขึ้นอยู่กับผลที่ได้จากขั้นตอนการตรวจจับตำแหน่งผู้ใช้เป็นหลัก

สำหรับผลที่ได้จากพีชแท่งก็วาร์ก็สามารถแสดงมุมมองของโมเดลได้สัมพันธ์กับตำแหน่งสีระของผู้ใช้ อาจจะมีกระตุกบ้างในบางครั้ง ก็เนื่องมาจากการตรวจหาตำแหน่งมีความผิดพลาดและไม่ต่อเนื่อง โดยเฉพาะเมื่อผู้ใช้อยู่ในระยะที่ห่างจากกล้องมากๆ สำหรับความเสมือนจริงที่ผู้ใช้รู้สึกได้จะมีมากน้อยเพียงใดนั้น มีปัจจัยที่เกี่ยวข้องดังนี้

1. ปัจจัยที่ส่งผลมากที่สุดคือความนิ่มนวลของการเปลี่ยนมุมมอง (ขึ้นอยู่กับอัตราเฟรมต่อวินาที และ ความต่อเนื่องของการตรวจจับจุดมุมมอง)
2. ปัจจัยที่ส่งผลกระทบต่อความแม่นยำของการตรวจจับตำแหน่งสีระผู้ใช้งาน
3. อีกปัจจัยหนึ่งคือ ความเสมือนจริงของตัวโมเดลสีระมนุษย์ หากสร้างโมเดลด้วยความประณีต และใช้พื้นผิวที่เหมือนจริง ให้สภาพแสงที่ดี ก็สามารถทำให้มองดูเหมือนสีระมนุษย์จริงๆ มากยิ่งขึ้น

ในส่วนของการพูดคุยกับโมเดล โมเดลสามารถที่จะพูดคุยและสนทนากับผู้ใช้ และมีการขยับปากที่สอดคล้องกับคำพูดตามรูปแบบที่ได้กำหนดไว้ สำหรับปัจจัยที่ส่งผลกระทบต่อความเสมือนจริงในการพูดคุย มีดังนี้

1. คุณภาพของไมโครโฟน หากใช้ไมโครโฟนที่มีคุณภาพดี มีความไวต่อการตอบสนองเสียงพูด การรู้จำเสียงพูดก็จะแม่นยำมากขึ้น
2. การสอนระบบให้รู้จำเสียงพูดของผู้ใช้แต่ละคน หากสอนระบบให้รู้จำเสียงมากขึ้น ผลจากกลไกการรู้จำเสียงพูดก็จะแม่นยำมากขึ้น

3. รูปแบบปาก ในการทดลองได้ทำการกำหนดรูปแบบปากยังไม่มากเพียงพอ เนื่องจากไม่มีฟันและลิ้น ทำให้ไม่สามารถที่จะกำหนดรูปแบบปากให้สอดคล้องกับเสียงบางแบบได้ ถ้าหากมีฟันและลิ้นด้วย การกำหนดรูปแบบปากจะทำได้หลากหลายและให้ความเหมือนจริงยิ่งกว่านี้
- 4.ฐานข้อมูลความรู้ของ A.L.I.C.E Bot หากเพิ่มรูปแบบประโยคสนทนาให้มาก และหลากหลาย จะส่งผลให้มีความสามารถในการพูดคุยได้เหมือนคนจริงมากยิ่งขึ้น

สำหรับในส่วนของการแสดงอารมณ์ ขณะนี้ระบบสามารถแสดงอารมณ์ได้ทั้งหมด 7 แบบตามที่ได้กำหนดไว้ โดยในระหว่างที่ไม่มีการพูดคุย ระบบจะสุ่มเลือกแสดงอารมณ์ระหว่าง ปกติ และ ยิ้ม (ดีใจ) สลับไปมา หากผู้ใช้เริ่มพูดคุยกับระบบ การแสดงอารมณ์จะแปรเปลี่ยนไปตามคำที่ใช้สนทนากัน เช่น หากในบทสนทนา มีคำว่า “love” โมเดลก็จะแสดงอารมณ์ดีใจออกมา และหากมีคำว่า “fail” โมเดลก็จะแสดงอารมณ์เสียใจ เป็นต้น สำหรับปัจจัยต่างๆ ที่มีผลต่อความเหมือนจริงในการแสดงอารมณ์ มีดังนี้

1. ความหลากหลายของอารมณ์ ขณะนี้ระบบได้กำหนดอารมณ์ไว้เจ็ดแบบ ถ้าหากแสดงอารมณ์ได้หลากหลายกว่านี้ ก็จะสมจริงมากยิ่งขึ้น
2. จำนวนกล่อมเนื่อบนใบหน้า ถ้าหาก มีจำนวนกล่อมเนื่อมากยิ่งขึ้น ก็จะทำให้มีความสามารถในการแสดงออกทางใบหน้ามากขึ้น
3. การกำหนดเวลาในการแสดงอารมณ์ ขณะนี้ระบบแสดงอารมณ์โดยอิงจากการสุ่ม และจากคำในบทสนทนาเท่านั้น ถ้าหากเพิ่มความสัมพันธ์ระหว่างรูปแบบประโยคสนทนาและอารมณ์ที่เกี่ยวข้องกับประโยคนั้นๆ ให้มากขึ้น ก็จะสามารถที่จะแสดงอารมณ์ได้คล้ายมนุษย์มากขึ้น

จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปการวิจัย ข้อเสนอแนะ และงานวิจัยในอนาคต

5.1 สรุปการวิจัย

งานวิจัยนี้เริ่มต้นจากแนวคิดที่ว่าอยากให้การติดต่อกับคอมพิวเตอร์ง่ายเหมือนการติดต่อกับมนุษย์ด้วยตนเอง ผู้ใช้ไม่จำเป็นต้องเสียเวลาในการเรียนรู้การใช้งาน ระบบนี้จะทำให้เกิดประโยชน์มากมายมหาศาลและประยุกต์ใช้งานได้หลากหลาย เช่น ระบบให้บริการอัตโนมัติตามสถานที่ต่างๆ ระบบให้คำปรึกษา ระบบประชาสัมพันธ์ หรือจะนำไปประยุกต์ใช้กับการโฆษณาที่มีปฏิสัมพันธ์กันคนได้เช่นกัน ซึ่งจะช่วยให้เพิ่มความน่าสนใจ ผู้ใช้ไม่จำเป็นต้องจำเมนูคำสั่งหรือเคยใช้คอมพิวเตอร์มาก่อนก็สามารถติดต่อสื่อสารใช้งานได้อย่างมีประสิทธิภาพ

ระบบที่ได้ออกแบบไว้ในงานวิจัยนี้ก็เป็นลักษณะโมเดลสามมิติของศีรษะมนุษย์เสมือนที่อยู่ภายในจอภาพ มีความสามารถในการพูดคุยโต้ตอบโดยใช้ปัญญาประดิษฐ์ที่ชื่อ A.L.I.C.E Bot ซึ่งเป็นปัญญาประดิษฐ์ที่มีมาตรฐานสูงและใช้กันอย่างแพร่หลาย ในการสนทนา ผู้ใช้จะพูดผ่านไมโครโฟน ระบบรู้จำเสียงพูดจะแปลงจากเสียงเป็นข้อความตัวอักษร แล้วส่งให้กับ A.L.I.C.E Bot ซึ่งจะส่งคำโต้ตอบออกมาเป็นข้อความเช่นเดียวกัน จากนั้นระบบจะแปลงข้อความโต้ตอบเป็นเสียงพูดให้ผู้ใช้ได้ยินผ่านทางลำโพงโดยใช้กลไกการสังเคราะห์เสียง และจะมีการขยับปากให้สัมพันธ์กับคำพูดที่เปล่งออกไป ทั้งยังสามารถแสดงอารมณ์แบบต่างๆ ได้ โดยใช้หลักการของระบบกล้ามเนื้อบนใบหน้า นอกจากนี้ระบบสามารถรับรู้ตำแหน่งของผู้ใช้งานผ่านกล้องที่ติดตั้งบนจอภาพ เพื่อตรวจหาตำแหน่งของรูปแบบตรวจจับ (Pattern) บนหมวกที่ผู้ใช้สวมอยู่ เมื่อได้ตำแหน่งบนภาพ (x,y) จากกล้องแล้วจึงนำไปคำนวณร่วมกับพารามิเตอร์แบบภายใน เพื่อหาค่าพารามิเตอร์แบบภายนอกคือ เวกเตอร์การเคลื่อนที่ ที่มีข้อมูลพิกัด (x,y,z) ของตำแหน่งรูปแบบตรวจจับเทียบกับกล้อง จากนั้นจึงนำข้อมูลที่ได้ออกไปคำนวณตำแหน่งศีรษะผู้ใช้เทียบกับกึ่งกลางจอภาพ ระบบก็จะสามารถรับรู้ได้ว่าขณะนี้ศีรษะผู้ใช้อยู่ตำแหน่งใด ก็จะนำไปเป็นข้อมูลสำหรับการทำให้ตาของโมเดลจำลองไปยังตำแหน่งของผู้ใช้งาน ในขณะเดียวกัน ก็จะปรับเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้งานด้วยโดยใช้เทคโนโลยีพีชคณิตเชิงเส้น

สำหรับผลจากโปรแกรมที่คำนวณได้มีความแม่นยำอยู่ในระดับที่ดี นั่นคือระบบสามารถที่จะคำนวณตำแหน่งรูปแบบตรวจจับได้ใกล้เคียงกับตำแหน่งจริงโดยมีค่าความคลาดเคลื่อนไม่เกิน 10 เซนติเมตร ถึงแม้ว่ารูปแบบตรวจจับเพียง ทำมุมไม่ตั้งฉากกับกล้อง หรือ

