

# การประมวลผลเอกสารภาษาไทยเป็นทรีโครงสร้างของเอกสาร

IMAGE EXTRACTING PROCESS TO IDENTIFY THAI DOCUMENT STRUCTURE



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของงานวิจัยที่สนับสนุนโดยสำนักงานเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
สาขาวิชาวิศวกรรมศาสตร์และเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2545

ISBN 974-949-243-2

การประมวลผลเอกสารภาษาไทยเพื่อหาโครงสร้างของเอกสาร

IMAGE EXTRACTING PROCESS TO IDENTIFY THAI DOCUMENT  
STRUCTURE



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

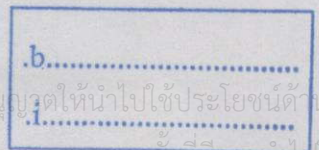
บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2544

40007

ISBN 974-648-243-2



พิมพ์.....  
ทะเบียน.....  
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาติให้นำไปใช้ประโยชน์ด้านการค้า  
เดือน, ปี 20 ก.ค. 2544

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรณีไปใช้

**IMAGE EXTRACTING PROCESS TO IDENTIFY THAI DOCUMENT  
STRUCTURE**



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENT FOR THE DEGREE OF  
MASTER OF SCIENCE IN COMPUTER SCIENCE AND INFORMATION TECHNOLOGY  
SCHOOL OF GRADUATE STUDIES  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**2001**

**ISBN 974-648-243-2**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**COPYRIGHT 2001**

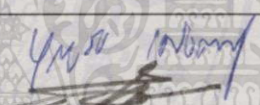
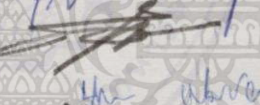


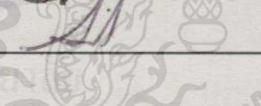
**SCHOOL OF GRADUATE STUDIES**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บัณฑิตวิทยาลัย  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การประมวลผลเอกสารภาษาไทย เพื่อหาโครงสร้างของเอกสาร  
IMAGE EXTRACTING PROCESS TO IDENTIFY THAI DOCUMENT  
STRUCTURE  
ชื่อนักศึกษา นายพิเศษพงศ์ สุธาพันธ์  
รหัสประจำตัว 37064422  
ปริญญา วิทยาศาสตรมหาบัณฑิต  
สาขาวิชา วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ  
อาจารย์ผู้ควบคุมวิทยานิพนธ์ รศ.ดร.บุญธีร์ เครือตราชู

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
รศ.ดร.บุญธีร์	เครือตราชู	
รศ.ดร.วิเชียร	เปรมชัยสวัสดิ์	
รศ.ดร.นุชรี	เปรมชัยสวัสดิ์	
ดร.ประจวบ	วานิชชัชวาล	
ดร.อาริต	ธรรมโน	

วัน/เดือน/ปี ที่สอบ 2 พฤษภาคม 2544 เวลา 9.00 น. เป็นต้นไป  
สถานที่สอบ ห้อง LAB 316 ชั้น 3 คณะเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัยรับรองแล้ว  
  
(รศ.ดร.บุญวัฒน์ อัทธนะ)  
คณบดีบัณฑิตวิทยาลัย

วันที่ 31 เดือน พฤษภาคม พ.ศ. 2544

หัวข้อวิทยานิพนธ์	การประมวลผลเอกสารภาษาไทยเพื่อหาโครงสร้างของเอกสาร
นักศึกษา	นาย พิเศษพงศ์ สุธาพันธ์
รหัสประจำตัว	37064422
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
พ.ศ.	2544
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ. ดร. บุญธีร์ เครือตราฐ

### บทคัดย่อ

การแบ่งส่วนประกอบภาพเอกสารไบนารี เป็นขั้นตอนการทำงานเพื่อใช้ในการวิเคราะห์โครงสร้างของเอกสารที่ประกอบด้วยคอลัมน์ ซึ่งอาจเป็นข้อความหรือรูปภาพ และสามารถกำหนดขอบเขตข้อความแบบอัตโนมัติไปใช้ในระบบ OCR แทนวิธีเดิมที่ใช้มนุษย์ ซึ่งวิทยานิพนธ์ฉบับนี้ได้เสนอวิธีการการแบ่งส่วนประกอบเอกสารที่รวดเร็วและมีประสิทธิภาพ โดยใช้แนวคิดผสมผสานระหว่างการแบ่งส่วนประกอบใหญ่ไปสู่ส่วนประกอบย่อย(top-down) และส่วนประกอบย่อยไปสู่ส่วนประกอบใหญ่(bottom-up) โดยการแบ่งส่วนประกอบใหญ่(คอลัมน์)ของหน้าเอกสารออกจากกัน ใช้การเดินรอบขอบ(contour following) ซึ่งใช้กรอบหน้าต่างขนาด  $32 * 32$  พิกเซล เพื่อที่จะสามารถเดินรอบคอลัมน์ที่มีตัวอักษรห่างกันน้อยกว่า 32 จุด ในภาพเอกสารไบนารีที่ความละเอียด 300 จุดต่อนิ้วได้ สำหรับขั้นตอนแรกเป็นการหาขอบเขตของส่วนประกอบใหญ่ โดยใช้หน้าต่างขนาด  $32*32$  พิกเซลวนรอบขอบส่วนประกอบ เพื่อหาโครงร่างคร่าวๆ ส่วนขั้นตอนที่สองจะทำการเดินรอบขอบส่วนประกอบย่อย(ตัวอักษร) โดยใช้หน้าต่างขนาด  $1*1$  พิกเซลหาขอบเขตย่อยเพื่อนำไปใช้วิเคราะห์ในขั้นตอนถัดไป ขั้นตอนสุดท้ายคือการวิเคราะห์ส่วนประกอบ ซึ่งเป็นขั้นตอนที่จัดการแบ่งขอบเขตเพื่อแยกคอลัมน์ที่อยู่ในส่วนประกอบใหญ่แยกออกจากกัน และแบ่งประเภทส่วนประกอบที่เป็นข้อความ เพื่อนำไปใช้เป็นข้อมูลในระบบ OCR ต่อไป จุดเด่นของวิธีการนี้คือไม่จำเป็นต้องตรวจสอบทุกจุดในภาพเอกสาร ไบนารีจึงทำให้มีความเร็วกว่าวิธีที่ต้องตรวจทุกจุด

<b>Thesis Title</b>	Image Extracting Process to Identify Thai Document Structure
<b>Student</b>	Mr. Phisetphong Suthaphan
<b>Student ID.</b>	37064422
<b>Degree</b>	Master of Science
<b>Programme</b>	Computer Science and Information Technology
<b>Year</b>	2001
<b>Thesis Advisor</b>	Assoc. Prof. Dr. Boontee Kruatrachue

## ABSTRACT

Partitioning of binary documents image is one of the document structure analysis processes, which may be text or images. Also, It could define the text boundary automatically for using in OCR systems. This thesis proposes a quick and efficient document-partitioning algorithm. By applying both top-down and bottom-up, to split a large component (column) in the document apart. The contour following method, is used window size of 32\*32 pixel for contour following column that has text space less than 32 points with apart (in resolution is 300 pixels/inch). The first process is to find the main components' boundary by using the 32\*32 pixel window over the border, for determining the draft structure of document. Secondly, the contours border of sub-components, or letters with 1\*1 pixel-sized window to separate individually letter. Finally, the step of managing the letter boundaries in orders to separate the columns in the main component and then classify types of paragraph. The benefit of this method is that it is not necessary to access every pixel in document images. So the process of segmentation goes through more speed in faster.

## กิตติกรรมประกาศ

การจัดทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี เพราะได้รับคำปรึกษาจากท่านอาจารย์ที่ปรึกษา ร.ศ. ดร. บุญธีร์ เครือตราชู ซึ่งได้ให้คำปรึกษาในการแก้ไขปัญหาต่างๆ และชี้แนวทางในการทำวิจัยตลอดมา ผู้ทำวิจัยรู้สึกซาบซึ้งในความกรุณาจากท่าน ขอกราบขอบคุณเป็นอย่างสูง

ขอกราบขอบคุณอาจารย์ในสถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง ทุกท่านที่ได้สั่งสอนวิชาความรู้ต่างๆ ซึ่งเป็นรากฐานที่ก่อให้เกิดงานวิจัยนี้สำเร็จ รวมทั้งขอขอบคุณคณะเทคโนโลยีสารสนเทศที่สนับสนุนเครื่องมือ หนังสือและอุปกรณ์ต่างๆที่ใช้ในการทดลอง รวมถึงขอบคุณเจ้าหน้าที่ทุกคนที่คอยอำนวยความสะดวกต่างๆ ในระหว่างการศึกษา

กราบขอบคุณพ่อแม่ที่คอยห่วงใยและเป็นกำลังใจที่สนับสนุนด้านปัจจัยต่างๆ ขอบใจน้องและคนรู้จักที่ช่วยเป็นแรงกระตุ้นให้ทำงานทดลองนี้ตลอดมา ขอขอบใจเพื่อนๆ รุ่นเดียวกัน รุ่นพี่ และรุ่นน้องที่คอยช่วยเหลือเกื้อกูลกัน

สุดท้ายนี้ขอขอบคุณทุกท่านที่ไม่ได้กล่าวถึงแต่ต้น ที่เคยช่วยเหลือกันเกื้อกูลกัน

พิเศษพงศ์ สุธาพันธ์

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	1
1.3 สมมุติฐานของการศึกษา.....	2
1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย.....	2
1.5 ขอบเขตของการศึกษา.....	3
1.6 ขั้นตอนการศึกษา.....	3
1.7 ข้อจำกัดของการศึกษา.....	3
1.8 รายละเอียดเครื่องคอมพิวเตอร์ เครื่องมือและอุปกรณ์ที่ใช้ในการทำวิจัย.....	3
1.9 รายละเอียดในแต่ละบท.....	4
บทที่ 2 ทฤษฎีและแนวคิดที่เกี่ยวข้อง.....	5
2.1 การแสดงภาพดิจิทัล.....	5
2.2 ขั้นตอนพื้นฐานในการประมวลผลภาพ.....	5
2.3 ความรู้พื้นฐานเกี่ยวกับข้อมูลภาพอิเล็กทรอนิกส์.....	7
2.4 รหัสแนว (Chain Code).....	8
2.5 คุณลักษณะพื้นฐานของภาพเอกสาร.....	8
2.6 ประเภทของเอกสารและการประยุกต์ใช้.....	9
2.7 ขั้นตอนการแบ่งหน้าเอกสารและวิธีการแบ่งประเภทขอบเขต.....	15
2.8 การแบ่งประเภทขอบเขตเอกสาร.....	21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

	หน้า
2.9 เปรียบเทียบวิธีการทำงาน .....	23
2.10 วิเคราะห์วิธีการทำงาน.....	26
2.11 สรุป.....	29
<b>บทที่ 3 แนวคิดและการนำมาประยุกต์ใช้ในการแบ่งส่วนประกอบภาพเอกสารใบนารี .....</b>	<b>32</b>
3.1 แนวคิดในการแบ่งส่วนประกอบภาพเอกสารใบนารี .....	32
3.2 การแบ่งส่วนประกอบภาพเอกสาร เป็นส่วนประกอบใหญ่.....	34
3.3 การแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ .....	36
3.4 การแบ่งคอลัมน์และปรับปรุงขอบเขตส่วนประกอบใหญ่.....	38
<b>บทที่ 4 อัลกอริทึมและขั้นตอนการทำงาน .....</b>	<b>40</b>
4.1 หลักการที่ใช้ในการทดลอง .....	40
4.2 วิธีการและขั้นตอนการทดลอง.....	49
4.3 ผลการทดลอง.....	49
4.4 วิเคราะห์ผลการทดลอง.....	60
4.5 สรุปผลการทดลอง .....	61
<b>บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ.....</b>	<b>62</b>
5.1 สรุปผลการทดลอง .....	62
5.2 ข้อเสนอแนะ.....	62
5.3 แนวทางการดำเนินการ .....	63
<b>เอกสารอ้างอิง.....</b>	<b>64</b>
<b>ภาคผนวก ก บทความที่ดีพิมพ์ในวารสาร .....</b>	<b>70</b>
<b>ภาคผนวก ข ภาพเอกสารที่ใช้ในการทดลอง .....</b>	<b>82</b>
<b>ประวัติผู้เขียน .....</b>	<b>88</b>

# สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงคุณสมบัติที่สำคัญของวิธีการต่างๆ ที่ใช้ในการวิเคราะห์แผนผังเอกสาร .....	24
2.2 แสดงเวลาที่ใช้ในการทำงานสำหรับวิธีการวิเคราะห์เอกสาร.....	25
2.3 แสดงผลงานวิจัยที่ผ่านมา .....	30
4.1 อ้างอิงตัวอย่างที่นำมาทดลองจากภาคผนวก .....	49
4.2 แสดงผลการทดลอง .....	59

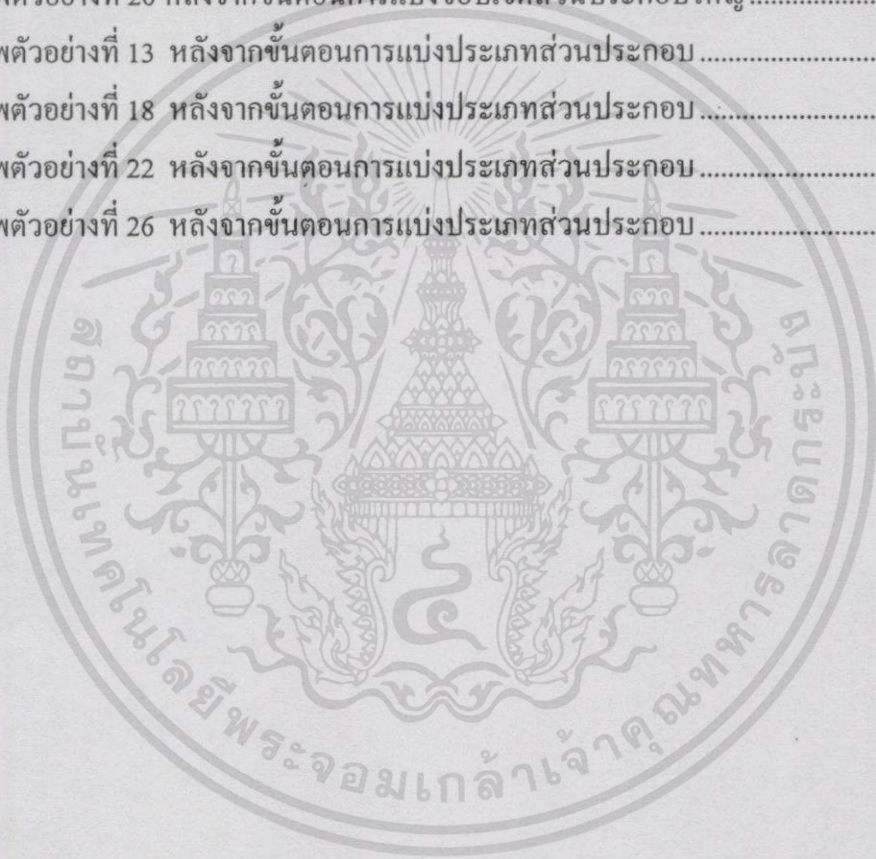


# สารบัญรูป

รูปที่	หน้า
2.1 แนวแกน X, Y ที่ใช้อย่างอิงพิกัด.....	5
2.2 ขั้นตอนพื้นฐานของกระบวนการประมวลผลภาพดิจิทัล .....	6
2.3 ทิศทางของรหัสแนว .....	8
2.4 ตัวอย่างภาพเอกสารชนิดต่างๆ.....	10
2.5 ชุดจุดภาพก่อนทำการรีนแรงค์สมูท .....	21
2.6 ชุดจุดภาพหลังจากทำการรีนแรงค์สมูท .....	22
2.7 การฉายเงาจุดดำในแนวแกนตั้ง และช่วงที่แบ่ง.....	22
2.8 กราฟที่ใช้แสดงการจัดเรียงไวยากรณ์ .....	23
3.1 ตัวอย่างแผนผังแบบต่างๆ .....	32
3.2 การเชื่อมต่อตัวอักษร.....	33
3.3 ขอบเขตภาพ.....	34
3.4 ตำแหน่งหน้าต่างเริ่มต้นและลักษณะการวนรอบส่วนประกอบใหญ่.....	35
3.5 ภาพส่วนประกอบที่ไม่อยู่ในลักษณะสี่เหลี่ยม .....	35
3.6 ขอบเขตของส่วนประกอบในเอกสารที่แบ่งโดยหน้าต่างขนาด 32*32 จุด.....	36
3.7 ส่วนประกอบใหญ่ที่มีรายละเอียดที่เป็นทั้งข้อมูลและภาพ .....	37
3.8 ส่วนประกอบใหญ่ที่มีรายละเอียดเป็นข้อความเพียงอย่างเดียว.....	37
3.9 การนำขอบเขตภาพมาหักออกจากขอบเขตส่วนประกอบใหญ่.....	38
3.10 ภาพจำลองในการนำกลุ่มบรรทัดมาตัดแนวแกน Y เพื่อแบ่งส่วนคอลัมน์.....	39
4.1 แผนผังขั้นตอนการทำงาน.....	40
4.2 แผนผังขั้นตอนการทำงานของการแบ่งส่วนประกอบใหญ่.....	42
4.3 โครงสร้างข้อมูลที่เก็บขอบเขตส่วนประกอบใหญ่ .....	43
4.4 ตัวอย่างข้อมูลขอบเขตของส่วนประกอบที่เป็นตัวอักษรในลักษณะ การจัดเรียงตามขอบซ้าย.....	44
4.5 แผนผังการทำงานการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่.....	45
4.6 การจัดข้อมูลที่แยกประเภทเป็นข้อความภาพและสัญญาณรบกวน .....	46
4.7 ขอบเขตส่วนประกอบใหญ่บางส่วนที่ครอบคลุมทั้งข้อความและรูปภาพ .....	47
4.8 ขอบเขตส่วนประกอบใหญ่บางส่วนที่นำขอบเขตภาพออกไป .....	47

## สารบัญรูป(ต่อ)

รูปที่	หน้า
4.9 แผนผังการทำงานแบ่งขอบเขตส่วนประกอบ และแบ่งประเภทส่วนประกอบ.....	48
4.10 ภาพตัวอย่างที่ 13 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่.....	50
4.11 ภาพตัวอย่างที่ 18 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่.....	51
4.12 ภาพตัวอย่างที่ 22 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่.....	52
4.13 ภาพตัวอย่างที่ 26 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่.....	53
4.14 ภาพตัวอย่างที่ 13 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ.....	55
4.15 ภาพตัวอย่างที่ 18 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ.....	56
4.16 ภาพตัวอย่างที่ 22 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ.....	57
4.17 ภาพตัวอย่างที่ 26 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ.....	58



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในยุคเริ่มแรกที่เครื่องคอมพิวเตอร์เข้ามามีบทบาทต่อการทำงานในกิจการต่างๆเพิ่มขึ้น ทางด้านการศึกษา และการวิจัย โดยเริ่มจากความต้องการที่จะเก็บบันทึกข้อมูลที่เป็นข้อความ ซึ่งต้องนำข้อมูลเข้าทางแป้นพิมพ์ จากนั้นก็พัฒนามาเป็นการบันทึกข้อมูลเป็นภาพ โดยใช้เครื่องสแกนรูปภาพ(scanner)เป็นตัวนำข้อมูลเข้า ซึ่งข้อมูลภาพที่จัดเก็บนั้นมีทั้งเอกสาร และรูปภาพ ซึ่งเอกสารนั้นเมื่อสแกนเข้าไปในลักษณะของภาพจะไม่สามารถนำมาแก้ไขหรือตัดแปลงได้อย่างสะดวก เหมือนกับการบันทึกเข้าไปเป็นตัวอักษร และด้วยความเจริญก้าวหน้าทางคมนาคมและการสื่อสาร ทำให้การรับรู้และการใช้ข้อมูลข่าวสารมีเพิ่มมากขึ้น ซึ่งทำให้ปริมาณงานในการจัดเก็บและค้นหาเอกสารมีมากขึ้น จึงมีความต้องการวิธีการจัดเก็บเอกสารที่รวดเร็ว ซึ่งทำให้เกิดแนวคิดที่ต้องการให้มีความสะดวกในการนำเข้าสู่ข้อมูลเอกสาร โดยเก็บเอกสารในลักษณะของภาพ แต่การเก็บเอกสารในรูปแบบนี้มีข้อด้อยหลายประการ ข้อแรกคือการจัดเก็บเอกสารในลักษณะภาพมีขนาดใหญ่กว่าการจัดเก็บเอกสารที่เป็นข้อความมาก ข้อสองเอกสารที่เก็บในลักษณะภาพไม่สามารถนำมาแก้ไขเปลี่ยนแปลงได้อย่างเอกสารที่เป็นตัวอักษร และข้อสุดท้ายเอกสารที่จัดเก็บเป็นรูปภาพ ไม่สามารถนำไปใช้ในการค้นหาเนื้อหาสาระข้างในเอกสารเป็นอย่างไร จากทั้งสามกรณีข้างต้น กระตุ้นให้เกิดความต้องการที่จะกำจัดข้อด้อยเหล่านั้น จึงมีการพัฒนา ระบบจดจำตัวอักษร(OCR : Optical Character Recognition) ในตอนแรกนั้นยังกำหนดขอบเขตที่จะให้ทำงานตามที่ต้องการด้วยมืออยู่ เพราะเอกสารบางอย่างมีหลายคอลัมน์ หรือ มีส่วนประกอบที่ซับซ้อนซึ่งระบบจดจำตัวอักษร ไม่สามารถแบ่งแยกได้ว่ารายละเอียดพวกนี้จะเป็นกลุ่มเดียวกันหรือไม่ ซึ่งทำให้เกิดความยุ่งยากและล่าช้าในงานที่ต้องจัดเก็บเอกสารจำนวนมาก จึงมีความต้องการกำหนดขอบเขตที่จะนำไปใช้ในระบบจดจำตัวอักษรแบบอัตโนมัติขึ้นมา เพื่อให้เกิดความสะดวกและทำงานได้รวดเร็วยิ่งขึ้น ทำให้เกิดงานวิจัยทางการแบ่งขอบเขตเอกสารภาพขึ้นอย่างมาก

### 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

ความมุ่งหมายและวัตถุประสงค์ของการศึกษางานวิทยานิพนธ์นี้ เพื่อพัฒนาวิธีการแบ่งส่วนประกอบภาพเอกสาร ไปนารีแบบอัตโนมัติที่รวดเร็วและถูกต้อง เพื่อให้ได้ขอบเขตส่วนประกอบที่จะนำไปใช้กับระบบจัดเก็บเอกสารและจดจำตัวอักษร เพื่อให้ระบบทำงานได้มีประสิทธิภาพยิ่งขึ้น โดยงานวิจัยนี้จะนำภาพเอกสารไปนารีมาแยกส่วนประกอบเอกสารซึ่งมีรายละเอียดเอกสารเป็นเอกสารที่ส่งวนเวสสำหรับบริการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลายชนิด โดยเน้นความสำคัญของส่วนประกอบที่เป็นประเภทตัวอักษร และสามารถกำหนดขอบเขตของส่วนประกอบที่มีลักษณะไม่เป็นสี่เหลี่ยมได้ ซึ่งสามารถทำให้ได้เฉพาะส่วนที่ต้องการเท่านั้น และเพิ่มความสามารถของระบบจัดเก็บภาพเอกสารให้ดียิ่งขึ้น

### 1.3 สมมุติฐานของการศึกษา

ในงานวิจัยนี้จะนำเสนอวิธีการแบ่งส่วนประกอบของเอกสารและผลการทดลองที่ได้ทำกับเอกสารหลายรูปแบบ เช่น วารสารทางวิชาการ นิตยสาร โดยผลลัพธ์ที่ได้จะเป็นขอบเขตของส่วนประกอบต่างๆภายในเอกสารที่ถูกแบ่งประเภทเป็นข้อความหรือรูปภาพได้อย่างถูกต้อง ใช้เวลาในการทำงานน้อย และสามารถนำไปประยุกต์ใช้งานกับระบบงานการจัดเก็บและค้นหาเอกสารได้จริง

### 1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

ภาพเอกสารที่นำมาใช้ในการทดลองนี้ เป็นภาพเอกสารไบนารีที่มีการจัดเก็บแบบบิตแมป โดยการจัดเก็บข้อมูลให้อยู่ในรูปแบบลำดับของพิกเซล ซึ่งค่าของข้อมูลในแต่ละพิกเซลจะมีค่าเป็น 0 หรือ 1 เท่านั้น (ขาวหรือดำ)

การแบ่งส่วนประกอบและจัดประเภทส่วนประกอบของภาพเอกสารนั้น ขึ้นแรกต้องกำหนดขอบเขตส่วนประกอบชนิดที่เป็นลักษณะเดียวกันอยู่ในขอบเขตเดียวกัน ซึ่งมีวิธีการทั้งแบบบนลงล่าง(top-down) เช่น วิธีการตัดซ้ำแนวแกน X-Y (recursive projection X-Y cut) ส่วนวิธีการแบบล่างขึ้นบน (bottom-up) เช่น วิธีการรวมส่วนประกอบเล็กที่มีลักษณะที่สัมพันธ์กันมาเป็นส่วนประกอบที่ใหญ่ขึ้น(connected components) ซึ่งในงานวิจัยนี้ใช้หน้าตาขนาด 32\*32 จุดในการกำหนดขอบเขตส่วนประกอบที่อยู่ติดกันเข้าไว้ด้วยกัน ส่วนการแบ่งประเภทของส่วนประกอบนั้นมีการใช้กฎ (rule) ในการแบ่งประเภทส่วนประกอบ เช่น ลักษณะของส่วนประกอบความสูงต่อความกว้างของส่วนประกอบ อัตราส่วนความหนาแน่นของจุดภาพ เป็นต้น

แนวทางที่คิด คือ หาวิธีการที่สามารถแบ่งขอบเขตส่วนประกอบภาพเอกสารไบนารี ที่มีลักษณะโครงสร้างรายละเอียดซับซ้อน และมีประสิทธิภาพในการทำงานที่รวดเร็ว จึงมีความคิดที่จะใช้วิธีการเดินรอบขอบในการกำหนดส่วนประกอบเอกสาร และมีการเดินรอบขอบส่วนประกอบย่อยเพื่อนำมาวิเคราะห์เพื่อแก้ไขแบ่งส่วนประกอบที่ผิดพลาดในขั้นตอนแรก โดยใช้ขอบเขตของกลุ่มที่เป็นตัวอักษรมาทำการจัดเรียงเป็นกลุ่มเพื่อแบ่งคอลัมน์ที่อยู่ในส่วนประกอบใหญ่เดียวกันออกจากกัน

## 1.5 ขอบเขตของการศึกษา

สำหรับขอบเขตในการทำวิทยานิพนธ์นี้ คือ การพัฒนาและประยุกต์ใช้แนวความคิดที่ใช้ในการแบ่งส่วนประกอบภาพเอกสารไบนารี เพื่อที่จะนำไปใช้ต่อในระบบจดจำตัวอักษร เพื่อที่จะส่งเสริมให้ระบบจัดเก็บเอกสารทำได้ดียิ่งขึ้น สามารถที่จะทำให้ระบบจัดเก็บเอกสารจัดเก็บเอกสารในรูปแบบเอกสารที่หลากหลายและซับซ้อนยิ่งขึ้น

## 1.6 ขั้นตอนการศึกษา

1. ศึกษาวิธีการแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบภาพเอกสาร เพื่อใช้ในการกำหนดหัวข้อ เป้าหมาย วัตถุประสงค์ และขอบเขตของการทำวิทยานิพนธ์
2. ทำการทดลอง นำวิธีการที่ศึกษามาลองทำ เพื่อหาแนวทางใหม่ในการแบ่งส่วนประกอบและวิเคราะห์ส่วนประกอบภาพเอกสาร
3. ทดลองตามแนวทางใหม่ที่วางไว้ เพื่อพิสูจน์แนวคิดว่าได้ผลตามที่ต้องการหรือไม่
4. จัดทำเอกสารประกอบวิทยานิพนธ์

## 1.7 ข้อจำกัดของการศึกษา

ภาพเอกสารที่ใช้ในการทดลอง เป็นรูปแบบไบนารี ที่สแกนด้วยความละเอียด 300 จุดต่อนิ้ว โดยที่ภาพเอกสารไม่เอียง และภาพเอกสารต้องไม่มีกรอบ ตัวอย่างภาพเอกสารที่ใช้เป็นวารสารเชิงวิชาการ หนังสือ นิตยสาร ที่มีตัวอักษรขนาดไม่เกิน 100 พิกเซล และในการพัฒนาโปรแกรมที่พัฒนาขึ้นมาเพื่อใช้ในการทดลอง โครงสร้างของการเขียนโปรแกรมมีความซับซ้อน (โครงสร้างข้อมูลแบบลิงก์ลิสต์) ซึ่งอาจเกิดความผิดพลาดเนื่องจากการจัดการกับหน่วยความจำที่ไม่สมบูรณ์ (จองหน่วยความจำแล้วคืนหน่วยความจำให้แก่ระบบไม่หมด) อาจทำให้การจับเวลาที่ได้จากการทำงานไม่เที่ยงตรง

## 1.8 รายละเอียดเครื่องคอมพิวเตอร์และเครื่องมือและอุปกรณ์ที่ใช้ในการทำวิจัย

### - เครื่องคอมพิวเตอร์

#### คุณสมบัติทางฮาร์ดแวร์

CPU	:	Pentium II 350 MHz
RAM	:	64 MB
HARD DISK	:	4 GB

#### ระบบปฏิบัติการ (Operating System)

Windows 98 SE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- อุปกรณ์สแกนภาพ (Scanner)  
HP Scanjet
- เครื่องมือที่พัฒนา  
Microsoft VC++ 5.0  
Photoshop 4.0  
UltraEdit32

### 1.9 รายละเอียดในแต่ละบท

วิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บท ประกอบด้วย

บทที่ 1 กล่าวถึง ความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ และสมมติฐานของการศึกษา รวมทั้งทฤษฎีหรือแนวคิด ขอบเขต และวิธีที่ใช้ในการทำวิจัยนี้

บทที่ 2 อธิบายทฤษฎีและแนวคิดที่เกี่ยวข้อง กล่าวถึงหลักการเดิมที่มีอยู่แล้วในลักษณะอธิบายและวิเคราะห์การทำงานทั้งด้านที่ดีและด้อย

บทที่ 3 แนวคิดและการนำมาประยุกต์ใช้ของการแบ่งส่วนประกอบภาพเอกสารโบราณ กล่าวถึงแนวทางที่ได้คิดประยุกต์ขึ้นมาใหม่ เพื่อนำมาทดลองตามแนวความคิด ทั้งในด้านการแบ่งขอบเขตและแบ่งประเภทส่วนประกอบของภาพเอกสาร

บทที่ 4 อัลกอริทึมและขั้นตอนการทำงาน กล่าวถึงอัลกอริทึมที่ใช้ในการแบ่งขอบเขตและแบ่งประเภทส่วนประกอบ และแผนผังขั้นตอนการทำงาน

บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ เป็นการสรุปผลการทดลองและแนวทางการพัฒนาต่อไปในอนาคต

ภาคผนวก ก บทความที่ตีพิมพ์ในวารสาร

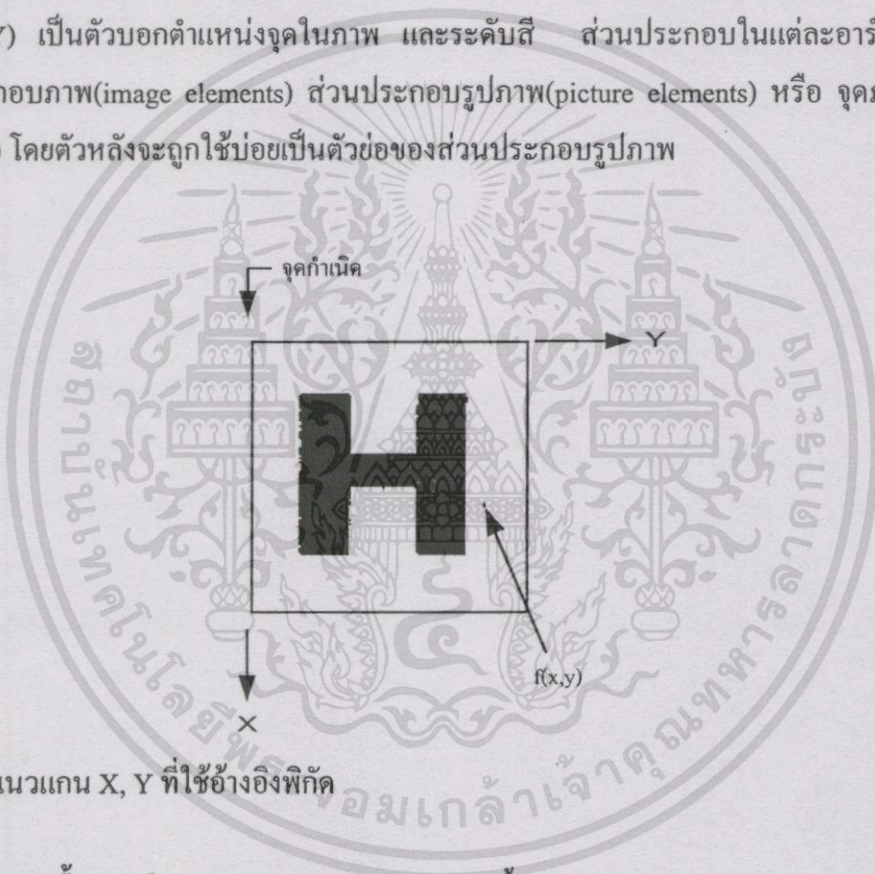
ภาคผนวก ข ภาพเอกสารที่ใช้ในการทดลอง

## บทที่ 2

# ทฤษฎีและแนวคิดที่เกี่ยวข้อง

### 2.1 การแสดงภาพดิจิทัล(Digital Image Representation)

ในกรณีภาพสองมิติทั่วไป การอ้างสีโดยใช้ฟังก์ชัน  $f(x,y)$  โดย  $x$  และ  $y$  เป็นพิกัดในแนวแกน  $X, Y$  ตามลำดับ ซึ่งค่าของ  $f$  ที่  $(x,y)$  จะเป็นระดับสี ดังรูปที่ 2.1 อธิบายแกนและพิกัดที่จะใช้ในวิทยานิพนธ์เล่มนี้ โดยภาพดิจิทัลสามารถนำไปพิจารณาให้อยู่ในรูปเมตริกซ์ที่ใช้แถว( $X$ )และคอลัมน์( $Y$ ) เป็นตัวบอกตำแหน่งจุดในภาพ และระดับสี ส่วนประกอบในแต่ละอาร์เรย์เรียกว่า ส่วนประกอบภาพ(image elements) ส่วนประกอบรูปภาพ(picture elements) หรือ จุดภาพ(pixels หรือ pels) โดยตัวหลังจะถูกใช้บ่อยเป็นตัวย่อของส่วนประกอบรูปภาพ



รูปที่ 2.1 แนวแกน  $X, Y$  ที่ใช้อ้างอิงพิกัด

### 2.2 ขั้นตอนพื้นฐานในการประมวลผลภาพ มีดังนี้

1. การทำภาพ (image acquisition) เป็นขั้นตอนนำภาพที่อยู่ในสื่ออื่นๆซึ่งไม่ใช่ดิจิทัล เช่น ภาพถ่าย ซองจดหมาย มาแปลงให้เป็นดิจิทัลโดยผ่านอุปกรณ์ เช่น สแกนเนอร์(scanner) , กล้องดิจิทัล(digital camera) เป็นต้น

2. การปรับปรุงภาพ (preprocessing) เป็นขั้นตอนในการเพิ่มเนื้อหาหรือแก้ไขรายละเอียดให้ชัดเจนยิ่งขึ้น เช่น การแก้ไขความเข้มของแสง(enhancing contrast) การปรับความเอียง (skew correction) และการลบสัญญาณรบกวน (removing noise) เป็นต้น

3. การแบ่งส่วนประกอบ (segmentation) เป็นขั้นตอนการแบ่งส่วนข้อมูลภาพให้เป็นส่วน

ที่สัมพันธ์กันหรือเป็นวัตถุ โดยการแบ่งส่วนประกอบอัตโนมัติเป็นงานที่ยากที่สุดในงานประมวล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

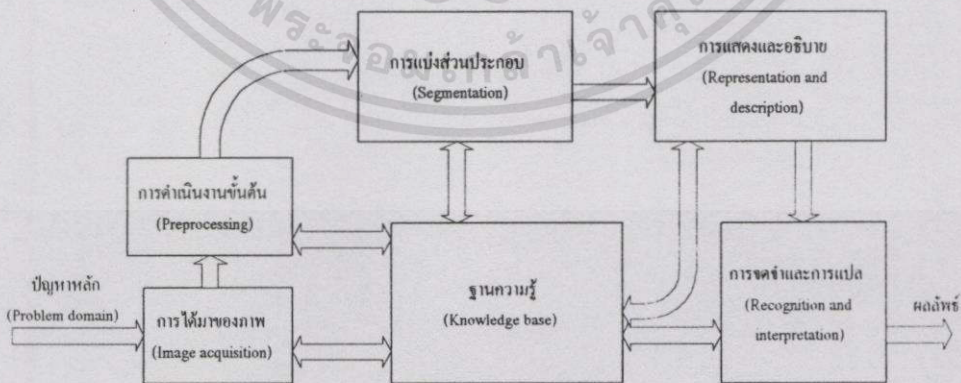
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลภาพดิจิทัล ถ้าขั้นตอนการแบ่งได้ดีและเหมาะสมจะทำให้ผลลัพธ์ที่ได้ดี แต่ถ้าอัลกอริทึมที่ใช้ในการแบ่งไม่ดี ก็จะทำให้ผลลัพธ์ที่ได้ไม่ดีตามไปด้วย ตัวอย่างเช่น ในกรณีการจดจำตัวอักษร (character recognition) ต้องแบ่งแยกเฉพาะตัวอักษรและคำออกมาจากฉากพื้น(black ground)

4. การแสดงและอธิบายผลลัพธ์(representation and description) ผลลัพธ์ของขั้นตอนการแบ่ง ปกติเป็นข้อมูลจุดภาพ(raw pixel data) ส่วนประกอบหรือขอบเขตของพื้นที่(constituting or boundary of region) หรือ จุดทั้งหมดภายในพื้นที่ การเปลี่ยนข้อมูลให้ไปอยู่ในรูปแบบสำหรับการประมวลผลเป็นสิ่งจำเป็น ขั้นแรกต้องทำข้อมูลให้อยู่ในรูปขอบเขตหรือพื้นที่ที่สมบูรณ์ โดยขอบเขตแสดงได้โดยรูปร่างของรูปทรง เช่น มุม หรือ เส้นขอบ พื้นที่ที่สามารถแสดงได้โดยพื้นผิว(texture) และรูปร่างของโครงสร้าง(skeletal shape) ขั้นตอนนี้ควรจะอธิบายข้อมูลถึงลักษณะและจุดเด่น ซึ่งจะสามารถแยกออกกว่าวัตถุนี้แตกต่างจากวัตถุนั้นอย่างไร เช่น วัตถุที่เป็นชุดตัวอักษร ซึ่งจะมีการกระจายของรายละเอียดค่อนข้างสม่ำเสมอ และจัดเรียงเป็นแนวตามแนวตั้งหรือแนวนอน ส่วนวัตถุที่เป็นภาพจะมีการกระจายของรายละเอียดไม่สม่ำเสมอ มีความหนาแน่นเป็นบางที่ และไม่ได้จัดเรียงตามแนว เป็นต้น

5. การจดจำและการแปล(recognition and interpretation) ขั้นตอนนี้ เป็นการกำหนดค่า(label) ให้กับวัตถุซึ่งจะขึ้นอยู่กับรายละเอียด และการกำหนดความหมาย(meaning)ของวัตถุนั้นๆ ที่ได้กำหนดไว้ในฐานความรู้

ขั้นตอนทั้งหมดที่กล่าวข้างต้นมีความสัมพันธ์กับ ฐานความรู้ (Knowledge base) คูในรูปที่ 2.2 เริ่มตั้งแต่การนำปัญหาที่มีมาจัดอยู่ในฐานข้อมูลความรู้(knowledge database) เพื่อให้ระบบประมวลผลภาพเข้าใจ ความรู้ในที่นี้หมายถึงรายละเอียดขอบเขตทั่วไปของภาพ ซึ่งความซับซ้อนของฐานความรู้ขึ้นอยู่กับเรื่องที่สนใจ



รูปที่ 2.2 ขั้นตอนพื้นฐานของกระบวนการประมวลผลภาพดิจิทัล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.3 ความรู้พื้นฐานเกี่ยวกับข้อมูลภาพอิเล็กทรอนิกส์

2.3.1 รูปแบบกลุ่มข้อมูลและปริมาณ(Uniform Sampling and Quantization) รูปแบบการประมวลผลของคอมพิวเตอร์ในฟังก์ชันภาพ  $f(x,y)$  บ่งบอกถึง พิกัด(spatially) และความถี่หรือแอมพลิจูด(amplitude) โดยส่วนพิกัด  $(x,y)$  จะเรียกว่า ข้อมูลภาพ(image sampling) ขณะที่ค่าของแอมพลิจูดเรียกว่าระดับเฉดสี (gray-level) โดยข้อมูลภาพจัดเรียงอยู่ในรูปแบบ อาร์เรย์  $N * M$  เมตริกซ์ ซึ่ง  $N$  คือ จำนวนแถว  $M$  คือ จำนวนคอลัมน์ ดังแสดงในสมการ 2.1 ซึ่งแต่ละส่วนประกอบในอาร์เรย์จะถูกอ้างเป็นจุดภาพและสี

$$f(x,y) = \begin{pmatrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{pmatrix} \dots (2.1)$$

2.3.2 ความสัมพันธ์ระหว่างจุด (Relationships between pixels) ที่อยู่ใกล้ชิดกันโดยกำหนดให้จุด  $p$  อยู่ที่พิกัด  $(x,y)$  จะมีจุดที่อยู่ใกล้ในแนวตั้งและแนวนอนได้แก่  $(x+1,y)$  ,  $(x-1,y)$  ,  $(x,y+1)$  ,  $(x,y-1)$  ซึ่งจุดนี้เรียกว่า ส่วนที่ติดกัน 4 ทิศของ  $p$  (4-neighbors) หรือ  $N_4(p)$  ซึ่งแต่ละจุดจะมีระยะห่างจาก  $(x,y)$  แล้วยังมีส่วนที่ติดกันของ  $p$  ที่อยู่นอกเหนือจากนี้คือ  $(x+1,y+1)$  ,  $(x+1,y-1)$  ,  $(x-1,y+1)$  ,  $(x-1,y-1)$  ซึ่งเป็น  $N_8(p)$  และทั้งหมดเรียกว่า  $N_8(p)$

2.3.3 การเชื่อมต่อกันระหว่างจุด ซึ่งใช้ในการหาค่าขอบเขตพื้นที่ของวัตถุและส่วนประกอบ โดยการเชื่อมต่อกันระหว่างจุด กำหนดได้สามชนิดดังนี้

