

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็นการจำแนกคลิปภาพยนตร์โดยแบ่งประเภทตามอารมณ์ของผู้ชมเป็นเกณฑ์ ซึ่งแบ่งลักษณะของอารมณ์ไว้สี่ประเภทได้แก่ อารมณ์กลัว (Fear) อารมณ์ตื่นเต้น (Excitement) อารมณ์สนุกสนาน (Joy) และอารมณ์เศร้า (Sadness) ดังนั้นในบทนี้จะอธิบายพื้นฐานข้อมูลวิดีโอ คุณลักษณะสำคัญของวิดีโอ แบบจำลองสี และงานวิจัยอื่นที่เกี่ยวข้องกับการจำแนก

2.1 พื้นฐานข้อมูลวิดีโอ

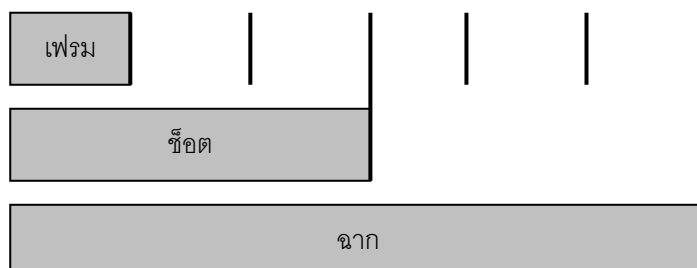
2.1.1 โครงสร้างพื้นฐานข้อมูลวิดีโอ

วิดีโอ คือการเรียงต่อกันของภาพเป็นเรื่องราว เนื้อหาภายในวิดีโอ แสดงดังภาพที่ 2.1 มีการแบ่งออกเป็นส่วนต่างๆ ดังนี้

- 1) เฟรม (Frame) หมายถึง ภาพนิ่งแต่ละภาพที่เรียงต่อกันจนเป็นเรื่องราว
- 2) ช็อต (Shot) หมายถึง กลุ่มของเฟรมภาพเคลื่อนไหวในวิดีโอที่มีเนื้อหาเดียวกัน และปรากฏต่อเนื่องกันโดยตลอด
- 3) ฉาก (Scene) หมายถึง การรวมกันของช็อตหลายๆ ช็อต โดยในทางภาพยนตร์มีความหมาย คือ สถานที่ที่เหตุการณ์เกิดขึ้น และได้หมายความว่าเฉพาะสถานที่และสิ่งต่างๆ ในสภาพแวดล้อมนั้น แต่ยังรวมถึงอารมณ์และบรรยากาศของสถานที่ด้วย

ภาพที่ 2.1

โครงสร้างพื้นฐานข้อมูลวิดีโอ



2.1.2 มาตรฐานการแพร่ภาพของวิดีโอ

ระบบวิดีโอ มีความสัมพันธ์กับการนำไฟล์วิดีโอไปเผยแพร่ในรูปแบบต่างๆ ในปัจจุบัน มาตรฐานการแพร่ภาพของวิดีโอ แบ่งออกเป็น 3 รูปแบบดังนี้

1) ระบบ PAL (Phase Alternate Line) เป็นรูปแบบวิดีโอที่มีความละเอียดในแนวนอนเท่ากับ 625 เส้น ใช้อัตราการแสดงผลภาพ (Frame Rate) เท่ากับ 25 เฟรม ต่อ 1 วินาที วิดีโอระบบนี้จะใช้ในประเทศออสเตรเลีย เบลเยียม จีน ฟินแลนด์ เยอรมัน ฮังการี อิตาลี สิงคโปร์ มาเลเซีย และไทย เป็นต้น

2) ระบบ NTSC (National Television System Committee) เป็นรูปแบบวิดีโอที่มีความละเอียดในแนวนอนเท่ากับ 525 เส้น ใช้อัตราการแสดงผลภาพ (Frame Rate) เท่ากับ 29.97 เฟรม ต่อ 1 วินาที วิดีโอระบบนี้จะใช้ในประเทศสหรัฐอเมริกา ญี่ปุ่น แคนาดา โคโลัมเบีย เม็กซิโก ไต้หวัน ฟิลิปปินส์ เป็นต้น

3) ระบบ SECAM (Sequential Color and Memory) เป็นรูปแบบวิดีโอที่มีความละเอียดในแนวนอนเท่ากับ 816 เส้น ใช้อัตราการแสดงผลภาพ (Frame Rate) เท่ากับ 25 เฟรม ต่อ 1 วินาที วิดีโอระบบนี้จะใช้ในประเทศบัลแกเรีย ฝรั่งเศส อิรัก อิหร่าน รัสเซีย โรมาเนีย เป็นต้น

2.1.3 คุณภาพของวิดีโอ

การวัดคุณภาพของวิดีโอสามารถวัดได้จาก อัตราการแสดงผลภาพ และความละเอียดของภาพ

1) อัตราการแสดงผลภาพ (Frame Rate) คืออัตราความเร็วในการแสดงผลภาพเคลื่อนไหวต่อหนึ่งหน่วยเวลา มีหน่วยเป็นเฟรมต่อวินาที (fps : frame per second เช่น อัตราเฟรมของภาพยนตร์เท่ากับ 24 เฟรม ต่อ 1 วินาที อัตราเฟรมระบบ PAL เท่ากับ 25 เฟรม ต่อ 1 วินาที และอัตราเฟรมของ NTSC เท่ากับ 30 เฟรมต่อ 1 วินาที

2) ความละเอียด หรือขนาดของเฟรม หมายถึง ความชัดของภาพที่แสดงผลออกมาทางจอภาพ ความละเอียดของจอภาพขึ้นอยู่กับจำนวนจุดภาพทั้งหมดที่เกิดบนจอภาพ เช่น ความละเอียดของจอภาพ 640 x 480 จุดภาพ หมายถึง มีจำนวนจุดภาพแสดงผลเรียงกันอยู่บนจอภาพในแนวนอน 640 จุดภาพและแนวตั้ง 480 จุดภาพ