สภาพแสงเปลี่ยน ก็ยังสามารถตรวจจับได้โดยมีความเร็วในการประมวลผล 24.9 เฟรมต่อวินาที ซึ่งเร็วเพียงพอสำหรับการนำไปใช้ปรับมุมมองภาพสามมิติให้มีความนุ่มนวล และเฟรมที่ตรวจจับไม่ได้มีไม่เกิน 5% ของจำนวนเฟรมทั้งหมด ความแม่นยำในขั้นตอนนี้สำคัญที่สุดเพราะจะส่งผลต่อความแม่นยำและความถูกต้องของการทำงานในส่วนอื่นๆ ที่เหลือ สำหรับการกำหนดให้โมเดลจำลองมองไปยังจุดต่างๆ รอบตัวผู้ใช้ที่ระยะห่างจากจอภาพ 60 เซนติเมตร ระบบสามารถจำลองมองไปได้อย่างถูกต้องโดยมีค่าความคลาดเคลื่อนไม่เกิน 2 เซนติเมตร จากการวัดและสังเกตโดยคนในแนวแกน x และ y สำหรับการแสดงอารมณ์และการขยับปากพูดก็ทำได้อย่างถูกต้องตามรูปแบบที่ได้กำหนดไว้ แต่อย่างไรก็ตามยังมีปัญหาต่างๆ เกิดขึ้นเช่น รูปแบบตรวจจับไม่อยู่ในมุมมองรับภาพของกล้อง สภาพแสงเปลี่ยนทำให้ไม่สามารถตรวจจับได้ การเทียบมาตรฐานกล้องยังไม่แม่นยำเพียงพอจนทำให้เกิดข้อผิดพลาดในการคำนวณตำแหน่งบริเวณขอบของภาพอันเกิดจากการบิดเบี้ยวในแนวรัศมีที่ยังเหลืออยู่ การพูดคุยโดยใช้เสียงมีข้อผิดพลาดเนื่องจากคุณภาพของไมโครโฟนและการสอนระบบให้รู้จำเสียงพูดที่ยังไม่มากเพียงพอ รูปแบบการสนทนาหรือคำที่ไม่มีในฐานข้อมูล ระบบก็จะไม่เข้าใจและไม่สามารถตอบออกมาได้อย่างถูกต้อง แต่ในอนาคตเทคโนโลยีต่างๆ ที่เกี่ยวข้องย่อมที่จะดีขึ้นและก้าวหน้าขึ้น เช่น เทคนิคการรู้จำเสียงพูดที่แม่นยำมากขึ้น เทคโนโลยีกล้องที่ละเอียดมากขึ้นและมีมุมมองที่ดีกว่าเดิม ไมโครโฟนที่สามารถตัดเสียงรบกวนได้ ไรต่อเสียงพูดมากขึ้นและมีราคาถูกลง รวมถึงระบบปัญญาประดิษฐ์ที่ก้าวหน้ามากขึ้น ถ้าหากเทคโนโลยีพัฒนาถึงจุดนั้น ก็สามารถนำมาใช้ร่วมกับงานวิจัยนี้เพื่อสร้างเป็นระบบที่สมบูรณ์แบบและสามารถใช้งานในเชิงพาณิชย์ได้จริง

5.2 ข้อเสนอแนะ

จากผลการทดลองที่มีข้อผิดพลาด ทางผู้พัฒนาอยากจะให้ข้อเสนอแนะเพื่อให้ได้ผลลัพธ์ที่ดีและแม่นยำยิ่งกว่าเดิม ด้วยเทคโนโลยีที่มีอยู่ ณ ตอนนี้ ได้แก่

1. ปัญหาเรื่องมุมมองรับภาพของกล้องที่ยังแคบ อาจจะได้แก้ไขได้โดยการเปลี่ยนไปใช้กล้องที่มีมุมมองกว้างขึ้น แต่กล้องที่มีมุมมองกว้างจะมีความบิดเบี้ยวในแนวรัศมีที่มากขึ้นตามไปด้วย เพราะฉะนั้นการหาค่าสัมประสิทธิ์ความบิดเบี้ยวในแนวรัศมีควรกระทำด้วยความประณีตและแม่นยำกว่าเดิม อีกวิธีหนึ่งคือใช้กล้องหลายตัว เช่น ใช้กล้องสี่ตัว ติดขอบบน ขอบซ้าย ขอบขวา และขอบล่าง ของจอภาพ ให้ทำงานประสานกัน จะช่วยในการเพิ่มมุมมองรับภาพของระบบโดยรวมได้ รวมถึงการเลือกใช้กล้องที่มีความละเอียดมากขึ้น จะช่วยให้การตรวจจับมีความแม่นยำมากขึ้น

2. การพูดคุยโดยใช้เสียงยังมีความผิดพลาดสูง การแก้ไขทำได้โดยการเทรนระบบให้รู้จำเสียงพูดมากกว่าเดิม และการใช้ไมโครโฟนที่มีคุณภาพสูง มีความไวในการตอบสนองต่อเสียงที่ความถี่ต่างๆ แต่ไมโครโฟนคุณภาพสูงจะมีราคาแพงตามไปด้วย

3. สามารถเพิ่มความสามารถในการพูดคุยของระบบโดยการเพิ่มประโยคและรูปแบบการสนทนาลงในฐานข้อมูล AIML สำหรับฐานข้อมูล AIML ที่ดีที่สุดที่สามารถพูดคุยได้อย่างเป็นธรรมชาติ รวมถึงฐานข้อมูล AIML เฉพาะทาง เช่น ระบบผู้เชี่ยวชาญหรือระบบให้คำปรึกษาในด้านต่างๆ นั้นมีขายในเชิงพาณิชย์ สามารถหาซื้อเพื่อนำมาใช้กับงานวิจัยนี้ได้ทันที

5.3 งานวิจัยในอนาคต

งานวิจัยชิ้นนี้สามารถนำไปพัฒนาต่อยอดในอนาคตได้อีกมาก เช่น นำไปพัฒนาให้มีความสามารถในการประมวลผลภาพที่ดียิ่งขึ้น สามารถตรวจจับตำแหน่งผู้ใช้ได้โดยตรวจจับจุดต่างๆ บนใบหน้าของผู้ใช้โดยตรง โดยที่ผู้ใช้ไม่จำเป็นต้องสวมหมวกที่ติดรูปแบบตรวจจับ หรือพัฒนาให้สามารถรับรู้อารมณ์ของผู้ใช้งานได้โดยมีความเร็วในการประมวลผลอยู่ในระดับที่ดี ในกรณีที่มีหลายคนอยู่ในเฟรม อาจจะพัฒนาให้สามารถพูดคุยกับหลายๆ คนได้พร้อมกัน เสมือนคุยกันเป็นกลุ่ม

ในด้านของการพูดคุย งานวิจัยในอนาคตควรจะเพิ่มความสามารถในการขยับปากของโมเดลให้มีส่วนของฟันและลิ้นด้วย เพื่อที่จะได้ขยับปากได้อย่างเป็นธรรมชาติมากที่สุด และอาจจะพัฒนาเพิ่มในส่วนของความจำ ให้ระบบสามารถจดจำได้ว่าเคยคุยอะไรกับผู้ใช้คนนี้ และผู้ใช้คนไหนชอบอะไร ไม่ชอบอะไร จะทำให้ผู้ใช้เกิดความประทับใจต่อระบบมากยิ่งขึ้น

ในด้านของคอมพิวเตอร์กราฟฟิก เมื่อคอมพิวเตอร์มีความเร็วในการประมวลผลมากขึ้น อาจจะเพิ่มจำนวนโพลีกอน เพิ่มรายละเอียดเทกซ์เจอร์ หรือใส่เอฟเฟกต์ต่างๆ ให้กับโมเดล เพื่อให้มีความเสมือนจริงมากที่สุด

รายการอ้างอิง

1. Ivan Sutherland. A head-mounted three dimensional display. Proceedings of the Fall Joint Computer Conference, AFIPS Conference Proceedings, 33 (1968) : 757-764
2. SimVis Head-mounted display (Online). Available from:
<http://www2.dcs.hull.ac.uk/simmod/Technology/HMD.htm>
3. Colin Ware, Kevin Arthur, Kellogg S. Booth. Fish Tank Virtual Reality, Proceedings of the SIGCHI conference on Human factors in computing systems, 1993
4. Welcome to Vislab (Online). Available From:
<http://www.ccom.unh.edu/vislab/FishTankVR.html>
5. DocShow-VR (Online). Available From: <http://www.dfn-expo.de/Technologie/DocShow-VR/stereo.html>
6. Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. Proceedings of the 20th annual conference on Computer graphics and interactive techniques, 1993
7. Virtual Skiing Environment (Online). Available From: <http://www-personal.umich.edu/~galtland/skiVR/video/tech.htm>
8. Cagatay Demiralp, David H. Laidlaw, Cullen Jackson, Daniel Keefe, and Song Zhang. Subjective usefulness of cave and fish tank vr display systems for a scientific visualization application. In Poster Compendium IEEE Conference on Visualization, Seattle, WA, 2003
9. Jun Rekimoto, A Vision-Based Head Tracker for Fish Tank Virtual Reality - VR without Head Gear. Proceedings of VRAIS'95.(1995) : 94-100
10. Paula Salgado Lucena, Marcelo Gattass, and Luiz Velho. Expressive Talking Heads: A Study on Speech and Facial Expression in Virtual Characters. Revista SCIENTIA, 2002.

11. Cynthia Breazeal, Brian Scassellati. Robot in Society: Friend of Appliance? In Agents99 Workshop on Emotion-based Agent Architectures, Seattle, WA. (1999) : 18-26
12. Cynthia Breazeal, Brian Scassellati, How to build robots that make friends and influence people., IROS99, Kyonjiu, Korea, 1999.
13. R. C. Gonzalez, and R. E. Woods. Digital Image Processing. U.S.A.: Prentice-Hall, 2002.
14. Dave Shreiner, OpenGL(R) Reference Manual: The Official Reference Document to OpenGL, Version 1.2 (3rd Edition)
15. Frederic Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David Salesin. Synthesizing Realistic Facial Expressions from Photographs. Proceedings of SIGGRAPH 98, in Computer Graphics Proceedings, Annual Conference Series, 1998.
16. Intel Corporation. Open Source Computer Vision Library Reference Manual. 2001. Available from: <http://sourceforge.net/projects/opencvlibrary/>
17. Microsoft Corporation, Microsoft Speech SDK 5.1 Documentation, 2003
18. Richard S. Wallace. A.L.I.C.E. and AIML Documentation . A.L.I.C.E. Artificial Intelligence Foundation, 2000
19. Michael Treadgolf, Kevin Novins, Geoff Wyvill, and Brian Niven. What do you think you're doing? Measuring perception in Fish Tank Virtual Reality. Computer Graphics International Conference, 2001
20. Peter M. Plantec, Ray Kurzweil. Virtual Humans: A Build-It-Yourself Kit, Complete With Software and Step-By-Step Instructions. AMACOM Book and CD-ROM edition , November 2003
21. Richard S. Wallace. Be Your Own Botmaster. Alice Foundation , 2003
22. Richard S. Wallace. AIML Overview. A.L.I.C.E. Artificial Intelligence Foundation.
23. Richard S. Wallace. The Anatomy of A.L.I.C.E., A.L.I.C.E. Artificial Intelligence Foundation.
24. Bill Freming, Darris Dobbs. Animating Facial Features & Expressions, Charles River Media. 1999

25. C. Pelachaud, N. Badler, and M. Steedman. Generating facial expressions for speech, Cognitive Science. 20(1)(1996) : 1-46.
26. Frederic I. Parke, Keith Waters. Computer Facial Animation, A K Peters. 1996
27. S. Pasquariello, C. Pelachaud. Greta: A Simple Facial Animation Engine, 6th Online World Conference on Soft Computing in Industrial Applications, Session on Soft Computing for Intelligent 3D Agents, September 2001