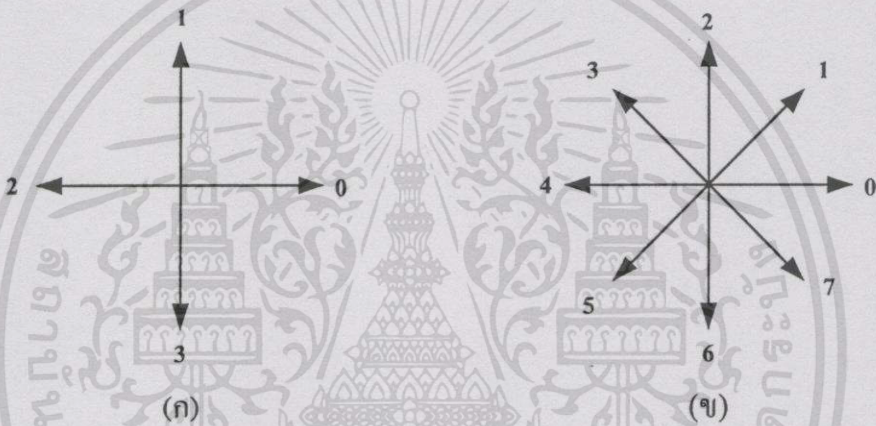
1. ส่วนติดกัน 4 ส่วน(4-connectivity) ของจุด  $p$  และ  $q$  กำหนดให้  $V$  เป็น 4 ส่วนที่ติดกัน ถ้า  $q$  อยู่ในเซตของ  $N_4(p)$
2. ส่วนติดกัน 8 ส่วน(8-connectivity) ของจุด  $p$  และ  $q$  กำหนดให้  $V$  เป็น 8 ส่วนที่ติดกัน ถ้า  $q$  อยู่ในเซตของ  $N_8(p)$
3. ส่วนติดกัน  $m$  ส่วน( $m$ -connectivity) ของจุด  $p$  และ  $q$  กำหนด  $V$  เป็น  $m$  ส่วนที่ติดกัน ถ้า
  - $q$  อยู่ใน  $N_4(p)$  หรือ
  - $q$  อยู่ใน  $N_8(p)$  และ ชุดของ  $N_4(p) \cap N_4(q)$  ว่างเปล่า

จุดภาพที่ใช้อ้างอิงในการเก็บก็ขึ้นอยู่กับสีของภาพ อย่างน้อยภาพจะต้องมี 2 สี โดยมีการอ้างข้อมูล 1 บิต ต่อ 1 พิกเซล คือภาพขาวดำหรือไบนารี(Binary) ซึ่งมีค่า 0 และ 1 และความหลากหลาย

หลายของสี่จะเพิ่มขึ้นเมื่อเพิ่มขนาดการเก็บข้อมูลต่อ 1 พิกเซล เช่น 8 บิตต่อ 1 พิกเซลจะมีการอ้าง  
 สี่ได้ถึง 256 สี คือ การอ้างถึงค่าสีจะอ้างเท่ากับ  $2^n$  เมื่อ  $n$  คือ ขนาดที่เก็บข้อมูลต่อ 1 พิกเซล

## 2.4 รหัสแนว (Chain Codes)

รหัสแนวถูกใช้ในการแสดงขอบเขตส่วนประกอบที่เรียงติดต่อกันในส่วนของเส้นตรงที่มี  
 ความยาวและทิศทาง อยู่บนหลักเกณฑ์ 4 หรือ 8 ทิศทาง โดยส่วนประกอบติดกันของข้อ(segment)  
 กำหนดทิศทางแต่ละข้อเป็นรหัสใช้ตัวเลขดังแสดงในรูปที่ 2.3 ซึ่งขอบเขตส่วนประกอบส่วนมาก  
 จะถูกอธิบายในรูปแบบของพิกัดจุด (x,y)



รูปที่ 2.3 ทิศทางของรหัสแนว 4 ทิศทาง (ก) และ รหัสแนว 8 ทิศทาง (ข)

## 2.5 คุณลักษณะพื้นฐานของภาพเอกสาร

เอกสารที่ถูกสแกนและนำมาแสดงจะอยู่ในรูปภาพแบบไบนารี(binary : 1 บิตต่อจุด) เกรย์  
 สเกล(gray-scale : 8 บิตต่อจุด) หรือ สี(color : 8-24 บิตต่อจุด) ชนิดของภาพที่ใช้ขึ้นอยู่กับนำไป  
 ใช้และการแปลงระหว่างชนิด เช่น เปลี่ยนภาพสีไปเป็นเกรย์สเกลหรือไบนารี หรือเปลี่ยนภาพเกรย์  
 สเกลเป็นไบนารี โดยขึ้นกับความต้องการที่จะเพิ่มความเร็วในการคำนวณ ส่วนประกอบพื้นฐานที่  
 แสดงในภาพ ได้แก่

1. ข้อความ (ตัวอักษรหรือตัวเลขในส่วนของข้อความ หัวเรื่อง ชื่อเรื่อง เซลล์ของตาราง  
 คำอธิบายภาพ) ที่เกาะกลุ่มในภาพหรือกราฟิกเป็นคำอธิบาย
2. ตาราง (มีเส้นแบ่งและไม่มีเส้นแบ่ง)
3. สมการคณิตศาสตร์
4. รูปไบนารี รูปฮาฟโทน และรูปสี่
5. กราฟิก (แผนผังงาน , ภาพลายเส้น , กราฟ , ไดอะแกรม , โลโก้ , เป็นต้น)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับบางขอบเขตซึ่งเป็นข้อความ บ่อยครั้งที่รายละเอียดไม่ชัดเจนพอที่จะแบ่งประเภทไปเป็นข้อความหรือรูปภาพ ยกตัวอย่างเช่น ถ้าส่วนที่เป็นข้อความอยู่ใกล้ชิดกับรูปภาพหรือกราฟิกมาก ซึ่งอาจถูกกำหนดให้เป็นขอบเขตรูปภาพ ในกรณีอื่น บางขอบเขตของส่วนประกอบสามารถแบ่งให้เป็นประเภทข้อความได้ ตัวอย่างเช่น ภาพวาดลายเส้น ซึ่งหากเกิดกรณีแบบนี้มากๆ จำเป็นต้องมีการประมวลผลพื้นที่เฉพาะ (Zone-specific processing) เพื่อใช้แก้ไขปัญหาขอบเขตส่วนประกอบที่ถูกแบ่งประเภทผิดพลาด

แผนผังทางภูมิศาสตร์ (Geometric document layout) และความเอียง (skew) สองลักษณะที่สำคัญของภาพเอกสาร แผนผังเอกสารเป็นแนวทางในการจัดเรียงของส่วนประกอบ (ข้อความ, รูปภาพ, กราฟิก, อื่นๆ) ของเอกสาร แผนผังพื้นฐานสามชนิดได้แก่ แผนผังแมนฮัตตัน (Manhattan layout) ที่ขอบเขตเป็นรูปหลายเหลี่ยมที่คั่นที่ซึ่งขอบเขตนั้นเป็นเส้นตรงแนวนอนและแนวตั้ง แผนผังสี่เหลี่ยม (rectangular layout) ซึ่งเป็นชนิดหนึ่งในแผนผังแมนฮัตตัน สุดท้ายแผนผังไร้หลักเกณฑ์ (arbitrary layout) เป็นขอบเขตหลายเหลี่ยมที่ไม่คงที่หรือมีพื้นที่ซ้อนทับกัน ความเอียงของเอกสารที่ทำให้ภาพเอกสารเอียงเป็นปัญหาหลักของการกำหนดตำแหน่งหน้าเอกสาร ความเอียงจะทำให้เอกสารอยู่ไม่ถูกตำแหน่งที่ควรจะเป็นตั้งแต่สแกนเข้ามา ในหลายกรณีทำให้ประสิทธิภาพในการแบ่งหน้าเอกสารและการแบ่งประเภทขอบเขตแย่ลง

วิธีวิเคราะห์แผนผังส่วนมากแบ่งขอบเขตและจัดประเภทภาพออกเป็น 3 ประเภทคือ ข้อความ กราฟิก และ รูปภาพ โดยสมการคณิตศาสตร์จะถูกกำหนดเป็นประเภทข้อความ และ ตารางจะถูกกำหนดเป็นกราฟิก เมื่อจัดประเภทโดยใช้ข้อมูลที่ละเอียดขึ้น ขั้นตอนนี้เรียกว่าการทำความเข้าใจภาพเอกสาร (document understanding) ซึ่งข้อความจะถูกจัด ไปอยู่ในส่วนประกอบทางตรรกะ (logical layout)

## 2.6 ประเภทของเอกสารและการประยุกต์ใช้

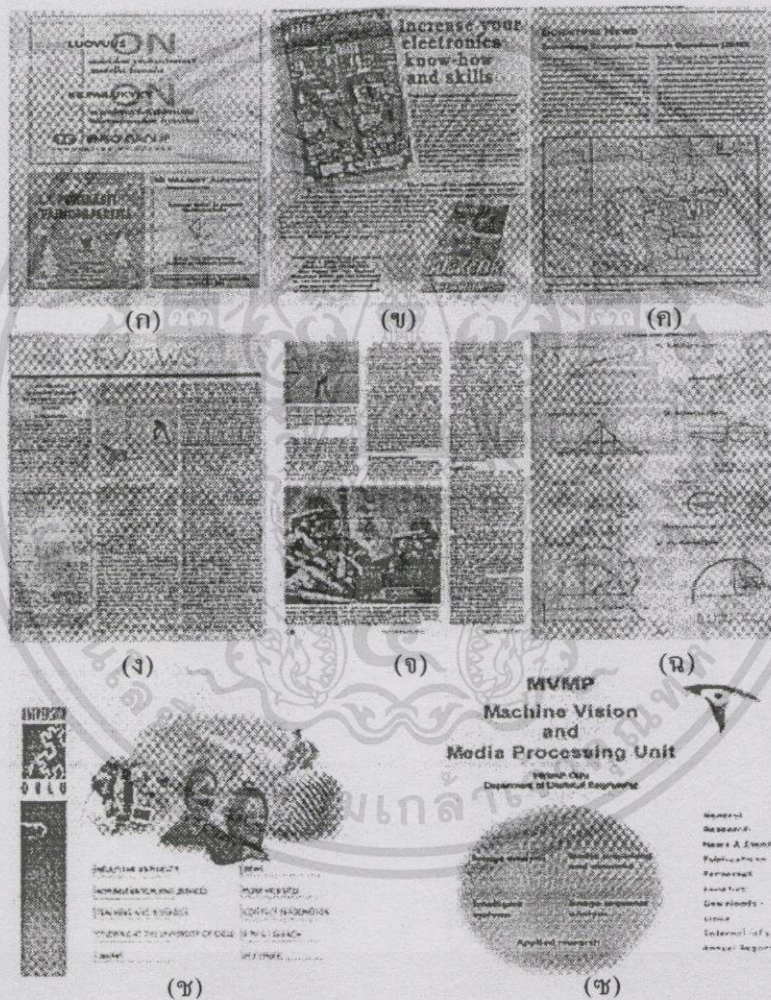
เนื่องจากมีความเป็นไปได้ยากที่จะใช้วิธีการแบบเดียวกันในการจัดประเภทของส่วนประกอบเอกสารที่มีส่วนประกอบหลากหลายและซับซ้อน ทำให้มีการเสนอวิธีการต่างๆ ที่ออกแบบมาเพื่อวิเคราะห์ภาพเอกสารที่มีส่วนประกอบภายในหลากหลายประเภท เห็นได้ชัดว่าประโยชน์ส่วนมากเริ่มมาจากการแบ่งเอกสารโดยใช้แผนผังทางภูมิศาสตร์ ดังนั้นเมื่อมีวิธีการใหม่ๆ เพิ่มขึ้น สามารถที่จะประเมินคุณค่าจากวิธีการวิเคราะห์แผนผังที่ใช้จากคุณสมบัติพื้นฐานของประเภทส่วนประกอบเป้าหมาย แม้กระนั้นก็ยังไม่สามารถที่จะพิจารณาตามลำดับของใจความได้ หรือได้แต่พิจารณาในประเภทที่จำกัด ในการจัดประเภทของชนิดเอกสารแบ่งตามลักษณะได้ดังนี้

1. เอกสารรูปแบบโครงสร้าง (วารสาร หนังสือพิมพ์ และ จดหมายข่าว)
2. เอกสารที่มีรูปแบบไม่แน่นอน (โฆษณา, ปกซีดี, หนังสือ และ นิตยสาร)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เอกสารกึ่งรูปแบบ (ของจดหมาย , จดหมายและนามบัตร , เช็คนาคาร , แบบฟอร์ม และ เอกสารที่คล้ายตาราง)
4. แผนที่และภาพวาดเชิงวิศวกรรม
5. เอกสารที่ไม่ปกติ(หน้าโฮมเพจ และ รูปจากวิดีโอ)

ในส่วนนี้เป็นการสรุปภาพรวมของหลักการแบ่งส่วนประกอบและการแบ่งประเภทส่วนประกอบที่ขึ้นอยู่กับตำแหน่งส่วนประกอบของตัวอย่างในแต่ละชนิดที่กล่าวถึงทั้งหมดรวมถึงปัญหาที่เกิดขึ้น จากรูปที่ 2.4 แสดงตัวอย่างภาพเอกสารชนิดต่างๆ



รูปที่ 2.4 ตัวอย่างภาพเอกสารชนิดต่างๆ

- |                   |                   |                 |
|-------------------|-------------------|-----------------|
| ก. สิ่งพิมพ์โฆษณา | ข. สิ่งพิมพ์โฆษณา | ค. หนังสือพิมพ์ |
| ง. บทความ         | จ.บทความ          | ฉ. หนังสือ      |
| ช. หน้าเว็บ       | ซ. หน้าเว็บ       |                 |

\* ที่มา Okun O, Doermann D & Pietikäinen M, "Page Segmentation and Zone Classification: the state of the art"

เอกสาร, Technical Report LAMP-TR-036(1999) , page 2 เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.6.1 เอกสารรูปแบบโครงสร้าง(structure articles)

วารสาร หนังสือพิมพ์ และจดหมายข่าว เป็นแหล่งข้อมูลข่าวสารแรกและเป็นสิ่งพิมพ์ที่แพร่หลายและหาง่าย ภาพทั้งหมดที่เก็บมีหลากหลาย (ข้อความที่มีขนาด รูปทรง และแบบที่แตกต่างกัน ความหลากหลายของกราฟิก ภาพไบนารี ภาพฮาฟโทนและภาพสี) ซึ่งสามารถสแกนเป็นไบนารี เกรย์สเกลหรือสี ขอบเขตนั้นจะถูกแบ่งตามชนิดกายภาพของแต่ละประเภท อย่างไรก็ตาม ถ้าช่องว่างระหว่างขอบเขตส่วนประกอบที่แตกต่างกันแคบมาก และในบางกรณีจะมีข้อความที่อยู่ในกราฟิกหรือรูป บางที่ตัวหนังสือก็สีเข้ม(สิ่งพิมพ์ปกติ) หรือ สีสว่าง(สีตรงข้ามกับปกติ)กว่าฉากพื้น บทความในวารสาร หนังสือพิมพ์ และ จดหมายข่าวเป็นตัวอย่างของเอกสารโครงสร้าง เพราะทุกส่วนประกอบบนหน้ามีการจัดเรียงลำดับตามชนิด เช่น หัวเรื่อง คำนำ และเนื้อหา โดยจะเชื่อมต่อกันด้วยกัน ด้วยกฎต่างๆไป ตัวอย่างของกฎเช่น การอ่านตามลำดับของบล็อกข้อความจากบนลงล่าง และจากซ้ายไปขวา คำอธิบายภาพจะอยู่ใต้ภาพ แบบฟอร์มข้อความจะสัมพันธ์กับขนาดของพื้นที่ รูปหรือตารางจะอยู่หลังจากที่ข้อความอ้างอิง หนังสือพิมพ์จะมีความซับซ้อนมากกว่าวารสารหรือจดหมายข่าว แต่องค์ประกอบของโครงสร้างทางตรรกะจะน้อยกว่า ซึ่งทั้งหมดก็สามารถแสดงในแผนผังเมนฮัตตัน และแผนผังไม่มีกฎเกณฑ์ได้ ซึ่งจะรวมถึงหน้าเอกสารของวารสารที่คล้ายตาราง แต่แผนผังนี้ใช้ได้ไม่บ่อยดีกว่าวารสารที่ไม่ใช่เชิงเทคนิค ถึงกระนั้นรูปแบบของแผนผังไม่ว่าจะเป็นภาษาหรือสิ่งพิมพ์พิเศษ ซึ่งปกติแล้วจะมีรูปแบบของเนื้อหาและรายละเอียดมากมาย โดยประเด็นงานหลัก คือการกำหนดตำแหน่งขอบเขตรวมถึง การประมาณและแก้ไขความเอียง(บางครั้งเป็นตัวเลือก) การแยกเอกสารเฉพาะส่วนที่เหมือนกัน และนำไปจัดแบ่งประเภทขอบเขตเป็นฉากพื้น ข้อความ กราฟิก และ รูปภาพ โดยทั่วไปบรรทัดข้อความที่ไม่มีความเอียงมีรูปแบบการจัดเรียงตามแนวตั้งและแนวนอน แต่สามารถเอียงได้ถ้า แกน X หรือ แกน Y เอียง ปกติทุกขอบเขตบนภาพจะมีความเอียงเดียวกันทั้งหน้าเอกสาร แต่บางกรณีอาจมีบางขอบเขตที่มีทิศทางแตกต่างจากขอบเขตอื่นๆ

แม้จะมีแผนผังโครงสร้างหลายรูปแบบ แต่ปัญหาหลักที่ยังคงอยู่คือความแตกต่างของชนิดภาพ เงื่อนไขการนำไปใช้ซึ่งควรปรับปรุงให้ดี(ความเอียง , รูปแบบแผนผังเอกสาร) และสัญญาณรบกวน

### 2.6.2 เอกสารที่มีรูปแบบไม่แน่นอน

นิยาม เอกสารที่มีรูปแบบไม่แน่นอน สะท้อนให้เห็นความขาดแคลนของกฎทั่วไปที่จะนำมากำหนดส่วนประกอบเหล่านี้ แต่ส่วนมากการกำหนดแผนผังจะขึ้นอยู่กับลักษณะเอกสารและข้อกำหนดซึ่งออกแบบมาใช้สำหรับรายละเอียดที่สนใจ

- โฆษณา

โฆษณาปกติจะพิมพ์อยู่ในนิตยสารหรือหนังสือพิมพ์ แต่ก็สามารถอยู่บนเว็บเช่น เอกสารอิเล็กทรอนิกส์ แผนผังของเอกสารดังกล่าวจะไม่คงที่เหมือนกับหัวเรื่องวารสารและหนังสือพิมพ์ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พื้นที่ข้อความภาพมีได้หลายความเอียง บรรทัดข้อความเป็นรูปโค้ง ซึ่งตัวอักษรไม่เรียงลำดับกัน เป็นแนวเส้นตรง และกลุ่มข้อความข้างในหรือที่ติดกับรูปภาพเกิดได้บ่อยในการแบบที่เป็นข้อมูลสำคัญ และเพื่อที่ดึงดูดสายตาของผู้อ่าน ดังนั้นโฆษณาเป็นเอกสารแบบไม่เป็นโครงสร้าง [42,48]

#### - หน้าปก

ภาพของเอกสารที่บ่อยๆจะใช้สีและมีตำแหน่งข้อความที่หลากหลาย และรูปแบบอักษรหลายขนาดทั้งฉากพื้นมีหลายสีหรือหลายลวดลาย

การวิเคราะห์แผนผังของภาพเพื่อจัดประเภทข้อความและจดจำ ซึ่งมีประโยชน์สำหรับการเรียกค้นข้อมูล การเรียกค้นข้อความง่ายกว่าใช้สูตรอย่างใดอย่างหนึ่งในการค้นหาภาพเช่น รูปทรงพื้นผิว หรือสี ความทำทลายอยู่ที่แผนผังไม่แน่นอนและพื้นฉากมีหลายสี ซึ่งทำการตรวจหาตัวอักษรให้ถูกต้องได้ยาก

#### 2.6.3 เอกสารกึ่งรูปแบบ (semi-structured layouts)

แผนผังกึ่งโครงสร้างมีส่วนประกอบโครงสร้างข้างในซึ่งผู้ใช้จะใส่ข้อมูลลงตรงที่กำหนดให้ แต่ชนิดของมันไม่มีกำหนดว่าจะอยู่ตรงส่วนไหน หมายความว่าเอกสารสองชุดสามารถที่มีแผนผังที่แตกต่างกันทั้งๆที่มีส่วนประกอบข้างในเหมือนกัน

#### - ของจดหมายและเช็คธนาคาร

ไปรษณียบัตรมีรูปแบบเดียวกับเช็คธนาคารซึ่งเป็นตัวอย่างเอกสารแบบกึ่งโครงสร้าง ซึ่งสามารถกำหนดได้ก่อนว่าฟิลด์แต่ละฟิลด์เป็นบล็อกที่อยู่ หรือ ผลรวมจำนวนเงิน แต่ขอบเขตของฟิลด์นั้นๆที่อยู่ในเอกสารเป็นระเบียบแบบแผนหลักที่สำคัญ นามบัตรก็ถูกรวมอยู่ในประเภทนี้ด้วย เพราะ มีโครงสร้างทางตรรกะคล้ายคลึงกัน ภาพของเอกสารชนิดนี้จะเป็น ไบนารี เกรย์สเกล หรือ สี และข้อความจะถูกพิมพ์บนพื้นผิวฉากที่ซับซ้อน

งานหลักที่ต้องดำเนินการ คือ การกำหนดความแตกต่างของฟิลด์ซึ่งมีลักษณะเฉพาะของแต่ละเอกสาร ดังเช่นเช็คธนาคารรายละเอียดนั้นจะรวมถึง ลายเซ็น หมายเลขเช็ค วันที่ ผลรวมยอดเงินเป็นตัวเลข ผลรวมยอดเงินที่เป็นตัวอักษร เลขที่บัญชี ชื่อผู้ส่งจ่าย ที่อยู่ของสถาบันการเงิน และ/หรือตรา แต่สำหรับนามบัตรประกอบด้วย ชื่อผู้ถือ กิจการ และที่อยู่ ส่วนไปรษณียบัตรมีบล็อกที่อยู่ แสคัมปี บาร์โค้ด(barcode) และสถานที่รับไปรษณีย์

ปัญหาหลักที่สังเกตได้

1. พื้นฉากที่ซับซ้อนทำให้ยากต่อการทำแปลงเป็นภาพไบนารีหรือแบ่งส่วนประกอบภาพ
2. ตัวอักษรตัวพิมพ์ ตัวอักษรลายมือ สัมผัสหรือซ้อนทับกับเส้นบรรทัด
3. การเปลี่ยนแปลงของความสว่าง
4. ตำแหน่งเอกสารที่ไม่แน่นอน ที่เคลื่อนย้ายตามการจัดวางรูปแบบ
5. ข้อจำกัดของเวลาที่ใช้ในการประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แบบฟอร์มและเอกสารที่เป็นตาราง

แบบฟอร์มและเอกสารที่เป็นตาราง เช่น กระดาษคำตอบ หรือ ใบส่งของ ที่ประกอบด้วยฟิลด์หรือเซลล์ข้างในซึ่งข้อมูลจะถูกเขียนหรือพิมพ์เข้าไป ในการวิเคราะห์ภาพเริ่มต้นปกติใช้ในรูปแบบภาพไบนารี จนถึงแบบฟอร์มสีก็มี อย่างไรก็ตามอันหลังถ้าหากพื้นภาพเป็นรูปแบบเดียว ก็ไม่ยากที่จะแบ่งข้อความออกมาจากแบบฟอร์ม ในส่วนเพิ่มเติมของข้อความ ใบส่งของสามารถเก็บรูปเล็กๆ เช่น ตราสัญลักษณ์ รูปแบบและเอกสารแบบตารางเป็นเอกสารกิ่งโครงสร้างเพื่อสร้างจำกัดโดยใช้ตัวแบ่งเป็นตัวกำหนดขอบเขตข้อมูลตัวอักษร

งานพื้นฐานที่จะใช้ในการแยกขอบเขตของฟิลด์ ขอบเขตบล็อก หรือเส้นบรรทัดออก เพื่อที่จะสกัดเฉพาะข้อความภายในแต่ละฟิลด์ ใบส่งของโดยปกติมีที่อยู่ของบริษัท และเลขที่บัญชีธนาคาร ซึ่งการดำเนินงานก็เหมือนกับการประมวลผลเซ็คชันการ

การวิเคราะห์แบบฟอร์มทั้งสองชนิด อันดับแรกต้องรู้แบบฟอร์มที่จะวิเคราะห์ กำหนดฟิลด์ในฟอร์มให้อยู่รูปแบบประเภทใดประเภทหนึ่ง และรูปแบบ(model)ของฟิลด์ที่ว่างหรือเซลล์ที่ยังไม่ได้ใส่ข้อมูล ในกรณีนี้งานแรกคือการกำหนดฟอร์มโดยใช้ค่าที่กำหนดหรือ โครงสร้างที่สามารถถูกสกัดจากแบบได้ มีอยู่สองชนิดคือ อย่างแรกชนิดแบบฟอร์มถูกกำหนดเพียงครั้งเดียวเพื่อที่กำหนดพิกัดขอบเขตของฟิลด์ข้อมูลที่ต้องการ อย่างที่สองเป็นการวิเคราะห์แบบฟอร์มที่ไม่รู้เกี่ยวกับรูปแบบที่นำมาใช้ เมื่อไม่รู้เกี่ยวกับชนิดของแบบฟอร์มที่ต้องใช้มาก่อน ต้องใช้อัลกอริทึมที่ใช้สกัดแยกพื้นฐาน โครงสร้างแบบฟอร์มเป็นพื้นฐานของตัวแบ่งระหว่างเซลล์หรือฟิลด์ของชนิดที่สองนี้จะคิดแปลงได้ยืดหยุ่น แต่จะใช้เวลาคำนวณมาก และบางครั้งได้ความถูกต้องน้อย

ปัญหาจะเพิ่มขึ้น เมื่อข้อมูลหรือข้อความถูกเขียนนอกขอบเขตฟิลด์ที่กำหนดไว้ ซึ่งเป็นการยากที่จะกำหนดข้อมูลได้ถูกต้อง การแปลงภาพ (การหมุน การย่อขยาย และ การเปลี่ยนแปลงรูปร่าง) ดังนั้นจึงมีความยุ่งยากที่ต้องดำเนินการจับคู่รูปแบบก่อนเพื่อหาข้อมูลเข้าของภาพ สำหรับใบส่งของที่ส่งทางเครื่องแฟกซ์จากผู้ขายถึงลูกค้า ลักษณะของภาพเหมือนเดิมแต่มีคุณภาพของรายละเอียดต่ำลงเพราะอาจเกิดการสูญเสียข้อมูลเนื่องจากการส่งแฟกซ์ และการประทับตราลงบนพื้นที่ที่เป็นข้อมูล คือสาเหตุหนึ่งที่ทำให้เกิดผลกระทบที่ไม่ดีต่อการสกัดข้อมูลหรือข้อความของเอกสาร

#### 2.6.4 แผนที่และภาพวาด

- แผนที่

ภาพแผนที่ปกติจะเป็นภาพเกรย์สเกลหรือสีที่ใช้บอกรายละเอียด แผนที่สีจะประกอบด้วยชั้นสีหลายชั้น แต่ละชั้นสีจะตรงกับข้อมูลชนิดเบื้องต้นที่กำหนดไว้ ข้อมูลที่เป็นข้อความ(ชื่อเมือง ประเทศ ทะเลสาบ และอื่นๆ) รายละเอียดที่กำหนดไว้ในลักษณะกราฟิก(ถนน ความแตกต่างของพื้นผิว และอื่นๆ)

ข้อมูลที่เป็นข้อความบนแผนที่หรือภาพวาดมีขนาดรูปแบบที่เล็กและมีจำนวนน้อย สามารถผสมกับส่วนที่เป็นกราฟิก ดังนั้นจึงมีบางกฎที่ใช้สำหรับแยกสิ่งที่เหมาะสมกันอยู่ โดยสามารถที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะแยกได้ถูกต้องว่าเป็นตัวหนังสือหรือขอบเขตที่แคบของแผนที่ เพื่อที่จะทำให้ไม่คลุมเครือ และเพื่อกำหนดให้เป็นส่วนประกอบที่อยู่ในโครงสร้าง ปกติข้อความในแผนที่จะปรากฏได้ทุกแห่งในแผนที่และแนวบรรทัดของข้อความมีทิศทางไม่แน่นอน

ดังนั้นรูปแบบที่มากมายของแผนที่ (เกี่ยวกับที่ดิน , อุตสาหกรรม , ภูมิประเทศ , อื่นๆ) การวิเคราะห์ของภาพแผนที่ แบ่งงานได้สามประเภทคือ

1. การแบ่งข้อความและกราฟิก
2. กำหนดค่าให้ตัวอักษรและสัญลักษณ์พิเศษ
3. การกำหนดทิศทางของกราฟิก

สำหรับภาพสีการแบ่งสีจะถูกทำตั้งแต่การสแกนจากภาพทีละสีโดยสแกนเนอร์ที่ออกแบบเพื่อใช้ในงานนี้โดยเฉพาะ ตัวอักษรและสัญลักษณ์จะถูกแบ่งแยกออกจากกราฟิกที่ระดับสีโดยใช้ขนาดของตัวอักษรและสัญลักษณ์ โดยทั้งหมดจะถูกส่งไปยังกระบวนการจดจำ ขณะที่กราฟิกจะถูกทำให้เล็กลงและกำหนดขนาดและทิศทางของกราฟิก

เนื่องจากแผนที่ที่มีรูปแบบที่หลากหลายจึงก่อให้เกิดวิธีการและระบบที่ใช้งานมากมาย โดยแต่ละวิธีอาจจะถูกนำไปประมวล โดยเฉพาะกับแผนที่ชนิดใดชนิดหนึ่งได้เท่านั้น ปัญหาต่างๆคือตัวอักษรและสัญลักษณ์สัมผัสหรือซ้อนทับกราฟิก ข้อความเอียงหรือโค้ง ระยะห่างระหว่างตัวอักษรที่ใกล้กันและในข้อความเดียวกันมีระยะห่างที่ไม่แน่นอน ความแตกต่างของขนาดตัวอักษร (แม้ว่าจะรู้ล่วงหน้าแบบอักษรที่ใช้ในการสร้างแผนที่) และสัญลักษณ์กราฟิกบางตัวที่คล้ายกับตัวอักษร

#### - ภาพวาดเชิงวิศวกรรม

ภาพวาดเชิงวิศวกรรมนั้นมีความน่าท้าทายในการวิเคราะห์พอๆกับแผนที่ ซึ่งในเอกสารของทั้งสองประเภทเก็บข้อมูลตัวอักษรขนาดเล็กและอยู่อย่างกระจัดกระจายไปทั่วพื้นที่ซึ่งกราฟิกมีขนาดใหญ่ อย่างไรก็ตามภาพวาดทางวิศวกรรมโดยทั่วไปเป็นภาพไบนารีหรือเกรย์สเกล มีรายละเอียดที่อยู่ห่างกัน และมีภูมิศาสตร์แบบปิด การวิเคราะห์แผนผังทางภูมิศาสตร์ของภาพวาดเชิงวิศวกรรม การทำงานคือต้องการที่จะแบ่งตัวอักษรหรือกราฟิก และแบ่งประเภทส่วนประกอบของกราฟิกเช่น เส้น (เส้นประ หรือเส้นจุด) เส้นโค้ง รูปทรง เป็นต้น

การแบ่งชนิดของภาพวาดเชิงวิศวกรรมที่ใช้ส่วนมากคือ วิธีการเฉพาะที่สามารถนำมาใช้ได้กับส่วนประกอบหนึ่งหรือสองชนิดที่แตกต่างกัน ปัญหาที่เป็นปัญหาหลักของเอกสารแบบนี้ คือข้อความที่ติดกับกราฟิก หรือ เอกสารต้นฉบับมีคุณภาพแย่

#### 2.6.5 หน้าเว็บ และ ภาพจากวิดีโอ

หน้าเว็บและภาพจากวิดีโอที่มีรายละเอียดเป็นอักษรถูกจัดให้เป็นเอกสารประเภทใหม่ คือ เอกสารอิเล็กทรอนิกส์ ซึ่งจะมีคุณลักษณะไม่เหมือนกับเอกสารอื่นที่อธิบายมาแล้วก่อนหน้านี้ เอกสารนี้โดยทั่วไปเป็นหน้าสีและภาพจะมีความละเอียดต่ำ มีขนาดที่เล็กซึ่งจะเรียกดูภาพได้รวดเร็ว โดยรูปภาพจะช่วยให้เกิดความเข้าใจที่ชัดเจนยิ่งขึ้น ส่วนประกอบหลักของเอกสารมีทั้งข้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความและรูปภาพที่มีขอบเขตและตำแหน่งไม่แน่นอน ตัวอักษรในหน้าเว็บที่เห็นอาจเป็นได้ทั้งตัวอักษรในรูปแบบตัวอักษรและตัวอักษรและรูปแบบรูปภาพอย่างใดอย่างหนึ่ง สำหรับภาพจากวิดีโอ นั้นส่วนมากข้อความจะอยู่ในส่วนของภาพที่ค่อนข้างชัดชัด สิ่งที่น่าประหลาดใจจะเป็นฉากหรือข้อความกราฟิกอย่างใดอย่างหนึ่ง และความละเอียดของภาพที่น้อยกว่า 200-300 จุดต่อนิ้ว สำหรับเอกสารชนิดนี้จะทำการตรวจสอบและจดจำส่วนประกอบได้ยาก

หน้าเว็บและภาพจากวิดีโอ จัดอยู่ในประเภทเอกสารที่ไม่มีโครงสร้าง เพราะแผนผังของหน้าเว็บและภาพจากวิดีโอ ไม่สามารถจำกัดขอบเขตโดยการกำหนดกฎไว้ก่อน และในกรณีของภาพจากวิดีโอ บ่อยครั้งที่ไม่สามารถกำหนดกฎได้ งานหลักของเอกสารประเภทนี้คือ การกำหนดขอบเขตของข้อความที่อยู่ในภาพ มีวัตถุประสงค์เพื่อการสืบค้นหาข้อมูลและทำดัชนีวิดีโอ

ปัญหาหลักของเอกสารประเภทนี้ได้แก่

1. สีของข้อความและฉากพื้นจะใกล้เคียงกันเมื่อความละเอียดของภาพต่ำ หรือข้อความที่มีความเข้มน้อยเมื่อเทียบกับฉากพื้น (ตัวอย่างข้อความที่อยู่บนลวดลายฉากพื้นที่ชัดชัด) ดังนั้นการเรโซลิว (thresholding) ไม่สามารถแบ่งข้อความและฉากพื้นได้
2. ส่วนประกอบข้อความ อาจวิเคราะห์ผิดพลาดเมื่อเป็นบรรทัดโค้ง หรือในกรณีที่แย่มากที่สุด คือ ไม่มีการจัดเรียงกันเป็นแนว (บรรทัดที่มีลักษณะเป็นคลื่น)
3. ส่วนประกอบข้อความที่ไม่ได้เป็นสีเดียวกัน และมีการขาดตอนไปเป็นส่วนย่อยหลายส่วน หรือเป็นสัญญาณรบกวนอันเนื่องมาจากการออกแบบตัวอักษร
4. ผลกระทบที่เกิดจากการเคลื่อนไหวของภาพในวิดีโอ

## 2.7 ขั้นตอนการแบ่งหน้าเอกสารและวิธีการแบ่งประเภทขอบเขต

ในส่วนนี้แสดงขั้นตอนการแบ่งหน้าเอกสารหรือส่วนประกอบของเอกสารและวิธีการแบ่งประเภทขอบเขตของส่วนประกอบ โดยประเภทของเอกสารนั้นได้กล่าวมาแล้วในข้อที่ 2.6 มีเอกสารรูปแบบโครงสร้าง (วารสาร , หนังสือพิมพ์ และจดหมายข่าว) เอกสารที่มีรูปแบบไม่แน่นอน (โฆษณา , ปกซีดี , หนังสือ และ วารสาร) และ เอกสารที่ไม่ใช่รูปแบบเก่า เช่น หน้าจอเว็บ เป็นต้น โดยทั่วไปเอกสารจะมีการอธิบายรายละเอียดเป็นหน้า (ในรูปแบบกระดาษหรืออิเล็กทรอนิกส์) และประกอบด้วยขอบเขตรายละเอียดที่ต่างกัน ซึ่งต่างจากภาพวาดลายเส้น แผนที่ เช็ครถยนต์ และแบบฟอร์ม เพราะโครงสร้างขอบเขตที่เหมือนกัน โดยทั้งหมดมีส่วนประกอบที่ผสมกันทั้งที่เป็นข้อความและกราฟิก และบ่อยครั้งที่ข้อความกับกราฟิกชิดกันหรือทับกัน

ในขั้นการแบ่งส่วนประกอบหมายถึงการแบ่งภาพออกเป็นส่วนๆ ซึ่งในแต่ละส่วนต้องเป็นชนิดเดียวกัน หรือ พื้นที่แต่ละขอบเขตต้องเป็นข้อมูลเพียงหนึ่งชนิดเท่านั้น เช่น ข้อความหรือกราฟิก และจะมีการจัดประเภทส่วนประกอบหลังจากแบ่งขอบเขตส่วนประกอบเอกสารแล้ว ดังนั้นการจัดแบ่งประเภทส่วนประกอบจะกำหนดประเภทได้จากพื้นฐานที่เป็นจุดเด่น ซึ่งหลายเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรรมวิธีนั้น คำเนิการแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบ โดยแต่ละวิธีการจะกล่าว  
ย่อๆไม่เห็นรายละเอียดของการนำไปใช้ แล้วจะยกตัวอย่างบางวิธีที่น่าสนใจในหัวข้อถัดไป

### 2.7.1 การแบ่งส่วนประกอบเอกสาร

การแบ่งส่วนประกอบเอกสารสามารถแบ่งตามแนวคิดได้เป็น วิธีการล่างขึ้นบน(bottom-up ; data-driven) วิธีการบนลงล่าง(top-down , model-driven) และ วิธีการผสมผสาน(hybrid เป็น  
การผสมระหว่างวิธีการบนลงล่างและวิธีการล่างขึ้นบน) วิธีการแบบบนลงล่าง ใช้มานานและมีการ  
ทำงานที่รวดเร็ว เมื่อรู้รายละเอียดของลำดับแผนผังเอกสาร ในกรณีนี้กระบวนการทำงานจะเริ่มทำ  
จากหน้าเอกสารทั้งหมดที่อยู่ในระดับสูงสุดนำมาแบ่งทั้งหมดให้เป็นขอบเขตที่เล็กลงเช่น บล็อก  
บรรทัด คำ และตัวอักษร วิธีการล่างขึ้นบนจะเริ่มจากจุดและรวมกลุ่มเข้าด้วยกันเป็นขอบเขต แล้ว  
รวมขอบเขตที่ใกล้ชิดกันเข้าเป็น โครงสร้างที่ใหญ่ขึ้น ซึ่งวิธีการนี้มีความแม่นยำสามารถใช้ได้กับ  
เอกสารที่มีรูปแบบที่หลากหลาย แต่เวลาในการประมวลผลส่วนมากใช้เวลานาน วิธีการแบบผสม  
ผสานเป็นวิธีที่อยู่ระหว่างกลางของวิธีการบนลงล่างและวิธีการล่างขึ้นบน โดยพยายามที่จะรวม  
ความสามารถในการทำงานด้วยความเร็วแบบวิธีการบนลงล่างและความมั่นคงถูกต้องในการทำงาน  
ของวิธีล่างขึ้นบน และจัดแบ่งประเภทของส่วนประกอบให้เป็นข้อความหรือกราฟิก โดยวิธีการ  
เหล่านี้สามารถแบ่งขอบเขตเอกสาร โดยใช้การแบ่งลดสายพื้นที่แตกต่างกันเป็นข้อกำหนดในการ  
แบ่งขอบเขต ซึ่ง O. Okun และ M. Pietikäinen [1] ได้สรุปวิธีการใช้พื้นที่การแบ่งจากลดสาย

#### 2.7.1.1 วิธีการล่างขึ้นบน

วิธีการนี้ใช้พื้นฐานของการวิเคราะห์ส่วนประกอบของภาพไบนารีและรวมส่วนประกอบ  
เหล่านั้นเข้าเป็นตัวอักษร บรรทัด และบล็อกโดยใช้ความใกล้ชิดระหว่างส่วนประกอบที่ใกล้กัน  
และขนาดของส่วนประกอบ การทำเบลอเพื่อเชื่อมส่วนประกอบที่ใกล้ชิดกันเข้าด้วยกัน และไดอะ  
แกรมโวโรนอย(Voronoi diagrams) เป็นวิธีการจับกลุ่มหลัก ระยะเวลาที่ใช้ในการทำงานในแต่ละ  
ภาพจะต่างกันขึ้นอยู่กับวิธีการที่ใช้ วิธีการในแบบล่างขึ้นบนนี้ส่วนมากมีความคงทน(tolerant)กับ  
เอกสารที่มีความเอียง(ซึ่งบางทีอาจจะมีหลายความเอียง คือ ส่วนประกอบในภาพแต่ละส่วนเอียง ไป  
คนละทิศทาง)และเอกสารที่มีแผนผังไม่แน่นอน ความเอียงจะถูกคำนวณจากผลรวมของบรรทัดข้อ  
ความที่แยกออกออกมา ถ้าความเร็วในการทำงานหรือหลักเกณฑ์ทั่วไปเป็นตัวกำหนด การเลือกตัว  
อย่างข้อมูลเป็นหลักสำคัญสำหรับวิธีการในกลุ่มนี้ ทุกขอบเขตจะถูกแสดงในโครงร่างต้นไม้แบบ  
ขึ้นลำดับ สำหรับประมวลผลเอกสารที่หลากหลายชนิด ที่เสนอโดย S.Y. Wang และ T. Yagasaki  
[2] สามารถใช้กับ แบบฟอร์ม หรือ วารสาร อักษรประกอบด้วยตัวภาษาอังกฤษ และตัวคันจิ บล็อก  
ที่ใกล้ชิดกันใช้เวลาในการประมวลผลที่เร็วในการวิเคราะห์คุณสมบัติของบล็อกและการอ้างอิงของ  
ต้นไม้

การวิเคราะห์ส่วนที่ติดกันของบล็อกสีเหลี่ยมก่อนที่จะแบ่งประเภทเข้าด้วยกัน โดยใช้หน้า กากเล็กๆ จำนวนมากที่อยู่ในขอบเขตใหญ่ในการเชื่อมต่อระดับบล็อก สามารถลดเวลาที่ใช้ในการ ทำงานได้อย่างมากเมื่อเปรียบเทียบกับ การเชื่อมต่อระหว่างจุด[3]

วิธีการที่ T. Pavlidis และ H. Zhou [4] เสนอคือ การสกัดแยกข้อมูลโดยการคำนวณค่าของ จุดภายในหน้าต่าง  $5 \times 5$  ที่แต่ละจุดจะถูกสแกนภาพโดยแถวต่อแถว วิธีการตัดส่วนที่อยู่ใกล้กันโดย ใช้บล็อกขนาดเล็กมารวมจุดเข้าเป็นพื้นที่เดียวกัน ประโยชน์ของวิธีการนี้สามารถทำได้ทั้งภาพไบ นารีและเกรย์สเกล แต่ไม่คงทนกับเอกสารที่เอียงหรือมีรูปแบบไม่แน่นอน