ในงานวิจัยนี้จะนำไฟล์วิดีโออยู่ในรูปแบบไฟล์ MPEG1 เป็นระบบ PAL (Phase Alternate Line) มีอัตราการแสดงผลภาพ (Frame Rate) เท่ากับ 25 เฟรมต่อวินาที (fps) โดยกำหนดขนาด

ความกว้างคูณความยาวเท่ากับ 352 x 288 pixels ตัวอย่างที่แสดงในภาพ 2.2 เป็นตัวอย่างฉากที่อยู่ในคลิปวิดีโอ ในแต่ละฉากก็จะประกอบไปด้วยจำนวนของช็อตอย่างน้อย 1 ช็อต

ภาพที่ 2.2
ส่วนประกอบต่างๆ ของคลิปวิดีโอ



2.1.4 องค์ประกอบหลักในการสร้างภาพยนตร์แนวสยองขวัญ

การศึกษาองค์ประกอบหลักในการสร้างภาพยนตร์เป็นสิ่งสำคัญปัจจัยหนึ่ง ในหัวข้อนี้เป็นการยกตัวอย่างองค์ประกอบหลักในการสร้างภาพยนตร์แนวสยองขวัญ (Horror Film) เนื่องจากภาพยนตร์แนวนี้เป็นแนวที่ให้ความรู้สึก หรืออารมณ์กลัว (Fear) แก่ผู้ชมมากที่สุด (ศิรินาฏ กุณธร, 2550) ดังนั้นการศึกษาองค์ประกอบในการสร้างภาพยนตร์มีส่วนช่วยในการดึงคุณลักษณะสำคัญเพื่อมาใช้ในการจำแนกต่อไป

องค์ประกอบหลักในการสร้างภาพยนตร์แนวสยองขวัญที่สำคัญ (กฤษดา เกิดดี ,2547) ดังนี้

1) ช่วงเวลาของความตื่นเต้นระทึกใจ (Moments of suspense) เป็นระยะเวลาช่วงหนึ่งในภาพยนตร์ที่เกิดเหตุการณ์ที่ทำให้คนดูต้องเกิดความตื่นเต้นระทึกใจ หรืออาจถึงกับอาการผวาสุดขีด ซึ่งจะมากน้อยเพียงใดขึ้นอยู่กับบุคลิกภาพของคนดู และความสามารถของผู้กำกับ

2) องค์ประกอบในส่วนของตัวละคร (Helpless character) เป็นตัวละครซึ่งตกอยู่ในสถานะที่ไม่สามารถช่วยเหลือตัวเองได้อันเนื่องมาจากสาเหตุต่างๆ ตามแต่จะกำหนด ส่วนมากมักเป็นเพราะการมีชีวิตอยู่อย่างโดดเดี่ยว และถูกตัดขาดจากสังคมภายนอก

3) สถานการณ์ที่จมนม (Trapped situation) เป็นสถานการณ์ที่เกิดขึ้นใช้เวลาสั้นๆ ช่วงใดช่วงหนึ่งในภาพยนตร์ หรืออาจเป็นระยะเวลาทั้งหมดตลอดเรื่อง

4) เหตุผลทางด้านจิตวิทยา (Psychic element) เป็นการกำหนดที่ตัวละคร เช่น หากเป็นมนุษย์ส่วนมากจะกำหนดให้เป็นผู้ร้าย ฆาตกร หรือป่วยเป็นโรคจิตเสมอ

2.2 คุณลักษณะสำคัญของวิดีโอ

การดึงลักษณะสำคัญของข้อมูลวีดิโอนั้นเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งเพื่อนำไปใช้ในการจำแนกคลิปภาพยนตร์ ข้อมูลวีดิโอสามารถแบ่งตามลักษณะออกได้เป็นสามแบบ คือ

2.2.1 ลักษณะจากสิ่งที่เห็น (Visual Feature)

ลักษณะจากสิ่งที่เห็นหมายถึง ภาพ สี แสง ข้อความหรือสิ่งต่างๆ ที่ปรากฏบนข้อมูลวีดิโอที่สามารถมองเห็นได้ โดยคุณลักษณะสำคัญจากสิ่งที่เห็นของภาพสามารถสื่อความคิดและอารมณ์โดยการรับรู้ทางตาของผู้ชมด้วยภาพที่เคลื่อนไหว(เสนห์ ปัญโญ, 2547) ในทางภาพยนตร์สื่อความหมายด้วยภาพขนาดต่างๆ กัน ตั้งแต่ช็อตไกล (Long shot) ซึ่งเป็นช็อตมุมกว้าง (Wide-Angle shot) ที่เห็นภาพภูมิทัศน์กว้างจนกระทั่งถึงช็อตใกล้ (Close-up shot) ที่เห็นเฉพาะตาของคนเรา ในระหว่างช็อตทั้งสองดังกล่าวยังมีช็อตต่างๆ อีกหลายขนาดซึ่งในแต่ละขนาดของภาพต่างก็ให้ความหมายและสื่อถึงอารมณ์แตกต่างกันไปจากการสื่อความหมายด้วยภาพขนาดต่างๆ แล้วในทางภาพยนตร์ยังให้ความสำคัญเกี่ยวกับสีและแสงเพื่อเป็นการสื่อถึงอารมณ์แก่ผู้ชม

สี (color) คือส่วนที่แสดงถึงสีที่กระจายอยู่ภายในภาพ (Wang, & Huang, 2000) สีของภาพจะประกอบด้วย 3 ส่วนคือ สีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) มีงานวิจัยหลายงานที่นำลักษณะสำคัญของสีมาใช้ในด้านการจำแนก แต่การใช้สีอย่างเดียวยังไม่เพียงพอ เช่น สีเขียวอาจจะไม่ได้หมายถึงต้นไม้อย่างเดียว อาจจะหมายถึงลูกบอลสีเขียวหรือวัตถุอื่นๆ สีเขียวเป็นต้น การตอบสนองต่อสีต่างๆ ของมนุษย์จะซับซ้อนตามวัฒนธรรมและประสบการณ์ (นิรุต ภูวิชิต, 2549) เป็นผลให้สีสามารถสร้างความพอใจหรือไม่พอใจและสร้างความรู้สึกลงในแง่บวกหรือสร้างความรู้สึกลงในแง่ลบ เช่น ทำให้รู้สึกหดหู่หรืออาจมีความหมายอื่นๆ แตกต่างกันไป เช่น ใน