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

โปรแกรมต้นแบบ

ภาคผนวก ก จะอธิบายถึงรายละเอียดของโปรแกรมต้นแบบที่ได้พัฒนาขึ้นสำหรับงานวิจัยนี้ โดยจะประกอบไปด้วยการเตรียมระบบที่เหมาะสมกับการใช้งานโปรแกรม การเรียกใช้งาน และตัวอย่างหน้าจอของโปรแกรมต้นแบบ

ก.1 ระบบที่เหมาะสมสำหรับการใช้งานโปรแกรม

เพื่อการใช้งานโปรแกรมได้อย่างมีประสิทธิภาพ ก่อนการใช้งานโปรแกรม ควรเตรียมระบบคอมพิวเตอร์ให้พร้อมดังรายการต่อไปนี้

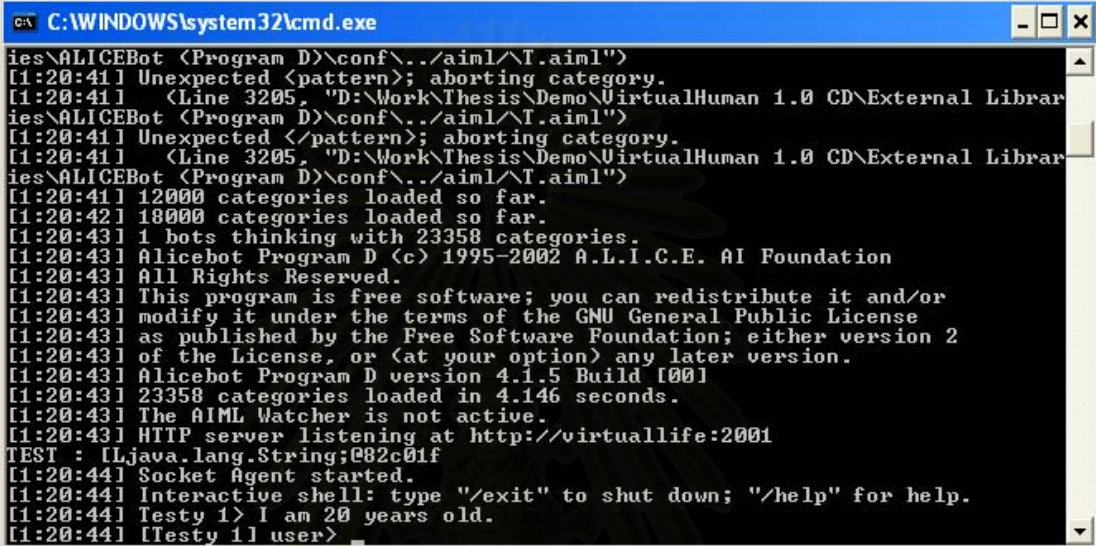
1. ใช้ซีพียูที่มีความเร็วในการประมวลผลอย่างต่ำ 1 GHz
2. มีหน่วยความจำไม่น้อยกว่า 256 เมกะไบต์
3. มีพื้นที่ในฮาร์ดดิสก์คงเหลือไม่น้อยกว่า 50 เมกะไบต์
4. เชื่อมต่อกับไมโครโฟนและลำโพง
5. ติดตั้งกล่องสำหรับจับภาพใบหน้า อาจเป็นกล่องเว็บแคมหรือกล่องวงจรถัดก็ได้ ในกรณีของกล่องวงจรถัดจะต้องเชื่อมต่อกับการ์ดจับภาพวิดีโอ
6. เตรียมหมวกที่ติดแพทเทิร์นลายตารางหมากรุกขนาด 2 x 2
7. ใช้ระบบปฏิบัติการ Microsoft Windows XP
8. ติดตั้ง DirectX 8.1 Runtime หรือเวอร์ชันสูงกว่าเรียบร้อยแล้ว
9. ติดตั้งโปรแกรมสำหรับรู้จำเสียงพูดเรียบร้อยแล้ว เช่น Dragon Naturally Speaking เป็นต้น

ก.2 การเรียกใช้งานโปรแกรม

การเรียกใช้งานโปรแกรมมีขั้นตอนดังนี้

1. ใส่แผ่นซีดีโปรแกรมเข้าไปในเครื่องอ่านแผ่นซีดีรอม
2. สำเนาข้อมูลทั้งหมดในแผ่นซีดีลงในฮาร์ดดิสก์ในชื่อโฟลเดอร์ “Virtual Human 1.0”

3. เข้าไปในโฟลเดอร์ “Virtual Human 1.0\External Libraries\ALICEBot (Program D)”
4. ดับเบิลคลิกที่ชื่อไฟล์ “run.bat” เมื่อได้หน้าจอตั้งรูปที่ ก.1 แสดงว่า ณ เวลานั้น A.L.I.C.E Bot ได้ทำงานเป็นเซิร์ฟเวอร์พร้อมรับการติดต่อมาที่พอร์ต หมายเลข 2001 แล้ว



```

c:\WINDOWS\system32\cmd.exe
ies\ALICEBot <Program D>\conf\..\aiml\T.aiml")
[1:20:41] Unexpected <pattern>; aborting category.
[1:20:41] <Line 3205, "D:\Work\Thesis\Demo\VirtualHuman 1.0 CD\External Librar
ies\ALICEBot <Program D>\conf\..\aiml\T.aiml")
[1:20:41] Unexpected </pattern>; aborting category.
[1:20:41] <Line 3205, "D:\Work\Thesis\Demo\VirtualHuman 1.0 CD\External Librar
ies\ALICEBot <Program D>\conf\..\aiml\T.aiml")
[1:20:41] 12000 categories loaded so far.
[1:20:42] 18000 categories loaded so far.
[1:20:43] 1 bots thinking with 23358 categories.
[1:20:43] Alicebot Program D (c) 1995-2002 A.L.I.C.E. AI Foundation
[1:20:43] All Rights Reserved.
[1:20:43] This program is free software; you can redistribute it and/or
[1:20:43] modify it under the terms of the GNU General Public License
[1:20:43] as published by the Free Software Foundation; either version 2
[1:20:43] of the License, or <at your option> any later version.
[1:20:43] Alicebot Program D version 4.1.5 Build [00]
[1:20:43] 23358 categories loaded in 4.146 seconds.
[1:20:43] The AIML Watcher is not active.
[1:20:43] HTTP server listening at http://virtuallife:2001
TEST : [Ljava.lang.String;@82c01f
[1:20:44] Socket Agent started.
[1:20:44] Interactive shell: type "/exit" to shut down; "/help" for help.
[1:20:44] Testy 1> I am 20 years old.
[1:20:44] [Testy 1] user>

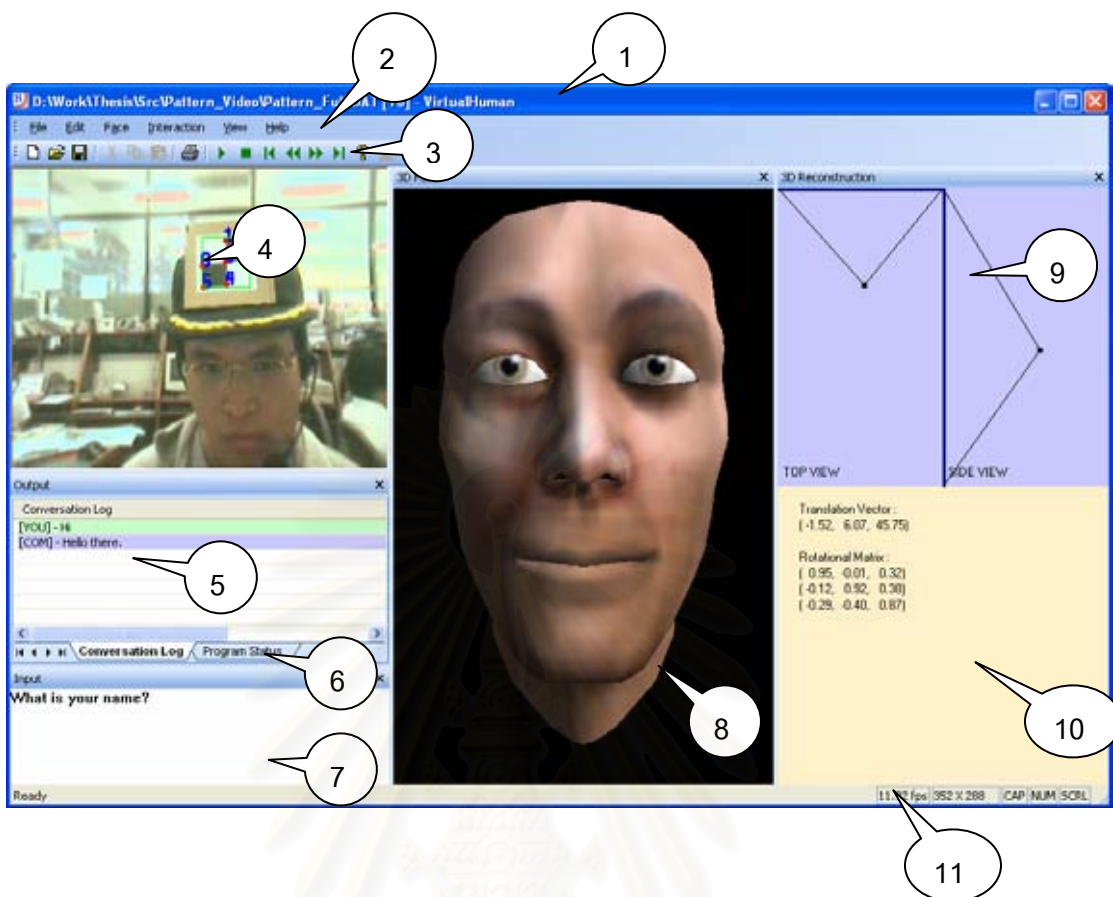
```

รูปที่ ก.1 แสดงหน้าจอเมื่อสั่งให้ A.L.I.C.E Bot ทำงาน

5. เข้าไปในโฟลเดอร์ “VirtualHuman 1.0\VirtualHuman 1.0 Software”
6. เปิดโปรแกรมผู้จำลองพฤติกรรมให้ทำงาน
7. ดับเบิลคลิกที่ไฟล์ชื่อ “VirtualHuman.exe” โปรแกรมก็จะถูกเปิดขึ้นมาและพร้อมที่จะทำงาน