แนวคิดพื้นฐานของการวิเคราะห์ฉากพื้นของภาพไบนารีถูกนำมาเสนอใน [5], [6], [7], [8] วิธีการที่ใช้ได้ประโยชน์ที่สุด คือวิธีการที่เรียกว่า กระเบื้องขาว(white tiles) [5], [6] ในการ ทำงานอนแรกจะสร้างโครงข่ายของสีเหลี่ยม ซึ่งแต่ละอันจะกว้างเท่ากับความกว้างของช่องว่างสี ขาว(ฉากพื้น) และเป็นสีเหลี่ยมที่อยู่ติดกันวนรอบขอบเขต วิธีการนี้เป็นการแสดงที่ยืดหยุ่น ที่ สามารถแสดงได้ทั้งแผนผังเมนฮัตตันและแบบที่ไม่แน่นอน โดยไม่มีการแตกขอบเขตรูปทรงที่ซับซ้อน วิธีการนี้ไม่ต้องแก้ไขความเอียงของภาพเอกสารก่อน และสามารถที่จะกำหนดขอบเขตพื้นที่ข้อ ความที่เอียงในตำแหน่งที่แตกต่างกันได้ วิธีการนี้ทำงานได้รวดเร็ว เพราะเป็นการทำงานกับจุดเพียง ครั้งเดียว

การดำเนินงานเกี่ยวกับรูปร่างลักษณะ แบบเปิดหรือปิด ถูกนำมาใช้ใน [9], [10], [11] เพื่อ จัดกลุ่มจุดภาพรวมเข้าไปเป็นขอบเขตใหญ่ก่อนจัดประเภท การเบลอภาพโดยใช้วิธีการรันแรงค์ส มูทอัลกอริทึม(Run Length Smoothing (RLS)) ที่ใช้ใน [12], [13], [14], [15], [16] การดำเนินงานนี้ คล้ายกับการขยายรูปร่าง บางวิธีในกลุ่มนี้ใช้ได้กับเอกสารที่มีความเอียงและแผนผังที่ไม่ใช่รูปแบบแน่นอน [10], [13], [17] ในขณะที่วิธีการอื่น[11], [12], [14], [15] ไม่สามารถที่จะใช้ได้

การสกัดข้อความจากภาพเอกสารที่ซับซ้อน ใน [18], [19], [20], [21], [22], [23] ซึ่งใน หลายวิธีนี้กำหนดให้ตัวอักษรมีสีเดียวกันทั้งหมดและตัวอักษรจัดเรียงอยู่ตามแนวนอน โดยสีตัว อักษรและฉากพื้นต้องแยกออกจากกันได้ วิธีการที่เสนอโดย W.Y. Chen และ S.Y. Chen[6] ใช้การ ประมวลผลหน้าปกวารสารเทคนิค ซึ่งประกอบด้วยการลดคุณภาพสีโดยลดจำนวนสีลงตามการ แบ่งขอบเขตและสีพื้น ส่วนวิธีการที่ A.K. Jain และ B. Yu[19] สามารถสกัดข้อความโดยการ กำหนดค่าต่างๆ เพื่อนำมาทำการประมวลผลภาพ วิธีการนี้สามารถนำไปใช้กับภาพเว็บ หน้าปกซีดี และหนังสือ หรือ ภาพจากวิดีโอ ภาพหลายค่าจะเป็นภาพไบนารี(โฆษณา) ภาพเกรย์สเกล ภาพสี (ภาพเว็บในรูปแบบ GIF) หรือ ภาพสีสรร(วิดีโอ หนังสือ หรือ ปกซีดี) ซึ่งจะถูกรวบรวมไปเป็น ส่วนรายละเอียดที่แสดงด้านหน้าและฉากพื้นที่ซับซ้อน ดังนั้นการวิเคราะห์ส่วนประกอบที่ใกล้ชิด กันสามารถที่จะใช้กำหนดส่วนประกอบที่เป็นข้อความได้

วิธีการของ S. Messelodi และ C.M. Modena[20] มีการใช้ประโยชน์จากตัวแทน (significant) โดยวิธีการนี้ไม่ขึ้นกับความเอียงของบรรทัดข้อความ และสามารถที่จะดำเนินการรวม

ตัวอักษรเข้าด้วยกันอย่างถูกต้องรวมถึงบรรทัดข้อความที่ใกล้เคียงกันด้วย วิธีการนี้ตอนแรกจะตรวจสอบส่วนประกอบที่ใกล้เคียงกันในภาพไบนารีหนึ่งหรือหลายภาพ จำนวนของภาพขึ้นอยู่กับชนิดของภาพ(ไบนารี เกรย์สเกล หรือ สี) ตัวอย่าง สำหรับภาพไบนารี ส่วนประกอบที่ใกล้เคียงกันจะถูกตรวจสอบบนสองภาพที่จะให้ผลบวกและผลลบของความเข้มของข้อความที่สอดคล้องกับฉากพื้น บรรทัดข้อความจะถูกสกัดโดยโดยการใส่แบ่งแยกความแตกต่างเป็นขั้นลำดับในการแก้ปัญหา ภาพเกรย์สเกลของปกหนังสือเป็นตัวอย่างที่แสดงการทำงานของวิธีการนี้ ดังนั้นจะเห็นว่าวิธีการนี้สามารถจัดการกับประเภทเอกสารอื่นได้ดีเช่นเดียวกัน

การวิเคราะห์ภาพเว็บของ T. Park, D. Kim และ K. Chung [21] ในตอนแรกทำการลดจำนวนของสี ปริมาณของสีจะตรวจสอบได้จากส่วนประกอบที่ใกล้เคียงกันสำหรับแต่ละสีที่คงเหลือ หลังจากนั้นตัวอักษรของส่วนประกอบที่สัมพันธ์กันที่สนใจจะถูกเลือกโดยการทำให้บางลง และคำนวณทั้งความกว้างและความสูง ตัวอักษรประกอบเป็นชุดแบบเส้นร่างที่ต่อกันโดยมีความสัมพันธ์กับความกว้างที่กำหนดและอัตราส่วนระหว่างความกว้างและความสูง เส้นร่างชุดข้อมูลจะพบในขอบเขตตลอดแนวตัวอักษรและจะถูกใช้เป็นข้อมูลเข้าที่การเสนอขอบเขตของกลุ่มตัวอักษรที่ใกล้เคียงกันเข้าเป็นบรรทัด ซึ่งเหมือนกันกับวิธีการของ S. Messelodi และ C.M. Modena [20] โดยวิธีการนี้สามารถที่จะตรวจพบข้อความของทิศทางที่ไม่แน่นอนตลอดจนบรรทัดข้อความที่เป็นแนวโค้ง

### 2.7.1.2 วิธีบนลงล่าง

จุดเด่นทั่วไปของวิธีการบนลงล่างจะอธิบายดังข้างล่าง ซึ่งเป็นการดำเนินงานจากส่วนประกอบใหญ่ไปเป็นส่วนประกอบย่อย อย่างไรก็ตามวิธีการประเภทนี้ยังมีแนวความคิดที่แตกต่างอย่างหลากหลาย

- เริ่มต้นจากการประมวลผลภาพทั้งหมดโดยการแยกไปเป็นบล็อกเล็กๆ [24], [25]
- การแปลงส่วนประกอบใหญ่โดยการใส่ภาพทั้งหมดและจุดของภาพที่เปลี่ยนแปลงถูกนำมารวมเข้าด้วยกันเป็นกลุ่ม[26]
- การดำเนินงานสามารถเริ่มทำจากตัวแทนภาพอย่างหยาบหรือขอบเขตใหญ่ไปยังตัวแทนภาพแบบละเอียดหรือขอบเขตเล็ก[9], [27] โดยขอบเขตใหญ่จะถูกประมวลผลได้รวดเร็วในตัวแทนอย่างหยาบ และจะทำละเอียดอีกทีในตัวแทนที่ละเอียด

การวิเคราะห์โดยการฉายภาพเงาเป็นการนับจำนวนจุดไปตามแนวทิศทางใดทิศทางหนึ่ง เป็นวิธีแบบบนลงล่างที่รู้จักกันดีวิธีหนึ่ง ซึ่งใช้กับภาพไบนารีที่ไม่มีความเอียงและทำการตัดแบ่งภาพซ้ำๆจนมีขอบเขตที่เล็กลง โดยใช้พื้นฐานความสูงต่ำในการฉายภาพเงาในแนวตั้งและแนวนอน ซึ่งจะสอดคล้องกับตัวแบ่งระหว่างบล็อกที่เหลี่ยม การฉายเงาจุดถูกนำมาใช้ประโยชน์โดย

T.Akiyama และ N. Hagita [24] ขณะรูปร่างของบล็อกขอบเขตจะถูกใช้จากการนำเสนอของ J. Ha, R.M. Haralick และ I.T. Phillips [25] เพื่อเพิ่มความเร็วในการทำงาน

การเปลี่ยนแปลงภาพใหญ่ใช้ฟิวเจอร์คาร์บอ(Gabor filters) ที่ใช้กับตำแหน่งที่หลากหลายที่เสนอโดย A.K. Jain และ S. Bhattacharjee [26]. ใช้กับภาพเกรย์สเกลซึ่งแบ่งจุดจากการเปลี่ยนแปลงภาพกับอัลกอริทึมสแควร์เออเรอร์(squared-error) เพื่อตรวจจับขอบเขตตัวอักษรและไม่ใช่ตัวอักษร(ฉากพื้น รูป) วิธีการนี้ใช้ได้ดีกับเอกสารที่มีความเอียงแต่ใช้เวลาในการทำงานมาก

การแบ่งเอกสารภาพเกรย์สเกลแบบปริามิตีชั้นของเกาส์เซียน(Gaussian) ทั้งสี่ระดับได้ถูกนำมาใช้โดย L. Cinque, L. Lombardi และ G. Mazini [9] ใช้ปริามิตในการแสดงจุดเด่นทั้งสี่แผนที่ที่จุดเด่นคือ ค่าเฉลี่ย ค่าความแปรปรวน ค่าคงที่ ค่ามัธยฐาน การทำงานจะทำจากภาพรายละเอียดระดับต่ำไปยังรายละเอียดที่มากกว่า วิธีการนี้ไม่ขึ้นกับความเอียงและรูปแบบของเอกสาร

การตรวจจับพื้นที่ข้อความบนพื้นผิวของฉากพื้นทีเสนอ โดย J.L. Chen[27] เป็นแนวคิดพื้นผิวที่ใช้การสกัดรายละเอียดจุดเด่นด้วยกฎหน้ากาก(Laws' masks) การแบ่งประเภทแบบหยาบๆโดยใช้บล็อกขนาด 8\*8 พิกเซลและบล็อกมีการจัดเรียงที่ไม่ซ้อนทับกัน สามารถแบ่งได้เป็น ข้อความฉากพื้น และส่วนยังไม่ตัดสินใจ(fuzzy) (ความน่าจะเป็นขอบเขตระหว่างข้อความหรือฉากพื้น) และการแบ่งประเภทบล็อกของส่วนที่ยังไม่ตัดสินใจจะใช้การตรวจสอบที่ระดับจุดภาพ ทั้งการแบ่งแบบหยาบและแบบละเอียดใช้ตามแบบ HMMs(Hidden Markov Models) ที่ใช้ 4 ถึง 8 สถานะ(states) ประโยชน์ของ HMMs ที่เหนือกว่ากระบวนการเรียนรู้ของนิเวรอนเน็ตเวิร์กอื่นๆ คือ แต่ละโมเดลสามารถเรียนรู้ได้อย่างอิสระจากตัวอื่นๆ ซึ่งเมื่อเพิ่มข้อมูลประเภทใหม่เข้าไป HMMs จะถูกสร้างขึ้นและเรียนรู้ เฉพาะตัวอย่างของประเภทนั้น(ส่วน HMMs ตัวอื่นๆไม่มีการเรียนรู้ใหม่) วิธีการนี้ไม่ไว(sensitive)ต่อข้อความ รูปแบบเอกสาร หรือ ชนิดตัวร่าง

### 2.7.1.3 วิธีแบบผสมผสานและวิธีอื่นๆ

D. Wang และ S.N. Srihari [28] นำอัลกอริทึม RLS ของแนวคิดวิธีการต่างชั้นบน มาใช้กับภาพไบนารีเพื่อตรวจข้อมูลบรรทัดข้อความและข้อมูลที่ไม่ใช่ข้อความ(รูปภาพ) และใช้วิธีบนลงล่างคือ วิธีการตัดซ้ำตามแนว X-Y (recursive X-Y cuts) โดยจะมีการแยกบรรทัดข้อความที่ไม่ติดกันไปเป็นบล็อก วิธีการนี้ง่ายต่อการนำไปใช้และเวลาที่ใช้การทำงานค่อนข้างเร็ว แต่สามารถใช้วิเคราะห์ได้ในรูปแบบที่เป็นสี่เหลี่ยมเท่านั้นและต้องแปลงภาพเกรย์สเกลให้เป็นไบนารีก่อน

การคิดแปลงการแบ่งส่วนประกอบแบบแบ่งแล้วรวบรวม(split-and-merge)ของภาพเอกสารเกรย์สเกลไปเป็นขอบเขตที่มีเนื้อหาเดียวกัน เช่นการแสดงของต้นไม้สี่เหลี่ยมที่ถูกพัฒนาโดย J. Liu, Y.Y. Tang และ C.Y. Suen [29] ทำการตัดแบ่งและการรวมส่วนประกอบเข้าด้วยกันเป็นการดำเนินงานที่เวลาเดียวกัน ถ้าบางขอบเขตไม่เป็นเนื้อหาเดียวกัน โดยจะถือแบ่งเป็นสี่เหลี่ยมย่อยสี่ส่วนโดยการใช้ค่าเรอร์โธร์ของการฉายเงา ถ้าสองพื้นที่ที่ติดกัน(อาจจะเป็นส่วนประกอบระดับเดียวกันหรือไม่เป็นก็ได้) มีรายละเอียดแบบเดียวกันจะถูกนำเข้ามารวมกัน ค่าเฉลี่ยและความแปรปรวนของความเข้มข้นของจุดในแต่ละขอบเขตที่ถูกใช้ในการทำตัดสินใจเกี่ยวกับการรวมและการแบ่ง รูปแบบเอกสารอาจมีรูปแบบที่ไม่แน่นอน แต่การประมาณและแก้ไขความเอียงเป็นสิ่งเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำเป็น วิธีการนี้สามารถนำไปใช้กับภาพของวารสาร เช็ครณาการที่มีฉากพื้นที่ที่ซับซ้อน และเอกสารที่เป็นตาราง เหมือนกับวิธีการการของ S.N. Sribri, T. Hong และ G. Srikantan [30]

วิธีการใน [31] การสกัดข้อความจากภาพที่ซับซ้อนของปกหนังสือและวารสารโดยใช้การวิเคราะห์แบบผสมผสาน เทคนิคของวิธีการบนลงล่างจะเป็นการตัดแบ่งภาพแบบซ้ำๆไปเป็นบล็อกสี่เหลี่ยมเล็กๆ และจะหยุดการแบ่ง เมื่อจุดตัวอย่างมีสีน้อยกว่าสองสีที่แตกต่างกันภายในบล็อก บล็อกที่มีรูปข้อมูลเดียวกันถูกนำมาพิจารณาตลอดทั้งหมดของฉากพื้นชนิดเดียวกัน ขณะที่บล็อกที่ไม่มีรายละเอียดเหมือนกันมีสีที่แตกต่างกันอย่างน้อยสองสีที่สอดคล้องกับข้อความ วิธีการข้างขึ้นบนใช้ตรวจหาขอบเขตที่มีรายละเอียดเหมือนกันที่มีรูปร่างไม่แน่นอนโดยใช้คุณสมบัติของการเจริญเติบโตของขอบเขต(region growing) ผลลัพธ์ของทั้งสองวิธีเมื่อนำมารวมกันเพื่อลำดับถูกต้องที่หลากหลายเพื่อให้ได้ขอบเขตข้อความและที่ไม่ใช่ข้อความโดยสมมุติฐานว่าข้อความจัดเรียงกันตามแนวนอน

### 2.7.2 การแบ่งประเภทขอบเขตเอกสาร

วิธีจำนวนมาก [2], [19], [20], [23], [24], [32], [33], [34], [35] ใช้จุดเด่นของส่วนประกอบที่สัมพันธ์กันที่จะแบ่งส่วนข้อความและรูปภาพในภาพไบนารี จุดเด่นนี้ประกอบด้วยขนาดของส่วนประกอบที่ติดกัน และส่วนที่ใกล้ชิดกัน แนวการเรียง ส่วนเริ่มต้นที่อยู่ใกล้กัน(proximity) และสิ่งที่ต่อกันยาวขึ้น(elongation)

รูปร่างที่ได้จากการรันแรงจูงใจจุดดำและจุดขาวที่เห็นภายในแต่ละส่วนประกอบที่ถูกแบ่งใน [16], [28] การคำนวณค่ารันแรงจูงใจจากสี่ทิศทาง (แนวนอน แนวตั้ง เอียงซ้าย เอียงขวา) การแบ่งประเภทโดยใช้ต้นไม้ตัดสินใจ(decision tree) ที่นำเสนอใน [36] และจุดเด่นพื้นผิวของกระเบื้องขาวที่ใช้ใน [37]

วิธีการ [38] เป็นการแบ่งประเภทโดยใช้วิธี คออสคอรีเลชัน (cross-correlation) ระหว่างเส้นสแกนที่ใกล้กัน และ การอ้างอิงจากระยะห่างระหว่างบรรทัด การรวมจุดเด่นของรันแรงจูงใจและคออสคอรีเลชันระหว่างจุดที่ใช้ใน [3]

วิธีการ [12], [13] ใช้การฉายเงาทางแนวตั้งเพื่อแบ่งข้อความและส่วนที่ไม่ใช่ข้อความออกจากกัน(การประมาณข้อความ) โดยการกระจายของจุดขาวและจุดดำนั้นช่วยให้รู้ถึงลักษณะว่าควรจะเป็นกราฟิกหรือรูปภาพ (ซึ่งตัวอักษรจะอยู่ห่างกันพอสมควรในเอกสารที่เป็นรูปแบบ ส่วนจดหมายหรือเอกสารที่ไม่ใช่รูปแบบตัวหนังสืออาจจะอยู่ติดกันมาก)

การคำนวณระดับสีเทาที่ภาพเกรย์สเกลซึ่งทำให้รู้ความถึงแตกต่างระหว่าง ฉากพื้น ข้อความ กราฟิก และรูปภาพ ใน [9]

การจัดแบ่งประเภทโดยใช้วิธีการคำนวณแบบละมุนที่อธิบายใน [11], [15], [17], [39] ปกติใช้นิวรอลเน็ตเวิร์ก(ในกรณีที่มีเงื่อนไขมาก ก็จะมีขั้นตอนหลายระดับมากขึ้น) ตอนแรกจะถูกให้เรียนรู้ก่อน เริ่มจากตัวอย่างข้อความ กราฟิก และรูปภาพ ที่ต้องใช้ในการจัดแบ่งประเภท ขนาดเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

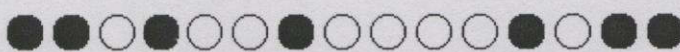
บล็อกและจุดเด่นของรันเรงค์เป็นข้อมูลเข้าของนิวรอลเน็ตเวิร์ก[15] ขณะที่ในระดับต่ำของส่วนประกอบชุดเวฟเล็ต(wavelet packet components) ซึ่งจะถูกลำดับจากหน้าต่ำเล็ก ๆ ที่จุดเด่น(feature) จะถูกใช้สำหรับการแบ่งประเภทโดยนิวรอล[10], [17] เพื่อที่จะทำการแบ่งประเภทของแต่ละหน้าต่ำได้อย่างถูกต้อง การรวมค่าสัดส่วนความน่าจะเป็น(fuzzy) เพื่อใช้ตัดสินใจจากข้อมูลที่ได้มาจากหน้าต่ำใกล้เคียงกันที่เสร็จแล้ว การตัดสินใจเป็นการรวมการวัดที่มาจากความละเอียดภาพและที่อยู่ภายในมาตรวัด แนวคิดที่ [11] นิวรอลเน็ตเวิร์กเรียนรู้จากชุดหน้าต่ำเล็ก ๆ ซึ่งจะแยกความแตกต่างระหว่างข้อความ ฉากพื้น ภาพวาดลายเส้นและรูปได้ดีที่สุด การรวมเข้าด้วยกัน(convolving) ของหน้าต่ำนี้กับจุดเด่นของพื้นผิวที่ได้จากภาพนำเข้า ซึ่งถูกใช้ในการแบ่งประเภทของแต่ละจุดภาพ โดยนิวรอลเน็ตเวิร์กไปเป็นหนึ่งในสามประเภท(ข้อความ+ภาพวาดลายเส้น รูป ภาพฮาฟโทนและฉากพื้น) ขอบเขตที่ถูกจัดไปเป็นประเภทแรก เป็นการทำแนวสูงต่ำกับค่ากลางซึ่งถูกกำหนด และถูกแบ่งโดยขนาดของส่วนประกอบที่เกี่ยวข้อง ซึ่งวิธีการนี้คงทนกับภาษาต่างๆ และสามารถแยกความแตกต่างของข้อความในแต่ละภาษา เช่น ภาษาอังกฤษ หรือภาษาจีน วิธีใน[10], [11], [17] จะแตกต่างจากวิธีอื่นๆ เพราะการแบ่งประเภทจุด/หน้าต่ำ จะทำก่อนขั้นตอนการแบ่งส่วนประกอบไปเป็นขอบเขต

## 2.8 การแบ่งประเภทขอบเขตเอกสาร

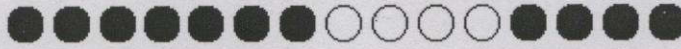
โดยในส่วนนี้จะอธิบายการทำงานอย่างคร่าวๆของบางวิธีการแบ่งขอบเขตส่วนประกอบเอกสารที่กล่าวถึงในหัวข้อ 2.7 ตามต่อไปนี้

### 1. วิธีการรันเรงค์สมูทอัลกอริทึม (run length smoothing)

เป็นวิธีการที่ใช้การเชื่อมต่อระหว่างจุดในแนวตั้งหรือแนวนอน เพื่อเชื่อมตัวจุดเข้าด้วยกัน หรือเรียกว่าวิธีการทำเบลอ โดย กำหนดค่าขึ้นมาเพื่อเป็นค่าระยะห่าง โดยกำหนดค่าระยะห่างของจุดค่าสองจุดทั้งในแนวตั้งหรือแนวนอน โดยมีเงื่อนไขว่าจะเพิ่มจุดเข้าระหว่างจุดค่าสองจุดที่มีระยะห่างกันน้อยกว่า ค่าที่กำหนด ตัวอย่างในรูปที่ 2.5 เป็นจุดภาพก่อนการทำงาน รูปที่ 2.6 เป็นชุดของจุดภาพหลังจากการทำงานรันเรงค์สมูท ด้วยค่าที่กำหนดไว้เป็น 3



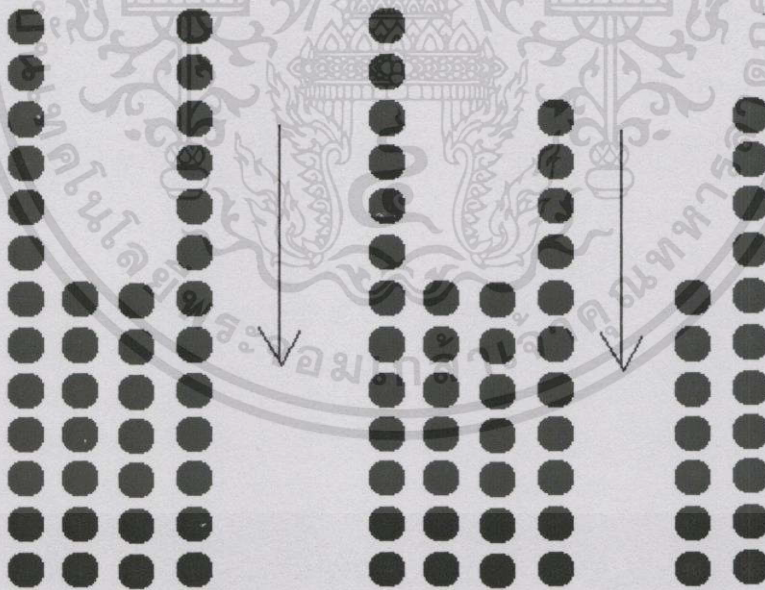
รูปที่ 2.5 ชุดจุดภาพก่อนทำการรันเรงค์สมูท



รูปที่ 2.6 ชุดจุดภาพหลังจากทำการรันเรจค์สมูท ด้วยค่าที่กำหนดไว้คือ 3

## 2. วิธีการตัดซ้ำแกน X-Y (recursive projection X-Y cut)

วิธีการตัดซ้ำแกน X-Y มีพื้นฐานมาจากการฉายเงาภาพ ซึ่งลักษณะการทำงานจะเหมือนกับการเอาตาที่มีลูกแก้วอยู่อย่างกระจัดกระจาย เมื่อเอียงตานั้นไปข้างใดข้างหนึ่ง จะพบว่าลูกแก้วจะกองอยู่ที่ข้างที่เอียงลงไป นั่นคือหลักการทำงานของการฉายภาพเงา โดยลักษณะภาพเงาของเอกสาร ไบนารีที่ได้ จะมีลักษณะเป็นช่วงความถี่ของจุดดำ หรือการฮิต โดแกรมจุดดำ ซึ่งในการตัดแบ่งขอบเขตนั้นจะกำหนดค่าที่ยอมรับขึ้นมา เมื่อช่วงใดที่มีค่าน้อยกว่าที่กำหนดไว้ ช่วงนั้นจะถูกตัดแบ่งออกไป โดยการทำงานแบบตัดซ้ำแกนนั้น อ้างขอบเขตในลักษณะแผนผังต้นไม้ โดยโหนดรากจะแสดงเป็นภาพเอกสารทั้งหน้า โดยโหนดลูกของแต่ละโหนดนั้นจะแสดงส่วนประกอบย่อยของโหนดเมื่อนั้นๆ ดังรูปที่ 2.7 แสดงการตัดตามแนวตั้ง โดยมีที่ลูกศรคือตำแหน่งที่ตัดแบ่งส่วนประกอบ



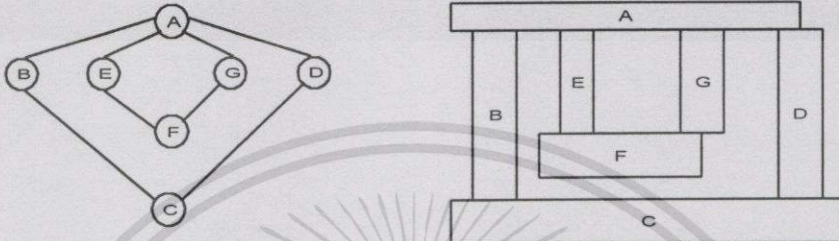
รูปที่ 2.7 การฉายเงาจุดดำในแนวตั้ง และช่วงที่จะแบ่ง

## 3. วิธีการเดินวนรอบขอบ (contouring)

วิธีการนี้เป็นการเดินวนรอบขอบส่วนประกอบโดยใช้จุดภาพที่ใกล้ชิดกันในทิศทาง 4 ทิศ หรือ 8 ทิศ แบบใน 2.4 รหัสแนว ซึ่งการเดินวนรอบขอบส่วนนี้ ทำการเดินรอบขอบส่วนประกอบจนวนกลับมาที่จุดเริ่มต้น ผลลัพธ์ที่ได้คือขอบเขตของส่วนประกอบ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4. วิธีการแบบไวท์ไทน์ (white tile)

วิธีการนี้คือการนำส่วนที่เป็นพื้นฉากเป็นตัวกำหนดขอบเขตของส่วนประกอบภาพแบบไบนารี โดยการทำงานจะสร้างบล็อกสี่เหลี่ยมที่มีขนาดเท่ากับช่องว่างสีขาวหรือฉากพื้น โดยข้อมูลบล็อกสี่เหลี่ยมช่องขาวจะแสดงด้วยกราฟในรูปที่ 2.8 เพื่อใช้ในการบอกขอบเขตส่วนประกอบในรูปแบบปิด



รูปที่ 2.8 กราฟที่ใช้แสดงการจัดเรียงไวท์ไทน์

### 2.9 เปรียบเทียบวิธีการทำงาน

ตารางการเปรียบเทียบวิธีการของการทำงานแนวคิดต่างๆ ได้มาจากรายงานเชิงเทคนิค ของ Okun O, Doermann D & Pietikäinen M [40] ซึ่งผลลัพธ์จะมีอยู่ 2 ตาราง โดยในตารางที่ 2.1 “อ้างอิง” จะอ้างอิงวิธีการ “ชนิดเอกสาร” อ้างอิงประเภทของภาพเอกสารที่ถูกนำมาประมวลผลโดยวิธีการนั้น “ชนิดภาพ” ค่าที่เป็นไปได้มี B, G และ C คือแทน ภาพไบนารี ภาพเกรย์สเกล และ ภาพสี “ฉากพื้น” ค่าที่สามารถเป็นได้คือ U (พื้นขาวหรือดำ) T (พื้นผิวที่เป็นลวดลาย) หรือ C (สี) “แผนผัง” ค่าที่เป็นไปได้มี R (แผนผังสี่เหลี่ยม) M (แผนผังแมนฮัตตัน) หรือ A (แผนผังที่ไม่มีรูปแบบ) “ความเอียง” บอกว่าวิธีการนี้ใช้งานกับเอกสารที่มีความเอียงบ้างได้(Y) หรือ ไม่ได้(N) “ชุดทดสอบ” และ “%ความถูกต้อง” อ้างถึงชุดขนาดที่นำมาทดสอบและความถูกต้องที่ได้รับจากวิธีการนั้นๆ “-” หมายความว่าจุดเด่นที่ให้มาไม่ได้ถูกอ้างอิงในบทความต้นฉบับ สัญลักษณ์ “+” แสดงถึงผลลัพธ์ที่ไม่ถูกระบุโดยรูปเดียว แต่มีหลายมาตรฐานที่นำมาใช้เช่น การแบ่งเป็นส่วน (fragmentation) และอัตราการรวมเข้าด้วยกัน พวกนี้เป็นกรณีมูลค่าการทำงานของวิธีการ แต่บางวิธีการง่ายหรือยากต่อการตัดแปลงแก้ไข เพราะฉะนั้นอาจจะมีการคิดแปลงซึ่งทำให้ประสิทธิภาพการทำงานเปลี่ยนแปลงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 แสดงคุณสมบัติที่สำคัญของวิธีการต่างๆ ที่ใช้ในการวิเคราะห์แผนผังของเอกสาร

อ้างอิง	ชนิดเอกสาร	ชนิดภาพ	ฉากพื้น	แผนผัง	ความเอียง	ชุดทดสอบ	%ความถูกต้อง
[2]	วารสาร	B	U	A	Y	-	-
[3]	วารสาร	B	U	R	N		99-100
[4]	วารสาร	B,G	U	M	N	-	-
[6]	วารสาร หนังสือพิมพ์ จดหมายข่าว โฆษณา	B	U	A	Y	40	+
[8]	วารสาร	B	U	A	Y	-	-
[9]	วารสาร	G	U	A	Y	100	-
[10]	วารสาร	G	U	A	Y		
[11]	วารสาร	G	U	M	N	-	-
[12]	วารสาร	B	U	R	N	-	-
[13]	วารสาร	B	U	R	Y	30	92-97
[14]	หนังสือพิมพ์	B	U	M	N	100	96
[15]	วารสาร	B	U	R	N	50	98.18-99.61
[16]	วารสาร หนังสือพิมพ์	B	U	R	N	100	-
[18]	ปกวารสาร	C	C	A	N	100	95.2-98
[19]	โฆษณา ภาพหน้าเว็บ ปกหนังสือ ภาพจากวิดีโอ	B, G, C	U, C	A	N	26 54 30 6,952	99.2 97.6 72.0 94.7
[20]	ปกหนังสือ	G	U	A	Y	100	91.2
[21]	หน้าเว็บ	C	C	A	Y	200	88.8-92
[23]	ภาพหน้าเว็บ	C	C	A	N	262	90
[25]	วารสาร	B	U	R	N	150	
[26]	หนังสือพิมพ์	G	U	A	Y	-	-
[28]	หนังสือพิมพ์	B	U	R	N	-	78-100
[29]	วารสาร แบบฟอร์ม เช็คราชการ	G	U,T	A	N	-	-
[31]	หนังสือและปกวารสาร	C	C	A	N	16	
[32]	วารสาร นามบัตร รายงานเชิงวิชาการ	B	U	A	N	30	93-100
[33]	โฆษณา จดหมาย ของจดหมาย	B	U	A	Y	150	-
[34]	วารสาร	B	U	M	N	150	19.1-99.4
[35]	ปกหนังสือและปกซีดี ภาพจากวิดีโอ	C	C	A	N		
[36]	วารสาร	B	U	R	N	979	97
[37]	ไม่ทราบ	G	T	A	Y		
[38]	วารสาร	B	U	R	Y		
[41]	วารสาร หนังสือพิมพ์	B	U	A	Y	114	+
[42]	วารสาร	G	U	A	N		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในตาราง 2.2 ประกอบด้วย “เวลา” บอกถึงเวลาที่ใช้ในการทำงานหน่วยเป็นวินาที ที่ทำงานบนเครื่องคอมพิวเตอร์รุ่นอะไร “ชนิด” “ขนาดภาพ” บอกขนาดของภาพ โดย “ความละเอียด” บอกถึงความละเอียดของภาพ โดยขนาดภาพเป็นจุดต่อหน้า และความละเอียดเป็นจุดต่อนิ้ว ส่วนรายละเอียดอื่นๆจะเหมือนกับในตารางที่ 2.1

ตารางที่ 2.2 แสดงเวลาที่ใช้ในการทำงานสำหรับวิธีการวิเคราะห์เอกสาร

อ้างอิง	ชนิดเอกสาร	เวลา	ขนาดภาพ	ความละเอียด	ชนิด
[6]	วารสาร หนังสือพิมพ์ จดหมายข่าว โฆษณา	0.55	810x1151	100	HP 9000/735
[6]	วารสาร หนังสือพิมพ์ จดหมายข่าว โฆษณา	1.2	1215x1727	150	HP 9000/735
[6]	วารสาร หนังสือพิมพ์ จดหมายข่าว โฆษณา	6.5	2431x3455	300	HP 9000/735
[8]	วารสาร	20	1278x1746		Sun Sparc2
[10]	วารสาร	22	A4	300	Sun Sparc 20
[11]	วารสาร	60-85	264x332 ถึง 780x1080	100	Sun Sparc 20
[14]	หนังสือพิมพ์	~ 9	6592x9890		Pentium 350 II
[16]	วารสาร หนังสือพิมพ์	2.37	A4	300	Sun Sparc
[18]	ปกวารสาร	95	2000x2679	-	Pentium 100
[18]	ปกวารสาร	180	1719x2476	-	Pentium 100
[19]	โฆษณา	0.15	548x769	150	Sun Ultra-Sparc I
[19]	ภาพหน้าเว็บ	0.11	385x234	-	Sun Ultra-Sparc I
[19]	ปกหนังสือ	0.4	763x537	50	Sun Ultra-Sparc I
[19]	ภาพจากวีดีโอ	0.09	160x120		Sun Ultra-Sparc I
[20]	ปกหนังสือ	0.01- 2.86	512x512		Pentium 200 Pro
[24]	วารสาร หนังสือพิมพ์	80	A4	400	-
[25]	วารสาร	~ 2	Letter-sized	300	Sun Sparc 10
[26]	หนังสือพิมพ์	~ 120	512x512	75	Sun Sparc 2
[28]	หนังสือพิมพ์	2.6	-	100	Sun 3/60
[28]	หนังสือพิมพ์	9.5		200	Sun 3/60
[31]	ปกหนังสือและวารสาร	31.31	1600x2400	200	Sun Ultra Sparc 5/10
[32]	วารสาร นามบัตร รายงานเชิงเทคนิค	4.8			PC 486
[33]	โฆษณา ซองจดหมาย จดหมาย	306.9- 563.6	< A4	300	Sun Sparc IPX
[34]	วารสาร	1.3	A4	300	SG Indigo
[35]	ปกหนังสือและซีดี ภาพจากวีดีโอ	5.5-6	256x256		Sun Sparc 20
[38]	วารสาร	0.9-1.9	A4	300	Sun Sparc
[41]	วารสาร หนังสือพิมพ์	2.93	1053x1149	90	Pentium 200 Pro
[41]	วารสาร หนังสือพิมพ์	5.37 – 7.03	2592x3300- 3114x3554	300	Pentium 200 Pro

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.10 วิเคราะห์วิธีการทำงาน

ในส่วนนี้จะอธิบายแนวคิดเกี่ยวกับเทคนิคการวิเคราะห์แผนผังเอกสาร ทั้งหมดของบางวิธี [2], [19], [20], [29], [33], [35], [43] ขั้นตอนวิธีการทั้งหมดนำไปใช้ในการแบ่งประเภทเอกสารต่างๆ และชนิดของภาพ ซึ่งแสดงให้เห็นว่าไม่มีการแก้ปัญหาแบบธรรมดาที่ใช้ในปัญหาการแบ่งส่วนประกอบเอกสารและแบ่งประเภทส่วนประกอบ เพราะว่างานมีขอบเขตกว้าง ยากต่อการจัดการ มีเงื่อนไขในการปรับแต่งมาก และทำนายผลลัพธ์ได้ยาก

ทั้งหมดต้องมีกระบวนการเฉพาะเพื่อพัฒนาวิธีการสำหรับแยกตัวอักษรจากภาพสีหรือภาพเกรย์สเกล วิธีการโดยทั่วไปจะใช้กฎที่ได้จากการเรียนรู้สำหรับการจัดแบ่งประเภทส่วนประกอบ จุดเด่นส่วนใหญ่ของวิธีการคือจำนวนเงื่อนไขที่พอที่จะทำให้เกิดความมั่นใจ เช่น

1. สีของตัวอักษรต้องเป็นสีเดียวกันเท่านั้น
2. สีจะเป็นตัวแบ่งข้อความและพื้นฉาก
3. ตัวอักษรควรจะเป็นรูปแบบที่เรียงต่อกันเป็นบรรทัดในแนวนอน (อย่างน้อยเงื่อนไขที่ไม่จำเป็นในวิธีการต่อไปนี้ [20], [21], [44], [45])

การแยกข้อความจากภาพสีไม่จำเป็นต้องใช้ OCR ในการตรวจสอบความถูกต้องของผลลัพธ์ที่สกัดแยกได้ ซึ่งที่มีการแบ่งขอบเขตและประเภทของส่วนประกอบก่อนเพื่อให้การทำงานของ OCR ได้ลดจำนวนกฎเกณฑ์ที่ใช้และเป็นการเพิ่มความถูกต้องของการทำงาน

วิธีการวิเคราะห์แผนผังกับภาพที่เป็นวารสาร หนังสือพิมพ์ หรือเอกสารที่เป็นข้อความอื่นๆ สามารถแบ่งได้เป็นสองกลุ่มคือ กลุ่มแรกต้องทำการประมาณและแก้ไขความเอียงก่อนการวิเคราะห์แผนผัง วิธีการชนิดนี้จะเป็นแผนผังชนิดสี่เหลี่ยมหรือแมนฮัตตัน เพราะการแก้ไขความเอียงก่อนจะทำให้การวิเคราะห์แผนผังได้ง่ายขึ้น อย่างไรก็ตามความผิดพลาดที่เกิดจากการหาความเอียงจะลดความถูกต้องของวิธีการวิเคราะห์แผนผัง ถ้าวิธีการนั้นไม่สามารถจัดการกับขอบเขตที่เอียงได้ ส่วนอีกประเภทหนึ่งจะจัดอยู่ในกลุ่มที่จะทำการแบ่งขอบเขตส่วนประกอบและจัดประเภทส่วนประกอบก่อน แล้วจึงมาประมาณความเอียงเฉพาะขอบเขตที่เป็นข้อความเท่านั้น ปกติวิธีการชนิดนี้จะใช้กับแผนผังที่มีความซับซ้อนและมีรูปแบบที่ไม่แน่นอน หรือภาพเอกสารที่มีหลากหลายความเอียง ในกรณีสุดท้าย การประมาณความเอียงจะถูกนำไปใช้กับบล็อกข้อความทุกบล็อก บางครั้งวิธีการนี้ต้องใช้กระบวนการดำเนินงานที่พิเศษ เพราะว่าการประมาณความเอียงจะไม่ทำในส่วนที่ไม่ใช่ข้อความ อย่างไรก็ตามการวิเคราะห์แผนผังยังเป็นงานที่ยากกว่าการประมาณความเอียงและไม่ง่ายที่จะทำให้ได้ถูกต้องเหมือนการประมาณความเอียง นั้นหมายความว่าความผิดพลาดที่เกิดจากการวิเคราะห์แผนผังที่ซับซ้อนจะพบได้บ่อยมากกว่าการประมาณค่าความเอียง คือการหาผลลัพธ์หลังจากการประมาณความเอียง ตัวเลือกสุดท้ายขึ้นอยู่กับนำไปใช้และความซับซ้อนของปัญหา

การกำหนดรายละเอียดภาพจะแตกต่างกันในแอปพลิเคชันที่ต่างกัน สำหรับตัวอย่างภาพรายละเอียดสำหรับภาพหน้าเว็บ[23] แต่ต้องการความละเอียดของภาพสูงสำหรับภาพวารสารหรือหนังสือพิมพ์ การประมวลผลที่ความละเอียดต่ำจะใช้เวลาในคำนวณได้รวดเร็วกว่า แต่เอกสารที่ต้องการรายละเอียดที่แม่นยำจะทำให้เกิดความผิดพลาดและจะทำให้ขั้นตอนการทำงานต่อไปได้ผลลัพธ์ที่ไม่ดีเช่น การแบ่งส่วนประกอบที่เป็นข้อความ หรือ OCR ทางเลือกของความละเอียดที่เหมาะสมไม่ใช่งานที่ง่าย เพราะความละเอียดที่ไม่เหมาะสมจะทำให้เกิดขั้นตอนประมวลผลเอกสารมากมาย

โครงสร้างข้อมูลเป็นส่วนที่สำคัญที่สุดสำหรับวิธีการวิเคราะห์แผนผัง เพราะโครงสร้างข้อมูลที่ดียิ่งสามารถลดเวลาในการทำงานและกำหนดความยากง่ายในการเข้าถึงข้อมูลในการทำงานหลังจากการวิเคราะห์แผนผัง ตัวอย่างโครงสร้างที่ยืดหยุ่นและมีประสิทธิภาพ [34], [46] ใช้วิธีการ BAG, [5], [6], [37] ใช้วิธีการ while tile, [29] ใช้วิธีการ quadtrees และ [3], [10], [17], [27] ใช้วิธีการ square block tessellation โดยวิธีทั้งหมดเป็นการแสดงข้อมูลที่ตี ไม่เพียงแต่เข้าถึงข้อมูลได้ง่ายแต่ยังเป็นอิสระต่อความเอียงหรือแผนผัง

ประเด็นนี้มีเพียงสองหรือสามแอปพลิเคชันที่ใช้เทคนิคการคำนวณแบบละมุน (soft computing techniques) มาใช้ในการวิเคราะห์แผนผังเอกสาร [10], [11], [17] ในบทความ [15] อธิบายผลงานวิจัยที่ใช้นิเวศน์เน็ตเวิร์กสำหรับการแบ่งประเภทส่วนประกอบเอกสารอย่างครอบคลุม อย่างไรก็ตามจำนวนของการสอนให้จดจำตัวอย่างมีขนาดได้ใหญ่มาก (มากถึง 1,000,000 ใน [11]) และยังคงคลุมเครือ ถึงกระนั้นตัวอย่างที่มากยังจำเป็นเพราะว่าความคลาดเคลื่อนทั้งระหว่างคลาสและระหว่างภายในคลาสมักมีมาก