วัฒนธรรมทางตะวันตกถือว่าสีแดงแสดงความรุนแรงและไม่ปลอดภัย ในขณะที่วัฒนธรรมทางตะวันออกถือว่าสีแดงจะนำสิ่งที่เป็นมงคลและความสุขมาสู่ครอบครัวเป็นต้น มีตัวอย่างงานวิจัยเกี่ยวกับการเชื่อมโยงสีกับสภาวะอารมณ์ (ซูฟิยา เจอารง , 2547) ผลการวิจัยคือสีดำกับสภาวะอารมณ์โศกเศร้า สีแดงกับสภาวะอารมณ์ตื่นเต้น สีส้ม,เหลืองกับสภาวะอารมณ์สนุกสนาน เป็นต้น

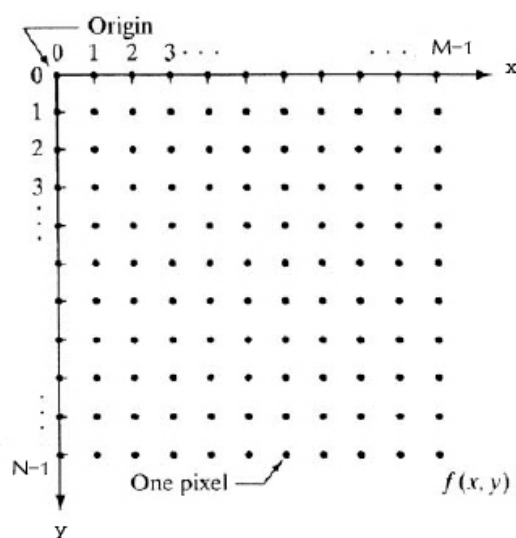
แสง (Lighting) กล่าวคือ ในภาพยนตร์จะมีการจัดแสงเพื่อให้เกิดแสงและเงาในฉาก เพื่อสร้างอารมณ์และบรรยากาศให้กับผู้ชม (เสนห์ ปัญโญ, 2547) ตัวอย่างการสื่อถึงอารมณ์ ซึมเศร้า หรือต้องการความลึกลับซับซ้อน จะมีการจัดแสงในฉากนั้นให้เป็นภาพโทนมืด (Low Key) หมายถึงภาพที่บริเวณส่วนใหญ่ของภาพเป็นสีแก่อาจจะมีสีอ่อนปนอยู่บ้างเล็กน้อย แต่หากต้องการสื่อถึงความรู้สึกบริสุทธิ์ อ่อนโยน และโรแมนติกจะมีการจัดแสงในฉากให้เป็นภาพโทนสว่าง (High Key) หมายถึง ภาพที่บริเวณส่วนใหญ่ของภาพเป็นสีอ่อน อาจจะมีสีแก่ปนอยู่บ้างเล็กน้อย

พื้นผิว (Texture) คือ ส่วนที่ปรากฏครอบคลุมรูปภาพทั้งหมด แสดงถึงลวดลายต่างๆ ภายในภาพและเป็นส่วนที่สื่อออกมาให้รับรู้ถึงความรู้สึก ซึ่งก็แบ่งเป็น 2 ลักษณะใหญ่ๆ คือ พื้นผิวเชิงบวก ละเอียด เรียบ ให้ความรู้สึกใหม่ สะอาด ปลอดภัย เป็นระเบียบ สงบ โดด และ พื้นผิวเชิงลบ หยาบ ขรุขระ ให้ความรู้สึกไร้ระเบียบ ไม่สงบ สับสน น่ากลัว ไม่น่าเข้าใกล้ ในทางภาพยนตร์พื้นผิวอาจหมายถึง ความสลับซับซ้อน หรือความเรียบโล่งของภาพ ตัวอย่างเช่น ฉากโกดังสินค้าที่มีกล่อง ลัง ม้วนกระดาษ ท่อ วางอยู่กระจัดกระจายไม่เป็นระเบียบ ฉากนี้ก็เรียกว่ามีพื้นผิวขรุขระ เป็นต้น

การดึงลักษณะสำคัญของภาพ เช่น สี แสง พื้นผิว นั้นจำเป็นต้องเข้าใจหลักการประมวลผลภาพเชิงตัวเลข (Digital Image Processing) (จิรา แก้วสุวรรณ, 2549) ดังนี้

จุดภาพ (Pixel) คือ ความเข้มแสงที่รวมกันทำให้เกิดเป็นภาพ ภาพหนึ่งๆ จะประกอบด้วยจุดภาพมากมาย ซึ่งภาพแต่ละภาพที่สร้างขึ้นจะมีความหนาแน่นของจุดภาพเหล่านี้แตกต่างกันออกไป ความหนาแน่นนี้เป็นตัวบอกถึงความละเอียด (Resolution) ของภาพซึ่งมีหน่วยเป็น ppi (Pixel Per Inch) คือ จำนวนจุดภาพต่อนิ้ว ดังภาพที่ 2.3

ภาพที่ 2.3
ตำแหน่งของจุดภาพ



N = จำนวนจุดภาพที่มากที่สุดในการแกน Y

M = จำนวนจุดภาพที่มากที่สุดในการแกน X

ในภาพหนึ่ง ๆ เราสามารถอธิบายได้ในรูปแบบเมตริกซ์ของจุดภาพขนาด $N \times M$ ดังภาพที่ 2.4 โดยใช้คู่ลำดับ $f(i, j)$ แทนค่าของแต่ละจุดภาพ และบ่งชี้ความเข้มแสงที่จุดภาพนั้น ๆ ของภาพ

ภาพที่ 2.4
เมตริกซ์ของจุดภาพในภาพ

	0	1	2	3	4	$\dots$	M
0	$f(0,0)$	$f(0,1)$	$f(0,2)$	$f(0,3)$	$f(0,4)$	$\dots$	
1	$f(1,0)$	$f(1,1)$	$f(1,2)$	$f(1,3)$	$f(1,4)$	$\dots$	
2	$f(2,0)$	$f(2,1)$	$f(2,2)$	$f(2,3)$	$f(2,4)$	$\dots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
i	$f(i,0)$	$f(i,1)$	$f(i,2)$	$f(i,3)$	$f(i,4)$	$\dots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
N	$f(N,0)$	$f(N,1)$	$f(N,2)$	$f(N,3)$	$f(N,4)$	$\dots$	

ค่าที่กำกับแต่ละจุดภาพจะแสดงถึงค่าเฉลี่ยของความเข้มแสงในภาพที่จุดภาพนั้น โดยค่าของจุดภาพดังกล่าวจะเขียนแทนด้วย $f(i, j)$ โดยทั่วไปแล้วภาพจะมีความเข้มตั้งแต่ 2 ระดับขึ้นไปแต่ที่นิยมใช้กันมากคือค่าระดับความเข้มของจุดภาพที่เท่ากับ 256 ระดับ ซึ่งจะทำให้ค่าของจุดภาพอยู่ในช่วง (0-255) โดยใช้เนื้อที่ในการเก็บข้อมูลขนาด 1 ไบต์หรือ 8 บิต ในกรณีที่ต้องการภาพที่มีความเข้มสูงอาจจะต้องการจำนวนบิตสำหรับการเก็บข้อมูลมากกว่า 8 บิต คือ อาจจะเป็น 16 หรือ 24 บิต โดยจะแยกความแตกต่างของภาพแต่ละประเภทให้เห็นอย่างชัดเจนได้ ดังนี้

- 1) ภาพ 2 ระดับ คือมีจุดภาพสีขาวกับสีดำเท่านั้น โดยแต่ละจุดภาพจะมีขนาดของข้อมูลเท่ากับ 1 บิต
- 2) ภาพ 16 ระดับ คือในแต่ละจุดภาพจะมีขนาดของข้อมูล 4 บิต ซึ่งทำให้สามารถแสดงได้ 16 ระดับสี
- 3) ภาพ 256 ระดับ คือในแต่ละจุดภาพจะมีขนาดของข้อมูล 8 บิต ซึ่งทำให้สามารถแสดงภาพได้ ความเข้มถึง 256 ระดับ
- 4) ภาพทูล์คัลเลอร์ (True Color) คือในแต่ละจุดภาพจะมีขนาดของข้อมูล 24 บิตทำให้ สามารถแสดงภาพออกมาได้เหมือนจริงมากที่สุด เพราะสามารถแสดงสีได้ถึง 16,777,216 สี

2.2.2 ลักษณะเสียง (Audio Feature)

เสียงในภาพยนตร์เป็นการสื่อความหมายช่วยให้ผู้ชมเข้าใจสิ่งที่เห็น และอารมณ์ของ ภาพยนตร์ได้อย่างสมบูรณ์ เสียงในภาพยนตร์แบ่งออกเป็น 3 ชนิด คือ

- 1) เสียงพูด (Speech) คือ เสียงที่สัมพันธ์พร้อมกับภาพ ผู้แสดงพูดพร้อมๆ กับได้ยิน คำพูดนั้น หรืออาจกล่าวได้ว่าภาพและเสียงสัมพันธ์กันตามความเป็นจริง
- 2) เสียงประกอบ (Sound Effect) คือ เสียงที่เกิดขึ้นรอบๆ ผู้แสดง โดยหลักแล้วใช้ เพื่อช่วยให้เกิดความสมจริงที่สัมพันธ์กับสิ่งที่ปรากฏบนจอ เช่น ภาพคนเดินขึ้นบันไดจะมาพร้อมๆ กับเสียงฝีเท้า หรือเสียงบางอย่างที่ไม่ได้เกิดขึ้นในชีวิตจริงแต่ไปเสริมความรู้สึกให้กับภาพได้ เช่น เสียงฮัมต่ำ ๆ ก็อาจใช้แทนความรู้สึกถึงอันตรายที่กำลังจะเกิดขึ้น
- 3) เสียงดนตรี (Music) เป็นเสียงในลักษณะหนึ่งเพื่อใช้เสริมอารมณ์มากกว่าการ บอกเรื่องราวเป็นส่วนที่ให้ความรู้สึกมากกว่าการได้ยิน

ในงานวิจัยนี้คลิปเสียงที่ใช้ถูกแยกออกจากไฟล์วีดีโออยู่ในรูปแบบไฟล์ wav อัตราการสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 44,100 เฮิร์ตซ์ (Hz) ระดับความลึกของเสียง (Bit Depth) 16 บิต ช่องของเสียง (Channel) เป็นแบบหนึ่งช่องทาง (mono) โดยเสียงที่ได้จะรวมทั้งเสียงพูด เสียงประกอบและเสียงดนตรี

หลักการเบื้องต้นเกี่ยวกับเสียง (อาณนท นามสนิท 2549) มีดังนี้ เสียงคือพลังงานอย่างหนึ่ง เกิดจากสั่นสะเทือนของวัตถุ การสั่นสะเทือนนี้จะแผ่กระจายออกจากวัตถุที่สั่นแล้วแผ่รัศมีออกไปโดยรอบผ่านตัวกลาง ซึ่งอาจจะเป็นอากาศ น้ำ หรือของแข็งก็ได้ ประกอบด้วยคุณสมบัติพื้นฐานดังต่อไปนี้

1) ระดับเสียง (Pitch) คือ ระดับความสูงต่ำของเสียงโดยเราสามารถเปรียบเทียบเสียงสองเสียงได้ว่าเสียงไหนมีระดับสูงกว่า และเสียงไหนมีระดับเสียงต่ำกว่าโดยการฟัง ระดับเสียงเกิดจากความถี่ (Frequency) ของการสั่นสะเทือนของวัตถุหรืออาจกล่าวได้ว่าความถี่ของการสั่นสะเทือนเป็นตัวกำหนดระดับเสียงวัตถุที่สั่นสะเทือนเร็วกว่าทำให้เกิดระดับเสียงสูงกว่า ในขณะที่วัตถุที่สั่นสะเทือนช้ากว่าทำให้เกิดระดับเสียงต่ำกว่าความถี่ซึ่งเป็นตัวกำหนดระดับเสียงได้มาจากการวัดความถี่ของวัตถุโดยมีหน่วยเป็นรอบต่อวินาที