ก.3 หน้าจอหลักของโปรแกรม

เมื่อเปิดโปรแกรม จะเข้าสู่หน้าจอหลักดังรูปที่ ก.2



รูปที่ ก.2 หน้าจอหลักของโปรแกรม

ส่วนประกอบต่างๆ ของหน้าจอหลัก มีดังนี้คือ

1. แถบชื่อโปรแกรม (Title Bar)
2. แถบเมนู (Menu Bar)
3. แถบเครื่องมือ (Tools Bar)
4. ภาพที่ได้จากกล้องและผลการตรวจจับจุดมุมของแพทเทิร์น (Camera View)
5. หน้าจอบันทึกการสนทนา (Conversation Log Pane)
6. หน้าจอรายงานสถานะการทำงานของโปรแกรม (Program Status Pane)
7. ช่องรับอินพุตข้อความ (เมื่อเปิดโปรแกรมรู้จำเสียงพูดแล้ว ให้คลิกตรงช่องนี้เพื่อรับข้อความที่ได้จากโปรแกรมรู้จำเสียงพูด (Input Pane)
8. หน้าจอแสดงกราฟฟิกระยะสามมิติของมนุษย์เสมือนจริง (3D Face Pane)
9. หน้าจอจำลองตำแหน่งผู้ใช้งานเทียบกับจอภาพจากผลการตรวจจับตำแหน่งผู้ใช้ที่คำนวณได้ (3D Reconstruction Pane)
10. หน้าจอรายงานเวกเตอร์การเคลื่อนที่ และเมตริกซ์ของการหมุนของศีรษะผู้ใช้

11. แถบสถานะของหน้าต่างหลัก (Status Bar)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

บทความที่นำเสนอในงานการประชุมวิชาการ

1. บทความเรื่อง “ศิระมะมนุษย์เสมือนแบบมีปฏิสัมพันธ์โดยใช้เทคโนโลยีพีชแท่งกีวีอาร์”
นำเสนอในงานการประชุมวิชาการด้านวิทยาการและวิศวกรรมคอมพิวเตอร์แห่งชาติ
ประจำปี 2548 จัดโดยมหาวิทยาลัยหอการค้าไทย กรุงเทพมหานคร



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ศึกษามนุษย์เสมือนแบบมีปฏิสัมพันธ์โดยใช้เทคโนโลยีฟิชแทงก์วีอาร์

Interactive Virtual Human Head with Fish Tank VR Technology

อรรถกร พูนศิลป์

อรรถวิทย์ สูดแสง

Attakorn Poonsilp

Attawith Sudsang

Department of Computer Engineering, Chulalongkorn University
Bangkok, 10330, Thailand

{g45aps,attawith}@cp.eng.chula.ac.th

บทคัดย่อ

บทความนี้นำเสนอการออกแบบส่วนติดต่อกับผู้ใช้ในลักษณะโมเดลสามมิติของศึกษามนุษย์เสมือนจริงที่สามารถพูดคุยโต้ตอบกับผู้ใช้ได้อย่างเป็นธรรมชาติ การขยับปากสอดคล้องกับคำพูด ดวงตาของโมเดลสามารถมองมายังผู้ใช้งาน รวมถึงการปรับเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในแบบเวลาจริงโดยใช้เทคโนโลยีฟิชแทงก์วีอาร์

คำสำคัญ: ความจริงเสมือน, ศึกษามนุษย์เสมือน, ส่วนติดต่อกับผู้ใช้, การมองเห็นด้วยคอมพิวเตอร์, ฟิชแทงก์วีอาร์

Abstract

This paper proposes the designing and architecture of user interface using 3D human head model. The proposed user interface can chat and interact to user in a natural way such as lip-sync according to speech. The model's eyes can gaze to user. And model's viewing perspective is adjusted by user's eyes position in real-time using Fish Tank VR technology.

Key-Words: Virtual Reality, Virtual Human Head, User Interface, Computer Vision, Fish Tank VR

1. บทนำ

ระบบคอมพิวเตอร์ตั้งแต่อดีตมาจนถึงปัจจุบันได้มีการพัฒนาการอย่างมากมาย ทั้งความเร็วในการประมวลผล

หน่วยความจำ รวมถึงระบบเพิ่มขยายต่างๆ เพื่อเพิ่มความสามารถให้มากยิ่งขึ้น แต่สำหรับการปฏิสัมพันธ์ (User Interface) กับผู้ใช้นั้น ระบบส่วนใหญ่ยังคงกำหนดให้ใช้แป้นพิมพ์ และเมาส์ในการป้อนข้อมูลเป็นหลัก ทำให้ในบางกรณีผู้ใช้ไม่สามารถใช้งานได้สะดวกและเป็นธรรมชาติเท่าที่ควร เช่น กรณีของผู้ที่ไม่คุ้นเคยกับคีย์บอร์ดและเมาส์ หรือคนพิการ นอกจากนั้นเมื่อผ่านการประมวลผลแล้วระบบคอมพิวเตอร์ก็จะแสดงผลพรีออกมาในรูปแบบตัวอักษร ภาพสองมิติ หรือภาพสามมิติที่ส่วนใหญ่ไม่มีปฏิสัมพันธ์ใดๆ กับผู้ใช้ จึงทำให้การนำเสนอขาดความน่าสนใจ ด้วยปัญหาต่างๆ เหล่านี้จึงได้มีการนำทฤษฎีหลายกลุ่มเริ่มพัฒนาระบบติดต่อกับผู้ใช้งานที่น่าสนใจและเป็นธรรมชาติมากขึ้น ดังเช่นในงานวิจัยของ Magy S. N. [13], W.Muller et al. [17], Jonathan et al. [11] และ Paula et al. [18] ที่ได้เสนอแนวคิดในการสร้างส่วนติดต่อกับผู้ใช้ โดยใช้ใบหน้ามนุษย์เสมือนที่สามารถพูดคุยกับผู้ใช้ได้ สามารถขยับปากให้สอดคล้องกับคำพูด และแสดงอารมณ์แบบต่างๆ ได้ Michel D. Bondy et al. [15] และ Byoungwon Choe et al. [2] ได้เพิ่มความสามารถให้กับการแสดงอารมณ์ด้วยการออกแบบระบบกล่อมเนื้อตามจุดต่างๆ บนใบหน้าทำให้การแสดงอารมณ์และการขยับปากมีความสมจริงมากยิ่งขึ้น แต่อย่างไรก็ตาม งานวิจัยที่กล่าวมาได้นั้นเน้นส่วนการแสดงผลของโมเดลใบหน้าเท่านั้น การพูด

ของโมเดลจะพูดจากสคริปต์ที่ได้กำหนดไว้ล่วงหน้า และโมเดลไม่สามารถรับรู้ความเป็นไปของผู้ใช้ หรือสภาวะแวดล้อมรอบๆ ได้ ทำให้มีข้อจำกัดด้านความสมจริงอย่างมาก ยังมีงานวิจัยอีกชิ้นหนึ่งที่น่าสนใจ งานวิจัยชิ้นนี้เป็น การสร้างหุ่นยนต์ (เฉพาะส่วนของศีรษะและใบหน้า) พัฒนาโดย Cynthia Breazeal [6] [7] หุ่นยนต์สามารถแสดงออกทางอารมณ์บนใบหน้า มีปฏิกริยากับผู้ใช้ได้ในแบบต่างๆ กัน รับรู้สภาพแวดล้อมรอบตัวได้ เรียกว่า Social Robot มีจุดมุ่งหมายหลักคือ ต้องการให้เป็นหุ่นยนต์ที่สามารถสื่อสารกับมนุษย์ และสามารถเป็นเพื่อนเล่นกับมนุษย์ได้ โดยหุ่นยนต์มีชื่อว่า “Kismet” จะมีลักษณะการแสดงออกทางอารมณ์และความสามารถเทียบได้กับเด็กทารกอายุหกเดือน แต่การสร้างเป็นหุ่นยนต์กลไกแบบนี้จะมีต้นทุนการสร้างที่สูง ไม่สามารถทำให้ดูเหมือนคนจริงได้เท่ากับระบบคอมพิวเตอร์กราฟิก และมีกลุ่มเป้าหมายที่จำกัด

จากปัญหาในงานวิจัยต่างๆ ที่กล่าวมาทั้งหมด ผู้วิจัยจึงได้เกิดแนวคิดที่จะออกแบบและวางรากฐานส่วนติดต่อกับผู้ใช้บนคอมพิวเตอร์พีซีธรรมดาให้มีความเสมือนจริง เป็นธรรมชาติ ผู้ใช้ไม่จำเป็นต้องใช้เวลาเรียนรู้การใช้งาน เพราะผู้ใช้สามารถสื่อสารกับคอมพิวเตอร์ได้โดยไม่ต่างอะไรกับการสื่อสารกับมนุษย์ด้วยกันเอง ในงานวิจัยนี้จะนำเสนอส่วนของการพัฒนาโมเดลสามมิติของศีรษะมนุษย์ให้สามารถมีปฏิกริยาต่างๆ ต่อผู้ใช้ได้ เช่น รับรู้สภาพแวดล้อมรอบๆ ผู้ใช้ มองไปยังผู้ใช้ ขยับปากสนทนากับผู้ใช้ได้อย่างเป็นธรรมชาติ โดยติดต่อกับ A.I ที่มีมาตรฐานสูงคือ A.L.I.C.E BOT [23] รวมถึงการเปลี่ยนมุมมองของโมเดลให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในแบบเวลาจริงโดยนำเทคโนโลยี Fish Tank VR [4][5][14] มาประยุกต์ใช้ ซึ่งจะทำให้ผู้ใช้เกิดความรู้สึกเสมือนว่ามีศีรษะมนุษย์อยู่ภายในจอมอนิเตอร์จริงๆ ด้วยส่วนติดต่อกับผู้ใช้แบบนี้ จะช่วยเพิ่มความน่าสนใจในการใช้คอมพิวเตอร์ได้อีกมาก นอกจากนี้การใช้โมเดลศีรษะมนุษย์จะมีผลต่อจิตใจทำให้ผู้ใช้รู้สึกว่าคอมพิวเตอร์เป็น

มิตรมากขึ้นและสามารถเป็นเพื่อนกับเราได้ (โดยเฉพาะเด็กๆ) ทำให้ชอบที่จะใช้งาน และไม่รู้สึกลบเนื้อ สำหรับส่วนที่เหลือของบทความมีส่วนต่างๆ ดังนี้ ส่วนที่ 2 การออกแบบ ส่วนที่ 3 รายละเอียดการทำงาน ส่วนที่ 4 ผลการทดลองและการใช้งาน และส่วนที่ 5 สรุปผล

