สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การแบ่งส่วนประกอบภาพเอกสารไบนารี เพื่อใช้ในระบบจดจำตัวอักษร

BINARY DOCUMENT SEGMENTATION FOR OCR SYSTEM

นายชวินทร์ จุลกรังคะ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตร์บัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2545

เลขหมู่.....
เลขทะเบียน 49988
วัน,เดือน,ปี 16 เม.ย. 2547
.b.....
.i.....
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแบ่งส่วนประกอบภาพเอกสารไบนารี
เพื่อใช้ในระบบจดจำตัวอักษร

BINARY DOCUMENT SEGMENTATION
FOR OCR SYSTEM

โดย

นายชวินทร์ จุลกรังคะ

อาจารย์ที่ปรึกษา

รศ.ดร. บุญธีร์ เครือตราชู

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตร์บัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2545

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2545

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การแบ่งส่วนประกอบภาพเอกสารไบนารีเพื่อใช้ในระบบจดจำตัวอักษร

BINARY DOCUMENT SEGMENTATION FOR OCR SYSTEM

ผู้จัดทำ

1. นาย ชวินทร์ จุลกรังคะ รหัสประจำตัว 42010075

_____ อาจารย์ที่ปรึกษา

(รศ.ดร. บุญธีร์ เครือตราชู)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแบ่งส่วนประกอบภาพเอกสารไบนารีเพื่อใช้ในระบบจดจำตัวอักษร

นายชวินทร์ จุลกรังคะ 42010075
รศ.ดร. บุญธีร์ เครือตราชู อาจารย์ที่ปรึกษา
ปีการศึกษา 2545

บทคัดย่อ

ในปัจจุบัน คนเราต้องการเก็บข้อมูลไว้ในคอมพิวเตอร์มากขึ้น เนื่องจากว่าคอมพิวเตอร์สามารถค้นหาข้อมูลได้อย่างสะดวกและง่ายกว่าหนังสือมาก การที่จะบันทึกข้อมูลมากๆด้วยแป้นพิมพ์นั้นยากที่จะทำได้ และในปัจจุบันการนำภาพสู่คอมพิวเตอร์กระทำได้ง่ายขึ้น แต่ก็มีข้อด้อยหลายประการ ข้อแรกคือการจัดเก็บเอกสารในลักษณะภาพมีขนาดใหญ่กว่าการจัดเก็บเอกสารที่เป็นข้อความมาก ข้อสองเอกสารที่เก็บในลักษณะภาพไม่สามารถนำมาแก้ไขเปลี่ยนแปลงได้อย่างเอกสารที่เป็นตัวอักษร และข้อสุดท้ายเอกสารที่จัดเก็บเป็นรูปภาพ ไม่สามารถนำไปใช้ในการค้นหาว่าเนื้อหาสาระข้างในเอกสารเป็นอย่างไร จากทั้งสามกรณีข้างต้น กระตุ้นให้เกิดความต้องการที่จะกำจัดข้อด้อยเหล่านั้น จึงมีการพัฒนา ระบบจดจำตัวอักษร(OCR : Optical Character Recognition) ในตอนแรกนั้นยังกำหนดขอบเขตที่จะให้ทำงานตามที่ต้องการด้วยมืออยู่ เพราะเอกสารบางอย่างมีหลายคอลัมน์ หรือ มีส่วนประกอบที่ซับซ้อนซึ่งระบบจดจำตัวอักษร ไม่สามารถแบ่งแยกได้ว่ารายละเอียดพวกนี้จะเป็นกลุ่มเดียวกันหรือไม่ ซึ่งทำให้เกิดความยุ่งยากและล่าช้าในงานที่ต้องจัดเก็บเอกสารจำนวนมาก จึงมีความต้องการกำหนดขอบเขตที่จะนำไปใช้ในระบบจดจำตัวอักษรแบบอัตโนมัติขึ้นมา เพื่อให้เกิดความสะดวกและทำงานได้รวดเร็วยิ่งขึ้น ทำให้เกิดงานวิจัยทางด้านการแบ่งขอบเขตเอกสารภาพขึ้นอย่างมาก

ปริญญานิพนธ์ฉบับนี้จัดทำขึ้นเพื่อพัฒนาวิธีการแบ่งส่วนประกอบภาพเอกสารไบนารีแบบอัตโนมัติที่รวดเร็วและถูกต้อง เพื่อให้ได้ขอบเขตส่วนประกอบที่จะนำไปใช้กับระบบจัดเก็บเอกสารและจดจำตัวอักษร เพื่อให้ระบบทำงานได้มีประสิทธิภาพยิ่งขึ้น โดยจะนำภาพเอกสารไบนารีมาแยกส่วนประกอบเอกสารซึ่งมีรายละเอียดหลายชนิด โดยเน้นความสำคัญของส่วนประกอบที่เป็นประเภทตัวอักษร และสามารถกำหนดขอบเขตของส่วนประกอบที่มีลักษณะไม่เป็นสี่เหลี่ยมได้ ซึ่งสามารถทำให้ได้เฉพาะส่วนที่ต้องการเท่านั้น และเพิ่มความสามารถของระบบจัดเก็บภาพเอกสารให้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

BINARY DOCUMENT SEGMENTATION FOR OCR SYSTEM

Chawin Julkrangka

Assoc. Prof. Dr. Boontee Chertrachoo Advisor

ABSTRACT

Nowadays, most of data is stored in a computer because of its comfortably. Computer can find us our desired data just in a second while we have to spend many hours looking for it in a record book. Therefore , typing a lot of data with keyboard is not a good idea because it will take a lot of time and it is not comfortable. As the computer technology is growing up move and more, storing images into computer is not so hard anymore. But it also has many disadvantages. First , image files take more disk space than text file do. Second, image files can't be edit while text files can. Last, we can't search anything from image files. Because of these disadvantages, the OCR (Optical Character Recognition) is developed to get rid of them. In the first place, the OCR's range is limited only for text because it can't distinguish types of data. It can't know if there are same of images in the document. This makes our works more complicated. So we try to extend the range of the OCR so that it can distinguish types of data automatically.

This thesis is made for developing how the OCR will distinguish an image part of binary document automatically which is fast and accurate so we can apply it with the OCR to make it to be more effective. This thesis will distinguish a lot of data type from a binary document and focus only on the text part and we also can define the range of a non-rectangular data type so we can focus on our desired data directly. That makes the OCR to be more effective

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้คงไม่อาจสำเร็จไปได้ด้วยดี หากไม่ได้รับความช่วยเหลือ และร่วมมือจากหลายๆ ฝ่ายด้วยกัน

ขอขอบคุณ อาจารย์ บุญธีร์ เครือตราชู ที่ช่วยบอกแนวคิด แนะนำ และให้ความช่วยเหลืออย่างดียิ่ง แม้ท่านจะไม่ค่อยมีเวลาว่างซักเท่าไหร่นัก ก็เอาใจใส่กระผมเป็นอย่างดี

ขอขอบคุณ บิดา มารดา ที่ช่วยให้กำลังใจ และช่วยให้ผมมีความพยายามที่จะทำให้ปริญญานิพนธ์นี้สำเร็จไปได้ภายในปีเดียว

ขอขอบคุณ ดร. อรัญญา วลัยรัชต์ ขอบคุณ ที่ช่วยสอน VC++ ได้อย่างเข้าใจมากกว่าอ่านจากหนังสือทั่วไปหลายเท่านัก

ขอขอบคุณอาจารย์นเรศ มาลัย ขอบคุณครับที่ทำให้เราได้มีห้องโปรเจ็คอยู่ และได้ใช้เน็ตได้ทั้งวัน ทั้งคืน และบางครั้งก็ได้ยินแนวคิดในการทำโปรเจ็คด้วย

ขอขอบคุณอาจารย์กฤตวัน ศิริบูรณ์ ที่ช่วยดูแล เป็นที่ปรึกษาโปรเจ็คแทนอาจารย์บุญธีร์ เวลาที่อาจารย์ไม่อยู่

และขอขอบใจเพื่อนๆ นักศึกษาทุกคนที่ช่วยเหลือให้คำแนะนำต่าง ๆ จนสำเร็จสมบูรณ์และยังให้กำลังใจต่อข้าพเจ้าอย่างใกล้ชิดตลอดมา

ชวินทร์ จุลกรังคะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I
บทคัดย่อภาษาอังกฤษ II
กิตติกรรมประกาศ III
สารบัญ IV
สารบัญตาราง VI
สารบัญภาพ VII
บทที่ 1 บทนำ 1
1.1 ความสำคัญและที่มา 1
1.2 ขอบเขตของโครงงาน 1
1.3 ความมุ่งหมายและจุดประสงค์ของโครงงาน 1
1.4 ประโยชน์ที่คาดว่าจะได้รับ 2
1.5 ทฤษฎีหรือแนวคิดที่ใช้ในการศึกษา 2
1.6 วิธีการดำเนินงาน 2
1.7 ข้อจำกัดของการศึกษา 3
1.8 รายละเอียดในแต่ละบท 3
บทที่ 2 ทฤษฎีและแนวคิดที่เกี่ยวข้อง 4
2.1 การแสดงภาพดิจิตอล 4
2.2 ขั้นตอนพื้นฐานในการประมวลผลภาพ 4
2.3 ความรู้พื้นฐานเกี่ยวกับข้อมูลภาพอิเล็กทรอนิกส์ 6
2.4 รหัสแนว (Chain Codes) 7
2.5 คุณลักษณะพื้นฐานของภาพเอกสาร 7
2.6 ประเภทของเอกสารและการประยุกต์ใช้ 8
2.6.1 เอกสารรูปแบบโครงสร้าง(structure articles) 10
2.6.2 เอกสารที่มีรูปแบบไม่แน่นอน 10
2.6.3 เอกสารกึ่งรูปแบบ (semi-structured layouts) 11
2.6.4 แผนที่และภาพวาด 12
2.6.5 หน้าเว็บ และ ภาพจากวีดีโอ 13
2.7 ขั้นตอนการแบ่งหน้าเอกสารและวิธีการแบ่งประเภทขอบเขต 14
2.7.1 การแบ่งส่วนประกอบเอกสาร 14
2.7.1.1 วิธีการล่างขึ้นบน 15
2.7.1.2 วิธีบนลงล่าง 16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

2.7.1.3 วิธีแบบผสมผสานและวิธีอื่นๆ 17
2.7.2 การแบ่งประเภทขอบเขตเอกสาร 18
2.8 การแบ่งประเภทขอบเขตเอกสาร 19
2.8.1 วิธีการรันเรงค์สมูทอัลกอริทึม (run length smoothing) 19
2.8.2 วิธีการตัดซ้ำแกน X-Y (recursive projection X-Y cut) 20
2.8.3 วิธีการเดินวนรอบขอบ (contouring) 20
2.8.4 วิธีการแบบไวท์ไทน์ (white tile) 21
2.9 เปรียบเทียบวิธีการทำงาน 21
2.10 วิเคราะห์วิธีการทำงาน 24
2.11 สรุป 27
บทที่ 3 แนวคิดและการนำมาประยุกต์ใช้ในการแบ่งส่วนประกอบภาพเอกสารไบนารี 30
3.1 แนวคิดในการแบ่งส่วนประกอบภาพเอกสารไบนารี 30
3.2 การแบ่งส่วนประกอบภาพเอกสารเป็นส่วนประกอบใหญ่ 32
3.3 การแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ 34
3.4 การแบ่งย่อหน้าของเอกสารในส่วนประกอบใหญ่ 35
บทที่ 4 อัลกอริทึมและขั้นตอนการทำงาน 36
4.1 วิธีการแบ่งส่วนประกอบใหญ่ 37
4.2 วิธีการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ 43
4.3 วิธีการแบ่งคอลัมน์ภายในขอบเขตส่วนประกอบใหญ่ 45
บทที่ 5 ผลการทดลอง 46
5.1 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่หนึ่ง 46
5.2 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่สอง 54
บทที่ 6 สรุปผลการทดลองและเสนอแนะแนวทางการดำเนินงานต่อไป 59
6.1 สรุปผลการทดลอง 59
6.2 ข้อเสนอแนะ 59
บรรณานุกรม 61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 แสดงคุณสมบัติที่สำคัญของวิธีการต่างๆ ที่ใช้ในการวิเคราะห์แผนผังของเอกสาร 22

2.2 แสดงเวลาที่ใช้ในการทำงานสำหรับวิธีการวิเคราะห์เอกสาร 23

2.3 แสดงผลงานวิจัยที่ผ่านมา 27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

ภาพที่ หน้า

2.1 แนวแกน X, Y ที่ใช้อ้างอิงพิกัด 4
2.2 ขั้นตอนพื้นฐานของกระบวนการประมวลผลภาพดิจิตอล 5
2.3 ทิศทางของรหัสแนว 4 ทิศทาง (ก) และ รหัสแนว 8 ทิศทาง (ข) 7
2.4 ตัวอย่างภาพเอกสารชนิดต่างๆ 9
2.5 ชุดจุดภาพก่อนทำการรันเรงค์สมูท 19
2.6 ชุดจุดภาพหลังจากทำการรันเรงค์สมูท ด้วยค่าที่กำหนดไว้คือ 3 20
2.7 การฉายเงาจุดดำในแนวตั้ง และช่วงที่จะแบ่ง 20
2.8 กราฟที่ใช้แสดงการจัดเรียงไวท์ไทน์ 21
3.1 ตัวอย่างแผนผังแบบต่างๆ 30
3.2 ลักษณะการเชื่อมต่อตัวอักษร 31
3.3 ขอบเขตภาพ (ก) ภาพแสดงเมื่อมองจากระยะไกล (ข) ภาพที่แบ่งขอบเขตด้วยสายตา 32
3.4 การวางตำแหน่งหน้าต่างเริ่มต้นและลักษณะการวนรอบส่วนประกอบใหญ่ 33
3.5 ลักษณะภาพส่วนประกอบที่ไม่อยู่ในลักษณะสี่เหลี่ยม 33
3.6 ขอบเขตของส่วนประกอบในภาพเอกสารที่แบ่งโดยหน้าต่างขนาด 32*32 พิกเซล 34
4.1 แผนผังขั้นตอนการทำงาน 36
4.2 แผนผังการขั้นตอนการทำงานของการแบ่งส่วนประกอบใหญ่ 43
4.3 แผนผังการทำงานการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ 44
4.4 ตัวอย่างการวางรูปแบบของหน้าต่างในย่อหน้าต่างๆ 45
5.1 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ 47
5.2 แสดงภาพตัวอย่างที่ 2 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ 48
5.3 แสดงภาพตัวอย่างที่ 10 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ 49
5.4 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ 50
5.5 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel 51
5.6 แสดงภาพตัวอย่างที่ 2 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel 52
5.7 แสดงภาพตัวอย่างที่ 10 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel 53
5.8 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel 54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.9 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย
และแสดงขอบเขตของตัวอักษรในแต่ละตัว 55
5.10 แสดงภาพตัวอย่างที่ 5 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย 56
5.11 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย 57
5.12 แสดงภาพตัวอย่างที่ 9 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย
และแสดงขอบเขตของตัวอักษรในแต่ละตัว 57

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มา

ปัจจุบันข้อมูลเป็นสิ่งสำคัญ ความต้องการที่จะนำเข้าข้อมูลสู่คอมพิวเตอร์เพื่อใช้งานมีมากขึ้น การนำเข้าข้อมูลโดยการพิมพ์ มีความล่าช้ากว่าการนำเข้าด้วยภาพมาก แต่การเก็บเอกสารในรูปแบบนี้มีข้อด้อยหลายประการ ข้อแรกคือการจัดเก็บเอกสารในลักษณะภาพมีขนาดใหญ่กว่าการจัดเก็บเอกสารที่เป็นข้อความมาก ข้อสองเอกสารที่เก็บในลักษณะภาพไม่สามารถนำมาแก้ไขเปลี่ยนแปลงได้อย่างเอกสารที่เป็นตัวอักษร และข้อสุดท้ายเอกสารที่จัดเก็บเป็นรูปภาพ ไม่สามารถนำไปใช้ในการค้นหาว่าเนื้อหาสาระข้างในเอกสารเป็นอย่างไร เราจึงต้องการข้อมูลที่เป็นตัวอักษร มากกว่าเป็นรูปภาพ นำไปสู่การพัฒนาระบบจดจำตัวอักษร(OCR : Optical Character Recognition) ในตอนแรกนั้นยังกำหนดขอบเขตที่จะให้ทำงานตามที่ต้องการด้วยมืออยู่ โดยเฉพาะเอกสารที่มีหลายคอลัมน์ หรือ มีส่วนประกอบที่ซับซ้อนซึ่งระบบจดจำตัวอักษร ไม่สามารถแบ่งแยกได้ว่ารายละเอียดพวกนี้จะเป็นกลุ่มเดียวกันหรือไม่ ซึ่งทำให้เกิดความยุ่งยากและล่าช้าในงานที่ต้องจัดเก็บเอกสารจำนวนมาก จึงมีความต้องการกำหนดขอบเขตที่จะนำไปใช้ในระบบจดจำตัวอักษรแบบอัตโนมัติขึ้นมา เพื่อให้เกิดความสะดวกและทำงานได้รวดเร็วยิ่งขึ้น ทำให้เกิดงานวิจัยทางด้านการแบ่งขอบเขตเอกสารภาพขึ้นอย่างมาก

1.2 ขอบเขตของโครงงาน

สำหรับขอบเขตของงานวิจัยนี้ คือแบ่งส่วนประกอบภาพเอกสารไบนารีที่มีลักษณะซับซ้อน ให้สามารถใช้งานกับระบบจดจำตัวอักษรได้ ซึ่งจะทำให้การนำเข้าข้อมูลสู่คอมพิวเตอร์สะดวกและทำงานได้รวดเร็วยิ่งขึ้น ไม่จำเป็นต้องแบ่งเอกสารด้วยมืออีกต่อไป

1.3 ความมุ่งหมายและวัตถุประสงค์ของโครงงาน

ความมุ่งหมายและวัตถุประสงค์ของการศึกษางานปริญญานิพนธ์นี้ เพื่อพัฒนาวิธีการแบ่งส่วนประกอบภาพเอกสารไบนารีแบบอัตโนมัติที่รวดเร็วและถูกต้อง เพื่อให้ได้ขอบเขตส่วนประกอบที่จะนำไปใช้กับระบบจัดเก็บเอกสารและจดจำตัวอักษร เพื่อให้ระบบทำงานได้มีประสิทธิภาพยิ่งขึ้น โดยงานวิจัยนี้จะนำภาพเอกสารไบนารีมาแยกส่วนประกอบเอกสารซึ่งมีรายละเอียดหลายชนิด โดยเน้นความสำคัญของส่วนประกอบที่เป็นประเภทตัวอักษร และสามารถกำหนดขอบเขตของส่วนประกอบที่มีลักษณะไม่เป็นสี่เหลี่ยมได้ ซึ่งสามารถทำให้ได้เฉพาะส่วนที่ต้องการเท่านั้น และเพิ่มความสามารถของระบบจัดเก็บภาพเอกสารให้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

การนำเข้าข้อมูลประเภทตัวอักษรด้วยใช้การแสกนเอกสารจะสามารถทำได้ด้วยความสะดวกและรวดเร็วยิ่งขึ้น โปรแกรมสามารถแบ่งเอกสารเองได้โดยอัตโนมัติ ประหยัดเวลายิ่งขึ้น และโครงงานนี้ได้นำไปสู่การพัฒนาการแบ่งเอกสารต่างๆ จากภาพ ได้มีประสิทธิยิ่งขึ้น

1.5 ทฤษฎีหรือแนวคิดที่ใช้ในการศึกษา

ภาพเอกสารที่นำมาใช้ในการทดลองนี้ เป็นภาพเอกสารไบนารีที่มีการจัดเก็บแบบบิตแมป โดยการจัดเก็บข้อมูลให้อยู่ในรูปแบบลำดับของพิกเซล ซึ่งค่าของข้อมูลในแต่ละพิกเซลจะมีค่าเป็น 0 หรือ 1 เท่านั้น (ขาวหรือดำ)

การแบ่งส่วนประกอบและจัดประเภทส่วนประกอบของภาพเอกสารนั้น ขั้นแรกต้องกำหนดขอบเขตส่วนประกอบชนิดที่เป็นลักษณะเดียวกันอยู่ในขอบเขตเดียวกัน ซึ่งจะพูดถึงวิธีการทั้งแบบบนลงล่าง(top-down) เช่น วิธีการตัดซ้ำแนวแกน X-Y (recursive projection X-Y cut) ส่วนวิธีการแบบล่างขึ้นบน (bottom-up) เช่น วิธีการรวมส่วนประกอบเล็กที่มีลักษณะที่สัมพันธ์กันมาเป็นส่วนประกอบที่ใหญ่ขึ้น (connected components) ว่ามีข้อดี ข้อเสีย อย่างไร ซึ่งในงานวิจัยนี้ใช้หน้าต่างในการเดินวนรอบส่วนประกอบ ซึ่งจะต้องศึกษาว่าหน้าต่างความมีขนาดเท่าใด ในการกำหนดขอบเขตส่วนประกอบที่อยู่ติดกันเข้าไว้ด้วยกัน เพื่อดูว่าขนาดใดจึงจะเหมาะสม ส่วนการแบ่งประเภทจะเทียบจากความสูงและความกว้างของส่วนประกอบ ว่าส่วนประกอบอักษร ไม่ควรเป็นเท่าใด เพื่อตัดทิ้งส่วนประกอบภาพ และสิ่งรบกวน

แนวทางที่คิด คือ หาวิธีการที่สามารถแบ่งขอบเขตส่วนประกอบภาพเอกสารไบนารี ที่มีลักษณะโครงสร้างรายละเอียดซับซ้อน และมีประสิทธิภาพในการทำงานที่รวดเร็ว จึงมีความคิดที่จะใช้วิธีการเดินรอบขอบในการกำหนดส่วนประกอบเอกสาร และมีการเดินรอบขอบส่วนประกอบย่อยเพื่อนำมาวิเคราะห์เพื่อแก้ไขแบ่งส่วนประกอบที่ผิดพลาดในขั้นตอนแรก โดยใช้ขอบเขตของกลุ่มที่เป็นตัวอักษรมาทำการจัดเรียงเป็นกลุ่มเพื่อแบ่งคอลัมน์ที่อยู่ในส่วนประกอบใหญ่เดียวกันออกจากกัน

1.6 วิธีการดำเนินงาน

ขั้นแรก ต้องศึกษาวิธีการแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบภาพเอกสาร เพื่อใช้ในการกำหนดหัวข้อ เป็นหมาย วัตถุประสงค์ และขอบเขตของการทำปริญญานิพนธ์ เมื่อได้วิธีการแบ่งส่วนประกอบเอกสารที่ดีที่สุด นั้นคือ การเดินรอบส่วนประกอบเอกสาร จึงเขียนโปรแกรมตามแนวทางนั้น เพื่อให้ได้โปรแกรมที่แบ่งเอกสารได้อย่างมีประสิทธิภาพ สุดท้ายจึงจัดทำเอกสารประกอบปริญญานิพนธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7 ข้อจำกัดของการศึกษา

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

1.8 รายละเอียดในแต่ละบท

วิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บท ประกอบด้วย

บทที่ 1 กล่าวถึง ความสำคัญและที่มา วัตถุประสงค์ รวมทั้งทฤษฎีหรือแนวคิด ขอบเขต และวิธีที่ใช้ในการทำวิจัยนี้

บทที่ 2 อธิบายทฤษฎีและแนวคิดที่เกี่ยวข้อง กล่าวถึงหลักการเดิมที่มีอยู่แล้วในลักษณะอธิบายและวิเคราะห์การทำงานทั้งด้านที่ดีและด้อย

บทที่ 3 แนวคิดและการนำมาประยุกต์ใช้ของการแบ่งส่วนประกอบภาพเอกสารไบนารี กล่าวถึงแนวทางที่ได้คิดประยุกต์ขึ้นมาใหม่ เพื่อนำมาทดลองตามแนวความคิด ทั้งในด้านการแบ่งขอบเขตและแบ่งประเภทส่วนประกอบของภาพเอกสาร

บทที่ 4 อัลกอริทึมและขั้นตอนการทำงาน กล่าวถึงอัลกอริทึมที่ใช้ในการแบ่งขอบเขตและแบ่งประเภทส่วนประกอบ และแผนผังขั้นตอนการทำงาน

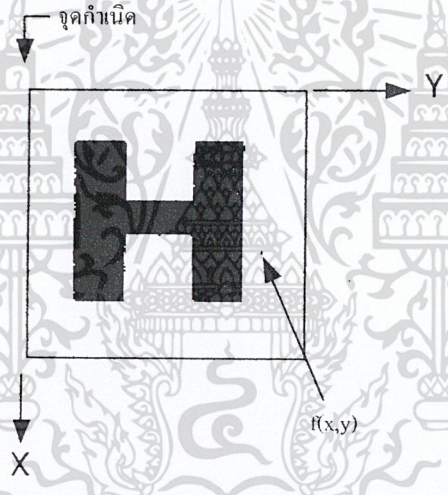
บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ เป็นการสรุปผลการทดลองและแนวทางการพัฒนาต่อไปในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและแนวคิดที่เกี่ยวข้อง