2) ความเข้มของเสียง (Volume หรือ Intensity) หมายถึง เสียงเบาเสียงดัง ความเข้มเสียงเกิดจากแรงสั่นสะเทือนของวัตถุที่เป็นแหล่งกำเนิดเสียงสั่นสะเทือนมากเสียงจึงดังมากและในทางกลับกันถ้าวัตถุสั่นสะเทือนน้อยเสียงก็จะเบาความเข้มของเสียงขึ้นอยู่กับความแรงที่ส่งจากแหล่งกำเนิดเสียงไปยังหู ความเข้มของเสียงวัดได้จากความสูงคลื่นเสียง (Amplitude) ถ้าคลื่นเสียงยิ่งสูงมากเสียงก็จะยิ่งดัง

2.2.3 ลักษณะการเคลื่อนไหว (Motion Feature)

ลักษณะการเคลื่อนไหวเป็นส่วนหนึ่งที่นิยมนำไปเป็นการสร้างดัชนีข้อมูลวีดีโอ เนื่องจากจะมีทิศทางการเคลื่อนที่เข้ามาเกี่ยวข้อง การเคลื่อนไหวเกิดขึ้น 2 ลักษณะ (นิรุท อำนวยศิริ, 2546) คือ

(1) Subject Movement คือ กล้องอยู่กับที่แต่มีวัตถุเคลื่อนไหว

(2) Camera Movement คือ วัตถุอยู่กับที่แล้วกล้องเคลื่อนไหว หรือกล้องเคลื่อนไปตามวัตถุที่กำลังเคลื่อนไหว เช่น Panning เป็นการหมุนกล้องที่ตั้งอยู่กับที่ไปรอบๆ แกนในแนวนอน Tilting เป็นการหมุนกล้องที่ตั้งอยู่กับที่ไปรอบๆ แกนในแนวตั้ง เป็นต้น

มีหลายวิจัยที่ศึกษาลักษณะการเคลื่อนไหวเพื่อนำมาใช้ในการหาจุดเปลี่ยนแปลงของช็อต (Shot Detection) การเปลี่ยนช็อตซึ่งมีอยู่ด้วยกันสองประเภท คือการเปลี่ยนช็อตแบบทันทีทันใด (Abrupt Change) และการเปลี่ยนช็อตแบบค่อยเป็นค่อยไป (Gradual Change) (บุรินทร์ นรินทร์, 2549) รูปแบบการเปลี่ยนช็อตทั้งสองลักษณะนี้ สามารถแยกออกได้เป็นประเภทต่างๆ ดังนี้

(1) การเปลี่ยนช็อตแบบตัดเป็นรูปแบบของการเปลี่ยนช็อตแบบทันทีทันใดรูปแบบหนึ่ง โดยจะเปลี่ยนระหว่างภาพสองภาพในวิดีโอแบบทันทีทันใด

(2) การเปลี่ยนช็อตแบบเลื่อนเป็นรูปแบบการเปลี่ยนช็อตรูปแบบหนึ่งในการเปลี่ยนช็อตแบบค่อยเป็นค่อยไปมีอยู่ด้วยกันสองประเภทคือ Fade In หมายถึงภาพในช็อตแรกจะค่อยๆ มีดลงเรื่อยๆ จนกลายเป็นสีดำไปในที่สุดและ Fade Out หมายถึง ภาพในเฟรมแรกที่มีการเปลี่ยนช็อตจะเป็นสีดำและค่อยๆ ปรากฏภาพในเฟรมต่อไปชัดเจนเรื่อยๆ

(3) การเปลี่ยนช็อตแบบจางหายเป็นการเปลี่ยนช็อตแบบค่อยเป็นค่อยไปโดยภาพในเฟรมแรกจะเริ่มเปลี่ยนช็อตจะค่อยๆ จางหายไป ในขณะที่เดียวกันภาพในช็อตต่อไปก็จะค่อยๆ ปรากฏชัดขึ้นมาจนกลายเป็นภาพของช็อตต่อไปทั้งหมด

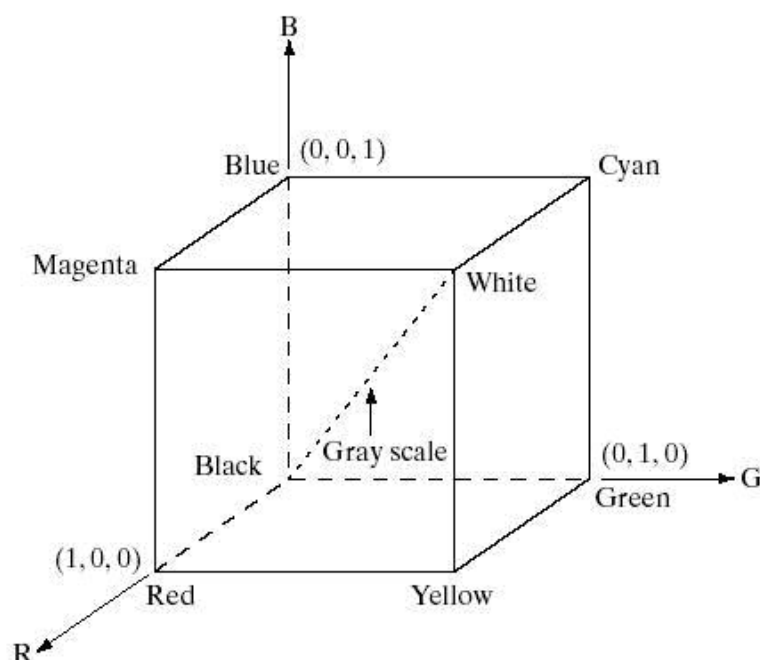
(4) การเปลี่ยนช็อตแบบกวาดเป็นลักษณะของการเปลี่ยนช็อตแบบค่อยเป็นค่อยไปอีกรูปแบบหนึ่ง โดยในช่วงที่มีการเปลี่ยนช็อตจะมีอีกช็อตหนึ่งวิ่งผ่านจากด้านหนึ่งของจอภาพไปสู่อีกด้านหนึ่งของจอภาพเข้ามาแทนที่ช็อตวิดีโอที่มีอยู่ก่อนหน้าจนกระทั่งกลายเป็นวิดีโอช็อตใหม่ทั้งหมด

2.3 แบบจำลองสี (Color Model)

2.3.1 แบบจำลองสี RGB (RGB Color Model)

แบบจำลองสี RGB เป็นแบบจำลองที่ใช้กันโดยทั่วไปใช้ในการแสดงผลบนจอมอนิเตอร์ (จิรา แก้วสุวรรณ, 2549) เนื่องจากการกำหนดความเข้มของลำอเล็กตรอนที่ยิงไปกระทบบนจอภาพประกอบด้วยประกอบด้วยค่าของแม่สีหลัก 3 สี ได้แก่ ค่าของความหนาแน่นของสีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) ดังภาพที่ 2.5 รูปภาพที่นำมาทำการประมวลผล เพื่อจัดเก็บลงฐานข้อมูลใช้ภาพดิจิทัลอลสีขนาด 24 บิตต่อจุดภาพ นั่นคือ แต่ละสีมีข้อมูลได้ 8 บิต ดังนั้นแต่ละจุดภาพสามารถแทนค่าสีได้ถึง 16,777,216 สี

ภาพที่ 2.5
แบบจำลองสี RGB



2.3.2 แบบจำลองสี HSV (HSV Color Model)

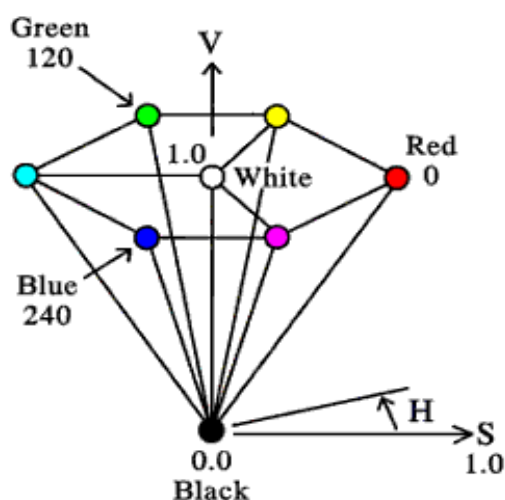
แบบจำลองสี HSV เป็นแบบจำลองที่ถูกสร้างขึ้นเพื่อเป็นทางเลือก (นิรุต ภูวิชาติ ,2549) แสดงในภาพที่ 2.6 จะให้ความหมายที่ดีกว่าเมื่อกล่าวถึงสีต่างๆในเชิงศิลปะเช่นเมื่อพูดถึงสีเหลืองในทางศิลปะจะมีความแตกต่างกัน เมื่อพิจารณาสีเหลืองอ่อน สีเหลืองแก่ หรือสีน้ำตาลว่ามีความแตกต่างกันอย่างไรจะพบว่าทุกสีคือสีเหลืองนั่นเองที่มีระดับความเข้มหรือมีความอิ่มตัวที่ต่างกัน ดังนั้นในแบบจำลองสี HSV จึงให้ความรู้สึกที่เข้าใจได้มากกว่าสำหรับมนุษย์ซึ่งแบบจำลองสี HSV ประกอบด้วยสามส่วนคือ

H หมายถึง Hue หรือสีที่มีค่าที่แตกต่างออกไปตามความถี่ของแสง เช่น แดง เหลือง เขียว น้ำเงิน หรือ ม่วง เป็นต้น

S หมายถึง Saturation หรือความอิ่มตัวของ Hue นั้น ๆ เช่น สีแดง และสีชมพูก็คือสีแดงเพียงแต่สีชมพูมีความอิ่มตัวน้อยกว่า

V หมายถึง Value หรือค่าความสว่างของสี โดยที่ค่า Value ต่ำสุดหมายถึง สีดำ ไม่ว่าจะมี Hue หรือ Saturation เท่าใดและค่า Value สูงสุดหมายถึงสีขาวซึ่งเป็นสีที่สว่างที่สุดของ Hue และ Saturation นั้นๆ

ภาพที่ 2.6
แบบจำลองสี HSV



2.4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้จึงมีเป้าหมายหลักในการศึกษาปัจจัยที่นำมาใช้ในการจำแนกอารมณ์กลัว (Fear) และอารมณ์อื่นๆ (Non-Fear) คือ อารมณ์ตื่นเต้น (Excitement) อารมณ์สนุกสนาน (Joy) อารมณ์เศร้า (Sadness) โดยแบ่งประเภทตามอารมณ์ของผู้ชมภาพยนตร์เป็นเกณฑ์ งานวิจัยเกี่ยวกับการจำแนกอารมณ์กลัวนั้นจำเป็นต้องอาศัยคุณลักษณะที่เหมาะสม ทั้งนี้งานวิจัยจึงรวบรวมแนวคิด และคุณลักษณะที่สำคัญในการจำแนกเพื่อเป็นแนวทางในการทำวิจัยโดยจะกล่าวถึงรายละเอียดต่อไปในส่วนนี้

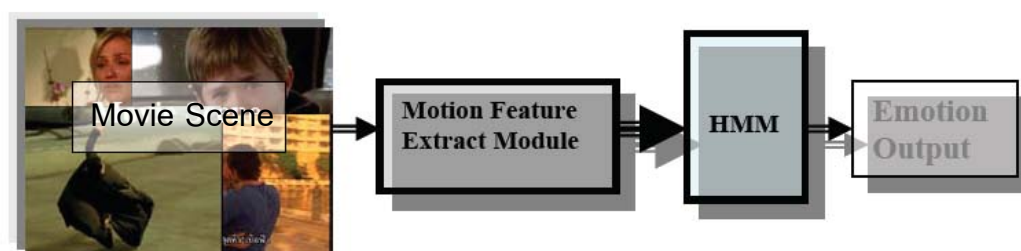
2.4.1 กลุ่มงานวิจัยด้านการจำแนกอารมณ์จากภาพ

งานวิจัยระบบรู้จำอารมณ์จากภาพโดยใช้ตัวอย่างจากภาพยนตร์ (Prasert Lorwinitnan, 2003) ใช้องค์ประกอบของการเคลื่อนไหวของภาพ (Motion Vector) สี่องค์ประกอบ

คือ ปริมาณ (Volume) ทิศทาง (Director) ความเร็ว (Speed) และพลังงาน (Energy) โดยใช้หลักการจำการเปลี่ยนแปลงค่าของข้อมูลที่ซับซ้อนของการเคลื่อนไหวภาพในภาพยนตร์เพื่อจำแนกกลุ่มของอารมณ์ตื่นเต้น (Exciting) อารมณ์สนุกสนาน (Joy) อารมณ์รัก (Romance) และอารมณ์เศร้า (Sadness) โดยใช้โมเดลฮิดเดนมาร์คอฟ (Hidden Markov Model) ดังภาพที่ 2.7 ซึ่งโมเดลนี้เป็นที่รู้จักแพร่หลาย และประสบผลสำเร็จในการใช้กับระบบรู้จำเสียงและลายมือ ผลสรุปการจำแนกวิธีดังกล่าวไม่สามารถใช้องค์ประกอบของทิศทางการเคลื่อนไหวได้แต่สามารถใช้องค์ประกอบ ปริมาณ ความเร็ว และพลังงาน โดยผลการทดลองจำแนกกลุ่มอารมณ์ตื่นเต้นได้ร้อยละ 88 และกลุ่มอารมณ์เศร้าร้อยละ 85 และจากวิธีดังกล่าวไม่สามารถแยกตัวอย่างภาพยนตร์ในกลุ่มของอารมณ์สนุกสนานและอารมณ์รักได้

ภาพที่ 2.7

โมเดลของระบบงานวิจัยรู้จำอารมณ์จากภาพ

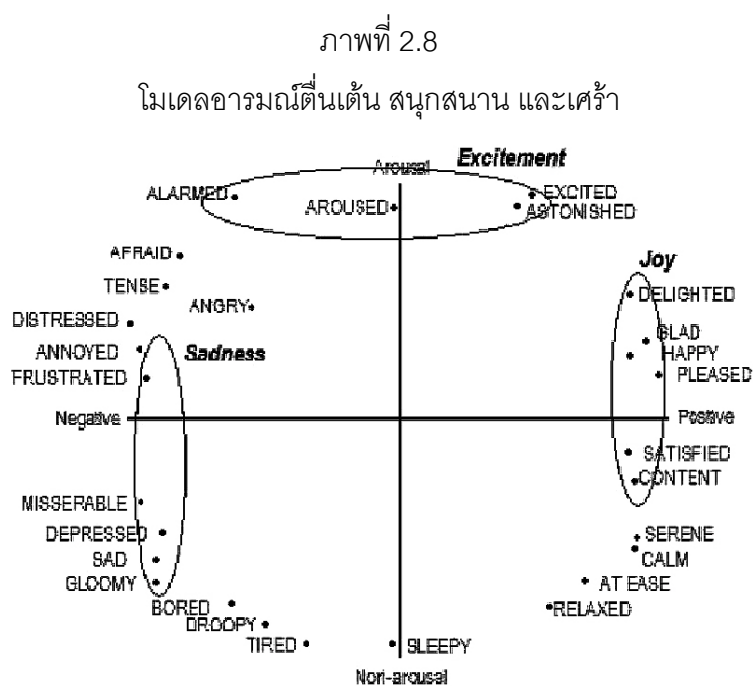


2.4.2 กลุ่มงานวิจัยด้านการจำแนกอารมณ์จากเสียง

งานวิจัยการเรียนรู้และแยกแยะระดับอารมณ์ในเสียงพูดโดยใช้ตัวอย่างเสียงจากภาพยนตร์ (Wanchai Loturatana ,2005) คุณลักษณะที่ใช้ในงานวิจัยคือความถี่พื้นฐาน (Pitch parameter) ในอังกถึงจุดสูงสุด (maximum pitch) จุดต่ำสุด (minimum pitch) ของรูปร่างความถี่พื้นฐาน เพื่อทำการศึกษาลักษณะของกลุ่มอารมณ์โกรธ (anger) อารมณ์สนุกสนาน (happiness) อารมณ์เศร้า (sadness) และอารมณ์ปกติ (unemotional) โดยวิธีในงานวิจัยนี้ใช้สามารถแบ่งแยกหรือจำแนกลักษณะของอารมณ์ได้ แต่จะไม่สามารถแยกเสียงที่มีหลายอารมณ์ได้ ผลการทดลองจำแนกกลุ่มอารมณ์โกรธร้อยละ 72.97 อารมณ์สนุกสนานร้อยละ 88 อารมณ์เศร้าร้อยละ 77.78 และอารมณ์ปกติร้อยละ 73.63

2.4.3 กลุ่มงานวิจัยด้านการจำแนกอารมณ์จากภาพ และเสียง

งานวิจัยการจำแนกคลิปภาพยนตร์โดยใช้ปัจจัยทางอารมณ์ (Saowaluk, 2008) ใช้ค่าของแสง (Lighting) สี (Color) เวลาการเปลี่ยนช็อต (pace) ความดังของเสียง (Loudness) มาเป็นเกณฑ์ในการจำแนกกลุ่มของอารมณ์ตื่นเต้น (Excitement) อารมณ์สนุกสนาน (Joy) และอารมณ์เศร้า (sadness) ดังภาพที่ 2.8



การจำแนกอารมณ์ในงานวิจัยนี้ใช้การหาคุณลักษณะของภาพและเสียง โดยสรุปได้จากรูป 2.9 ดังนี้

ภาพที่ 2.9

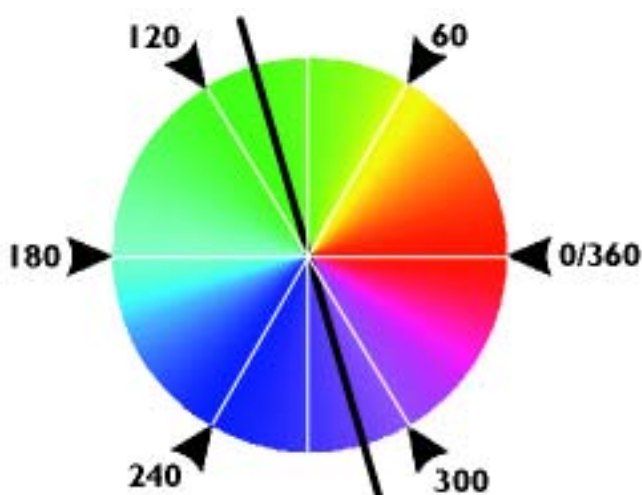
คุณลักษณะที่สำคัญของงานวิจัยจำแนกสามอารมณ์

		Perceived Emotion:		
		Excitement	Joy	Sadness
Film Element:	Lighting Scheme	-	High-Key	Low-Key
	Color Scheme	-	Warm	Cold
	Movement	Great	Semi-Great	Little
	Pace	Fast	Semi-Fast	Slow
	Changes in Sound Loudness	Great	Great	Little

- Lighting Scheme กำหนดให้เป็นสองค่าคือ ค่าเปอร์เซ็นต์เฉลี่ยของแสงสว่างในคลิป (APB : The average percent of pixels in the clip with high brightness) และค่าเปอร์เซ็นต์เฉลี่ยของแสงมืดในคลิป (APD: The average percent of pixels in the clip with dark value) หากค่า APB มากกว่า 50 % จัดอยู่ในอารมณ์สนุกสนาน และค่า APD น้อยกว่า 20 % จัดอยู่ในอารมณ์เศร้า

- Color Scheme เป็นการแบ่งโทนสีระหว่างสีโทนร้อน (Warm) คือ เหลือง เหลืองส้ม ส้ม แดงส้ม แดง และแดงม่วง และ สีโทนเย็น (Cold) คือ เหลืองเขียว เขียว น้ำเงินเขียวน้ำเงิน น้ำเงิน ม่วง และม่วง ดังภาพที่ 2.10

ภาพที่ 2.10
การแบ่งสีโทนร้อนและสีโทนเย็น



การแบ่งโทนสีช่วยในการจำแนกอารมณ์ระหว่างสนุกสนาน และอารมณ์เศร้า โดยหาจากค่าเฉลี่ยสีของคลิป (APC : Average percent of pixels in the clip with cold hue)

- Movement ค่าเฉลี่ยการเคลื่อนไหวในคลิป (AMM : Average squared motion magnitude) ค่าที่ได้เป็นค่าจากการคำนวณของการเคลื่อนไหวของภาพ (Motion Vector) โดยจะแบ่งเฟรมเป็นบล็อกขนาด 16x16 pixels ยกตัวอย่างกลุ่มอารมณ์ตื่นเต้นจะมีค่าเฉลี่ยมากกว่ากลุ่มอารมณ์เศร้า

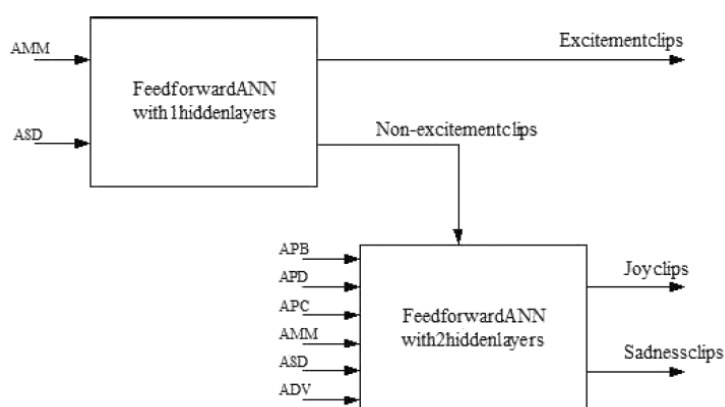
- Pace ค่าเฉลี่ยความยาวของช็อตในคลิป (ASD : Average shot duration) โดยค่าเฉลี่ยจะแสดงเวลาในการเปลี่ยนช็อต ยกตัวอย่างกลุ่มอารมณ์ตื่นเต้นจะมีค่าเฉลี่ยน้อยกว่ากลุ่มอารมณ์เศร้า เนื่องจากกลุ่มอารมณ์ตื่นเต้นจะใช้เวลาในการเปลี่ยนช็อตเร็วกว่า

- Changes in Sound Loudness ค่าเฉลี่ยความแตกต่างของเสียง (AVD : Average volume difference) เสียงที่นำมาวิเคราะห์เป็นเสียงช่วงใดช่วงหนึ่งในคลิปหลังจากนั้นนำมาแบ่งเป็นเฟรมมีความยาวประมาณ 10-40 ms เพื่อนำมาวิเคราะห์หาความแตกต่างของเสียงจากเฟรมที่ติดกันยกตัวอย่างกลุ่มอารมณ์ตื่นเต้นจะมีค่าเฉลี่ยมากกว่ากลุ่มอารมณ์เศร้า

หลังจากนั้นนำคุณลักษณะมาจำแนกโดยใช้โครงข่ายประสาทเทียมตามโมเดลดังภาพที่ 2.11

ภาพที่ 2.11

การจำแนกโดยใช้โครงข่ายประสาทเทียมของสามกลุ่มอารมณ์



ผลการทดลองจำแนกอารมณ์สนุกสนานได้ค่าความถูกต้อง 93 เปอร์เซ็นต์ ส่วนอารมณ์ตื่นเต้นและอารมณ์เศร้าจำแนกได้ค่าความถูกต้อง 100 เปอร์เซ็นต์