2. การออกแบบ

ในงานวิจัยนี้ได้มีแนวคิดในการพัฒนาส่วนติดต่อกับผู้ใช้แสดงได้ดังรูปที่ 1



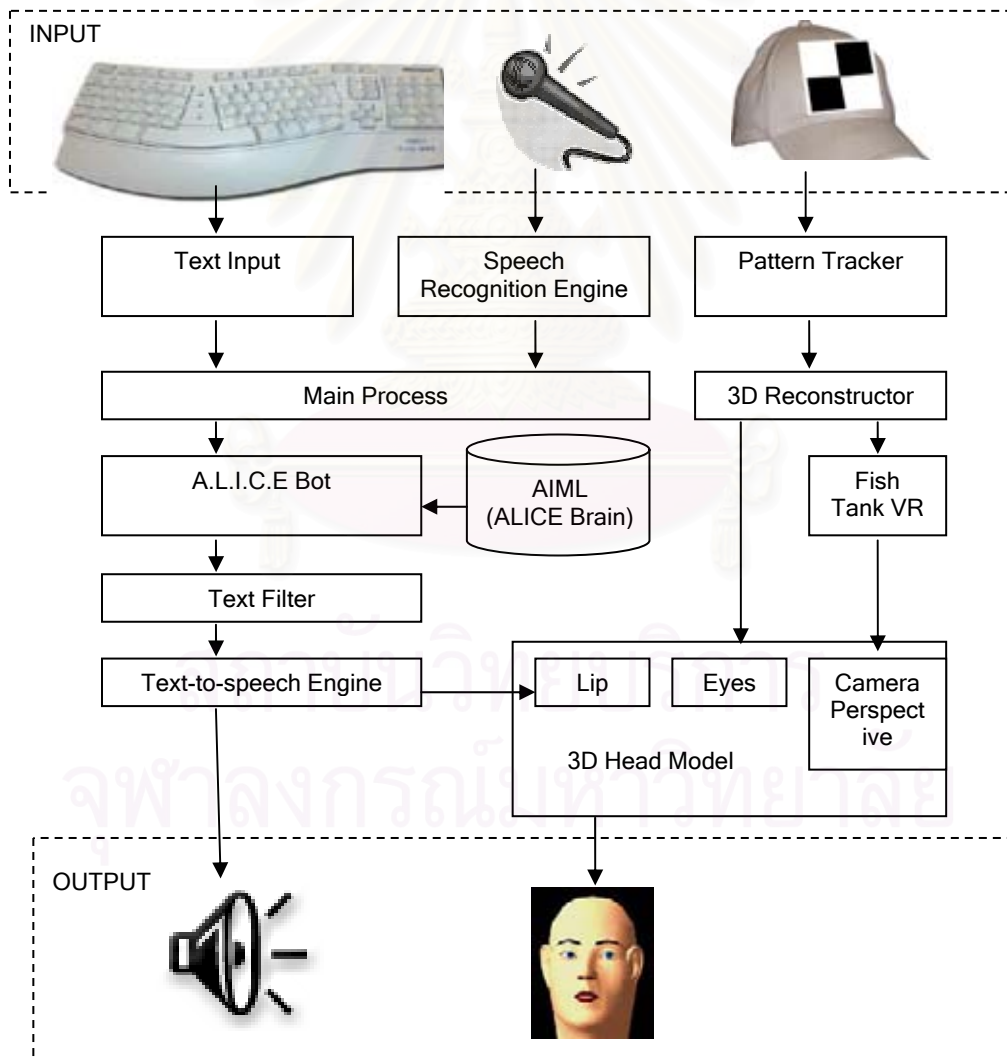
รูปที่ 1 แนวคิดในการนำโมเดลศีรษะมนุษย์มาใช้ในการติดต่อกับผู้ใช้

จากแนวคิดในรูปที่ 1 นำมากำหนดเป็นรายละเอียดและความสามารถของระบบได้ดังนี้

1. ระบบที่พัฒนาขึ้นสามารถแสดงโมเดลศีรษะมนุษย์เสมือนจริง โดยที่ระบบจะมีโมเดลศีรษะให้ผู้ใช้ได้เลือกหลากหลายรูปแบบตามความต้องการ
2. โมเดลศีรษะสามารถมองไปยังผู้ใช้หรือสบตาผู้ใช้ได้
3. สามารถสนทนาพูดคุยกับผู้ใช้ได้
4. สามารถขยับปากให้สอดคล้องกับคำพูดได้
5. สามารถแสดงอารมณ์บนใบหน้า ได้แก่ ปกติ, ยิ้ม, เศร้า, หวาดกลัว, โกรธ เป็นต้น
6. สามารถปรับมุมมองของโมเดลให้สอดคล้องกับตำแหน่งดวงตาของผู้ใช้ในแบบเวลาจริง โดยผู้ใช้ต้องสวมหมวกติด Pattern เอาไว้ และต้องเคลื่อนไหวให้อยู่ในมุมรับภาพของกล้อง

เมื่อจะคุยกับระบบ ผู้ใช้สามารถที่จะพิมพ์ หรือพูดโดยใช้เสียงก็ได้ กรณีที่ผู้ใช้พูดไมโครโฟน เสียงจะถูกแปลงเป็นข้อความเสียก่อนโดยใช้ Speech Recognition Engine [16] เมื่อได้เป็นข้อความตัวอักษร ข้อความนี้จะถูกนำไปเป็นอินพุตให้กับ A.L.I.C.E Bot จากนั้น A.L.I.C.E Bot ก็จะให้คำตอบรับออกมาเป็นข้อความเช่นเดียวกัน จึงต้องนำข้อความนี้เข้าสู่ Text-to-speech Engine เพื่อสร้างเป็นเสียงให้ผู้ใช้ได้ยิน ในขณะที่ Text-to-speech Engine ทำงาน จะส่งสัญญาณไปยังโมเดลของโมเดลใบหน้าสามมิติ เพื่อกำหนดรูปแบบการขยับปากให้สอดคล้องกับเสียงที่เปล่งออกมา

ส่วนสำคัญอีกส่วนหนึ่งคือส่วนของการรับรู้ตำแหน่งผู้ใช้ ระบบสามารถรับรู้ตำแหน่งของผู้ใช้ได้จากการตรวจจับรูป Pattern ที่ติดไว้บนหมวกที่ผู้ใช้สวมอยู่ผ่านกล้องเว็บแคม เมื่อได้ตำแหน่งของรูป Pattern ก็สามารถนำไปคำนวณหาตำแหน่งของผู้ใช้และมุมมองของผู้ใช้ที่กระทำกับจอภาพโดยผ่านโมดูล 3D Reconstructor จากนั้นจะนำข้อมูลที่คำนวณได้ไปปรับดวงตาของโมเดล และส่งต่อข้อมูลไปให้โมดูล Fish Tank VR เพื่อปรับมุมมองการแสดงผลโมเดลสามมิติให้สอดคล้องกับตำแหน่งผู้ใช้ สำหรับภาพแสดงการทำงานของส่วนประกอบต่างๆ แสดงดังรูปที่ 2



รูปที่ 2 แสดงส่วนประกอบต่างๆ ของระบบ

3. รายละเอียดการทำงาน

รายละเอียดการทำงานในแต่ละส่วน มีดังนี้

3.1 การตรวจหาตำแหน่งของผู้ใช้เทียบกับจอภาพ

การตรวจหาตำแหน่งในระบบพิกัดสามมิติของผู้ใช้จากภาพดิจิทัลสองมิติสามารถทำได้โดยใช้เทคนิค Camera Calibration [10] ซึ่งเกี่ยวข้องกับพารามิเตอร์ของกล้องแบบ Intrinsic และ Extrinsic ดังนี้

1. Intrinsic camera parameters คือค่าที่บ่งบอกลักษณะเฉพาะของกล้องนั้นๆ ซึ่งได้แก่

- ความยาวโฟกัส (Focal length) ซึ่งก็คือระยะระหว่างเลนส์ของกล้องและฉากรับภาพ
- ตำแหน่งของจุดกึ่งกลางภาพในระบบพิกัดจุด (Principal point)
- ขนาดของแต่ละจุดภาพ (Effective pixel size)
- สัมประสิทธิ์การบิดเบี้ยวในแนวรัศมีของเลนส์ (Radial distortion coefficient)

2. Extrinsic camera parameters คือค่าที่บ่งบอกถึงความสัมพันธ์เชิงระยะทาง (Spatial Relationship) ระหว่างกล้อง (Camera) กับโลกภายนอก (World) ได้แก่

- Rotational Matrix คือ เมตริกซ์การหมุน
- Translation Vector คือ เวกเตอร์การเคลื่อนที่

ถ้าหากใช้กล้องแบบรูเข็มต่างๆ ไปแล้ว จุดในพิกัดสามมิติ M และจุดที่ปรากฏบนภาพ m เป็นไปตามความสัมพันธ์ดังนี้

$$m = A[Rt]M$$

โดยที่ A คือเมตริกซ์ของ Intrinsic parameters:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \text{ เมื่อ}$$

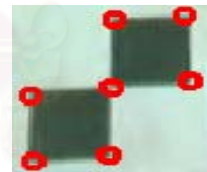
(c_x, c_y) คือพิกัดของจุดกึ่งกลางภาพ หรือ Principal point

(f_x, f_y) คือระยะโฟกัสบนแกน x และ y

(R, t) คือ Extrinsic parameters: เมตริกซ์การหมุน R และเวกเตอร์การเคลื่อนที่ t

เมื่อทราบถึงความสัมพันธ์ระหว่างจุดในโลกของผู้ใช้กับจุดที่ปรากฏบนภาพก็สามารถแปลงตำแหน่งของจุดสองมิติบนภาพไปเป็นจุดบนพิกัดสามมิติในโลกของผู้ใช้ได้ แต่ในการแปลงจุดในพิกัดสองมิติให้อยู่ในพิกัดสามมิตินั้นใช้จุดเดียวไม่เพียงพอ (เพราะการแปลงโดยใช้จุดสองมิติเพียงจุดเดียวจะได้ผลลัพธ์เพียงแค่แนวโปรเจกชันของจุดในระนาบสามมิติ) จำเป็นต้องใช้หลายๆ จุดที่มีรูปแบบการเรียงตัวที่แน่นอน (Pattern) จึงจะเพียงพอสำหรับการคำนวณ

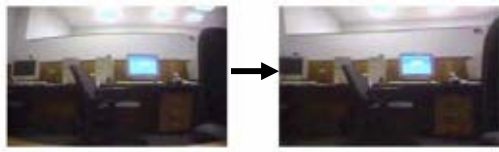
สำหรับในงานวิจัยนี้ กำหนดให้คอมพิวเตอร์รับภาพผู้ใช้โดยกล้องวงจรปิดที่ต่อกับการ์ดจับภาพวิดีโอ (Video capture card) กล้องจะถูกติดตั้งอยู่บนจอมอนิเตอร์ (เป็นตำแหน่งที่กล้องสามารถจับภาพศีรษะได้ดีที่สุด) ก่อนที่จะนำกล้องไปใช้งาน จะต้องผ่านการ Calibration โดยการนำเช็ทของจุด (x, y, z) บน Reference frame (เช่น Pattern จุดมุมตารางหมากรุกดังรูปที่ 3) และเช็ทของจุด (x, y) บนภาพที่ได้จากกล้องมาคำนวณหา Intrinsic parameters