ในบทความ [11], [27], [29] แสดงตัวอย่างประสิทธิภาพของวิธีการวิเคราะห์แผนผังเมื่อค้นงานทางแก้ไขงานที่ยากมากของการสกัดข้อความจากพื้นที่ที่มีลวดลายซับซ้อน [27], [29] และการแบ่งภาษาที่แตกต่างกันภายในภาพเดียวกัน [11] นักวิจัยในด้านการวิเคราะห์ภาพเอกสารยังคงไม่เอาใจใส่ในด้านนี้

การใช้ฉากพื้นมาช่วยในการแบ่งได้มากที่สุด เพราะว่าปกติแล้วจะเป็นตัวแบ่งขอบเขตระหว่างขอบเขตที่แตกต่างกัน ความละเอียดของภาพสามารถถูกลดลงถึง 75-100 จุดต่อนิ้วก่อนทำการประมวลผล เป็นความคิดที่ดีถ้ารวมการแบ่งส่วนประกอบเอกสารกับการแบ่งประเภทส่วนประกอบเอกสารเข้าไว้ด้วยกัน จะทำให้ประหยัดเวลาในการทำงานลงเพราะข้อมูลจะถูกนำไปใช้เพียงหนึ่งครั้งเท่านั้น

เวลาที่ใช้ในการประมวลผลและความถูกต้องเป็นจุดเด่นที่สำคัญของวิธีการวิเคราะห์แผนผังเอกสาร เวลาที่ใช้ในการทำงานจะมีช่วงความแตกต่างที่มากเริ่มตั้งแต่ประมาณ 1 วินาทีไปจนถึงหลายนาทีต่อหนึ่งภาพ ซึ่งจะพบว่าวิธีการที่ทำงานได้เร็วเช่น [6], [41] ซึ่งจะทำการประมาณความเอียงไปพร้อมกัน และไม่ขึ้นกับแผนผังเอกสาร ใช้เวลาในการประมวลผลภาพไปบนรีความละเอียดเอกสารเป็นเอกสารที่ส่งวนเวลาสำหรับการแข่งขันเพื่อการศึกษาเท่านั้น เหมือนญาติเหินาไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

300 จุดต่อนี้ประมาณ 5.5 – 7 วินาที และวิธีการที่เร็วกว่าคือ [34], [38], [46] ใช้เวลาทำงานเพียง 1-2 วินาทีในงานเดียวกัน แต่ในกรณีนี้ไม่มีทั้งการแก้ไขความเอียงของเอกสารซึ่งเป็นสิ่งจำเป็นหรือแม้แต่แผนผังที่เป็นรูปแบบไม่แน่นอนก็ทำไม่ได้ การคำนวณด้วยความเร็วเป็นสิ่งที่สำคัญมากสำหรับการเรียกค้นข้อมูลข่าวสาร เช่นข้อความที่สกัดแยกจากภาพหน้าเว็บหรือจากวิดีโอ มากกว่าที่ต้องการจากการวิเคราะห์ของภาพวารสารหรือหนังสือพิมพ์ ซึ่งขึ้นอยู่กับว่าต้องการการตอบสนองทันทีหรือไม่ ในกรณีหลัง ผู้ใช้จะมีอิสระในการแก้ไขผลลัพธ์ได้นอกเหนือจากการกำหนดค่าตัวแปรใหม่ๆในกรณีที่เป็นรูปแบบ การประมวลผลที่สามารถปรับเปลี่ยนค่าใหม่ได้ มีความจำเป็นถ้าผลลัพธ์ก่อนหน้าไม่เป็นที่น่าพอใจ ดังนั้นการกำหนดตัวแปรที่เหมาะสมจะขึ้นอยู่กับรายละเอียดของแต่ละแอปพลิเคชัน

เป็นการค่อนข้างยากที่จะเปรียบเทียบความถูกต้องของวิธีการที่แตกต่างกัน เพราะการทดสอบจะใช้ชุดข้อมูลที่แตกต่างกันและเงื่อนไขตั้งต้นที่ต่างกัน ในหลายกรณี ยกเว้นการวิเคราะห์ภาพที่เป็นข้อความล้วนๆ ผลลัพธ์จะถูกประเมินคุณค่าจากสายตา และประเมินคุณค่าจากประสิทธิภาพที่ตั้งไว้บางทีไม่ได้ นอกจากนี้ยังไม่มีข้อกำหนดที่เป็นหนึ่งเดียวในการวัดความถูกต้องสำหรับตัวอย่าง การวัดอาจใช้อัตราส่วนของจำนวนขอบเขตที่แบ่งได้ถูกต้องต่อหน้า ในการแบ่งส่วนประกอบ และ จำนวนขอบเขตที่ถูกจัดประเภทต่อหน้า หรือถูกกำหนดโดยจำนวนของกรณีที่ขอบเขตถูกแบ่งหรือถูกรวมเข้าด้วยกันผิดพลาด การแบ่งขอบเขตที่มีรายละเอียดข้อมูลประเภทเดียวกันออกเป็นส่วนย่อยๆหลายส่วน ซึ่งง่ายต่อการแก้ไขในขั้นต่อไปมากกว่าการที่รวมกับขอบเขตที่ไม่ใช่ประเภทเดียวกันมารวมไว้ด้วยกันซึ่งยากต่อการตรวจสอบ ความถูกต้องจะมีค่าจาก 70% ไปจนถึงเกือบ 100% สำหรับวิธีการที่แตกต่างและเอกสารที่หลากหลายประเภท ความถูกต้องจะมีค่าสูงถ้าเป็นวารสารและหนังสือพิมพ์ และจะมีค่าต่ำเมื่อเป็นเว็บสื่และภาพวิดีโอ

รูปแบบที่ถูกต้อง(Ground-truthing) และ การวัดประสิทธิภาพของวิธีการวิเคราะห์แผนผังเอกสารยังไม่มีทางออกที่สมบูรณ์ ทุกวิธีการจะขึ้นขึ้นในความเอียงที่ไม่เปลี่ยนแปลง แต่เป็นการยากมากที่จะพิสูจน์ความจริงในประสิทธิภาพของวิธีทั้งหมด โดยใช้ชุดข้อมูลขนาดใหญ่ของภาพที่เอียง เพราะรูปแบบที่ถูกต้องปกคิจะถูกสร้างด้วยภาพที่ตั้งตรงปราศจากความเอียง ระบบวัดประสิทธิภาพความถูกต้องจะประเมินคุณค่าจากผลลัพธ์ที่ได้จาก OCR [47] (ซึ่งตรงนี้จะไม่ชัดเจนว่าความผิดพลาดนั้นเกิดจากการแบ่งส่วนประกอบที่ไม่ดีหรือการจำจำตัวอักษร ไม่ถูกต้อง) หรือ ทั้งหมดจำเป็นที่จะต้องแบ่งรูปแบบที่ถูกต้องสำหรับแต่ละภาพที่เอียง [48], [49], [50], [51] ในกรณีหลัง ภาพที่ถูกสแกนเป็นไปได้อาจมีความเอียง และหลังจากที่สร้างรูปแบบที่ถูกต้องสำหรับพวกเอกสาร การทำงานแบบอัตโนมัติและการทำรูปแบบที่ถูกต้องให้ถูกต้องกินเวลาในการทำงานมาก (อาจจะ 5 – 10 นาที ต่อภาพ ที่ระดับพิกเซล เช่นที่รายงานใน [50], [51] และ ขึ้นไปถึง 5 นาทีต่อหน้าในกรณีที่ทำการระดับขอบเขตบล็อกดังรายงานใน [49] ดังนั้นถ้ามีจำเป็นต้องให้คุณค่าประสิทธิภาพของวิธีการต้องใช้หนึ่งภาพที่สแกนเข้าด้วยความเอียงมุมต่างๆเริ่มตั้งแต่ 1 องศาถึง 100 องศา โดยมีการเพิ่มขึ้นเอกสารแบบเอกสารทรวงานวิสาทรภกรเชิงานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตเินาไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทีละ 1 องศา ซึ่งจำเป็นต้องสแกนถึง 100 ครั้งในมุมที่แตกต่างกัน ในกรณีนี้การตรวจสอบรูปแบบที่ถูกต้องจะใช้เวลาหลายชั่วโมง นั่นเป็นคำตอบว่าทำไมผู้แต่งหลายคนถึงใช้วิธีเก่าในการประเมินคุณค่า โดยใช้การมองเห็นด้วยสายตามาประเมินคุณค่า อย่างไรก็ตาม ยังมีหวังกับผลลัพธ์ที่ได้จากการให้คุณค่าของอัลกอริทึมวิเคราะห์แผนผังและการสร้างรูปแบบที่ถูกต้องเริ่มปรากฏใน [48], [52] แนวคิดที่ [49] เป็นการใช้จ่ายเงินสำหรับการสร้างรูปแบบที่ถูกต้องเพราะยินยอมให้เป็นการทำงานแบบอัตโนมัติ แต่วิธีการนี้ยังทำได้แค่เอกสารข้อความที่ไม่มีรูปภาพหรือกราฟิก

ในการปฏิบัติ เป็นการดีหากมีการใช้อัลกอริทึมเฉพาะเจาะจงสำหรับชนิดเอกสารที่แตกต่างกัน และงานที่ทำในลำดับในประสิทธิภาพการทำงานอย่างเหมาะสมที่สุด ลักษณะเด่นต่อไปนี้ควรจะถูกใช้ในกรณีต่างๆ

- ความคงทนต่อความเอียง (เอียงแบบเดียว หรือ เอียงหลายแบบ)
- ไม่ขึ้นกับรูปแบบแผนผังของเอกสาร
- สามารถสกัดข้อความได้ทั้งฉากพื้นสีขาวและฉากพื้นสีตรงข้าม
- ง่ายต่อการเข้าถึงข้อมูล
- มีความเร็วและความถูกต้องสูง
- ไม่ขึ้นกับ ชนิด ขนาด และ ลักษณะของรูปแบบตัวอักษร

## 2.11 สรุป

การสำรวจนี้เป็นกรกล่าวขั้นตอนต่างๆไปของการวิเคราะห์แผนผังเอกสาร และอธิบายขั้นตอนการทำงานเริ่มแรกตั้งแต่ปี 1990 ในตอนแรกสรุปการวิเคราะห์งานที่จำเป็นต่อประเภทเอกสารที่หลากหลาย สามารถจัดได้เป็นสามประเภท

1. กลุ่มที่มีโครงสร้าง
2. เอกสารที่มีแผนผังโครงสร้างไม่แน่นอน
3. เอกสารที่ไม่ใช่แบบดั้งเดิม

ในกรณีเพิ่มเติมอธิบายวิธีการที่ใช้สำหรับกลุ่มนี้ อธิบายถึงจุดเด่นเช่นภาพ และ ชนิดของฉากพื้น ความละเอียดของภาพ เวลาที่ใช้ในการประมวลผล ความทนทานต่อความเอียง และ แผนผังที่แตกต่าง

ถึงอย่างไรก็ตามงานวิจัยที่จำนวนมากมายในด้านนี้ ยังไม่มีวิธีที่เป็นพื้นฐานที่จะใช้ประมวลผลภาพในประเภทเอกสารที่ต่างกัน ทั้งความถูกต้องและอัตโนมัติ จุดเด่นที่สำคัญในวิธีการที่มีคือ ความเอียง แผนผัง และ ความอิสระต่อแบบเขียน ความเร็วในการทำงาน ความถูกต้องสูง และยืดหยุ่น การแสดงตัวแทนภาพสามารถเข้าถึงข้อมูลได้ง่าย รูปแบบที่เหมาะสมของความเชื่อของกราฟิก และรูปภาพ การสร้างรูปแบบที่ถูกต้องแบบอัตโนมัติ การวัดประสิทธิภาพของวิธีการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิเคราะห์แผนผังสำหรับภาพไบนารีและเกรย์สเกล และระบบฐานความรู้อัตโนมัติหรือไม่อัตโนมัติ การประมวลผลภาพสีจะกลายเป็นเป้าหมายสำหรับการวิจัยในอนาคต สำหรับงานวิจัยที่ผ่านมาจะแสดงในตารางที่ 2.3 ซึ่งจะแสดงถึง ชื่อผู้วิจัย วิธีการที่ใช้ และคุณลักษณะเด่นของแต่ละวิธีการ

ตารางที่ 2.3 แสดงผลงานวิจัยที่ผ่านมา\*

ลำดับ	ผู้แต่ง	ปี	วิธีการ	จุดเด่น
1	Wahl et al.[54]	1982	Run length smoothing	ใช้เวลามากและไวต่อความเอียง
2	Nagy et al.[55]	1984	X-Y tree cut	ไวต่อความเอียง และเป็นบล็อกสี่เหลี่ยม
3	Wang et al.[28]	1989	Run Length smoothing and recursive X-Y cut	วิเคราะห์หนังสือพิมพ์ : ไวต่อความเอียง
4.	Fujisawa et al. [56]	1990	Top-down	เอกสารสิทธิบัตรประเทศญี่ปุ่น
5	Fisher et al.[57]	1990	Run-length smoothing and connected component extraction	กำหนดขอบเขตข้อความและรูปภาพ ไวต่อความเอียง
6	Pavlidis et al.[58]	1991	Column oriented projection	กำหนดขอบเขตข้อความและรูปภาพ เหมาะกับความเอียงปานกลาง
7	Baird[59]	1992	Global-to-local strategy	เหมาะกับภาษาต่างๆ และแก้ไขความเอียง
8	Jain et al.[26]	1992	Gabor filtering	ใช้จุดเด่นพื้นผิวของภาพเฉดเทา ใช้เวลาในการทำงานมาก
9	Labourgeois et al. [60]	1992	8*3 windows Filtering	เอกสารไม่เป็นรูปแบบ ความเอียง ไม่มีผล
10	Pavlidis et al.[38]	1992	Horizontal smearing and bottom-up	กำหนดค่าคงที่ เหมาะกับความเอียง
11	Akindele et al. [61]	1993	White space tracing	บล็อกหลายเหลี่ยม สนใจเฉพาะขอบเขตข้อความเท่านั้น
12	Amamoto et al. [62]	1993	Morphological operation on white space	กำหนดตัวเขียนทั้งแนวตั้งและแนวนอน ความเอียงไม่มีผล
13	Ittner et al.[63]	1993	White space and minimum spanning tree	ภาษาและตำแหน่งเป็นอิสระ ใช้เวลาในการคำนวณมาก

\* ที่มา A. Jain, B. Yu, "Document Representation and Its Application to Page Decomposition", IEEE

ตารางที่ 2.3 (ต่อ)

ลำดับ	ผู้แต่ง	ปี	วิธีการ	จุดเด่น
14	O’Gorman[64]	1993	k-nearest neighbor clustering	ใช้กับเอกสารที่มีโครงร่างไม่แน่นอน ให้ความถูกต้องสูง และใช้การคำนวณมาก
15	Antonacopoulos et al.[5], [65]	1994	Contours from white tiles	หาขอบเขตที่ไม่เป็นสี่เหลี่ยมและเอียง มีความผิดพลาดในการแบ่งตัวอักษรที่มีขนาดใหญ่
16	Zlatopolsky[66]	1994	Connected component extraction	ไวต่อค่าต่างๆ เอกสารมีหลายความเอียง
17	Doermann[67]	1995	Wavelet multiscale analysis	การคำนวณซับซ้อนมาก แบ่งเอกสารที่มีความซับซ้อน
18	Drivas et al.[32]	1995	Connected component grouping	แก้ไขความเอียง ใช้เวลาในการทำงานมาก
19	Ha et al.[68]	1995	Connected component-based projection profile	เร็วกว่าการฉายภาพของจุด และไวต่อความเอียง
20	Sylwester et al. [69]	1995	Trainable X-Y cut	มันคงมาก ความเอียงและสัญญาณรบกวนไม่มีผล
21	Tang et al.[42]	1995	Modified fractal signature	ใช้กับเอกสารที่มีหลายภาษา ใช้การคำนวณมาก
22	Jain et al.[11], [70]	1996	Masks and neural network	ใช้จัดการกับเอกสารที่มีความซับซ้อน ใช้เวลามาก
23	Kise et al.[71]	1996	Background Thinning	ขอบเขตไม่เป็นสี่เหลี่ยม ความเอียงไม่มีผล
24	Liu et al.[72]	1996	Adaptive top-down and bottom-up	ขอบเขตไม่เป็นสี่เหลี่ยม ความเอียงไม่มีผล
25	Yamaahita et al. [73]	1996	Run length smearing and adaptive thresholding	ขนาดของอักษรและช่องว่างมีผล กระทบน้อย ความเอียงไม่มีผล

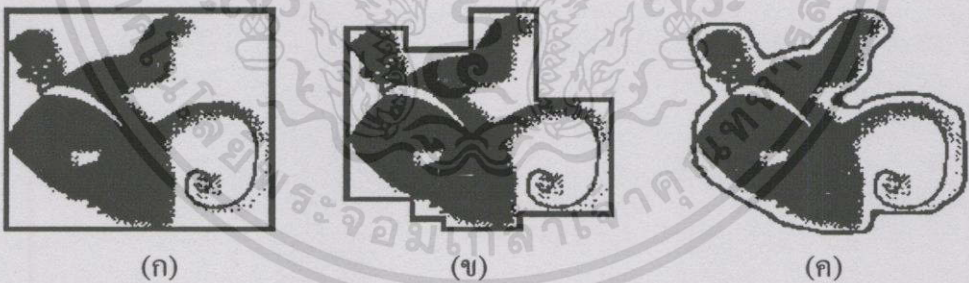
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### บทที่ 3

## แนวคิดและการนำมาประยุกต์ใช้ในการแบ่งส่วนประกอบภาพ เอกสารใบนารี

### 3.1 แนวคิดในการแบ่งส่วนประกอบภาพเอกสารใบนารี

จากการแบ่งรูปแบบแผนผังโครงสร้างเอกสารในบทที่ 2 ซึ่งแบ่งแผนผังของส่วนประกอบเอกสารออกเป็นรูปแบบที่เป็นสี่เหลี่ยม รูปแบบแมนฮัตตัน และรูปแบบที่ไม่แน่นอน ดังแสดงในรูปที่ 3.1 โดยมีรูปแบบที่ยืดหยุ่นที่สุดคือ รูปร่างที่ไม่แน่นอน ซึ่งเหมาะสมในการแสดงการแยกส่วนประกอบของเอกสารต่างๆ ได้กว้างขวางและเหมาะสมกว่า นอกจากนี้ยังมีลักษณะที่พิเศษคือไม่จำกัดว่าขอบเขตของส่วนประกอบต้องเป็นสี่เหลี่ยมเท่านั้น จึงเป็นทางเลือกที่ต้องการในการศึกษาหาแนวทางที่จะแยกขอบเขตของส่วนประกอบที่มีรูปแบบไม่แน่นอน เพราะเอกสารไม่ได้มีรูปร่างส่วนประกอบที่เป็นชนิดสี่เหลี่ยมเท่านั้น แต่ยังมีรูปทรงเป็นวงกลม หรือรูปหลายเหลี่ยม จากรูปจะเห็นได้ว่าแผนผังรูปร่างไม่แน่นอน สามารถบอกขอบเขตของส่วนประกอบได้ดีที่สุด โดยใช้ข้อมูลในการอ้างอิงขอบเขตมากที่สุด รองลงมาเป็นแผนผังแมนฮัตตัน ส่วนสี่เหลี่ยมจะบอกขอบเขตได้หยาบที่สุด แต่ใช้ข้อมูลในการอ้างอิงขอบเขตน้อยที่สุด



รูปที่ 3.1 ตัวอย่างแผนผังแบบต่างๆ

- ก. แผนผังสี่เหลี่ยม
- ข. แผนผังแมนฮัตตัน
- ค. แผนผังแบบรูปร่างไม่แน่นอน

แนวคิดการทำเบลอเป็นการทำให้รายละเอียดที่อยู่ใกล้กันในระยะทางที่กำหนดไว้มาเชื่อมรวมเข้าด้วยกัน ทำให้รายละเอียดนั้นมีขนาดใหญ่ขึ้น วิธีการนี้เป็นการทำงานที่คล้ายหนึ่งเพื่อที่จะรวมส่วนประกอบย่อยที่อยู่ในระยะทางที่กำหนดเข้าด้วยกันเป็นส่วนประกอบที่ใหญ่ขึ้น เป็นแนวทางหนึ่งที่จะรวบรวมรายละเอียดของส่วนประกอบที่อยู่ใกล้กัน เช่น ตัวอักษร ดังนั้นจึงมีการสร้างเอกสารแบบเอกสารที่แสดงรายละเอียดการเขียนเพื่อที่จะทำเช่นนี้ เมื่อผู้ให้เนื้อหาเขียนเสร็จสิ้นการดำเนินการก็ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

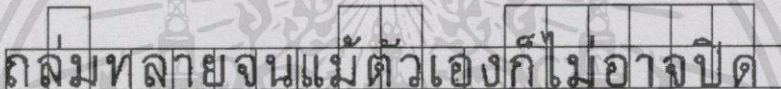
หน้าต่างขึ้นมาเพื่อที่จะเชื่อมตัวอักษรที่อยู่ใกล้กันในระยะช่องหน้าต่างเชื่อมเข้าด้วยกันเป็นส่วนประกอบเดียวกัน ซึ่งวิธีการนี้จะคล้ายกับวิธีการทำเบลอ เปลี่ยนการทำงานด้วยจุดไปเป็นหน้าต่างแทนซึ่งทำให้ลดเวลาในการเชื่อมส่วนประกอบเข้าด้วยกันลง ดังแสดงในรูปที่ 3.2

## กลมทลายจนแม้ตัวเองก็ไม่อาจปิด

(ก)



(ข)



(ค)

รูปที่ 3.2 ลักษณะการเชื่อมต่อตัวอักษร

- ก. ข้อความต้นฉบับ
- ข. การเชื่อมต่อตัวอักษรด้วยการทำเบลอ
- ค. การเชื่อมต่อตัวอักษรด้วยหน้าต่าง

จากรูปแสดงให้เห็นการเชื่อมเพื่อรวมกลุ่มของตัวอักษรจากข้อความต้นฉบับ(ก) ด้วยวิธีการทำเบลอ(ข) ซึ่งจะใช้การเบลอระดับจุดภาพในการขยายรายละเอียดออกไปเท่าขนาดที่กำหนดทำให้รวมกันเป็นขอบเขต ในขณะที่หน้าต่าง(ค)จะใช้วิธีการนับจำนวนจุดภาพในหน้าต่างเพื่อเชื่อมต่อหน้าต่างที่มีจุดสีมากกว่าค่าที่กำหนดไว้เข้าด้วยกัน ซึ่งใช้เวลาในการทำงานน้อยกว่าการเบลอแบบจุด

แนวคิดวิธีการเดินรอบขอบ เป็นวิธีการหาขอบเขตของรายละเอียดได้โดยไม่สนใจจุดที่อยู่ในขอบเขตแต่จะสนใจเฉพาะส่วนที่เป็นขอบของรายละเอียดเท่านั้น ซึ่งสามารถลดเวลาในการทำงานลงได้

เมื่อนำจุดเด่นของทั้งสามแนวทางที่พิจารณามารวมกันเพื่อหาแนวทางในการทำงานที่รวดเร็วและดียิ่งขึ้นซึ่งเป็นวิธีการแบ่งประเภทที่ใช้ได้กับเอกสารที่มีรูปแบบที่หลากหลาย เช่น คอลัมน์ที่ไม่อยู่ในรูปแบบสี่เหลี่ยม ซึ่งอาจเป็นรูปหลายเหลี่ยม หรือวงกลม แต่ยังมีข้อสังเกตอีกอย่างที่สามารถจะแบ่งส่วนประกอบได้ คือ ช่องว่าง และ เส้นกัน โดยส่วนมากส่วนประกอบของเอกสาร

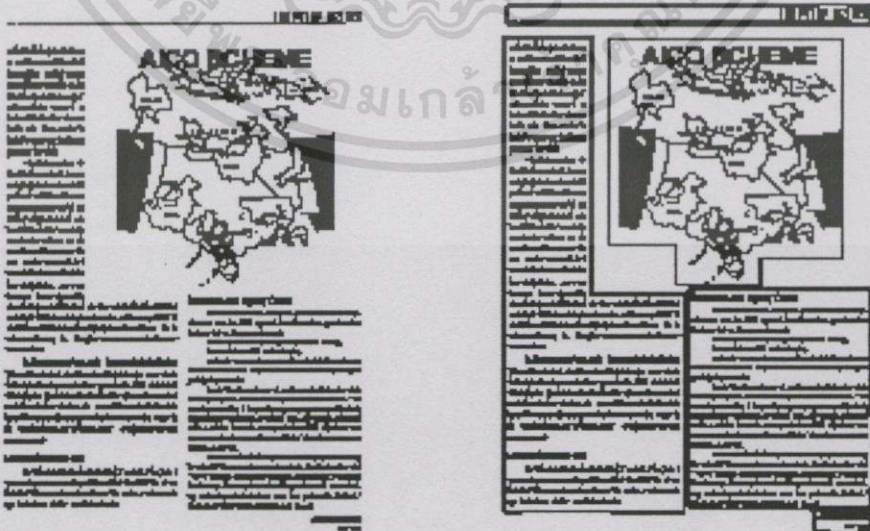
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่มีรายละเอียดแตกต่างกันจะมีระยะห่างจากกันพอสมควร เพื่อความสวยงาม หรือ มีเส้นกันไว้เพื่อบอกขอบเขตที่สิ้นสุดของรายละเอียด ในการที่จะแบ่งขอบเขตส่วนประกอบของแต่ละส่วน หลังจากที่ได้พิจารณาแล้ว ได้แนวคิดว่าการทำหน้าต่างเดิวนรอบขอบส่วนประกอบและเก็บข้อมูลในลักษณะแผนผังแบบแมนฮัตตัน

ในการเดินรอบขอบเขตโดยใช้หน้าต่างนี้จะได้ขอบเขตของส่วนประกอบใหญ่ซึ่งรู้ตำแหน่งและขนาด แต่ยังไม่สามารถบอกได้ว่าส่วนประกอบที่แบ่งขอบเขตมานั้นเป็นส่วนข้อความหรือรูปภาพ ดังนั้นจึงต้องมี การเดินรอบส่วนรายละเอียดในส่วนประกอบใหญ่ที่ได้จากการแบ่งขอบเขตจากหน้าต่าง ซึ่งในขั้นตอนนี้สามารถบอกได้อย่างคร่าวๆว่าส่วนประกอบใดน่าจะเป็นข้อความ ส่วนประกอบใดน่าจะเป็นรูปภาพ แต่เนื่องจากการแบ่งขอบเขตในตอนแรก อาจจะได้รายละเอียดที่ไม่ใช่ชนิดเดียวกันมาอยู่รวมกัน เพราะฉะนั้นต้องมีขั้นตอนสุดท้ายคือการนำส่วนประกอบย่อยหรือขอบเขตรายละเอียดมาทำการแบ่งส่วนประกอบใหญ่ที่มีรายละเอียดหลายชนิดรวมกันให้แบ่งแยกรายละเอียดที่ไม่ใช่ประเภทเดียวกันออกจากกัน ซึ่งจะอธิบายทั้ง 3 ขั้นตอนในหัวข้อ 3.2 ถึง 3.4

### 3.2 การแบ่งส่วนประกอบภาพเอกสารเป็นส่วนประกอบใหญ่

ในขั้นตอนนี้เปรียบได้กับการมองเอกสารที่อยู่ไกลๆ แสดงในรูปที่ 3.3(ก) โดยไม่สามารถมองเห็นรายละเอียดที่ชัดเจน แต่จะมองเห็นขอบเขตของส่วนประกอบเท่านั้น ว่าอยู่ตรงตำแหน่งไหน และมีขอบเขตของส่วนประกอบขนาดเท่าใด โดยรับรู้จากการมองเห็นผ่านทางสายตา และนำมาวิเคราะห์โดยผ่านประสบการณ์ จะได้ขอบเขตดังแสดงในรูปที่ 3.3(ข)

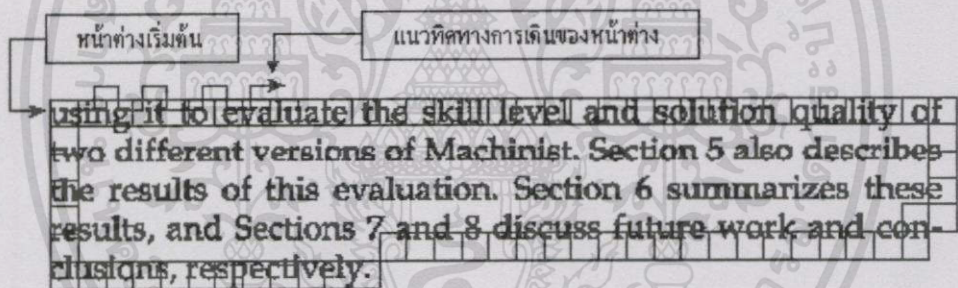


(ก)

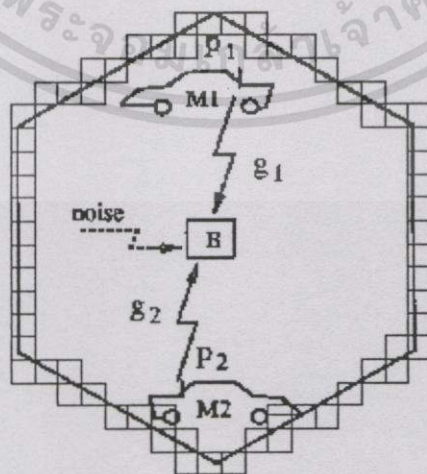
(ข)

รูปที่ 3.3 ขอบเขตภาพ (ก) ภาพแสดงเมื่อมองจากระยะไกล (ข) ภาพที่แบ่งขอบเขตด้วยสายตา เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในหลักการเดียวกัน แนวคิดการวางหน้าต่างขนาด  $32 \times 32$  พิกเซลจะวางที่ตำแหน่งซ้ายบนสุดของรายละเอียดของส่วนประกอบ แล้ววางหน้าต่างถัดไปเรียงตามขอบของส่วนประกอบ ขอบเขตที่ได้จะเป็นขอบเขตส่วนประกอบใหญ่ โดยจะนำขอบเขตดังกล่าวจากขั้นตอนนี้ไปใช้ในการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ต่อไป เหตุผลที่ต้องกำหนดหน้าต่างขนาดเท่ากับ  $32 \times 32$  พิกเซล ซึ่งขนาดของหน้าต่างๆจะใกล้เคียงกับขนาดตัวอักษรปกติที่ใช้ในเอกสารทั่วไป ซึ่งขนาดของส่วนประกอบนี้สามารถที่จะเชื่อมตัวอักษรที่อยู่ใกล้ชิดกันเข้าด้วยกันรวมเป็นกลุ่ม และมีขนาดเล็กพอที่จะแบ่งส่วนประกอบที่มีระยะห่างกันออกจากกันได้ โดยการวางหน้าต่างนี้จะวางเรียงต่อกันไปตามรอบขอบส่วนประกอบ ดังนั้นขอบเขตส่วนประกอบที่ได้คือขอบเขตของหน้าต่างที่เรียงต่อกันจากจุดเริ่มต้นของหน้าต่างแรกจนวนกลับมาถึงหน้าต่างเริ่มต้น ดังนั้นขอบเขตที่ได้จึงเป็นขอบเขตส่วนประกอบขนาดใหญ่ที่มีรายละเอียดต่างๆ มาอยู่รวมกัน ดังรูปที่ 3.4 แสดงรูปแบบการเดินรอบส่วนประกอบของหน้าต่างและทิศทางที่ใช้ในการเดินรอบส่วนประกอบใหญ่ และรูปที่ 3.5 แสดงลักษณะส่วนประกอบที่ได้จากหน้าต่างที่เดินรอบขอบที่มีลักษณะแผนผังไม่อยู่ในรูปแบบสี่เหลี่ยม



รูปที่ 3.4 การวางตำแหน่งหน้าต่างเริ่มต้นและลักษณะการวนรอบส่วนประกอบใหญ่



รูปที่ 3.5 ลักษณะภาพส่วนประกอบที่ไม่อยู่ในลักษณะสี่เหลี่ยม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

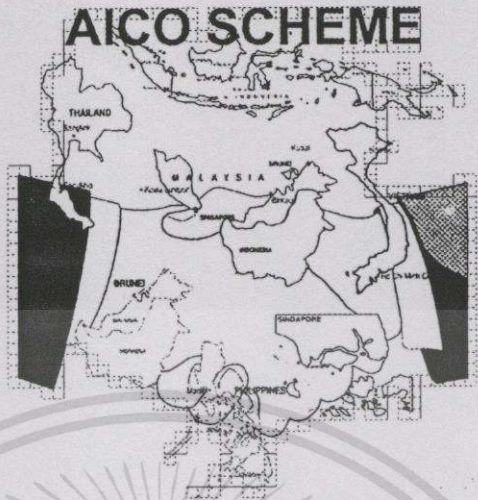
ภาคตะวันออกเฉียงเหนือของอาเซียน  
 3) หากมีระยะเวลาให้ใช้สิทธิบัตร  
 สิทธิบัตรสั้นๆ หากพิจารณาจาก  
 ปัญหาที่ยังมีผลกระทบ และต้อง  
 มีผู้สามัคคีหรือเจรจาเพื่อแก้ไข  
 จึงสมควรจะสั้นๆ มีอยู่ 1 ปี 1.5  
 ปี 2 ปี หรือ 3 ปี หรือ 4 ปี หรือ 5 ปี  
 ไม่เกิน เช่น มีเวลาให้สิทธิบัตร  
 สิทธิบัตรความร่วมมือร่วมสมัย  
 สิทธิบัตรร่วมสมัย 1 ปี 1.5 ปี 2 ปี

ข้อที่ 10 บัญชีทรัพย์สิน  
 สิทธิบัตรสิทธิบัตรจะเขียนในภาษา  
 หนึ่งหรือสองภาษาเป็นปกติหรือได้  
 ภายใต้อาณัติในเบื้องต้น 10  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรเขียนในสองภาษาได้  
 สิทธิบัตรเขียนในสองภาษาได้

ข้อที่ 11 สิทธิบัตร  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้

ข้อที่ 12 สิทธิบัตร  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้

ข้อที่ 13 สิทธิบัตร  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้  
 สิทธิบัตรจะเขียนในสองภาษาได้



Each ASEAN member country has specific criteria that companies must meet for AICO approval, but all members agreed to three basic provisions. The companies must:  
 - be incorporated and operating in an ASEAN country;  
 - have a minimum 30% national equity;  
 - undertake resource sharing or pooling of industrial capabilities with other industrial cooperation activities that are accepted by the participating countries.  
 For justification and approval, applicants should clearly explain how they cooperate, and how such cooperation will be beneficial. The most important thing is the participating company must specify clearly to the analysts what kind of cooperation it has, and what the results will achieve and that the results correspond with AICO's objectives. This means strengthening competitive capabilities," the assistant permanent secretary suggested.  
 Mr. Dhawatchai further explained that the equity condition may be waived if the companies meet other criteria. For Thailand, he said "we will take into consideration whether its proposed plan is beneficial in other ways, for example, merchandising products with new technology and such products that have never been produced in Thailand. Such situations are looked at case by case," he said.

AS INDUSTRY MAG  
 March 2004

รูปที่ 3.6 ขอบเขตของส่วนประกอบในภาพเอกสารที่แบ่งโดยหน้าต่างขนาด 32\*32 พิกเซล

จากรูปที่ 3.6 แสดงขอบเขตของส่วนประกอบหลังจากทำการวางหน้าต่างขนาด 32\*32 พิกเซลรอบขอบเขต ซึ่งการวางหน้าต่างนี้สามารถแบ่งขอบเขตแต่ละของส่วนประกอบได้ ถ้าส่วนประกอบเหล่านั้นมีระยะที่ห่างกันเกินกว่าหน้าต่างจะสามารถเชื่อมกันก็ได้ และหากส่วนประกอบอยู่ใกล้กันจนทำให้หน้าต่างมาเชื่อมต่อกัน ซึ่งบางที่ขอบเขตส่วนประกอบใหญ่ที่ได้จะมีทั้งข้อความและภาพรวมกัน หรือมีคอลัมน์ข้อความเกินกว่าหนึ่งคอลัมน์ก็ได้

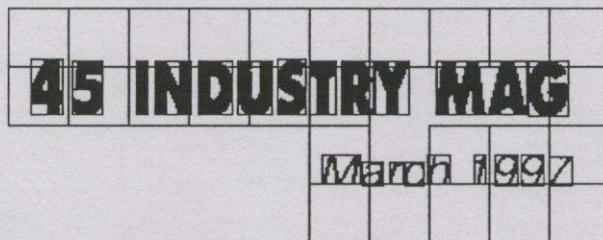
3.3 การแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่

หลังจากที่ได้ขอบเขตรูปร่างของส่วนประกอบใหญ่แล้ว ยังไม่สามารถที่จะนำมาจัดแบ่งประเภทส่วนประกอบว่าเป็นชนิดข้อความหรือรูปภาพ ดังนั้นต้องมีการแบ่งส่วนประกอบย่อยในระดับรายละเอียด เพื่อแบ่งแยกส่วนต่างๆที่อยู่ในส่วนประกอบใหญ่ โดยจะมีการเดินขอบด้วยเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้าต่างขนาด 1 พิกเซลแทนหน้าต่าง 32\*32 พิกเซล เพื่อแยกส่วนรายละเอียดออกมา โดยขอบเขตของรายละเอียดจะถูกเก็บเป็นบล็อกสี่เหลี่ยม(มุมบนซ้าย, มุมล่างขวา) เพื่อประหยัดเนื้อที่ในการจัดเก็บข้อมูล และมีข้อกำหนดว่า ถ้าความกว้างและความสูงของขอบเขตส่วนประกอบมีขนาดน้อยกว่าหรือเท่ากับ 5 พิกเซล จะถูกกำหนดให้เป็นสัญลักษณ์รบกวน ถ้าความสูงหรือความกว้างของขอบเขตส่วนประกอบมีขนาดมากกว่าหรือเท่ากับ 100 พิกเซล จะถูกกำหนดให้เป็นภาพ ส่วนขอบเขตที่ไม่อยู่ในช่วงที่กำหนดข้างต้น จะถูกกำหนดให้เป็นตัวอักษร ซึ่งหลังจากที่ได้ขอบเขตที่เป็นตัวอักษรตัวแรกแล้ว นำมาตรวจสอบหาตัวอักษรระดับบรรทัด(ตัวอักษรที่คาดว่าจะอยู่ในระดับบรรทัดเช่น ก หรือ ข เป็นต้น) ซึ่งเป็นพยัญชนะในภาษาไทย ถ้าเป็นตัวอักษรภาษาอังกฤษได้แก่ a หรือ b เป็นต้น) โดยทำการสแกนหาจุดค่าที่อยู่ใต้ขอบเขตตัวอักษรแรกนั้น แล้วทำการเดินวนรอบเพื่อหาขอบเขตว่าเป็นตัวอักษรหรือไม่ จากนั้นทำการตรวจสอบความสูงของขอบเขตทั้งสองว่าขอบเขตใดที่สูงที่สุด จากนั้นทำซ้ำไปจนกระทั่งไม่เจอขอบเขตที่เป็นตัวอักษรที่อยู่ข้างล่างขอบเขตนั้นๆ จึงนำขอบเขตที่มีความสูงที่สุดมากำหนดเป็นตัวอักษรระดับบรรทัด ต่อไปทำการตรวจสอบหาขอบเขตตัวอักษรที่อยู่ทางด้านซ้าย ด้านบน ด้านล่าง และด้านขวา ถ้าเจอขอบเขตที่เป็นตัวอักษรให้นำมาจัดเรียงกันเป็นกลุ่มตามลำดับขอบซ้าย และทำงานกระทั่งไม่เจอขอบเขตตัวอักษร สำหรับการเก็บขอบเขตจะแยกเก็บ โดยตัวอักษรจะถูกเก็บเป็นกลุ่มตัวอักษร รูปภาพจะถูกเก็บทีละภาพเรียงต่อกัน และสัญลักษณ์รบกวนจะถูกจัดเก็บเรียงต่อกัน โดยการดำเนินงานทั้งหมดจะทำภายในขอบเขตส่วนประกอบใหญ่เท่านั้น และแยกเก็บขอบเขตของรายละเอียดภายในของแต่ละกลุ่มแยกกัน พิกัดของขอบเขตรายละเอียดจะนำไปใช้ในการวิเคราะห์ เพื่อหารูปแบบหรือประเภทของส่วนประกอบและแบ่งส่วนประกอบที่ไม่สัมพันธ์กันออกจากกัน ส่วนรายละเอียดที่อยู่ภายในส่วนประกอบใหญ่แสดงดังรูปที่ 3.7 เป็นส่วนประกอบที่มีทั้งข้อความและรูปภาพที่เป็นเส้นอยู่ด้วยกัน และรูปที่ 3.8 เป็นส่วนประกอบที่มีเฉพาะข้อความ

NEWSTECH ISSUES

รูปที่ 3.7 ส่วนประกอบใหญ่ที่มีรายละเอียดที่เป็นทั้งข้อความและภาพอยู่ด้วยกัน



รูปที่ 3.8 ส่วนประกอบใหญ่ที่มีรายละเอียดเป็นข้อความเพียงอย่างเดียว

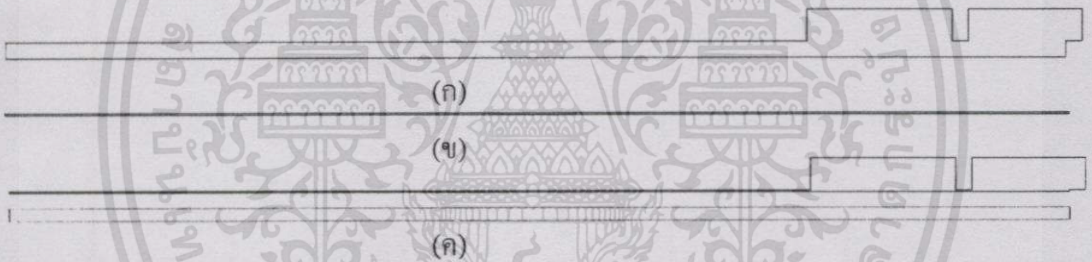
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.4 การแบ่งคอลัมน์และปรับปรุงขอบเขตส่วนประกอบใหญ่

ข้อมูลต่างๆที่ได้มาจากขั้นตอนที่ 3.3 จะนำมาวิเคราะห์เพื่อตรวจสอบว่ารายละเอียดข้อมูลที่อยู่ในส่วนประกอบใหญ่เป็นประเภทเดียวกันหรือไม่ ถ้าส่วนประกอบใหญ่ไม่ได้เป็นประเภทเดียวกันจะทำการแบ่งแยกส่วนประกอบที่ไม่สัมพันธ์กันออกไป โดยจะมีขั้นตอนดังต่อไปนี้

#### 3.4.1 การแบ่งโดยนำรายละเอียดที่เป็นภาพมาตัดแบ่งส่วนประกอบใหญ่

นำขอบเขตส่วนรายละเอียดที่เป็นภาพ มาหักออกจากขอบเขตใหญ่ ซึ่งใช้ขอบเขตส่วนประกอบใหญ่ในแนวบรรทัดหน้าค่างที่ขอบซ้ายขวาตั้งเป็นตัวหลักเพื่อใช้ ขอบเขตภาพรายละเอียด ข้ายขวามาเป็นตัวลบนในแนวบรรทัดจุดภาพ ดังแสดงในรูปที่ 3.9 โดยนำขอบเขตจากรูปที่ 3.7 มาเป็นตัวอย่าง ซึ่งการนำภาพมาหักจากส่วนประกอบใหญ่จะทำทุกภาพในส่วนประกอบใหญ่นั้น และทำทุกส่วนประกอบใหญ่ที่มีภาพ โดยการหักจะมีเงื่อนไข ถ้าขอบเขตบรรทัดหน้าค่างข้ายขวา ถูกหักโดยขอบเขตของภาพแล้วจำนวนจุดของแต่ละด้านเหลือไม่เกิน 100 พิกเซลให้ถือว่าส่วนบน กับส่วนล่างของภาพถูกแบ่งออกจากกัน



รูปที่ 3.9 แสดงการนำขอบเขตภาพมาหักออกจากขอบเขตส่วนประกอบใหญ่

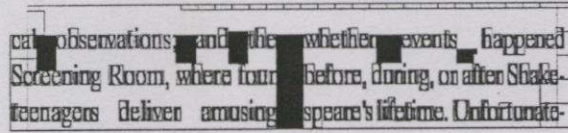
- ก. ขอบเขตของส่วนประกอบใหญ่
- ข. ขอบเขตของภาพ
- ค. ขอบเขตส่วนประกอบหลังจากหักขอบเขตส่วนประกอบใหญ่

#### 3.4.2 การนำขอบเขตกลุ่มตัวอักษรมาหาคอลัมน์ในส่วนประกอบใหญ่

กลุ่มตัวอักษรที่ได้จากขั้นตอนที่ 3.3 มีลักษณะที่อยู่รวมกันเป็นกลุ่ม โดยแต่ละกลุ่มจะมีระยะทางระหว่างกลุ่มแตกต่างกันไป โดยกลุ่มตัวอักษรต้องมีการจัดเรียงขอบเขตค่าพิกัดในแนวแกนนอนและแกนตั้งจากน้อยไปมาก แล้วจัดเรียงข้อมูลทั้งหมดให้อยู่ในรูปบรรทัด ในการทำงานถัดไป จะนำระยะทางระหว่างที่กลุ่มตัวอักษร จากนั้นนำกลุ่มบรรทัดนี้มาทำการตัดแนวแกน Y (ตามแนวตั้ง) ของส่วนประกอบใหญ่ ซึ่งจะทำให้เราทราบแนวคอลัมน์ ดังแสดงในรูปที่ 3.10 โดยจะแสดงกลุ่มขอบเขตตัวอักษรซึ่งจะมีช่องว่างระหว่างกลุ่ม และการทำงานจะสนใจเฉพาะกลุ่มบรรทัดที่อยู่บนส่วนประกอบที่ถือว่าเป็นบรรทัดแรก มาทำการดำเนินการทำการฉายเงาตัดจุดขาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

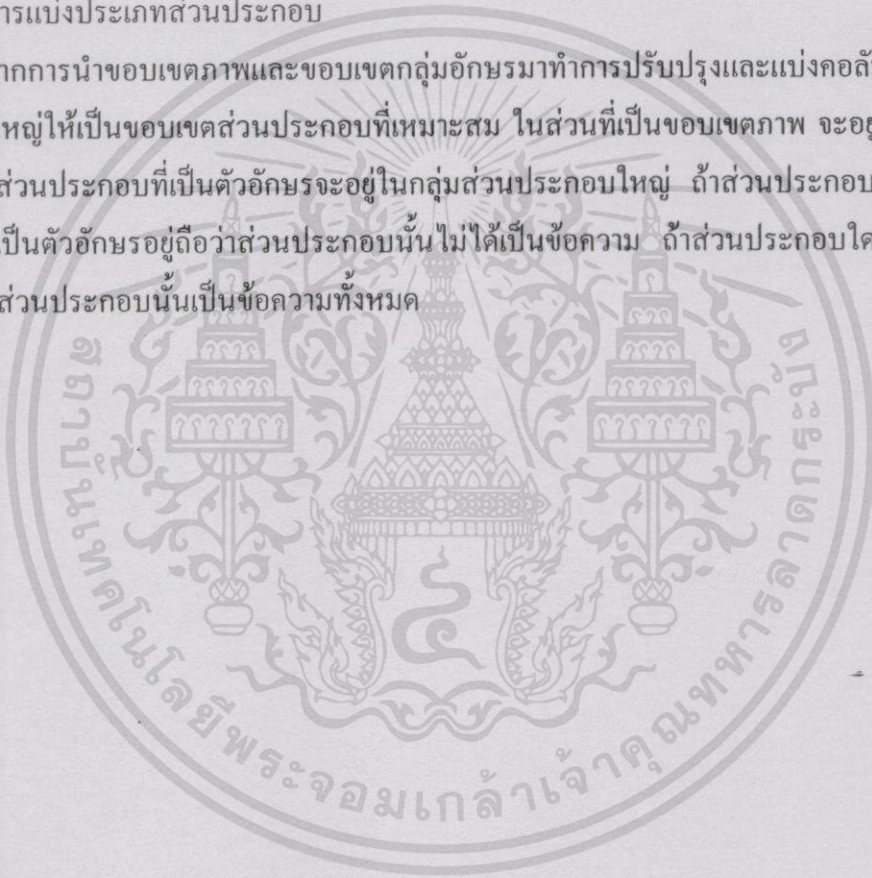
แกน Y เพื่อหาขอบเขตคอลัมน์ โดยมีการกำหนดไว้ว่า ถ้ามีการฉายเงาจุดขาวได้ค่าเกินกว่า 3 บรรทัด ให้ถือว่าสามารถแบ่งคอลัมน์ของกลุ่มตัวอักษรออกเป็นสองคอลัมน์ได้



รูปที่ 3.10 ภาพจำลองในการนำกลุ่มบรรทัดมาตัดแนวแกน Y เพื่อแบ่งส่วนคอลัมน์

### 3.4.3 การแบ่งประเภทส่วนประกอบ

จากการนำขอบเขตภาพและขอบเขตกลุ่มอักษรมาทำการปรับปรุงและแบ่งคอลัมน์ในส่วนประกอบใหญ่ให้เป็นขอบเขตส่วนประกอบที่เหมาะสม ในส่วนที่เป็นขอบเขตภาพ จะอยู่ในชุดของภาพ และส่วนประกอบที่เป็นตัวอักษรจะอยู่ในกลุ่มส่วนประกอบใหญ่ ถ้าส่วนประกอบใหญ่ใดไม่มีข้อมูลที่เป็นตัวอักษรอยู่ถือว่าส่วนประกอบนั้นไม่ได้เป็นข้อความ ถ้าส่วนประกอบใดมีตัวอักษรอยู่ก็ถือว่าส่วนประกอบนั้นเป็นข้อความทั้งหมด



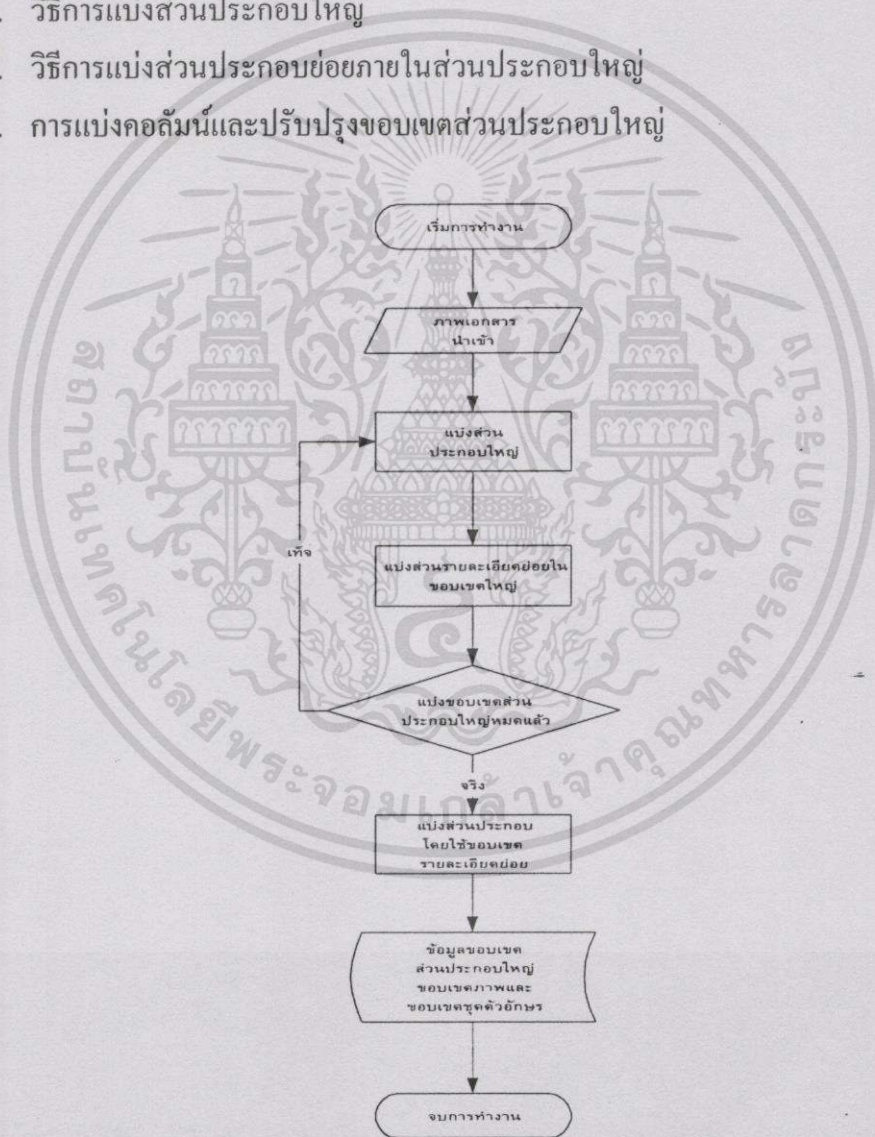
## บทที่ 4

# อัลกอริทึมและขั้นตอนการทำงาน

### 4.1 หลักการที่ใช้ในการทดลอง

จากแนวคิดที่ผ่านมาในบทที่ 3 เรานำมาทำการประยุกต์เพื่อทดสอบตามสมมุติฐานที่ตั้งไว้ โดยจะแบ่งวิธีการทำงานออกเป็น 3 ขั้นตอนดังนี้ และแสดงในแผนผังการดำเนินงานรูปที่ 4.1

1. วิธีการแบ่งส่วนประกอบใหญ่
2. วิธีการแบ่งส่วนประกอบย่อยภายในส่วนประกอบใหญ่
3. การแบ่งคอลัมน์และปรับปรุงขอบเขตส่วนประกอบใหญ่



รูปที่ 4.1 แผนผังขั้นตอนการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.1.1 วิธีการแบ่งส่วนประกอบใหญ่ มีขั้นตอนการทำงานดังนี้

4.1.1.1 หาดำแหน่งเริ่มต้นของหน้าต่างขนาด  $32 \times 32$  พิกเซล การหาดำแหน่งเริ่มแรกของหน้าต่างขนาด  $32 \times 32$  พิกเซลนี้ จะสแกนหาจุดดำไปที่ละแถว (X) โดยไปที่ละจุดในแนวคอลัมน์ (Y) เมื่อเจอจุดดำจุดแรก ให้กำหนด  $x$  เป็นตำแหน่งของขอบบน และ  $y$  เป็นตำแหน่งของขอบซ้าย ของหน้าต่าง ต่อจากนั้นจะทำการหาดำแหน่งของจุดดำทางซ้ายสุดที่อยู่ในช่วงแถว  $x+1$  ถึง  $x+32-1$  และตามแนวนอน  $0$  ถึง  $y - 1$  โดยตำแหน่งทางซ้ายสุดที่เจอ จะกำหนดให้เป็นขอบซ้ายของหน้าต่างเริ่มต้น และทำการตรวจนับจุดดำในหน้าต่างเริ่มต้นว่ามีค่าเกินกว่า 10 พิกเซลหรือไม่ ถ้าเกินจะกำหนดให้หน้าต่างนี้เป็นหน้าต่างเริ่มต้นในการวางรอบส่วนประกอบ ถ้าไม่ใช่ให้ถือว่ารายละเอียดภายในหน้าต่างนี้เป็นสัญญาณรบกวน และต้องทำการหาดำแหน่งหน้าต่างเริ่มต้นใหม่

4.1.1.2 เดินรอบขอบเขตส่วนประกอบใหญ่ ทำการเดินรอบส่วนประกอบใหญ่โดยเริ่มจากหน้าต่างบนสุดซ้ายสุด เดินวนรอบขอบเขตโดยทิศทาง 4 ทิศ กำหนด  $top$  คือ ขอบบนของหน้าต่าง  $left$  คือ ขอบซ้ายของหน้าต่าง โดยให้ทิศแต่ละค่าดังต่อไปนี้ 0 คือ ไปทางขวา ( $left = left+32$ ) 1 คือ ถัดล่าง ( $top = top+32$ ) 2 คือ ไปทางซ้าย ( $left = left-32$ ) 3 คือ ขึ้นข้างบน ( $top = top-32$ ) หน้าต่างที่เรากำหนดจะมี 2 ลักษณะคือ หน้าต่างดำและหน้าต่างขาว ซึ่งหน้าต่างดำคือ มีจุดดำที่อยู่ในหน้าต่างนั้นมากกว่า 10 จุด ถ้าไม่อยู่ในเงื่อนไขให้เป็นหน้าต่างขาว โดยขั้นตอนการทำงานทิศทางเดินของหน้าต่างเป็นดังนี้

กรณีหน้าต่าง เป็นหน้าต่างดำ

กรณีทิศทางเท่ากับ 0 : ทิศทางเท่ากับ 3

กรณีทิศทางเท่ากับ 1 : ทิศทางเท่ากับ 0

กรณีทิศทางเท่ากับ 2 : ทิศทางเท่ากับ 1

กรณีทิศทางเท่ากับ 3 : ทิศทางเท่ากับ 2

กรณีหน้าต่าง เป็นหน้าต่างขาว

กรณีทิศทางเท่ากับ 0 : ทิศทางเท่ากับ 1

กรณีทิศทางเท่ากับ 1 : ทิศทางเท่ากับ 2

กรณีทิศทางเท่ากับ 2 : ทิศทางเท่ากับ 3

กรณีทิศทางเท่ากับ 3 : ทิศทางเท่ากับ 0

ข้อมูลพิกัดตำแหน่งของหน้าต่าง(ขอบบน ขอบซ้าย ขอบขวา และ ขอบล่าง)ทั้งหมดจะเก็บเป็นลิสต์ลิสต์เพื่อใช้ในหาขอบเขตส่วนประกอบด้านซ้ายและด้านขวา

#### 4.1.1.3 การหาขอบเขตส่วนประกอบใหญ่ (ขอบซ้ายขอบขวา)

1. นำข้อมูลพิกัดหน้าต่างมาทำการจัดเรียงลำดับตามแนวแกน X (ขอบบน) และ Y

(ขอบซ้าย) ด้วยวิธีการ merge sort จากนั้นย่อไปมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. จัดกลุ่มค่าแนวแกน  $x$  ที่มีค่าเดียวกัน และมีค่า  $y$  ของแต่ละพิกัดที่ต่อเนื่องกัน เช่น ชุดแรก  $\{(30,20,61,51),(30,52,61,83),(30,84,61,115),(30,116,61,147),(30,148,61,179)\}$  ชุดที่สอง  $\{(62,20,93,51)\}$  ชุดที่สาม  $\{(62,148,93,179),(62,180,93,211)\}$  เป็นต้น

3. กำหนดขอบซ้ายขวาของแต่ละหน้าต่างแนวแกน  $x$  โดยมีกฎดังนี้

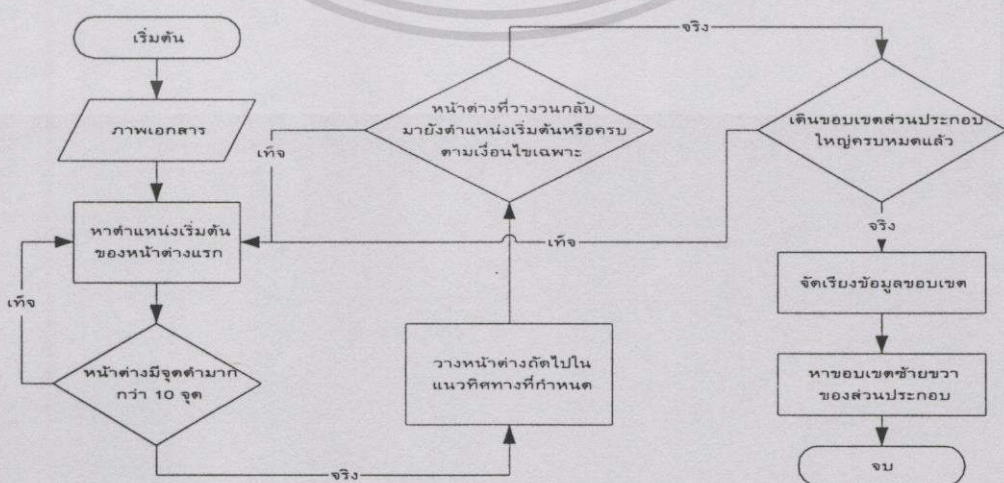
- กำหนดให้ กลุ่มพิกัด( $b$ ) คือ กลุ่มค่าแนวนอน( $x$ ) ที่มีค่าเดียวกัน และมีค่าแนวตั้ง( $y$ ) ของแต่ละพิกัดต่อเนื่องกัน และ ชุดกลุ่มพิกัด( $ub$ ) คือ ชุดของ  $b$  ตั้งแต่ 1 ชุดขึ้นไปที่มีค่า  $x$  เดียวกัน

- กลุ่มพิกัด( $b$ ) เดียวกันที่ไม่ติดกับชุดกลุ่มพิกัด( $ub$ )อื่นที่อยู่เหนือกลุ่มนั้น จะถือว่าเป็นชุดกลุ่มพิกัด( $ub$ ) โดยขอบซ้ายของชุดกลุ่มพิกัด( $ub$ ) คือ ค่า  $y$  ที่น้อยที่สุดของ  $ub$  และ ขอบขวาของ  $ub$  คือ ค่า  $y$  ที่มากที่สุด

- ระยะห่างของกลุ่มพิกัดกับกลุ่มพิกัดก่อนหน้าที่อยู่ในแนวแกน  $x$  เดียวกัน ถ้าขอบบนติดกับขอบล่างชุดกลุ่มพิกัดที่อยู่เหนือมัน และจำนวนจุดค่าที่อยู่ในช่วงที่ห่างระหว่างสองกลุ่มพิกัดมีค่าเกินกว่าจำนวนจุดที่กำหนดไว้ จะรวมสองกลุ่มพิกัดนั้นเข้าด้วยกันแล้วกำหนดขอบซ้ายของ  $ub$  คือ ค่า  $y$  ที่น้อยที่สุด และขอบขวาคือค่า  $y$  ที่มากที่สุด

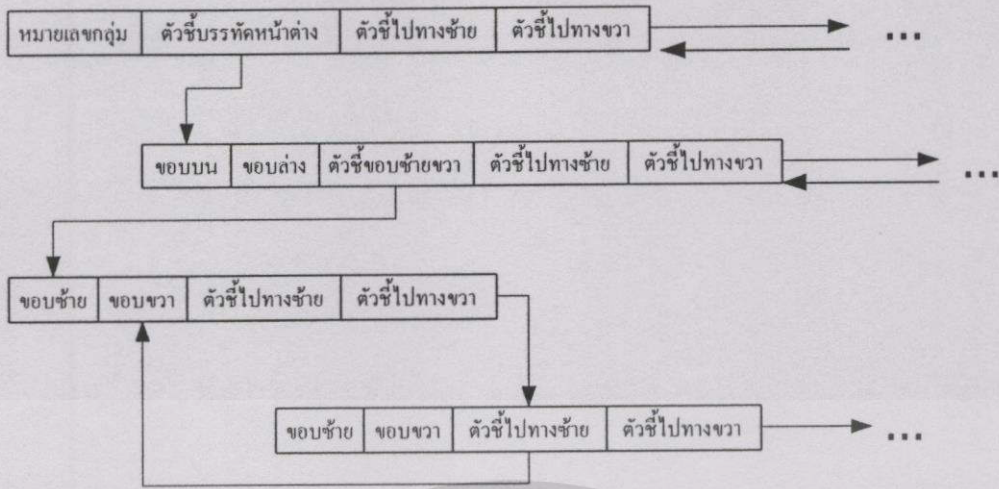
ในรูปที่ 4.2 จะเป็นแผนผังการขั้นตอนการทำงานในการหาขอบเขตใหญ่ และข้อมูลขอบเขตส่วนประกอบใหญ่จะเก็บในโครงสร้างแบบลิงค์ลิสต์ ที่ประกอบไปด้วยกลุ่มส่วนประกอบขอบบนของหน้าต่าง และขอบซ้ายขอบขวา ของกลุ่มหน้าต่างในแต่ละบรรทัดหน้าต่างดังแสดงในรูปที่ 4.3

### แผนผังการทำงานแบ่งส่วนประกอบใหญ่



รูปที่ 4.2 แผนผังการขั้นตอนการทำงานของการแบ่งส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น มิอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 โครงสร้างข้อมูลที่เก็บขอบเขตส่วนประกอบใหญ่

#### 4.1.2 วิธีการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ มีขั้นตอนการทำงานดังนี้

4.1.2.1 การเดินขอบรอบส่วนประกอบย่อย เป็นการเดินรอบขอบแบบ 8 ทิศทาง เพื่อที่จะวนหาพิกัดขอบเขตบนซ้ายและล่างขวา โดยเริ่มทำสแกนจุดที่ละบรรทัดจากทางบนซ้ายของส่วนประกอบใหญ่ ไปจนสุดทางขวา โดยการเดิน 8 ทิศ ได้กำหนดค่าทิศทางดังนี้ 0 คือ ไปทางขวา( $y = y+1$ ) 1 คือ ไปทางขวาล่าง( $x = x+1, y = y+1$ ) 2 คือ ไปข้างล่าง( $x = x+1$ ) 3 คือ ไปซ้ายล่าง( $x = x+1, y = y-1$ ) 4 คือ ไปทางซ้าย( $y = y-1$ ) 5 คือ ไปทางซ้ายบน( $x = x-1, y = y-1$ ) 6 คือ ไปข้างบน( $x = x-1$ ) 7 คือ ไปทางขวาบน( $x = x-1, y = y+1$ ) มีอัลกอริทึมในการทำงานดังนี้

กรณี เป็นจุดดำ

กรณีทิศทางเท่ากับ 0 : ทิศทางเท่ากับ 6

กรณีทิศทางเท่ากับ 1 : ทิศทางเท่ากับ 7

กรณีทิศทางเท่ากับ 2 : ทิศทางเท่ากับ 0

กรณีทิศทางเท่ากับ 3 : ทิศทางเท่ากับ 1

กรณีทิศทางเท่ากับ 4 : ทิศทางเท่ากับ 2

กรณีทิศทางเท่ากับ 5 : ทิศทางเท่ากับ 3

กรณีทิศทางเท่ากับ 6 : ทิศทางเท่ากับ 4

กรณีทิศทางเท่ากับ 7 : ทิศทางเท่ากับ 5

กรณี เป็นจุดขาว

กรณีทิศทางเท่ากับ 0 : ทิศทางเท่ากับ 1

กรณีทิศทางเท่ากับ 1 : ทิศทางเท่ากับ 2

กรณีทิศทางเท่ากับ 2 : ทิศทางเท่ากับ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีทิศทางเท่ากับ 3 : ทิศทางเท่ากับ 4

กรณีทิศทางเท่ากับ 4 : ทิศทางเท่ากับ 5

กรณีทิศทางเท่ากับ 5 : ทิศทางเท่ากับ 6

กรณีทิศทางเท่ากับ 6 : ทิศทางเท่ากับ 7

กรณีทิศทางเท่ากับ 7 : ทิศทางเท่ากับ 0

4.1.2.2 ทำการจัดเก็บพิกัดที่ได้เป็นช่วงของแกน x เรียงต่อกันเป็นแนวบรรทัด และจะเก็บขอบเขตบล็อกส่วนประกอบย่อย โดยข้อมูลพิกัดจะประกอบด้วยนี้ P(ขอบบน , ขอบซ้าย , ขอบล่าง , ขอบขวา , ความสูง , ความกว้าง) ซึ่งข้อมูลทั้งหมดจะเก็บเรียงกันลงถึงคลัสเตอร์ โดยมีเงื่อนไขดังนี้

1. ขอบเขตที่ความสูงและ ความกว้างน้อยกว่า 5 พิกเซล ถือว่าเป็นสัญญาณรบกวน จะไม่นำมาพิจารณา แต่จะเก็บไว้ต่างหาก
2. ขอบเขตที่มีความสูงหรือความกว้างมากกว่า 100 พิกเซลขึ้นไปจะกำหนดให้เป็นรูปภาพ และเก็บไว้ในลิงค์ลิสต์ที่เป็นรูปภาพ
3. ขอบเขตที่ไม่อยู่ในข้อ 1 และ 2 จะถูกกำหนดให้เป็นขอบเขตตัวอักษร ซึ่งจะถูกนำมาจัดกลุ่มกันตามเงื่อนไข ขอบบนหรือขอบล่างจะต้องอยู่ในส่วนขอบเขตใหญ่ของกลุ่ม ซึ่งขอบเขตใหญ่ของกลุ่ม หามาจากขอบบนที่มีค่าน้อยที่สุด และขอบล่างที่มีค่ามากที่สุดในกลุ่ม โดยขอบเขตทั้งหมดจะนำมาเรียงขอบเขตซ้าย จากน้อยไปมาก ดังแสดงในรูปที่ 4.3

{ 93, 843, 173, 923 } h 81 w 81
{ 92, 924, 175, 1015 } h 84 w 92
{ 116, 1016, 175, 1079 } h 60 w 64
{ 117, 1086, 174, 1141 } h 58 w 56
{ 94, 1173, 174, 1257 } h 81 w 85
{ 117, 1266, 176, 1334 } h 60 w 69
{ 119, 1340, 175, 1405 } h 57 w 66
{ 118, 1415, 175, 1471 } h 58 w 57
{ 118, 1473, 176, 1534 } h 59 w 62

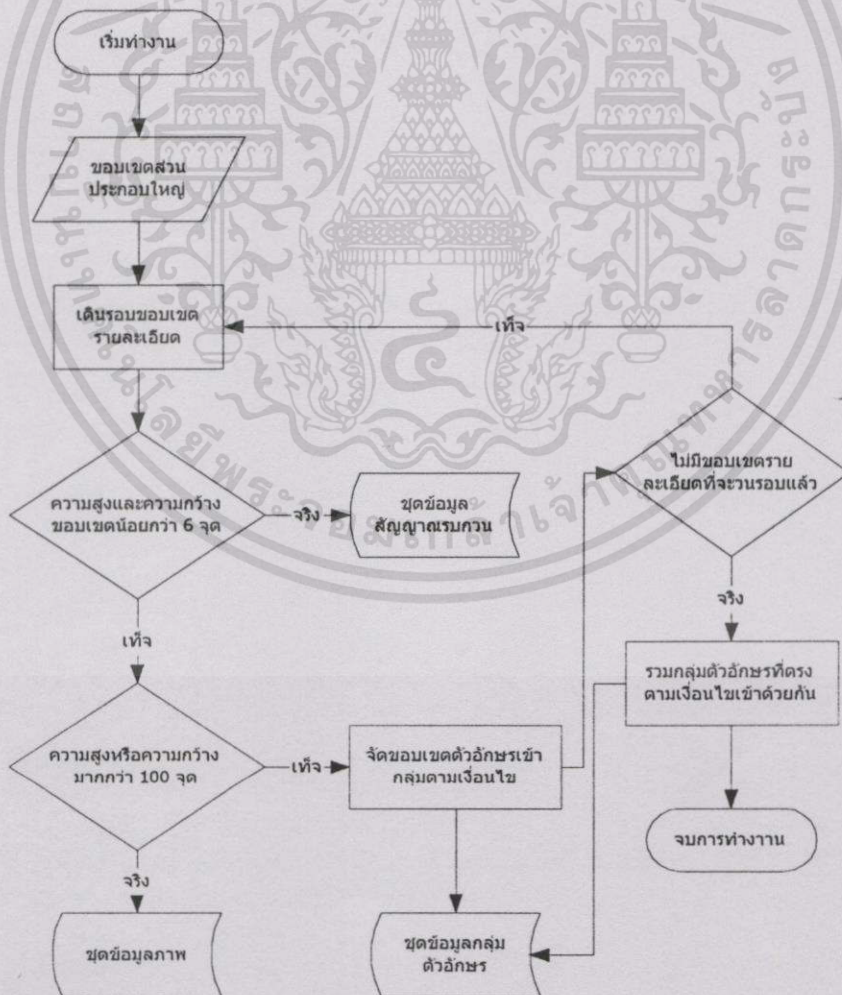
รูปที่ 4.4 ตัวอย่างข้อมูลขอบเขตของส่วนประกอบตัวอักษรในลักษณะการจัดเรียงตามขอบซ้าย

ในกรณีที่ เป็นขอบเขตตัวอักษร ถ้าเป็นขอบเขตตัวอักษรตัวแรกต้องนำไปหาตัวอักษรระดับบรรทัด โดยการสแกนหาจุดค่า 5 แนวทิศทางจากด้านล่างของขอบเขตตัวอักษรลงไปเป็นระยะทางเท่ากับ ความสูงของตัวอักษร แล้วนำขอบเขตตัวอักษรที่สูงที่สุดมาเป็นตัวอักษรระดับบรรทัด เพื่อนำไปใช้หากลุ่มตัวอักษร ซึ่งการหากลุ่มตัวอักษรมีวิธีการดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- สแกนจุดคำ 5 แนว ทิศทางจากด้านบนขวาของขอบเขตตัวอักษรไปเป็นระยะทางเท่ากับความกว้างของขอบเขตตัวอักษร
- สแกนจุดคำ 5 แนว ทิศทางจากด้านบนของขอบเขตตัวอักษรไปเป็นระยะทางเท่ากับครึ่งหนึ่งของความสูงของขอบเขตตัวอักษร
- สแกนจุดคำ 5 แนว ทิศทางจากด้านล่างของขอบเขตตัวอักษรไปเป็นระยะทางเท่ากับครึ่งหนึ่งของความสูงของขอบเขตตัวอักษร
- สแกนจุดคำ 5 แนว ทิศทางจากด้านซ้ายของขอบเขตตัวอักษรไปเป็นระยะทางเท่ากับความกว้างของขอบเขตตัวอักษร

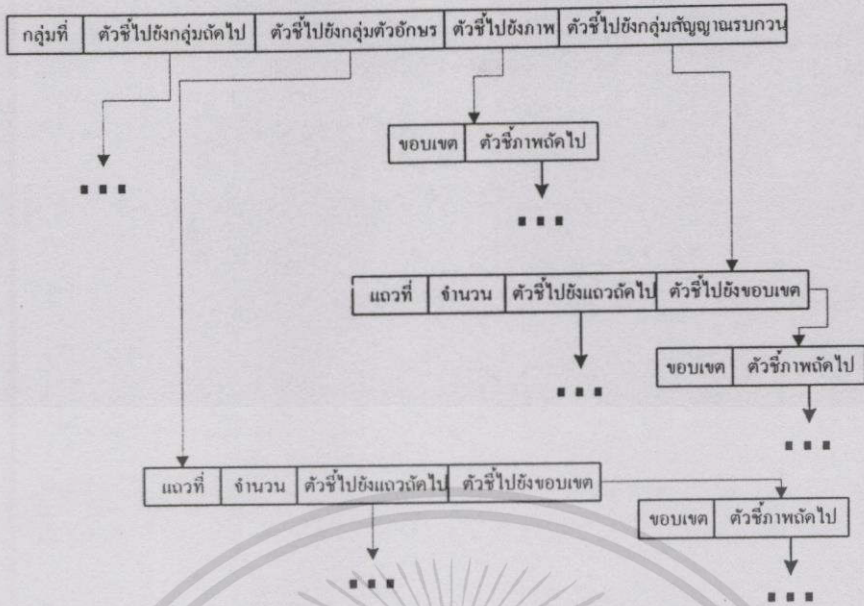
โดยการทำงานนี้จะทำจนกระทั่งไม่เจอขอบเขตตัวอักษร จึงกลับไปเริ่มต้นการทำงานเหมือนเดิมต่อไปจนหมดภาพเอกสาร จากขอบเขตกลุ่มตัวอักษรที่ได้นำมาจัดเรียงกันในระดับบรรทัดเดียวกัน โดยจะแสดงเป็นแผนผังดังในรูปที่ 4.5 ซึ่งขอบเขตรายละเอียดทั้งข้อความ ภาพ และสัญญาณทั้งหมดจะถูกเก็บในลิงค์ลิสต์ที่มีลักษณะดังรูปที่ 4.6



รูปที่ 4.5 แผนผังการทำงานการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนเวลาสำหรับการศึกษาเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า.

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 แสดงการจัดข้อมูลที่แยกประเภทเป็นข้อความ ภาพ และสัญญาณรบกวน

4.1.3. การวิเคราะห์เพื่อแบ่งขอบเขตส่วนประกอบและแบ่งประเภทส่วนประกอบ จากข้อมูลกลุ่มขอบเขตที่ได้จากข้อ 4.2 จะนำมาคำนวณงานวิเคราะห์เพื่อแบ่งประเภทส่วนประกอบให้ถูกต้อง และแบ่งประเภทส่วนประกอบเป็นประเภทข้อความ และส่วนประกอบที่เป็นภาพ ด้วยการดำเนินงานต่อไปนี้

#### 4.1.3.1 การแบ่งส่วนประกอบใหญ่โดยใช้ขอบเขตภาพ

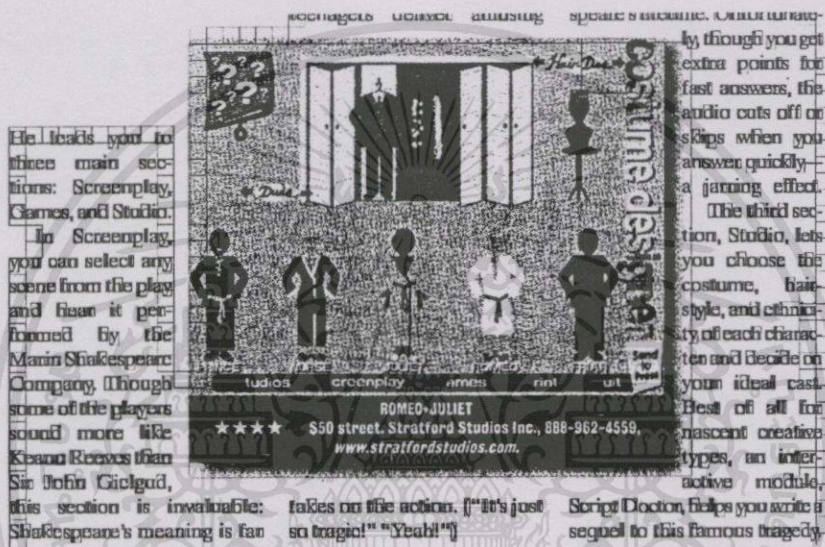
ขั้นตอนการดำเนินงานนี้เริ่มจากนำขอบเขตของภาพที่ได้จาก 4.2 มาทำการลบขอบเขตออกจากขอบเขตส่วนประกอบใหญ่ วัตถุประสงค์เพื่อต้องการให้ขอบเขตใหญ่ที่เหลือเพียงขอบเขตของข้อความและสัญญาณรบกวน โดยวิธีการลบขอบเขตภาพออกจากขอบเขตส่วนประกอบใหญ่จะกระทำโดยนำขอบเขตของภาพ(ขอบเขตของภาพในที่นี้หมายถึง ขอบเขตบน ล่าง ซ้าย ขวา (U, B, L, R)) มาลบกับขอบเขตส่วนประกอบหลักซึ่งประกอบด้วยบรรทัดหน้าต่าง ที่มีขอบเขตซ้ายขวา ดังที่แสดงในรูปที่ 4.7 ซึ่งแสดงให้เห็นแสดงบางส่วนของส่วนประกอบใหญ่ที่มีทั้งภาพและข้อความอยู่ในขอบเขตเดียวกัน ในการแบ่งส่วนประกอบใหญ่ด้วยภาพ ซึ่งจะทำให้จำนวนกลุ่มของส่วนประกอบใหญ่เปลี่ยนแปลงหรือไม่ก็ได้ และชุดของขอบเขตบรรทัดหน้าต่าง อาจจะมีเพิ่มขึ้นหรือลดลงก็ได้ ผลที่ได้จากขั้นตอนนี้จะเป็นขอบเขตส่วนประกอบใหญ่ที่ไม่มีขอบเขตภาพอยู่ภายใน การลบขอบเขตมีเงื่อนไขดังนี้

- ขอบเขตบน ล่าง ซ้าย ขวา ต้องอยู่ในขอบเขตกลุ่มที่ต้องทำการลบ
- ผลต่างในขอบเขตซ้ายหรือขวาของขอบเขตภาพ และขอบเขตในชุดบรรทัดหน้าต่าง ถ้าน้อยกว่า 100 พิกเซลให้ทำการลบขอบเขตชุดบรรทัด

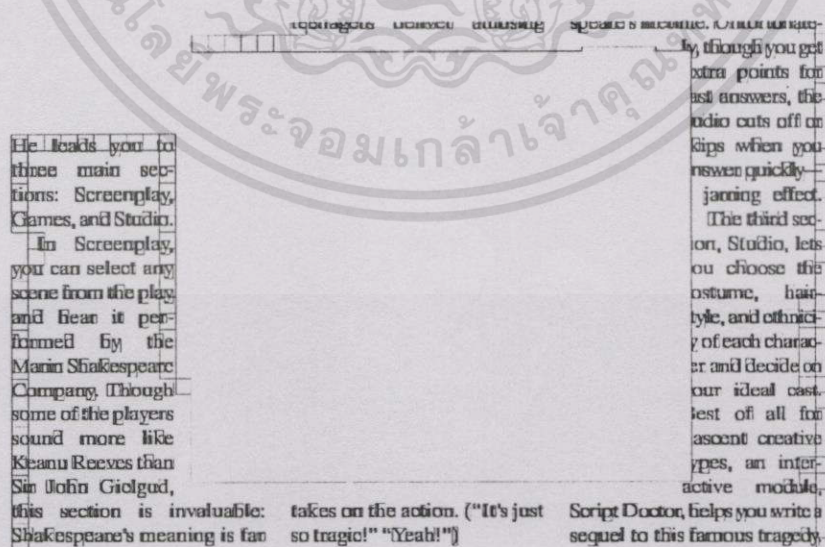
หน้าต่างในด้านที่น้อยกว่า 100 พิกเซลออกไป ถ้า น้อยกว่าทั้งสองด้านให้  
ลบชุดบรรทัดหน้าตานั้นออกไปเลย

- กลุ่มส่วนประกอบถ้ามีบรรทัดหน้าตาดังใดๆ ไม่มีขอบเขตซ้ายขวา ให้ถือ  
ว่าส่วนบรรทัดนั้นเป็นตัวแบ่งกลุ่มส่วนประกอบใหญ่ให้ออกจากกัน

รูปที่ 4.8 แสดงตัวอย่างส่วนประกอบของรูปที่ 4.7 ที่ผ่านขั้นตอนการตัดภาพออกจากส่วน  
ประกอบใหญ่ซึ่งจะได้ขอบเขตของส่วนประกอบใหญ่ใหม่โดยขอบเขตที่เปลี่ยนแปลงถูกเก็บใน  
โครงสร้างข้อมูลแบบลิงค์ลิสต์



รูปที่ 4.7 ขอบเขตส่วนประกอบใหญ่บางส่วนที่ครอบคลุมทั้งข้อความและรูปภาพ



รูปที่ 4.8 ขอบเขตส่วนประกอบใหญ่บางส่วนที่นำขอบเขตภาพออกไป

#### 4.1.3.2 การจัดกลุ่มข้อความให้ตรงกับส่วนประกอบใหญ่

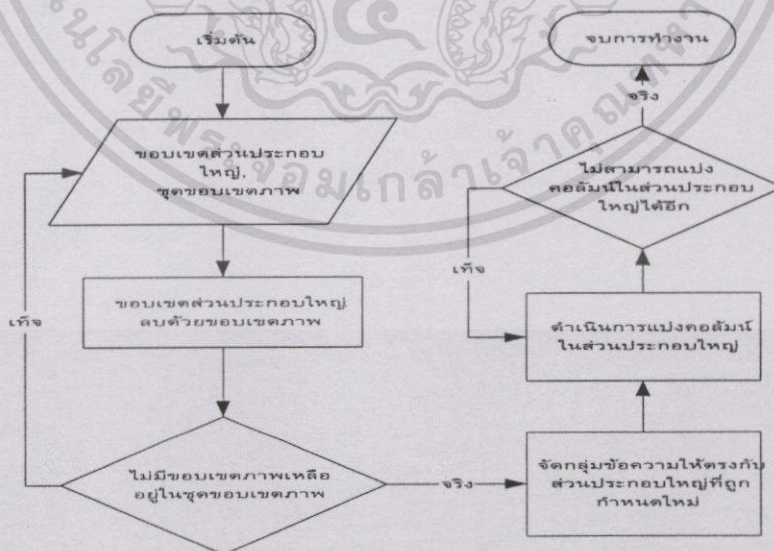
ทำการจัดกลุ่มข้อความให้อยู่ในกลุ่มส่วนประกอบใหญ่อีกครั้ง เนื่องจากอาจจะมี การเปลี่ยนแปลงขอบเขตของกลุ่มข้อความเพราะฉะนั้นต้องมีการจัดกลุ่มส่วนประกอบ ใหญ่ใหม่ เพื่อความถูกต้อง และจัดเรียงหมายเลขกลุ่มส่วนประกอบใหญ่ด้วยอีกครั้ง

#### 4.1.3.3 การแบ่งคอลัมน์ในส่วนประกอบใหญ่

การแบ่งคอลัมน์ในส่วนประกอบใหญ่มีการทำงานดังนี้ นำเอากลุ่มข้อความที่ถูกจัดเรียง บรรทัดจากน้อยไปมาก มาทำการดูขอบเขตของระดับบรรทัดบน แล้วทำการเปรียบเทียบกับ บรรทัดถัดไปว่าช่องว่างระหว่างกลุ่ม(ความกว้างไม่น้อยกว่าขอบเขต 1 ตัวอักษร) สามารถผ่านทะลุไปยังสุดขอบเขตของส่วนประกอบใหญ่หรือไม่ โดยมีเงื่อนไขว่า

- ถ้าช่องว่างนั้นสามารถทะลุไปเกินสามบรรทัดข้อความหรือจนกระทั่งไปถึง สิ้นสุดที่ขอบล่างของส่วนประกอบใหญ่ ถือว่าส่วนที่วิ่งผ่านไปในั้นเป็นคอลัมน์
- ถ้าช่องว่างไม่สามารถทะลุบรรทัดข้อความที่อยู่ข้างล่าง ให้บรรทัดที่อยู่ข้าง ล่างนั้นเป็นบรรทัดเริ่มแรกในการตรวจสอบต่อไป

ดังนั้นเมื่อสิ้นสุดขั้นตอนการในงานนี้แล้ว เราได้ขอบเขตการทำงานส่วนประกอบที่เป็น ขอบเขตข้อความใหญ่ที่มีเฉพาะขอบเขตของข้อความที่รูปร่างของขอบเขตไม่แน่นอน และ ขอบเขตภาพแต่ละภาพที่เป็นอยู่ในลักษณะสี่เหลี่ยม แต่กระนั้นยังมีการเก็บสัญญาณรบกวนไว้เพื่อ ชดเชยยามที่ส่งขอบเขตตัวอักษร ไประบบ OCR แล้วข้อมูลตัวอักษรมีไม่ครบ ซึ่งขั้นตอนการดำเนินงานนี้จะแสดงในรูปที่ 4.9



รูปที่ 4.9 แผนผังการทำงานการแบ่งขอบเขตส่วนประกอบและแบ่งประเภทส่วนประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.2 วิธีการทดลองและขั้นตอนการทดลอง

วิธีการทดลอง นำภาพใบนารีที่ผ่านการสแกนแล้วมาทำการประมวลผล โดยจับเวลาเป็น 2 ช่วง คือ ช่วงที่หนึ่งจับเวลาในการทำงานการแบ่งส่วนประกอบใหญ่ด้วยหน้าต่าง 32\*32 และ ช่วงที่สองจับเวลาในการทำงานงานแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่รวมทั้งการแบ่งคอลัมน์และปรับปรุงขอบเขตส่วนประกอบ และนำมาวิเคราะห์หาความถูกต้องโดยการนับขอบเขตส่วนประกอบที่ถูกต้องหลังจากผ่านกระบวนการปรับปรุงขอบเขตส่วนประกอบแล้วเปรียบเทียบกับส่วนประกอบทั้งหมด โดยในการทดลองนี้จะแบ่งส่วนประกอบออกเป็น 2 ส่วน คือ ขอบเขตของข้อความและขอบเขตของรูปภาพ

## 4.3 ผลการทดลอง

ส่วนนี้เป็นผลการทดลองที่ได้ทำการทดสอบกับตัวอย่างเอกสารต่างๆ 28 ตัวอย่าง จากวารสาร 3 หัวเรื่องและนิตยสาร 3 หัวเรื่อง ถูกนำมาสแกนเป็นภาพขาวดำ ด้วยความละเอียด 300 จุดต่อนิ้ว ดังที่แสดงในภาคผนวก ข. จากตารางที่ 4.1ใช้อ้างอิงตัวอย่างที่นำมาทดลองจากภาคผนวก

ตารางที่ 4.1 อ้างอิงตัวอย่างที่นำมาทดลองจากภาคผนวก

ตัวอย่างที่	รูปที่(ภาคผนวก ก)
1, 2, 3, 4, 5, 6	ข.1 { (ก) (ข) (ค) (ง) (จ) (ฉ) }
7, 8, 9, 10	ข.2 { (ก) (ข) (ค) (ง) }
11, 12, 13, 14, 15	ข.3 { (ก) (ข) (ค) (ง) (จ) }
16, 17, 18, 19	ข.4 { (ก) (ข) (ค) (ง) }
20, 21, 22, 23, 24, 25	ข.5 { (ก) (ข) (ค) (ง) (จ) (ฉ) }
26, 27, 28	ข.6 { (ก) (ข) (ค) (ง) }

การทดลองนี้จะวัดความเร็วในการทำงานของอัลกอริทึม และความถูกต้องในการแบ่งส่วนประกอบจากสาขาของผู้ทดลอง โดยความเร็วในการทำงานจะมีหน่วยเป็นวินาที และวัดเวลาในการทำงานเป็น 2 ช่วงเวลา คือ ช่วงแรกการแบ่งขอบเขตส่วนประกอบใหญ่ ช่วงที่สองการแบ่งส่วนประกอบย่อย รวมถึงการวิเคราะห์แบ่งขอบเขตส่วนประกอบและการแบ่งประเภทส่วนประกอบ ส่วนความถูกต้องในการแบ่งส่วนประกอบจะแบ่งให้เป็น 5 ระดับ น้อยที่สุด(0-80%) น้อย(81 – 85%) ปานกลาง(86 - 90%) มาก(91 – 95%) และ มากที่สุด(96 – 100%) โดยเงื่อนไขการแบ่งประเภทส่วนประกอบเราจะแบ่งเป็น 2 ประเภทเท่านั้น คือ ส่วนประกอบชนิดข้อความ หรือ ส่วนประกอบที่ไม่ใช่ข้อความ ผลการทดลองจะแสดงเป็นข้อ้อยู่ตามลำดับต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.3.1 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่หนึ่ง

ผลลัพธ์ที่ได้จากการทดลองช่วงที่หนึ่ง คือ ขั้นตอนการทำงานแบ่งส่วนประกอบใหญ่ เป็นการเดินรอบขอบเขตส่วนประกอบโดยใช้หน้าตาขนาด 32 x 32 จุด ผลลัพธ์ที่ได้จะแสดงในตัวอย่างรูปที่ 4.10 ถึง รูปที่ 4.13

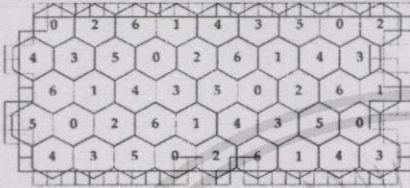


Figure 9. Iteration step assignments.

in a fixed interference environment, therefore calculating the power budget is simplified. Finally, the remaining cells within the neighborhood have exactly one interfering cell which is changing its power; therefore, estimating intracell and intercell interference coefficients ( $\gamma_j$  from section 3.4) is simplified.

So far, we have updated cell power budgets for only a fraction of cells. Our ultimate goal is to update power budgets for all cells in the system. This is achieved by iterating according to a periodic schedule,<sup>8</sup> where the system adjusts different subsets of cells at each iteration; after one period, all cells in the system are updated at least once.

We will demonstrate this algorithm using our example of first order interference and hexagonal cell topology. Figure 9 shows the assignment of cells to iteration steps, and the iteration period equals to 7 in this case. Notice at every time step, the subset of cells which change power levels have non-overlapping neighborhoods.

We next describe how we determine the cell power budgets. The procedure is the same for all cells; however, we will focus on cell #0, with neighboring cells 1 through 6.

When updating the power budget for cell #0, its power level is chosen so as to maximize its overall neighborhood utility. We know from section 3.4 that the maximal cell utility for each cell in the neighborhood of cell #0 depends on  $P_0$ . Therefore, the total neighborhood utility, which is the sum of the maximal cell utilities, is also a function of  $P_0$ , as shown below:

$$U_{\text{neighborhood}}(P_0, P_1, \dots, P_6) = U_{\text{cell},0}(P_0) + \sum_{i=1}^6 U_{\text{cell},i}(P_i) \quad (16)$$

While  $P_0$  is upper bounded by the implementation limit, the goal of adjusting  $P_0$  is to maximize the total neighborhood utility.  $P_0$  thus has to satisfy:

$$\frac{\partial}{\partial P_0} [U_{\text{neighborhood}}(P_0, P_1, \dots, P_6)] = 0 \quad (17)$$

This is equivalent to

$$\frac{\partial}{\partial P_0} [U_{\text{cell},0}(P_0)] - \sum_{i=1}^6 \frac{\partial}{\partial P_0} [U_{\text{cell},i}(P_i)] = 0 \quad (18)$$

<sup>8</sup> This schedule is fixed and determined by the system designer.

Notice that on the left side of (18) is cell #0's marginal utility as a function of  $P_0$ , and on the right is the total marginal utility of the neighbor cells. At the optimal point, with respect to  $P_0$ , the marginal utility of cell #0 offsets the marginal utility of the neighbor cells.

The system utility optimization algorithm we described above is indeed an iterative, one-dimensional search algorithm, in which we optimize along each  $P_i$  iteratively until reaching the maximum. There is a known numerical algorithm [10] that meets this purpose; furthermore, this algorithm has been proven to converge.

### 4. Summary and conclusions

This paper investigated design and control issues faced by multimedia indoor CDMA systems. We have proposed a distributed algorithm that fully utilizes the system resources such as bandwidth and power. Our algorithm integrates three techniques for multimedia downlink transmission: these techniques are: power control, FEC and scheduling. As a multimedia system is designed to satisfy users, our objective is to maximize the overall user satisfaction, which we call "system utility".

We divided this system optimization problem into a hierarchy of three levels: user level, cell level and system level. This partitioning allows us to localize the problem, so that we can perform independent and local optimizations for each user and then each cell. Because users and cells in a CDMA system are subject to bandwidth and power constraints, we achieve the optimization by applying the Lagrange multiplier. The optimization results are then presented to the next level in the hierarchy. At the topmost level, system level, we allocated cell power to one set of cells while keeping their neighboring cell power levels constant. The cell power level is determined based on its effect to the overall utility for its entire neighborhood of seven cells. This layering approach yields a distributed algorithm.

### Acknowledgement

The first author wishes to thank Richard Edell for the helpful discussions.

### References

- [1] S. Sheng, A. Chandrakasan and R. Brodersen, A portable multimedia terminal, IEEE Communications Magazine (1992) 64-75.
- [2] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver and C. E. Wheatley III, On the capacity of a cellular CDMA system, IEEE Transactions on Vehicular Technology, 40 (1991), 311-322.
- [3] R. Yates, A framework for uplink power control in cellular radio systems, IEEE Journal on Selected Areas in Communications 13 (1995) 1341-1347.
- [4] L. Yun and D. Messerschmitt, Power control and coding for variable quality of service on a CDMA channel, in: Proceedings of IEEE MILCOM Conference (1994), pp. 178-182.

รูปที่ 4.10 ภาพตัวอย่างที่ 13 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## Computer Telephony Architectures: MVIP, H-MVIP, and SCbus

Robb Flegg, Mifel

**T**oday, everything seems to revolve around standards. The methods of the past are not good enough for the future. All of the proprietary architectures of the past — all the private branch exchanges (PBXs), central office (CO) switches, and multiplexers (muxes) — are now looked down upon as through inferior. However, the systems of the past did perform functions, and the standards of the future must perform the same functions. Figure 1 shows the architecture currently used by most large systems today. Voice mail applications, file and print services, and others are all separate boxes integrated in an ad hoc manner.

In the late '80s and early '90s many different technologies emerged and became quite specialized. They were predominantly digital signal processing (DSP)-based versions of fax, text to speech, speech recognition and video, generally used as adjunct resources to an existing switching system or for specialized functions such as voice mail or a service bureau. Individual companies tended to specialize in only one area; and it became useful for the different technologies to interact, which led to the development of the first Industry Standard Architecture (ISA) card interconnection bus on the PC. Dialogic's AEB and Peripheral Expansion Bus (PEB) were the first buses of this type; these schemes allowed different technologies to integrate through a common intercard bus.

As the variety of resources, network cards, and DSP resource cards increased, it became clear that the building blocks of the traditional proprietary switch, and many additional high-value blocks (fax, speech recognition) were being created. This spurred the next generation of buses to appear, namely the Multi-Vendor Integration Protocol (MVIP) bus and, later, the SCbus, each with enough bandwidth to be used in almost any type of switching system.

Because it is difficult for one company to supply all the technology it requires, even large multinational corporations are eagerly embracing computer-telephony integration (CTI) standards. Currently the Enterprise Computer Telephony Forum (ECTF) is the most active group in this area, working on both hardware and software standards for building the next generation of computer telephony (CT) systems. This new set of standard application programming interfaces (APIs) and service programming interfaces (SPIs) will be used by all vendors and will have economies of scale to allow small developers or value-added resellers (VARs) to enter the

*With standards and open systems it will be easier for suppliers to meet the changing requirements of the customer.*

*Good standards will allow suppliers to respond faster and more effectively to customer needs.*

telecommunications market in a credible way.

The large proprietary system of the past was not scalable enough to be economic. Today's marketplace requires call centers and voice mail for all companies, no matter how small. In the past, if your current vendors did not support a feature you needed, such as voice processing, it was often expensive for them to integrate it into their proprietary systems, if you could get them to do it at all. With standards and open

systems it will be easier for suppliers to meet the wide range, and changing requirements, of the customer. Good standards will allow suppliers to respond faster and more effectively to customer needs.

In this article, the buses examined will be limited to those that apply to the PC business, and their use in a client-server system. Client-server is the focus of this discussion because a telecommunications bus is not required if each client has its own resources (i.e., the MVIP bus, the High-Capacity MVIP — H-MVIP — bus, and the SCbus). Each of these buses has a function in a client-server setup. The MVIP bus and SCbus are primarily server technology, because they are usually high-density boards; therefore, this article concentrates on client-server systems.

### OBJECTIVES OF A CTI BUS FOR THE PC

A telecommunications bus is made up of three basic components: multiple time division multiplexed (TDM) streams, clock signals, and a mechanism for maintaining network synchronization. We merely assume that such a bus is digital. While it is true that the most common applications are voice, fax, and modems, they perform better digitally. Data streams are more easily manipulated by digital signal processing (DSP) cards, and digital switching is easier and more economic than analog switching. A digital bus also provides a common point of understanding between any two cards on the bus. Without digital technology, we would be unable to build the cost-effective voice mail systems so popular today.

### TIME DIVISION MULTIPLEXING

The TDM portion of a CTI bus is needed to transport data from card to card. The most feasible and cost effective way to do this is still using a bus separate from the computer's

รูปที่ 4.11 ภาพตัวอย่างที่ 18 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## After Hours

### O Romeo, Romeo

A new study aid brings the words of Shakespeare to life for teens.

Remember reading Shakespeare in high school—or trying to? Great stories once you figured out how to leap across the 400-year cultural gap. *Romeo+Juliet*, from Stratford Studios, aims to remedy this problem by using games, interactive studios, and a full audio performance of the play to draw students in and keep them interested. Despite occasionally choppy sound and a sometimes annoying tendency to try to be cool, this CD-ROM proves absorbing and should be helpful to young scholars.

As you begin the program, you're greeted by Lester, an animated jester who serves as a help and navigation guide.

He leads you to three main sections: Screenplay, Games, and Studio.

In Screenplay, you can select any scene from the play and hear it performed by the Marin Shakespeare Company. Though some of the players sound more like Keanu Reeves than Sir John Gielgud, this section is invaluable: Shakespeare's meaning is far easier to grasp when heard than read. Other help includes Myrna, the "director," who sets up the scene for you; Dr. Wacko, who contributes criti-

cal observations; and the Screening Room, where four teenagers deliver amusing

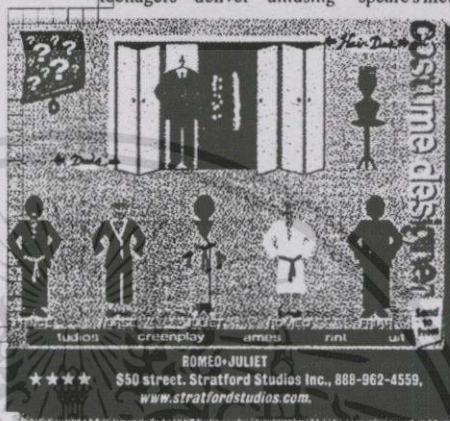
whether events happened before, during, or after Shakespeare's lifetime. Unfortun-

ately, though you get extra points for fast answers, the audio cuts off or skips when you answer quickly—a jarring effect.

The third section, Studio, lets you choose the costume, hairstyle, and ethnicity of each character and decide on your ideal cast. Best of all for nascent creative types, an interactive module,

Script Doctor, helps you write a sequel to this famous tragedy.

Motivated students will have fun with *Romeo+Juliet* and come away with a deeper appreciation for the play. —Carol Mangis



takes on the action. ("It's just so tragic!" "Yeah!")

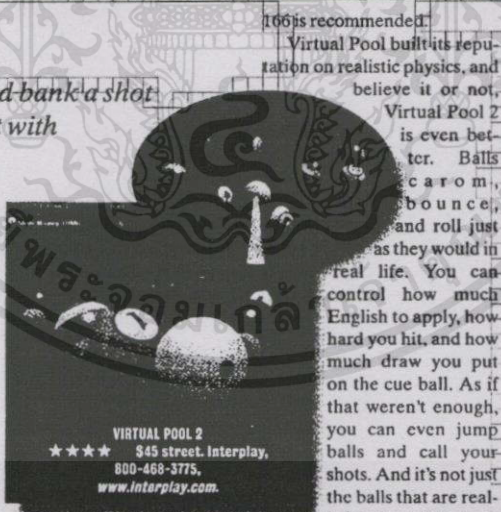
After listening to the play, you can proceed to the five games, which test your knowledge of the play, of literary devices, and of history—

### Rack 'Em Up

Chalk up your cue and bank a shot into the corner pocket with *Virtual Pool 2*.

Billiards is accessible to the athlete and the couch potato, the young and the old. It is this everyman appeal, along with the ultrarealistic look, physics, and game play, that made *Virtual Pool* a classic. And now we have the unexpected sequel, *Virtual Pool 2*, which takes the game one step further.

While not much ink is spilled hyping the graphics of a pool game, rest assured that the detail in this title is fantastic. The balls are as realistic as you're ever going to get, with light sourcing that accurately reflects off their shiny sur-



faces. And that's without the enabling 3-D acceleration. The game runs on a Pentium/90-based PC, but a Pentium

166 is recommended. *Virtual Pool* built its reputation on realistic physics, and believe it or not, *Virtual Pool 2* is even better. Balls carom, bounce, and roll just as they would in real life. You can control how much English to apply, how hard you hit, and how much draw you put on the cue ball. As if that weren't enough, you can even jump balls and call your shots. And it's not just the balls that are realistic; the table is as well. The game gives you the ability to adjust almost everything from table speed to rail-cushioning, allowing you to re-

create your favorite real-world table on your desktop.

But this game isn't just for pros. VR Sports saw to it that everything was provided for the beginner as well as the pool maven. There are on-screen rules for all nine games, along with a new instructional video from pool guru Mike Sigel that tells you everything you need to know, from basic play to advanced strategies.

We found a few glitches in the game, but only after we switched over to 3-D-accelerated mode. The problems consisted of melding graphics and occasional lockups. However, the game looks so good in normal mode that there's really no need to change. Considering how rewarding this game is and the money-back guarantee, we'd say it's time to stop waiting and give *Virtual Pool* a try. —Shane Mooney

รูปที่ 4.12 ภาพตัวอย่างที่ 22 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

using it to evaluate the skill level and solution quality of two different versions of Machinist. Section 5 also describes the results of this evaluation. Section 6 summarizes these results, and Sections 7 and 8 discuss future work and conclusions, respectively.

## 2. A SHORT HISTORY OF THE DEVELOPMENT

### 2.1 Motivations

The development of QUEM arose from the desire to measure the quality of Machinist, a knowledge-based system for manufacturing planning [12]. Machinist creates manufacturing plans for machined parts when given a geometric specification of the part. Initially, we implemented a prototype system to test the basic method.

However, before pushing ahead with a large effort to improve the completeness and robustness of the system, we wanted to be able to evaluate the basic soundness of the approach. In other words, if evaluation were to show that the best solutions produced by the system were of low quality, then we would conclude that our efforts should be focused on re-evaluating our basic approach and problem-solving architecture. On the other hand, if the solutions produced proved to be of high quality then we could feel confident in putting our efforts into further development of the current system architecture. To make this judgment, we needed to find a way to measure the solution quality produced by our evolving system.

### 2.2 Challenges in Developing a Quality Metric

Quality is in general hard to measure because it is hard to quantify. Even if one can generate a function to describe quality, it may be equally hard to quantify the component factors. In our initial attempts to estimate solution quality we attempted to construct just such a quality function composed of factors that our experts believed to be important: plan cost, feasibility, and reliability. However, we soon found this approach to be inadequate. After much tweaking and adjusting of the quality function we found that its assessments still did not agree very closely with those of the experts. Furthermore, we realized that it was unlikely that it would be possible to come up with an accurate quality function for the reason that many of the component factors, such as reliability, were very difficult to quantify accurately.

To describe the problem further, *Plan reliability* is the likelihood that the operations within the plan will fail or will produce marginal results. Plans can fail in catastrophic ways, for example when tools break, or in subtler ways, such as when the resulting product does not meet tolerance requirements. Predicting reliability requires knowledge of a wide variety of situations, which are hard to capture without a large body of empirical data. Because of these difficult to quantify factors, the task of constructing a solution quality metric proved very difficult.

However, we found that experts were able to make quality assessments, and that they tended to agree strongly with each other in those assessments. One reason that experts can succeed in assessing quality where a quality function fails is that experts are able to estimate hard-to-quantify quality factors, such as reliability, because they have a broad range of empirical experience to draw on. Human experts vary some in their assessments, but that

variability can be measured (for example, by having several experts independently rate the same solution) and taken into account. Advantages of this approach are that an experimenter can still measure quality without explicitly knowing the quality function.

Similar difficulties in measuring solution quality arise in many other domains, such as design and scheduling, in which solution quality is judged on a broad range of ill-defined characteristics. The measures described here are well suited for measuring quality in any domain in which a strong correlation between experience and solution quality can be demonstrated.

Next, we needed to devise a scoring system in which human judges could report their quality assessments. The scoring system must allow the quality assessments of different judges to be compared. Initially, we considered having the judges assign quality scores between 1 and 10, like Olympic sports judges, indicating the absolute quality of each plan. However, we decided against such an approach because machinists do not have a standard or agreed upon method for assigning numerical quality measures to plans. We were concerned that it might be difficult to compare scores assigned by two judges; if 10 is the best quality score, an enthusiastic judge might give many 10s while a conservative judge may rarely give a score better than 6. However, the first judges' 10 may mean the same thing as the second judges' 6. We decided, instead, to have the expert judges rank order the plans from best to worst. This makes it easier to compare the quality assessments of different judges.

## 3. RELATED WORK

As mentioned earlier, most literature on knowledge-based system evaluation deals with validation, verification and testing (VVT) [18] in which the primary concern is with correctness, circularity, inconsistency, redundancy, modifiability, ease of use, and cost [14], [15]. However, these properties alone may not be enough to describe a system's competence. A complete, consistent, and nonredundant system may not solve very sophisticated problems, nor cover a broad range of problems. Conversely, a system which performs at a high level of competence, may be neither complete nor consistent.

Clancey [6] describes 4 perspectives useful for evaluating a system's competence: performance, articulation, accuracy (in terms of closeness to human reasoning), and completeness. Other parameters important to system competence are: solution feasibility, solution quality, problem-solving range, computer effort, and user effort. Computer effort refers to the speed at which the computer solves a problem and the number of decisions it must make [17]. User effort refers to the degree to which an expert computer tool increases (or decreases) the effectiveness of a human expert in problem-solving.

MacMillan et al. [16] report on an experiment in which expertise was assessed in a complex domain in which there are no agreed upon expertise measures, and no single right answers to problems. This study has certain similarities with QUEM in the nature of the domain studied and in the way in which data was collected. They studied various solution properties and problem-solving behaviors, and correlated them with expertise. Their goal was to eventually

รูปที่ 4.13 ภาพตัวอย่างที่ 26 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

ผลลัพธ์ที่ได้ นำมาวิเคราะห์แต่ละรูปได้ดังนี้

1. รูปที่ 4.10 ส่วนประกอบที่มีระยะห่างกันเกินกว่าหน้าตาสองช่อง ซึ่งอัลกอริทึมสามารถแยกส่วนประกอบที่อยู่ห่างกันได้ แต่มีส่วนประกอบตอนกลางนับจากบนซ้ายเรียงไปทางขวาล่าง นับได้เป็นส่วนประกอบที่หก ซึ่งเป็นส่วนของข้อความสองคอลัมน์ซึ่งอยู่ใกล้กัน แต่ถูกนำมารวมเข้าด้วยกัน ซึ่งสามารถวิเคราะห์ได้ว่าการทำงานนี้ถ้าส่วนประกอบที่อยู่ใกล้ชิดกันมากจะถูกนำมารวมกันหมด นอกจากนี้ยังสามารถสังเกตได้ว่าส่วนประกอบที่อยู่ส่วนล่างของภาพ ที่เป็นข้อความที่เชื่อมกับสมการ สังเกตได้ว่า จะมีส่วนคอดในแนวนอนระหว่างส่วนประกอบที่อาจจะไม่เป็นส่วนประกอบเดียวกัน

2. รูปที่ 4.11 ส่วนประกอบต่างๆ อยู่ห่างกันและสามารถแยกกันได้ดี แต่เมื่อมาดูตอนล่างของภาพจะพบว่า มีเส้นปิดกันทำให้ส่วนประกอบเชื่อมกันเป็นส่วนประกอบเดียว แต่ขอบเขตที่อยู่ตรงกลางระหว่างคอลัมน์จะเป็นหลุมลงไป

3. รูปที่ 4.12 รายละเอียดของภาพอยู่ติดกันมากทำให้การแบ่งส่วนประกอบของภาพไม่ดี ทำให้ข้อความและรูปภาพ รวมถึงเส้นแบ่งมารวมด้วยกัน ทำให้ไม่สามารถแบ่งรายละเอียดได้ ส่วนที่แบ่งได้ก็คือส่วนประกอบที่อยู่ห่างกัน แต่เมื่อมาดูลักษณะขอบเขตจะพบว่า ขอบเขตจะมีส่วนเว้าเข้าที่มากในส่วนที่จะสามารถแบ่งได้ ที่เห็นได้ชัดคือ ตรงระหว่างคอลัมน์

4. รูปที่ 4.13 ภาพมีส่วนประกอบที่อยู่ห่างกันชัดเจน จึงทำให้ผลลัพธ์ขอบเขตที่ได้ดีที่สุด ซึ่งสามารถ แยกคอลัมน์ออกจากกันได้ แต่ในส่วนย่อหน้า ถ้าห่างกันไม่มากก็ยังมี การเชื่อมต่อกันอยู่แต่เป็นลักษณะคอดกึ่ง

#### 4.3.2 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่สอง

การทดลองช่วงที่สอง คือ การทำการแบ่งขอบเขตส่วนประกอบย่อยในส่วนประกอบใหญ่ รวมถึงการวิเคราะห์เพื่อแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบ โค้ชรูปตัวอย่างผลลัพธ์ จะแสดงในรูปที่ 4.14 ถึง 4.17

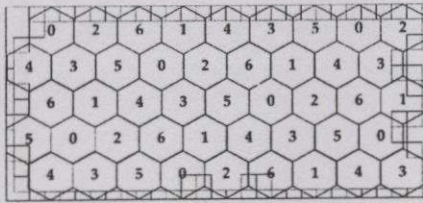


Figure 9. Iteration step assignment.

a fixed interference environment; therefore, calculating the power budget is simplified. Finally, the remaining cells within the neighborhood have exactly one interfering cell, which is changing its power; therefore, estimating intracell and intercell interference coefficients ( $\gamma_{ij}$  from section 3.4) is simplified.

So far, we have updated cell power budgets for only a fraction of cells. Our ultimate goal is to update power budgets for all cells in the system. This is achieved by iterating according to a periodic schedule,<sup>8</sup> where the system adjusts different subsets of cells at each iteration; after one period, all cells in the system are updated at least once.

We will demonstrate this algorithm using our example of first order interference and hexagonal cell topology. Figure 9 shows the assignment of cells to iteration steps, and the iteration period equals to 7 in this case. Notice at every time step, the subset of cells which change power levels have non-overlapping neighborhoods.

We next describe how we determine the cell power budgets. The procedure is the same for all cells; however, we will focus on cell #0, with neighboring cells 1 through 6.

When updating the power budget for cell #0, its power level is chosen so as to maximize its overall neighborhood utility. We know from section 3.2 that the maximal cell utility for each cell in the neighborhood of cell #0 depends on  $U_0$ . Therefore, the total neighborhood utility, which is the sum of the maximal cell utilities, is also a function of  $U_0$ , as shown below:

$$U_{\text{total}}(U_0) = \sum_{i=1}^6 U_{\text{cell},i}(U_0) \quad (16)$$

While  $U_0$  is upper bounded by the implementation limit, the goal of adjusting  $U_0$  is to maximize the total neighborhood utility,  $U_0$  thus has to satisfy:

$$\frac{\partial}{\partial U_0} [U_{\text{total}}(U_0)] = 0 \quad (17)$$

(16) is equivalent to

$$\frac{\partial}{\partial U_0} [U_{\text{cell},i}^*(U_0)] = \frac{\partial}{\partial U_0} \sum_{i=1}^6 [U_{\text{cell},i}^*(U_0)] \quad (18)$$

<sup>8</sup> This schedule is fixed and determined by the system designer.

Notice that on the left side of (18) is cell #0's marginal utility as a function of  $U_0$ , and on the right is the total marginal utility of the neighbor cells. At the optimal point with respect to  $U_0$ , the marginal utility of cell #0 matches the marginal utility of the neighbor cells.

The system utility optimization algorithm we described above is indeed an iterative, one-dimensional search algorithm, in which we optimize along each  $U_i$  iteratively until reaching the maximum. There is a known numerical algorithm [10] that meets this purpose. Furthermore, this algorithm has been proven to converge.

#### 4. Summary and conclusions

This paper investigated design and control issues faced by multimedia in-cell CDMA systems. We have proposed a distributed algorithm that fully utilizes the system resources such as bandwidth and power. Our algorithm integrates three techniques for multimedia downlink transmission; those techniques are: power control, FEC and scheduling. As a multimedia system is designed to satisfy users, our objective is to maximize the overall user satisfaction, which we call "system utility".

We divided this system optimization problem into a hierarchy of three levels: user level, cell level and system level. This partitioning allows us to localize the problem, so that we can perform independent and local optimizations for each user and then each cell. Because users and cells in a CDMA system are subject to bandwidth and power constraints, we achieved this optimization by applying the Lagrange multiplier. The optimization results are then presented to the next level in the hierarchy. At the topmost level, system level, we allocated cell power to a set of cells while keeping their neighboring cell power levels constant. The cell power level is determined based on its effect to the overall utility for its entire neighborhood of seven cells. This layering approach yields a distributed algorithm.

#### Acknowledgement

The first author wishes to thank Richard Bell for the helpful discussions.

#### References

- [1] S. Sriniv, M. Chandrasekar and R. Henderson, Adaptive multimedia terminal, IEEE Communications Magazine (1992) 69-75.
- [2] K. S. Chih-Chen, L. M. Jacob, R. Padovani, B. I. Meech, C. S. Weaver and G. B. Westley III, On the capacity of a cellular CDMA system, IEEE Transactions on Vehicular Technology (1991) 803-810.
- [3] R. Yates, W framework for uplink power control in cellular radio systems, IEEE Journal on Selected Areas in Communications (3) (1993) 1341-1342.
- [4] G. Mun and D. Messerschmitt, Power control and coding for service quality of service on a CDMA channel, in: Proceedings of VECSE & NECTAM, Chulalongkorn (1994) pp. 108-112.

รูปที่ 4.14 ภาพตัวอย่างที่ 13 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ

# Computer Telephony Architectures: MVIP, H-MVIP, and SCbus

Robb Flegg, Mitel

Today, everything seems to revolve around standards. The methods of the past are not good enough for the future. All of the proprietary architectures of the past — all the private branch exchanges (PBXs), central office (CO) switches, and multiplexers (muxes) — are now looked down upon as though inferior. However, the systems of the past did perform functions, and the standards of the future must perform the same functions. Figure 1 shows the architecture currently used by most large systems today. Voice mail, applications, file and print services, and others are all separate boxes integrated in an ad hoc manner.

In the late '80s and early '90s many different technologies emerged and became quite specialized. They were predominantly digital signal processing (DSP)-based versions of fax, text to speech, speech recognition and video, generally used as adjunct resources to an existing switching system or for specialized functions such as voice mail or a service bureau. Individual companies tended to specialize in only one area; and it became useful for the different technologies to interact, which led to the development of the first Industry Standard Architecture (ISA) card interconnection bus on the PC. Dialogic's NEB and Percipical Expansion Bus (PEB) were the first buses of this type; these schemes allowed different technologies to integrate through a common intercard bus.

As the variety of resources, network cards, and DSP resource cards increased, it became clear that the building blocks of the traditional proprietary switch, and many additional high-value blocks (fax, speech recognition) were being created. This spawned the next generation of buses to appear, namely the Multi-Vendor Integration Protocol (MVIP) bus and, later, the SCbus, each with enough bandwidth to be used in almost any type of switching system.

Because it is difficult for one company to supply all the technology it requires, even large multinational corporations are eagerly embracing computer-telephony integration (CTI) standards. Currently the Enterprise Computer Telephony Forum (ECTF) is the most active group in this area, working on both hardware and software standards for building the next generation of computer telephony (CT) systems. The new set of standard application programming interfaces (APIs) and service programming interfaces (SPIs) will be used by all vendors and will have economies of scale to allow small developers or value-added resellers (VARs) to enter the

*With standards and open systems it will be easier for suppliers to meet the changing requirements of the customer. Good standards will allow suppliers to respond faster and more effectively to customer needs.*

telecommunications market in a credible way.

The large proprietary system of the past was not scalable enough to be economic. Today's marketplace requires call centers and voice mail for all companies, no matter how small. In the past, if you current vendors did not support a feature you needed, such as voice processing, it was often expensive for them to integrate it into their proprietary systems, if you could get them to do it at all. With standards and open systems it will be easier for suppliers to meet the wide range, and changing requirements, of the customer. Good standards will allow suppliers to respond faster and more effectively to customer needs.

In this article, the buses examined will be limited to those that apply to the PC business, and their use in a client-server system. Client-server is the focus of this discussion because a telecommunications bus is not required if each client has its own resources (i.e., the MVIP bus, the High-Capacity MVIP — H-MVIP — bus, and the SCbus). Each of these buses has a function in a client-server setup. The MVIP bus and SCbus are primarily server technology, because they are usually high-density boards; therefore, this article concentrates on client-server systems.

## OBJECTIVES OF A CTI BUS FOR THE PC

A telecommunications bus is made up of three basic components: multiple time division multiplexed (TDM) streams, clock signals, and a mechanism for maintaining network synchronization. We merely assume that such a bus is digital. While it is true that the most common applications are voice, fax, and modems, they perform better digitally. Data streams are more easily manipulated by digital signal processing (DSP) cards, and digital switching is easier and more economical than analog switching. A digital bus also provides a common point of understanding between any two cards on the bus. Without digital technology, we would be unable to build the cost-effective voice mail systems so popular today.

## TIME DIVISION MULTIPLEXING

The TDM portion of a CTI bus is needed to transport data from card to card. The most feasible and cost effective way to do this is still using a bus separate from the computers,

รูปที่ 4.15 ภาพตัวอย่างที่ 18 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# After Hours

## O Romeo, Romeo

*A new study aid brings the words of Shakespeare to life for teens.*

Remember reading Shakespeare in high school—or trying to? Great stories, once you figured out how to leap across the 400-year cultural gap. Romeo+Juliet, from Stratford Studios, aims to remedy this problem by using games, interactive studios, and a full audio performance of the play to draw students in and keep them interested. Despite occasionally choppy sound and a sometimes annoying tendency to try to be cool, this CD-ROM proves absorbing and should be helpful to young scholars.

As you begin the program, you're greeted by Dexter, an animated jester who serves as a help and navigation guide.

He leads you to three main sections: Screenplay, Games, and Studio.

In Screenplay, you can select any scene from the play and hear it performed by the Main Shakespeare Company. (Though some of the players sound more like Keanu Reeves than Sir John Gielgud, this section is invaluable: Shakespeare's meaning is far easier to grasp when heard than read. Other help includes Myrna, the "Director," who sets up the scene for you; Dr. MacRo, who contributes crit-

ical observations; and the Screening Room, where four teenagers believe amusing

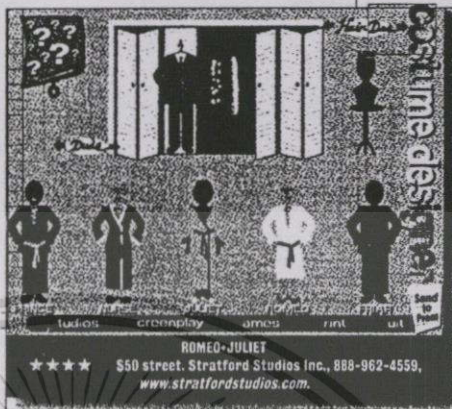
whether the events happened before, during, or after Shakespeare's lifetime. Unfortunately,

though you get extra points for fast answers, the audio cuts off on slips when you answer quickly—a jarring effect.

(The third section, Studio, lets you choose the costume, hair style, and ethnicity of each character and decide on your ideal cast. Best of all for nascent creative types, an interactive module,

Script Doctor, helps you write a sequel to this famous tragedy.

Motivated students will have fun with Romeo+Juliet and come away with a deeper appreciation for the play. —Carol Morley



iques on the action. ("It's just so tragic!" "Yeah!")

After listening to the play, you can proceed to the five games, which test your knowledge of the play, of literary devices, and of history.

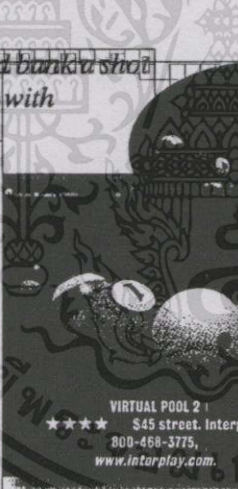
## Rack 'Em Up

*Chalk up your cue and bank a shot into the corner pocket with Virtual Pool 2.*

Billiards is accessible to the athlete and the couch potato, the young and the old. It is this evergreen appeal, along with the classicistic look, physics, and game play, that made Virtual Pool a classic. And now we have the unexpected sequel, Virtual Pool 2, which takes the game one step further.

While not much ink is spilled hyping the graphics of a pool game, rest assured that the detail in this title is fantastic. The balls are as realistic as you're ever going to get, with light scattering that accurately reflects off their shiny sur-

faces. And that's without enabling 3-D acceleration.



faces. And that's without enabling 3-D acceleration. The game runs on a Pentium 90-based PC, but a Pentium

166 is recommended. Virtual Pool built its reputation on realistic physics, and believe it or not,

Virtual Pool 2 is even better. Balls bounce, roll, and spin just as they would in real life. You can control how much English to apply, how hard you hit, and how much draw you put on the cue ball. As if that weren't enough, you can even jump balls and call your shots. And it's not just the balls that are realistic; the table is as

well. The game gives you the ability to adjust almost everything from table speed to rail cushioning, allowing you to re-

create your favorite real-world table on your desktop.

But this game isn't just for pros. VR Sports saw to it that everything was provided for the beginner as well as the pool maven. There are on-screen rules for all nine games, along with a new instructional video from pool guru Mike Sigel that tells you everything you need to know, from basic play to advanced strategies.

We found a few glitches in the game, but only after we switched over to 3-D-accelerated mode. (The problems consisted of melting graphics and occasional lockups. However, the game looks so good in normal mode that there's really no need to change. Considering how rewarding this game is and the money-back guarantee, we'd say it's time to stop wasting and give Virtual Pool a try. —Shane McInay

รูปที่ 4.16 ภาพตัวอย่างที่ 22 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

using it to evaluate the skill level and solution quality of two different versions of Machinist. Section 5 also describes the results of this evaluation. Section 6 summarizes these results, and Sections 7 and 8 discuss future work and conclusions, respectively.

## 2. A SHORT HISTORY OF THE DEVELOPMENT

### 2.1 Motivations

The development of QUEM arose from the desire to measure the quality of Machinist, a knowledge-based system for manufacturing planning [12]. Machinist creates manufacturing plans for machined parts when given a geometric specification of the part. Initially, we implemented a prototype system to test the basic method.

However, before pushing ahead with a large effort to improve the completeness and robustness of the system, we wanted to be able to evaluate the basic soundness of the approach. In other words, if evaluation were to show that the best solutions produced by the system were of low quality, then we would conclude that our efforts should be focused on re-evaluating our basic approach and problem-solving architecture. On the other hand, if the solutions produced proved to be of high quality, then we could feel confident in putting our efforts into further development of the current system architecture. To make this judgment, we needed to find a way to measure the solution quality produced by our evolving system.

### 2.2 Challenges in Developing a Quality Metric

Quality is in general hard to measure because it is hard to quantify. Even if one can generate a function to describe quality, it may be equally hard to quantify the component factors. In our initial attempts to estimate solution quality, we attempted to construct just such a quality function composed of factors that our experts believed to be important: plan cost, feasibility, and reliability. However, we soon found this approach to be inadequate. After much tweaking and adjusting of the quality function we found that its assessments still did not agree very closely with those of the experts. Furthermore, we realized that it was unlikely that it would be possible to come up with an accurate quality function for the reason that many of the component factors, such as reliability, were very difficult to quantify accurately.

To describe this problem further, *Plan reliability* is the likelihood that the operations within the plan will fail or will produce marginal results. Plans can fail in catastrophic ways, for example when tools break, or in subtler ways, such as when the resulting product does not meet tolerance requirements. Predicting reliability requires knowledge of a wide variety of situations, which are hard to capture without a large body of empirical data. Because of these difficult to quantify factors, the task of constructing a solution quality metric proved very difficult.

However, we found that experts were able to make quality assessments, and that they tended to agree strongly with each other in those assessments. One reason that experts can succeed in assessing quality when a quality function fails is that experts are able to estimate hard-to-quantify quality factors, such as reliability, because they have a broad range of empirical experience to draw on. Human experts vary some in their assessments, but that

variability can be measured (for example, by having several experts independently rate the same solution) and taken into account. Advantages of this approach are that an experimenter can still measure quality without explicitly knowing the quality function.

Similar difficulties in measuring solution quality arise in many other domains, such as design and scheduling, in which solution quality is judged on a broad range of ill-defined characteristics. The measures described here are well suited for measuring quality in any domain in which a strong correlation between experience and solution quality can be demonstrated.

Next, we needed to devise a scoring system in which human judges could report their quality assessments. The scoring system must allow the quality assessments of different judges to be compared. Initially, we considered having the judges assign quality scores between 1 and 10, like Olympic sports judges, indicating the absolute quality of each plan. However, we decided against such an approach because machinists do not have a standard or agreed upon method for assigning numerical quality measures to plans. It was soon obvious that it might be difficult to compare scores assigned by two judges; if 10 is the best quality score, an enthusiastic judge might give many 10s while a conservative judge may rarely give a score better than 8. However, the first judges' 10 may mean the same thing as the second judges' 8. We decided, instead, to have the expert judges rank order the plans from best to worst. This makes it easier to compare the quality assessments of different judges.

## 3. RELATED WORK

As mentioned earlier, most literature on knowledge-based system evaluation deals with validation, verification and testing [VMT] [18] in which the primary concern is with correctness, circularity, inconsistency, redundancy, modifiability, ease of use, and cost [14], [15]. However, those properties alone may not be enough to describe a system's competence. A complete, consistent, and nonredundant system may not solve very sophisticated problems, nor cover a broad range of problems. Conversely, a system which performs at a high level of competence may be neither complete nor consistent.

Glancey [6] describes 8 perspectives useful for evaluating a system's competence: performance, articulation, accuracy (in terms of closeness to domain reasoning), and completeness. Other parameters important to system competence are: solution feasibility, solution quality, problem-solving range, computer effort, and user effort. Computer effort refers to the amount of which the computer solves a problem and the number of decisions it must make [12]. User effort refers to the degree to which an expert computer user increases (or decreases) the effectiveness of a human expert in problem-solving.

MacMillan et al. [16] report on an experiment in which expertise was assessed in a complex domain in which there are no agreed upon expertise measures, and no single right answers to problems. This study has certain similarities with QUEM in the nature of the domain studied and in the way in which data was collected. They studied various solution properties and problem-solving behaviors, and correlated them with expertise. Their goal was to experimentally

รูปที่ 4.17 ภาพตัวอย่างที่ 26 หลังจากขั้นตอนการแบ่งประเภทส่วนประกอบ

ผลลัพธ์ที่ได้ นำมาวิเคราะห์แต่ละรูปได้ดังนี้

1. รูปที่ 4.14 สามารถแยกส่วนประกอบภาพและข้อความออกจากกัน ได้สมบูรณ์ และขอบเขตที่ได้ถูกต้อง
2. รูปที่ 4.15 สามารถแยกส่วนประกอบภาพและข้อความออกจากกัน ได้สมบูรณ์ และขอบเขตที่ได้ถูกต้อง
3. รูปที่ 4.16 สามารถแยกส่วนประกอบภาพและข้อความออกจากกันสมบูรณ์ แต่ข้อความบางส่วนจะถูกตัดออกไป และขอบเขตที่ได้ถูกต้อง
4. รูปที่ 4.17 สามารถแยกส่วนประกอบภาพและข้อความออกจากกัน ได้สมบูรณ์ และขอบเขตที่ได้ถูกต้อง

#### 4.3.3 ผลการทดลอง

ผลการทดลองที่ได้ แสดงในตารางที่ 5.2 โดย “ตัวอย่าง” แสดงถึง ลำดับตัวอย่างที่นำมาใช้ในการทดลอง “ขนาดภาพ” แสดงถึง ความสูงและกว้างของภาพหน่วยเป็นจุด “เวลาในช่วงที่หนึ่ง” หมายความว่า เวลาที่ใช้ในการดำเนินการแบ่งขอบเขตส่วนประกอบใหญ่ทั้งหมดต่อหน้าเอกสาร (วินาที) “เวลาในช่วงที่สอง” หมายความว่า เวลาที่ใช้ในการดำเนินการแบ่งขอบเขตรายละเอียดย่อยภายในส่วนประกอบใหญ่ รวมถึงการแบ่งคอลัมน์ในส่วนประกอบใหญ่และการปรับปรุงขอบเขตของส่วนประกอบทั้งหมดต่อหน้าเอกสาร (วินาที) “เวลารวม” คือ เวลาที่ใช้ในช่วงที่หนึ่งร่วมกับเวลาที่ใช้ในช่วงที่สอง (วินาที) “ระดับความถูกต้อง” แบ่งเป็น 5 ระดับคือ มากที่สุด (96-100%) มาก(91-95%) ปานกลาง(86-90%) น้อย(81-85%) และ น้อยที่สุด(0-80%)

ตารางที่ 4.2 แสดงผลการทดลอง

ตัวอย่างที่	ขนาดภาพ(จุด)	เวลาในช่วงที่หนึ่ง (วินาที)	เวลาในช่วงที่ สอง(วินาที)	เวลารวม (วินาที)	ระดับความถูกต้อง
1	3342 x 2104	2.325	10.563	12.888	มากที่สุด
2	3342 x 2128	2.925	13.21	16.135	มากที่สุด
3	3342 x 2096	2.805	10.23	13.035	มากที่สุด
4	3342 x 2128	2.835	11.54	14.375	มากที่สุด
5	3342 x 2128	2.67	10.42	13.09	มากที่สุด
6	3342 x 2120	2.156	13.21	15.366	มากที่สุด
7	3065 x 2223	2.68	13.28	15.96	มากที่สุด
8	3090 x 2236	2.41	10.452	12.862	มากที่สุด
9	3090 x 2148	2.97	13.21	16.18	มากที่สุด
10	3090 x 2211	1.025	12.342	13.367	มากที่สุด
11	3355 x 2311	1.564	10.45	12.014	มากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 (ต่อ)

ตัวอย่างที่	ขนาดภาพ(จุด)	เวลาในช่วงที่หนึ่ง (วินาที)	เวลาในช่วงที่ สอง(วินาที)	เวลารวม (วินาที)	ระดับความถูกต้อง
12	3292 x 2362	1	9.26	10.26	มากที่สุด
13	3342 x 2324	1.41	13.58	14.99	มากที่สุด
14	3153 x 2387	2.795	11.423	14.218	มากที่สุด
15	3254 x 2337	3.66	13.63	17.29	มากที่สุด
16	3153 x 2399	3.04	12.42	15.46	มากที่สุด
17	3355 x 2437	2.01	11.75	13.76	มากที่สุด
18	3355 x 2437	2.235	11.52	13.755	มากที่สุด
19	3355 x 2437	2.18	12.41	14.59	มากที่สุด
20	3000 x 2352	0.61	8.72	9.33	มากที่สุด
21	3108 x 2202	0.976	8.862	9.838	มากที่สุด
22	3173 x 2349	0.96	8.63	9.59	มากที่สุด
23	3219 x 2392	0.79	8.43	9.22	มากที่สุด
24	3194 x 2379	1.055	10.532	11.587	มากที่สุด
25	3172 x 2201	1.545	10.253	11.798	มากที่สุด
26	3228 x 2275	0.865	8.962	9.827	มากที่สุด
27	3238 x 2362	2.4	13.23	15.63	มากที่สุด
28	3238 x 2311	1.47	12.567	14.037	มากที่สุด

#### 4.4 วิเคราะห์ผลการทดลอง

1. ผลการทดลองที่ได้จากการวัดเวลาการทำงานในช่วงที่หนึ่ง แสดงให้เห็นว่า เวลาที่ใช้ในการทำงานนั้นมีความแตกต่างกันในภาพเอกสารแต่ละภาพ ซึ่ง สาเหตุมาจาก ขนาดของภาพ และรายละเอียดของภาพ โดยขนาดของภาพจะอยู่ที่ประมาณ 3300 x 2100 ซึ่งเวลาที่ใช้ในภาพไม่ควรแตกต่างกันมาก แต่เวลาที่ได้จากการทดลองในการทำงานกับตัวอย่างชุดละ 5 ครั้งแล้วนำเวลากลางที่ได้มาบันทึก พบว่าเวลาน้อยที่สุด คือ 0.61 และ เวลาที่มากที่สุดคือ 3.66 วินาที เมื่อลองวิเคราะห์หาสาเหตุแล้ว สาเหตุเกิดจากการที่เดินวนขอบโดยหน้าต่าง แต่หน้าต่างมีขนาดใหญ่กว่าส่วนประกอบทำให้ต้องวางหน้าต่างทับที่เดิมในการเดินวนกลับไปยังจุดเริ่มต้น อัลกอริทึมไม่ได้มีการตรวจสอบในส่วนการวางหน้าต่างซ้ำจุดเดิม จึงทำให้ต้องใช้เวลาในการทำงานเพิ่มขึ้น และเวลาในการทำการในการวนรอบขอบ ขึ้นอยู่กับจำนวนของส่วนประกอบใหญ่ภายในเอกสาร และตำแหน่งในการวางหน้าต่างเริ่มต้นของแต่ละส่วนประกอบ

2. ผลการทดลองที่ได้จากการวัดเวลาการทำงานในช่วงที่สอง แสดงให้เห็นว่า เวลาที่ใช้ในการทำงานมาก เนื่องมาจากการวนรอบขอบรายละเอียดภายในส่วนประกอบใหญ่ทุกส่วน ซึ่งแม้ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การหาขอบเขตนี้ปกติจะถูกหาโดยระบบ OCR เพราะฉะนั้นจึงเป็นการถึงงานในส่วน OCR มาทำในขั้นตอนการแบ่งประเภท จึงทำให้การทำงานในส่วนนี้ใช้เวลามาก

3. รูปแบบที่ได้หลังจากการแบ่งส่วนประกอบ จากขอบเขตส่วนประกอบที่ได้ขอบเขตตามที่ต้องการ เป็นไปตามแนวคิดที่ว่า การหาขอบเขตของส่วนประกอบใหญ่ก่อน เพื่อใช้กำกับการแบ่งส่วนประกอบย่อยๆ ให้ถูกต้อง เป็นไปตามทิศทางที่วางไว้ ผลที่ได้จึงมีการแยกส่วนที่เป็นข้อความ และรูปภาพออกจากกันได้อย่างถูกต้องแม่นยำ

#### 4.5 สรุปผลการทดลอง

การแบ่งขอบเขตส่วนประกอบออกเป็นขนาดใหญ่ก่อน ต่อจากนั้นมากำหนดขอบเขตส่วนประกอบย่อยซึ่งอยู่ภายในส่วนประกอบใหญ่ จะทำให้แบ่งขอบเขตได้ถูกต้อง เพราะมีขอบเขตของส่วนประกอบใหญ่เป็นกรอบกำหนด เปรียบเสมือนกับการเดินทางที่ต้องใช้เข็มทิศมาคอยกำกับ ซึ่งส่วนประกอบต่างๆที่แบ่งได้ จะมีลักษณะขอบเขตที่พอดีกับส่วนประกอบ และไม่ซ้อนทับกัน จึงทำให้นำขอบเขตของส่วนประกอบที่ต้องการไปใช้ได้อย่างถูกต้อง โดยที่การใช้หน้าต่างเดินรอบขอบส่วนประกอบใหญ่ ทำงานได้เร็ว และเหมาะกับเอกสารที่มีส่วนประกอบที่มีระยะห่างกันเกินกว่า 2 หน้าต่างซึ่งสามารถบอกขอบเขตส่วนประกอบได้ทันที โดยไม่ต้องมาผ่านขั้นตอนการแบ่งคอลัมน์ และปรับปรุงขอบเขตส่วนประกอบอีก ผลที่ได้จากการทำงานขอบเขตรายละเอียดที่ได้อยู่ในรูปแบบแผนผังแมนฮัตตันซึ่งมีการอ้างขอบเขตที่พอดีกับส่วนประกอบ และจะไม่มีการอ้างซ้อนทับกันแบบแผนผังรูปสี่เหลี่ยม และการทำงานไม่จำเป็นต้องตรวจสอบทุกจุดในภาพเอกสาร

## สรุปผลการทดลองและเสนอแนะแนวทางการดำเนินงานต่อไป

จากผลการทดลองที่ได้ แสดงให้เห็นว่า การแบ่งข้อมูลได้ถูกต้อง แต่เวลาที่ใช้ในการทำงานนั้นมากกว่าที่ประเมินไว้ เมื่อวิเคราะห์แล้ว สาเหตุที่ทำให้ใช้เวลาในการทำงานนาน คือ การวางหน้าต่างซ้ำกับหน้าต่างเดิมที่วางไว้ก่อนแล้ว เงื่อนไขตรวจสอบขอบเขตส่วนประกอบที่ใช้ในการทำงานมาก การใช้หน่วยความจำในการจัดเก็บข้อมูลมากเกินไป ดังนั้นจึงมีข้อสรุปผลการทดลอง และการเสนอแนะแนวทางแก้ไขดังนี้

### 5.1 สรุปผลการทดลอง

1. การจับเวลาในช่วงที่หนึ่ง มีเวลาที่ใช้ในการทำงานแตกต่างกันมาก ความแตกต่างกันน่าจะมาจากการอัลกอริทึมในการทำงานยังไม่มีควมมั่นคง ต้องมีการปรับปรุงการตรวจสอบการวางหน้าต่างทับไปยังหน้าต่างที่มีอยู่แล้ว เพื่อลดเวลาในการทำงานที่ซับซ้อน
2. การจับเวลาในช่วงที่สอง เวลาที่ใช้ในการดำเนินการช่วงนี้มาก เพราะเป็นการดึงงานการหาขอบเขตตัวอักษรในระบบ OCR มาทำในขั้นตอนี่ ซึ่งเป็นารลดการทำงานในส่วนระบบ OCR เพราะการหาส่วนประกอบที่เป็นตัวอักษร ต้องทำอยู่แล้วดังนั้นเวลาการทำงานที่เพิ่มในขั้นตอนี่ มีเหตุผลเพียงพอที่ยอมรับในเวลาการทำงานเพิ่มขึ้น
3. การสร้างโปรแกรมที่ใช้ในการทดสอบ ลักษณะโครงสร้างของข้อมูลเป็นแบบไม่คงที่ ฉะนั้นการใช้โครงสร้างข้อมูลแบบลิงค์ลิสต์ อาจจะมีการลบหน่วยความจำที่ไม่ใช้ไม่หมด ในการทำงานต่างๆ จึงทำให้เกิดการเขียนข้อมูลลงบนฮาร์ดดิสก์แทนหน่วยความจำหลัก ซึ่งทำให้เพิ่มเวลาทำงานอีก
4. การเขียนโปรแกรมที่มีการเก็บรายละเอียดข้อมูลต่างๆเพื่อใช้ในการตรวจสอบมากเกินไป ทำให้ต้องเปลืองหน่วยความจำ
5. ขอบเขตส่วนประกอบที่ได้เป็นรูปแบบแผนผังแมนฮัตตัน
6. การตรวจสอบจุดภายในภาพเอกสารไม่จำเป็นต้องทำทุกจุด

### 5.2 ข้อเสนอแนะ

1. ปรับปรุงอัลกอริทึมในการนำหน้าต่างมาเดินรอบขอบส่วนประกอบใหญ่

การสร้างขอบเขตจากส่วนประกอบใหญ่นั้น จากผลการทดลองแสดงให้เห็นว่า เวลาที่ใช้ในการทำงานมีช่วงเวลาที่ต่างกันมาก ทั้งที่รายละเอียดของส่วนประกอบใกล้เคียงกัน ขนาดภาพใกล้เคียงกัน แสดงว่าการจัดวางหน้าต่างรอบขอบเขตยังมีอัลกอริทึมในการทำงานไม่ดีพอ เนื่องจากไม่มีการตรวจสอบการวางซ้ำหน้าต่างที่ทำการเดินไปแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. ลดการใช้หน่วยความจำ

ในการแบ่งข้อมูลให้เล็กลงๆ ไปเรื่อยๆ จะทำให้ต้องเก็บข้อมูลมากขึ้นเรื่อยๆ ทำให้ระบบต้องดูแลหน่วยความจำมากขึ้น ซึ่งเป็นผลให้ต้องมีการเขียนข้อมูลลงมาไว้ที่ฮาร์ดดิสก์เพื่อต้องการดึงข้อมูลอื่นไปไว้ในหน่วยความจำแทน เพราะฉะนั้นถ้าลดการจัดเก็บขอบเขตข้อมูลลง สามารถที่จะลดการใช้หน่วยความจำลงได้ โดยออกแบบโครงสร้างในการจัดเก็บข้อมูลและต้องมีการจัดการหน่วยความจำให้ดี

## 3. การลดการสูญหายจากการอ้างอิงส่วนประกอบย่อยเป็นขอบเขตสี่เหลี่ยม

การเดินทางรอบขอบเขตส่วนประกอบย่อยเมื่อมีการตรวจสอบว่ามีขอบเขตที่กว้างหรือสูงเกิน 100 พิกเซล ให้ทำการเดินรอบและเก็บขอบเขตแนวทางเดินแบบ 8 ทิศทางตามเข็มนาฬิกา เพื่อมาหาค่าขอบซ้ายขวา จากแนวการเดิน โดยการเดินไปทางซ้าย และการเดินไปทางขวา สามารถหาขอบซ้ายขวาได้จาก ค่าที่น้อยที่สุดทางซ้ายและมากที่สุดทางขวา และค่าครึ่งบนจะเป็นขอบซ้าย ส่วนค่าครึ่งล่างจะเป็นขอบขวา ซึ่งถ้าภาพมีรูปร่างไม่แน่นอนบางทีอาจจะไปทับขอบเขตตัวอักษรได้ เพราะฉะนั้นถ้าเก็บขอบเขตที่มีรูปแบบไม่แน่นอนจะดีกว่าการเก็บขอบเขตแบบสี่เหลี่ยม แต่จะเพิ่มเนื้อที่ในการเก็บข้อมูลมากขึ้นด้วย

## 5.3 แนวทางการดำเนินการ

จากวิธีการที่เสนอไปนี้สามารถนำไปเพิ่มในระบบจัดเก็บเอกสารและระบบ OCR เพื่อเพิ่มประสิทธิภาพในการทำงานให้สะดวกขึ้น โดยมีการสร้างฐานข้อมูลรวม ในการจัดเก็บแผนผังของภาพเอกสารต่างๆ เพื่อนำมาเปรียบเทียบแบบแผนผังที่ได้จากภาพเอกสารใหม่ ส่วนแนวทางปรับปรุงความเร็วในการแบ่งคอลลัมน์ให้เร็วขึ้น ทำโดยการหาพิกัดบรรทัดข้อความแรกของส่วนประกอบ โดยจะทำการเดินขอบรอบตัวอักษรเฉพาะส่วนที่บรรทัดแรกเท่านั้นแล้วทำการฉายเงาภาพจุดขาว เพื่อหาแนวคอลลัมน์ที่จะทำการแบ่ง โดยการไปจนสุดเป็นบรรทัดที่กำหนดไว้ จะเริ่มนำบรรทัดนั้นเป็นบรรทัดเริ่มต้นของส่วนประกอบใหม่ และทำการแบ่งต่อไปเรื่อยๆจนหมดขอบเขตส่วนประกอบ หรือสิ้นสุดที่ขอบเขตล่างของส่วนประกอบ ซึ่งจะสามารถเพิ่มความเร็วในการแบ่งคอลลัมน์ได้ เพราะไม่ต้องทำการเดินขอบรอบตัวอักษร แต่ความถูกต้องในการแบ่งคอลลัมน์ก็จะลดลงตามไปด้วย

## เอกสารอ้างอิง

- [1] Okun O., Pietikäinen M. "A Survey of Texture-based Methods for Document Layout Analysis." Proceedings of Workshop on Texture Analysis in Machine Vision, Oulu, Finland, 1999. pp. 137-148.
- [2] Wang S.Y., Yagasaki T. "Block Selection: A Method for Segmenting Page Image of Various Editing Styles." Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995. pp. 128-133.
- [3] Sauvola J., Pietikäinen M. "Page Segmentation and Classification using Fast Feature Extraction and Connectivity Analysis." Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995. pp. 1127-1131.
- [4] Payne J.S., Stonham T.J., Patel D. "Document Segmentation using Texture Analysis." Proceedings of the 12<sup>th</sup> ICPR, Jerusalem, Israel, 1994. pp. 380-382.
- [5] Antonacopoulos A., Ritchings R.T. "Flexible Page Segmentation using the Background." Proc. of the 12<sup>th</sup> ICPR, Jerusalem, Israel, 1994. pp. 339-344.
- [6] Antonacopoulos A. "Page Segmentation using the Description of the Background." Computer Vision and Image Understanding, vol. 70, 1998. pp. 350-369.
- [7] Baird H.S. "Background Structure in Document Images." Advances in Structural and Syntactic Pattern Recognition, World Scientific, Singapore, 1992. pp. 253-269.
- [8] Normand H.S., Viard-Gaudin C. "A Background Based Adaptive Page Segmentation Algorithm." Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995. pp. 138-141.
- [9] Cinque L., Lombardi L., Manzini G. "A Multiresolution Approach for Page Segmentation." Pattern Recognition Letters, vol. 19, 1998. pp. 217-225.
- [10] Etemad K., Doermann D., Chellappa R. "Multiscale Segmentation of Unstructured Document Pages using Soft Decision Integration." IEEE Trans. on PAMI, vol. 19, 1997. pp. 92-96.
- [11] Jain A.K., Zhong Y. "Page Segmentation using Texture Analysis." Pattern Recognition, vol. 29, 1996. pp. 743-770.
- [12] Fan K.C., Liu C.H., Wang Y.K. "Segmentation and Classification of Mixed Text/Graphics/Image Documents." Pattern Recognition Letters, vol. 15, 1994. pp. 1201-1209.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [13] Fan K.C., Wang L.S. "Classification of Document Blocks using Density Feature and Connectivity Histogram." *Pattern Recognition Letters*, vol. 16, 1995. pp. 955–962.
- [14] Gatos B., Mantzaris S.L., Chandrinos K.V., Tsigris A., Perantonis S.JJ. "Integrated Algorithms for Newspaper Page Decomposition and Article Tracking." *Proceedings of the 5<sup>th</sup> ICDAR*, Bangalore, India, 1999. pp. 559–562.
- [15] Le D.X., Thoma G.R., Wechsler H. "Classification of Binary Document Images into Textual or Non-textual Data Blocks using Neural Network Models." *Machine Vision and Applications*, vol. 8, 1995. pp. 289–304.
- [16] Shih F.Y., Chen S.S. "Adaptive Document Block Segmentation and Classification." *IEEE Trans. on SMC – Part B, Cybernetics*, vol. 26, 1996. pp. 797–802.
- [17] Etemad K., Doermann D., Chellappa R. "Page Segmentation using Decision Integration and Wavelet Packets." *Proceedings of the 12<sup>th</sup> ICPR*, Jerusalem, Israel, 1994. pp. 345–349.
- [18] Chen W.Y., Chen S.Y. "Adaptive Page Segmentation for Color Technical Journals' cover Images." *Image and Vision Computing*, vol. 16, 1998. pp. 855–877.
- [19] Jain A.K., Yu B. "Automatic Text Location in Images and Video Frames." *Pattern Recognition*, vol. 31, 1998. pp. 2055–2076.
- [20] Messelodi S., Modena C.M. "Automatic Identification and Skew Estimation of Text Lines in Real Scene Images." *Pattern Recognition*, vol. 32, 1999. pp. 791–810.
- [21] Park T., Kim D., Chung K. "Orientation and Scale Invariant Text Region Extraction in WWW Images." *Proceedings IAPR Workshop on Machine Vision and Applications*, Chiba, Japan, 1998. pp. 290–293.
- [22] Zhong Y., Karu K., Jain A.K. "Locating Text in Complex Color Images." *Proceedings of the 3<sup>rd</sup> ICDAR*, Montréal, Canada, 1995. pp. 146–149.
- [23] Zhou J., Lopresti D. "Extracting Text from WWW Images." *Proceedings of the 4<sup>th</sup> ICDAR*, Ulm, Germany, 1997. pp. 248–252.
- [24] Akiyama T., Hagita N. "Automated Entry System for Printed Documents." *Pattern Recognition*, vol. 23, 1990. pp. 1141–1154.
- [25] Ha J., Haralick R.M., Phillips I.T. "Recursive X-Y Cut using Bounding Boxes of Connected Component." *Proceeding of the 3<sup>rd</sup> ICDAR*, Montréal, Canada, 1995. pp. 952–955.
- [26] Jain A.K., Bhattacharjee S. "Text Segmentation using Gabor Filters for Automatic Document Processing." *Machine Vision and Application*, vol. 5, 1992. pp. 169–184.

- [27] Chen J.L. "A Simplified Approach to the HMM Based Texture Analysis and Its Application to Document Segmentation." *Pattern Recognition Letters*, vol. 18, 1997. pp. 993-1007.
- [28] Wang D., Srihari S.N. "Classification of Newspaper Image Blocks using Texture Analysis." *Computer Vision, Graphics, and Image Processing*, vol. 47, 1989. pp. 327-352.
- [29] Liu J., Tang Y.Y., Suen C.Y. "Chinese Document Layout Analysis Based on Adaptive Split-and-merge and Qualitative Spatial Reasoning." *Pattern Recognition*, vol. 30, 1997. pp. 1265-1278.
- [30] Srihari S.N., Hong T., Srikantan G. "Machine Printed Japanese Document Recognition." *Pattern Recognition*, vol. 30, 1997. pp. 1301-1313.
- [31] Sobottka K., Bunke H., Kronenberg H. "Identification of Text on Colored Book and Journal Covers." *Proceedings of the 5<sup>th</sup> ICDAR, Bangalore, India, 1999.* pp. 57-62.
- [32] Drivas D., Amin A. "Page Segmentation and Classification Utilizing Bottom-up Approach." *Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995.* pp. 610-614.
- [33] Hönes F., Lichter J. "Layout Extraction of Mixed Mode Documents." *Machine Vision and Applications*, vol. 7, 1994. pp. 237-246.
- [34] Jain A.K., Yu B. "Page Segmentation using Document Model." *Proceedings of the 4<sup>th</sup> ICDAR, Ulm, Germany, 1997.* pp. 34-38.
- [35] Zhong Y., Karu K., Jain A.K. "Locating Text in Complex Color Images." *Pattern Recognition*, vol. 28, 1995. pp. 1523-1535.
- [36] Sivaramakrishnan R., Phillips I.T., Ha J., Subramaniam S., Haralick R.M. "Zone Classification in a Document using the Method of Feature Vector Generation." *Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995.* pp. 541-544.
- [37] Antonacopoulos A., Ritchings R.T. "Representation and Classification of Complex-shaped Printed Regions using White Tiles." *Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995.* pp. 1132-1135.
- [38] Pavlidis T., Zhou H. "Page Segmentation and Classification." *Computer Vision Graphics and Image Processing*, vol. 54, 1992. pp.484-496.
- [39] Hori O. "A Video Text Extraction Method for Character Recognition." *Proceedings of the 5<sup>th</sup> ICDAR, Bangalore, India, 1999.* pp. 25-28.
- [40] Okun O, Doermann D., Pietikäinen M. "Page Segmentation and Zone Classification: the State of the Art." *Technical Report LAMP-TR-036, University of Maryland, Language and Media Processing Laboratory, 1999.* 34 p.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [41] Kise K., Sato A., Iwata M. "Segmentation of Page Images using the Area Voronoi Diagram." *Computer Vision and Image Understanding*, vol. 70, 1998. pp. 370-382.
- [42] Tang Y.Y., Ma H., Mao X., Liu D., Suen C.Y. "A New Approach to Document Analysis Based on Modified Fractal Signature." *Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995.* pp. 567-570.
- [43] Ishitani Y. "Document Layout Analysis Based on Emergent Computation." *Proceedings of the 4<sup>th</sup> ICDAR, Ulm, Germany, 1997.* pp. 45-50.
- [44] Hase H., Shinokawa T., Yoneda M., Sakai M., Maruyama H. "Character String Extraction from a Color Document." *Proceedings of the 5<sup>th</sup> ICDAR, Bangalore, India, 1999.* pp. 75-78.
- [45] Li H., Doermann D., Kia O. "Automatic Text Detection and Tracking in Digital Video." to appear in *IEEE Trans. on Image Processing-Special Issue on Image and Video Processing for Digital Libraries, 1999.*
- [46] Jain A.K., Yu B. "Document Representation and Its Application to Page Decomposition." *IEEE Trans. on PAMI*, vol. 20, 1998. pp. 294-308.
- [47] Kanai J., Rice S.V., Nartker T.A., Nagy G. "Automatic Evaluation of OCR Zoning." *IEEE Trans. on PAMI*, vol. 17, 1995. pp. 86-90.
- [48] Antonacopoulos A., Brough A. "Methodology for Flexible Analysis of the Performance of Page Segmentation Algorithms." *Proceedings of the 5<sup>th</sup> ICDAR, Ranganore, India, 1999.* pp. 451-454.
- [49] Kanungo T., Haralick R.M. "An Automatic Closed-loop Methodology for Generating Character Groundtruth for Scanned Documents." *IEEE Trans. on PAMI*, vol. 21, 1999. pp. 179-183.
- [50] Yanikoglu B.A., Vincent L. "Ground-truthing and Benchmarking Document Page Segmentation." *Proceedings of the 3<sup>rd</sup> ICDAR, Montréal, Canada, 1995.* pp. 601-604.
- [51] Yanikoglu B.A., Vincent L. "Pink Panther: A Complete Environment for Ground-truthing and Benchmarking Document Page Segmentation." *Pattern Recognition*, vol. 31, 1998. pp. 1191-1204.
- [52] Junker M., Hoch R., Dengel A. "On the Evaluation of Document Analysis Components by Recall Precision and Accuracy." *Proceedings of the 5<sup>th</sup> ICDAR, Bangalore, India, 1999.* pp. 713-716.

- [53] Jain A., Yu B. "Document Representation and Its Application to Page Decomposition." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, March 1998.
- [54] Wahl F., Wong K., Casey R. "Block Segmentation and Text Extraction in Mixed Text/Image Document." Computer Vision, Graphics, and Image Processing, vol. 20, 1982. pp. 375-390.
- [55] Nagy G., Seth S. "Hierarchical Representation of Optically Scanned Documents." Proc. Seventh Int'l Conf. Pattern Recognition, Montreal, 1984. pp. 347-349.
- [56] Fujisawa H., Nakano Y. "A Top-Down Approach for the Analysis of Document." Proc. 10<sup>th</sup> Int'l Conf. Pattern Recognition, Atlantic City, N.J., 1990. pp. 113-122.
- [57] Fisher J., Hinds S., D'Amato K. "A Rule-Based System for Document Image Segmentation." Proc. 10<sup>th</sup> Int'l Conf. Pattern Recognition, Atlantic City, N.J., 1990. pp. 567-572.
- [58] Pavlidis T., Zhou J. "Page Segmentation by White Streams." Proc. First Int'l Conf. Document Analysis and Recognition, Saint-Malo, France, 1991. pp. 945-953.
- [59] Baird H. "Anatomy of a Versatile Page Reader." Proc. IEEE, vol. 80, 1992. pp. 1,059-1,065.
- [60] Legourgiois F., Bublinski Z., Emptoz H. "A Fast and Efficient Method for Extracting Text Paragraphs and Graphics From Unconstrained Documents." Proc. 11<sup>th</sup> Int'l Conf. Pattern Recognition, The Hague, 1992. pp. 272-276.
- [61] Akindele O., Belaid A. "Page Segmentation by Segment Tracing." Proc. Second Int'l Conf. Document Analysis and Recognition, Tsukuba, Japan, 1993. pp. 341-344.
- [62] Amamoto N., Torigoe S., Hirogaki Y. "Block Segmentation and Text Area Extraction of Vertically/Horizontally Written Document." Proc. Second Int'l Conf. Document Analysis and Recognition, Tsukuba, Japan, 1993. pp. 739-742.
- [63] Ittner D., Baird H. "Language-Free Layout Analysis." Proc. Second Int'l Conf. Document Analysis and Recognition, Tsukuba, Japan, 1993. pp. 336-340.
- [64] O'Gorman L. "The Document Spectrum for Page Layout Analysis." IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, 1993. pp. 1,162-1,173.
- [65] Antonacopoulos A., Ritchings R. "Representation and Classification of Complex-Shaped Printed Regions Using White Tiles." Proc. Third Int'l Conf. Document Analysis and Recognition, Montreal, 1995. pp. 1,132-1,135.

- [66] Zlatopolsky A. "Automated Document Segmentation." *Pattern Recognition Letters*, vol. 15, 1994. pp. 699-704.
- [67] Doermann D. "Page Decomposition and Related Research." *Proc. Symp. Document Image Understanding Technology*, Bowie, Md., 1995. pp. 39-55.
- [68] Ha J., Haralick R., Phillips I. "Document Page Decomposition by the Bounding-Box Projection Technique." *Proc. Third Int'l Conf. Document Analysis and Recognition*, Montreal, 1995. pp. 1,119-1,122.
- [69] Sylwester D., Seth S. "Atrainable, Single-Pass Algorithm for Column Segmentation." *Proc. Third Int'l Conf. Document Analysis and Recognition*, Montreal, 1995. pp. 615-618.
- [70] Jain A., Karu K. "Learning Texture Discrimination Masks." *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, 1995. pp. 195-205.
- [71] Kise K., Yanagida O., Takamatsu S. "Page Segmentation Based on Thinning of Background." *Proc. 13<sup>th</sup> Int'l Conf. Pattern Recognition*, Vienna, 1996. pp. 788-792.
- [72] Liu J., Tang Y., He Q., Suen C. "Adaptive Document Segmentation and Geometric Relation Labeling: Algorithms and Experimental Results." *Proc. 13<sup>th</sup> Int'l Conf. Pattern Recognition*, Vienna, 1996. pp. 763-767.
- [73] Yamachita A., Amano T., Hirayama Y., Itoh N., Katoh S., Mano T., Toyokawa K. "A Document Recognition System and Its Applications." *IBM J. Research and Development*, vol. 40, 1996. pp. 341-352.

ภาคผนวก ก  
บทความที่ตีพิมพ์ในวารสาร

บทความที่ตีพิมพ์ในวารสาร มี 1 บทความคือ

- [1] พิเศษพงศ์ สุราพันธ์ และ บุญธิร์ เครือตราชู “วิธีเร็วและมีประสิทธิภาพในการแบ่งส่วนประกอบเอกสารสำหรับระบบจดจำอักษร.” สารเนคเทค, ฉบับที่ 40, 2544.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ECTI: Electronic•Computer•Telecommunication•Information

สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ  
73/1 ถนนพระราม 6 เขตราชเทวี กรุงเทพฯ 10400  
โทร. (662) 644-8150-90 โทรสาร. (662) 644-8137-8

National Science and Technology Development Agency  
73/1 Rama VI rd., Ratchathewi, Bangkok 10400, THAILAND.  
Tel. (662) 644-8150-90 Fax. (662) 644-8137-8

ที่ วว. 5204.0104/018

๒๐ กุมภาพันธ์ ๒๕๕๔

เรื่อง การขอรับบทความ

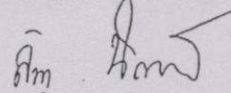
เรียน คุณ พิเศษพงศ์ สุธาพันธ์

เนื่องด้วยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ได้อนุมัติให้ งานประชาสัมพันธ์และนิติตีมีเดีย จัดทำวารสาร "สารเนคเทค" ขึ้น เพื่อใช้เป็นเวทีในเชิงวิชาการและเผยแพร่ข้อมูลข่าวสารด้านคอมพิวเตอร์ เทคโนโลยีสารสนเทศ แก่บุคคลทั่วไปที่สนใจ

ในการนี้ทางกองบรรณาธิการจึงเห็นสมควรนำบทความของท่าน นักศึกษาปริญญาโท คณะเทคโนโลยีสารสนเทศ สาขาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เรื่อง "วิธีเร็วและมีประสิทธิภาพในการแบ่งส่วนประกอบเอกสารสำหรับระบบจดจำอักษร : A Fast and Efficient Method for Document Segmentation for OCR) ลงตีพิมพ์ในวารสารฉบับที่ ๕๐ เดือนพฤษภาคม-มิถุนายน ๒๕๕๔

จึงเรียนมาเพื่อทราบ

ขอแสดงความนับถือ



( ลัญจนา นิตยพัฒน์ )

หัวหน้างานประชาสัมพันธ์และนิติตีมีเดีย

งานประชาสัมพันธ์และนิติตีมีเดีย

โทรศัพท์ 644-8150-9 ต่อ 725

โทรสาร 644-8129

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# วิธีเร็วและมีประสิทธิภาพในการแบ่งส่วนประกอบเอกสารสำหรับระบบจดจำตัวอักษร

## A Fast and Efficient Method for Document Segmentation for OCR

พิเศษพงศ์ สุธาพันธ์ \*

ดร.บุญธีร์ เครือตราชู\*\*

\*นักศึกษาระดับปริญญาโท สาขาวิชาวิทยาการคอมพิวเตอร์ และเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สจล.

\*\*ผู้ช่วยศาสตราจารย์ ภาควิชาคอมพิวเตอร์ คณะวิศวกรรมศาสตร์, สำนักวิจัยการสื่อสารและเทคโนโลยีสารสนเทศ สจล.

### บทคัดย่อ

บทความนี้นำเสนอวิธีการแบ่งส่วนประกอบเอกสารที่รวดเร็วและมีประสิทธิภาพ โดยใช้แนวคิดผสมผสานระหว่างการแบ่งส่วนใหญ่ไปสู่ส่วนย่อย(top-down) และส่วนย่อยไปสู่ส่วนใหญ่(bottom-up) การแบ่งส่วนประกอบใหญ่(คอลัมน์) ของหน้าเอกสารออกจากกันใช้การเดินรอบขอบ(contour following) โดยใช้กรอบหน้าต่างขนาด  $32 \times 32$  พิกเซล เพื่อที่จะสามารถเดินรอบคอลัมน์ที่มีตัวอักษร ห่างกันน้อยกว่า 32 จุด ภาพเอกสารที่ใช้เป็นแบบไบนารีที่มีความละเอียด 300 จุดต่อนิ้ว และอาจมีความห่างระหว่างคอลัมน์น้อยกว่า 32 จุด จึงเป็นไปได้ที่ส่วนประกอบใหญ่อาจประกอบด้วยคอลัมน์มากกว่า 1 คอลัมน์ ในการแยกคอลัมน์ในส่วนประกอบใหญ่ออกจากกัน โดยใช้การหาส่วนประกอบย่อย(ตัวอักษร)ที่อยู่ในส่วนประกอบใหญ่ จากนั้นทำการฉายภาพเงาทางแนวตั้ง (vertical projection) ของช่องว่างระหว่างส่วนประกอบย่อย เพื่อหาแนวแบ่งคอลัมน์ วิธีที่นำเสนอใหม่นี้ มีข้อได้เปรียบจากวิธีอื่น เนื่องจากไม่จำเป็นต้องตรวจสอบทุกจุดภายในภาพเอกสาร โดยในขั้นตอนแรก ถ้าตรวจพบจุดดำในหน้าต่างขนาด  $32 \times 32$  ก็จะถือว่าอยู่ในขอบของเอกสารและจะทำการเลื่อนหน้าต่างต่อไปโดยที่ไม่จำเป็นต้องดูทุกจุดภายในหน้าต่าง ส่วนขั้นตอนที่สองซึ่งเป็นการหาส่วนประกอบย่อยโดยใช้การเดินรอบขอบด้วยหน้าต่างขนาด 1 พิกเซล ซึ่งเข้าถึงเฉพาะขอบของตัวอักษรเท่านั้น ไม่ได้เข้าถึงทุกจุดของภาพเอกสาร ซึ่งขั้นตอนนี้จำเป็นต้องมีอยู่แล้วในระบบ OCR และส่วนขั้นตอนสุดท้ายเป็นการแยกคอลัมน์ข้อความของส่วนประกอบใหญ่ออกจากกัน เพราะวิธีการที่นำเสนอไม่ได้เข้าถึงจุดภาพทุกจุดจึงทำให้ได้การทำงานที่เร็วในการแบ่งส่วนประกอบรายละเอียดของภาพเพื่อใช้กับระบบ OCR

### Abstract

This paper outlines fast and efficient method for page segmentation. The presented method is based on mixed top-down and bottom-up approach to document analysis. The basic idea is that a block of column (paragraph) is extracted by a modified edge following algorithm. Instead of a pixel, a window of 32 by 32 pixel is used in the algorithm so that a paragraph can be extracted instead of a character. The document is scanned at 300 dpi and it is possible to extract multiple columns. In order to separate them, edge following algorithm and their boundaries are used (bottom-up). Since the block extraction only scans through border pixels of paragraphs and characters need to be extracted in the OCR process, this algorithm is faster with fewer overheads than algorithms that need to access all pixel of a document.

### 1. บทนำ

ในปัจจุบันการจัดเก็บเอกสารแบบอัตโนมัติเริ่มมีใช้กันอย่างแพร่หลาย ซึ่งในการทำงานของระบบนี้ การแบ่งส่วนประกอบเอกสารเป็นการทำงานขั้นตอนแรก โดยจะแบ่งส่วนของหน้าเอกสารเป็น บล็อก หรือ ขอบเขตข้อมูลที่มีชนิดเดียวกัน เช่น ข้อความ รูปภาพ ตาราง หรือ อื่นๆ โดยในขั้นตอนนี้จึงมีการศึกษาอย่างกว้างขวาง และมีวิธีการที่ใช้สำหรับการแบ่งส่วนเอกสารหลากหลาย ซึ่งสามารถจัดประเภทการทำงานออกเป็น 2 กลุ่ม คือ วิธีบนลงล่าง(Top-Down) และวิธีล่างขึ้นบน(Bottom-Up) วิธีบนลงล่างเป็นการแบ่งเอกสารให้เป็นส่วนประกอบใหญ่ก่อนแล้วนำไปแบ่งเป็นส่วนประกอบย่อยในส่วนประกอบใหญ่อีก โดยวิธีการนี้ต้องใช้ความรู้เกี่ยวกับโครงสร้างเอกสาร แผนผังที่ได้ส่วนมากจะอยู่ในรูปบล็อกสี่เหลี่ยม ส่วนวิธีการล่างขึ้นบน เป็นการดำเนินงานกับข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์การเขียนเพื่อการศึกษาเท่านั้น เมื่ออยู่ใต้เห็นไปจะยี่ระเษยนต้นกักรัก

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระดับเล็กที่สุดก่อน(ตัวอักษร) แล้วค่อยนำมาประกอบกันเป็นระดับใหญ่ขึ้น (บล็อก , ย่อหน้า) โดยสามารถดูแนวคิดและจุดเด่นของแต่ละวิธีได้จาก [1]

### 1.1 วิธีการบนลงล่าง

วิธีการนี้ได้แก่ การทำให้ภาพเบลอ (Smearing) การตัดโดยวิธีฉายเงาภาพ (Projection profile cuts) โดยโครงสร้างส่วนใหญ่จะอธิบายในรูปแบบผังต้นไม้(tree)

1.1.1 วิธีการทำเบลอ วิธีนี้ภาพต้นฉบับจะต้องเป็นภาพเอกสารสีขาวดำ(Binary) และใช้วิธีการของ Wong et al. [2] ซึ่งได้เสนออัลกอริทึมที่เรียกว่า รั้นเรงส์สมูทติ้งอัลกอริทึม (Run-Length Smoothing Algorithm (RLSA)) หรือ รั้นเรงส์เมออัลกอริทึม(Run-Length Smearing Algorithm) โดยการทำงานจะใช้จุดภาพสีค่าสองจุดใดๆ (1's) ซึ่งมีระยะห่างน้อยกว่าค่าที่กำหนด (t) ที่กำหนดไว้จะถูกรวมกันเป็นจุดค่าที่อยู่ต่อเนื่องกันเป็นแนว ส่วนจุดขาว (0's) ที่ไม่อยู่ในขอบเขตนั้นจะไม่เปลี่ยนแปลง ตัวอย่าง ข้อมูล 00011000001100100001 และ ค่า t เป็น 3 ดังนั้นผลลัพธ์ของ RLSA จะเป็น 11111000001111100001 อัลกอริทึมนี้นำมาใช้กับจุดค่าที่อยู่ห่างกันในแต่ละแถว และจุดค่าที่อยู่ห่างกันในแต่ละคอลัมน์ โดยจุดค่าที่อยู่ใกล้กันภายในค่าที่กำหนดจะต้องถูกเชื่อมเข้าด้วยกัน ผลลัพธ์ของภาพที่ใช้ RLSA จะมีลักษณะเบลอทุกจุดที่ปรากฏบนภาพต้นฉบับ โดยค่ากำหนด  $t_x$  และ  $t_y$  ในสองทิศทางไม่จำเป็นต้องเท่ากัน ส่วนประกอบที่คาดหวังไว้ในแต่ละบล็อกควรเป็นข้อมูลประเภทเดียวกัน ถ้าเป็นข้อความต้องเป็นข้อความทั้งหมด (ข้อความ รูปภาพ ตาราง และ อื่นๆ)

### 1.1.2 การตัดโดยการฉายเงาภาพ (Projection profile cuts)

เทคนิคการแบ่งส่วนประกอบที่ใช้วิธีการตัดโดยการฉายเงาภาพเงาชนิดหนึ่งที่รู้จักกันดี คือ การตัดซ้ำแนว X-Y (recursive X-Y cut) โดยจะนำมาใช้แยกส่วนประกอบเอกสารให้เล็กลงไปเรื่อยๆ ในลักษณะของบล็อกสี่เหลี่ยม โดยแต่ละรอบการทำงานภาพเอกสารจะมีการคำนวณจุดค่าทั้งในแนวตั้งและแนวนอน ซึ่งการแบ่งขอบเขตจะทำตรงช่วงที่ต่ำที่สุด(valley)ของการฉายเงาภาพในแต่ละครั้ง (เพื่อให้เห็นภาพได้ชัดเจน ภาพเอกสารเปรียบได้กับถาดที่มีลูกแก้วอยู่อย่างกระจัดกระจายเมื่อ เอียงถาดไปด้านใดด้านหนึ่ง ทำให้ลูกแก้วไหลมาเรียงกันแน่นที่บริเวณขอบที่เอียงนั้น และจะมีบางช่วงที่ลูกแก้วเรียงต่อกันมีความสูงน้อยกว่าช่วงอื่นๆ หรือไม่มีลูกแก้วเลย ส่วนนั้นก็จะเป็นตำแหน่งที่แบ่ง เมื่อลูกแก้วเปรียบได้กับจุดค่า) และดำเนินการทำซ้ำจนกระทั่งไม่สามารถแบ่งแยกส่วนประกอบออกเป็นส่วนประกอบย่อยได้อีก ส่วนประกอบที่ได้จากการตัดแนว X-Y สามารถนำมาแสดงโครงสร้างส่วนประกอบของภาพเอกสาร ในรูปแบบต้นไม้ซึ่งเก็บข้อมูลขอบเขตส่วนประกอบเป็นขอบเขตสี่เหลี่ยม หรือที่เรียกว่าต้นไม้ X-Y โดยต้นไม้นี้เป็นโครงสร้างที่แสดงขอบเขตส่วนประกอบของภาพเอกสาร ย่อลงไปตามลำดับชั้นของส่วนประกอบที่ได้จากการตัดซ้ำแนว X-Y โหนดราก(root node)ของต้นไม้ X-Y เป็นขอบเขตทั้งหมดของหน้าเอกสาร และโหนดลูก(child node)เป็นขอบเขตส่วนประกอบย่อยจากโหนดแม่(parent node) Nagy และ Seth[3,4] ได้เสนอวิธีการใช้ต้นไม้ X-Y ในการแสดงรูปแบบหน้าเอกสาร ซึ่งการแสดงต้นไม้ X-Y มีคุณสมบัติดังนี้

1. หน้าจะถูกจัดให้เป็นขอบเขตสี่เหลี่ยม โดยจะไม่มีขอบเขตที่ไม่ใช่รายละเอียด ส่วนประกอบย่อยที่แบ่งจากส่วนประกอบใหญ่ จะเก็บเป็นโหนดลูกถัดลงมาจกโหนดที่เก็บส่วนประกอบใหญ่ไว้
2. มีขอบเขตของส่วนประกอบเป็นสี่เหลี่ยมเท่านั้น
3. ในแต่ละระดับ ส่วนประกอบย่อยจะอยู่ถัดลงมาจกส่วนประกอบใหญ่เท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ต้นไม้ X-Y สามารถมีได้มากกว่า 2 มิติได้ เพื่อจัดการกับเอกสารที่มีหลายหน้า หรือนำไปใช้ประยุกต์  
 อย่างอื่นได้

## 1.2 วิธีการล่างขึ้นบน

การวิเคราะห์ล่างขึ้นบนเป็นวิธีพื้นฐานวิธีหนึ่งของการประมวลผลภาพ โดยเริ่มจากการนำส่วนประกอบ  
 ระดับต่ำที่สุดมารวมกันเป็นกลุ่มส่วนประกอบชนิดเดียวกันที่ใหญ่ขึ้น โดยชนิดส่วนประกอบได้แก่ข้อความ และ  
 รูปภาพ ในกรณีของข้อความ ส่วนประกอบเล็กที่สุดคือตัวอักษร ซึ่งตัวอักษรจะถูกรวมกันเข้าเป็นคำ แล้วคำจะ  
 ถูกรวมเป็นบรรทัด และบรรทัดจะรวมเข้าเป็นย่อหน้า และ ย่อหน้าก็จะไปสู่อัลกอริทึมที่ใหญ่กว่า ถ้าการรวมนั้น  
 สามารถทำได้ ในวิธีการล่างขึ้นบน ปกติเป็นวิธีพื้นฐานที่ใช้ในการกำหนดส่วนประกอบที่ติดกันเป็นส่วน เช่น ข้อ  
 ความ บรรทัด รูปภาพ หรือ ภาพฮาฟโทน [5] ยกตัวอย่างของ วิธีของ D. Drivas and A. Amin[6] ใช้วิธีการสร้าง  
 ส่วนประกอบที่ติดกันเข้าด้วยกัน โดยใช้วิธีการสแกนทีละบรรทัด เมื่อเจอจุดดำที่อยู่เดี่ยวหรือรวมกันเป็นกลุ่ม ก็  
 จะทำการติกรอบสี่เหลี่ยมรอบกลุ่มนั้น โดยขอบเขตจะขยายต่อไป เมื่อบรรทัดถัดลงมา มีจุดสีดำอยู่ติดกับ  
 ขอบเขตสี่เหลี่ยมเดิมที่มีอยู่ จะทำการรวมติกรอบสี่เหลี่ยมครอบคลุมส่วนที่ติดกันทั้งหมด ขอบเขตจะหยุดการ  
 ขยายก็ต่อเมื่อบรรทัดถัดไปไม่มีจุดดำอื่นอยู่ใกล้ขีด และจะมีการสร้างขอบเขตใหม่เมื่อมีจุดดำที่เริ่มต้น โดย  
 บรรทัดก่อนหน้าไม่มีขอบเขตสี่เหลี่ยมของส่วนประกอบไปอยู่ และนำส่วนประกอบที่ติดกันมารวมกลุ่มเข้าด้วย  
 กันโดยใช้ขนาดของส่วนประกอบนั้นๆ ที่จัดอยู่ในสามระดับ คือ ขนาดเล็ก กลาง ใหญ่ เพื่อนำส่วนประกอบที่มี  
 ขนาดใกล้เคียงกันมารวมกันเพื่อจะนำไปดำเนินการในการแบ่งชนิดส่วนประกอบโดยใช้อัตราส่วนขาวต่อดำใน  
 การใช้แบ่งชนิดว่าเป็น ข้อความ หรือ กราฟิก ต่อไป

## 2 การแบ่งส่วนประกอบเอกสาร

การแบ่งส่วนประกอบเอกสาร คือ การแบ่งขอบเขตรายละเอียดของภาพเอกสารให้เป็นส่วนๆ โดยแต่ละ  
 ส่วนจะมีรายละเอียดภายในที่เป็นชนิดเดียวกัน เช่น ขอบเขตที่เป็นข้อความ หรือขอบเขตที่เป็นรูปภาพ โดยวิธี  
 การนำเสนอจะเป็นวิธีการสกัดส่วนที่เป็นข้อความเพื่อนำไปใช้ในระบบ OCR ซึ่งแบ่งขั้นตอนการทำงานได้ 3  
 ขั้นตอนดังนี้

### 2.1 การแบ่งบล็อกส่วนประกอบใหญ่(คอลัมน์)โดยใช้หน้าต่างขนาด 32\*32 พิกเซล (top-down)

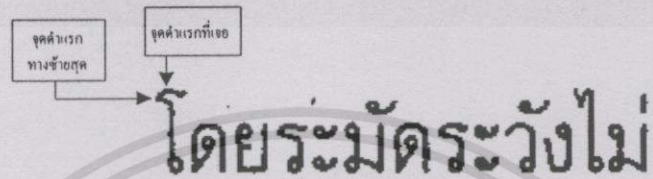
เอกสารโดยทั่วไปอาจจะมีทั้งข้อความและรูปภาพอยู่ภายในหน้าเอกสารเดียวกัน เพื่อที่จะแบ่งแยกส่วน  
 ประกอบที่แตกต่างออกจากกัน จึงได้เสนอการวางขอบเขตส่วนประกอบด้วยหน้าต่างขนาด 32\*32 พิกเซล เหตุ  
 ผลเนื่องจากข้อความในเอกสารปกติจะมีระยะห่างระหว่างตัวอักษรและระหว่างบรรทัด ดังนั้นในการที่จะรวมตัว  
 อักษรเข้าด้วยกันเป็นกลุ่ม หน้าต่างที่ใช้ในการตรวจสอบรายละเอียดของเอกสารต้องมีขนาดใหญ่พอที่สามารถ  
 ครอบคลุมและเชื่อมต่อตัวอักษรหรือรายละเอียดที่เป็นชนิดเดียวกันที่ใกล้กันมารวมเป็นกลุ่มเดียวกันได้ และ  
 ขนาดหน้าต่างจะต้องเล็กกว่าระยะห่างระหว่างรายละเอียดที่ไม่ใช่ประเภทเดียวกัน การใช้หน้าต่าง 32\*32 พิก  
 เซลในการเดินรอบขอบเขต เริ่มจากต้องหาจุดดำแรกของรายละเอียดของเอกสารและจุดดำทางซ้ายสุดภายใน  
 ขอบเขต 32 พิกเซล ดังแสดงในรูปที่ 1 แล้วกำหนดหน้าต่างตั้งต้นที่มุมบนซ้ายของรายละเอียด ดังแสดงในรูปที่ 2  
 โดยจุดดำในหน้าต่างต้องมากกว่า 10 จุด เพราะถ้าเท่ากับหรือต่ำกว่า 10 จุด จะจัดให้เป็นสัญญาณรบกวน ต่อ  
 จากนั้นทำการวางแนวหน้าต่างด้วยวิธีการเดินวนรอบ (contour following) ตามแนวขอบเขตส่วนประกอบจนลึ้น  
 สุดที่ตำแหน่งหน้าต่างเริ่มต้น ดังแสดงในรูปที่ 3 และเนื่องจากการเดินรอบขอบนี้จะมีแนวการเดินวนเป็นขอบ  
 ของส่วนประกอบ ดังนั้นจึงมีรูปร่างที่หลากหลาย ขึ้นกับส่วนประกอบที่ทำการเดินวน ดังตัวอย่างที่แสดงในรูปที่

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ภายใต้การดำเนินงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่เชิงพาณิชย์บนเว็บไซต์  
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4 และการแบ่งส่วนประกอบนี้ความเอียงของเอกสารไม่มีผลต่อการแบ่งส่วนประกอบดังแสดงในรูปที่ 5 และในรูปที่ 6 จะแสดงภาพเอกสารที่ผ่านการแบ่งแยกส่วนประกอบแล้ว

ในขั้นตอนหาส่วนประกอบใหญ่นี้ เมื่อเปรียบเทียบกับวิธีอื่นๆ ที่ได้ทำมาจะเห็นว่ามีความเร็วกว่ามาก เพราะ

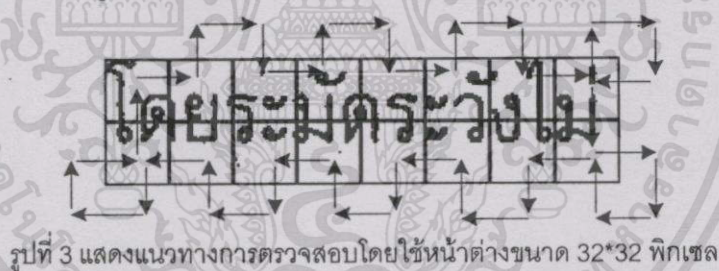
1. หน้าต่างจะวนรอบเฉพาะขอบเท่านั้น มิได้เข้าไปในส่วนเนื้อของส่วนประกอบใหญ่
2. หน้าต่างที่อยู่ตรงขอบส่วนประกอบใหญ่ ไม่จำเป็นต้องนับจุดดำทุกจุดในหน้าตานั้น ถ้านับได้มากกว่า 10 จุดขึ้นไป แสดงว่าหน้าตานั้นเป็นหน้าตารายละเอียด



รูปที่ 1 จุดดำแรกขอบบน และจุดดำซ้ายสุดที่เจอ ที่นำมาเป็นขอบตำแหน่งหน้าต่าง 32\*32 พิกเซล



รูปที่ 2 หน้าต่างขนาด 32\*32 พิกเซลที่ตำแหน่งเริ่มต้น



รูปที่ 3 แสดงแนวทางการตรวจสอบโดยใช้หน้าต่างขนาด 32\*32 พิกเซล

# High-Order Pattern Discovery from Discrete-Valued Data

Andrew K.C. Wong, Member, IEEE, and Yang Wang

**Abstract**—To uncover qualitative and quantitative patterns in a data set is a challenging task for research in the area of machine learning and data analysis. Due to the complexity of real-world data, high-order (polynomial) patterns or event associations, in addition to first-order class-dependent relationships, have to be acquired. Once the patterns of different orders are found, they should be represented in a form appropriate for further analysis and interpretation. In this paper, we propose a novel method to discover qualitative and quantitative patterns (or event associations) inherent in a data set. It uses the adjusted residual analysis in statistics to test the significance of the occurrence of a pattern candidate against its expectation. To avoid exhaustive search of all possible combinations of primary events, techniques of eliminating the impossible pattern candidates are developed. The detected patterns of different orders are then represented in an attributed hypergraph which is used for pattern interpretation and analysis. Test results on artificial and real-world data are discussed toward the end of the paper.

**Index Terms**—Adjusted residual, attributed hypergraph, data analysis, database mining, machine learning, pattern discovery, pattern representation.

## 1. INTRODUCTION

ONE of the basic tasks of machine learning and data analysis is to automatically uncover the qualitative and quantitative patterns in a data set. When a large data base is given, to discover the inherent patterns and regularities becomes a challenging task, especially when no domain knowledge is available or when the domain knowledge is too weak. Because of the size of the database, it is almost impossible for a decision maker to manually abstract from it the useful patterns. Hence, it is desirable to have automatic pattern discovery tools to support various types of decision making tasks.

Hitherto, machine learning research has focused more on supervised learning or learning by being told. With such an approach, the learner is given explicit information on the target mapping in the form of example pairs. Algorithms are developed to find the relation between the examples and the predefined classes. Hence, these methods are largely classification-oriented. In self-supervised or unsupervised learning (which is more general), the learner receives no explicit information on the pairs. Further, in a fully unsupervised learning setting, improvements in the behavior of the learner do not involve the evaluation of the actual outputs. Instead, the desired behavior is achieved via a direct, algorithmic process, which usually can be described in terms of an objective function [32]. Obviously, then, for pattern discovery in data, unsupervised learning is more useful than supervised strategies when there is no explicit classification information.

Consider that we have a set  $D$  of  $M$  observations or samples obtained from a database. Each observation is described by  $N$  attributes. Thus, for any distinct attribute, there is a domain of possible values which can be numeric, Boolean, and/or symbolic. Here, we assume that all the possible values are discrete. In compliance with the general nature of machine learning, though our method is applicable to mixed-mode data when discretizing techniques are used on numerical data [37].

The purpose of pattern discovery, be it referred to as conceptual clustering [26] or rule induction [31], is to find the relations among the attributes and/or among their values. We seek to find those events whose occurrences are significantly different from those based on a "random" model. Thus, the co-occurrence of events detected may reflect the nature of the data set and provide useful information for future inferences and reasoning.

Several open problems in machine learning and automatic knowledge acquisition are considered in this paper. The first is in regard to learning in the presence of uncertainty. Early machine learning systems were developed in the noise-free or deterministic environment [23]. The training data set is assumed to be entirely correct. However, for a real-world database, this assumption hardly stands. The instances in the database may include attributes based on measurement or subjective judgments, both of which may give rise to errors in the value of attributes [21]. Nonsystematic errors are usually referred to as noise. In a pattern discovery process, noise appears as the inconsistency of some samples against certain patterns. In a classification problem, some of the samples may be misclassified. The use of misclassified instances in learning can be considered as introducing noise which would affect classification accuracy. Missing values of attributes are also considered as noise. Now, many researchers are more interested in

The authors are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. E-mail: fekwong@uwaterloo.ca; yang@uwaterloo.ca.

Manuscript received 9 May 1995.  
For information on obtaining reprints of this article, please contact the IEEE Publishing Dept. at IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854.

0893-4005/97/010076-11\$05.00/0

รูปที่ 4 ตัวอย่างขอบเขตที่ได้จากการเดินวนรอบส่วนประกอบ

**ความปลอดภัยในโรงงานโซนี่ เช่นคอนดัคเตอร์**  
**การปลูกฝังให้พนักงานมีสำนึก และกระตุ้นให้มีส่วนร่วม**  
**ในการบริหารความปลอดภัยของโรงงาน ช่วยลดอุบัติเหตุจากการทำงานให้เหลือศูนย์**  
**Sony Semiconductor Plays It Safe**  
**Employees' safety consciousness and participation in safety management**  
**help reduce in-factory accidents to zero.**

*Mr. Sombon Sombon*  
*Dr. Anoop Pathumthani*

เมื่อพูดถึงเรื่องความปลอดภัย... ผู้ช่วยผู้จัดการทั่วไป บริษัท โซนี่ เซมิคอนดัคเตอร์ (ประเทศไทย) จำกัด คุณสมบุญ อีร์รอรอนันต์ กล่าวไว้ การจัดการด้านความปลอดภัยในระยะยาว... บริษัทฯ ได้ริเริ่มโครงการความปลอดภัย... บริษัทฯ ได้กำหนดมาตรการความปลอดภัยอย่างรอบคอบ เพื่อให้คนเราไป... มีข้อดีคือช่วยปรับประสิทธิภาพ การวางแผน... หรือบุคคลอื่นควรระวังความเสียหายทั้งหมด... โปรดคำนึงถึงทั้งในหน้าที่และความรับผิดชอบ... สิทธิที่หมด จุดศูนย์กลางเรื่องนี้เป็นหัวใจของ... ความสนใจในเรื่องความปลอดภัยของโรงงาน... บริษัทฯ ในโรงงานอุตสาหกรรมบางแห่งที่... ผู้ช่วยประธาน พนักงานบริษัท โซนี่ เซมิคอนดัคเตอร์ (ประเทศไทย) จำกัด กว่า 1,100 คน... อธิษฐานกันว่า บริษัทฯ จะจัดการจัดการ... ความปลอดภัยได้เป็นอย่างดีไว้ให้กับบริษัท... ความสำเร็จของโครงการความปลอดภัยและอุปกรณ์ส่วนบุคคล รวมถึงการ... การพัฒนาบุคลากรให้ยอมรับ... ทั้งให้ศึกษาเรื่องความปลอดภัยที่เป็น... ปลอดภัย

“If we can preclude accidents in our plant, our employees will have good morale and can produce quality products.” Mr. Sombon explains. “On top of that, we do not accrue expenses caused by accidents. At the same time, employees have confidence in the company safety system and equipment and therefore they will work safely.” He explains, “Safety management at Sony Semiconductor is the same as at other companies abiding to well-known principles. Our safety activities conform to the rules and regulations set by the Government. But, we place great emphasis on protection procedures and building employee's safety consciousness.”

**27 INDUSTRY MAG**  
MAY 1997

รูปที่ 5 ตัวอย่างขอบเขตที่ได้จากการเดินรอบคุณส่วนประกอบจากหน้าเอกสารที่เรียง 5 ฮังคา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

oriented programming languages.

MTS objects can be written using virtually any object-oriented programming language and typically reside in the middle tier of a three-tier client/server system. When designed correctly, these middle-tier components let you centralize the application logic that accesses your databases. More important, by centralizing your data access in components that can take advantage of connection pooling and higher-speed connections to the database servers (particularly when the client applications connect over the Internet), the total time that database locks are held by these components is minimized. The amount of contention for data in the data tier is also minimized.

NETWORK ROUND TRIPS

Another factor that can limit the scalability of your application is the total number of network round trips required to do some work on your server. In today's object-oriented client/server applications, you will typically perform three atomic operations, each requiring at least one network round trip, to perform a given operation. These steps are creating an instance of an object on the server computer, invoking one or more methods on the object, and destroying the object on the server computer.

Since these are middle-tier objects, they are usually acquiring expensive resources to do their work. Typically these expensive resources are database connections and database locks.

For your application to scale, it is essential that you conserve these resources. Therefore, object-oriented programmers have traditionally created the objects just when they needed to use them and destroyed them immediately thereafter. The implication of this tactic is that we must pay the price of at least three network round trips to invoke a method on a server-side object.

By requiring three round trips to perform a single logical operation, we introduce an

MTS reduces the number of round trips required to perform a single operation to almost one, through its use of a context wrapper object.



other potential bottleneck and thus limit the scalability of the application. Additional remote procedure call (RPC) runtime resources are required to handle the overhead of processing three distinct messages where only one does meaningful work. In theory, the operations required to create and destroy the object can be optimized out of existence.

MTS reduces the number of round trips required to perform a single operation to almost one: it optimizes the creation and destruction operations out of existence through its use of a context wrapper object.

The context wrapper object, along with its associated context object (discussed below) forms an object runtime environment (Figure 2), which is conceptually the object that the client application sees; that there are three separate cooperating objects is an implementation detail.

For MTS to eliminate the need for a client application to create and destroy MTS objects, the client application must follow a complete

times that you invoke methods on the object. The MTS network round-trip optimization is called just-in-time activation, and the key is the MTS context wrapper object, which implements the methods for the interfaces implemented by the actual object. It is created dynamically at runtime by MTS and is inserted transparently between the client application and the real MTS object. The context wrapper creates an instance of the MTS object at the moment it receives an incoming call from a client application—hence the term just-in-time activation.

For MTS to create a context wrapper object for any MTS object dynamically, it must know the intimate details of the object's interface signature. That requires that an MTS object meet a certain set of requirements. We'll discuss those requirements a little later when we examine what constitutes an MTS object.

TRANSACTIONS

Building a distributed application involves more than just scalability. As the complexity of the computing system increases, so does the probability of failure. Well-designed distributed applications are not only scalable but can also recover automatically from failure.

Most distributed systems use the concept of a transaction to implement automatic recovery. Transactions allow a complex set of actions to appear as a single unit of work.

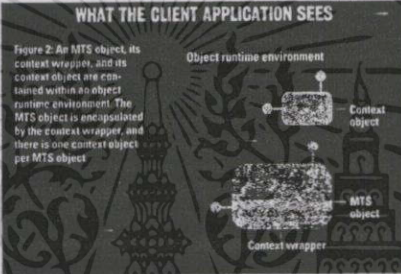
Most modern transaction systems will pass the so-called "ACID" test, which consists of the following four key properties:

- Atomic: All of the work gets done, or none of the work gets done.
- Consistent: The operation being performed is legal.
- Isolated: The behavior of a transaction is independent of the actions of any other transaction.
- Durable: Once a transaction has been completed, the work has been committed to durable storage.

A transaction forms a binding contract between a client application and the code running on the server. Once the server code has signaled a client that a transaction has succeeded, it has effectively guaranteed the client that the work has been committed to durable storage. Likewise, if the server code signals the client that the transaction has been aborted, the client code is promised that as far as the data is concerned, everything is as though the operation were never

WHAT THE CLIENT APPLICATION SEES

Figure 2: An MTS object, its context object, and its context wrapper are contained within an object runtime environment. The MTS object is encapsulated by the context wrapper, and there is one context object per MTS object.



By requiring three round trips to perform a single logical operation, we introduce an

รูปที่ 6 ตัวอย่างขอบเขตที่ส่วนประกอบมีคอลัมน์มากกว่า 1 คอลัมน์

2.2 การหาส่วนประกอบย่อยภายในส่วนประกอบใหญ่ โดยการเดินขอบตรวจหาจุดดำในส่วนประกอบใหญ่ หาส่วนประกอบย่อยในส่วนประกอบใหญ่ที่ได้จาก 2.1 โดยใช้เทคนิคการเดินวนรอบกรอบหน้าต่างขนาด 1 พิกเซล ซึ่งจะเดินรอบภาพจุดดำที่ติดกันออกมา ซึ่งอาจจะเป็นตัวอักษร เส้น หรือ รูปภาพ การพิจารณาค่าแห่งสูงสุดและต่ำสุดในพิกัด X, Y ของแนวขอบ(contour ของส่วนประกอบย่อย) จะสามารถหากรอบสี่เหลี่ยมรอบส่วนประกอบย่อยได้ (bounding box)

การเดินวนรอบส่วนประกอบย่อยที่อยู่ในคอลัมน์ จะทำทั้งคอลัมน์ โดยเริ่มจากขอบซ้ายไปขอบขวาของคอลัมน์ และในการเดินวนรอบ มีการจัดกลุ่มความสูง และความกว้างของกรอบส่วนประกอบย่อย เพื่อหาแนวบรรทัด ผลของขั้นตอนนี้แสดงในรูปที่ 7 (โดยถ้าเอกสารมีความเอียง จะสามารถปรับภาพเอียงในขั้นตอนนี้ได้ โดยใช้ขอบเขตความสูงต่ำ ที่เก็บอยู่ในช่วงมาวิเคราะห์แบบสมการถดถอย แต่ในบทความนี้เราไม่มีขั้นตอนนี้ เพราะเอกสารโดยปัจจุบันนำเข้าด้วยเครื่องสแกนเนอร์ที่มีตัวป้อนกระดาษที่ดี จึงไม่มีความจำเป็นในขั้นตอนนี้แก้ไขความเอียง)



ขั้นตอนนี้มีความรวดเร็วเมื่อเทียบกับอัลกอริทึมอื่น (X-Y cut) เพราะการทำการฉายภาพเงาจะทำเฉพาะจุดขาวที่อยู่ระหว่างขอบเขตตัวอักษรเท่านั้น และไม่ได้เป็นการทำทุกบรรทัด ทำเฉพาะบรรทัดแรกของแนวคอลัมน์ย่อยต่างๆในส่วนประกอบใหญ่

### 3 ผลการทดลอง

เอกสารที่ใช้ในการทดลองมีขนาดประมาณกระดาษ A4 และความละเอียดในการสแกน 300จุดต่อนิ้ว ประมวลผลโดยเครื่องไมโครคอมพิวเตอร์ Pentium II 350 MHz หน่วยความจำ 64 MB ระบบปฏิบัติการ MS Windows 98 SE เครื่องมือที่ใช้พัฒนา Visual C++ 5.0

3.1 จากภาพต้นฉบับในรูปที่ 4 และ 5 สามารถแบ่งขอบเขตของส่วนประกอบออกจากกันได้อย่างถูกต้อง ถ้าวัตถุในเอกสารแต่ละส่วนแยกจากกัน ในระยะที่ห่างกันพอสมควร แต่บางกรณีอาจมีคอลัมน์มากกว่าหนึ่งคอลัมน์อยู่ในขอบเขตเดียวกันดังแสดงในรูปที่ 6 ในขั้นตอนนี้ใช้เวลาประมาณ 1 วินาที ขึ้นอยู่กับขนาดและรายละเอียดภายในภาพ

3.2 แบ่งส่วนประกอบย่อยโดยการเดินขอบด้วยรอบจุดดำในแต่ละส่วนย่อยที่ได้จากข้อ 3.1 ขนาดส่วนประกอบที่เล็กมากจะถูกกำหนดให้เป็นขยะและถูกกำจัดไป ซึ่งข้อมูลพวกนี้ขอบเขตบล็อกตัวอักษรที่มีขนาดใกล้เคียงกันจะถูกนำไปใช้ในระบบงาน OCR ต่อไป โดยจะมีขอบเขตใหญ่ในขั้นที่ 2.1 และ 2.3 เป็นตัวกำหนดขอบเขตข้อมูลที่จะนำไปใช้กับระบบ OCR ต่อไป ในส่วนนี้ใช้เวลาในการทำงานประมาณ 1 วินาที

### 4 สรุปผลการทดลอง

วิธีการนี้ได้ทดสอบกับวารสารทางวิทยาศาสตร์และนิตยสาร ที่ประกอบด้วย ภาพ ตาราง และตัวอักษร ผลลัพธ์ที่ได้ถูกต้อง และสามารถทำกับเอกสารที่มีคอลัมน์ไม่เป็นสี่เหลี่ยมได้ วิธีการนี้เร็วเพราะไม่จำเป็นต้องเข้าถึงทุกจุดของภาพทั้งส่วนที่ทาสวนประกอบใหญ่และส่วนที่ทาสวนย่อยในส่วนประกอบใหญ่ และในการทาสวนประกอบย่อยก็จำเป็นต้องทำในการทำงาน OCR อยู่แล้ว ซึ่งวิธีการนี้ใช้เวลาการทำงานรวมประมาณ 2 วินาที ขึ้นอยู่กับความละเอียดของเอกสารที่สแกนเข้าและความซับซ้อนของเอกสาร หากแต่มีข้อจำกัดบางประการคือเอกสารต้องไม่เอียง รูปที่ 9 แสดงรูปที่แบ่งส่วนประกอบขอบเขตเรียบร้อยแล้ว

**oriented programming languages.**  
 MTS objects can be written using virtually any object-oriented programming language and typically reside in the middle tier of a three-tier client/server system. When designed correctly, these middle-tier components let you centralize the application logic that accesses your databases. More important, by centralizing your data access in components that can take advantage of connections pooling and higher-speed connections to the database servers (particularly when the client applications connect over the Internet), the total time that database locks are held by these components is minimized. The amount of contention for data in the data tier is also minimized.

**MTS reduces the number of round trips required to perform a single operation to almost one, through its use of a context wrapper object.**

other potential bottleneck and thus limit the scalability of the application. Additional remote procedure call (RPC) runtime resources are required to handle the overhead of processing three distinct messages where only one does meaningful work. In theory, the operations required to create and destroy the object can be optimized out of existence.

MTS reduces the number of round trips required to perform a single operation to almost one. It optimizes the creation and destruction operations out of existence through its use of a context wrapper object. The context wrapper object, along with its associated context object (discussed below) forms an object runtime environment (Figure 2), which is conceptually the object that the client application sees; that there are three separate cooperating objects is an implementation detail.

For MTS to eliminate the need for a client application to create and destroy MTS objects, the client application must follow a complete set of rules, requiring a reference to the MTS object as early as possible and releasing it as late as possible. This lets MTS create an instance of the MTS object whenever an incoming call comes in. When the object signals to MTS that it has finished doing what it was doing, MTS destroys it but keeps the context wrapper alive to service the next incoming request (Figure 2). This effectively allows MTS to amortize an object's creation and destruction overhead across the total number of times that you invoke methods on the object.

The MTS network round-trip optimization is called *just-in-time activation*, and the key is the MTS context wrapper object, which implements the methods for the interfaces implemented by the actual object. It is created dynamically at runtime by MTS and is inserted transparently between the client application and the real MTS object. The context wrapper creates an instance of the MTS object at the moment it receives an incoming call from a client application—hence the term *just-in-time activation*.

For MTS to create a context wrapper object for any MTS object dynamically, it must know the intimate details of the object's interface signature. That requires that an MTS object meet a certain set of requirements. We'll discuss those requirements a little later when we examine what constitutes an MTS object.

**TRANSACTIONS**  
 Building a distributed application involves more than just scalability. As the complexity of the computing system increases, so does the probability of failure. Well-designed distributed applications are not only scalable but can also recover automatically from failure.

Most distributed systems use the concept of a *transaction* to implement automatic recovery. Transactions allow a complex set of actions to appear as a single unit of work.

Most modern transaction systems will pass the so-called "ACID" test, which consists of the following four key properties:

- **Atomic:** All of the work gets done or none of the work gets done.
- **Consistent:** The operation being performed is legal.
- **Isolated:** The behavior of a transaction is independent of the actions of any other transaction.
- **Durable:** Once a transaction has been completed, the work has been committed to durable storage.

A transaction forms a binding contract between a client application and the code running on the server. Once the server code has signaled a client that a transaction has succeeded, it has effectively guaranteed the client that the work has been committed to durable storage. Likewise, if the server code signals the client that the transaction has been aborted, the client code is promised that as far as the data is concerned, everything is as though the operation were never

**NETWORK ROUND TRIPS**  
 Another factor that can limit the scalability of your application is the total number of network round trips required to do some work.

In today's object-oriented client/server applications, you will typically perform three atomic operations, each requiring at least one network round trip, to perform a given operation. These steps are creating an instance of an object on the server computer, invoking one or more methods on the object, and destroying the object.

Since these are middle-tier objects, they are usually acquiring expensive resources such as database connections and database locks.

For your application to scale, it is essential that you conserve these resources. Therefore, client application programmers have traditionally created the objects just when they needed to use them and destroyed them immediately thereafter. The implication of this tactic is that we must pay the price of at least three network round trips to invoke a method on a server-side object.

By requiring three round trips to perform a single logical operation, we introduce an-

**WHAT THE CLIENT APPLICATION SEES**

Figure 2: An MTS object, its context wrapper, and its context object are contained within an object runtime environment. The MTS object is maintained by the context wrapper, and there is one context object per MTS object.

212 PC MAGAZINE APRIL 1998

รูปที่ 9 แสดงขอบเขตหลังจากแบ่งส่วนประกอบเอกสารเรียบร้อยแล้ว

อ้างอิง

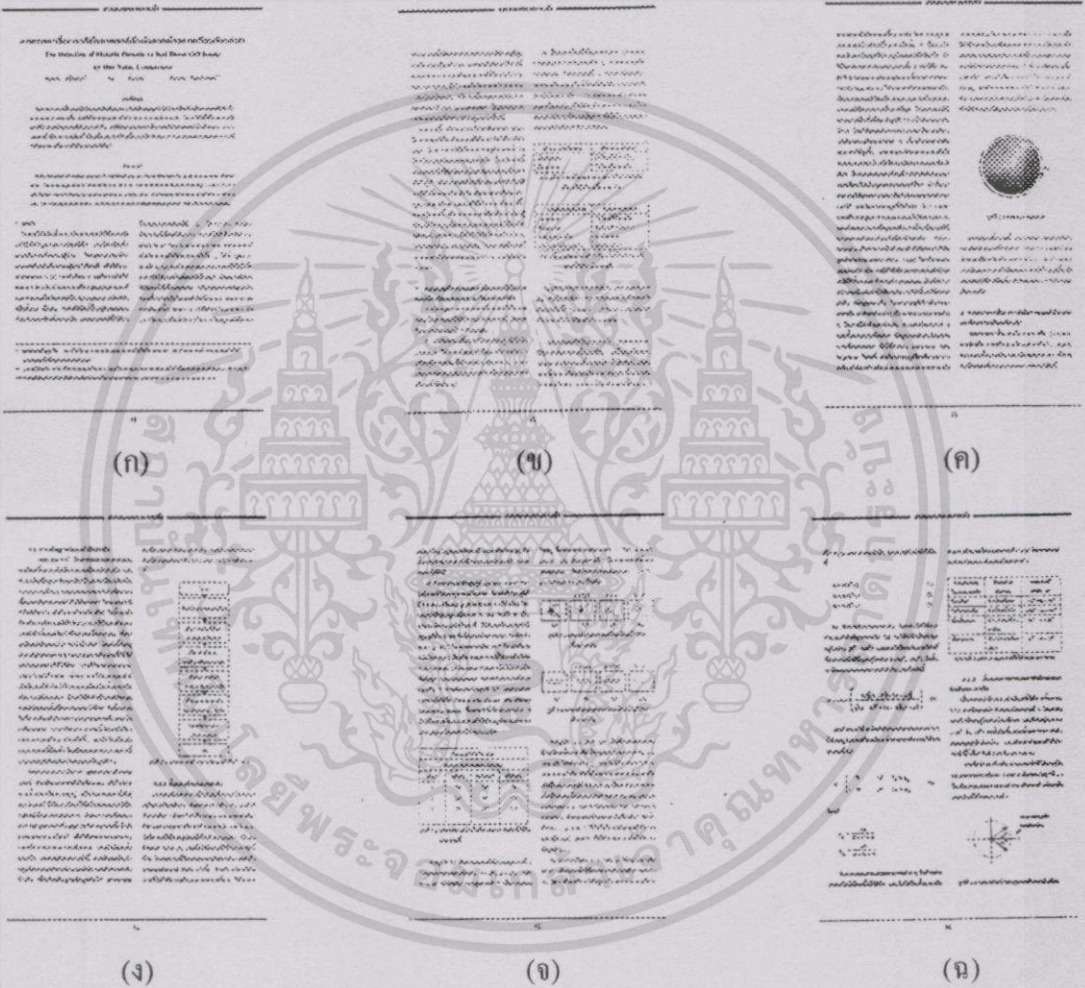
1. A. Jain, B. Yu, "Document Representation and Its Application to Page Decomposition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, March 1998.
2. K. Y. Wong, R. G. Casey and F. M. Wahl, "Document Analysis system.", IBM Journal of Research and Development 26, pp. 647-655, 1982
3. G. Nagy and S. Seth., "Hierarchical Representation of Optical Scanned Documents.", In Processing of the 7<sup>th</sup> International Conference on Pattern Recognition Montreal, pp. 347-349, 1984.
4. G. Nagy, S. C. Seth and S. D. Stoddard, "Document Analysis with an Expert System", Pattern Recognition in Practice II (E. S. Gelsema and C. N. Kanal, Eds) pp. 147-159, 1986.
5. Sargur N. Srihari, "Document Image Understanding", Proceedings Fall Joint Computer Conference, 1986.
6. D. Drivas and A. Amin, "Page segmentation and classification utilising bottom-up approach", Proceedings of the 3<sup>rd</sup> International Conference on Document Analysis and Recognition, Montreal, August 1995, pp. 610-614.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ภาคผนวก ข

## ภาพเอกสารที่ใช้ในการทดสอบ

### ภาพเอกสารทั้งหมดถูกสแกนในลักษณะภาพขาวดำ ด้วยความละเอียด 300 จุดต่อนิ้ว



รูปที่ ข.1 ภาพเอกสารจากวารสารสนเทศ ลาดกระบัง จำนวน 6 ภาพ

- (ก) 3342 x 2104 พิกเซล
- (ข) 3342 x 2128 พิกเซล
- (ค) 3342 x 2096 พิกเซล
- (ง) 3342 x 2128 พิกเซล
- (จ) 3342 x 2104 พิกเซล
- (ฉ) 3342 x 2120 พิกเซล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)



(ค)



(ง)

รูปที่ ข.2 ภาพจากวารสาร LIFE ENTERTAINMENT จำนวน 4 ภาพ

- (ก) 3065 x 2223 พิกเซล
- (ข) 3090 x 2236 พิกเซล
- (ค) 3090 x 2148 พิกเซล
- (ง) 3090 x 2211 พิกเซล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิได้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



### A Historical Perspective of CSTA

Richard A. DeLoach, Lowell Holmstrom

**H**ighly supported telecommunication applications (CSTA) is one of the more important standards for computer-telephony interfaces (CTI) and provides a common set of protocols and procedures used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

**CSTA: LOOKING ON THE BRIGHT SIDE**

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

### Design Considerations for Computer-Telephony Application Programming Interfaces and Related Components

July 1988, 20th Anniversary

**A** number of industry people have predicted that the telephony industry will follow the same path as the computer industry -- to create more standard connectivity functions.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

(ก)

(ข)

### Computer Telephony Architectures: MVIP, H-MVIP, and SCBUS

Mark Perry, IBM

**T**he number of users of computer-telephony interfaces (CTI) is growing rapidly. This growth is due to the increasing number of users who are using CTI to access information and services.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

### Overview of Implementing ATM-Based Enterprise Local Area Network for Desktop Multimedia Computing

Mark Perry, IBM

**A** number of industry people have predicted that the telephony industry will follow the same path as the computer industry -- to create more standard connectivity functions.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

Computer-telephony interfaces (CTI) is a term used to describe the interface between a computer and a telephone system. The interface is used to provide a link between computers and telecommunication systems. The interface is used to provide a link between computers and telecommunication systems.

(ค)

(ง)

รูปที่ ข.4 ภาพจากวารสาร IEEE Communication Magazine จำนวน 4 ภาพ

- (ก) 3153 x 2399 พิกเซล
- (ข) 3355 x 2437 พิกเซล
- (ค) 3355 x 2437 พิกเซล
- (ง) 3355 x 2437 พิกเซล

**MSI reduces the number of board tries required to perform a single operation in absentees, through its use of a patented memory aligner.**

MSI's new memory aligner is a small, portable device that allows users to quickly and accurately align memory modules in their systems. This is particularly useful for users who are upgrading their systems or who are troubleshooting memory-related problems. The aligner is designed to be used with any standard memory module, and it is easy to use and requires no special tools or equipment.

**What the Data Protection Bill**

The Data Protection Bill is a proposed piece of legislation that would regulate the collection, use, and disclosure of personal data. The bill is designed to protect the privacy of individuals and to ensure that their data is handled in a fair and transparent manner. The bill would apply to all organizations that collect or process personal data, regardless of the size of the organization or the nature of the data.

**From the Editor-in-Chief**

**Regulation: Not If but How**

When the House Judiciary Committee passed the bill on Tuesday, the Senate Judiciary Committee is likely to have a similar bill in the next few days. And in that time, the government will be working to get the bill passed. The bill is designed to regulate the collection, use, and disclosure of personal data, and it is a key piece of legislation that would protect the privacy of individuals.

**After Hours**

**O Rome, Rome**

A new study has found that the average person spends about 10 hours per week on the Internet. This is a significant increase from the 5 hours per week reported in a previous study. The study is designed to help researchers understand how people use the Internet and to identify areas where they need more support.



**Back 'Em Up**

IBM's new hard drive is the largest ever, with a capacity of 100 gigabytes. This is a significant increase from the 40 gigabytes reported in a previous study. The hard drive is designed to be used in servers and other high-performance systems, and it is a key component of IBM's new storage solutions.



**(ก) First Looks**

**Folder 15**

The new folder is designed to be used for organizing files and folders in a Windows 95 environment. It is a key component of the new operating system, and it is designed to be easy to use and to provide a clear and organized view of the user's files.



**(ข) First Looks**

**Win 4.0 Makes Drawing and Migrating Easier Than Ever**

The new version of Windows 4.0 is designed to make it easier for users to draw and migrate files. It includes a new drawing tool that allows users to create and edit drawings more easily, and a new migration tool that allows users to move files between different systems more easily.



**(ค) Bill Review**

**Divine Right and Divine Wrong**

The new bill is designed to regulate the collection, use, and disclosure of personal data. It is a key piece of legislation that would protect the privacy of individuals and ensure that their data is handled in a fair and transparent manner. The bill would apply to all organizations that collect or process personal data, regardless of the size of the organization or the nature of the data.



**(ง) First Looks**

The new folder is designed to be used for organizing files and folders in a Windows 95 environment. It is a key component of the new operating system, and it is designed to be easy to use and to provide a clear and organized view of the user's files.

**(จ) First Looks**

The new version of Windows 4.0 is designed to make it easier for users to draw and migrate files. It includes a new drawing tool that allows users to create and edit drawings more easily, and a new migration tool that allows users to move files between different systems more easily.

**(ฉ) Bill Review**

The new bill is designed to regulate the collection, use, and disclosure of personal data. It is a key piece of legislation that would protect the privacy of individuals and ensure that their data is handled in a fair and transparent manner. The bill would apply to all organizations that collect or process personal data, regardless of the size of the organization or the nature of the data.

รูปที่ ข.5 ภาพจากวารสาร PC MAGAZINE จำนวน 6 ภาพ

- (ก) 3000 x 2352 พิกเซล
- (ข) 3108 x 2202 พิกเซล
- (ค) 3194 x 2379 พิกเซล
- (ง) 3194 x 2379 พิกเซล
- (จ) 3194 x 2379 พิกเซล
- (ฉ) 3172 x 2201 พิกเซล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

...the first step in the process of ...

**2. A Brief Introduction to the Method**

...the second step in the process of ...

**3. Examples of Results**

...the third step in the process of ...

**4. CONCLUSION**

...the fourth step in the process of ...

**High-Order Pattern Discovery from Discrete-Valued Data**

Authors: K. Wang, H. Zhou, J. Li, and Y. Zhang

**1. Introduction**

...the first step in the process of ...

**2. Related Work**

...the second step in the process of ...

**3. Proposed Method**

...the third step in the process of ...

**4. Experimental Results**

...the fourth step in the process of ...

**5. Conclusion**

...the fifth step in the process of ...

**Abstract**

...the first step in the process of ...

**Keywords**

...the second step in the process of ...

**References**

...the third step in the process of ...

(ก)

(ข)

(ค)

รูปที่ ข.6 ภาพจากวารสาร IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING จำนวน 3 ภาพ

- (ก) 3228 x 2275 พิกเซล
- (ข) 2238 x 2362 พิกเซล
- (ค) 3238 x 2311 พิกเซล



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

นายพิเศษพงศ์ สุราพันธ์ เกิดเมื่อวันที่ 21 มีนาคม พ.ศ. 2516 ที่จังหวัดลำปาง สำเร็จการศึกษา วิทยาศาสตรบัณฑิต (สถิติ) จากมหาวิทยาลัยศรีนครินทรวิโรฒ วิทยาเขตภาคใต้ ปีการศึกษา 2536



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้