2.1 การแสดงภาพดิจิตอล(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 โดยภาพดิจิตอลสามารถนำไปพิจารณาให้อยู่ในรูปเมตริกซ์ที่ใช้แถว(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) เป็นขั้นตอนนำภาพที่อยู่ในสื่ออื่นๆซึ่งไม่ใช่ดิจิตอล เช่น ภาพถ่าย ซองจดหมาย มาแปลงให้เป็นดิจิตอลโดยผ่านอุปกรณ์ เช่น สแกนเนอร์(scanner) , กล้องดิจิตอล(digital camera) เป็นต้น

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

3. การแบ่งส่วนประกอบ (segmentation) เป็นขั้นตอนการแบ่งส่วนข้อมูลภาพให้เป็นส่วนที่สัมพันธ์กันหรือเป็นวัตถุ โดยการแบ่งส่วนประกอบอัตโนมัติเป็นงานที่ยากที่สุดในงานประมวลผลภาพดิจิตอล ถ้าขั้นตอนการแบ่งได้ดีและเหมาะสมจะทำให้ผลลัพธ์ที่ได้ดี แต่ถ้าอัลกอริทึมที่ใช้ในการแบ่งไม่ดี

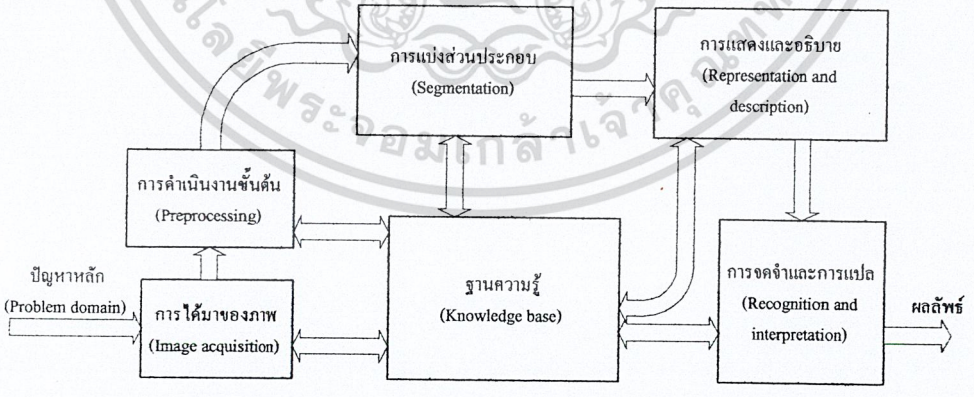
เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก็จะทำให้ผลลัพธ์ที่ได้ไม่ดีตามไปด้วย ตัวอย่างเช่น ในกรณีการจดจำตัวอักษร(character recognition) ต้องแบ่งแยกเฉพาะตัวอักษรและคำออกมาจากฉากพื้น(black ground)

4. การแสดงและอธิบายผลลัพธ์(representation and description) ผลลัพธ์ของขั้นตอนการแบ่งปกติเป็นข้อมูลจุดภาพ(raw pixel data) ส่วนประกอบหรือขอบเขตของพื้นที่(constituting or boundary of region) หรือ จุดทั้งหมดภายในพื้นที่ การเปลี่ยนข้อมูลให้ไปอยู่ในรูปแบบสำหรับการประมวลผลเป็นสิ่งจำเป็น ขั้นแรกต้องทำข้อมูลให้อยู่ในรูปขอบเขตหรือพื้นที่ที่สมบูรณ์ โดยขอบเขตแสดงได้โดยรูปร่างของรูปทรง เช่น มุม หรือ เส้นขอบ พื้นที่สามารถแสดงได้โดยพื้นผิว(texture) และรูปร่างของโครงสร้าง(skeletal shape) ขั้นตอนนี้ควรจะอธิบายข้อมูลถึงลักษณะและจุดเด่น ซึ่งจะสามารถแยกออกว่าวัตถุนี้แตกต่างจากวัตถุนั้นอย่างไร 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) ดูในรูปที่ 2.2 เริ่มตั้งแต่การนำปัญหาที่มีมาจัดอยู่ในฐานข้อมูลความรู้(knowledge database) เพื่อให้ระบบประมวลผลภาพเข้าใจ ความรู้ในที่นี้หมายถึงรายละเอียดขอบเขตทั่วไปของภาพ ซึ่งความซับซ้อนของฐานความรู้ขึ้นอยู่กับเรื่องที่สนใจ



รูปที่ 2.2 ขั้นตอนพื้นฐานของกระบวนการประมวลผลภาพดิจิตอล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 ความรู้พื้นฐานเกี่ยวกับข้อมูลภาพอิเล็กทรอนิกส์

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

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

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) ซึ่งชุดนี้เรียกว่า ส่วนที่ติดกัน 4 ทิศของ p (4-neighbors) หรือ $N_4(p)$ ซึ่งแต่ละจุดจะมีระยะห่างจาก (x,y) แล้วยังมีส่วนที่ติดกันของ p ที่อยู่นอกเหนือจากนี้อีก คือ (x+1,y+1) , (x+1,y-1) , (x-1,y+1) , (x-1,y-1) ซึ่งเป็น $N_D(p)$ และทั้งหมดเรียกว่า $N_8(p)$

2.3.3 การเชื่อมต่อกันระหว่างจุด ซึ่งใช้ในการหาค่าขอบเขตพื้นที่ของวัตถุและส่วนประกอบโดยการเชื่อมต่อกันระหว่างจุด กำหนดได้สามชนิดดังนี้

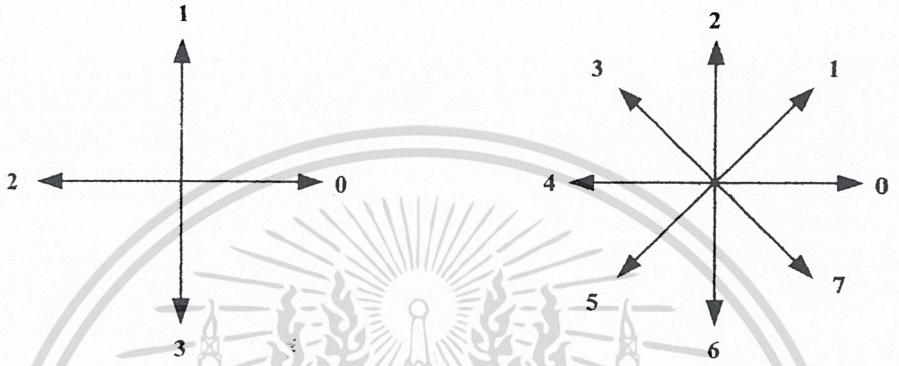
1. ส่วนติดกัน 4 ส่วน(4-connectivity) ของจุด p และ q กำหนดให้ V เป็น 4 ส่วนที่ติดกัน ถ้า q อยู่ในเซตของ $N_4(p)$
2. ส่วนติดกัน 8 ส่วน(8-connectivity) ของจุด p และ q กำหนดให้ V เป็น 8 ส่วนที่ติดกัน ถ้า q อยู่ในเซตของ $N_8(p)$
3. ส่วนติดกัน m ส่วน(m-connectivity) ของจุด p และ q กำหนด V เป็น m ส่วนที่ติดกัน ถ้า
   - q อยู่ใน $N_4(p)$ หรือ
   - q อยู่ใน $N_D(p)$ 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 และความหลากหลายของสีจะเพิ่มขึ้นเมื่อเพิ่มขนาดการเก็บข้อมูลต่อ 1 พิกเซล เช่น 8 บิตต่อ 1 พิกเซลจะมีการอ้างสีได้ถึง 256 สี คือ การอ้างถึงค่าสีจะอ้างเท่ากับ 2^n เมื่อ n คือ ขนาดที่เก็บข้อมูลต่อ 1 พิกเซล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 รหัสแนว (Chain Codes)

รหัสแนวถูกใช้ในการแสดงขอบเขตส่วนประกอบที่เรียงติดต่อกันในส่วนของเส้นตรงที่มีความยาวและทิศทาง อยู่บนหลักเกณฑ์ 4 หรือ 8 ทิศทาง โดยส่วนประกอบติดกันของข้อ(segment) กำหนดทิศทางแต่ละข้อเป็นรหัสใช้ตัวเลขดังแสดงในรูปที่ 2.3 ซึ่งขอบเขตส่วนประกอบส่วนมากจะถูกอธิบายในรูปแบบของพิกัดจุด (x,y)



(ก) (ข)

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

2.5 คุณลักษณะพื้นฐานของภาพเอกสาร

เอกสารที่ถูกสแกนและนำมาแสดงจะอยู่ในรูปภาพแบบไบนารี(binary : 1 บิตต่อจุด) เกรย์สเกล (gray-scale : 8 บิตต่อจุด) หรือ สี(color : 8-24 บิตต่อจุด) ชนิดของภาพที่ใช้ขึ้นอยู่กับการนำไปใช้และการแปลงระหว่างชนิด เช่น เปลี่ยนภาพสีไปเป็นเกรย์สเกลหรือไบนารี หรือเปลี่ยนภาพเกรย์สเกลเป็นไบนารี โดยขึ้นกับความต้องการที่จะเพิ่มความเร็วในการคำนวณ ส่วนประกอบพื้นฐานที่แสดงในภาพ ได้แก่

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

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนผังทางภูมิศาสตร์(Geometric document layout) และ ความเอียง(skew) สองลักษณะที่สำคัญของภาพเอกสาร แผนผังเอกสารเป็นแนวทางในการจัดเรียงของส่วนประกอบ(ข้อความ, รูปภาพ, กราฟิก, อื่นๆ) ของเอกสาร แผนผังพื้นฐานสามชนิดได้แก่ แผนผังแมนฮัตตัน(Manhattan layout) ที่ขอบเขตเป็นรูปหลายเหลี่ยมที่คงที่ซึ่งขอบเขตนั้นเป็นเส้นตรงแนวนอนและแนวตั้ง 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) ซึ่งข้อความจะถูกจัดไปอยู่ในส่วนประกอบทางตรรกะ(logical layout)

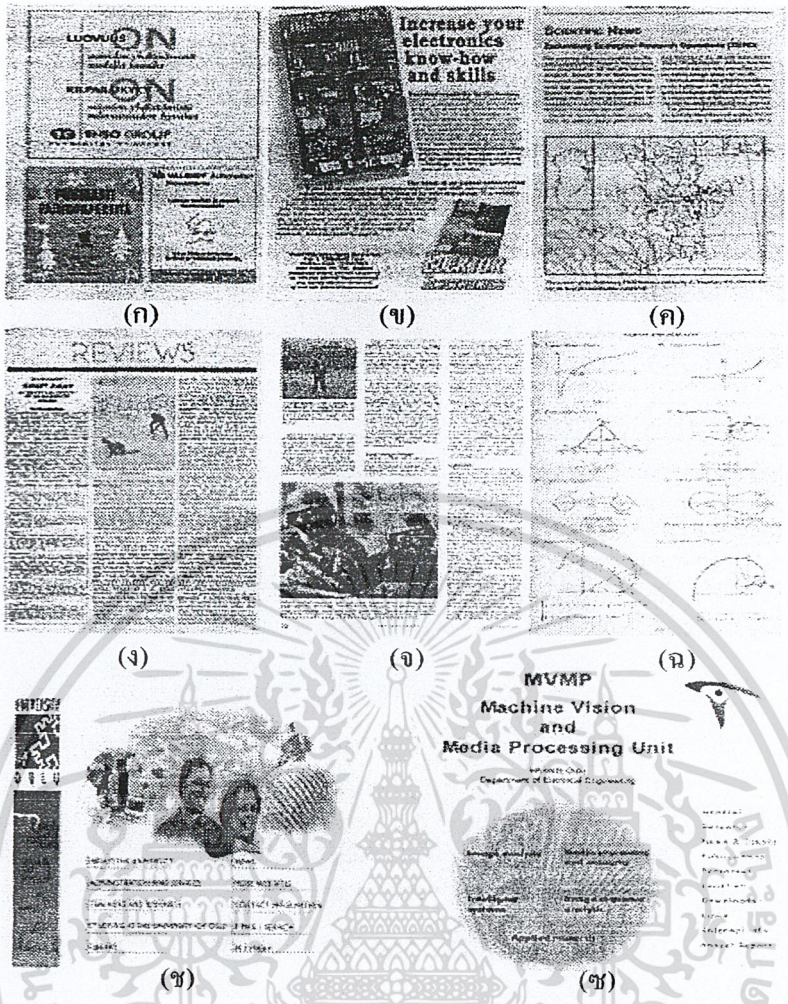
2.6 ประเภทของเอกสารและการประยุกต์ใช้

เนื่องจากมีความเป็นไปได้ยากที่จะใช้วิธีการแบบเดียวกันในการจัดประเภทของส่วนประกอบเอกสารที่มีส่วนประกอบหลากหลายและซับซ้อน ทำให้มีการเสนอวิธีการต่างๆ ที่ออกแบบมาเพื่อวิเคราะห์ภาพเอกสารที่มีส่วนประกอบภายในหลากหลายประเภท เห็นได้ชัดว่าประโยชน์ส่วนมากเริ่มมาจากการแบ่งเอกสารโดยใช้แผนผังทางภูมิศาสตร์ ดังนั้นเมื่อมีวิธีการใหม่ๆเพิ่มขึ้น สามารถที่จะประเมินคุณค่าจากวิธีการวิเคราะห์แผนผังที่ใช้จากคุณสมบัติพื้นฐานของประเภทส่วนประกอบเป้าหมาย แม้กระนั้นก็ยังไม่สามารถที่จะพิจารณาตามลำดับของใจความได้ หรือได้แต่พิจารณาในประเภทที่จำกัด ในการจัดประเภทของชนิดเอกสารแบ่งตามลักษณะได้ดังนี้

1. เอกสารรูปแบบโครงสร้าง (วารสาร หนังสือพิมพ์ และ จดหมายข่าว)
2. เอกสารที่มีรูปแบบไม่แน่นอน (โฆษณา , ปกซีดี , หนังสือ และ นิตยสาร)
3. เอกสารกึ่งรูปแบบ (ซองจดหมาย , จดหมายและนามบัตร , เช็คธนาคาร , แบบฟอร์ม และเอกสารที่คล้ายตาราง)
4. แผนที่และภาพวาดเชิงวิศวกรรม
5. เอกสารที่ไม่ปกติ(หน้าโฮมเพจ และ รูปจากวีดีโอ)

ในส่วนนี้เป็นการสรุปภาพรวมของหลักการแบ่งส่วนประกอบและการแบ่งประเภทส่วนประกอบที่ขึ้นอยู่กับตำแหน่งส่วนประกอบของตัวอย่างในแต่ละชนิดที่กล่าวถึงทั้งหมดรวมถึงปัญหาที่เกิดขึ้น จากรูปที่ 2.4 แสดงตัวอย่างภาพเอกสารชนิดต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก) (ข) (ค)

(ง) (จ) (ฉ)

(ช) (ซ)

รูปที่ 2.4 ตัวอย่างภาพเอกสารชนิดต่างๆ[*]

ก. สิ่งพิมพ์โฆษณา ข. สิ่งพิมพ์โฆษณา ค. หนังสือพิมพ์

ง. บทความ จ.บทความ ฉ. หนังสือ

ช. หน้าเว็บ ซ. หน้าเว็บ

2.6.1 เอกสารรูปแบบโครงสร้าง(structure articles)

วารสาร หนังสือพิมพ์ และจดหมายข่าว เป็นแหล่งข้อมูลข่าวสารแรกและเป็นสิ่งพิมพ์ที่แพร่หลายและหาง่าย ภาพทั้งหมดที่เก็บมีหลากหลาย (ข้อความที่มีขนาด รูปทรง และแบบที่แตกต่างกัน ความหลากหลายของกราฟิก ภาพไบนารี ภาพฮาฟโทนและภาพสี) ซึ่งสามารถสแกนเป็นไบนารี เกรย์สเกลหรือสี ขอบเขตนั้นจะถูกแบ่งตามชนิดกายภาพของแต่ละประเภท อย่างไรก็ตาม ถ้าช่องว่างระหว่าง

[*] ที่มา Okun O, Doermann D & Pietikäinen M, "Page Segmentation and Zone Classification: the state of the art" ,Technical Report LAMP-TR-036(1999) , page 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขอบเขตส่วนประกอบที่แตกต่างกันแตบมาก และในบางกรณีจะมีข้อความที่อยู่ในกราฟิกหรือรูป บางที ตัวหนังสือก็สีเข้ม(สิ่งพิมพ์ปกติ) หรือ สีสว่าง(สีตรงข้ามกับปกติ)กว่าฉากพื้น บทความในวารสาร หนังสือพิมพ์ และ จดหมายข่าวเป็นตัวอย่างของเอกสารโครงสร้าง เพราะทุกส่วนประกอบบนหน้ามีการจัดเรียงลำดับตามชนิด เช่น หัวเรื่อง คำนำ และเนื้อหา โดยจะเชื่อมต่อเข้าด้วยกัน ด้วยกฎทั่วๆไป ตัวอย่างของกฎเช่น การอ่านตามลำดับของบล็อกข้อความจากบนลงล่าง และจากซ้ายไปขวา คำอธิบายภาพจะอยู่ใต้ภาพ แบบฟอร์มข้อความจะสัมพันธ์กับขนาดของพื้นที่ รูปหรือตารางจะอยู่หลังจากที่ข้อความอ้างถึง หนังสือพิมพ์จะมีความซับซ้อนมากกว่าวารสารหรือจดหมายข่าว แต่องค์ประกอบของโครงสร้างทางตรรกะจะน้อยกว่า ซึ่งทั้งหมดก็สามารถแสดงในแผนผังแมนฮัตตัน และแผนผังไม่มีกฎเกณฑ์ได้ ซึ่งจะรวมถึงหน้าเอกสารของวารสารที่คล้ายตาราง แต่แผนผังนี้ใช้ได้ไม่ค่อยดีกับวารสารที่ไม่ใช่เชิงเทคนิค ถึงกระนั้นรูปแบบของแผนผังไม่ว่าจะเป็นภาษาหรือสิ่งพิมพ์พิเศษ ซึ่งปกติแล้วจะมีรูปแบบของเนื้อหาและรายละเอียดมากมาย โดยประเด็นงานหลัก คือการกำหนดตำแหน่งขอบเขตรวมถึง การประมาณและแก้ไขความเอียง(บางครั้งเป็นตัวเลือก) การแยกเอกสารเฉพาะส่วนที่เหมือนกัน และนำไปจัดแบ่งประเภทขอบเขตเป็นฉากพื้น ข้อความ กราฟิก และ รูปภาพ โดยทั่วไปบรรทัดข้อความที่ไม่มีความเอียงมีรูปแบบการจัดเรียงตามแนวตั้งและแนวนอน แต่สามารถเอียงได้ถ้า แกน X หรือ แกน Y เอียง ปกติทุกขอบเขตบนภาพจะมีความเอียงเดียวกันทั้งหน้าเอกสาร แต่บางกรณีอาจมีบางขอบเขตที่มีทิศทางแตกต่างจากขอบเขตอื่นๆ

แม้จะมีแผนผังโครงสร้างหลายรูปแบบ แต่ปัญหาหลักที่ยังคงอยู่คือความแตกต่างของชนิดภาพเงื่อนไขการนำไปใช้ซึ่งควรปรับปรุงให้ดี(ความเอียง , รูปแบบแผนผังเอกสาร) และสัญญาณรบกวน

2.6.2 เอกสารที่มีรูปแบบไม่แน่นอน

นิยาม เอกสารที่มีรูปแบบไม่แน่นอน สะท้อนให้เห็นความขาดแคลนของกฎทั่วไปที่จะนำมากำหนดส่วนประกอบเหล่านี้ แต่ส่วนมากการกำหนดแผนผังจะขึ้นอยู่กับลักษณะเอกสารและข้อกำหนดซึ่งออกแบบมาใช้สำหรับรายละเอียดที่สนใจ

- โฆษณา

โฆษณาปกติจะพิมพ์อยู่ในนิตยสารหรือหนังสือพิมพ์ แต่ก็สามารถอยู่บนเว็บเช่น เอกสารอิเล็กทรอนิกส์ แผนผังของเอกสารดังกล่าวจะไม่คงที่เหมือนกับหัวเรื่องวารสารและหนังสือพิมพ์ พื้นที่ข้อความภาพมีได้หลายความเอียง บรรทัดข้อความเป็นรูปโค้ง ซึ่งตัวอักษรไม่เรียงลำดับกันเป็นแนวเส้นตรง และกลุ่มข้อความข้างในหรือที่ติดกับรูปภาพเกิดได้บ่อยในการแบบที่เป็นข้อมูลสำคัญ และเพื่อที่ดึงดูดสายตาของผู้อ่าน ดังนั้นโฆษณาเป็นเอกสารแบบไม่เป็นโครงสร้าง [42,48]

- หน้าปก

ภาพของเอกสารที่บ่อยๆจะใช้สีและมีตำแหน่งข้อความที่หลากหลาย และรูปแบบอักษรหลายขนาดทั้งฉากพื้นมีหลายสีหรือหลายลวดลาย

การวิเคราะห์แผนผังของภาพเพื่อจัดประเภทข้อความและจดจำ ซึ่งมีประโยชน์สำหรับการเรียกค้นข้อมูล การเรียกค้นข้อความง่ายกว่าใช้สูตรอย่างใดอย่างหนึ่งในการค้นหาภาพเช่น รูปทรง พื้นผิว หรือสี ความท้าทายอยู่ที่แผนผังไม่แน่นอนและพื้นฉากมีหลายสี ซึ่งทำการตรวจหาตัวอักษรให้ถูกต้องได้ยาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.3 เอกสารกึ่งรูปแบบ (semi-structured layouts)

แผนผังกึ่งโครงสร้างมีส่วนประกอบโครงสร้างข้างในซึ่งผู้ใช้จะใส่ข้อมูลลงตรงที่กำหนดให้ แต่ชนิดของมันไม่มีกำหนดว่าจะอยู่ตรงส่วนไหน หมายความว่าเอกสารสองชุดสามารถที่มีแผนผังที่แตกต่างกันทั้งๆที่มีส่วนประกอบข้างในเหมือนกัน

- ซองจดหมายและเช็คธนาคาร

ไปรษณียบัตรมีรูปแบบเดียวกับเช็คธนาคารซึ่งเป็นตัวอย่างเอกสารแบบกึ่งโครงสร้าง ซึ่งสามารถกำหนดได้ก่อนว่าฟิลด์แต่ละฟิลด์เป็นบล็อกที่อยู่ หรือ ผลรวมจำนวนเงิน แต่ขอบเขตของฟิลด์นั้นๆที่อยู่ในเอกสารเป็นระเบียบแบบแผนหลักที่สำคัญ นามบัตรก็ถูกรวมอยู่ในประเภทนี้ด้วยเพราะ มีโครงสร้างทางตรรกะคล้ายคลึงกัน ภาพของเอกสารชนิดนี้จะเป็นไบนารี เกรย์สเกล หรือ สีและข้อความจะถูกพิมพ์บนพื้นผิวฉากที่ซับซ้อน

งานหลักที่ต้องดำเนินการ คือ การกำหนดความแตกต่างของฟิลด์ซึ่งมีลักษณะเฉพาะของแต่ละเอกสาร ดังเช่นเช็คธนาคารรายละเอียดนั้นจะรวมถึง ลายเซ็น หมายเลขเช็ค วันที่ ผลรวมยอดเงินเป็นตัวเลข ผลรวมยอดเงินที่เป็นตัวอักษร เลขที่บัญชี ชื่อผู้สั่งจ่าย ที่อยู่ของสถาบันการเงิน และ/หรือตรา แต่สำหรับนามบัตรประกอบด้วย ชื่อผู้ถือ กิจการ และที่อยู่ ส่วนไปรษณีย์บัตรมีบล็อกที่อยู่ แสตมป์ บาร์โค้ด (barcode) และสถานที่รับไปรษณีย์

ปัญหาหลักที่สังเกตได้

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

- แบบฟอร์มและเอกสารที่เป็นตาราง

แบบฟอร์มและเอกสารที่เป็นตาราง เช่น กระดาษคำตอบ หรือ ใบส่งของ ที่ประกอบด้วยฟิลด์หรือเซลล์ข้างในซึ่งข้อมูลจะถูกเขียนหรือพิมพ์เข้าไป ในการวิเคราะห์ภาพเริ่มต้นปกติใช้ในรูปแบบภาพไบนารี จนถึงแบบฟอร์มสีก็มี อย่างไรก็ตามอันหลังถ้าฉากพื้นภาพเป็นรูปแบบเดียว ก็ไม่ยากที่จะแบ่งข้อความออกมาจากแบบฟอร์ม ในส่วนเพิ่มเติมของข้อความ ใบส่งของสามารถเก็บรูปเล็กๆ เช่น ตรา สัญลักษณ์ รูปแบบและเอกสารแบบตารางเป็นเอกสารกึ่งโครงสร้างเพื่อสร้างจำกัดโดยใช้ตัวแบ่งเป็นตัวกำหนดขอบเขตข้อมูลตัวอักษร