รูปที่ 3 Pattern จุดมุมตารางหมากรุกที่ใช้ในงานวิจัย

เมื่อได้ Intrinsic parameters แล้วจึงเริ่มรับภาพผู้ใช้เข้ามาจากกล้อง แต่ภาพที่ได้จากกล้องส่วนใหญ่จะมีการบิดเบี้ยวตามแนวรัศมี หากนำไปใช้คำนวณหาตำแหน่งของศีรษะภายหลังอาจให้ผลที่ผิดพลาดได้ เพราะฉะนั้นจึงมีความจำเป็นที่จะต้องปรับแก้ภาพให้มีความเที่ยงตรงเสียก่อน โดยการนำ Radial distortion coefficient ที่ได้จาก Intrinsic parameters มาร่วมคำนวณเพื่อทำ Undistortion

ซึ่งได้ผลเป็นดังรูปที่ 4



รูปที่ 4 รูปจากกล้องที่มีการบิดเบี้ยวตามแนวรัศมี (ซ้าย) และภาพที่ปรับแก้การบิดเบี้ยวแล้ว (ขวา)

จากนั้นจึงเข้าสู่กระบวนการตรวจหาตำแหน่งของศีรษะ โดยระหว่างการใช้งานผู้ใช้จำเป็นต้องสวมหมวกคิด Pattern ที่มีลักษณะเป็นลายตารางหมากรุกขนาด 2 x 2 ช่อง เพื่อตรวจจับจุดมุมทั้ง 7 จุด ซึ่งจำเป็นสำหรับการคำนวณหา Extrinsic parameters

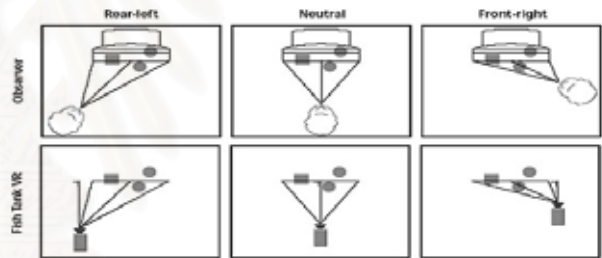
ข้อดีของการใช้การตรวจจุดมุมของตารางหมากรุกแทนการตรวจด้วยเซนเซอร์ในใบหน้า [12] คือ วิธีการนี้มีความรวดเร็วในการประมวลผล และมีความแม่นยำสูงกว่า และไม่ขึ้นกับลักษณะหน้าตาของผู้ใช้ ไม่ว่าผู้ใด เพียงแค่สวมอุปกรณ์สวมศีรษะก็สามารถใช้งานในระบบได้ทันที

เมื่อได้จุดมุมทั้ง 7 จุดแล้ว จึงนำไปคำนวณหา Extrinsic parameters จะได้ Translation Vector (ตำแหน่งของ Pattern เทียบกับกล้อง) และ Rotational Matrix (มุมที่ Pattern กระทำกับกล้อง) เมื่อรู้ตำแหน่งและมุมของ Pattern ก็จะสามารถประมาณตำแหน่งของตาและมุมมองที่ผู้ใช้กระทำกับกล้องได้ (การใช้หมวกทำให้ Pattern มีตำแหน่งคงที่เทียบกับตา) จากนั้นระบบจะต้องคำนวณการย้ายแกนเพื่อให้จุดอ้างอิงเปลี่ยนจากกล้องไปอยู่ที่ตำแหน่งกึ่งกลางจอภาพ (เนื่องจากผู้ใช้มองไปยังจอภาพ) จึงจะได้ตำแหน่งของผู้ใช้เทียบกับจอภาพตามต้องการ ในงานวิจัยนี้ตำแหน่งของผู้ใช้จะคำนวณในหน่วยของเซนติเมตร (cm) ทั้งหมด

3.2 การปรับมุมมองแบบ Fish Tank VR

Colin Ware et al. [5] ได้นำเสนอระบบความจริงเสมือนรูปแบบใหม่ที่ชื่อว่า Fish Tank VR ระบบ VR แบบนี้จะแสดงผลผ่านจอมอนิเตอร์ธรรมดาที่หาซื้อได้ทั่วไป

โดยใช้ระบบการติดตามตำแหน่งศีรษะของผู้ใช้ (Head Tracking) มาร่วมด้วย เพื่อปรับเพอซเพลกทีฟ (Perspective) สำหรับการแสดงภาพสามมิติให้สัมพันธ์กับตำแหน่งดวงตาของผู้ใช้ในขณะนั้น สำหรับในงานวิจัยชิ้นนี้จะสร้างภาพกราฟฟิกโดยใช้ OpenGL [8] เป็นหลัก เพราะฉะนั้นการปรับเพอซเพลกทีฟจะกระทำผ่านฟังก์ชันของ OpenGL ตำแหน่งศีรษะของผู้ใช้เทียบกับจอภาพที่ได้ในหัวข้อที่แล้วจะถูกนำมาเซตเป็นตำแหน่งให้กับกล้อง (ในงานวิจัยนี้ระยะห่างหนึ่งเซนติเมตรในโลกจริง จะถูกกำหนดให้เท่ากับหนึ่งหน่วยใน OpenGL) ถ้าหากผู้ใช้เคลื่อนศีรษะไปทางซ้ายหรือขวาและมองทำมุมเอียงกับจอภาพ จะต้องเคลื่อนตำแหน่งกล้องใน OpenGL ให้สัมพันธ์กัน (แต่ฉากรับภาพจะยังคงอยู่ที่เดิม ไม่เปลี่ยนแปลง) และกล้องต้องมีทิศตั้งฉากกับฉากรับภาพ Near plane เสมอ เพราะฉะนั้นการปรับเพอซเพลกทีฟจะเป็นดังรูปที่ 5



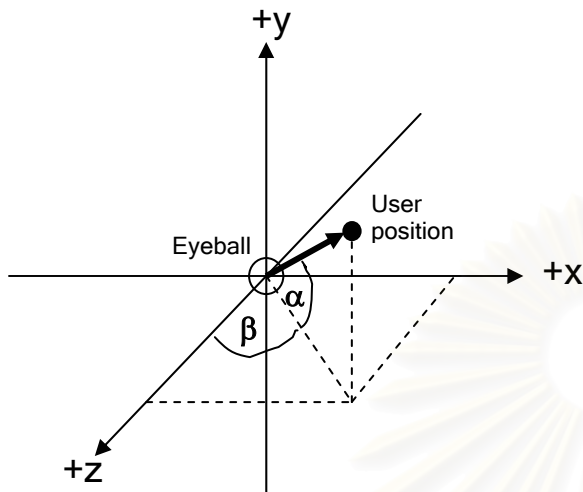
รูปที่ 5 การปรับเพอซเพลกทีฟเมื่อศีรษะผู้ใช้ทำมุมต่างๆ กันกับจอภาพ

จากรูปจะเห็นว่าฉากรับภาพ Near plane ไม่จำเป็นต้องสมมาตรกับกล้อง ซึ่ง OpenGL ได้เตรียมคำสั่งสำหรับการตั้งเพอซเพลกทีฟเช่นนี้ไว้แล้ว คือ glFrustum()

3.3 การทำให้โมเดลสบตาผู้ใช้

เพื่อให้โมเดลสามารถสบตากับผู้ใช้ได้ จึงได้ออกแบบโมเดลศีรษะมนุษย์ โดยใช้โมเดลสามชิ้นหลักๆ คือ ส่วนศีรษะ ดวงตาซ้าย และดวงตาขวา มาประกอบกัน เพื่อให้สามารถแยกควบคุมโมเดลแต่ละชิ้นได้อย่างอิสระ การทำให้โมเดลมองไปยังจุดต่างๆ สามารถทำได้โดยการหมุน

โมเดลดวงตาแต่ละข้าง รอบแกน x และ y เพื่อให้ลูกตาดำจ้องไปยังทิศทางที่ต้องการ



รูปที่ 6 แสดงมุมการหมุนของดวงตา

จากรูปที่ 6 จุดกำเนิดของแกนอยู่ตรงจุดศูนย์กลางของดวงตาที่ต้องการจะหมุน, User position คือตำแหน่งของผู้ใช้ที่คำนวณได้ เพราะฉะนั้น ดวงตากลูกนี้ต้องหมุนรอบแกน x เป็นมุม α และหมุนรอบแกน y เป็นมุมเท่ากับ β องศา โดยที่

$$\alpha = \tan^{-1} \left(\frac{y}{\sqrt{x^2 + z^2}} \right)$$

$$\beta = \tan^{-1} \left(\frac{x}{z} \right)$$

เมื่อได้มุมแล้ว ก็สามารถสั่งดวงตาให้หมุนโดยใช้คำสั่ง glRotatef() ดวงตาก็จะจ้องไปยังตำแหน่งของผู้ใช้งาน

3.4 การพูดคุยโต้ตอบ

งานวิจัยนี้ได้เลือกใช้ A.L.I.C.E Bot [19][20][21][22] เป็น A.I. ที่ทำหน้าที่สำหรับพูดคุยโต้ตอบ เนื่องจาก A.L.I.C.E Bot เป็น A.I. ที่มีมาตรฐานดี และได้รับความนิยมอย่างสูง มีการพัฒนาอย่างต่อเนื่อง ความสามารถในการโต้ตอบจะขึ้นอยู่กับข้อมูล AIML (Artificial Intelligence Markup Language) [23] ซึ่งเป็นที่เก็บความรู้

และประโยคพูดคุยที่ใช้บ่อยๆ เปรียบเสมือนเป็นมันสมองของ A.L.I.C.E Bot สำหรับไฟล์ AIML มาตรฐานที่สามารถดาวน์โหลดได้ฟรีจากอินเทอร์เน็ต จะเก็บรูปแบบประโยคสนทนาไว้กว่า 41,000 รูปแบบ ซึ่งเพียงพอต่อการพูดคุยทั่วๆ ไป หากผู้ใช้ต้องการเพิ่มรูปแบบประโยคใหม่ๆ ก็สามารถทำได้โดยไม่ยาก

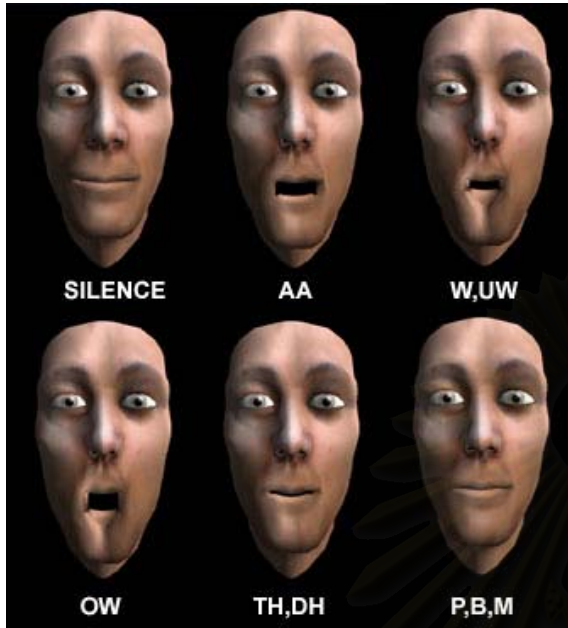
เนื่องจาก A.L.I.C.E Bot รับอินพุตและส่งเอาต์พุตเป็นข้อความตัวอักษร (Text) เพราะฉะนั้น ระบบจำเป็นต้องแปลงเสียงของผู้ใช้งานให้เป็นตัวอักษรเสียก่อนโดยใช้ Speech Recognition Engine [16] จากนั้นส่งข้อความที่ได้ให้เป็นอินพุตของ A.L.I.C.E Bot จะได้เอาต์พุตคำตอบออกมา คำตอบที่ได้ต้องเข้าสู่โมดูล Text Filter เพื่อปรับข้อความให้เหมาะสมกับการพูดออกเสียง (บางครั้ง A.L.I.C.E Bot อาจจะส่งประโยคออกมาในลักษณะนี้ “Yes..... : -)” จึงต้องปรับให้อยู่ในรูป “Yes.” มิฉะนั้นระบบจะออกเสียงผิดพลาด) จากนั้นจึงส่งต่อไปกับ Text-to-speech engine เพื่อแปลงข้อความเป็นเสียงพูดให้ผู้ใช้ได้ยิน

เพื่อให้ Speech Recognition Engine ทำงานได้ดี ผู้ใช้จำเป็นต้องเทรนระบบให้รู้จำเสียงพูดก่อน ขั้นตอนนี้อาจจะใช้เวลานาน งานวิจัยนี้จึงได้เพิ่มทางเลือกให้ผู้ใช้สามารถพิมพ์ข้อมูลเข้าโดยตรงได้ด้วย

3.5 การทำ Lip-sync

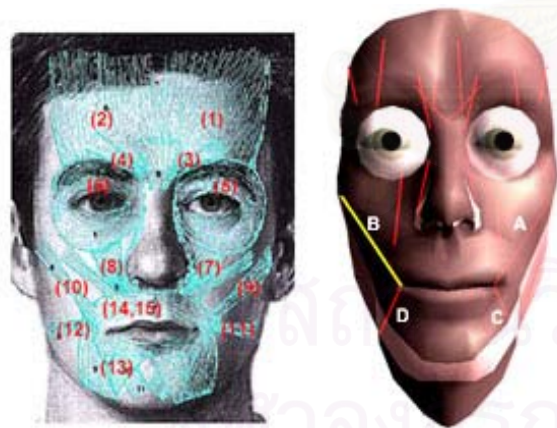
การทำ Lip-sync คือการขยับปากของโมเดลให้สอดคล้องกับเสียงพูด ในระหว่างที่ Text-to-speech Engine ทำงาน จะส่งสัญญาณที่เรียกว่า “Viseme” (รูปแบบของปากที่สัมพันธ์กับหน่วยเสียงที่เปล่งออกมาในขณะนั้น) ไปให้กับโมดูลศีรษะมนุษย์สามมิติเพื่อปรับรูปแบบปากให้สัมพันธ์กัน สำหรับ Viseme ทั้งหมดที่ Text-to-speech Engine สามารถส่งออกมาได้ มีอยู่ทั้งหมด 22 รูปแบบ คือ SILENCE, AE_AX_AH, AA, AO, EY_EH_UH, ER, Y_IY_IH_IX, W_UW, OW, AW, OY, AY, H, R, L, S_Z, SH_CH_JH_ZH, TH_DH, F_V, D_T_N, K_G_NG, และ

P_B_M สำหรับตัวอย่างปากที่สอดคล้องกับ Viseme แต่ละแบบเป็นดังรูปที่ 7



รูปที่ 7 ตัวอย่างปากที่สอดคล้องกับ Viseme แต่ละแบบ

รูปแบบปากแบบต่างๆ ของโมเดลสามมิติ สร้างขึ้นโดยใช้ระบบกล้ามเนื้อควบคุมบนใบหน้า (Muscle) [1][3][9][24] ซึ่งเป็นดังรูปที่ 8



รูปที่ 8 ระบบกล้ามเนื้อบนใบหน้า

โดยปกติแล้วเมื่อคนเราต้องการขยับปาก ขากรรไกรจะทำหน้าที่ขยับปากขึ้นลง ส่วนรูปทรงของปาก กล้ามเนื้อตามจุดต่างๆ บนใบหน้าจะบีบรัดหรือคลายตัวเพื่อให้รูปปากเป็นไปตามต้องการ เช่นเดียวกับโมเดล เมื่อจำลอง

กล้ามเนื้อ บริเวณปากจะมีกล้ามเนื้อสำคัญอยู่ 4 เส้นได้แก่ A,B,C,D ดังรูปที่ 8 (ขวา) จำลองมาจากกล้ามเนื้อหมายเลข 9,10,11,12 ตามลำดับ ดังรูปที่ 8 (ซ้าย) เพราะฉะนั้น การเก็บลักษณะปากที่สอดคล้องกับ Viseme แต่ละแบบ จะประกอบไปด้วยข้อมูลการยืดหดกล้ามเนื้อทั้ง 4 เส้น และองศาของการอ้าขากรรไกร เมื่อมีสัญญาณ Viseme เข้ามาก็เรียกใช้ข้อมูลเหล่านี้มาปรับกล้ามเนื้อและขากรรไกรให้เป็นไปตามต้องการ

3.6 การแสดงอารมณ์

การแสดงอารมณ์ จะใช้หลักการของกล้ามเนื้อควบคุม เช่นเดียวกับการทำ Lip-sync เพียงแต่การแสดงอารมณ์ต้องเกี่ยวข้องกับกล้ามเนื้อที่จุดอื่นๆ ด้วยนอกเหนือจากปาก เช่น กล้ามเนื้อที่แก้ม ตา หน้าผาก เป็นต้น เมื่อปรับค่าการยืดหดกล้ามเนื้อเหล่านี้ ก็จะได้เป็นลักษณะอารมณ์ที่แตกต่างกัน ดังรูปที่ 9

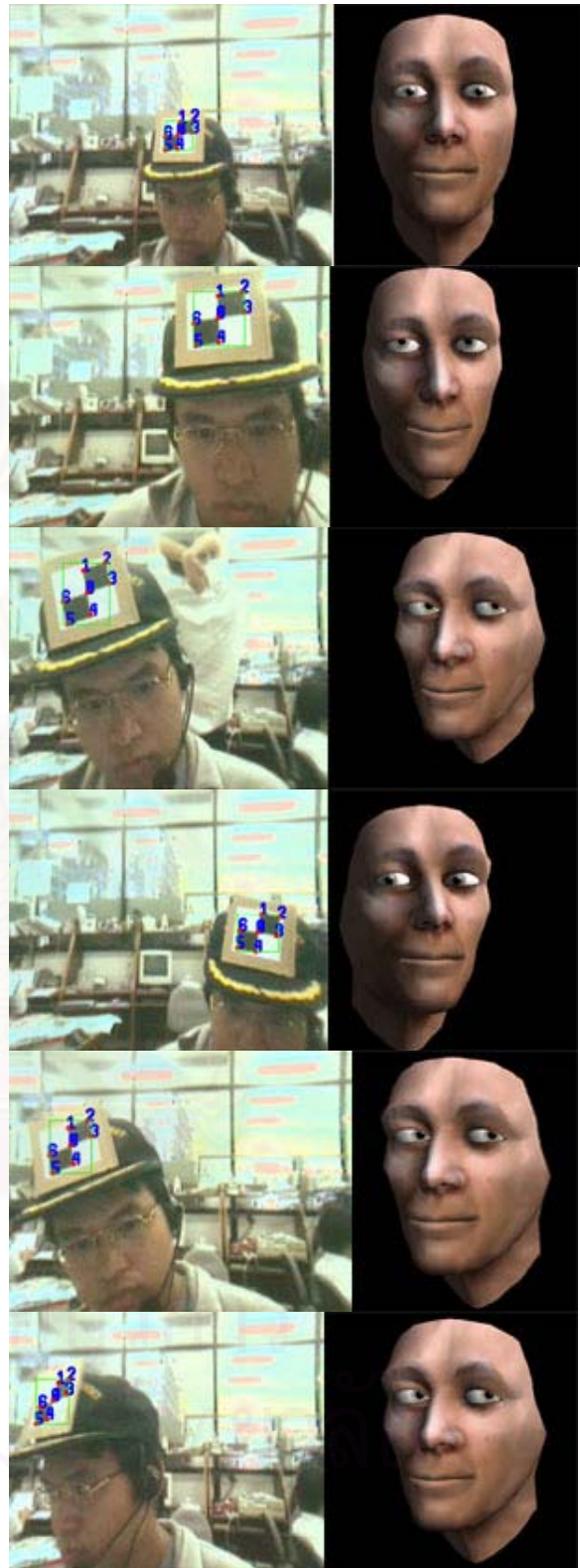
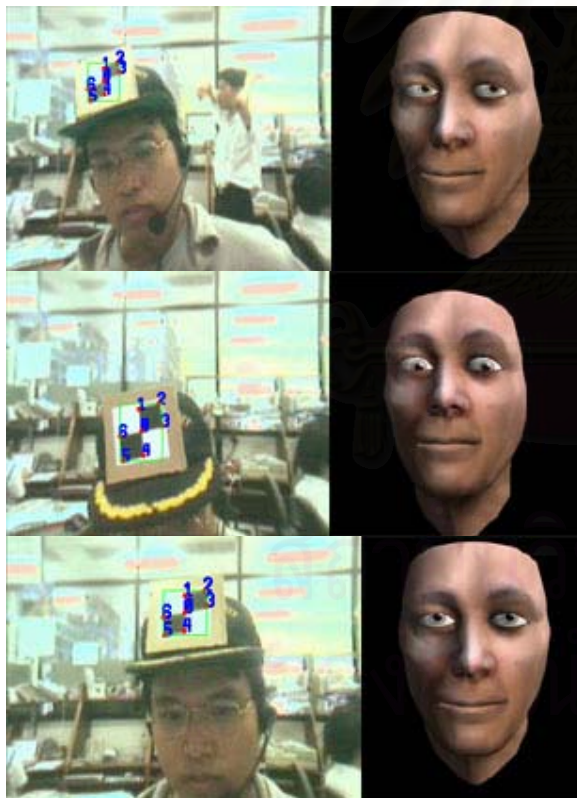