งานพื้นฐานที่จะใช้ในการแยกขอบเขตของฟิลด์ ขอบเขตบล็อก หรือเส้นบรรทัดออก เพื่อที่จะสกัดเฉพาะข้อความภายในแต่ละฟิลด์ ใบส่งของโดยปกติมีที่อยู่ของบริษัท และเลขที่บัญชีธนาคาร ซึ่งการดำเนินงานก็เหมือนกับการประมวลผลเช็คธนาคาร

เอกสารนี้เป็นทรัพย์สินทางปัญญาของจุฬาลงกรณ์มหาวิทยาลัย ห้ามมิให้ทำซ้ำ ดัดแปลง หรือนำไปใช้เพื่อการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น เว้นแต่จะได้รับอนุญาตจากจุฬาลงกรณ์มหาวิทยาลัยเป็นลายลักษณ์อักษร และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวิเคราะห์แบบฟอร์มทั้งสองชนิด อันดับแรกต้องรู้แบบฟอร์มที่จะวิเคราะห์ กำหนดฟิลด์ในฟอร์มให้อยู่รูปแบบประเภทใดประเภทหนึ่ง และรูปแบบ(model)ของฟิลด์ที่ว่างหรือเซลล์ที่มียังไม่ได้ใส่ข้อมูล ในกรณีนี้งานแรกคือการกำหนดฟอร์มโดยใช้ค่าที่กำหนดหรือโครงสร้างที่สามารถถูกสกัดจากแบบได้ มีอยู่สองชนิดคือ อย่างแรกชนิดแบบฟอร์มถูกกำหนดเพียงครั้งเดียว เพื่อที่กำหนดพิกัดขอบเขตของฟิลด์ข้อมูลที่ต้องการ อย่างที่สองเป็นการวิเคราะห์แบบฟอร์มที่ไม่รู้เกี่ยวกับรูปแบบที่นำมาใช้ เมื่อไม่รู้เกี่ยวกับชนิดของแบบฟอร์มที่ต้องใช้มาก่อน ต้องใช้อัลกอริทึมที่ใช้สกัดแยกพื้นฐานโครงสร้างแบบฟอร์มเป็นพื้นฐานของตัวแบ่งระหว่างเซลล์หรือฟิลด์ของชนิดที่สองนี้จะดัดแปลงได้ยืดหยุ่น แต่จะใช้การคำนวณมาก 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 ดังนั้นจึงมีบางกฎที่ใช้สำหรับแยกสิ่งที่ผสมกันอยู่ โดยสามารถที่จะแยกได้ถูกต้องว่าเป็นตัวหนังสือหรือขอบเขตที่แคบของแผนที่ เพื่อที่จะทำให้ไม่คลุมเครือ และเพื่อกำหนดให้เป็นส่วนประกอบที่อยู่ในโครงสร้าง ปกติข้อความในแผนที่จะปรากฏได้ทุกแห่งในแผนที่และแนวบรรทัดของข้อความมีทิศทางไม่แน่นอน

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

1. การแบ่งข้อความและกราฟิก
2. กำหนดค่าให้ตัวอักษรและสัญลักษณ์พิเศษ
3. การกำหนดทิศทางของกราฟิก

สำหรับภาพสีการแบ่งสีจะถูกทำตั้งแต่การสแกนจากภาพทีละสีโดยสแกนเนอร์ที่ออกแบบเพื่อใช้ในงานนี้โดยเฉพาะ ตัวอักษรและสัญลักษณ์จะถูกแบ่งแยกออกจากกราฟิกที่ระดับสีโดยใช้ขนาดของตัวอักษรและสัญลักษณ์ โดยทั้งหมดจะถูกส่งไปยังกระบวนการจดจำ ขณะที่กราฟิกจะถูกทำให้เล็กลงและกำหนดขนาดและทิศทางของกราฟิก

เอกสารนี้มีไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากแผนที่มีรูปแบบที่หลากหลายจึงก่อให้เกิดวิธีการและระบบที่ใช้งานมากมาย โดยแต่ละวิธีอาจจะถูกนำไปประมวลโดยเฉพาะกับแผนที่ชนิดใดชนิดหนึ่งได้เท่านั้น ปัญหารวมๆคือ ตัวอักษรและสัญลักษณ์สัมผัสหรือซ้อนทับกราฟิก ข้อความเอียงหรือโค้ง ระยะห่างระหว่างตัวอักษรที่ใกล้กันและในข้อความเดียวกันมีระยะห่างที่ไม่แน่นอน ความแตกต่างของขนาดตัวอักษร (แม้ว่าจะรู้ล่วงแบบอักษรที่ใช้ในการสร้างแผนที่) และสัญลักษณ์กราฟิกบางตัวที่คล้ายกับตัวอักษร

- ภาพวาดเชิงวิศวกรรม

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

การแบ่งชนิดของภาพวาดเชิงวิศวกรรมที่ใช้ส่วนมากคือ วิธีการเฉพาะที่สามารถนำมาใช้ได้กับส่วนประกอบหนึ่งหรือสองชนิดที่แตกต่างกัน ปัญหาที่เป็นปัญหาหลักของเอกสารแบบนี้ คือ ข้อความที่ติดกับกราฟิก หรือ เอกสารต้นฉบับมีคุณภาพแย่

2.6.5 หน้าเว็บ และ ภาพจากวีดีโอ

หน้าเว็บและภาพจากวีดีโอที่มีรายละเอียดเป็นอักษรถูกจัดให้เป็นเอกสารประเภทใหม่ คือ เอกสารอิเล็กทรอนิกส์ ซึ่งจะมีคุณลักษณะไม่เหมือนกับเอกสารอื่นที่อธิบายมาแล้วก่อนหน้า เอกสารนี้โดยทั่วไปเป็นหน้าสีและภาพจะมีความละเอียดต่ำ มีขนาดที่เล็กซึ่งจะเรียกดูภาพได้รวดเร็ว โดยรูปภาพจะช่วยให้เกิดความเข้าใจที่ชัดเจนยิ่งขึ้น ส่วนประกอบหลักของเอกสารมีทั้งข้อความและรูปภาพที่มีขอบเขตและตำแหน่งไม่แน่นอน ตัวอักษรในหน้าเว็บที่เห็นอาจเป็นได้ทั้งตัวอักษรในรูปแบบตัวอักษรและตัวอักษรและรูปแบบรูปภาพอย่างใดอย่างหนึ่ง สำหรับภาพจากวีดีโอนั้นส่วนมากข้อความจะอยู่ในส่วนของภาพที่ค่อนข้างซับซ้อน สิ่งที่ปรากฏอาจจะเป็นฉากหรือข้อความกราฟิกอย่างใดอย่างหนึ่ง และความละเอียดของภาพที่น้อยกว่า 200-300 จุดต่อนิ้ว สำหรับเอกสารชนิดนี้จะทำการตรวจสอบและจดจำส่วนประกอบได้ยาก

หน้าเว็บและภาพจากวีดีโอ จัดอยู่ในประเภทเอกสารที่ไม่มีโครงสร้าง เพราะแผนผังของหน้าเว็บและภาพจากวีดีโอไม่สามารถจำกัดขอบเขตโดยการกำหนดกฎไว้ก่อน และในกรณีของภาพจากวีดีโอบ่อยครั้งที่ไม่สามารถกำหนดกฎได้ งานหลักของเอกสารประเภทนี้คือ การกำหนดขอบเขตของข้อความที่อยู่ในภาพ มีวัตถุประสงค์เพื่อการสืบค้นหาข้อมูลและทำดัชนีวีดีโอ

ปัญหาหลักของเอกสารประเภทนี้ ได้แก่

1. สีของข้อความและฉากพื้นจะใกล้เคียงกันเมื่อความละเอียดของภาพต่ำ หรือข้อความที่มีความเข้มน้อยเมื่อเทียบกับฉากพื้น (ตัวอย่างข้อความที่อยู่บนลวดลายฉากพื้นที่ซับซ้อน) ดังนั้นการเธรโชร์(thresholding)ไม่สามารถแบ่งข้อความและฉากพื้นได้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ ขออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ส่วนประกอบข้อความ อาจจะวิเคราะห์ผิดพลาดเมื่อเป็นบรรทัดโค้ง หรือในกรณีที่แย่ที่สุดคือไม่มีการจัดเรียงกันเป็นแนว (บรรทัดที่มีลักษณะเป็นคลื่น)
3. ส่วนประกอบข้อความที่ไม่ได้เป็นสีเดียวกัน และมีการขาดตอนไปเป็นส่วนย่อยหลายส่วน หรือเป็นสัญญาณรบกวนอันเนื่องมาจากการออกแบบตัวอักษร
4. ผลกระทบที่เกิดจากการเคลื่อนไหวของภาพในวีดีโอ

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 เช็คธนาคาร และแบบฟอร์ม เพราะโครงสร้างขอบเขตที่เหมือนกัน โดยทั้งหมดมีส่วนประกอบที่ผสมกันทั้งที่เป็นข้อความและกราฟิก และบ่อยครั้งที่ข้อความกับกราฟิกชิดกันหรือทับกัน

ในที่นี้การแบ่งส่วนประกอบหมายถึงการแบ่งภาพออกเป็นส่วนๆ ซึ่งในแต่ละส่วนต้องเป็นชนิดเดียวกัน หรือ พื้นที่แต่ละขอบเขตต้องเป็นข้อมูลเพียงหนึ่งชนิดเท่านั้น เช่น ข้อความหรือกราฟิก และจะมีการจัดประเภทส่วนประกอบหลังจากแบ่งขอบเขตส่วนประกอบเอกสารแล้ว ดังนั้นการจัดแบ่งประเภทส่วนประกอบจะกำหนดประเภทได้จากพื้นฐานที่เป็นจุดเด่น ซึ่งหลายกรรมวิธีนั้น ดำเนินการแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบ โดยแต่ละวิธีการจะกล่าวย่อๆ ไม่เน้นรายละเอียดของการนำไปใช้ แล้วจะยกตัวอย่างบางวิธีที่น่าสนใจในหัวข้อถัดไป

2.7.1 การแบ่งส่วนประกอบเอกสาร

การแบ่งส่วนประกอบเอกสารสามารถแบ่งตามแนวคิดได้เป็น วิธีการล่างขึ้นบน(bottom-up , data-driven) วิธีการบนลงล่าง(top-down , model-driven) และ วิธีการผสมผสาน(hybrid เป็นการผสมระหว่างวิธีการบนลงล่างและวิธีการล่างขึ้นบน) วิธีการแบบบนลงล่าง ใช้มานานและมีการทำงานที่รวดเร็ว เมื่อรู้รายละเอียดของลำดับแผนผังเอกสาร ในกรณีนี้กระบวนการทำงานจะเริ่มทำจากหน้าเอกสารทั้งหมดที่อยู่ในระดับสูงสุดนำมาแบ่งทั้งหมดให้เป็นขอบเขตที่เล็กลงเช่น บล็อก บรรทัด คำ และตัวอักษร วิธีการล่างขึ้นบนจะเริ่มจากจุดและรวมกลุ่มเข้าด้วยกันเป็นขอบเขต แล้วรวมขอบเขตที่ใกล้ชิดกันเข้าเป็นโครงสร้างที่ใหญ่ขึ้น ซึ่งวิธีการนี้มีความแม่นยำสามารถใช้ได้กับเอกสารที่มีรูปแบบที่หลากหลาย แต่เวลาในการประมวลผลส่วนมากใช้เวลานาน วิธีการแบบผสมผสานเป็นวิธีที่อยู่ระหว่างกลางของวิธีการบนลงล่างและวิธีการล่างขึ้นบน โดยพยายามที่จะรวมความสามารถในการทำงานด้วยความเร็วแบบวิธีการบนลงล่างและความมั่นคงถูกต้องในการทำงานของวิธีล่างขึ้นบน และจัดแบ่งประเภทของส่วนประกอบให้เป็นข้อความหรือกราฟิก โดยวิธีการเหล่านี้สามารถแบ่งขอบเขตเอกสาร โดยใช้การแบ่งลวดลายพื้นที่แตกต่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กันเป็นข้อกำหนดในการแบ่งขอบเขต ซึ่ง O. Okun และ M. Pietikäinen [1] ได้สรุปวิธีการใช้พื้นการแบ่งจากลวดลาย

2.7.1.1 วิธีการล่างขึ้นบน

วิธีการนี้ใช้พื้นฐานของการวิเคราะห์ส่วนประกอบของภาพไบนารีและรวมส่วนประกอบเหล่านั้นเข้าเป็นตัวอักษร บรรทัด และบล็อกโดยใช้ความใกล้ชิดระหว่างส่วนประกอบที่ใกล้กันและขนาดของส่วนประกอบ การทำเบลอเพื่อเชื่อมส่วนประกอบที่ใกล้ชิดกันเข้าด้วยกัน และไดอะแกรมโวโรนอย (Voronoi diagrams) เป็นวิธีการจับกลุ่มหลัก ระยะเวลาที่ใช้ในการทำงานในแต่ละภาพจะต่างกันขึ้นอยู่กับวิธีการที่ใช้ วิธีการในแบบล่างขึ้นบนนี้ส่วนมากมีความคงทน(tolerant)กับเอกสารที่มีความเอียง(ซึ่งบางทีอาจจะมีหลายความเอียง คือ ส่วนประกอบในภาพแต่ละส่วนเอียงไปคนละทิศทาง)และเอกสารที่มีแผนผังไม่แน่นอน ความเอียงจะถูกคำนวณจากผลรวมของบรรทัดข้อความที่แยกออกออกมา ถ้าความเร็วในการทำงานหรือหลักเกณฑ์ทั่วไปเป็นตัวกำหนด การเลือกตัวอย่างข้อมูลเป็นหลักสำคัญสำหรับวิธีการในกลุ่มนี้ ทุกขอบเขตจะถูกแสดงในโครงร่างต้นไม้แบบชั้นลำดับ สำหรับประมวลผลเอกสารที่หลากหลายชนิดที่เสนอโดย S.Y. Wang และ T. Yagasaki [2] สามารถใช้กับ แบบฟอร์ม หรือ วารสาร อักษรประกอบด้วยตัวภาษาอังกฤษ และตัวคันจิ บล็อกที่ใกล้ชิดกันใช้เวลาในการประมวลผลที่เร็วในการวิเคราะห์คุณสมบัติของบล็อกและการอ้างอิงของต้นไม้

การวิเคราะห์ส่วนที่ติดกันของบล็อกสี่เหลี่ยมก่อนที่จะแบ่งประเภทเข้าด้วยกันโดยใช้หน้ากากเล็กๆ จำนวนมากที่อยู่ในขอบเขตใหญ่ในการเชื่อมต่อระดับบล็อก สามารถลดเวลาที่ใช้ในการทำงานได้อย่างมากเมื่อเปรียบเทียบกับการเชื่อมต่อระหว่างจุด[3]

วิธีการที่ T. Pavlidis และ H. Zhou [4] เสนอคือ การสกัดแยกข้อมูลโดยการคำนวณค่าของจุดภายในหน้าต่าง 5*5 ที่แต่ละจุดจะถูกสแกนภาพโดยแถวต่อแถว วิธีการตัดส่วนที่อยู่ใกล้กันโดยใช้บล็อกขนาดเล็กมารวมจุดเข้าเป็นพื้นที่เดียวกัน ประโยชน์ของวิธีการนี้สามารถทำได้ทั้งภาพไบนารีและเกรย์สเกล แต่ไม่คงทนกับเอกสารที่เอียงหรือมีรูปแบบไม่แน่นอน

แนวคิดพื้นฐานของการวิเคราะห์ฉากพื้นของภาพไบนารีถูกนำมาเสนอใน [5], [6], [7], [8] วิธีการที่ใช้ได้ประโยชน์ที่สุด คือวิธีการที่เรียกว่า กระเบื้องขาว(white tiles) [5], [6] ในการทำงานอนแรกจะสร้างโครงข่ายของสี่เหลี่ยม ซึ่งแต่ละอันจะกว้างเท่ากับความกว้างของช่องว่างสีขาว(ฉากพื้น) และเป็นสี่เหลี่ยมที่อยู่ติดกันวนรอบขอบเขต วิธีการนี้เป็นการแสดงที่ยืดหยุ่น ที่สามารถแสดงได้ทั้งแผนผังแมนฮัตตันและแบบที่ไม่แน่นอน โดยไม่มีการแตกขอบเขตรูปทรงที่ซับซ้อน วิธีนี้ไม่ต้องแก้ไขความเอียงของภาพเอกสารก่อน และสามารถที่จะกำหนดขอบเขตพื้นที่ข้อความที่เอียงในตำแหน่งที่แตกต่างกันได้ วิธีการนี้ทำงานได้รวดเร็ว เพราะเป็นการทำงานกับจุดเพียงครั้งเดียว

การดำเนินงานเกี่ยวกับรูปร่างลักษณะ แบบเปิดหรือปิด ถูกนำมาใช้ใน [9], [10], [11] เพื่อจัดกลุ่มจุดภาพรวมเข้าไปเป็นขอบเขตใหญ่ก่อนจัดประเภท การเบลอภาพโดยใช้วิธีการรันเรงค์สมูทอัลกอริทึม (Run Length Smoothing (RLS)) ที่ใช้ใน [12], [13], [14], [15], [16] การดำเนินงานนี้คล้ายกับการขยายรูปร่าง บางวิธีในกลุ่มนี้ใช้ได้กับเอกสารที่มีความเอียงและแผนผังที่ไม่มีรูปแบบแน่นอน [10], [13], [17] ในขณะที่วิธีการอื่น[11], [12], [14], [15] ไม่สามารถที่จะใช้ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสกัดข้อความจากภาพเอกสารสีที่ซับซ้อน ใน [18], [19], [20], [21], [22], [23] ซึ่งในหลายวิธีนี้กำหนดให้ตัวอักษรมีสีเดียวกันทั้งหมดและตัวอักษรจัดเรียงอยู่ตามแนวนอน โดยสีตัวอักษรและฉากพื้นต้องแยกออกจากกันได้ วิธีการที่เสนอโดย W.Y. Chen และ S.Y. Chen[6] ใช้การประมวลผลหน้าปกวารสารเทคนิค ซึ่งประกอบด้วยการลดคุณภาพสีโดยลดจำนวนสีลงตามการแบ่งขอบเขตและสีพื้น ส่วนวิธีการที่ A.K. Jain และ B. Yu[19] สามารถสกัดข้อความโดยการกำหนดค่าต่างๆ เพื่อนำมาทำการประมวลผลภาพ วิธีการนี้สามารถนำไปใช้กับภาพเว็บ หน้าปกซีดีและหนังสือ หรือ ภาพจากวีดีโอ ภาพหลายค่าจะเป็นภาพไบนารี(โฆษณา) ภาพเกรย์สเกล ภาพสี (ภาพเว็บในรูปแบบ GIF) หรือ ภาพสีสรร (วีดีโอ หนังสือ หรือ ปกซีดี) ซึ่งจะถูกแบ่งย่อยไปเป็นส่วนรายละเอียดที่แสดงด้านหน้าและฉากพื้นที่ซับซ้อน ดังนั้นการวิเคราะห์ส่วนประกอบที่ใกล้ชิดกันสามารถที่จะใช้กำหนดส่วนประกอบที่เป็นข้อความได้

วิธีการของ S. Messelodi และ C.M. Modena[20] มีการใช้ประโยชน์จากตัวแทน(significant) โดยวิธีการนี้ไม่ขึ้นกับความเอียงของบรรทัดข้อความ และสามารถที่จะดำเนินการรวมตัวอักษรเข้าด้วยกันอย่างถูกต้องรวมถึงบรรทัดข้อความที่ใกล้ชิดกันด้วย วิธีการนี้ตอนแรกจะตรวจสอบส่วนประกอบที่ใกล้กันในภาพไบนารีหนึ่งหรือหลายภาพ จำนวนของภาพขึ้นอยู่กับชนิดของภาพ(ไบนารี เกรย์สเกล หรือ สี) ตัวอย่าง สำหรับภาพไบนารี ส่วนประกอบที่ใกล้ชิดกันจะถูกตรวจสอบบนสองภาพที่จะให้ผลบวกและผลลบของความเข้มของข้อความที่สอดคล้องกับฉากพื้น บรรทัดข้อความจะถูกสกัดโดยโดยการใช้แบ่งแยกความแตกต่างเป็นขั้นลำดับในการแก้ปัญหา ภาพเกรย์สเกลของปกหนังสือเป็นตัวอย่างที่แสดงการทำงานของวิธีการนี้ ดังนั้นจะเห็นว่าวิธีการนี้สามารถจัดการกับประเภทเอกสารอื่นได้ดีเช่นเดียวกัน

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

2.7.1.2 วิธีบนลงล่าง

จุดเด่นทั่วไปของวิธีการบนลงล่างจะอธิบายดังข้างล่าง ซึ่งเป็นการดำเนินงานจากส่วนประกอบใหญ่ไปเป็นส่วนประกอบย่อย อย่างไรก็ตามวิธีการประเภทนี้ยังมีแนวความคิดที่แตกต่างอย่างหลากหลาย

- เริ่มต้นจากการประมวลภาพทั้งหมดโดยการแยกไปเป็นบล็อกเล็กๆ [24], [25]
- การแปลงส่วนประกอบใหญ่โดยการใช้ภาพทั้งหมดและจุดของภาพที่เปลี่ยนแปลงถูกนำมารวมเข้าด้วยกันเป็นกลุ่ม[26]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

- การดำเนินงานสามารถเริ่มทำจากตัวแทนภาพอย่างหยาบหรือขอบเขตใหญ่ไปยังตัวแทนภาพแบบละเอียดหรือขอบเขตเล็ก[9], [27] โดยขอบเขตใหญ่จะถูกประมวลผลได้รวดเร็วในตัวแทนอย่างหยาบ และจะทำละเอียดอีกทีในตัวแทนที่ละเอียด

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

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

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

การตรวจจับพื้นที่ข้อความบนพื้นผิวของฉากพื้นที่เสนอโดย J.L. Chen[27] เป็นแนวคิดพื้นผิวที่ใช้การสกัดรายละเอียดจุดเด่นด้วยกฎหน้ากาก(Laws' masks) การแบ่งประเภทแบบหยาบๆโดยใช้บล็อกขนาด 8*8 พิกเซลและบล็อกมีการจัดเรียงที่ไม่ซ้อนทับกัน สามารถแบ่งได้เป็น ข้อความ ฉากพื้น และส่วนยังไม่ตัดสินใจ(fuzzy) (ความน่าจะเป็นขอบเขตระหว่างข้อความหรือฉากพื้น) และการแบ่งประเภทบล็อกของส่วนที่ยังไม่ตัดสินใจจะใช้การตรวจสอบที่ระดับจุดภาพ ทั้งการแบ่งแบบหยาบและแบบระเอียดใช้ตามแบบ HMMs(Hidden Markov Models) ที่ใช้ 4 ถึง 8 สถานะ(states) ประโยชน์ของ HMMs ที่เหนือกว่ากระบวนการเรียนรู้ของนิวรอนเน็ตเวิร์กอื่นๆ 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 ของแนวคิดวิธีการล่างขึ้นบน มาใช้กับภาพไบนารีเพื่อตรวจข้อมูลบรรทัดข้อความและข้อมูลที่ไม่ใช่ข้อความ(รูปภาพ) และใช้วิธีบนลงล่างคือวิธีการตัดซ้ำตามแนว X-Y (recursive X-Y cuts) โดยจะมีการแยกบรรทัดข้อความที่ไม่ติดกันไปเป็นบล็อก วิธีการนี้ง่ายต่อการนำไปใช้และเวลาที่ใช้การทำงานค่อนข้างเร็ว แต่สามารถใช้วิเคราะห์ได้ในรูปแบบที่เป็นสี่เหลี่ยมเท่านั้นและต้องแปลงภาพเกรย์สเกลให้เป็นไบนารีก่อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

49988

การคัดแปลงการแบ่งส่วนประกอบแบบแบ่งแล้วรวบรวม(split-and-merge)ของภาพเอกสารเกรย์สเกลไปเป็นขอบเขตที่มีเนื้อหาเดียวกัน เช่นการแสดงของต้นไม้สี่เหลี่ยมที่ถูกพัฒนาโดย J. Liu, Y.Y. Tang และ C.Y. Suen [29] ทำการตัดแบ่งและการรวมส่วนประกอบเข้าด้วยกันเป็นการดำเนินงานที่เวลาเดียวกัน ถ้าบางขอบเขตไม่เป็นเนื้อหาเดียวกัน โดยจะถือแบ่งเป็นสี่เหลี่ยมย่อยสี่ส่วนโดยการใช้ค่าเธอร์โชร์ของการฉายเงา ถ้าสองพื้นที่ที่ติดกัน(อาจจะเป็นส่วนประกอบระดับเดียวกันหรือไม่เป็นก็ได้) มีรายละเอียดแบบเดียวกันจะถูกนำเข้ามารวมกัน ค่าเฉลี่ยและความแปรปรวนของความเข้มข้นของจุดในแต่ละขอบเขตที่ถูกใช้ในการทำตัดสินใจเกี่ยวกับการรวมและการแบ่ง รูปแบบเอกสารอาจมีรูปแบบที่ไม่แน่นอน แต่การประมาณและแก้ไขความเอียงเป็นสิ่งจำเป็น วิธีการนี้สามารถนำไปใช้กับภาพของวารสาร เช็คธนาคารที่มีฉากพื้นที่ซับซ้อน และเอกสารที่เป็นตาราง เหมือนกับวิธีการการของ S.N. Srihri, T. Hong และ G. Srikantan [30]

วิธีการใน [31] การสกัดข้อความจากภาพสีที่ซับซ้อนของปกหนังสือและวารสารโดยใช้การวิเคราะห์แบบผสมผสาน เทคนิคของวิธีการบนลงล่างจะเป็นการตัดแบ่งภาพแบบซ้ำๆไปเป็นบล็อกสีเหลี่ยมเล็กๆ และจะหยุดการแบ่ง เมื่อจุดตัวอย่างมีสีน้อยกว่าสองสีที่แตกต่างกันภายในบล็อก บล็อกที่มีรูปข้อมูลเดียวกันถูกนำมาพิจารณาตลอดทั้งหมดของฉากพื้นชนิดเดียวกัน ขณะที่บล็อกที่ไม่มีรายละเอียดเหมือนกันมีสีที่แตกต่างกันอย่างน้อยสองสีที่สอดคล้องกับข้อความ วิธีล่างขึ้นบนใช้ตรวจหาขอบเขตที่มีรายละเอียดเหมือนกันที่มีรูปร่างไม่แน่นอนโดยใช้คุณสมบัติของการเจริญเติบโตของขอบเขต(region growing) ผลลัพธ์ของทั้งสองวิธีเมื่อนำมารวมกันเพื่อลำดับถูกต้องที่หลากหลายเพื่อให้ได้ขอบเขตข้อความและที่ไม่ใช่ข้อความโดยสมมุติฐานว่าข้อความจัดเรียงกันตามแนวนอน

2.7.2 การแบ่งประเภทขอบเขตเอกสาร

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

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

วิธีการ [38] เป็นการแบ่งประเภทโดยใช้วิธี ครอสคอรีเลชัน (cross-correlation) ระหว่างเส้นสแกนที่ใกล้กัน และ การอ้างอิงจากระยะห่างระหว่างบรรทัด การรวมจุดเด่นของรันเรงค์และครอสคอรีเลชันระหว่างจุดที่ใช้ใน [3]

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคำนวณระดับสีเทาที่ภาพเกรย์สเกลซึ่งทำให้รู้ความถึงแตกต่างระหว่าง ฉากพื้น ข้อความ กราฟิก และรูปภาพ ใน [9]

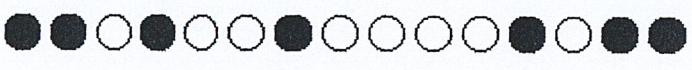
การจัดแบ่งประเภทโดยใช้วิธีการคำนวณแบบละมุนที่อธิบายใน [11], [15], [17], [39] ปกติใช้นิวรอลเน็ตเวิร์ก(ในกรณีที่มีเงื่อนไขมาก ก็จะมีขั้นตอนหลายระดับมากขึ้น) ตอนแรกจะถูกให้เรียนรู้ก่อน เริ่มจากตัวอย่างข้อความ กราฟิก และรูปภาพ ที่ต้องใช้ในการจัดแบ่งประเภท ขนาดบล็อกและจุดเด่นของรันเรงค์เป็นข้อมูลเข้าของนิวรอลเน็ตเวิร์ก[15] ขณะที่ในระดับต่ำของส่วนประกอบชุดเวฟเล็ต(wavelet packet components) ซึ่งจะถูกคำนวณจากหน้าต่างเล็กๆ ที่จุดเด่น(feature)จะถูกใช้สำหรับการแบ่งประเภทโดยนิวรอล[10], [17] เพื่อที่จะทำการแบ่งประเภทของแต่ละหน้าต่างได้อย่างถูกต้อง การรวมค่าสัดส่วนความน่าจะเป็น(fuzzy)เพื่อใช้ตัดสินใจจากข้อมูลที่ได้มาจากหน้าต่างใกล้กันที่เสร็จแล้ว การตัดสินเป็นการรวมการวัดที่มากมายของความละเอียดภาพ และที่อยู่ภายในมาตรวัด แนวคิดที่ [11] นิวรอลเน็ตเวิร์กเรียนรู้จากชุดหน้ากากเล็กๆ ซึ่งจะแยกความแตกต่างระหว่างข้อความ ฉากพื้น ภาพวาดลายเส้นและรูปได้ดีที่สุด การรวมเข้าด้วยกัน (convolving) ของหน้ากากนี้กับจุดเด่นของพื้นผิวที่ได้จากภาพนำเข้า ซึ่งถูกใช้ในการแบ่งประเภทของแต่ละจุดภาพ โดยนิวรอลเน็ตเวิร์กไปเป็นหนึ่งในสามประเภท(ข้อความ+ภาพวาดลายเส้น รูปภาพฮาฟโทนและฉากพื้น) ขอบเขตที่ถูกจัดไปเป็นประเภทแรก เป็นการทำแนวสูงต่ำกับค่ากลางซึ่งถูกกำหนด และถูกแบ่งโดยขนาดของส่วนประกอบที่เกี่ยวข้อง ซึ่งวิธีการนี้คงทนกับภาษาต่างๆ และสามารถแยกความแตกต่างของข้อความเป็นแต่ละภาษา เช่น ภาษาอังกฤษ หรือภาษาจีน วิธีใน[10], [11], [17] จะแตกต่างจากวิธีอื่นๆ เพราะการแบ่งประเภทจุด/หน้าต่าง จะทำก่อนขั้นตอนการแบ่งส่วนประกอบไปเป็นขอบเขต

2.8 การแบ่งประเภทขอบเขตเอกสาร

โดยในส่วนนี้จะอธิบายการทำงานอย่างคร่าวของบางวิธีการแบ่งขอบเขตส่วนประกอบเอกสารที่กล่าวถึงในหัวข้อ 2.7 ตามต่อไปนี้

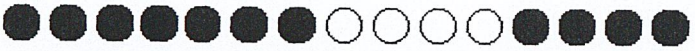
2.8.1 วิธีการรันเรงค์สมูทอัลกอริทึม (run length smoothing)

เป็นวิธีการที่ใช้การเชื่อมต่อระหว่างจุดในแนวตั้งหรือแนวนอน เพื่อเชื่อมตัวจุดเข้าด้วยกัน หรือเรียกว่าวิธีการทำเบลอ โดย กำหนดค่าขึ้นมาเพื่อเป็นค่าระยะห่าง โดยกำหนด ค่าระยะห่างของจุดดำสองจุดทั้งในแนวตั้งหรือแนวนอน โดยมีเงื่อนไขว่าจะเพิ่มจุดเข้าระหว่างจุดดำสองจุดที่มีระยะห่างกันน้อยกว่า ค่าที่กำหนด ตัวอย่างในรูปที่ 2.5 เป็นจุดภาพก่อนการทำงาน รูปที่ 2.6 เป็นชุดของจุดภาพหลังจากการทำงานรันเรงค์สมูท ด้วยค่าที่กำหนดไว้เป็น 3



รูปที่ 2.5 ชุดจุดภาพก่อนทำการรันเรงค์สมูท

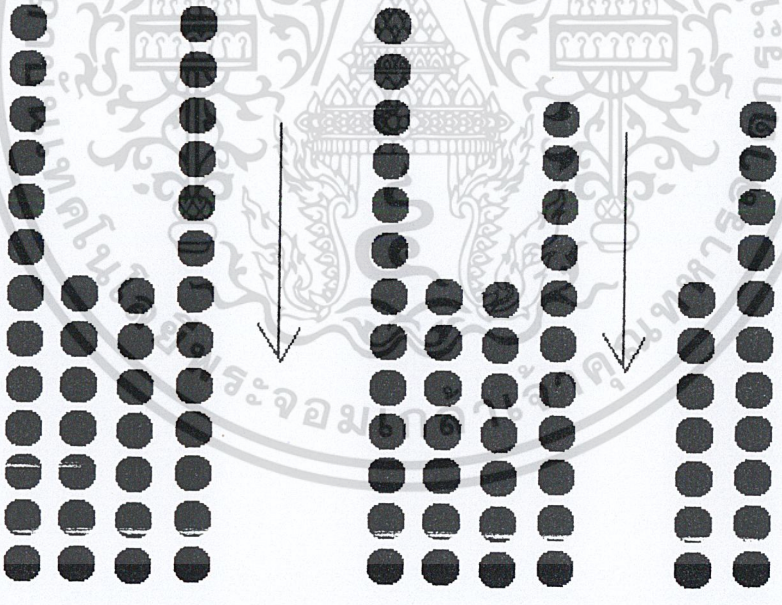
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 ชุดจุดภาพหลังจากทำการรันเรงค์สมูท ด้วยค่าที่กำหนดไว้คือ 3

2.8.2 วิธีการตัดซ้ำแกน X-Y (recursive projection X-Y cut)

วิธีการตัดซ้ำแกน X-Y มีพื้นฐานมาจากการฉายเงาภาพ ซึ่งลักษณะการทำงานจะเหมือนกับการเอาถาดที่มีลูกแก้วอยู่อย่างกระจัดกระจาย เมื่อเอียงถาดนั้นไปข้างใดข้างหนึ่ง จะพบว่าลูกแก้วจะกองอยู่ที่ข้างที่เอียงลงไป นั้นคือหลักการทำงานของการฉายภาพเงา โดยลักษณะภาพเงาของเอกสารไบนารีที่ได้ จะมีลักษณะเป็นช่วงความถี่ของจุดดำ หรือการฮิตโตแกรมจุดดำ ซึ่งในการตัดแบ่งขอบเขตนั้นจะกำหนดค่าที่ยอมรับขึ้นมา เมื่อช่วงใดที่มีค่าน้อยกว่าที่กำหนดไว้ ช่วงนั้นจะถูกตัดแบ่งออกไป โดยการทำงานแบบตัดซ้ำแกนนั้น อ้างขอบเขตในลักษณะแผนผังต้นไม้ โดยโหนดรากจะแสดงเป็นภาพเอกสารทั้งหน้า โดยโหนดลูกของแต่ละโหนดนั้นจะแสดงส่วนประกอบย่อยของโหนดแม่นั้นๆ ดังรูปที่ 2.7 แสดงการตัดตามแนวตั้งโดยมีที่ลูกศรคือตำแหน่งที่ตัดแบ่งส่วนประกอบ



รูปที่ 2.7 การฉายเงาจุดดำในแนวตั้ง และช่วงที่จะแบ่ง

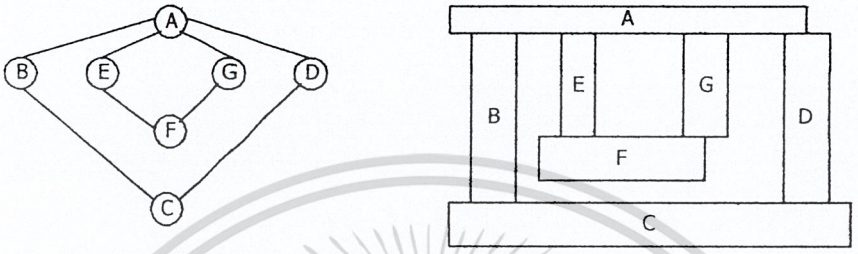
2.8.3 วิธีการเดินวนรอบขอบ (contouring)

วิธีการนี้เป็นการเดินวนรอบขอบส่วนประกอบโดยใช้จุดภาพที่ใกล้ชิดกันในทิศทาง 4 ทิศ หรือ 8 ทิศ แบบใน 2.4 รหัสแนว ซึ่งการเดินรอบขอบส่วนนี้ ทำการเดินรอบขอบส่วนประกอบจนวนกลับมาที่จุดเริ่มต้น ผลลัพธ์ที่ได้คือขอบเขตของส่วนประกอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8.4 วิธีการแบบไวท์ไทน์ (white tile)

วิธีการนี้ก็การนำส่วนที่เป็นพื้นฉากเป็นตัวกำหนดขอบเขตของส่วนประกอบภาพแบบไบนารี โดยการทำงานจะสร้างบล็อกสี่เหลี่ยมที่มีขนาดเท่ากับช่องว่างสีขาวหรือฉากพื้น โดยข้อมูลบล็อกสี่เหลี่ยมช่องขาวจะแสดงด้วยกราฟในรูปที่ 2.8 เพื่อใช้ในการบอกขอบเขตส่วนประกอบในรูปแบบปิด



รูปที่ 2.8 กราฟที่ใช้แสดงการจัดเรียงไวท์ไทน์

2.9 เปรียบเทียบวิธีการทำงาน

ตารางการเปรียบเทียบวิธีการของการทำงานแนวคิดต่างๆได้มาจากรายงานเชิงเทคนิค ของ Okun O, Doermann D & Pietikäinen M [40] ซึ่งผลลัพธ์จะมีอยู่ 2 ตาราง โดยในตารางที่ 2.1 “อ้างอิง” จะอ้างถึงวิธีการ “ชนิดเอกสาร” อ้างถึงประเภทของภาพเอกสารที่ถูกนำมาประมวลผลโดยวิธีการนั้น “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) หรือไม่ได้(N) “ชุดทดสอบ” และ “%ความถูกต้อง” อ้างถึงชุดขนาดที่นำมาทดสอบและความถูกต้องที่ได้รับจากวิธีการนั้นๆ “-” หมายความว่าจุดเด่นที่ให้มาไม่ได้ถูกอ้างอิงในบทความต้นฉบับ สัญลักษณ์ “+” แสดงถึง ผลลัพธ์ที่ไม่ถูกอธิบายโดยรูปเดียว แต่มีหลายมาตรฐานที่นำมาใช้เช่น การแบ่งเป็นส่วน(fragmentation) และอัตราการรวมเข้าด้วยกัน พวกนี้เป็นการตีมูลค่าการทำงานของวิธีการ แต่บางวิธีการง่ายหรือยากต่อการดัดแปลงแก้ไข เพราะฉะนั้นอาจจะมีการดัดแปลงซึ่งทำให้ประสิทธิภาพการทำงานเปลี่ยนแปลงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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 ภาพจากวีดีโอ	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 เช็คธนาคาร	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 ซองจดหมาย	B	U	A	Y	150	-
[34]	วารสาร	B	U	M	N	150	19.1-99.4
[35]	ปกหนังสือและปกซีดี ภาพจากวีดีโอ	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 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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		Prntium 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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.10 วิเคราะห์วิธีการทำงาน

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

ทั้งหมดต้องมีกระบวนการเฉพาะเพื่อพัฒนาวิธีการสำหรับแยกตัวอักษรจากภาพสีหรือภาพเกรย์สเกล วิธีการโดยทั่วไปจะใช้กฎที่ได้จากการเรียนรู้สำหรับการจัดแบ่งประเภทส่วนประกอบ จุดเด่นส่วนใหญ่ของวิธีการคือจำนวนเงื่อนไขที่พอที่จะทำให้เกิดความมั่นใจ เช่น

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

การแยกข้อความจากภาพสีไม่จำเป็นต้องใช้ OCR ในการตรวจสอบความถูกต้องของผลลัพธ์ที่สกัดแยกได้ ซึ่งที่มีการแบ่งขอบเขตและประเภทของส่วนประกอบก่อนเพื่อให้การทำงาน ของ OCR ได้ลดจำนวนกฎเกณฑ์ที่ใช้และเป็นการเพิ่มความถูกต้องของการทำงาน

วิธีการวิเคราะห์แผนผังกับภาพที่เป็นวารสาร หนังสือพิมพ์ หรือเอกสารที่เป็นข้อความอื่นๆ สามารถแบ่งได้เป็นสองกลุ่มคือ กลุ่มแรกต้องทำการประมาณและแก้ไขความเอียงก่อนการวิเคราะห์แผนผัง วิธีการชนิดนี้จะเป็นแผนผังชนิดสี่เหลี่ยมหรือแมนฮัตตัน เพราะการแก้ไขความเอียงก่อนจะทำให้การวิเคราะห์แผนผังได้ง่ายขึ้น อย่างไรก็ตามความผิดพลาดที่เกิดจากการหาความเอียงจะลดความถูกต้องของวิธีการวิเคราะห์แผนผัง ถ้าวิธีการนั้นไม่สามารถจัดการกับขอบเขตที่เอียงได้ ส่วนอีกประเภทหนึ่งจะจัดอยู่ในกลุ่มที่จะทำการแบ่งขอบเขตส่วนประกอบและจัดประเภทส่วนประกอบก่อน แล้วถึงมาประมาณความเอียงเฉพาะขอบเขตที่เป็นข้อความเท่านั้น ปกติวิธีการชนิดนี้จะใช้กับแผนผังที่มีความซับซ้อนและมีรูปแบบที่ไม่แน่นอน หรือภาพเอกสารที่มีหลากหลายความเอียง ในกรณีสุดท้าย การประมาณความเอียงจะถูกนำไปใช้กับบล็อกข้อความทุกบล็อก บางครั้งวิธีการนี้ต้องใช้กระบวนการดำเนินงานที่พิเศษเพราะว่าการประมาณความเอียงจะไม่ทำในส่วนที่ไม่ใช่ข้อความ อย่างไรก็ตามการวิเคราะห์แผนผังยังเป็นงานที่ยากกว่าการประมาณความเอียงและไม่ง่ายที่จะทำให้ได้ถูกต้องเหมือนการประมาณความเอียง นั้นหมายความว่าความผิดพลาดที่เกิดจากการวิเคราะห์แผนผังที่ซับซ้อนจะพบได้บ่อยมากกว่าการประมาณค่าความเอียง คือการหาผลลัพธ์หลังจากการประมาณความเอียง ตัวเลือกสุดท้ายขึ้นอยู่กับการนำไปใช้และความซับซ้อนของปัญหา

การกำหนดรายละเอียดภาพจะแตกต่างในแอพพลิเคชันที่แตกต่างกัน สำหรับตัวอย่าง ภาพรายละเอียดต่ำสำหรับภาพหน้าเว็บ[23] แต่ต้องการความละเอียดของภาพสูงสำหรับภาพวารสารหรือหนังสือพิมพ์ การประมวลผลที่ความละเอียดต่ำจะใช้เวลาในคำนวณได้รวดเร็วกว่า แต่เอกสารสารที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต้องการรายละเอียดที่แม่นยำจะทำให้เกิดความผิดพลาดและจะทำให้ขั้นตอนการทำงานต่อไปได้ผลลัพธ์ที่ไม่ดีเช่น การแบ่งส่วนประกอบที่เป็นข้อความ หรือ OCR ทางเลือกของความละเอียดที่เหมาะสมไม่ใช่งานที่ง่าย เพราะความละเอียดที่ไม่เหมาะสมจะทำให้เกิดขั้นตอนประมวลผลเอกสารมากมาย

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

ประเด็นนั้นมีเพียงสองหรือสามแอพพลิเคชันที่ใช้เทคนิคการคำนวณแบบละมุน(soft computing techniques) มาใช้ในการวิเคราะห์แผนผังเอกสาร [10], [11], [17] ในบทความ [15] อธิบายผลงานวิจัยที่ใช้นิวรอลเน็ตเวิร์กสำหรับการแบ่งประเภทส่วนประกอบเอกสารอย่างครอบคลุม อย่างไรก็ตามจำนวนของการสอนให้จดจำตัวอย่างมีขนาดได้ใหญ่มาก (มากถึง 1,000,000 ใน [11]) และยังคลุมเครือ ถึงกระนั้นตัวอย่างที่มากยังจำเป็นเพราะว่าความคลาดเคลื่อนทั้งระหว่างคลาสและระหว่างภายในคลาสมีมาก

ในบทความ [11], [27], [29] แสดงตัวอย่างประสิทธิภาพของวิธีการวิเคราะห์แผนผังเมื่อค้นงานทางแก้ไขงานที่ยากมากของการสกัดข้อความจากฉากพื้นที่มีลวดลายซับซ้อน [27], [29] และการแบ่งภาษาที่แตกต่างภายในภาพเดียวกัน [11] นักวิจัยในด้านการวิเคราะห์ภาพเอกสารยังคงไม่เอาใจใส่ในด้านนี้

การใช้ฉากพื้นมาช่วยในการแบ่งได้มากที่สุด เพราะว่าปกติแล้วจะเป็นตัวแบ่งขอบเขตระหว่างขอบเขตที่แตกต่างกัน ความละเอียดของภาพสามารถถูกลดลงถึง 75-100 จุดต่อนิ้วก่อนทำการประมวลผลเป็นความคิดที่ดีถ้ารวมการแบ่งส่วนประกอบเอกสารกับการแบ่งประเภทส่วนประกอบเอกสารเข้าไว้ด้วยกัน จะทำให้ประหยัดเวลาในการทำงานลงเพราะข้อมูลจะถูกนำไปใช้เพียงหนึ่งครั้งเท่านั้น

เวลาที่ใช้ในการประมวลผลและความถูกต้องเป็นจุดเด่นที่สำคัญของวิธีการวิเคราะห์แผนผังเอกสาร เวลาที่ใช้ในการทำงานจะมีช่วงความแตกต่างที่มากเริ่มตั้งแต่ประมาณ 1 วินาทีไปจนถึงหลายนาทีต่อหนึ่งภาพ ซึ่งจะพบว่าวิธีการที่ทำงานได้เร็วเช่น [6], [41] ซึ่งจะทำการประมาณความเอียงไปพร้อมกันและไม่ขึ้นกับแผนผังเอกสาร ใช้เวลาในการประมวลผลภาพไบนารีความละเอียด 300 จุดต่อนิ้วประมาณ 5.5 – 7 วินาที และวิธีการที่เร็วกว่าคือ [34], [38], [46] ใช้เวลาทำงานเพียง 1-2 วินาทีในงานเดียวกัน แต่ในกรณีนี้ไม่มีทั้งการแก้ไขความเอียงของเอกสารซึ่งเป็นสิ่งจำเป็นหรือแม้แต่แผนผังที่เป็นรูปแบบไม่แน่นอนก็ทำไม่ได้ การคำนวณด้วยความเร็วเป็นสิ่งที่สำคัญมากสำหรับการเรียกค้นข้อมูลข่าวสาร เช่นข้อความที่สกัดแยกจากภาพหน้าเว็บหรือจากวีดีโอ มากกว่าที่ต้องการจากการวิเคราะห์ของภาพวารสารหรือหนังสือพิมพ์ ซึ่งขึ้นอยู่กับว่าต้องการการตอบสนองทันทีหรือไม่ ในกรณีหลัง ผู้ใช้จะมีอิสระในการแก้ไขผลลัพธ์ได้นอกเหนือจากการกำหนดค่าตัวแปรใหม่ๆในกรณีที่เป็นรูปแบบ การประมวลผลที่สามารถปรับเปลี่ยนค่าใหม่ได้ มีความจำเป็นถ้าผลลัพธ์ก่อนหน้าไม่เป็นที่น่าพอใจ ดังนั้นการกำหนดตัวแปรที่เหมาะสมจะขึ้นอยู่กับรายละเอียดของแต่ละแอพพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

รูปแบบที่ถูกต้อง(Ground-truthing) และ การวัดประสิทธิภาพของวิธีการวิเคราะห์แผนผังเอกสารยังไม่มีทางออกที่สมบูรณ์ ทุกวิธีการจะยืนยันในความเอียงที่ไม่เปลี่ยนแปลง แต่เป็นการยากมากที่จะพิสูจน์ความจริงในประสิทธิภาพของวิธีทั้งหมดโดยใช้ชุดข้อมูลขนาดใหญ่ของภาพที่เอียง เพราะรูปแบบที่ถูกต้องปกติจะถูกสร้างด้วยภาพที่ตั้งตรงปราศจากความเอียง ระบบวัดประสิทธิภาพความถูกต้องจะประเมินคุณค่าจากผลลัพธ์ที่ได้จาก OCR [47] (ซึ่งตรงนี้จะไม่ชัดเจนว่าความผิดพลาดนั้นเกิดจากการแบ่งส่วนประกอบที่ไม่ดีหรือการจำจำตัวอักษรไม่ถูกต้อง) หรือ ทั้งหมดจำเป็นที่จะต้องแบ่งรูปแบบที่ถูกต้องสำหรับแต่ละภาพที่เอียง [48], [49], [50], [51] ในกรณีหลัง ภาพที่ถูกสแกนเป็นไปได้ว่าจะมีความเอียงและหลังจากที่สร้างรูปแบบที่ถูกต้องสำหรับพวกเอกสาร การทำงานแบบอัตโนมัติและการทำรูปแบบที่ถูกต้องให้ถูกต้องกินเวลาในการทำงานมาก (อาจจะ 5 – 10 นาที ต่อภาพ ที่ระดับพิกเซล เช่นที่รายงานใน [50], [51] และ ขึ้นไปถึง 5 นาทีต่อหน้าในกรณีที่ทำระดับขอบเขตบล็อกดังรายงานใน [49] ดังนั้นถ้ามีจำเป็นต้องให้คุณค่าประสิทธิภาพของวิธีการต้องใช้หนึ่งภาพที่สแกนเข้าด้วยความเอียงมุมต่างๆเริ่มตั้งแต่ 1 องศาถึง 100 องศา โดยมีการเพิ่มขึ้นทีละ 1 องศา ซึ่งจำเป็นต้องสแกนถึง 100 ครั้งในมุมที่แตกต่างกัน ในกรณีนี้การตรวจสอบรูปแบบที่ถูกต้องจะใช้เวลาหลายชั่วโมง นั่นเป็นคำตอบว่าทำไมผู้แต่งหลายคนถึงใช้วิธีเก่าในการประเมินคุณค่า โดยใช้การมองเห็นด้วยสายตามาประเมินคุณค่า อย่างไรก็ตาม ยังมีหวังกับผลลัพธ์ที่ได้จากการให้คุณค่าของอัลกอริทึมวิเคราะห์แผนผังและการสร้างรูปแบบที่ถูกต้องเริ่มปรากฏใน [48], [52] แนวคิดที่ [49] เป็นการใช้เจาะจงสำหรับการสร้างรูปแบบที่ถูกต้องเพราะยินยอมให้เป็นการทำงานแบบอัตโนมัติ แต่วิธีการนี้ยังทำได้แต่เอกสารข้อความที่ไม่มีรูปภาพหรือกราฟิก

ในการปฏิบัติ เป็นการดีหากมีการใช้อัลกอริทึมเฉพาะเจาะจงสำหรับชนิดเอกสารที่แตกต่างกันและงานที่ทำในลำดับในประสิทธิภาพการทำงานอย่างเหมาะสมที่สุด ลักษณะเด่นต่อไปนี้ควรจะถูกใช้ในกรณีต่างๆ

- ความคงทนต่อความเอียง (เอียงแบบเดียว หรือ เอียงหลายแบบ)
- ไม่ขึ้นกับรูปแบบแผนผังของเอกสาร
- สามารถสกัดข้อความได้ทั้งฉากพื้นสีขาวและฉากพื้นสีตรงข้าม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ง่ายต่อการเข้าถึงข้อมูล
- มีความเร็วและความถูกต้องสูง
- ไม่ขึ้นกับ ชนิด ขนาด และ ลักษณะของรูปแบบตัวอักษร

2.11 สรุป

การสำรวจนี้เป็นการกล่าวขั้นตอนทั่วๆไปของการวิเคราะห์แผนผังเอกสาร และอธิบายขั้นตอนการทำงานเริ่มแรกตั้งแต่ปี 1990 ในตอนแรกสรุปการวิเคราะห์งานที่จำเป็นต่อประเภทเอกสารที่หลากหลาย สามารถจัดได้เป็นสามประเภท

1. กลุ่มที่มีโครงสร้าง
2. เอกสารที่มีแผนผังโครงสร้างไม่แน่นอน
3. เอกสารที่ไม่ใช่แบบดั้งเดิม

ในกรณีเพิ่มเติมอธิบายวิธีการที่ใช้สำหรับกลุ่มนี้ อธิบายถึงจุดเด่นเช่นภาพ และ ชนิดของฉากพื้น ความละเอียดของภาพ เวลาที่ใช้ในการประมวลผล ความทนทานต่อความเอียง และ แผนผังที่แตกต่าง

ถึงอย่างไรก็ตามงานวิจัยที่จำนวนมากมายในด้านนี้ ยังไม่มีวิธีที่เป็นพื้นฐานที่จะใช้ประมวลผลภาพในประเภทเอกสารที่ต่างกัน ทั้งความถูกต้องและอัตโนมัติ จุดเด่นที่สำคัญในวิธีการที่มีคือ ความเอียงแผนผัง และ ความอิสระต่อแบบเขียน ความเร็วในการทำงาน ความถูกต้องสูง และยืดหยุ่น การแสดงตัวแทนภาพสามารถเข้าถึงข้อมูลได้ง่าย รูปแบบที่เหมาะสมของความเชื่อของ กราฟิก และรูปภาพ การสร้างรูปแบบที่ถูกต้องแบบอัตโนมัติ การวัดประสิทธิภาพของวิธีการวิเคราะห์แผนผังสำหรับภาพไบนารีและเกรย์สเกล 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

* ที่มา A. Jain, B. Yu, "Document Representation and Its Application to Page Decomposition", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, March 1998.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับ	ผู้แต่ง	ปี	วิธีการ	จุดเด่น
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 ความเอียงไม่มีผล
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
14	O'Gorman[64]	1993	k-nearest neighbor clustering	ใช้กับเอกสารที่มีโครงร่างไม่แน่นอน ให้ความถูกต้องสูง และใช้การคำนวณมาก
15	Antonacopoules 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	เร็วกว่าการฉายภาพเงาของจุด และไวต่อความเอียง
20	Sylwester et al.[69]	1995	Trainable X-Y cut	มั่นคงมาก ความเอียงและสัญญาณรบกวนไม่มีผล
21	Tang et al.[42]	1995	Modified fractal signature	ใช้กับเอกสารที่มีหลายภาษา ใช้การคำนวณมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับ	ผู้แต่ง	ปี	วิธีการ	จุดเด่น
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 lengtth smearing and adaptive thresholding	ขนาดของอักษรและช่องว่างมีผลกระทบน้อย ความเอียงไม่มีผล

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
แนวคิดและการนำมาประยุกต์ใช้ในการแบ่งส่วนประกอบภาพเอกสารไบนารี

3.1 แนวคิดในการแบ่งส่วนประกอบภาพเอกสารไบนารี

จากการแบ่งรูปแบบแผนผังโครงสร้างเอกสารในบทที่ 2 ซึ่งแบ่งแผนผังของส่วนประกอบเอกสารออกเป็นรูปแบบที่เป็นสี่เหลี่ยม รูปแบบแมนฮัตตัน และรูปแบบที่ไม่แน่นอน ดังแสดงในรูปที่ 3.1 โดยมีรูปแบบที่ยืนหยุ่นที่สุดคือ รูปร่างที่ไม่แน่นอน ซึ่งเหมาะสมในการแสดงการแยกส่วนประกอบของเอกสารต่างๆได้กว้างขวางและเหมาะสมกว่า นอกจากนี้ยังมีลักษณะที่พิเศษคือไม่จำกัดว่าขอบเขตของส่วนประกอบต้องเป็นสี่เหลี่ยมเท่านั้น จึงเป็นทางเลือกที่ต้องการในการศึกษาหาแนวทางที่จะแยกขอบเขตของส่วนประกอบที่มีรูปแบบไม่แน่นอน เพราะเอกสารไม่ได้มีรูปร่างส่วนประกอบที่เป็นชนิดสี่เหลี่ยมเท่านั้น แต่ยังมีรูปทรงเป็นวงกลม หรือรูปหลายเหลี่ยม จากรูปจะเห็นได้ว่าแผนผังรูปร่างไม่แน่นอน สามารถบอกขอบเขตของส่วนประกอบได้ดีที่สุด โดยใช้ข้อมูลในการอ้างอิงขอบเขตมากที่สุด รองลงมาเป็นแผนผังแมนฮัตตัน ส่วนสี่เหลี่ยมจะบอกขอบเขตได้หยาบที่สุด แต่ใช้ข้อมูลในการอ้างอิงขอบเขตน้อยที่สุด

(ก) (ข) (ค)

รูปที่ 3.1 ตัวอย่างแผนผังแบบต่างๆ
ก. แผนผังสี่เหลี่ยม
ข. แผนผังแมนฮัตตัน
ค. แผนผังแบบรูปร่างไม่แน่นอน

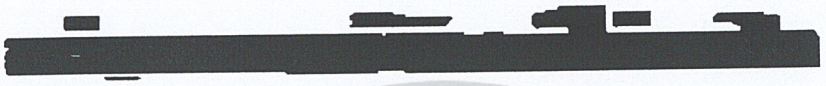
แนวคิดการทำเบลอเป็นการทำให้รายละเอียดที่อยู่ใกล้กันในระยะทางที่กำหนดไว้มาเชื่อมรวมเข้าด้วยกัน ทำให้รายละเอียดนั้นมีขนาดใหญ่ขึ้น วิธีการนี้เป็นการทำงานที่ดีวิธีหนึ่งเพื่อที่จะรวมส่วนประกอบย่อยที่อยู่ในระยะทางที่กำหนดเข้าด้วยกันเป็นส่วนประกอบที่ใหญ่ขึ้น เป็นแนวทางหนึ่งที่จะรวบรวมรายละเอียดของส่วนประกอบที่อยู่ใกล้กัน เช่น ตัวอักษร ดังนั้นจึงมีการสร้างหน้าต่างขึ้นมาเพื่อที่จะเชื่อมตัวอักษรที่อยู่ใกล้กันในระยะช่องหน้าต่างเชื่อมเข้าด้วยกันเป็นส่วนประกอบเดียวกัน ซึ่งวิธีการนี้ก็จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

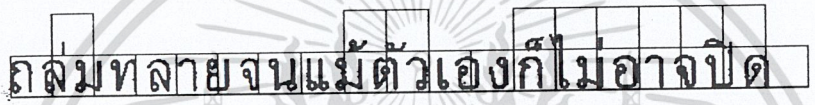
คล้ายกับวิธีการทำเบลอ เปลี่ยนการทำงานด้วยจุดไปเป็นหน้าต่างแทนซึ่งทำให้ลดเวลาในการเชื่อมส่วนประกอบเข้าด้วยกันลง ดังแสดงในรูปที่ 3.2

ถล่มทลายจนแม้ตัวเองก็ไม่อาจปิด

(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 เส้นกั้น โดยส่วนมากส่วนประกอบของเอกสารที่มีรายละเอียดแตกต่างกันจะมีระยะห่างจากกันพอสมควร เพื่อความสวยงาม หรือ มีเส้นกั้นไว้เพื่อบอกขอบเขตที่สิ้นสุดของรายละเอียด ในการที่จะแบ่งขอบเขตส่วนประกอบของแต่ละส่วน หลังจากที่พิจารณาแล้ว ได้แนวคิดว่าการทำหน้าต่างเดินวนรอบขอบส่วนประกอบและเก็บข้อมูลในลักษณะแผนผังแบบแมนฮัตตัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเดินรอบขอบเขตโดยใช้หน้าต่างนี้จะได้ขอบเขตของส่วนประกอบใหญ่ซึ่งรู้ตำแหน่งและขนาด แต่ยังไม่สามารถบอกได้ว่าส่วนประกอบที่แบ่งขอบเขตมานั้นเป็นส่วนข้อความหรือรูปภาพ ดังนั้นจึงต้องมี การเดินวนรอบส่วนรายละเอียดในส่วนประกอบใหญ่ที่ได้จากการแบ่งขอบเขตจากหน้าต่าง ซึ่งในขั้นตอนนี้สามารถบอกได้พอสมควรว่าส่วนประกอบใดน่าจะเป็นข้อความ ส่วนประกอบใดน่าจะเป็นรูปภาพ เมื่อเดินวนรอบส่วนรายละเอียดในส่วนประกอบใหญ่ และตัดส่วนประกอบที่เป็นรูปภาพ และเป็นสิ่งรบกวนออกไปพร้อมๆกันแล้ว สุดท้ายก็จะได้เฉพาะส่วนของตัวอักษรในแต่ละตัวเท่านั้น

3.2 การแบ่งส่วนประกอบภาพเอกสารเป็นส่วนประกอบใหญ่

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

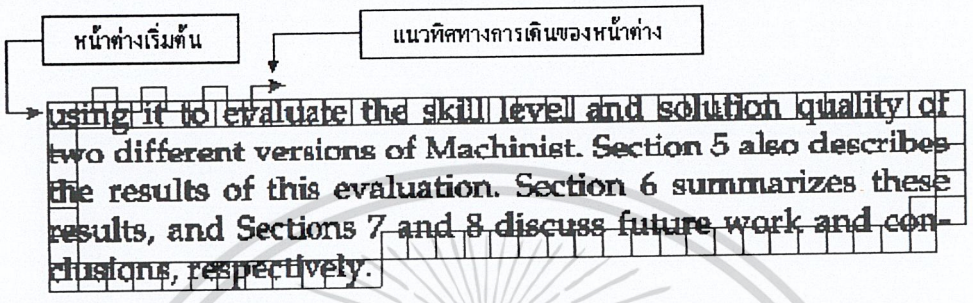
(ก) (ข)

รูปที่ 3.3 ขอบเขตภาพ (ก) ภาพแสดงเมื่อมองจากระยะไกล (ข) ภาพที่แบ่งขอบเขตด้วยสายตา

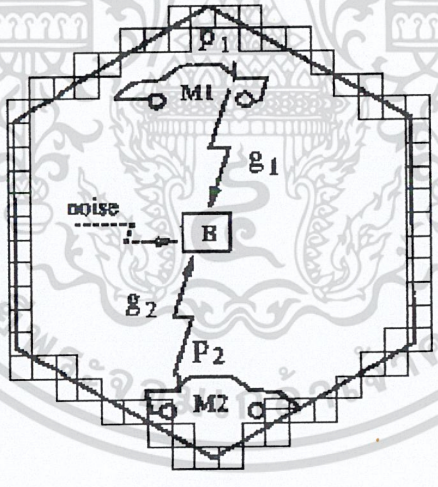
ในหลักการเดียวกัน แนวคิดการวางหน้าต่างจะวางที่ตำแหน่งซ้ายบนสุดของรายละเอียดของส่วนประกอบ แล้ววางหน้าต่างถัดไปเรียงตามขอบของส่วนประกอบ ขอบเขตที่ได้จะเป็นขอบเขตส่วนประกอบใหญ่ โดยจะนำขอบเขตดังกล่าวจากขั้นตอนนี้ไปใช้ในการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่ต่อไป สำหรับการกำหนดขนาดของหน้าต่างนั้น จะกำหนดได้ต้องได้ทำการทดลองก่อน ซึ่งในตอนต้น จะใช้ขนาด 32*32 ซึ่งขนาดของหน้าต่างๆจะใกล้เคียงกับขนาดตัวอักษรปกติที่ใช้ในเอกสารทั่วๆไป ขนาดของส่วนประกอบนี้สามารถที่จะเชื่อมตัวอักษรที่อยู่ใกล้ชิดกันเข้าด้วยกันรวมเป็นกลุ่ม และมีขนาดเล็กพอที่จะแบ่งส่วนประกอบที่มีระยะห่างกันออกจากกันได้ โดยการวางหน้าต่างนี้จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วางเรียงต่อกันไปตามรอบขอบส่วนประกอบ ดังนั้นขอบเขตส่วนประกอบที่ได้คือขอบเขตของหน้าต่างที่เรียงต่อกันจากจุดเริ่มต้นที่ของหน้าต่างแรกจนวนกลับมาถึงหน้าต่างเริ่มต้น ดังนั้นขอบเขตที่ได้จึงเป็นขอบเขตส่วนประกอบขนาดใหญ่ที่มีรายละเอียดต่างๆ มาอยู่รวมกัน ดังรูปที่ 3.4 แสดงรูปแบบการเดินรอบส่วนประกอบของหน้าต่างและทิศทางที่ใช้ในการเดินรอบส่วนประกอบใหญ่ และรูปที่ 3.5 แสดงลักษณะส่วนประกอบที่ได้จากหน้าต่างที่เดินรอบขอบที่มีลักษณะแผนผังไม่อยู่ในรูปแบบสี่เหลี่ยม



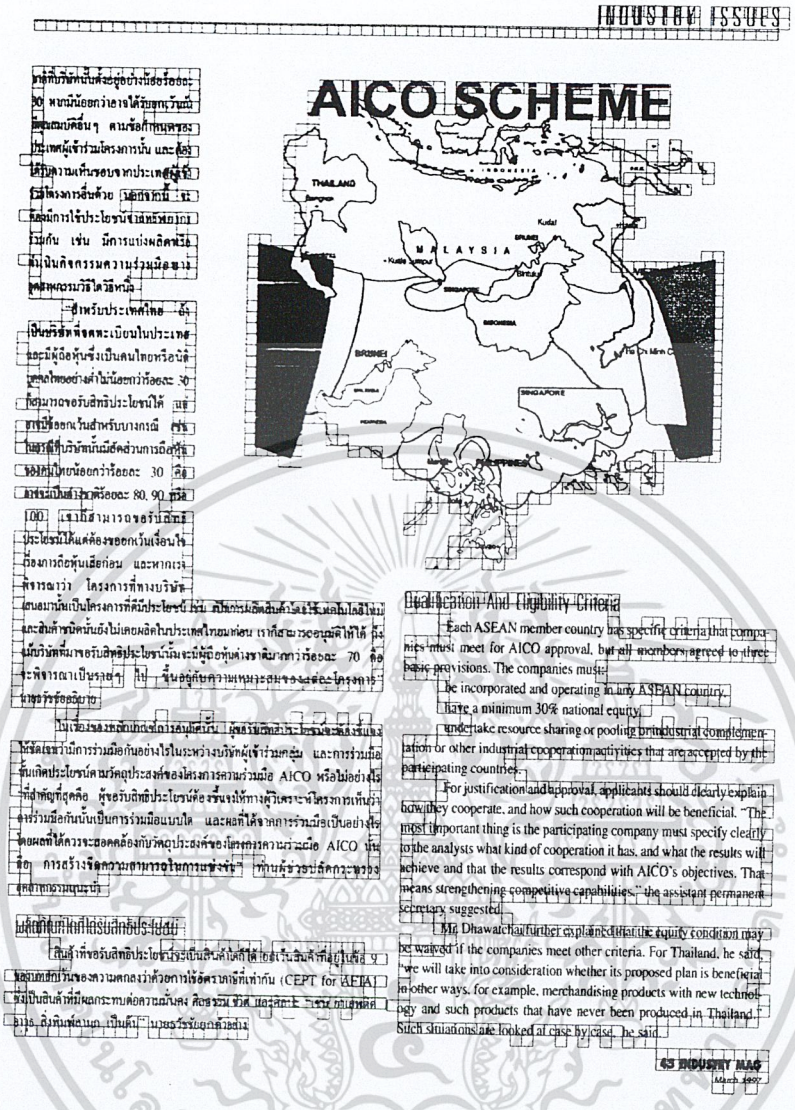
รูปที่ 3.4 การวางตำแหน่งหน้าต่างเริ่มต้นและลักษณะการวนรอบส่วนประกอบใหญ่



รูปที่ 3.5 ลักษณะภาพส่วนประกอบที่ไม่อยู่ในลักษณะสี่เหลี่ยม

จากรูปที่ 3.6 แสดงขอบเขตของส่วนประกอบหลังจากทำการวางหน้าต่างขนาด 32*32 พิกเซล รอบขอบเขต ซึ่งการวางหน้าต่างนี้สามารถแบ่งขอบเขตแต่ละของแต่ส่วนประกอบได้ ถ้าส่วนประกอบเหล่านั้นมีระยะที่ห่างกันเกินกว่าหน้าต่างจะสามารถเชื่อมกันได้ และหากส่วนประกอบอยู่ใกล้กันจนทำให้หน้าต่างมาเชื่อมต่อกัน ซึ่งบางที่ขอบเขตส่วนประกอบใหญ่ที่ได้จะมีทั้งข้อความและภาพรวมกัน หรือมีคอลัมน์ข้อความเกินกว่าหนึ่งคอลัมน์ก็ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



*รูปที่ 3.6 ขอบเขตของส่วนประกอบในภาพเอกสารที่แบ่งโดยหน้าต่างขนาด 32*32 พิกเซล*

3.3 การแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่

หลังจากที่ได้ขอบเขตรูปร่างของส่วนประกอบใหญ่แล้ว ยังไม่สามารถที่จะนำมาจัดแบ่งประเภทส่วนประกอบว่าเป็นชนิดข้อความหรือรูปภาพ ดังนั้นต้องมีการแบ่งส่วนประกอบย่อยในระดับรายละเอียดเพื่อแบ่งแยกส่วนต่างๆที่อยู่ในส่วนประกอบใหญ่ โดยจะมีการเดินขอบด้วยหน้าต่างขนาด 1 พิกเซลแทนหน้าต่าง 32*32 พิกเซล เพื่อแยกส่วนรายละเอียดออกมา โดยขอบเขตของรายละเอียดจะถูกเก็บเป็นบล็อกสี่เหลี่ยม(มุมบนซ้าย, มุมล่างขวา) การทำงานเป็นดังนี้

1. หาจุดเริ่มต้นของตัวอักษร
2. เดินวนรอบตัวอักษร จะได้ขอบเขตเสมือนเป็นบล็อกสี่เหลี่ยมครอบตัวอักษรใดๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ถ้าความกว้างและความสูงของขอบเขตส่วนประกอบมีขนาดน้อยกว่าหรือเท่ากับ 5 พิกเซล จะถูกกำหนดให้เป็นสัญญาณรบกวน ถ้าความสูงหรือความกว้างของขอบเขตส่วนประกอบมีขนาดมากกว่าหรือเท่ากับ 100 พิกเซล จะถูกกำหนดให้เป็นภาพ ส่วนขอบเขตที่ไม่อยู่ในช่วงที่กำหนดข้างต้น จะถูกกำหนดให้เป็นตัวอักษร
4. ซึ่งหลังจากที่ได้ขอบเขตที่เป็นตัวอักษรแล้ว นำทำการตัดส่วนประกอบนั้นทิ้งไปจากส่วนประกอบใหญ่
5. เดินต่อและข้ามส่วนที่ถูกตัด วนตามข้อ 1 – 5 จนสุดขอบเขตส่วนประกอบใหญ่

เมื่อจบจากขั้นตอนดังกล่าวแล้ว จะได้ขอบเขตของตัวอักษรเรียงต่อกัน จากบนลงล่าง และจากซ้ายไป

3.4 การแบ่งย่อหน้าของเอกสารในส่วนประกอบใหญ่

จากการทดลอง เห็นว่า เมื่อขึ้นย่อหน้าใหม่ กรอบการเดินของส่วนประกอบใหญ่จะมีลักษณะพิเศษดังนี้

1. บล็อกแรกของแถวของบล็อกในส่วนของย่อหน้าจะมีตำแหน่งไปทางขวาพอประมาณเมื่อเทียบกับซ้ายสุดของส่วนประกอบใหญ่
2. จะมีบล็อกอยู่ในส่วนบนสุดของย่อหน้าพอสมควร

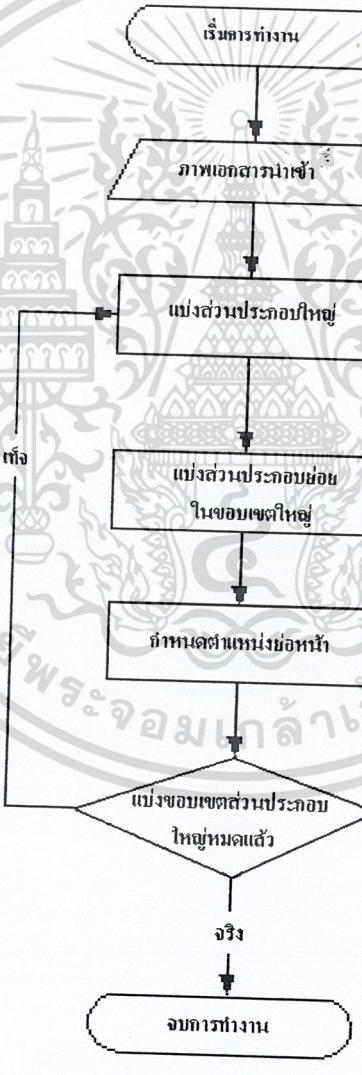
จากข้างต้น ได้ผลลัพธ์การกำหนดขอบเขตออกเป็นที่น่าพอใจ จากเทคนิคเล็กๆน้อยๆ เมื่อรู้ขอบเขตคล่าวๆ แล้ว จำเป็นต้องรูปบรรทัดของตัวอักษร แทนการรู้บรรทัดของบล็อกที่เยื้องเข้าไป ซึ่งจะสามารถรู้บรรทัดได้จากขนาดของขอบเขตของตัวอักษร และจากการแปลจากระบบรู้จำตัวอักษร ซึ่งไม่ได้อยู่ในขอบเขตของปริญญานิพนธ์นี้ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
อัลกอริทึมและขั้นตอนการทำงาน

จากแนวคิดที่ผ่านมาในบทที่ 3 เรานำมาทำการประยุกต์สร้างโปรแกรมแบ่งเอกสารให้มีประสิทธิภาพ โดยจะแบ่งการทำงานออกเป็น 3 ขั้นตอนดังนี้ และแสดงในแผนผังการดำเนินงานรูปที่ 4.1

1. การแบ่งส่วนประกอบใหญ่
2. การแบ่งส่วนประกอบย่อยภายในส่วนประกอบใหญ่
3. การแบ่งคอลัมน์ภายในขอบเขตส่วนประกอบใหญ่



รูปที่ 4.1 แผนผังขั้นตอนการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1 วิธีการแบ่งส่วนประกอบใหญ่

การแบ่งส่วนประกอบใหญ่ จุดประสงค์เพื่อให้รู้รูปร่างข้อมูลโดยรวม เพื่อการตัดส่วนที่เป็นสีขาวของเอกสารออก และสนใจเฉพาะตัวเนื้อข้อมูลเท่านั้น ดังนั้นการค้นหาส่วนของเนื้อเอกสารหลักจะกำหนดหน้าต่าง ไว้ในขั้นตอนให้มีขนาดเป็นรูปสี่เหลี่ยมจัตุรัส 32 จุด * 32 จุด ทั้งนี้เพื่อไม่ให้เดินวนรอบส่วนแบ่งระหว่างบรรทัดได้ ในการเดินหน้าต่างนั้นจะต้องมีโครงสร้างของข้อมูลว่าหน้าต่างนั้นควรอยู่ที่ใด โดยการโครงสร้างของข้อมูลไว้ดังนี้

โครงสร้างของข้อมูลของ หน้าต่าง 1 หน้าต่าง

```
class Block
{
public:
        long x;
        long y;
        int Checked;
        int count;
        Block();
        virtual ~Block();
};
```

ประกาศตัวแปรเพื่อเก็บตำแหน่งหน้าต่างไว้ดังนี้

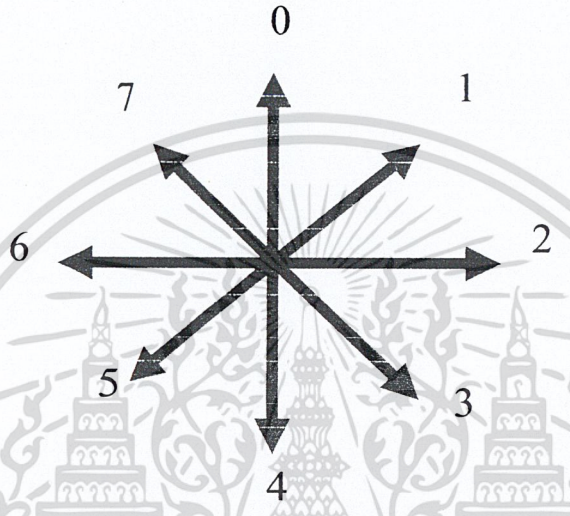
```
Block1 = new Block[(PicHeight/BlockHeight)*(PicWidth/BlockWidth)-1];
```

ให้ตัวแปร Block แต่ละตัว เก็บตำแหน่งในเอกสารไว้ ซึ่งเราจะเดินหน้าต่างไปตาม Block เหล่านี้

```
long x,y;
for(y=0;y<PicHeight/BlockHeight;y++)
{
        for(x=0;x<PicWidth/BlockWidth;x++)
        {
                Block1[y*(PicWidth/BlockWidth)+x].x = x*BlockWidth;
                Block1[y*(PicWidth/BlockWidth)+x].y = y*BlockHeight;
                Block1[y*(PicWidth/BlockWidth)+x].Checked = false;
                Block1[y*(PicWidth/BlockWidth)+x].count = -1;
        }
}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสร็จแล้วเราก็เริ่มเดินหน้าต่างตั้งแต่หน้าต่างแรกเรื่อยไป จนเจอหน้าต่างที่มีจุดดำภายในหน้าต่างนั้นมากกว่า 1% จะถือว่าเป็นหน้าต่างดำ และจำทำการกำหนดเป็นหน้าต่างเริ่มต้นในการเดินว่ารอบบริเวณตัวหนังสือ การเดินรอบส่วนประกอบใหญ่ จะวนตามเข็มนาฬิกาเพื่อตรวจเช็คว่าหน้าต่างรอบตัว หน้าต่างใดเป็นหน้าต่างดำต่อไป และการเช็คหน้าต่างรอบตัวเพื่อจะเดินต่อไปนั้น จะต้องตรวจสอบก่อนว่าหน้าต่างนั้นถูกตรวจสอบไปก่อนหน้าแล้วหรือยัง เพื่อที่จะลดเวลาที่ศูนย์เสีย วิธีการเดินรอบหน้าต่างจะใช้การเดินในลักษณะ 8 ทิศรอบ โดยให้ทิศแต่ละค่าดังต่อไปนี้



0 คือ บน	1 คือ บนขวา	2 คือ ขวา	3 คือ ล่างขวา
4 คือ ล่าง	5 คือ ล่างซ้าย	6 คือ ซ้าย	7 คือ บนซ้าย

โดยขั้นตอนการทำงานทิศทางการเดินของหน้าต่างเป็นดังนี้

กรณีหน้าต่างที่ถูกตรวจสอบ เป็นหน้าต่างดำ

กรณีทิศทางเท่ากับ 0 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 5

กรณีทิศทางเท่ากับ 1 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 6

กรณีทิศทางเท่ากับ 2 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 7

กรณีทิศทางเท่ากับ 3 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 0

กรณีทิศทางเท่ากับ 4 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 1

กรณีทิศทางเท่ากับ 5 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 2

กรณีทิศทางเท่ากับ 6 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 3

กรณีทิศทางเท่ากับ 7 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 4

กรณีทิศทางเท่ากับ 8 : เดินไปยังหน้าต่างดำ และเปลี่ยนทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 5

กรณีหน้าต่าง เป็นหน้าต่างขาว

กรณีทิศทางเท่ากับ 0 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 1

กรณีทิศทางเท่ากับ 1 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 2

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีทิศทางเท่ากับ 2 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 3

กรณีทิศทางเท่ากับ 3 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 4

กรณีทิศทางเท่ากับ 4 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 5

กรณีทิศทางเท่ากับ 5 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 6

กรณีทิศทางเท่ากับ 6 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 7

กรณีทิศทางเท่ากับ 7 : ทิศทางที่จะตรวจสอบต่อไปจะเท่ากับ 0

เมื่อพบหน้าต่างดำจะย้ายตำแหน่งของหน้าต่างไปเรื่อยๆ จนวนกลับมาพบหน้าต่างแรกจึงจะหยุด การเดินหน้าต่างนี้ จะต้องเก็บโครงสร้างข้อมูลไว้ดังนี้

```
class LinkList
{
public:
        char direction;
        long BlockPositionX;
        long BlockPositionY;
        LinkList* NextLinkList;
        LinkList();
        virtual ~LinkList();
};
```

โครงสร้างการวนรอบ 1 รอบ จะได้ Cparagraph 1 instance

```
class CParagraph
{
public:
        LinkListCharLine* HeadCharLine;
        LinkListPointer* LineList;
        int Right;
        int Left;
        int Bottom;
        int Top;
        long NumberofBlockround;
        LinkList* HeadPosition;
        LinkList* HeadOrder;
        LinkListChar* CharHead;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        LinkListChar* PictureHead;
        LinkListSubParagraph* HeadSubParagraph;
        int* Histogram;
        CParagraph();
        virtual ~CParagraph();
};
```

ตัวแปร HeadPosition จะเก็บตำแหน่งหน้าต่างรอบส่วนประกอบใหญ่เรียงตามการวนรอบ

ตัวแปร HeadOrder จะเก็บตำแหน่งหน้าต่างรอบส่วนประกอบใหญ่เรียงตาม BlockPositionX จากน้อยไปมาก

ตัวแปร Left , Right , Top, Bottom จะเก็บตำแหน่งสุดซ้าย ขวา บน ล่าง ของส่วนประกอบใหญ่

ระหว่างการเดินจะระบุทิศทางในตัวแปร LinkList : direction ตามการเดินรอบดังนี้

หน้าต่างเริ่มต้นจะมีทิศทางคือ L

กรณีทิศทางเท่ากับ 0 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น L

กรณีทิศทางเท่ากับ 1 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น L

กรณีทิศทางเท่ากับ 2 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น T

กรณีทิศทางเท่ากับ 3 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น L

ถ้าทิศก่อนหน้าไม่ใช่ L จะเป็นทิศทางก่อนหน้าเป็น R

กรณีทิศทางเท่ากับ 4 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น R

กรณีทิศทางเท่ากับ 5 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น R

ถ้าทิศก่อนหน้าไม่ใช่ L จะเป็นทิศทางก่อนหน้าเป็น R

กรณีทิศทางเท่ากับ 6 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น B

กรณีทิศทางเท่ากับ 7 ทิศทางของหน้าต่างถัดไปนั้นจะเป็น L

เมื่อเสร็จสิ้นการวนรอบ จะต้องทำการทำสัญลักษณ์กับข้อความภายในส่วนประกอบใหญ่ เพื่อไม่ให้ทำการเดินซ้ำอีก โดยการนำค่าในตัวแปร HeadOrder มาทำการเรียง Link List ต่อกันใหม่ได้ดังนี้

```
PtLinkList = Paragraph[ParagraphIndex].HeadOrder;
do
{
        BlockCopy = new LinkList;
        BlockCopy->BlockPositionX = PtLinkList->BlockPositionX;
        BlockCopy->BlockPositionY = PtLinkList->BlockPositionY;
        BlockCopy->direction = PtLinkList->direction;
        BlockCopy->NextLinkList = NULL;
        if (Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].Head == NULL)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        }
                Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].Head          =
                BlockCopy;
                Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].Tail          =
                BlockCopy;
        }
        else
        {
                Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].Tail-
        >NextLinkList = BlockCopy;
                Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].Tail          =
                BlockCopy;
        }
        Paragraph[ParagraphIndex].LineList[BlockCopy->BlockPositionY].ListSize++;
        PtLinkList = PtLinkList->NextLinkList;
}while(PtLinkList != NULL);
```

เมื่อได้ List ที่เป็นแถวแล้วก็จะนำมาระบุ ลงในตัวแปร Block1[x].Checked ว่าหน้าต่างที่ x ได้ถูกตรวจสอบไปแล้วดังนี้

```
i=0;
while(i<Paragraph[ParagraphIndex].LineList[j].ListSize)
{
        if (PtLineList->direction == 'L')
        {
                if (PtLineListOld != NULL)
                {
                for(m=PtLineListOld->BlockPositionX;m<=PtLineList->BlockPositionX;m++)
                {
                        Block1[j*(PicWidth/BlockWidth)+m].Checked = true;
                        BlocktoBitSetParagraph(j*(PicWidth/BlockWidth)+m,ParagraphIndex);
                }
                }
                else
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

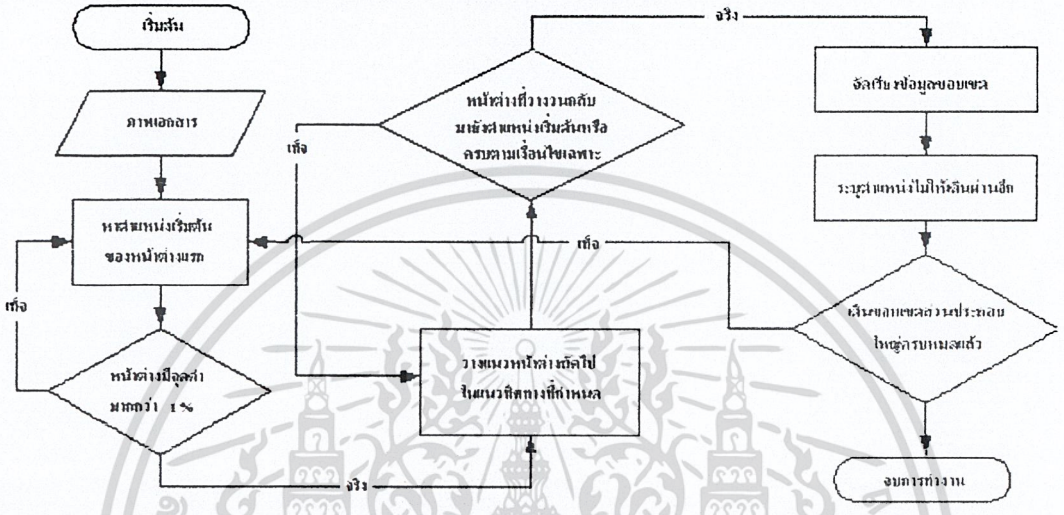
```
            {
                    // Mark Paragraph Bit not depend on PtLineListOld
                    m = PtLineList->BlockPositionX;
                    BlocktoBitSetParagraph(j*(PicWidth/BlockWidth)+m,ParagraphIndex);
            }
            PtLineListOld = PtLineList;
    }
    else if (PtLineList->direction == 'R')
    {
            if (PtLineListOld != NULL)
            {
            for(m=PtLineListOld->BlockPositionX;m<=PtLineList->BlockPositionX;m++)
            {
                    Block1[j*(PicWidth/BlockWidth)+m].Checked = true;
                    BlocktoBitSetParagraph(j*(PicWidth/BlockWidth)+m,ParagraphIndex);
            }
            PtLineListOld = NULL;
            }
            else
            {
                    // Mark Paragraph Bit not depend on PtLineListOld
                    m = PtLineList->BlockPositionX;
                    BlocktoBitSetParagraph(j*(PicWidth/BlockWidth)+m,ParagraphIndex);
            }
    }
    else
    {
            // Others Except direction L or R
            m=PtLineList->BlockPositionX;
            BlocktoBitSetParagraph(j*(PicWidth/BlockWidth)+m,ParagraphIndex);
    }
    PtLineList=PtLineList->NextLinkList;
    i++;
    }
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อระบุขอบเขตของส่วนประกอบใหญ่เรียบร้อยแล้ว ก็จะค้นหาส่วนประกอบใหญ่ต่อไปอีก และกระทำเหมือนกันดังข้างต้น จนสิ้นสุดเอกสาร

สรุปเป็นแผนผังการทำการแบ่งส่วนประกอบใหญ่ดังนี้

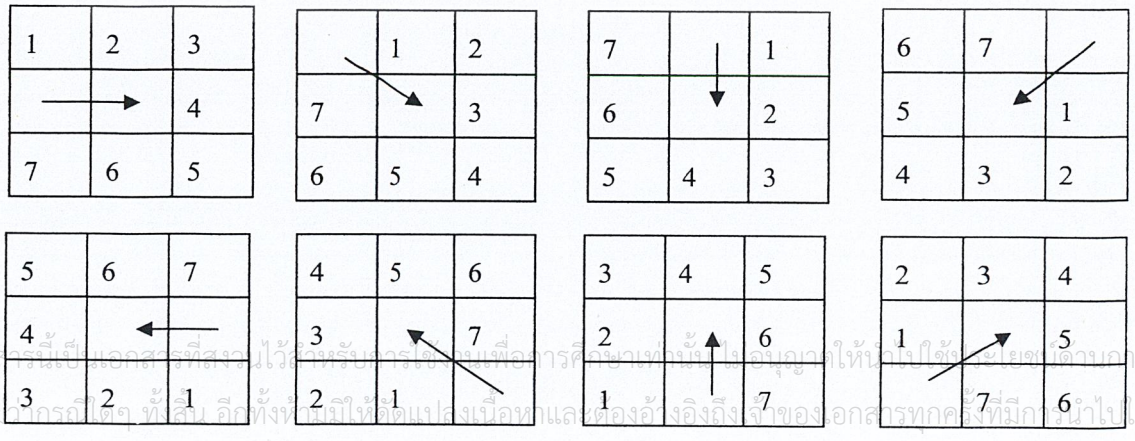
แผนผังการทำงานแบ่งส่วนประกอบใหญ่



รูปที่ 4.2 แผนผังการขั้นตอนการทำงานของการแบ่งส่วนประกอบใหญ่

4.2 วิธีการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่

เมื่อสิ้นสุดการหาขอบเขตของส่วนประกอบใหญ่แล้ว จะเริ่มต้นการหาขอบเขตส่วนประกอบย่อยเพื่อค้นหาตัวอักษรแต่ละตัวอักษร การหาขอบเขตส่วนประกอบกอบย่อยนั้น วิธีการเดินรอบตัวอักษรจะคล้ายกับวิธีการแบ่งส่วนประกอบใหญ่ แต่จะเดินด้วยจุด แทนการเดินด้วยหน้าต่าง



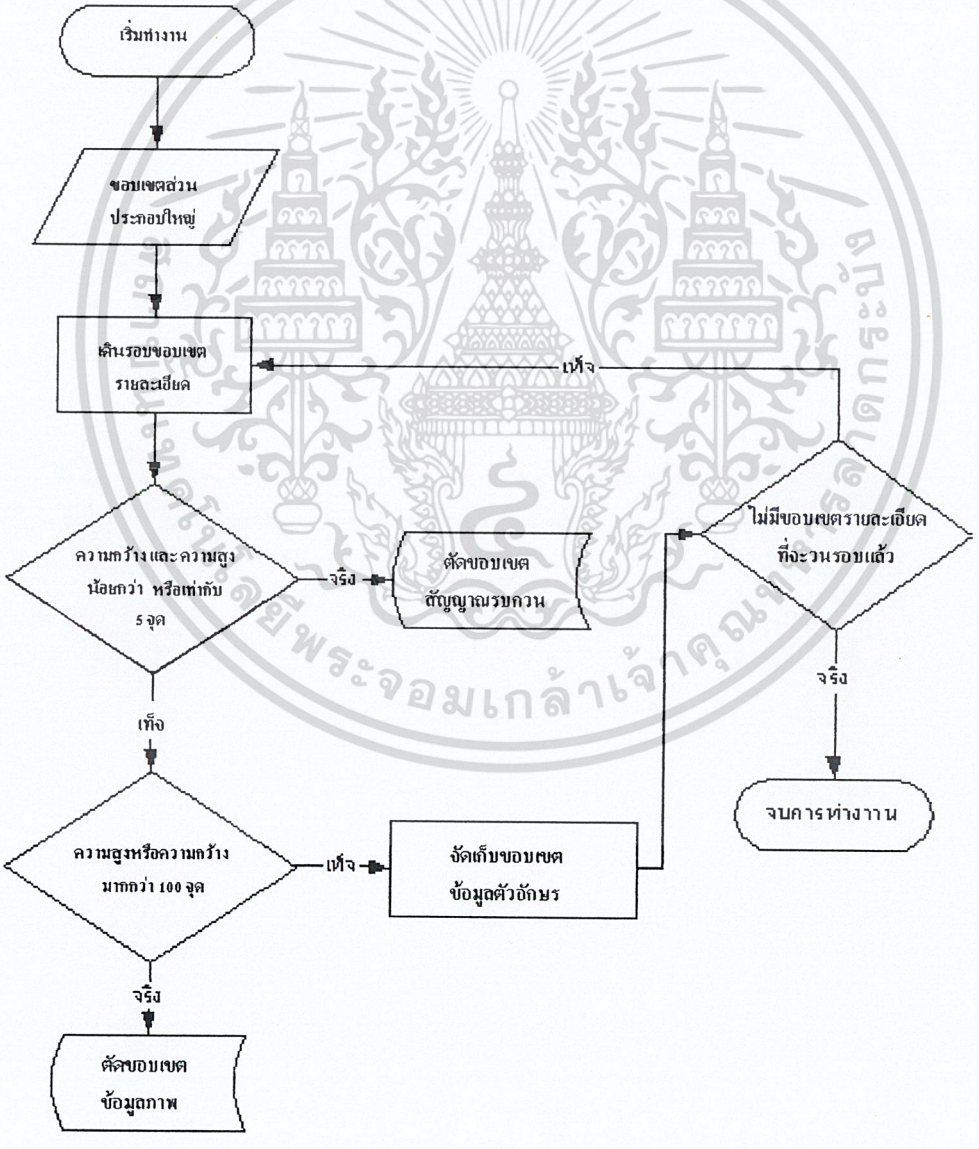
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป หัวลูกศรและปลายลูกศรแสดงลักษณะการเดินไปในทางทิศใด ซึ่งลักษณะการเดินจะมีผลต่อการค้นหาจุดหรือหน้าต่างดำต่อไป โดยจะหาจุดหรือหน้าต่างจากตำแหน่งที่ 1 เรื่อยไปจนถึง 7 หรือเมื่อเจอจุดหรือหน้าต่างดำก่อนที่จะครบ 7 ก็จะย้ายจุดหรือหน้าต่างไปยังตำแหน่งต่อไปทันที

ระหว่างการเดินรอบตัวอักษรจะทำการจัดการเก็บพิกัดซ้ายสุด ขวาสุด บนสุด และนำมาใช้ในการแยกแยะประเภทของภาพได้ดังนี้

1. ขอบเขตที่ความสูงและ ความกว้างน้อยกว่า 5 พิกเซล ถือว่าเป็นสัญญารบกวนจะไม่นำมาพิจารณา
2. ขอบเขตที่มีความสูงหรือความกว้างมากกว่า 100 พิกเซลขึ้นไปจะกำหนดให้เป็นรูปภาพ
3. ที่เหลือนอกจากข้อ 1 และข้อ 2 จะถือว่าเป็นตัวอักษร

การเก็บขอบเขตจะแบ่งเก็บ 2 ประเภท คือ ประเภทตัวอักษร และประเภทรูปภาพ



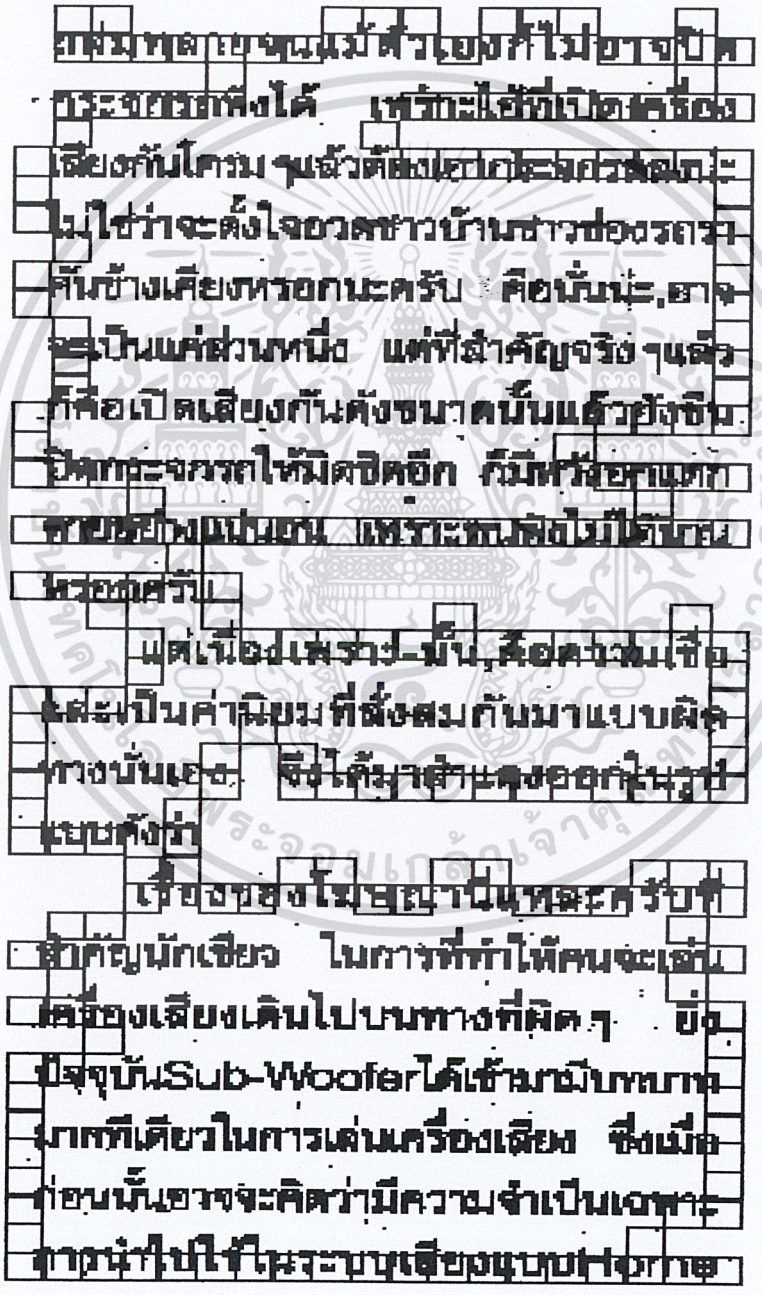
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.3 แผนผังการทำงานการแบ่งส่วนประกอบย่อยในส่วนประกอบใหญ่

4.3 วิธีการแบ่งคอลัมน์ภายในขอบเขตส่วนประกอบใหญ่

การแบ่งคอลัมน์ในส่วนประกอบใหญ่มีการทำงานดังนี้

ตรวจสอบหน้าต่างที่เดินวนรอบส่วนประกอบใหญ่ในแต่ละบรรทัดของหน้าต่างว่าหน้าต่างใดอยู่หน้าสุด ถ้าหน้าต่างนั้นอยู่ห่างจากขอบซ้ายสุดของส่วนประกอบใหญ่ ไปมากกว่า 3 หน้าต่าง และมีหน้าต่างอยู่ในบรรทัดหน้าต่างนั้น มากกว่า 4 จะถือว่าบรรทัดของบล็อกนั้น มีส่วนประกอบเป็น ย่อหน้า ทั้งนี้ได้ข้อมูลมาจากการทดลอง ดังรูปที่ 4.4



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.4 ตัวอย่างการวางรูปแบบของหน้าต่างในย่อหน้าต่าง ๆ

บทที่ 5
ผลการทดลอง

ในบทนี้กล่าวถึงผลการทดลองจากที่ได้ทำการทดสอบกับตัวอย่างเอกสารต่างๆ 12 ตัวอย่าง ถูกนำมาสแกนเป็นภาพขาวดำ ด้วยความละเอียด 300 จุดต่อนิ้ว

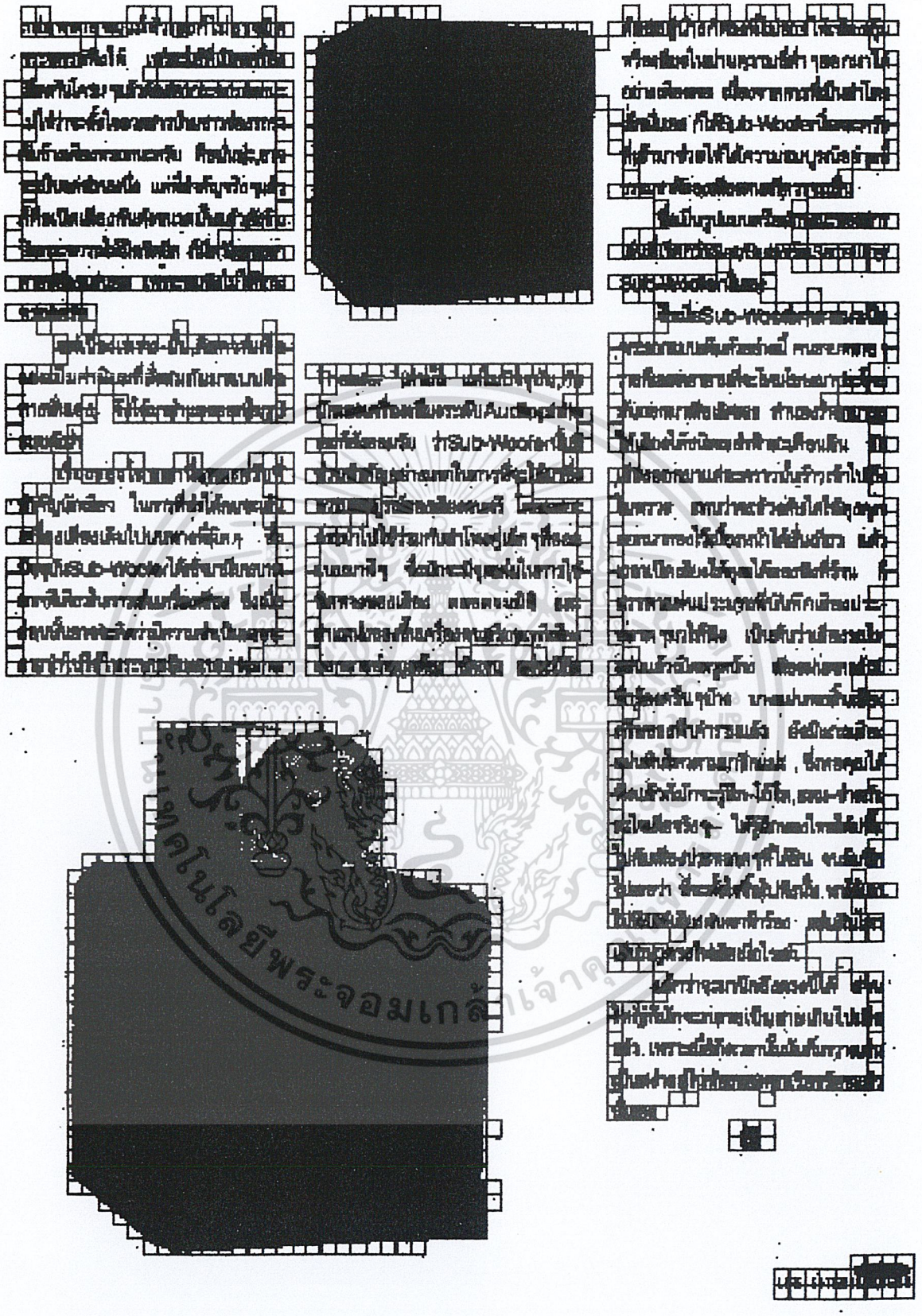
โครงงานจะทดสอบถูกต้องในการแบ่งส่วนประกอบจากสายตาของผู้ทดลอง ในการทำงานแบ่งเป็น 2 ช่วง คือ ช่วงแรกการแบ่งขอบเขตส่วนประกอบใหญ่ ช่วงที่สองการแบ่งส่วนประกอบย่อย รวมด้วยการวิเคราะห์แบ่งขอบเขตส่วนประกอบและการแบ่งประเภทส่วนประกอบ การทดลองจะแสดงเป็นข้อย่อยตามลำดับต่อไป

5.1 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่หนึ่ง

ผลลัพธ์ที่ได้จากการทดลองช่วงที่หนึ่ง คือ ขั้นตอนทำงานแบ่งส่วนประกอบใหญ่ เป็นการเดินรอบขอบเขตส่วนประกอบโดยใช้หน้าต่างขนาดใดๆ

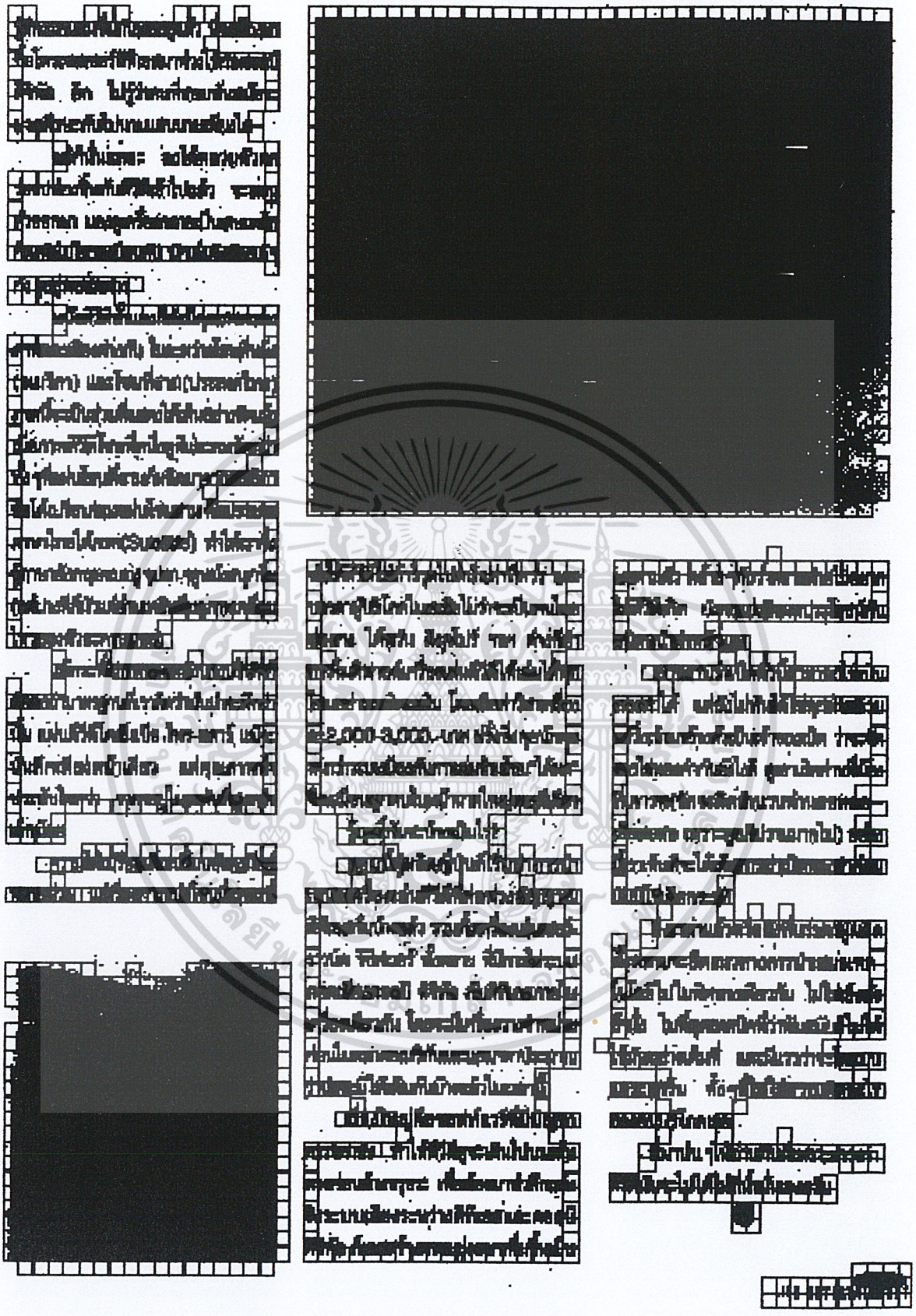
- เริ่มด้วยการเดินรอบขอบเขตส่วนประกอบโดยใช้หน้าต่าง ขนาด 32*32 ผลลัพธ์ที่ได้จะแสดงในตัวอย่างรูปที่ 5.1 ถึง รูปที่ 5.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



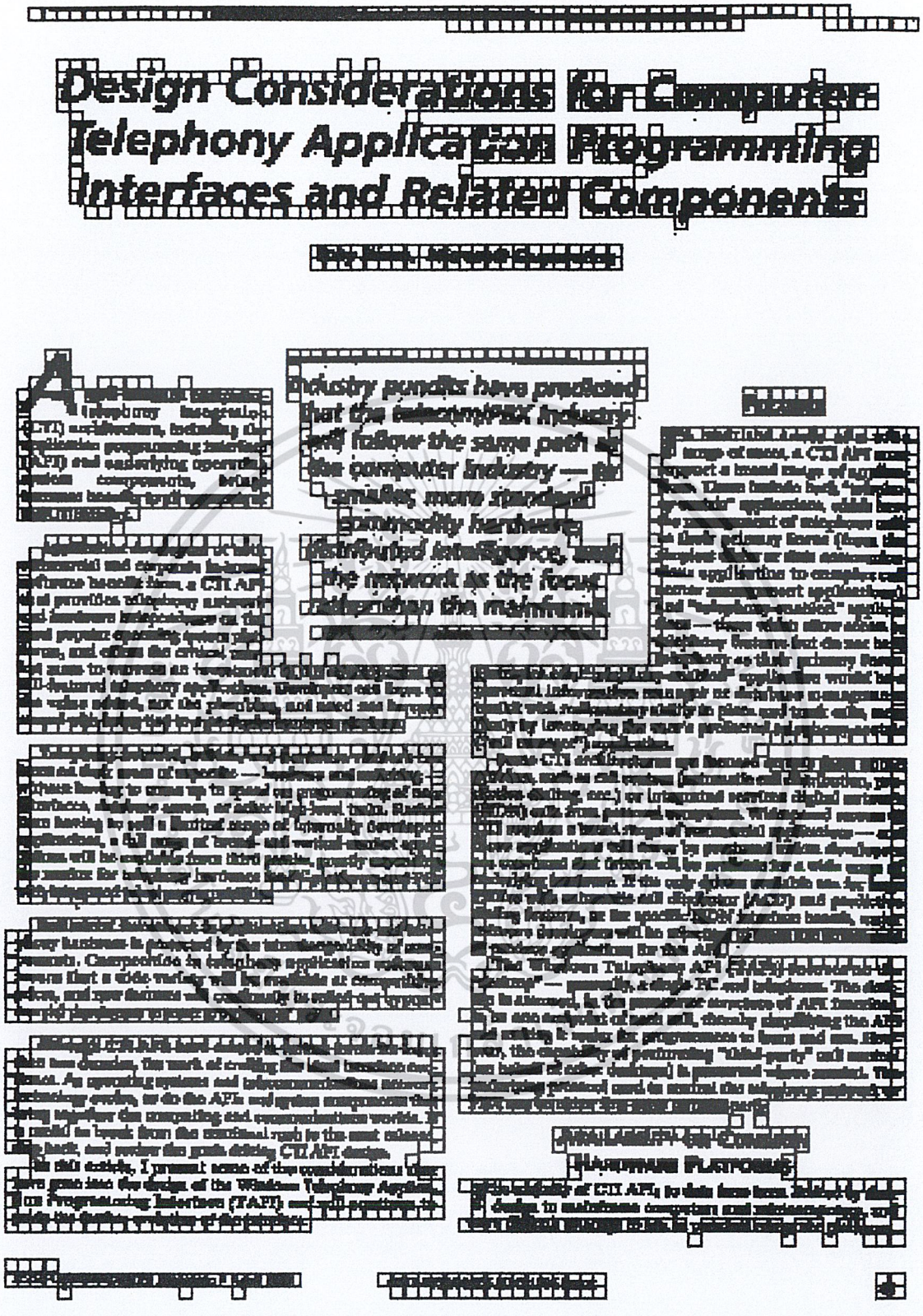
รูปที่ 5.1 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



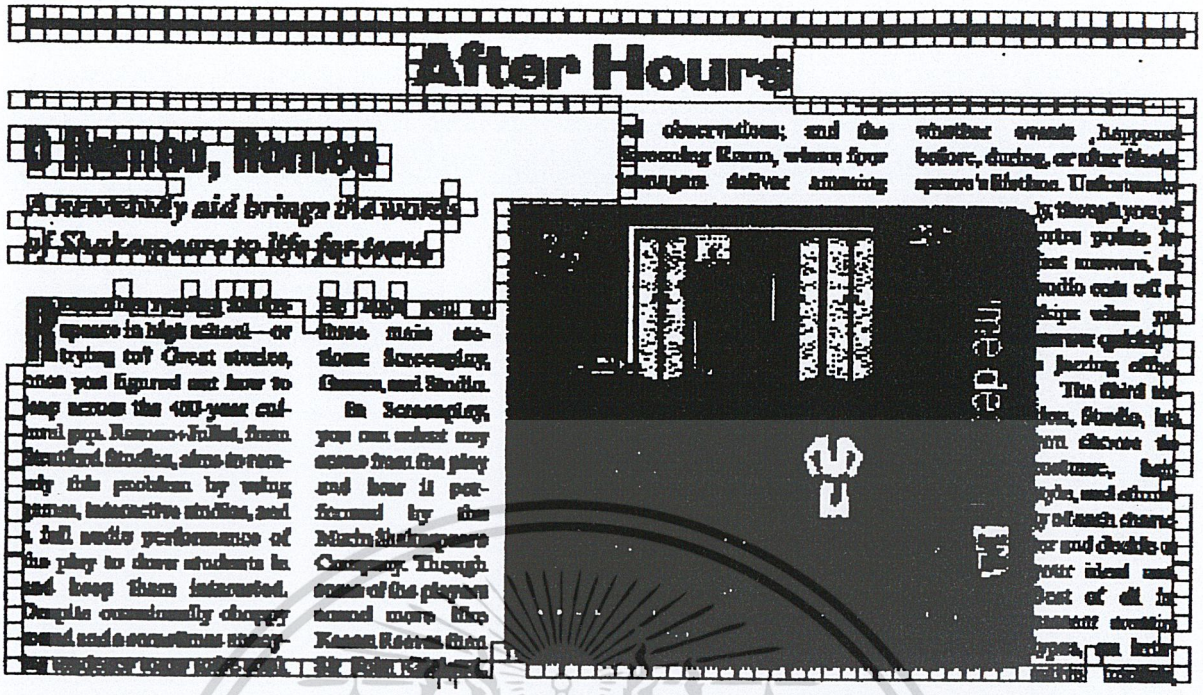
รูปที่ 5.2 แสดงภาพตัวอย่างที่ 2 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.3 แสดงภาพตัวอย่างที่ 10 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.4 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่

จากผลลัพธ์ที่ได้ นำมาวิเคราะห์แต่ละรูปได้ดังนี้

1. รูปที่ 5.1 ส่วนประกอบที่มีระยะห่างกันเกินกว่าหน้าต่างสองช่อง ซึ่งอัลกอริทึมสามารถแยกส่วนประกอบที่อยู่ห่างกันได้ แต่ก็มีส่วนประกอบบางส่วนประกอบที่อยู่ติดกัน ซึ่งควรแยกส่วนประกอบกัน แต่ถูกนำมารวมเข้าด้วยกัน ซึ่งสามารถวิเคราะห์ได้ว่าการทำงานนี้ถ้าส่วนประกอบที่อยู่ใกล้ชิดกันมากจะถูกนำมารวมกันหมด นอกจากนี้ยังสามารถสังเกตได้ว่าส่วนประกอบที่อยู่ส่วนล่างของภาพ ที่เป็นข้อความที่เชื่อมกับสมการ สังเกตได้ว่า จะมีส่วนคอดในแนวนอนระหว่างส่วนประกอบที่อาจจะไม่เป็นส่วนประกอบเดียวกัน

2. รูปที่ 5.2 ส่วนประกอบต่างๆ อยู่ห่างกันและสามารถแยกกันได้ดี แต่แต่ละย่อหน้ายังติดกันและไม่สามารถแยกออกจากกันได้

3. รูปที่ 5.3 รายละเอียดของภาพอยู่ติดกัน แต่ก็แยกส่วนประกอบได้ดี ยกเว้นการแยกส่วนประกอบย่อหน้า

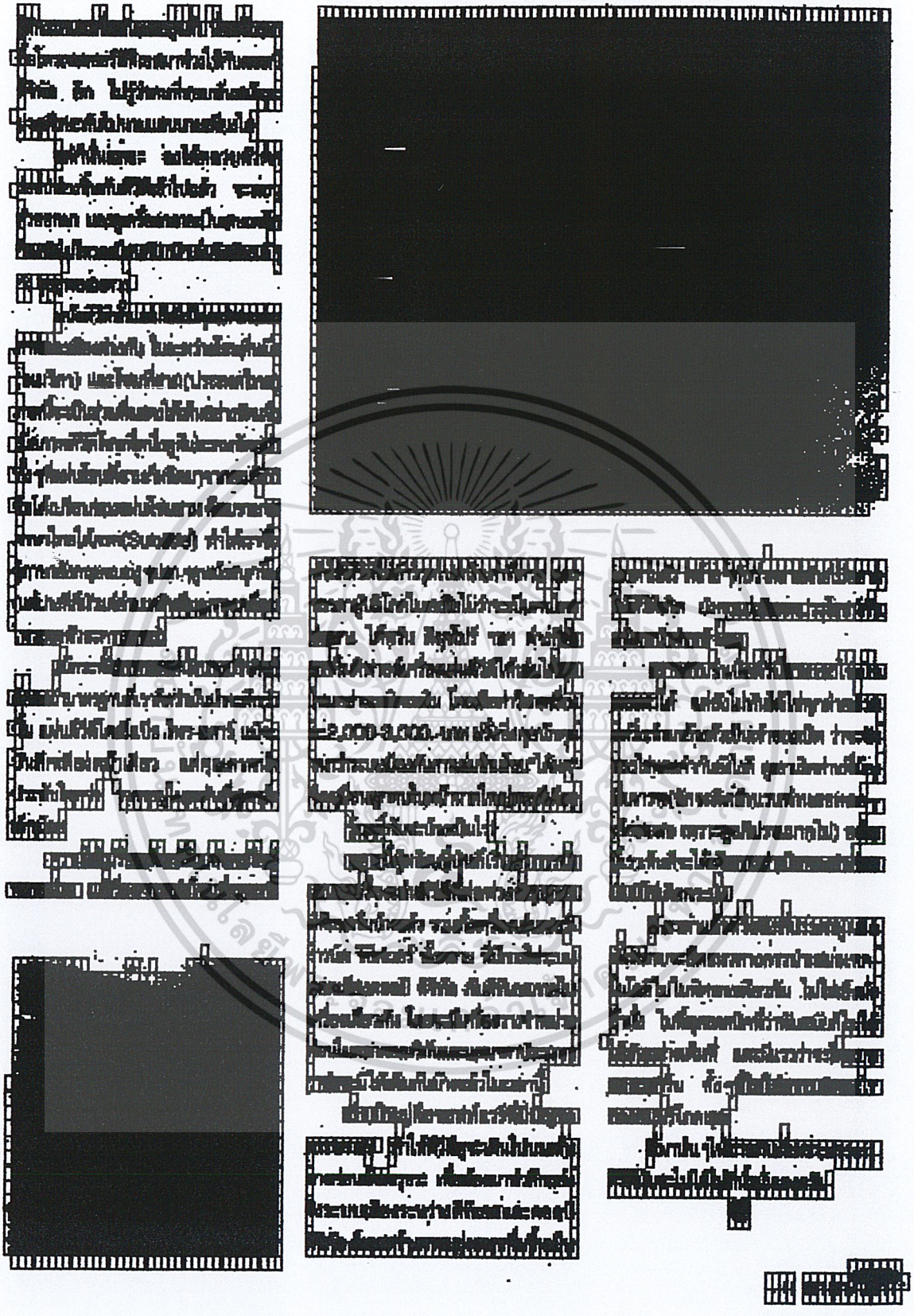
4. รูปที่ 5.4 ส่วนประกอบต่างๆ และยังมีเส้น ที่ทำให้ส่วนประกอบที่ไม่ควรติดกันติดกันภาพนี้จริงแยกได้ไม่ดีนัก

จากผลลัพธ์จึงควรปรับปรุงหน้าต่าง ให้สามารถแยกส่วนประกอบได้ดีขึ้น ได้ในตัวอย่างรูปที่ 5.5 ถึง รูปที่ 5.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

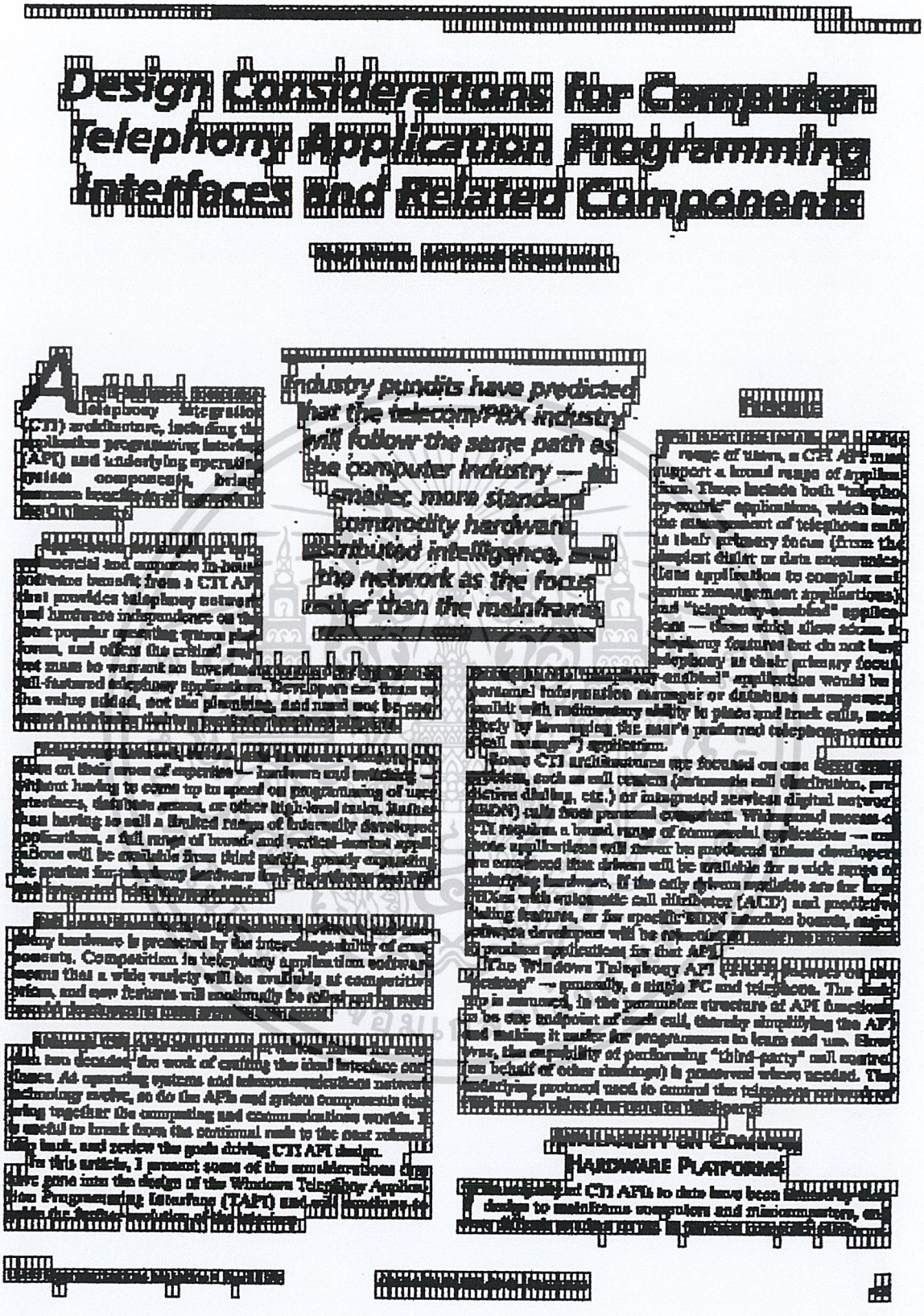
*รูปที่ 5.5 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



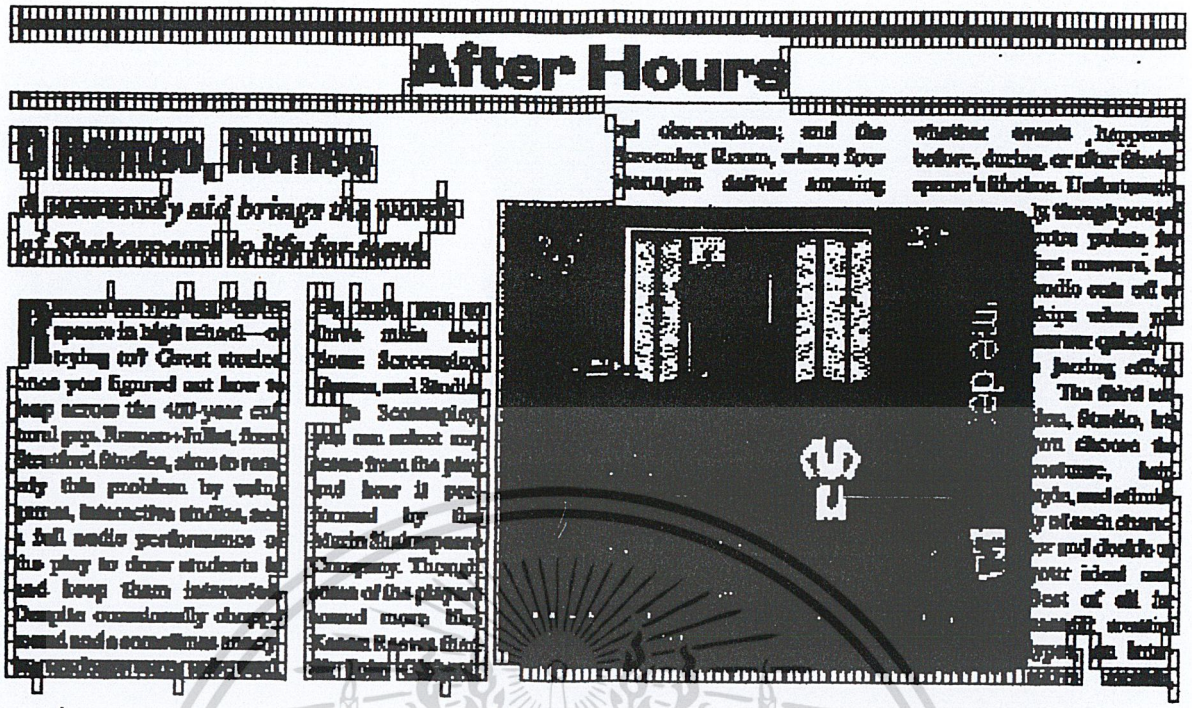
*รูปที่ 5.6 แสดงภาพตัวอย่างที่ 2 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



*รูปที่ 5.7 แสดงภาพตัวอย่างที่ 10 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาด หน้าต่าง กว้าง 16 pixel * ยาว 32 pixel*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



*รูปที่ 5.8 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบใหญ่ ด้วยขนาดหน้าต่าง กว้าง 16 pixel * ยาว 32 pixel*

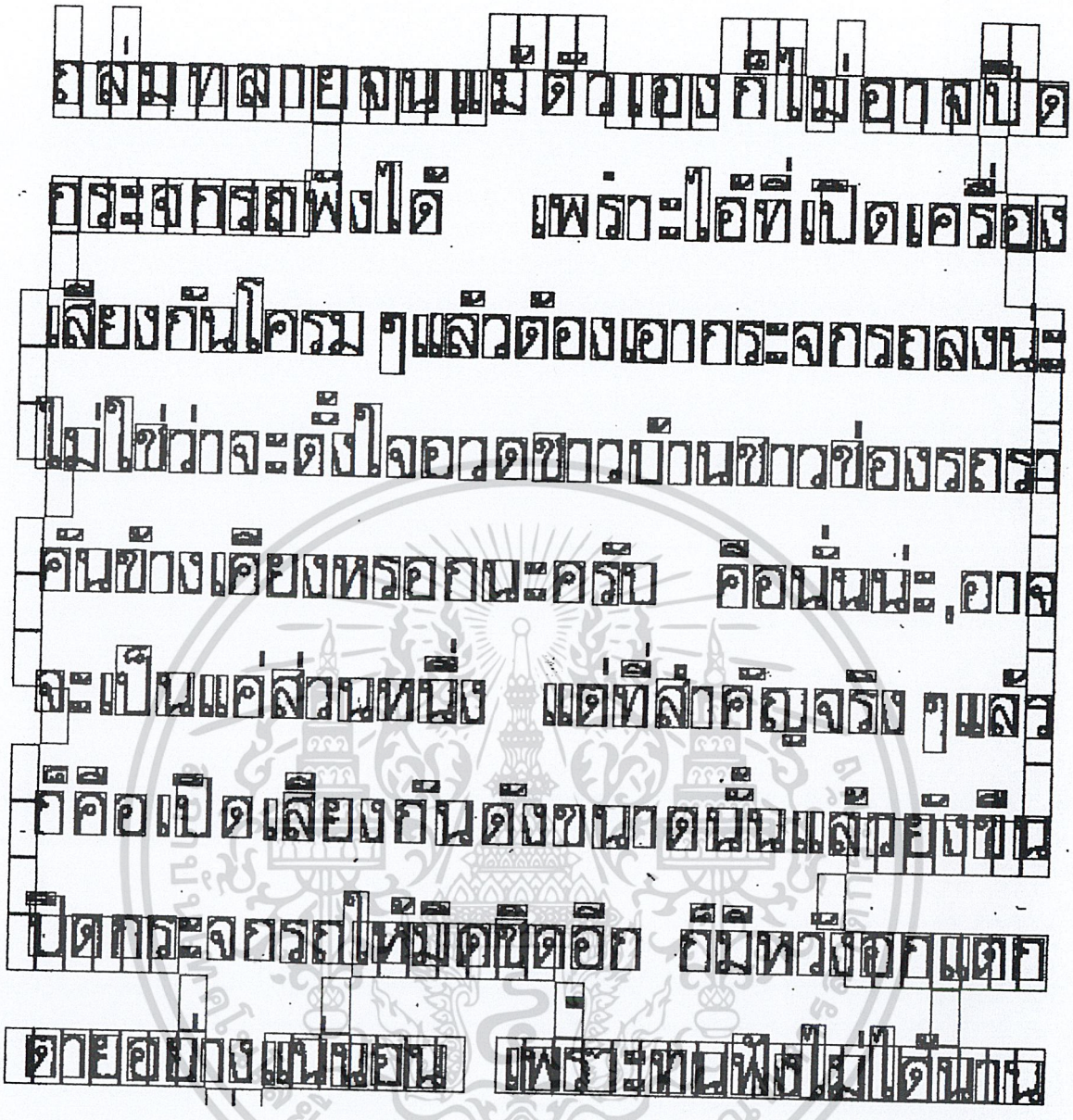
จากผลลัพธ์ที่ได้ นำมาวิเคราะห์แต่ละรูปได้ดังนี้

1. รูปที่ 5.5 ส่วนประกอบแยกกันได้ระหว่างคอลัมน์ได้ดี และจะต้องมีการแบ่งแยกย่อหน้าต่อไป
2. รูปที่ 5.6 ส่วนประกอบแบ่งแยกกันได้ดี เทียบเท่าหน้าต่างขนาด 32*32 ตารางพิกเซล
3. รูปที่ 5.7 ส่วนประกอบแบ่งแยกกันได้ด้อยลง เนื่องจากหน้าต่างแคบลงจึงมีการเชื่อมต่อกันน้อยลง เนื่องจากตัวอักษรมีขนาดใหญ่ ช่องว่างระหว่างตัวอักษรมาก จึงทำให้การแบ่งไม่ถูกต้อง แต่สำหรับโครงงานนี้ สามารถแบ่งเอกสารได้ขนาดเฉพาะเท่านั้น ทางแก้ไขทำได้ด้วยการเพิ่มความกว้างของหน้าต่างให้มีขนาดใหญ่ขึ้น เฉพาะบางเอกสารเท่านั้น
4. รูปที่ 5.8 ส่วนประกอบต่างๆ อยู่ใกล้ชิดกันมาก แต่ก็ยังสามารถแบ่งเอกสารได้ดี แต่มีเส้นลากจากซ้ายไปขวาด้านบน จึงทำให้ส่วนประกอบมาติดกัน ทางแก้ไขต่อไปควรตัดส่วนประกอบเอกสารที่เป็นภาพก่อนการเดิน ซึ่งอาจจะต้องทำช่วงที่สองของการทดลอง ก่อนช่วงที่หนึ่งของการทดลองก็เป็นได้

5.2 ตัวอย่างผลลัพธ์ที่ได้จากการทดลองช่วงที่สอง

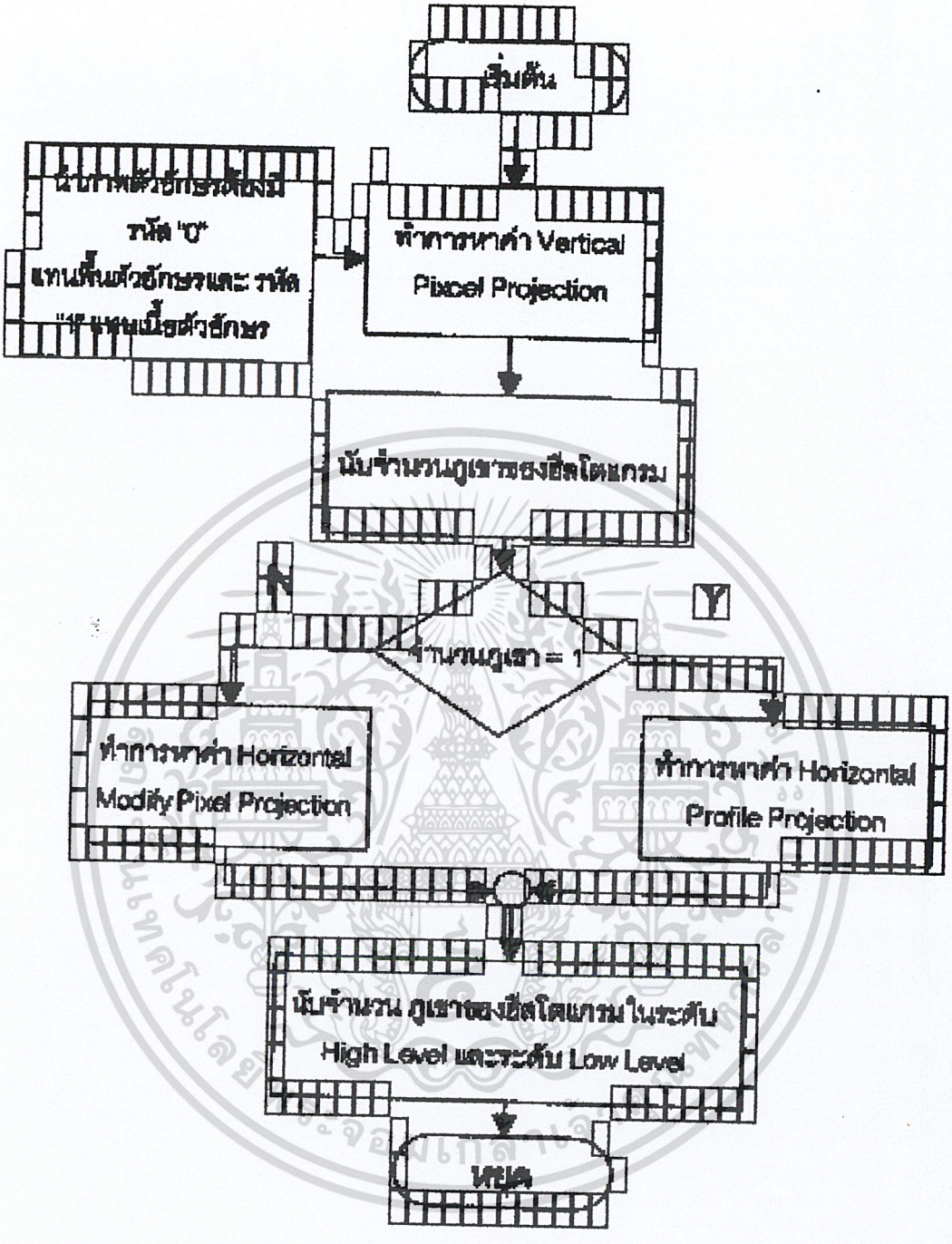
การทดลองช่วงที่สอง คือ การทำการแบ่งขอบเขตส่วนประกอบย่อยในส่วนประกอบใหญ่ รวมถึงการวิเคราะห์เพื่อแบ่งส่วนประกอบและแบ่งประเภทส่วนประกอบ โดยรูปตัวอย่างผลลัพธ์จะแสดงในรูปที่ 5.9 ถึง 5.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



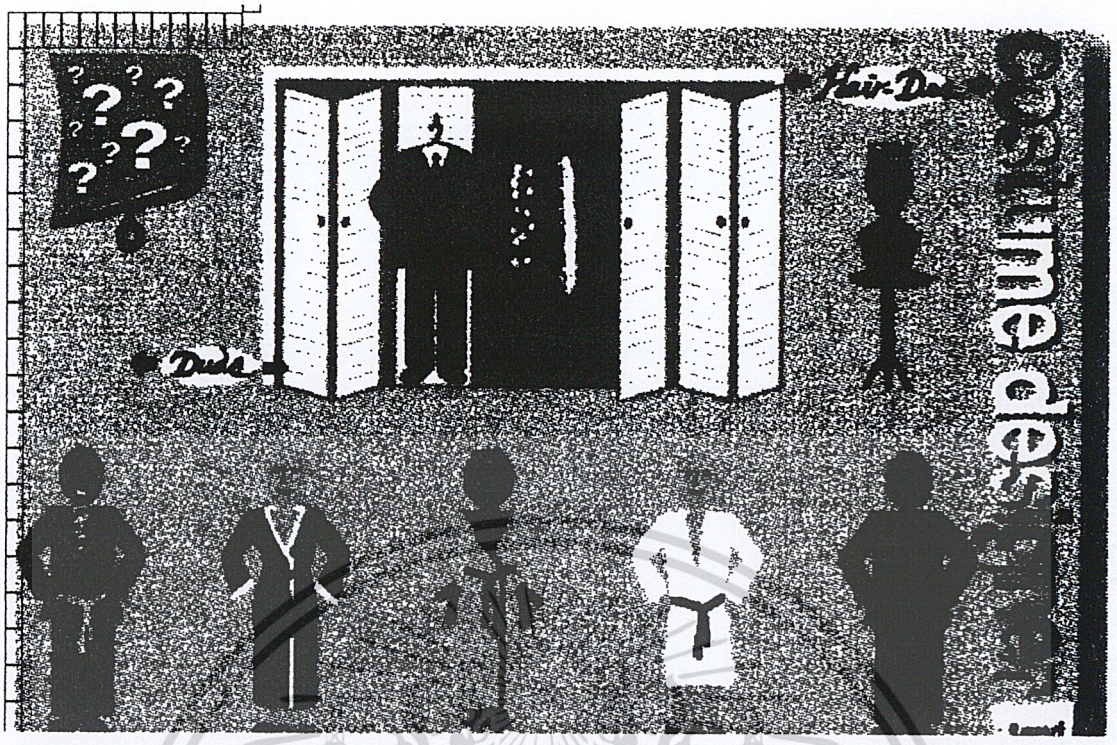
รูปที่ 5.9 แสดงภาพตัวอย่างที่ 1 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย และแสดงขอบเขตของตัวอักษรในแต่ละตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

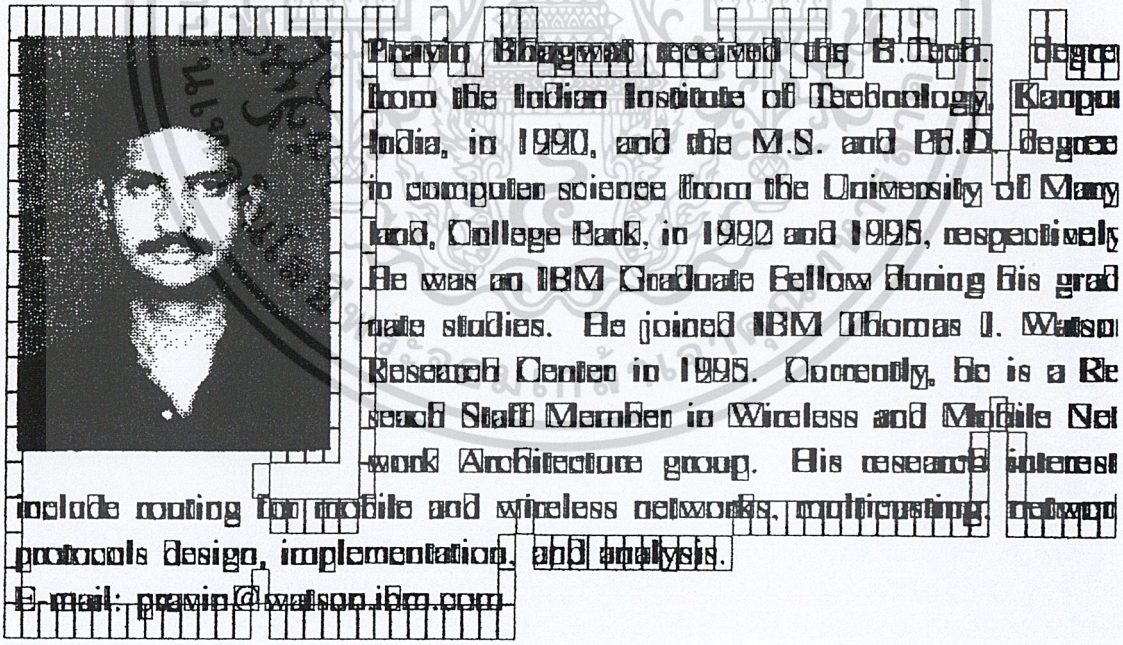


รูปที่ 5.10 แสดงภาพตัวอย่างที่ 5 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.11 แสดงภาพตัวอย่างที่ 8 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย



รูปที่ 5.12 แสดงภาพตัวอย่างที่ 9 หลังจากขั้นตอนการแบ่งขอบเขตส่วนประกอบย่อย และแสดงขอบเขตของตัวอักษรในแต่ละตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลลัพธ์ที่ได้ นำมาวิเคราะห์แต่ละรูปได้ดังนี้

1. รูปที่ 5.9 สามารถกำหนดขอบเขตของส่วนประกอบย่อยอย่างดีไม่มีปัญหา
2. รูปที่ 5.10 จะไม่สามารถเข้าไปยังตารางเพื่อจับตัวอักษรภายในตารางได้ ซึ่งเป็นไปตามที่คิดไว้
3. รูปที่ 5.11 ส่วนประกอบที่เป็นรูปภาพไม่มีเส้นรอบกรอบรูป และมีจุดมากมายภายในรูป แต่ถึงอย่างไรก็ไม่มีปัญหาในการเดินวนรอบ
4. รูปที่ 5.12 ส่วนประกอบที่เป็นรูปภาพไม่มีเส้นรอบกรอบรูป ไม่มีปัญหาในการเดินรอบขอบเขตระหว่างภาพ และตัวอักษรแต่อย่างไร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการทดลองและเสนอแนะแนวทางการดำเนินงานต่อไป

จากผลการทดลองที่ได้ แสดงให้เห็นว่า การแบ่งข้อมูลได้ถูกต้อง เวลาการทำงานน้อย แต่การใช้หน่วยความจำในการจัดเก็บข้อมูลอาจจะมากเกินไป เนื่องจากการทดลองบางภาพต้องรันในโหมดดีบักเท่านั้นดังนั้นจึงมีข้อสรุปผลการทดลอง และการเสนอแนวทางแก้ไขดังนี้

6.1 สรุปผลการทดลอง

1. การเดินรอบขอบเขตใหญ่ในช่วงที่หนึ่ง มีผลเป็นที่น่าพอใจ
2. การสร้างโปรแกรมที่ใช้ในการทดสอบ ลักษณะโครงสร้างของข้อมูลเป็นแบบไม่คงที่ ฉะนั้นการที่ใช้โครงสร้างข้อมูลแบบลิงค์ลิสต์ อาจจะมีการลบหน่วยความจำที่ไม่ใช้ไม่หมด ในการทำงานต่างๆ จึงทำให้เกิดการเขียนข้อมูลลงบนฮาร์ดดิสก์แทนหน่วยความจำหลัก ซึ่งทำให้เพิ่มเวลาทำงานอีก
3. การเขียนโปรแกรมที่มีการเก็บรายละเอียดข้อมูลต่างๆเพื่อใช้ในการตรวจสอบมากเกินไป ทำให้ต้องเปลืองหน่วยความจำ
4. ขอบเขตส่วนประกอบที่ได้เป็นรูปแบบแผนผังแมนฮัตตัน(ไม่เป็นสี่เหลี่ยม)
5. การตรวจสอบจุดภายในภาพเอกสารไม่จำเป็นต้องทำทุกจุด ทำให้แบ่งส่วนเอกสารได้อย่างรวดเร็ว
6. เมื่อปิดโปรแกรมจะ error เนื่องจากการจองพื้นที่หน่วยความจำไว้ และไม่ได้ยกเลิกการจอง เนื่องจากโปรแกรมดังกล่าว ยังขาดส่วนระบบรู้จำตัวอักษร ดังนั้น จึงต้องจองพื้นที่สำหรับเก็บข้อมูลไว้สำหรับทำต่อไป
7. เพื่อจะนำไปแปลเป็นตัวอักษรใช้งานในคอมพิวเตอร์ ยังไม่มีวิธีการส่งเรียงตัวอักษรให้เป็นประโยคที่ถูกต้องได้

6.2 ข้อเสนอแนะ

1. ลดการใช้หน่วยความจำ เพราะในการแบ่งข้อมูลให้เล็กลงๆ ไปเรื่อยๆ จะทำให้ต้องเก็บข้อมูลมากขึ้นเรื่อยๆ ทำให้ระบบต้องดูแลหน่วยความจำมากขึ้น ซึ่งเป็นผลให้ต้องมีการเขียนข้อมูลลงมาไว้ที่ฮาร์ดดิสก์เพื่อต้องการดึงข้อมูลอื่นไปไว้ในหน่วยความจำแทน เพราะฉะนั้นถ้าลดการจัดเก็บขอบเขตข้อมูลลง สามารถที่จะลดการใช้หน่วยความจำลงได้ โดยออกแบบโครงสร้างในการจัดเก็บข้อมูลและต้องมีการจัดการหน่วยความจำให้ดี
2. ปรับขนาดของหน้าต่างตามภาษา จะช่วยเพิ่มความถูกต้องของการแบ่งย่อหน้าได้มากขึ้น และสามารถทำให้เป็นอัตโนมัติได้ โดยการตรวจสอบภาษาก่อนสัก 1 ตัวอักษรก่อน
3. ประโยชน์สำหรับระบบจัดเก็บเอกสาร โดยขอบเขตส่วนประกอบที่ได้สามารถนำไปสร้างรูปแบบของเอกสารเพื่อเก็บไว้ในระบบจัดเก็บเอกสาร และให้ระบบเรียนรู้หน้าที่ของตัวอักษร เช่น ตัวเลขมุมบนขวา จะเป็นเลขหน้า เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ในการแยกขอบเขตส่วนประกอบ สามารถ ลดขนาดการจัดเก็บภาพลง และสามารถแปลงภาพเอกสารที่แยกส่วนแล้ว กลับมาเป็นภาพต้นฉบับได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] G. Nagy and S. Seth, "Hierarchical representation of optically scanned documents," *Proc. of ICPR*, 1984, pp. 347-349.

[2] T. Akiyama and N. Hagita, "Automated entry system for printed documents," *Pattern Recognition 23*, 1990, pp. 1141-1154.

[3] K. C. Fan, C. H. Liu, and Y. K. Wang, "Segmentation and classification of mixed text/graphics/image documents," *Pattern Recognition Letters* 15, 1994, pp. 1201-1209.

[4] S.N. Srihari, T. Hong, and G. Srikantan, "Machine printed Japanese document recognition," *Pattern recognition 30*, 1997, pp. 1301-1313.

A. Antonacopoulos and R. T. Ritchings, "Flexible page segmentation using the background," *Proc. of the 12th ICPR*, Jerusalem. Israel, 1994, pp. 339-344.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้