รูปที่ 9 แสดงอารมณ์แบบต่างๆ

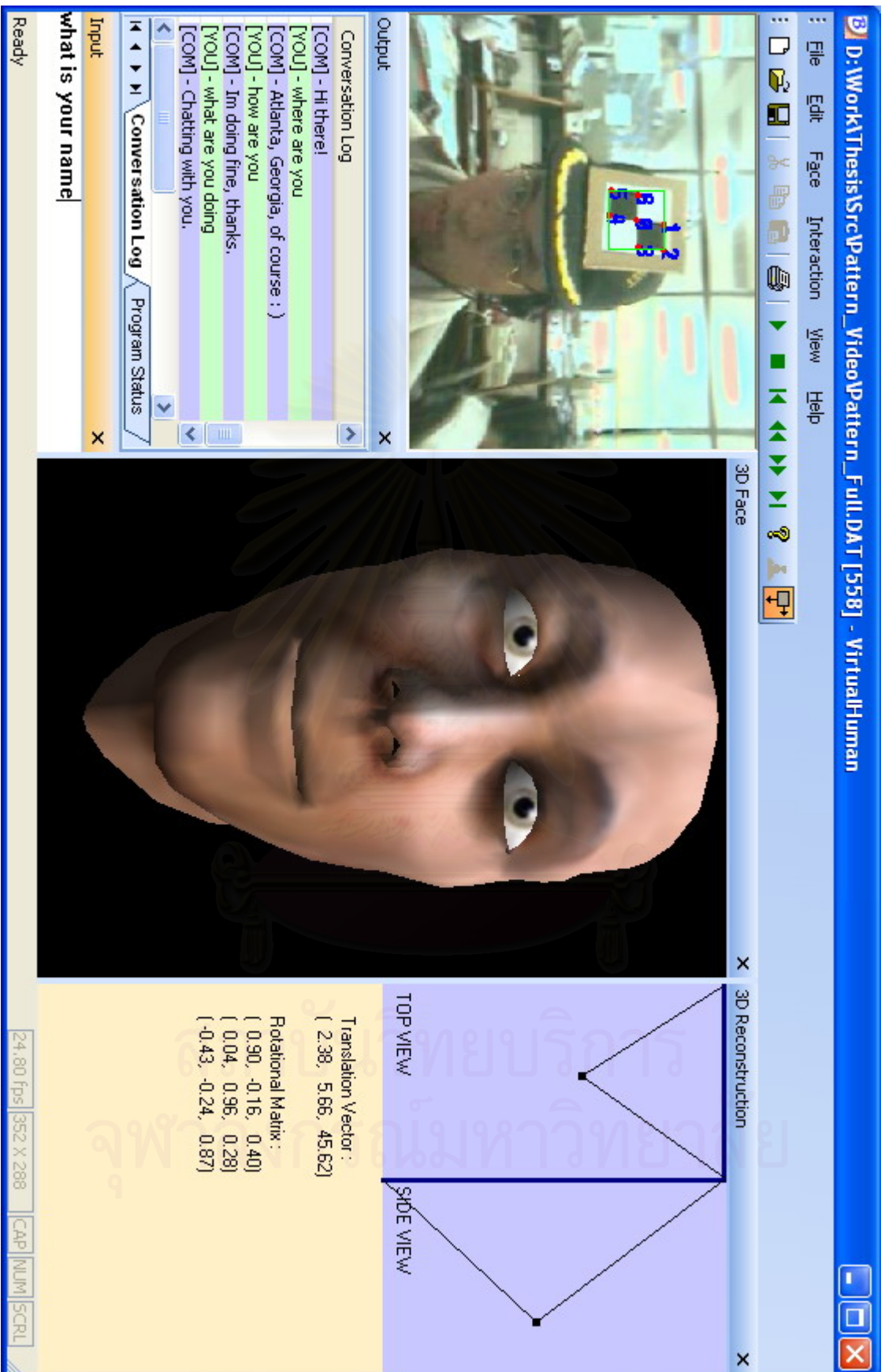
4. ผลการทดลองและการใช้งาน

จากการทดลอง โปรแกรมสามารถตรวจจับ Pattern พร้อมทั้งคำนวณตำแหน่งผู้ใช้ได้อย่างถูกต้องและรวดเร็ว โดยมีค่าความคลาดเคลื่อนไม่เกิน 5% ในการใช้งานจริงได้ ความเร็วสูงถึง 24.9 เฟรมต่อวินาที (fps) บนเครื่องที่ใช้ ซีพียู Intel Celeron 2.40 GHz, Ram 768MB ต่อกับกล้องวงจรปิดผ่าน Video Capture Card ในโหมด PAL (กล้องจะถูกติดตั้งอยู่บนจอมอนิเตอร์)

สำหรับโมเดลสามมิติ ก็สามารถขยับไปยังผู้ใช้และปรับมุมมองกล้องแบบ Fish Tank VR ได้อย่างถูกต้อง ในส่วนของการพูดคุยกับผู้ใช้ ก็สามารถคุยได้ในระดับที่ดี อาจจะมีผลจากความผิดพลาดของระบบรู้จำเสียงพูดอยู่บ้าง แต่ถ้าหากเทรนระบบให้รู้จำเสียงพูดมากกว่านี้ ข้อผิดพลาดก็จะน้อยลง สำหรับผลที่ได้ แสดงได้ดังรูปที่ 10



รูปที่ 10 ผลการทำงานของโปรแกรม



รูปที่ 11 แสดงการสนทนา และหน้าจอโปรแกรมทั้งหมด

5. สรุป

งานวิจัยชิ้นนี้พัฒนาขึ้นเพื่อให้ผู้ใช้ได้รับความสะดวกและความรู้สึกเสมือนจริงในการติดต่อสื่อสารกับคอมพิวเตอร์ ผู้ใช้จะรู้สึกว่ามีโมเดลศีรษะและใบหน้ามนุษย์สามมิติ อยู่ในจอมอนิเตอร์จริงๆ ซึ่งจะแปรเปลี่ยนมุมมองไปตามตำแหน่งของผู้ใช้ และมีปฏิสัมพันธ์กับผู้ใช้ในแบบต่างๆ กัน ทำให้การติดต่อกับคอมพิวเตอร์มีความน่าสนใจมากยิ่งขึ้น

6. เอกสารอ้างอิง

- [1] Bill Freming, Darris Dobbs, "Animating Facial Features & Expressions", Charles River Media Inc. 1999
- [2] Byoungwon Choe, Hyeong-Seok Ko, "Analysis and Synthesis of Facial Expressions with Hand-Generated Muscle Actuation Basis", Proceedings of Computer Animation 2001, pages 12–19, November 2001
- [3] C. Pelachaud, N. Badler, and M. Steedman, "Generating facial expressions for speech", Cognitive Science 20(1), pp. 1-46, 1996.
- [4] Gagatay Demiralp, David H. Laidlaw, Cullen Jackson, Daniel Keefe, and Song Zhang. "Subjective usefulness of cave and fish tank vr display systems for a scientific visualization application.", In Poster Compendium IEEE Conference on Visualization, Seattle, WA, 2003
- [5] Colin Ware, Kevin Arthur, Kellogg S. Booth, "Fish Tank Virtual Reality", Proceedings of the SIGCHI conference on Human factors in computing systems, 1993
- [6] Cynthia Breazeal, Brian Scassellati, "How to build robots that make friends and influence people.", IROS99, Kyonju, Korea, 1999.
- [7] Cynthia Breazeal, Brian Scassellati, "Robot in Society: Friend of Appliance?", In Agents99 Workshop on Emotion-based Agent Architectures, Seattle, WA. 18-26, 1999.
- [8] Dave Shreiner, "OpenGL(R) Reference Manual: The Official Reference Document to OpenGL", Version 1.2 (3rd Edition)
- [9] Frederic I. Parke, Keith Waters, "Computer Facial Animation", A K Peters Ltd. 1996
- [10] Intel Corporation, "Open Source Computer Vision Library Reference Manual", 2001, <http://sourceforge.net/projects/opencvlibrary/>
- [11] Jonathan Gratch, Jeff Rickel, Elisabeth Andr, Justine Cassell, Eric Petajan, Norman Badler, "Creating Interactive Virtual Humans: Some Assembly Required", IEEE Intelligent Systems, Volume 17, Issue 4, July-August 2002, pages 54-63.
- [12] Jun Rekimoto, "A Vision-Based Head Tracker for Fish Tank Virtual Reality - VR without Head Gear", Proceedings of VRAIS'95 ,1995, Pages 94-100
- [13] Magy Seif El-Nasr, T. Ioerger, J. Yen, D. House, and F. Parke., "Emotionally Expressive Agents." Proc. of Computer Animation '99, Geneva, Switzerland, 1999
- [14] Michael Treadgold, Kevin Novins, Geoff Wyvill, and Brian Niven, "What do you think you're doing? Measuring perception in Fish Tank Virtual Reality", Computer Graphics International Conference, 2001
- [15] Michel D. Bondy, Emil M. Petriu, Marius D. Cordea, "Model-based Face and Lip Animation for Interactive Virtual Reality Applications", Conference '01, Sept 30 - Oct. 5, 2001, Ottawa, Ont. Canada.
- [16] Microsoft Corporation, "Microsoft Speech SDK 5.1 Documentation"
- [17] Muller W, Spierling U, Alexa M, Rieger T. "Face-to-face with your assistant - realization issues of animated user interface agents for home appliances.", Computers and Graphics, 25:593–600, 2001.
- [18] Paula Salgado Lucena, Marcelo Gattass, and Luiz Velho. "Expressive Talking Heads: A Study on Speech and Facial Expression in Virtual Characters.", Revista SCIENTIA, 2002.
- [19] Peter M. Plantec, Ray Kurzweil , "Virtual Humans: A Build-It-Yourself Kit, Complete With Software and Step-By-Step Instructions", AMACOM; Book and CD-ROM edition , November 2003
- [20] Richard S. Wallace , "Be Your Own Botmaster", Alice Foundation , 2546
- [21] Richard S. Wallace, "AIML Overview", A.L.I.C.E. Artificial Intelligence Foundation, Inc
- [22] Richard S. Wallace, " The Anatomy of A.L.I.C.E.", A.L.I.C.E. Artificial Intelligence Foundation, Inc.
- [23] Richard S. Wallace, "A.L.I.C.E. and AIML Documentation" , A.L.I.C.E. Artificial Intelligence Foundation, Inc, 2000
- [24] S. Pasquariello, C. Pelachaud, "Greta: A Simple Facial Animation Engine", 6th Online World Conference on Soft Computing in Industrial Applications, Session on Soft Computing for Intelligent 3D Agents, September 2001

ประวัติผู้เขียนวิทยานิพนธ์

นายอรรถกร พูนศิลป์ เกิดวันที่ 4 สิงหาคม พ.ศ.2523 ที่จังหวัดปัตตานี สำเร็จการศึกษาระดับปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ จากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2545 หลังจากนั้นได้เข้ามาศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2545



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย