

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ระบบแยกตัวอักษร

CHARACTERS SEGMENTATION



เลขหน้.....
เลขทะเบียน..... 34092
วัน, เดือน, ปี-5 ๓.ค. 2542

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2541

ปริญญานิพนธ์ปีการศึกษา 2541

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบแยกตัวอักษร

CHARACTERS SEGMENTATION

ผู้จัดทำ

1. นาย กิตติศักดิ์ บัญชาวิมลเชษฐ
2. นาย บุญฤทธิ์ วงศ์สุฤทธิ์



(รศ.ดร. ชม กัมปาน)

อาจารย์ที่ปรึกษา

ระบบแยกตัวอักษร

นาย กิตติศักดิ์ บัญชาวิมลเชษฐ 39013230

นาย บุญฤทธิ์ วงศ์สุฤทธิ์ 39013239

รศ.ดร. ชม กัมปาน อาจารย์ที่ปรึกษา

ปีการศึกษา 2541

บทคัดย่อ

โครงการเป็นการแยกอักขระออกจากภาพ (Segmentation) ซึ่งเป็นองค์ประกอบหนึ่งที่สำคัญของการทำ Optical Character Recognition (OCR) ก็คือ การแยกเอาวัตถุที่ต้องการออกจากภาพซึ่งเป็น input หรือนำภาพผ่านกระบวนการ segmentation โดยวัตถุที่ได้นั้นก็คือ ตัวอักขระทุกตัวบนแผ่นกระดาษ

โดยได้นำเสนอวิธีการ segment ไว้ด้วยกัน 2 วิธีคือ Contour Following และ Contour with Matrix และใช้วิธีการของ Histogram ซึ่งเป็นทฤษฎีเสริมที่ช่วยตัดสินใจในการแยกตัวอักขระที่ติดกัน Contour Following เป็นการตามรอยขอบของตัวอักขระ Contour with Matrix เป็นการคัดลอกตามส่วนที่เป็นเนื้อของตัวอักขระ Histogram เป็นการช่วยตัดสินใจในแยกตัวอักขระที่ติดกัน โดยกำหนดค่าผิดพลาดให้กับ Histogram เพื่อทำการแยกตัวอักขระ

การทดลองในขั้นตอนการทำ segment ตัวอักขระภาษาไทย และภาษาอังกฤษ ข้อมูลที่นำมาเป็นอินพุต คือหน้าเอกสารทั่วไป ที่ผ่านการ scan เพื่อให้เครื่องคอมพิวเตอร์ สามารถนำมาทำการประมวลผล ซึ่งในการทดลองได้ทำการ scan ที่ค่าความละเอียด 300 dpi ข้อมูลที่ใช้ในการทดลองมีลักษณะของ Font ต่างกัน 3 Font คือ Angsana New , Browallia New และ Cordia New และมีลักษณะของตัวอักขระ 3 แบบ คือ ตัวอักขระปกติ ตัวเอียง และตัวหนา โดยใช้วิธีต่างๆ ข้างต้นมาใช้ในการทำ segment ตัวอักขระ

จากการทดลองวิธี Contour Following และ Contour with Matrix ให้ผลการทำ segment ที่เหมือนกัน โดยภาษาไทยมี 618 ตัว ในการ Segment Font Angsana New ตัวปกติถูก 616 ผิด 2 ตัวเอียงถูก 612 ผิด 6 ตัวหนาถูก 612 ผิด 6 Font Browallia New ตัวปกติถูก 612 ผิด 6 ตัวเอียงถูก 610 ผิด 8 ตัวหนาถูก 612 ผิด 6 Font Cordia New ตัวปกติถูก 606 ผิด 12 ตัวเอียงถูก 606 ผิด 12 ตัวหนาถูก 586 ผิด 32

ภาษาอังกฤษมี 356 ตัว ในการ Segment Font Angsana New ตัวปกติถูก 344 ผิด 12 ตัวเอียงถูก 348 ผิด 8 ตัวหนาถูก 352 ผิด 4 Font Browallia New ตัวปกติถูก 350 ผิด 6 ตัวเอียงถูก 350 ผิด 6 ตัวหนาถูก 356 ผิด 0 Font Cordia New ตัวปกติถูก 354 ผิด 2 ตัวเอียงถูก 350 ผิด 6 ตัวหนาถูก 350 ผิด 6

Characters Segmentation

Kittisak Banchawimoncheat

Boonrit Vongsurit

Assoc. Prof. Dr. Chom Kimpan Advisor

ABSTRACT

The purpose of this project is Characters Segmentation, the necessary part for Optical Character Recognition (OCR). Conceptually, separate object we want from input image or pass the image through the segmentation process. The object is every character in the paper.

Two Segmentation methods are presented in this thesis. Contour Following and Contour with Matrix. Contour Following is following along edge of the character. Contour with Matrix is copying the entire pixel of the character. Histogram supplementary method are used to separate characters when have more than 2 characters are connected, we can evaluate error value to the histogram to separate connected characters properly.

Experiment for segmentation on Thai and English characters. Input for experiment is general paper passed from scanner. We setting resolution for scanning at 300 dpi and we use three difference fonts for experimenting are Angsana New, Browallia New and Cordia New. Each form of fonts we use are Normal, Italic and Bold form, then Segment them by our above methods.

The result we approached from Contour Following and Contour with Matrix are the same. Thai language 618 characters in Angsana New , normal font correct 616 incorrect 2 characters, italic font correct 612 incorrect 6 characters, Bold font correct 612 incorrect 6. Browallia New , normal font correct 612 incorrect 6 characters, italic font correct 610 incorrect 8 characters, Bold font correct 612 incorrect 6. Cordia New , normal font correct 606 incorrect 12 characters, italic font correct 606 incorrect 12 characters, Bold font correct 582 incorrect 32.

In English 356 characters in Angsana New , normal font correct 344 incorrect 12 characters, italic font correct 348 incorrect 8 characters, Bold font correct 352 incorrect 4. Browallia New , normal font correct 350 incorrect 6 characters, italic font correct 350 incorrect 6 characters, Bold font correct 356 incorrect 0. Cordia New , normal font correct 354 incorrect 2 characters, italic font correct 350 incorrect 6 characters, Bold font correct 350 incorrect 6.

กิตติกรรมประกาศ

ขอขอบพระคุณ รองศาสตราจารย์ ดร. ชม กัมปาน ที่ได้ให้ความกรุณาช่วยเหลือ ทั้งทางด้าน
ความรู้ และตลอดจนให้คำแนะนำปรึกษาจนสำเร็จด้วยดี

ขอขอบพระคุณเหล่าคณาจารย์ และเจ้าหน้าที่ภาควิชาวิศวกรรมคอมพิวเตอร์



สารบัญ

หน้าที่

บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญภาพ	VI
สารบัญตาราง	VIII

บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 วิธีการดำเนินงาน	2
บทที่ 2 ทฤษฎีการแยกตัวอักษร	4
2.1 การหากรอบตัวอักษรด้วยวิธี Line Crossing	4
2.2 เทคนิคการตามรอยขอบภาพ (Contour Following)	6
2.3 Contour with Matrix	8
2.4 เทคนิคการแยกตัวอักษรที่ติดกัน โดยใช้ Histogram	11
2.5 การแยกภาพตัวอักษรออกจากข้อมูลภาพ	13
บทที่ 3 การใช้งานโปรแกรมแยกตัวอักษร	16
3.1 การเริ่มต้นใช้งานโปรแกรม	16
3.2 การอ่านข้อมูลจากไฟล์รูปภาพ	17
3.3 การทำ Segment ตัวอักษร	18
บทที่ 4 ผลการทดลอง	26
4.1 ขั้นตอนและวิธีที่ใช้ในการทดลอง	26
4.2 ผลการทดลอง	28
4.3 สรุปผลการทดลอง	57

สารบัญ(ต่อ)

หน้าที่

บทที่ 5 บทสรุป	62
5.1 ส่วนประกอบสำคัญที่ทำให้การ Segment มีความถูกต้องสูง	62
5.2 ข้อเสนอแนะ	62
ภาคผนวก ก โครงสร้างไฟล์รูปภาพ	63
ภาคผนวก ข ข้อมูลภาษาไทย และภาษาอังกฤษที่ใช้ในการทดสอบ	66
ภาคผนวก ค โปรแกรมการทำงานของระบบ	75
บรรณานุกรม	115



สารบัญภาพ

หน้าที่

รูปที่ 1-1 องค์ประกอบการทำ OCR	1
รูปที่ 2-1 แสดงการแยกตัวอักษรด้วยวิธี Line Crossing	4
รูปที่ 2-2 แสดง Flowchart การแยกตัวอักษรด้วยวิธี Line Crossing	5
รูปที่ 2-3 แสดงการ Segment ด้วยวิธี Contour Following	6
รูปที่ 2-4 แสดง Flowchart การ Segment ด้วยวิธี Contour Following	7
รูปที่ 2-5 Matrix ขนาด 3x3	8
รูปที่ 2-6 แสดงการ Segment ด้วยวิธี Contour with Matrix	9
รูปที่ 2-7 แสดง Flowchart การ Segment ด้วยวิธี Contour with Matrix	10
รูปที่ 2-8 แสดงการกำหนดจุดตัดด้วย Histogram	11
รูปที่ 2-9 แสดง Flowchart การแยกตัวอักษรที่ติดกันด้วยการใช้ Histogram	12
รูปที่ 2-10 แสดงการแยกภาพตัวอักษรออกจากข้อมูลภาพ	14
รูปที่ 2-11 แสดง Flowchart การแยกตัวอักษรออกจากข้อมูลภาพ	15
รูปที่ 3-1 แสดงหน้าจอการทำงานหลัก	16
รูปที่ 3-2 แสดงหน้าจอการ Load Image	17
รูปที่ 3-3 การแสดงรูปที่ทำการอ่านข้อมูลจากไฟล์	18
รูปที่ 3-4 การแสดงผลการแยกบรรทัดจากการทำ Segment	19
รูปที่ 3-5 การแสดงผลของการทำ Segment ตัวอักษร	20
รูปที่ 3-6 แสดงรูป Question Box เพื่อใช้ Save Image	20
รูปที่ 3-7 แสดงหน้าจอการ Save Image	21
รูปที่ 3-8 แสดงข้อความเมื่อชื่อไฟล์ในการ Save Image เหมือนกัน	21
รูปที่ 3-9 แสดงรูป Question Box เพื่อใช้ Histogram ในการ Segment	22
รูปที่ 3-10 แสดงรูป Question Box เพื่อใช้หยุดการทำงานของโปรแกรม	22
รูปที่ 3-11 แสดงตัวรับค่าเปอร์เซ็นต์ Error ในแนวนอนและแนวตั้ง	23
รูปที่ 3-12 แสดงรูป Question Box เพื่อการ Segment ข้อมูลใหม่	23
รูปที่ 3-13 แสดงรูป Question Box เพื่อการ Segment ข้อมูลเดิม	24
รูปที่ 3-14 แสดงรูป Question Box เมื่อทำการ Segment จนหมดบรรทัด	24
รูปที่ 3-15 แสดงรูป Message Box เมื่อทำการ Segment จนหมดหน้าเอกสาร	25
รูปที่ 4-1 ตัวอย่างการทำงานผิดพลาดของวิธีการตามรอยขอบของตัวอักษร	27
รูปที่ 4-2 ตัวอย่างเอกสารทดสอบขนาดเท่าจริง Font Angsana New ขนาด 18 ภาษาไทย	37
รูปที่ 4-3 ตัวอย่างเอกสารทดสอบขนาดเท่าจริง Font Angsana New ขนาด 18 ภาษาอังกฤษ	38
รูปที่ 4-4 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษที่ติดกัน	38

สารบัญภาพ(ต่อ)

หน้าที่

รูปที่ 4-5 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ติดกัน	38
รูปที่ 4-6 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษที่ใช้ในการทดลอง	57
รูปที่ 4-7 แสดงตัวอักษรภาษาอังกฤษที่ติดกันที่สามารถแยกได้	58
รูปที่ 4-8 แสดงตัวอักษรภาษาอังกฤษที่ติดกันที่ไม่สามารถแยกได้	58
รูปที่ 4-9 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษตัวเอียงที่ใช้ในการทดลอง	58
รูปที่ 4-10 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ใช้ในการทดลอง	59
รูปที่ 4-11 แสดงตัวอักษรภาษาไทยที่ติดกันที่สามารถแยกได้	59
รูปที่ 4-12 แสดงตัวอย่างการกำหนดค่า Error ที่สูงเกินไป	59
รูปที่ 4-13 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ใช้ในการทดลอง	60
รูปที่ 4-14 แสดงตัวอักษรภาษาไทยที่ติดกันที่ไม่สามารถแยกได้	60



สารบัญตาราง

หน้าที่

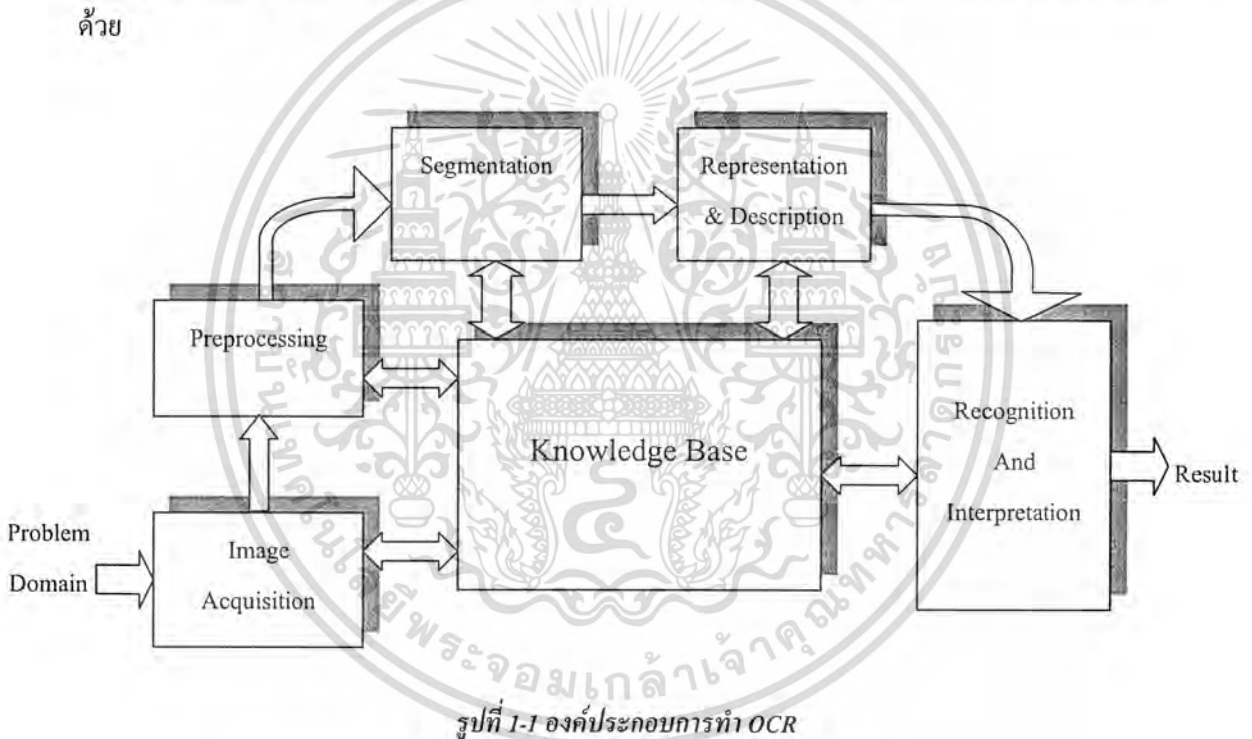
ตารางที่ 4-1 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย	26
ตารางที่ 4-2 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Angsana New	28
ตารางที่ 4-3 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Angsana New	29
ตารางที่ 4-4 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Angsana New	29
ตารางที่ 4-5 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Browallia New	30
ตารางที่ 4-6 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Browallia New	30
ตารางที่ 4-7 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Browallia New	31
ตารางที่ 4-8 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Crodia New	31
ตารางที่ 4-9 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Crodia New	32
ตารางที่ 4-10 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Crodia New	32
ตารางที่ 4-11 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Angsana New	33
ตารางที่ 4-12 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Angsana New	33
ตารางที่ 4-13 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Angsana New	34
ตารางที่ 4-14 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Browallia New	34
ตารางที่ 4-15 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Browallia New	35
ตารางที่ 4-16 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Browallia New	35
ตารางที่ 4-17 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Crodia New	36
ตารางที่ 4-18 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Crodia New	36
ตารางที่ 4-19 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Crodia New	37

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

ขั้นตอนต่าง ๆ ในการทำโครงการ OCR ควรต้องมีพื้นฐานขององค์ประกอบต่างๆ ดังรูปที่ 1-1 ซึ่งในแต่ละส่วนมีความสำคัญ และหน้าที่ที่แตกต่างกันไป ดังจะได้อธิบายต่อไป ส่วนที่เราจะเน้นมากที่สุดในการโปรเจกต์คือการทำ segmentation ซึ่งถือเป็นปัจจัยที่สำคัญที่สุดของการเรียนรู้ของคอมพิวเตอร์ ถ้ามีการ segment ได้ตามสัดส่วนของวัตถุ และตรงกับวัตถุประสงค์ที่ต้องการแล้ว ไม่เพียงส่งผลให้เกิดความถูกต้องในการรู้จำของคอมพิวเตอร์มากขึ้นเท่านั้น และยังจะช่วยทำให้ผลลัพธ์ที่ได้ไม่น่าผิดพลาด



หลักการทำงานในแต่ละส่วนที่ใช้ทำ OCR มีหน้าที่และความสำคัญดังต่อไปนี้

Problem domain : ขอบเขตของปัญหาที่เราคัดเลือกมาแล้ว โครงการของเราเน้นถือว่า Problem domain คือหน้ากระดาษเอกสารต่างๆ ไป ที่มีตัวอักษรเป็นหลัก ได้ทั้งภาษาไทย และภาษาอังกฤษ สีของเนื้อกระดาษ และตัวอักษรสามารถแยกแยะได้ชัดเจน ขอบเขตที่กำหนดค่อยลงไปอีกก็คือ ควรจะเป็นตัวอักษรที่ผ่านการพิมพ์มา และมีขนาดตัวอักษรที่ไม่เล็กจนเกินไปนัก คือไม่ต่ำกว่า 14 dot per inch

Image acquisition : เป็นอุปกรณ์ที่ใช้ในการรับข้อมูลภาพเข้ามา และเก็บไว้ในคอมพิวเตอร์ ซึ่งส่วนใหญ่แล้วในการทำ OCR จะนิยมใช้ Scanner ในการรับภาพ โดยโครงการนี้ไม่จำเป็นที่จะต้องเป็น Scanner ดี และมีความละเอียดในการ scan มากนัก และภาพที่นำมาใช้ต้องถูก scan ในโหมด ขาว - ดำ

Preprocessing : เป็นส่วนที่ใช้ทำการปรับแต่งภาพที่รับเข้ามาให้มีความชัดเจนมากยิ่งขึ้น เพราะการทำ segment เราอาจได้ส่วนเกินที่เกิดจากข้อผิดพลาดจากการ scan หรือต้นฉบับที่ไม่สะอาด ไม่มีความคมชัด ดังนั้นสิ่งที่ผิดปกติบนหน้ากระดาษจึงต้องถูกลบออกไป แต่ก็ควรระวังในการลบเพราะสิ่งที่เราลบอาจเป็นเนื้ออักษรที่สำคัญก็ได้ โดยเฉพาะภาษาไทยที่มีสระค่อนข้างมาก แต่ถ้าเราไม่ทำความสะอาดหรือกำจัดส่วนเกินนี้แล้ว เราก็อาจไม่สามารถทำขั้นตอนต่อไป เพื่อให้ได้ผลลัพธ์ที่ดีได้ ดังนั้นทั้งก่อนการ scan และหลัง scan จึงควรตรวจสอบจุดเพี้ยนหรือ noise ซึ่งมักเกิดขึ้นเสมอๆ

Segmentation : การแยกเอาวัตถุที่ต้องการในภาพออกจากฉากใดๆ เพื่อนำไปใช้ในการวิเคราะห์หรือหาความหมายของสิ่งที่วัตถุนั้นสื่อ ด้วยการนำสิ่งที่ได้จากการ segment ไปเทียบกับข้อมูลที่อยู่ในฐานข้อมูล วิธีการของการ segment เพื่อการรู้จำตัวอักษรได้กล่าวไว้ในบทที่ 2

Representation & description : การแทนรูปร่างของวัตถุด้วยสัญลักษณ์ , โครงสร้างข้อมูล หรือวิธีการใดๆ ที่ทำให้เราสามารถถอดรหัสกลับมา แล้วสื่อถึงวัตถุสิ่งนั้น แล้วเก็บตัวแทนนั้นไว้ในฐานข้อมูลเพื่อนำไปใช้ในการทำขั้นตอนต่อไป

Recognition & interpretation : เป็นส่วนที่ทำหน้าที่ คำนวณตัดสินว่าสิ่งที่ถูกป้อนเข้ามานั้นคืออะไร โดยนำเอาข้อมูลที่ได้รับเข้ามานั้น มาทำการเปรียบเทียบ ค้นหา กับความรู้ที่มีอยู่ในฐานข้อมูล

Knowledge base : เป็นส่วนที่สำคัญ ใช้เป็นฐานข้อมูล เก็บกฎต่างๆ เพื่อให้ระบบสามารถดึงเอาข้อมูล หรือความรู้ออกมาใช้ได้ โดยไม่ได้มีไว้เฉพาะส่วนใดส่วนหนึ่งของระบบ แต่ทุกๆ ส่วนก็สามารถใช้งานฐานความรู้ส่วนนี้ได้ตลอดเวลา

1.2 วัตถุประสงค์ของงานวิจัย

- 1.2.1 ศึกษาวิธีการทำการแยกตัวอักษระแบบภาพขาว-ดำด้วยวิธีการต่างๆ
- 1.2.2 นำวิธีการแยกตัวอักษระมาทดสอบกับตัวแบบอักษระที่มีใช้ใน โปรแกรม Windows
- 1.2.3 ทำบทสรุปถึงความสำเร็จในการทำการแยกตัวอักษระกับแต่ละตัวแบบอักษระใน Windows

1.3 ขอบเขตของงานวิจัย

แยกอักษระจากภาพขาว-ดำที่ได้ทำการ Scan ผ่านเครื่อง Scanner ด้วยวิธีการต่างๆ เช่น Contour with following , Contour with matrix พร้อมสร้างบทสรุปของผลลัพธ์ที่ได้จากการทำการแยกตัวอักษระกับตัวแบบอักษระแบบต่างๆ

1.4 วิธีการดำเนินงาน

ช่วงการดำเนินงานสามารถกำหนดให้เป็น 4 ช่วงดังนี้ คือ

1.4.1 กำหนดขอบเขตของงานวิจัยที่จะทำการศึกษา ตั้งวัตถุประสงค์ และวางแผนดำเนินงานตามวัตถุประสงค์ที่ได้ตั้งไว้ การดำเนินงานแบ่งได้เป็น 2 ช่วงดังต่อไปนี้

1.4.1.1 ศึกษาทฤษฎีที่จะนำมาประยุกต์ใช้ต่างๆ ในการทำ segmentation ตัวอักษระทั้งภาษาไทย และภาษาอังกฤษ ซึ่งทฤษฎีที่นำมาใช้ได้อธิบายไว้ในบทที่ 2

1.4.1.2 ทำการศึกษา และเลือกใช้ application ที่ช่วยสนับสนุนๆ การทำงานของการทำ segmentation ตัวอักษรทั้งภาษาไทย และภาษาอังกฤษ และทดสอบการใช้ tools ต่างๆ ที่จะช่วยให้การพัฒนาโครงการทำได้ง่าย และรวดเร็วยิ่งขึ้น

1.4.2 ทดสอบทฤษฎีแต่ละตัวที่นำมาประยุกต์ใช้ ในขั้นตอนของการทำ segmentation ตัวอักษรทั้งภาษาไทย และภาษาอังกฤษ

1.4.3 รวม Modules ต่างๆ เข้าเป็นระบบ การทำงานของระบบ และการใช้งานโปรแกรม ได้อธิบายไว้ในบทที่ 3

1.4.4 ทดสอบการทำงาน รวมทั้งทำรายงานผลการทดลอง ซึ่งผลการทดลองทั้งหมดในการทำ segment ตัวอักษรทั้งภาษาไทย และภาษาอังกฤษ ได้สรุปไว้ในบทที่ 4



บทที่ 2

ทฤษฎีการแยกตัวอักษร

2.1 การหากรอบตัวอักษรด้วยวิธี Line Crossing

เทคนิคนี้จะเป็นวิธีแรกของผู้ที่กำลังคิดเริ่มทำโครงการ OCR เสมอ โดยอาศัยหลักการง่าย ๆ คือ พยายามแบ่งอักษรแต่ละตัวออกมาให้ได้มากที่สุด โดยเริ่มจากการสแกนเป็นแนวเส้นตรง ทั้งแนวนอน และแนวตั้ง ตัดไปตามช่องว่างทำให้เกิดเป็นตารางกรอบตัวอักษร ซึ่งตารางที่เล็กที่สุดที่กรอบตัวอักษร อยู่ถือเป็นการ segment ได้หนึ่งอักษร ดังรูปที่ 2-1

แนวคิดของวิธีการ Line Crossing ดังรูปที่ 2-1 มีขั้นตอนดังต่อไปนี้

1. สแกนเป็นเส้นตรงตามแนวนอน
2. สแกนเป็นเส้นตรงตามแนวตั้ง
(ขั้นตอนที่ 1 กับ 2 สามารถที่จะทำขั้นตอนใดก่อนก็ได้)
3. คัดลอกข้อมูลที่อยู่ในกรอบที่มีอักษร เก็บไว้ใน buffer เพื่อนำไปแสดงผล



รูปที่ 2-1 แสดงการแยกตัวอักษรด้วยวิธี Line Crossing

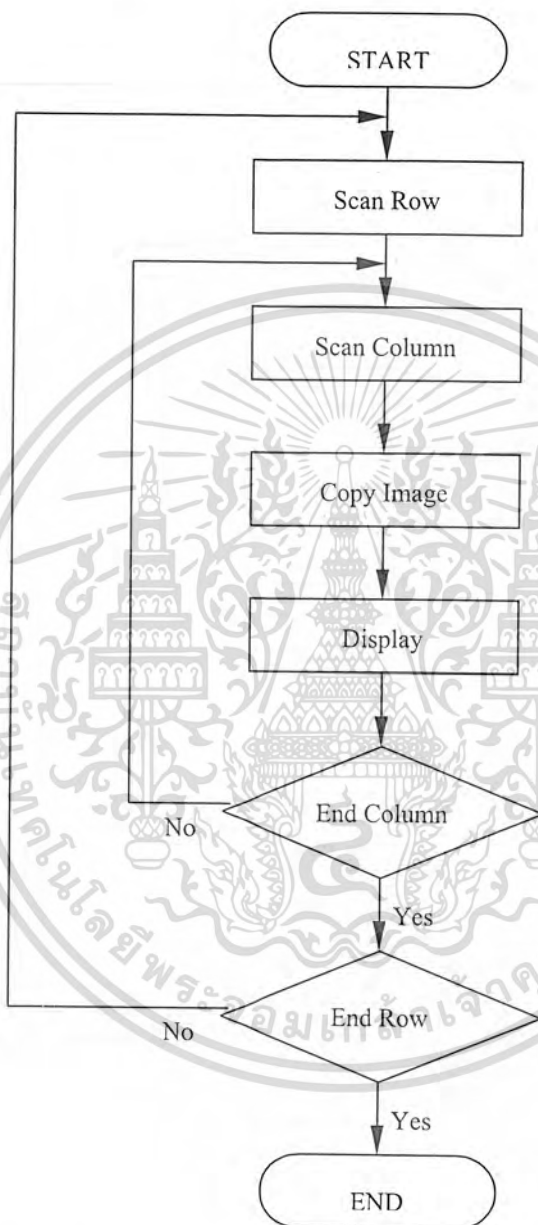
อธิบาย Flowchart การแยกตัวอักษรด้วยวิธีการ Line Crossing จากรูปที่ 2-2

1. Scan Row : หาช่องว่างในแนวนอน
2. Scan Column : หาช่องว่างในแนวตั้ง
3. Copy Image : คัดลอกข้อมูลของส่วนที่ตัดได้
4. Display : แสดงรูปที่ทำการตัดได้
5. End Column : การ scan ข้อมูลในแนวตั้งครบหมด
 - Yes : End Row
 - No : Scan Column

6. End Row : การ scan ข้อมูลในแนวนอนครบหมด

- Yes : END

- No : Scan Row



รูปที่ 2-2 แสดง Flowchart การแยกตัวอักษรด้วยวิธี Line Crossing

ข้อดี ของการทำ Line Crossing ก็คือ ง่ายต่อการเขียนโปรแกรม และมีประสิทธิภาพในการแยก อักษรกับแบบอักษรที่มีข้อผิดพลาดน้อย ตัวอักษรไม่ติดกันมาก และถ้าเป็นการ segment ตัวอักษร ภาษาอังกฤษ จะได้ผลลัพธ์ที่ดีกว่าการ segment ตัวอักษรภาษาไทย

ข้อเสีย ของวิธีการนี้ที่เห็นได้ชัดเจนนก็คือ ในการทำ OCR ภาษาไทยสิ่งที่เกิดขึ้นอยู่บ่อยครั้งก็คือ อักขระจะมีการเหลื่อมล้ำกัน ไม่ว่าจะเป็นสระเหลื่อมเข้าไปในตัวพยัญชนะ หรือพยัญชนะเหลื่อมล้ำกันเอง เทคนิค Line Crossing จะไม่สามารถแยกตัวอักขระออกจากกันได้เลย

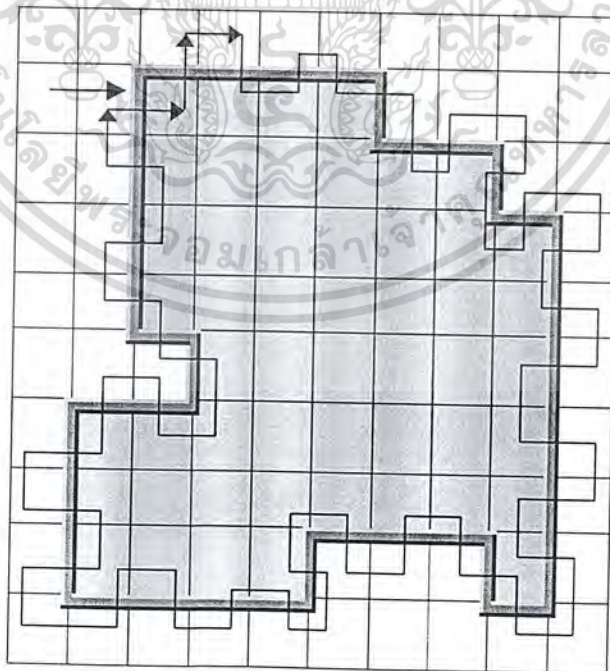
2.2 เทคนิคการตามรอยขอบภาพ (Contour Following)

เทคนิคการตามรอยขอบภาพ จะช่วยให้เราทำการ segment ได้ง่ายขึ้น โดยไม่ต้องคำนึงถึงรูปร่างของวัตถุที่จะทำการ segment เมื่อพิจารณาขอบของวัตถุจะมีค่าสีที่มีความแตกต่างกันกับสีของพื้นหลัง เช่น ภาพตัวอักษรสีดำที่วางอยู่บนพื้นกระดาษสีขาว โดยเนื้อตัวอักษรแทนด้วยค่าข้อมูลที่เป็นสีดำ และพื้นกระดาษแทนด้วยค่าข้อมูลที่เป็นสีขาว

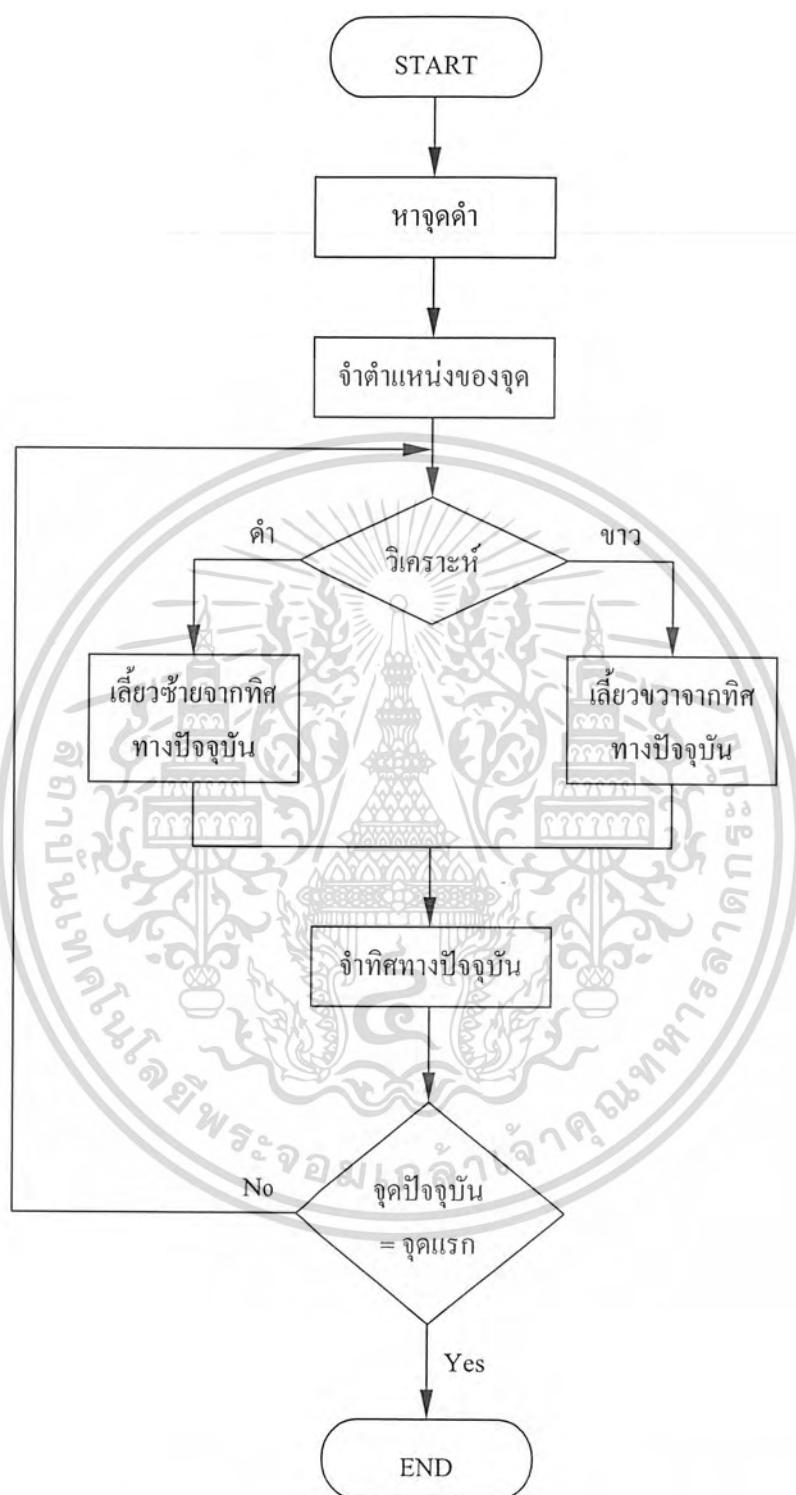
การทำงานของเทคนิคนี้ เป็นการเดินไปตามขอบระหว่างส่วนที่เป็นรูปภาพ กับส่วนที่เป็นพื้นหลัง รูปที่ 2-3 แสดงการตามรอยขอบภาพ เพื่อแยกวัตถุออกจากพื้นกระดาษด้วยแนวคิดดังนี้

แนวคิดในการทำ Contour Following ตามรูปที่ 2-3 ทำได้ตามขั้นตอนต่อไปนี้

1. จุดภาพปัจจุบันมีข้อมูลเป็นสีดำ ให้เลี้ยวซ้ายจากทิศทางปัจจุบัน
2. จุดภาพปัจจุบันมีข้อมูลเป็นสีขาว ให้เลี้ยวขวาจากทิศทางปัจจุบัน
3. สิ้นสุดเมื่อจุดภาพปัจจุบัน อยู่ที่ตำแหน่งเดียวกับจุดเริ่มต้นพอดี



รูปที่ 2-3 แสดงการ segment ด้วยวิธี Contour Following



รูปที่ 2-4 แสดง Flowchart การ segment ด้วยวิธี Contour Following

อธิบาย Flowchart การแยกอักขระด้วยวิธีการ Contour Following จากรูปที่ 2-4

1. หาจุดดำ : เป็นการหาจุดดำจุดแรกที่พบ โดยการ scan ตามแนวนอน
2. จำตำแหน่งของจุด : เป็นการจำตำแหน่งของจุดดำที่พบ

3. วิเคราะห์จุด : มี 2 กรณี

- ขาว : เลี้ยวขวาจากทิศทางปัจจุบัน
- ดำ : เลี้ยวซ้ายจากทิศทางปัจจุบัน

4. จำทิศทางปัจจุบัน : จดจำทิศทาง และตำแหน่งของจุดปัจจุบัน

5. จุดปัจจุบัน = จุดแรก : เป็นการเปรียบเทียบตำแหน่งปัจจุบัน กับจุดเริ่มต้น

- Yes : END
- No : วิเคราะห์จุด

ข้อดี ของการทำวิธี Contour following ก็คือ เราไม่ต้องสนใจว่ารูปร่างของตัวอักษรจะมีลักษณะเป็นอย่างไร เมื่อเราวิ่งไปตามขอบของภาพ เราก็จะได้ขอบของตัวอักษรออกมา

ข้อเสีย ของการทำวิธี Contour following ก็คือ ถ้าอักษรมีส่วนที่ติดกันอยู่ สิ่งที่เราได้มาก็คือกลุ่มของอักษรที่ติดกัน โดยวิธีนี้จะไม่สามารถแยกอักษรที่ติดกันออกจากกันได้

2.3 Contour with matrix

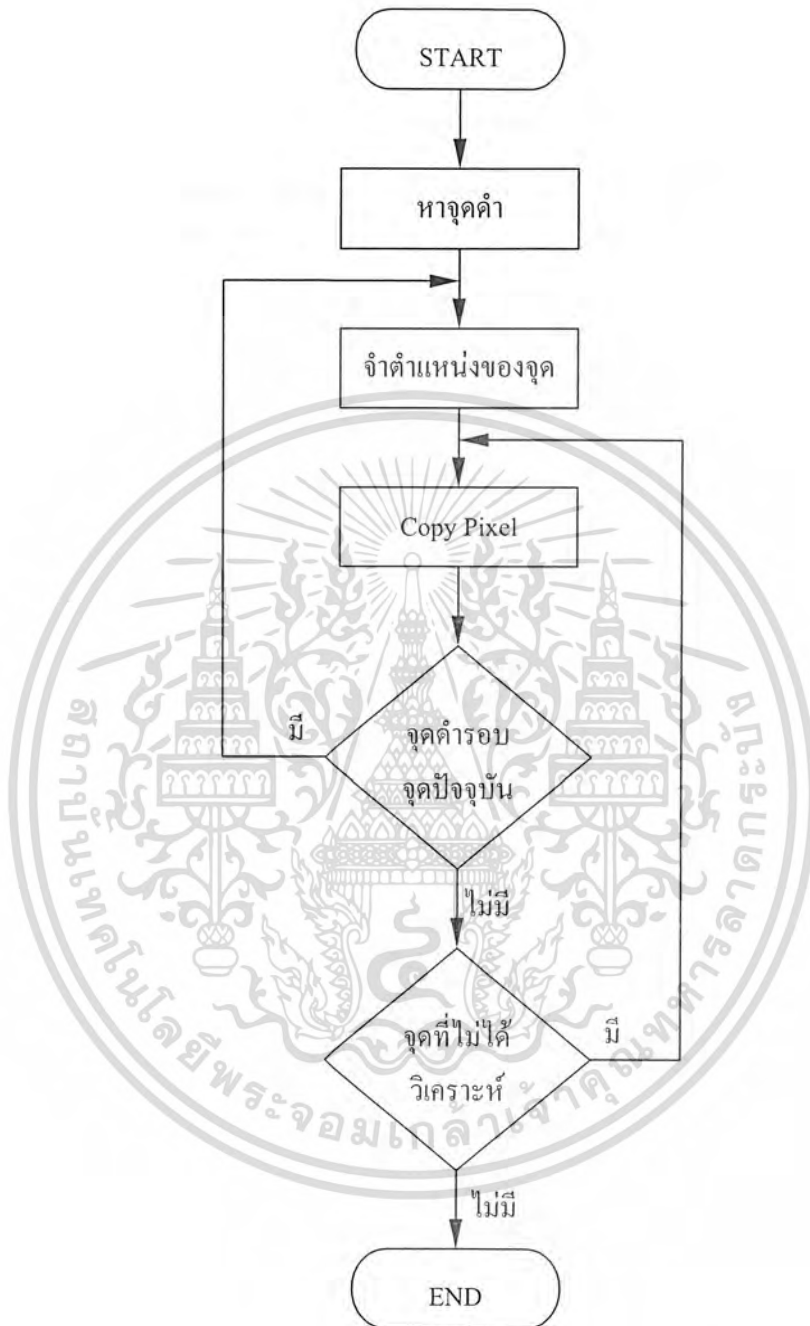
เทคนิคนี้คล้ายกับการทำ Contour Following แต่จะต่างกันตรงที่ เทคนิคนี้จะมีการดึงเอาข้อมูลภาพที่อยู่ในขอบเขตของการทำ contour ไปใช้ด้วย ทำให้เราได้ตัวอักษรทั้งตัวไปใช้งาน และการใช้งานไม่เพียงแค่ว่าข้อมูลตรงจุดนั้นเป็น 0 หรือ 1 แต่จะใช้ matrix ช่วยตรวจสอบแทน matrix ที่มีขนาด 3 x 3 ดังรูปที่ 2-5

P1	P2	P3
P4	P5	P6
P7	P8	P9

รูปที่ 2-5 matrix ขนาด 3x3

แนวคิดของการทำ Contour with matrix ดังรูปที่ 2-6 มีขั้นตอนการทำงานดังต่อไปนี้

1. ให้กำหนดจุดดำที่พบเป็นจุดกลางของ matrix (P5) คัดลอกจุดนี้ลงใน buffer แล้วลบจุดกลางนั้นออกจากรูปภาพ
2. หาจุดกลางถัดไป โดยตรวจสอบข้อมูลรอบๆ จุดกลางปัจจุบัน (P5) ซึ่งเป็นจุดสีดำ คือ ข้อมูลตัวอักษร การตรวจสอบจะเริ่มจากด้านบนซ้ายก่อน คือ ตำแหน่ง P1, P2, P3, P4, P6, P7, P8 และ P9 ตามลำดับ กำหนดจุดดำที่พบจุดแรกให้เป็นจุดกึ่งกลาง matrix
3. เมื่อรอบๆ จุดกลางนั้นไม่มีจุดดำแล้ว จะทำการ return กลับไปยังจุดกลางตัวก่อน
4. การทำงานจะเป็นเช่นนี้ไปเรื่อยๆ จนกระทั่งไม่มีจุดดำรอบจุดกลางทุกตัว



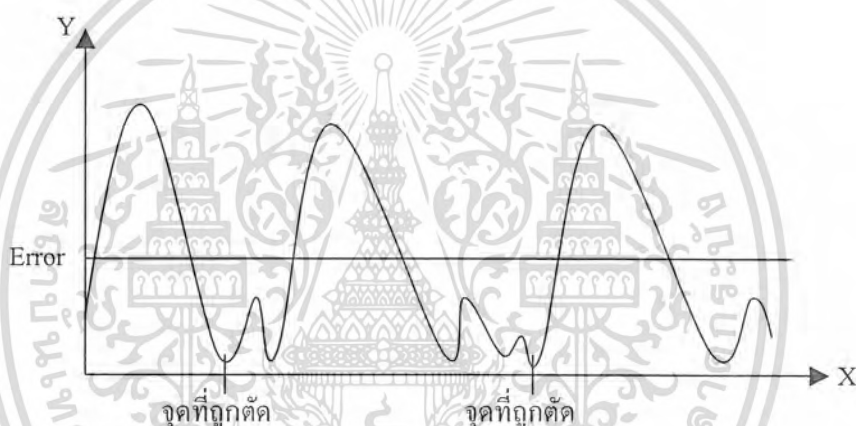
รูปที่ 2-7 แสดง Flowchart การ segment ด้วยวิธี Contour with Matrix

ข้อดี ของวิธี Contour with Matrix ก็คือ สามารถดึงเอาส่วนที่เป็นเนื้อตัวอักขระออกมาได้ทั้งหมด เป็นประโยชน์ต่อการนำไปใช้ในกระบวนการอื่นๆ ที่เนื้อของข้อมูลเป็นสิ่งจำเป็น

ข้อเสีย ของวิธี Contour with Matrix ก็คือ เมื่อเราต้องดึงเอาข้อมูลทุกส่วนของตัวอักขระออกมา ก็จะทำให้เราต้องใช้เวลาในการคัดลอก และลบข้อมูล วิธีการนี้ทำงานแบบ recursive ถ้ามีการ backtracking มากเกินไปอาจทำให้ stack overflow ได้

2.4 เทคนิคการแยกตัวอักษรที่ติดกันโดยใช้ Histogram

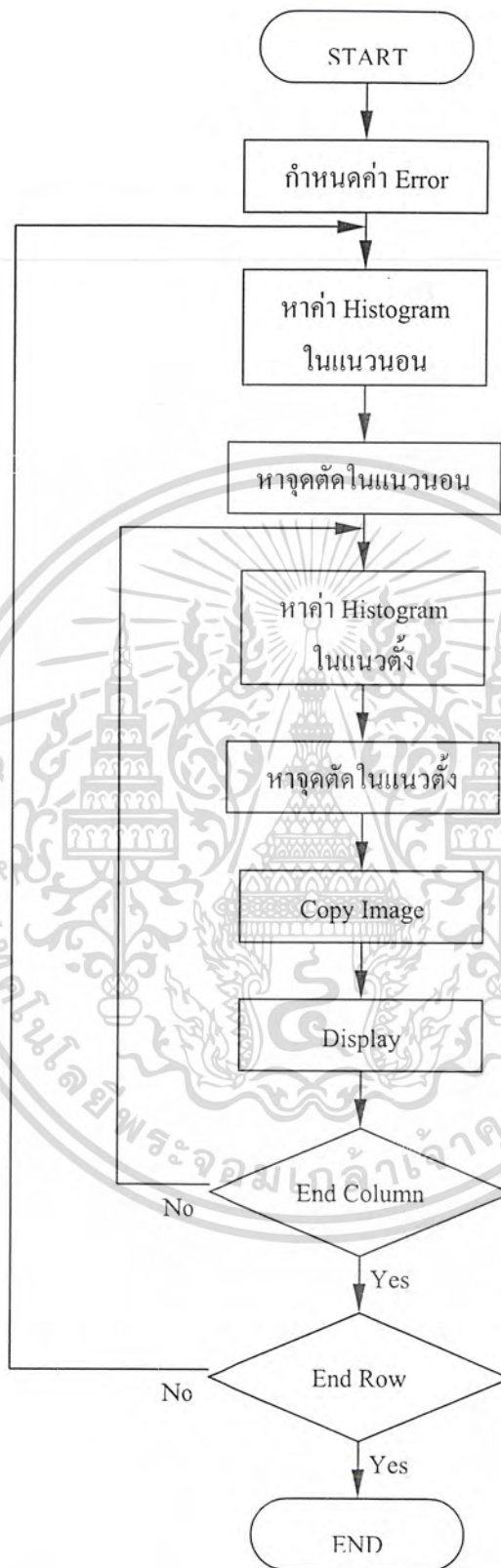
Histogram ใช้เพื่อวัดความหนาแน่นของข้อมูลภาพในช่วงที่กำหนด โดยค่าความหนาแน่นที่ใช้เป็นจำนวนจุดดำ เพื่อช่วยในการวิเคราะห์ภาพตัวอักษร โดยเก็บเป็นค่าเปอร์เซ็นต์ Error เพื่อนำไปใช้เปรียบเทียบกับตัวอักษรแต่ละตัว เพื่อหาตัวอักษรที่คาดว่าจะมีการติดกันอยู่ และการใช้ histogram จะช่วยให้โปรแกรมสามารถตัดสินใจได้ว่า ควรที่จะตัดส่วนที่ติดกันของอักษรที่ตรงส่วนไหน เพื่อให้สามารถแยกตัวอักษรที่ติดกัน ให้ออกมาเป็นตัวอักษรเดี่ยวให้ได้มากที่สุด รูปที่ 2-8 แสดงการเลือกจุดที่จะทำการแยกตัวอักษรออกจากกัน



รูปที่ 2-8 แสดงการกำหนดจุดตัดด้วย histogram

การแยกส่วนที่ติดกันของตัวอักษร โดยใช้ histogram มีวิธีการดังนี้

1. ทำการหาค่า histogram ของตัวอักษรที่สามารถแยกออกมาจากรูปภาพได้
2. กำหนดค่าเปอร์เซ็นต์ error เพื่อใช้ในการแยกตัวอักษรออกจากกัน
3. ทำการแยกตัวอักษรตรงส่วนที่มีค่าความหนาแน่นเท่ากับ หรือต่ำกว่าค่าเปอร์เซ็นต์ error ที่ได้กำหนดไว้ โดยจะเลือกแยกตัวอักษรตรงส่วนที่มีค่าความหนาแน่นต่ำที่สุด
4. ถ้าค่าความหนาแน่นที่ต่ำที่สุดมีหลายตำแหน่ง จะใช้ตำแหน่งที่อยู่ตรงกลาง



รูปที่ 2-9 แสดง Flowchart การแยกตัวอักษรที่ติดกันด้วยการใช้ Histogram

อธิบาย Flowchart การแยกตัวอักษรที่ติดกันด้วยการใช้ Histogram จากรูปที่ 2-9

1. กำหนดค่า Error : เป็นการกำหนดค่าเปอร์เซ็นต์ error เพื่อใช้ในการแยกตัวอักษร
2. หาค่า Histogram ในแนวนอน : หาค่าของ histogram โดยหาค่าความหนาแน่นในแนวนอน
3. หาจุดตัดในแนวนอน : เลือกตำแหน่งในการแยกตัวอักษรออกจากกันแนวนอน
4. หาค่า Histogram ในแนวตั้ง : หาค่าของ histogram โดยหาค่าความหนาแน่นแนวตั้ง
5. หาจุดตัดในแนวตั้ง : เลือกตำแหน่งในการแยกตัวอักษรออกจากกันแนวตั้ง
6. Copy Image : คัดลอกข้อมูลของส่วนที่ตัดได้
7. Display : แสดงรูปที่ทำการตัดได้
8. End Column : การ scan ข้อมูลในแนวตั้งครบหมด
 - Yes : End Row
 - No : หาค่า Histogram ในแนวตั้ง
9. End Row : การ scan ข้อมูลในแนวนอนครบหมด
 - Yes : End
 - No : หาค่า Histogram ในแนวนอน

2.5 การแยกภาพตัวอักษรออกจากข้อมูลภาพ

วิธีการนี้เป็นการเสริมการทำงาน ของระบบการแยกตัวอักษร โดยจะอาศัยเทคนิคการ Contour มาใช้ โดยจะมีการคัดลอกเอาตัวอักษรที่ scan เสร็จเก็บไว้ใน buffer ก่อน แล้วทำการลบตัวอักษรที่คัดลอกไว้แล้วออกจากข้อมูลภาพ ทั้งนี้ก็เพื่อทำให้เกิดพื้นที่ว่างมากที่สุด และจะทำให้การ segment ทำได้ง่ายยิ่งขึ้น และการเก็บอักขระแต่ละตัว จะเป็นการเรียงลำดับตามตำแหน่งก่อนหลังของตัวอักษร ทำให้เวลาทำ OCR แล้วอักขระเหล่านั้นจะถูกวางอยู่ในตำแหน่งที่เหมาะสม

แนวคิดการแยกภาพตัวอักษรออกจากข้อมูลภาพ ดังรูปที่ 2-10 มีวิธีดังนี้

1. Scan ตามคิงจากบนลงล่าง และจากซ้ายไปขวา
2. เมื่อเจอข้อมูลของตัวอักษร ให้ทำการกำหนดจุดเริ่มต้นการ contour
3. ทำขั้นตอนการ contour แล้วคัดลอกตัวอักษรเก็บไว้ใน buffer
4. ลบส่วนที่ทำการ segment แล้วด้วยการแทนส่วนนั้นด้วยข้อมูลเดียวกับพื้นกระดาษ
5. นำข้อมูลในส่วนที่อยู่ใน buffer ไปใช้ประโยชน์ต่อไป
6. สิ้นสุดการ scan เมื่อกระดาษว่างเปล่า

↓
วิทยาศาสตร์

↓
วิทยาศาสตร์

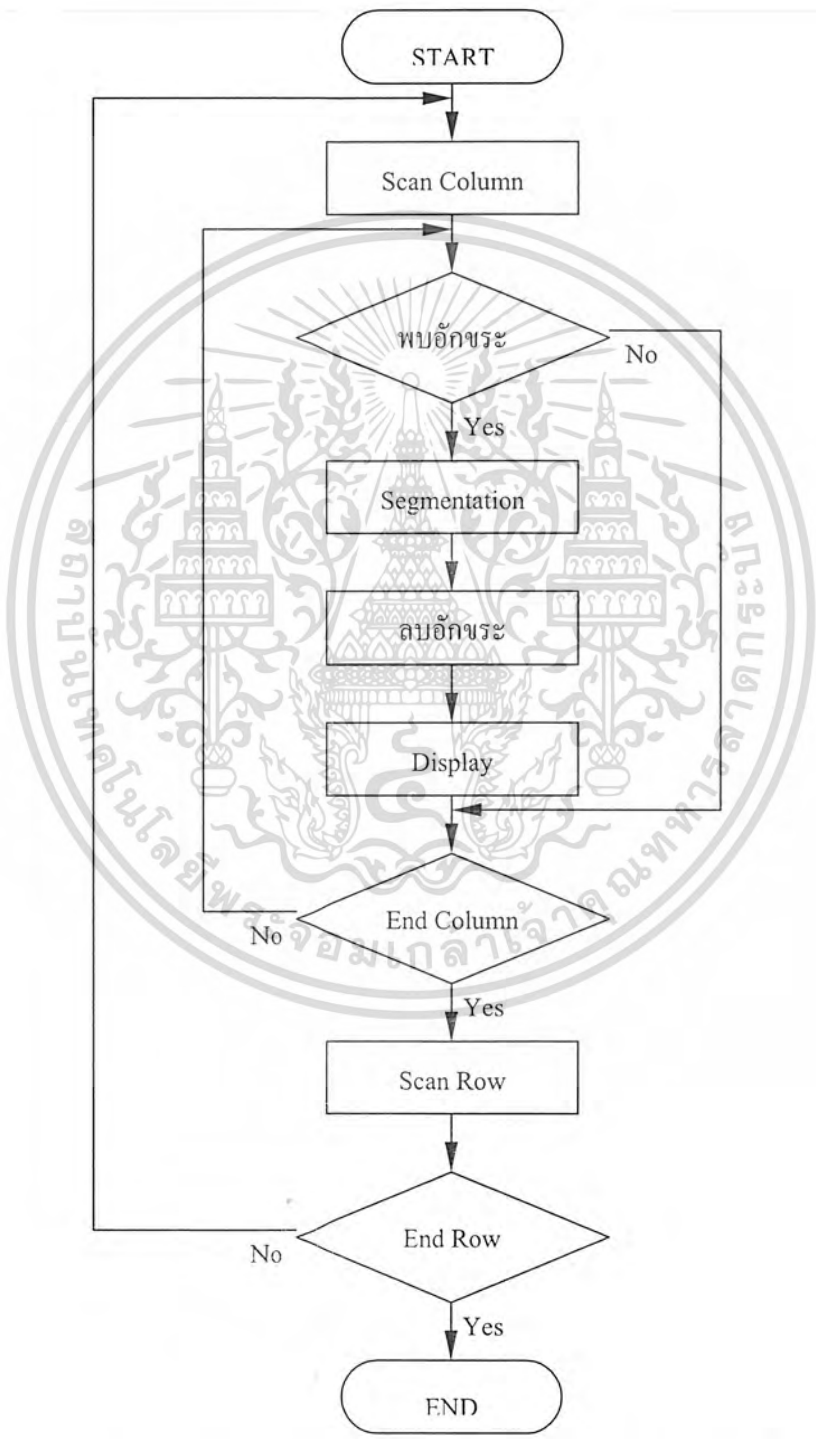
↓
วิทยาศาสตร์

รูปที่ 2-10 แสดงการแยกภาพตัวอักษรออกจากข้อมูลภาพ

อธิบาย Flowchart การแยกตัวอักษรออกจากข้อมูลภาพ จากรูปที่ 2-11

1. Scan Column : เป็นการ scan หาตัวอักษรในแนวดิ่ง
2. พบอักษร : ตรวจสอบว่าพบตัวอักษรในแนวดิ่งหรือไม่
 - Yes : Segmentation
 - No : End Column
3. Segmentation : ทำการ segment ตัวอักษร แล้วเก็บไว้ใน buffer
4. ลบอักษร : เป็นการลบตัวอักษรที่ถูก segment แล้วออกจากข้อมูลภาพ
4. Display : แสดงรูปที่ทำการตัดได้
5. End Column : การ scan ข้อมูลในแนวดิ่งครบหมด
 - Yes : Scan Row
 - No : พบอักษร

- 7. Scan Row : เลื่อนการ scan column ไปทางซ้าย
- 6. End Row : การ scan ข้อมูลในแนวนอนครบหมด
 - Yes : END
 - No : Scan Column



รูปที่ 2-11 แสดง Flowchart การแยกตัวอักขระออกจากข้อมูลภาพ

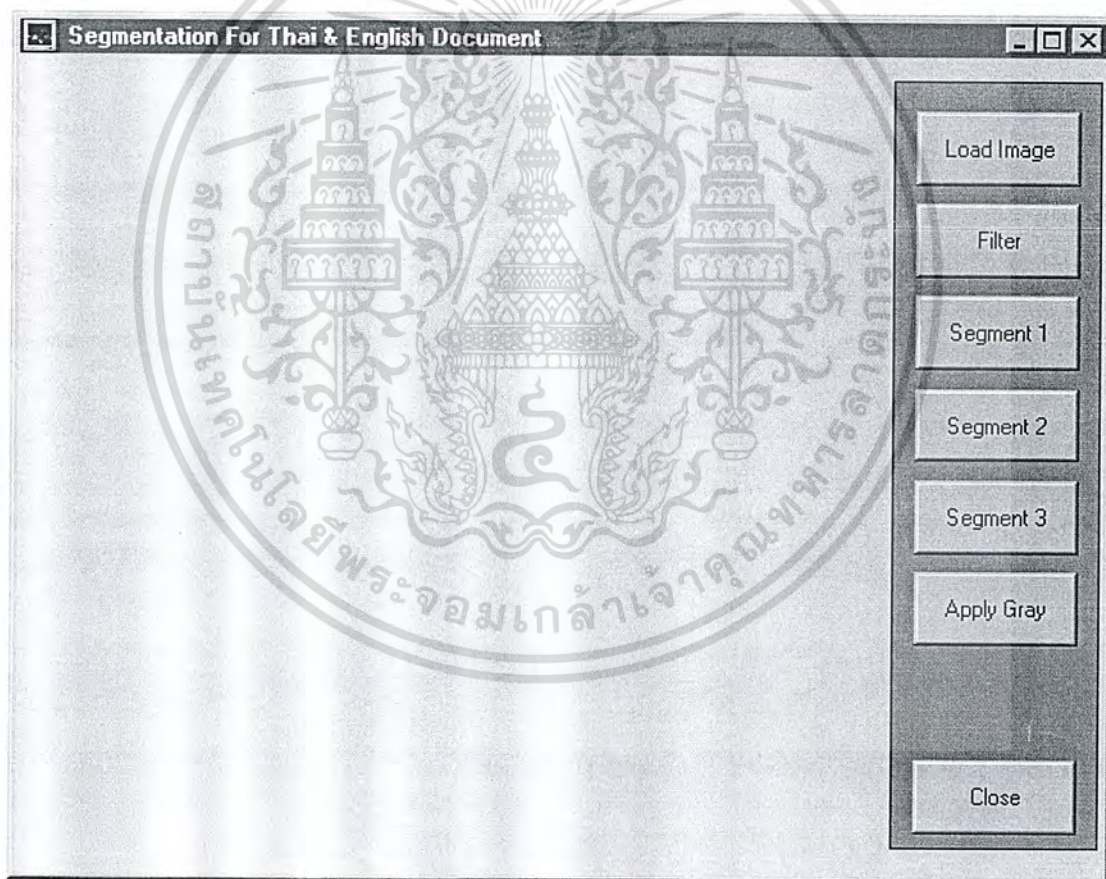
บทที่ 3

การใช้งานโปรแกรมแยกตัวอักษร

โปรแกรมแยกตัวอักษร เป็นโปรแกรมที่ทำงานบน windows และมีการสั่งงานโปรแกรมโดยผ่านทาง mouse และมีขั้นตอนการใช้งานโปรแกรม เพื่อใช้ในการแยกตัวอักษร ดังนี้

3.1 การเริ่มต้นใช้งานโปรแกรม

จากรูปที่ 3-1 เป็นหน้าจอเริ่มต้น เมื่อทำการ Run โปรแกรม โดยด้านขวาของหน้าจอเป็นปุ่มควบคุมการทำงานของโปรแกรม คือ Load Image , Filter , Segment 1 , Segment 2 , Segment 3 , Apply Gray และ Close โดยปุ่มต่างๆ มีหน้าที่การทำงานดังนี้



รูปที่ 3-1 แสดงหน้าจอการทำงานหลัก

Load Image : ใช้ในการอ่านข้อมูลจากไฟล์รูปภาพ เพื่อนำมาแสดงผลที่หน้าจอ และนำรูปภาพนั้นไปใช้ในการทำ segment ตัวอักษร

Filter : ใช้ในการลด noise ของรูปภาพ เพื่อช่วยให้ผลลัพธ์ในการทำ segment ตัวอักษรมีความถูกต้องสูงขึ้น

Segment 1 : ใช้ในการทดสอบ Histogram

Segment 2 : การทำ segment ตัวอักษร โดยใช้วิธี Contour Following ในการแยกตัวอักษรออกจากหน้าเอกสาร และใช้ histogram ในการแยกตัวอักษรที่ติดกัน

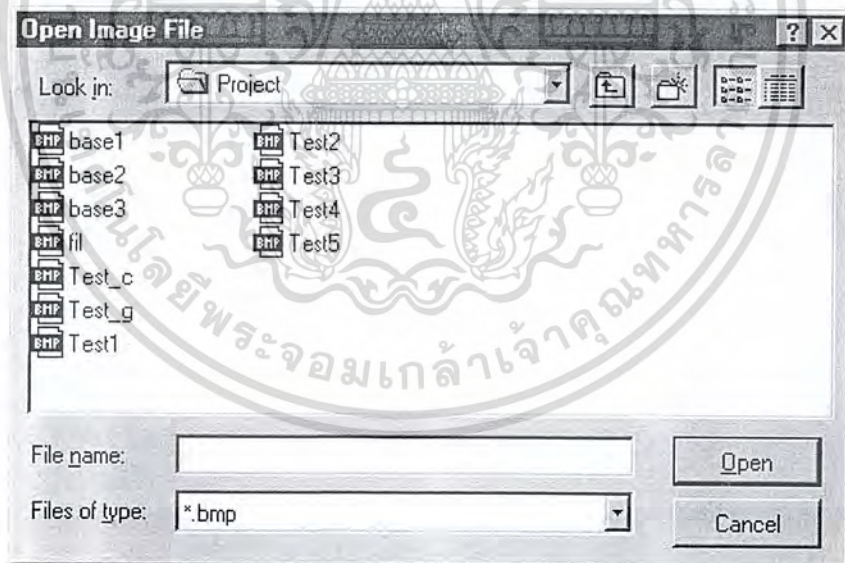
Segment 3 : การทำ segment ตัวอักษร โดยใช้วิธี Contour with Matrix ในการแยกตัวอักษรออกจากหน้าเอกสาร และใช้ histogram ในการแยกตัวอักษรที่ติดกัน

Apply Gray : การใช้วิธีของ Contour with Matrix ในการ segment รูปภาพที่เป็น gray level

Close : ออกจากการทำงานของโปรแกรม

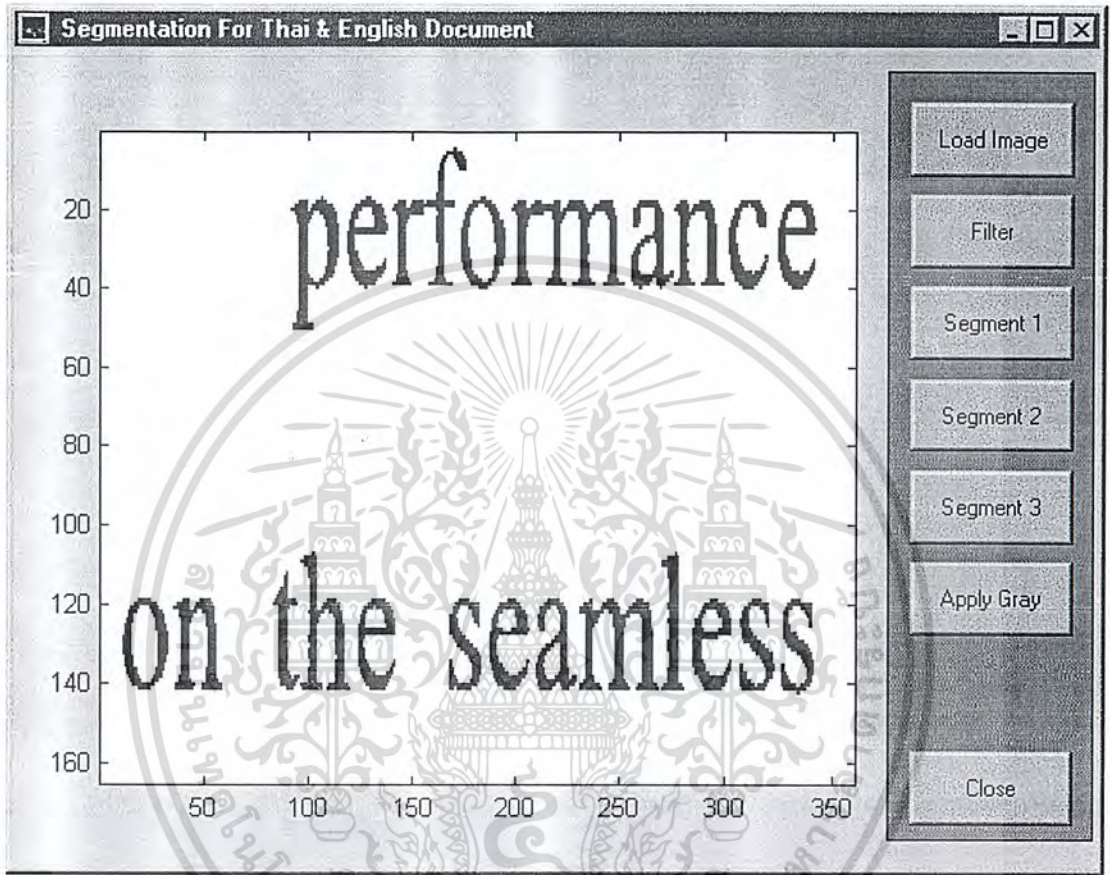
3.2 การอ่านข้อมูลจากไฟล์รูปภาพ

การอ่านข้อมูลจากไฟล์รูปภาพ ทำได้โดยการกดที่ปุ่ม Load Image โดยไฟล์รูปภาพที่โปรแกรมสามารถใช้งานได้ คือ ไฟล์รูปภาพชนิด BMP



รูปที่ 3-2 แสดงหน้าจอการ Load Image

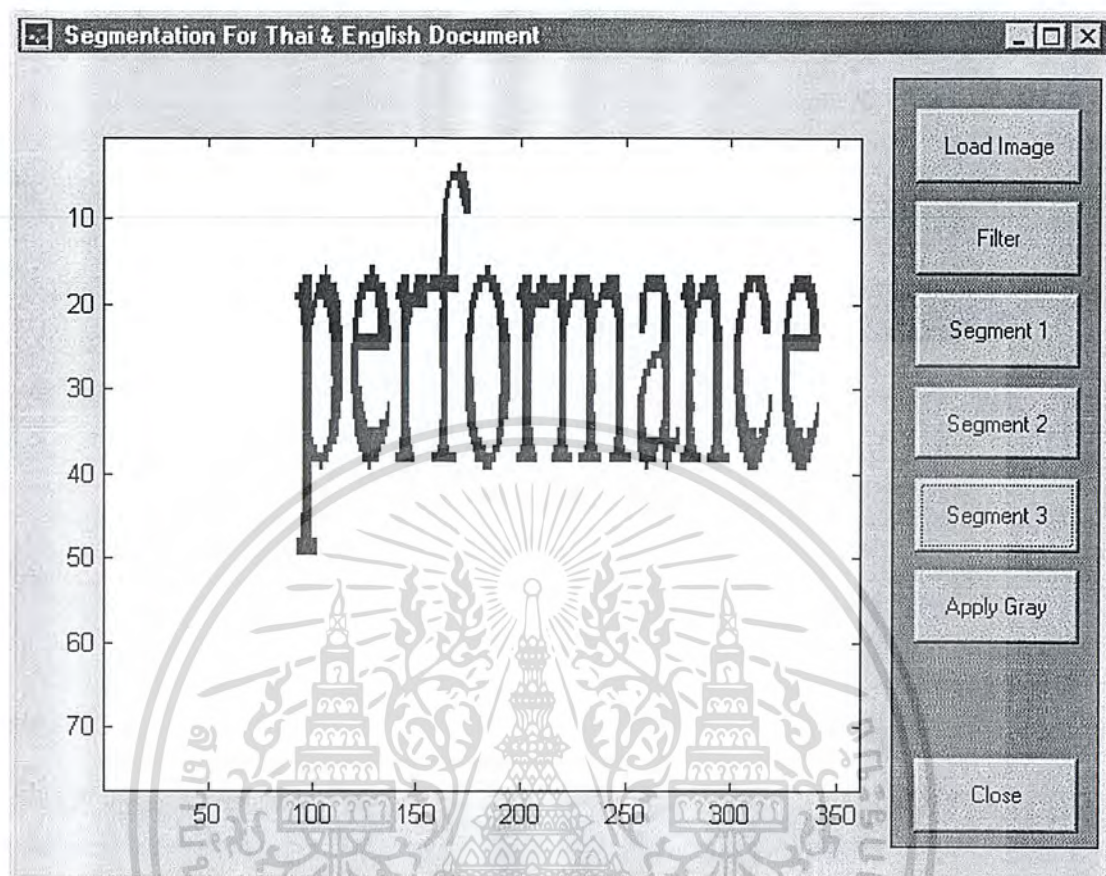
เมื่อทำการกดปุ่ม Load Image จะเกิด windows ดังรูปที่ 3-2 โดยการอ่านข้อมูลจากไฟล์รูปภาพ ให้ใส่ชื่อไฟล์ที่ช่อง File name หรือใช้ mouse เลือกไฟล์ที่ต้องการ และกดปุ่ม Open เพื่อทำการอ่านข้อมูล จากไฟล์ เมื่อทำการ Load Image เรียบร้อยแล้ว จะแสดงผลทางจอภาพ ดังรูปที่ 3-3



รูปที่ 3-3 การแสดงรูปที่ทำการอ่านข้อมูลจากไฟล์

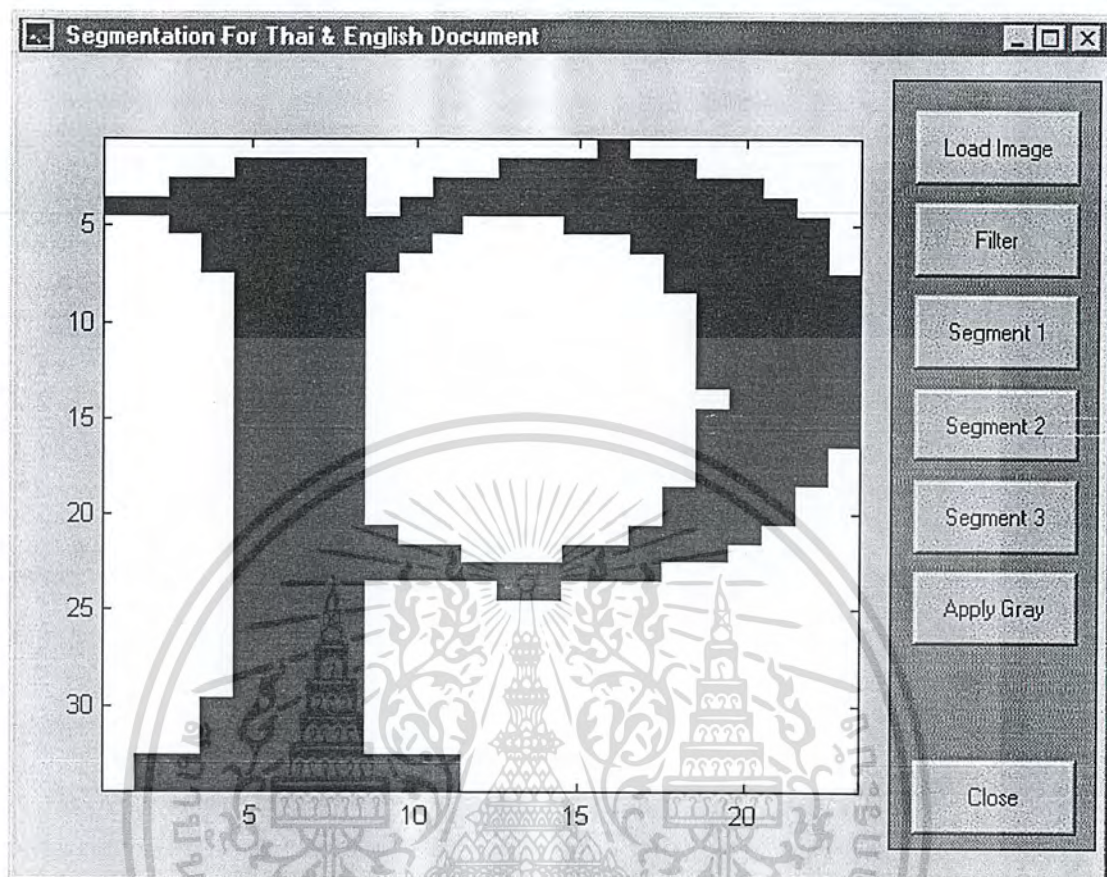
3.3 การทำ segment ตัวอักษร

เมื่อทำการ Load Image แล้ว ในการทำ segment ตัวอักษรให้กดปุ่ม Segment 2 หรือ Segment 3 ซึ่งทั้ง 2 เป็นการทำ segment ตัวอักษรเหมือนกัน แต่วิธีที่ใช้ในการทำ segment ต่างกัน การแสดงผลที่ หน้าจอ เมื่อทำการ segment ตัวอักษรแสดงในรูปที่ 3-4



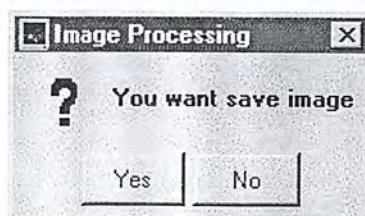
รูปที่ 3-4 การแสดงผลการแยกบรรทัดจากทำการ Segment

ในรูปที่ 3-4 แสดงผลการแยกบรรทัด เพื่อนำไปใช้ในการทำ segment ตัวอักษรต่อไป ในขั้นตอนของการทำ segment ตัวอักษร จะมีการหยุดรับ Key Enter ในการแสดงผลแต่ละครั้ง เพื่อให้สามารถตรวจสอบผลลัพธ์ในการทำ segment ตัวอักษรได้ เมื่อทำการกด Key Enter โปรแกรมจะทำการ segment ตัวอักษรจากบรรทัดที่ตัดได้ การแสดงผลที่หน้าจอ เมื่อทำการ segment ตัวอักษร แสดงในรูปที่ 3-5



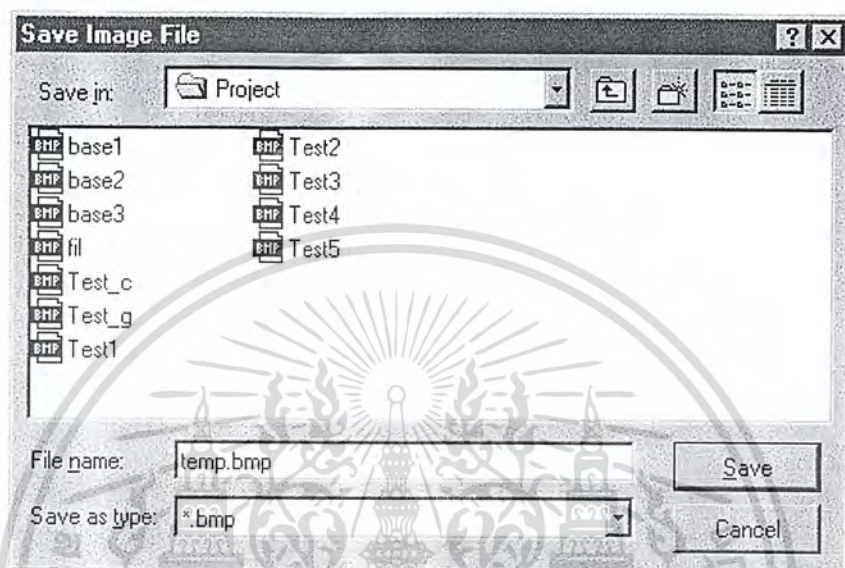
รูปที่ 3-5 การแสดงผลของการทำ Segment ตัวอักษร

การแสดงผลในรูปที่ 3-5 เป็นการแสดงตัวอักษรที่ segment ได้ในแต่ละครั้ง และเมื่อทำการ segment ตัวอักษรได้ในแต่ละครั้ง จะมีการแสดง Question Box เพื่อใช้ในการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล การแสดง Question Box ที่ใช้ในการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล แสดงในรูปที่ 3-6



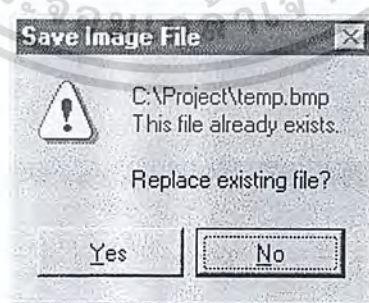
รูปที่ 3-6 แสดงรูป Question Box เพื่อใช้ Save Image

เมื่อกดปุ่ม Yes ที่ Question Box จะทำการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล โดยจะแสดง windows ที่ใช้ในการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล โดยแสดงหน้าจอ ดังรูปที่ 3-7 เพื่อใช้ในการ save รูปภาพลงไฟล์ข้อมูล โดยการใส่ชื่อไฟล์ที่ต้องการที่ช่อง File name หรือ ใช้ mouse เลือกไฟล์ที่ต้องการ และกดปุ่ม Save เพื่อทำการ save รูปภาพลงไฟล์ข้อมูล



รูปที่ 3-7 แสดงหน้าจอการ Save Image

เมื่อชื่อของไฟล์ที่ต้องการ save มีชื่อที่เหมือนกัน โปรแกรมจะแสดงข้อความดังรูปที่ 3-8 เพื่อให้ทราบว่าชื่อที่เหมือนกันอยู่ และจะให้เลือกว่าต้องการทำการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล โดยใช้ชื่อเดียวกันหรือไม่



รูปที่ 3-8 แสดงข้อความเมื่อชื่อไฟล์ในการ Save Image เหมือนกัน

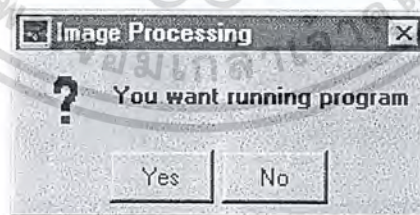
การกดปุ่ม Yes คือต้องการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล โดยใช้ชื่อเดียวกัน การกดปุ่ม No คือไม่ต้องการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูล โดยใช้ชื่อเดียวกัน และจะแสดง windows ที่ใช้ในการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูลอีกครั้ง เพื่อใส่ชื่อไฟล์ที่ต้องการใหม่

หลังจากการแสดงผล Question Box เพื่อใช้ในการ save รูปภาพที่ได้จากการทำ segment ลงไฟล์ข้อมูลแล้ว โปรแกรมจะแสดง Question Box เพื่อใช้ histogram ในการแยกตัวอักษรที่ติดกัน โดย Question Box เพื่อใช้ histogram ในการแยกตัวอักษรที่ติดกัน แสดงในรูปที่ 3-9



รูปที่ 3-9 แสดงรูป Question Box เพื่อใช้ Histogram ในการ Segment

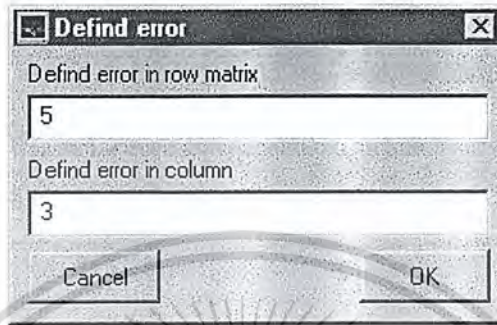
การกดปุ่ม Yes คือต้องการใช้ histogram ในการแยกตัวอักษรที่ติดกัน การกดปุ่ม No คือไม่ต้องการใช้ histogram ในการแยกตัวอักษร เมื่อทำการ segment ตัวอักษร โดยใช้ histogram ในการแยกตัวอักษรที่ติดกันแล้ว จะแสดง Question Box เพื่อใช้ในการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร โดย Question Box เพื่อใช้ในการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร แสดงในรูปที่ 3-10



รูปที่ 3-10 แสดงรูป Question Box เพื่อใช้หยุดการทำงานของโปรแกรม

จากรูป Question Box เพื่อใช้ในการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร การกดปุ่ม Yes จะทำการ segment ตัวอักษรตัวถัดไป การกดปุ่ม No เป็นการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร

การใช้ histogram เพื่อใช้ในการแยกตัวอักษรที่ติดกัน จะต้องมีการรับค่าเปอร์เซ็นต์ error ในแนวนอน และการรับค่าเปอร์เซ็นต์ error ในแนวตั้ง โดยค่าที่แสดงในตัวรับค่า จะเป็นค่าเปอร์เซ็นต์ error ที่มีค่าต่ำสุดที่โปรแกรมหาได้ โดยใช้ histogram เพื่อให้สามารถเลือกใช้ค่าเปอร์เซ็นต์ error นั้นไปตั้งค่าเปอร์เซ็นต์ error ใหม่ ดังแสดงในรูปที่ 3-11



รูปที่ 3-11 แสดงตัวรับค่าเปอร์เซ็นต์ error ในแนวนอนและแนวตั้ง

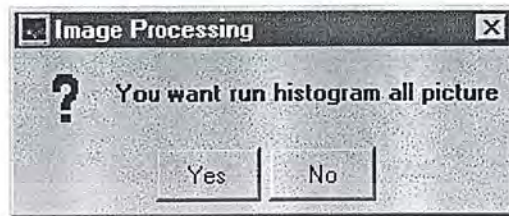
เมื่อใช้ histogram ทำการแยกตัวอักษรที่ติดกันแล้ว จะแสดง Question Box เพื่อใช้ในการ save รูปภาพที่ได้จากการทำ histogram ลงไฟล์ข้อมูล โดยการแสดง Question Box ดังรูปที่ 3-6 และจะแสดง Question Box เพื่อใช้ในการนำตัวอักษรที่แยกได้ โดยการใช้ histogram มาทำการ segment ต่อไปด้วยการใช้ histogram อีกครั้ง โดยการแสดง Question Box เพื่อใช้ในการนำตัวอักษรที่แยกได้ โดยการใช้ histogram มาทำการ segment ต่อไป แสดงในรูปที่ 3-12



รูปที่ 3-12 แสดงรูป Question Box เพื่อการ Segment ข้อมูลใหม่

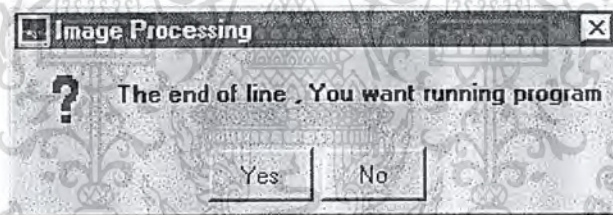
จากรูป Question Box เพื่อใช้ในการนำตัวอักษรที่แยกได้ โดยการใช้ histogram มาทำการ segment ต่อไปด้วยการใช้ histogram อีกครั้ง การกดปุ่ม Yes จะนำตัวอักษรที่แยกได้ โดยการใช้ histogram มาทำการ segment ต่อไป และการกดปุ่ม No คือไม่ต้องการที่จะนำตัวอักษรที่แยกได้ โดยการใช้ histogram มาทำการ segment ต่อไป

เมื่อทำการแยกตัวอักษรที่ติดกัน ด้วยการใช้ histogram แล้ว จะแสดง Question Box เพื่อใช้ในการทำการแยกตัวอักษรที่ติดกัน โดยการใช้ histogram กับข้อมูลเดิม โดยการแสดง Question Box เพื่อใช้ในการทำการแยกตัวอักษรที่ติดกัน โดยการใช้ histogram กับข้อมูลเดิม แสดงในรูปที่ 3-13



รูปที่ 3-13 แสดงรูป Question Box เพื่อการ Segment ข้อมูลเดิม

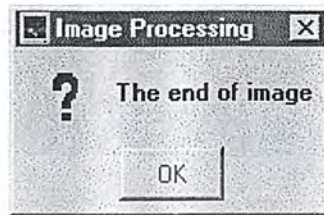
จาก Question Box เพื่อใช้ในการทำการแยกตัวอักษรที่ติดกัน โดยการใช้ histogram กับข้อมูลเดิม การกดปุ่ม Yes จะทำการแยกตัวอักษรที่ติดกัน โดยการใช้ histogram กับข้อมูลเดิม และการกดปุ่ม No คือไม่ต้องการที่จะทำการแยกตัวอักษรที่ติดกัน โดยการใช้ histogram กับข้อมูลเดิม



รูปที่ 3-14 แสดงรูป Question Box เมื่อทำการ segment จนหมดบรรทัด

เมื่อทำการ segment ตัวอักษรจนหมดในบรรทัดนั้น จะมีการแสดง Question Box เพื่อบอกให้ทราบ และเพื่อใช้ในการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร โดยการแสดง Question Box เพื่อบอกให้ทราบ และเพื่อใช้ในการหยุดการทำงานของโปรแกรมในการทำ segment ตัวอักษร แสดงในรูปที่ 3-14

จากรูปที่ 3-14 การกดปุ่ม Yes เป็นการสั่งให้โปรแกรมทำการ segment ตัวอักษรในบรรทัดถัดไป และการกดปุ่ม No เป็นการยกเลิกการทำงานของโปรแกรมในการทำ segment ตัวอักษรในส่วนที่เหลือ เมื่อโปรแกรมทำการ segment ตัวอักษรจนหมดทั้งหน้าเอกสารแล้วจะแสดง Message Box เพื่อบอกให้ทราบ ดังแสดงในรูปที่ 3-15



รูปที่ 3-15 แสดงรูป Message Box เมื่อทำการ segment จนหมดหน้าเอกสาร



บทที่ 4

ผลการทดลอง

4.1 ขั้นตอนและวิธีที่ใช้ในการทดลอง

การ Segment ตัวอักษรภาษาไทย และภาษาอังกฤษจากหน้าเอกสาร โดยการนำหน้าเอกสารที่ต้องการมาทำการ Scan เพื่อนำมาทำการ Segment โดยโปรแกรม ซึ่งโปรแกรมมีวิธีการทำ Segment ตัวอักษร 2 วิธี คือ

วิธีที่ 1 การตามรอยขอบของตัวอักษร (Contour Following)

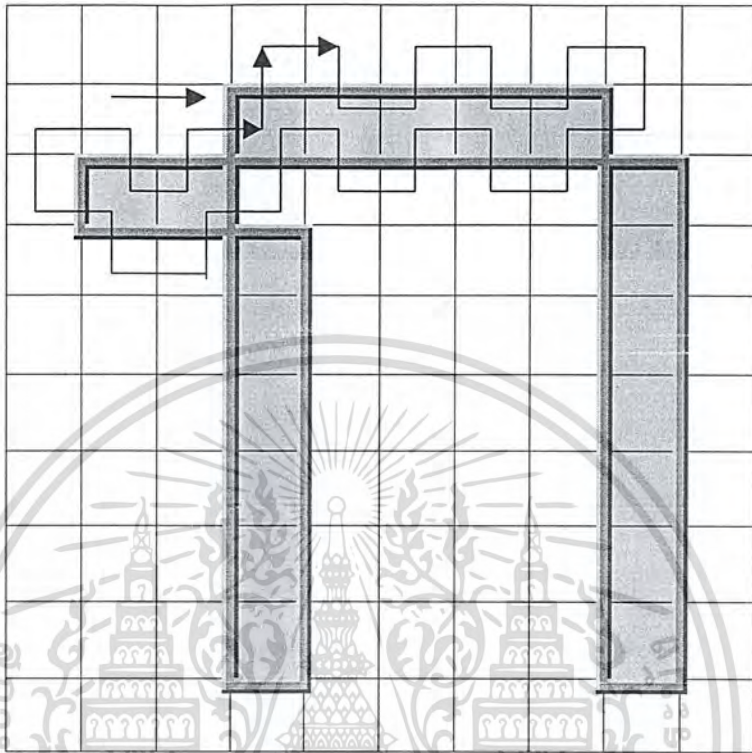
วิธีที่ 2 การตามส่วนที่เป็นเนื้อของตัวอักษร (Contour with Matrix)

จากการทดลอง เปรียบเทียบกับวิธีการ Line Crossing ซึ่งเป็นการ Scan ตามแนวตั้ง และ Scan ตามแนวนอน ผลการทดลองที่ได้จากการ Segment ตัวอักษรภาษาไทย และการ Segment ตัวอักษรภาษาอังกฤษ มีเปอร์เซ็นต์ความผิดพลาดสูงที่สุด ซึ่งไม่สามารถ Segment ตัวอักษรที่ติดกัน และตัวอักษรที่มีการ over lap กันของ pixel ได้

วิธีการ Segment	Resolution (dpi)	จำนวนตัว	ถูก		ผิด	
		อักษร	จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์
Line Crossing	150	618	582	94.2	36	5.8
Contour Following	150	618	585	94.7	33	5.3
Contour with Matrix	150	618	597	96.6	21	3.4

ตารางที่ 4-1 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย

วิธีที่ 1 เป็นการตามรอยขอบของตัวอักษร ซึ่งจากการทดลองสามารถจะทำการ Segment ตัวอักษรที่มีการ over lap กันของ pixel ได้ แต่ไม่สามารถแยกตัวอักษรที่ติดกันได้ จากการทดลองได้พบข้อเสียของวิธีการตามรอยขอบของตัวอักษรนี้ คือ ไม่สามารถใช้กับตัวอักษร ที่มีลักษณะบางได้ในบางกรณี ดังตัวอย่างในรูปที่ 4-1



รูปที่ 4-1 ตัวอย่างการทำงานผลิตผลของวิธีการตามรอยขอบของตัวอักษร

วิธีที่ 2 เป็นวิธีการตามส่วนที่เป็นเนื้อของตัวอักษร ซึ่งจากการทดลอง ได้ผลลัพธ์เหมือนกับวิธีที่ 1 คือ สามารถ Segment ตัวอักษรที่มีการ over lap กันของ pixel ได้ แต่ไม่สามารถแยกตัวอักษรที่ติดกันได้ แต่วิธีนี้สามารถใช้ได้กับตัวอักษรทุกขนาด

การทดลองในการทำ segment ตัวอักษรภาษาไทย และการทำ segment ตัวอักษรภาษาอังกฤษ โดยมีขั้นตอนในการทำการทดลองดังนี้

1. ข้อมูลอินพุต เป็นหน้าเอกสารที่ได้นำมาทำการ scan ซึ่งได้ทำการ scan ที่ค่าความละเอียด 300 dot per inch

2. ข้อมูลที่ใช้ในการทดลอง

- Font ที่ใช้ในการทดลองใช้ Font ต่างกัน 3 Font คือ Angsana New , Browallia New และ Cordia New โดยมีทั้งอักขระภาษาไทย และภาษาอังกฤษ

- ขนาดของ Font ทั้งหมดที่ใช้ในการทดลองมีขนาด 18

- การทดลองใช้ลักษณะของตัวอักษร 3 แบบ คือ ตัวอักษรปกติ ตัวเอียง และตัวหนา ทั้งตัวอักษรภาษาไทย และภาษาอังกฤษ

3. การทดลองใช้วิธีการตามรอยส่วนที่เป็นเนื้อของตัวอักษร (Contour with Matrix) ในการทำ segment ตัวอักษร และใช้ histogram เข้ามาช่วยในการแยกตัวอักษรที่ติดกัน ทั้งตัวอักษรภาษาไทย และภาษาอังกฤษ โดยการกำหนดค่าเปอร์เซ็นต์ error ของ histogram ให้มีค่าที่ต่างกัน เพื่อดูความแตกต่างของผลลัพธ์ที่ได้

4.2 ผลการทดลอง

ผลการทดลองที่แสดงในตารางมีจำนวนตัวอักษรภาษาไทยที่ใช้ในการทดลอง 618 ตัวอักษร และมีจำนวนตัวอักษรภาษาอังกฤษที่ใช้ในการทดลอง 356 ตัวอักษร โดยค่าเปอร์เซ็นต์ error ที่ใช้ในการทดลองกับอักษรภาษาไทยมี 4 ค่าคือ 0 3 5 และ 8 เปอร์เซนต์ ค่าเปอร์เซ็นต์ error ที่ใช้ในการทดลองกับอักษรภาษาอังกฤษมี 5 ค่าคือ 0 3 5 8 และ 10 เปอร์เซนต์ โดยผลลัพธ์ที่ได้ในการทำ segment ตัวอักษรภาษาไทย และภาษาอังกฤษ จะสรุปเป็นเปอร์เซ็นต์ความถูกต้อง และแสดงตัวอักษรที่เกิด error ขึ้นบ่อยๆ ซึ่งผลการทดลองเป็นดังนี้

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	616	99.7	2	0.3	-
300	3	618	612	99	6	1	๗
300	5	618	445	72	173	28	ไ ข ญ น ม บ ป ย ว
300	8	618	427	69.1	191	30.9	ไ ข ญ น ม บ ป ร ย ว

ตารางที่ 4-2 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	612	99	6	1	-
300	3	618	612	99	6	1	-
300	5	618	474	76.7	144	23.3	ก ญ น บ ห
300	8	618	466	75.4	152	24.6	ก ญ น บ ป ห

ตารางที่ 4-3 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	612	96.6	6	1	-
300	3	618	612	96.3	6	1	-
300	5	618	560	90.6	58	9.4	ไ น ป
300	8	618	554	89.6	64	10.4	ไ น ป

ตารางที่ 4-4 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	612	99	6	1	-
300	3	618	612	99	6	1	-
300	5	618	611	98.9	7	1.1	-
300	8	618	592	95.8	26	4.2	ไป

ตารางที่ 4-5 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Browallia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	610	98.7	8	1.3	-
300	3	618	609	98.5	9	1.5	-
300	5	618	607	98.2	11	1.8	-
300	8	618	581	94	37	6	ช ป

ตารางที่ 4-6 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Browallia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	612	99	6	1	-
300	3	618	612	99	6	1	-
300	5	618	611	98.9	7	1.1	-
300	8	618	598	96.8	20	3.2	ไ โป

ตารางที่ 4-7 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Browallia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	606	98.1	12	1.9	-
300	3	618	606	98.1	12	1.9	-
300	5	618	561	90.8	57	9.2	า ก ม ว
300	8	618	523	84.6	95	15.4	ไ ไก ก น ม ป ว

ตารางที่ 4-8 ผลการทดลองการทำ segment ตัวอักษรภาษาไทย Cordia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	606	98.1	12	1.9	-
300	3	618	606	98.1	12	1.9	-
300	5	618	551	89.2	67	10.8	าก น
300	8	618	526	85.1	92	14.9	โาก นป

ตารางที่ 4-9 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวเอียง Cordia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	618	586	94.8	32	5.2	-
300	3	618	586	94.8	32	5.2	-
300	5	618	586	94.8	32	5.2	-
300	8	618	579	93.7	39	6.3	ป

ตารางที่ 4-10 ผลการทดลองการทำ segment ตัวอักษรภาษาไทยตัวหนา Cordia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	340	95.5	16	4.5	-
300	3	356	344	96.6	12	3.4	-
300	5	356	333	93.5	23	6.5	h
300	8	356	267	75	89	25	a f g h m n u y
300	10	356	264	74.2	92	25.8	a f g h m n u y

ตารางที่ 4-11 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	348	97.8	8	2.2	-
300	3	356	345	96.9	11	3.1	f
300	5	356	341	95.8	15	4.2	fy
300	8	356	267	75	89	25	f m n r s u y
300	10	356	258	72.5	98	27.5	d f g m n r s u y

ตารางที่ 4-12 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	348	97.8	8	2.2	-
300	3	356	352	98.9	4	1.1	-
300	5	356	348	97.8	8	2.2	h
300	8	356	318	89.3	38	10.7	h m n u y
300	10	356	318	89.3	38	10.7	h m n u y

ตารางที่ 4-13 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Angsana New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	348	97.8	8	2.2	-
300	3	356	348	97.8	8	2.2	-
300	5	356	350	98.3	6	1.7	-
300	8	356	341	95.8	15	4.2	h
300	10	356	291	81.7	65	18.3	h m n u

ตารางที่ 4-14 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Browallia New

Resolution (dpi)	Error (%)	จำนวนตัวอักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	350	98.3	6	1.7	-
300	3	356	350	98.3	6	1.7	-
300	5	356	350	98.3	6	1.7	-
300	8	356	337	94.7	19	5.3	h
300	10	356	264	74.2	92	25.8	h m n r u

ตารางที่ 4-15 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Browallia New

Resolution (dpi)	Error (%)	จำนวนตัวอักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	352	98.9	4	1.1	-
300	3	356	352	98.9	4	1.1	-
300	5	356	352	98.9	4	1.1	-
300	8	356	356	100	0	0	-
300	10	356	343	96.3	13	3.7	h

ตารางที่ 4-16 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Browallia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	350	98.3	6	1.7	-
300	3	356	352	98.9	4	1.1	-
300	5	356	354	99.4	2	0.6	-
300	8	356	314	88.2	42	11.8	h k m n u
300	10	356	264	74.2	92	25.8	d h k m n p r u

ตารางที่ 4-17 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษ Cordia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	349	98	7	2	-
300	3	356	350	98.3	6	1.7	-
300	5	356	348	97.8	8	2.2	h
300	8	356	331	93	25	7	h n
300	10	356	229	64.3	127	35.7	b d g h m n p r u y

ตารางที่ 4-18 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวเอียง Cordia New

Resolution (dpi)	Error (%)	จำนวนตัว อักษร	ถูก		ผิด		ตัวอักษรที่เกิด error
			จำนวน	เปอร์เซ็นต์	จำนวน	เปอร์เซ็นต์	
300	0	356	347	97.5	9	2.5	-
300	3	356	347	97.5	9	2.5	-
300	5	356	350	98.3	6	1.7	-
300	8	356	339	95.2	17	4.8	h
300	10	356	295	98.9	82.9	17.1	h m n u

ตารางที่ 4-19 ผลการทดลองการทำ segment ตัวอักษรภาษาอังกฤษตัวหนา Cordia New

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

รูปที่ 4-2 ตัวอย่างเอกสารทดสอบขนาดเท่าจริง Font Angsana New ขนาด 18 ภาษาไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

รูปที่ 4-3 ตัวอย่างเอกสารทดสอบขนาดเท่าจริง Font Angsana New ขนาด 18 ภาษาอังกฤษ

จากการทดสอบการทำ Segment ตัวอักษร เปรอร์เซ็นต์ความถูกต้องของตัวอักษรภาษาไทย จะสูงกว่าเปอร์เซ็นต์ความถูกต้องของภาษาอังกฤษ ซึ่งจากการทดสอบ ตัวอักษรภาษาอังกฤษมักจะมีส่วนที่ติดกันตรงส่วนฐานของตัวอักษรเป็นส่วนใหญ่ ตัวอย่างดังรูปที่ 4-4



รูปที่ 4-4 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษที่ติดกัน

แต่ก็จะขึ้นอยู่กับลักษณะของตัวอักษรภาษาอังกฤษแต่ละแบบด้วย ตัวอักษรภาษาไทยก็จะมี ส่วนที่ติดกันของตัวอักษรเช่นกัน ตัวอย่างดังรูปที่ 4-5



รูปที่ 4-5 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ติดกัน

จากตัวอย่าง ตัวอักษรที่มักติดกันส่วนใหญ่ เป็นตัวอักษรที่มีความสูงกว่าตัวอักษรตัวอื่นๆ กับตัวสระที่อยู่ด้านบน ตัวสระที่อยู่ด้านบนด้วยกัน และตัวสระที่อยู่ด้านบนกับสระที่มีความสูง แต่ทั้งนี้ต้องขึ้นอยู่กับลักษณะตัวอักษรของ Font แต่ละแบบด้วย

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาไทย ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดช ได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาไทย ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้

เป็น ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาไทย ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

เป้า "ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้"

บติ "ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้"

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาไทย ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- ขึ้น สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 12 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- ฟุ้ง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 10 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาไทย ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- เป่า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- บัติ ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- ขึ้น สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 12 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- ฟุ้ง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 10 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคงยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความ เป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาไทย ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- พื้ง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 8 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- ขึ้น สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมายบนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาไทย ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

เป็น ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 บั๊ติ ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 ลั้ง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

ส่วน สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาไทย ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

เป็น ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 เป้า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 บัดิ ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 ส้ง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

ส่วน สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคงยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาไทย ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- เป็น ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- เป่า ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- บติ ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- สำ ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- ที่ สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 15 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- ขึ้น สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 10 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- เมื่อ สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 15 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- พื้น สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 15 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์
- พืง สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 0 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาอังกฤษ ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- fi ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- rm ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- rn ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- rf สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์
- ri สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาอังกฤษ ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

fi ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 gy ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
 tw สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Angsana New ภาษาอังกฤษ ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- fi ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- tw สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 5 เปอร์เซ็นต์
- ti สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาอังกฤษ ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- tw ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- rf สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 8 เปอร์เซ็นต์
- fi สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาอังกฤษ ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

rf ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้

ft ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Browallia New ภาษาอังกฤษ ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

rf สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 8 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาอังกฤษ ตัวอักษรปกติ ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- ft ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- rf สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 5 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาอังกฤษ ตัวอักษรตัวเอียง ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

rf ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
ft ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
tw สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 3 เปอร์เซ็นต์

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

จากเอกสารที่ใช้ในการทดสอบ Font Cordia New ภาษาอังกฤษ ตัวอักษรตัวหนา ขนาด 18 มีรูปแบบของตัวอักษรที่ไม่สามารถแยกออกจากกันได้ จากการ segment ด้วยวิธี Contour Following และ Contour with Matrix ดังนี้

- rf ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้ 8
- fi ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้ 5
- ft ไม่สามารถแยกตัวอักษรที่ติดกันด้วย histogram ได้
- tw สามารถแยกตัวอักษรที่ติดกันออกได้ โดยใช้ histogram ที่ค่าเปอร์เซ็นต์ error ในแนวนอนที่ 0 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ในแนวตั้งที่ 5 เปอร์เซ็นต์

4.3 สรุปผลการทดลอง

ค่าเปอร์เซ็นต์ error ที่เหมาะสมในการใช้ histogram เพื่อแยกตัวอักษรที่ติดกันนั้น จะขึ้นกับความละเอียดในการ scan ขนาดของตัวอักษรที่นำมา scan ความคมชัดของตัวอักษรที่นำมา scan และลักษณะของ font แต่ละแบบ ซึ่งมีผลทำให้ค่าเปอร์เซ็นต์ error ที่เหมาะสมมีค่าต่างกัน

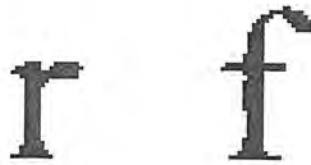
จากการทดลอง การ segment ตัวอักษรภาษาไทย การใช้ histogram ในการแยกตัวอักษรภาษาไทยที่ติดกัน ตัวอักษรภาษาไทยที่ติดกันไม่สามารถแยกได้โดยใช้ histogram เนื่องจากอักษรที่ติดกันส่วนใหญ่ในภาษาไทย มักจะเป็นการติดกันในแนวนอน ซึ่งเมื่อค่า error สูงจนสามารถแยกอักษรออกจากกันได้ ตัวอักษรปกติก็จะถูกแยกออกด้วย ซึ่งทำให้จำนวนผิดพลาดของตัวอักษรเพิ่มสูงขึ้น ซึ่งเมื่อค่า error สูงขึ้นจำนวนตัวอักษรที่ผิดพลาดในการทำ segment ก็จะมีมากขึ้น

จากการทดลอง การ segment ตัวอักษรภาษาอังกฤษ การใช้ histogram ในการแยกตัวอักษรภาษาอังกฤษที่ติดกัน วิธี histogram สามารถใช้ได้กับตัวอักษรภาษาอังกฤษ ในลักษณะของตัวอักษรปกติ และตัวหนา ในระดับเปอร์เซ็นต์ error ที่เหมาะสม ซึ่งจะช่วยให้สามารถแยกตัวอักษรที่ติดกันได้ ค่าเปอร์เซ็นต์ error ที่เหมาะสมของตัวอักษรปกติคือ 3 – 5 เปอร์เซ็นต์ และค่าเปอร์เซ็นต์ error ที่เหมาะสมของตัวอักษรตัวหนาจะอยู่ที่ 5 – 8 เปอร์เซ็นต์ ในตัวอักษรปกติจะมีค่า error ที่ต่ำกว่าตัวอักษรตัวหนา เมื่อค่า error สูงขึ้นจากระดับที่เหมาะสม จะทำให้จำนวนตัวอักษรที่ผิดพลาดที่เกิดจากการ segment สูงขึ้น และในลักษณะของตัวอักษรตัวเอียง ไม่สามารถใช้ histogram ในการแยกตัวอักษรที่ติดกันได้



รูปที่ 4-6 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษที่ใช้ในการทดลอง

จากตัวอย่างในรูปที่ 4-6 เป็นส่วนหนึ่งของหน้าเอกสารภาษาอังกฤษที่ใช้ในการทดลอง ซึ่งได้จากการ scan จากตัวอย่าง เมื่อตั้งค่า error ไว้ที่ 3 เปอร์เซ็นต์ ตัวอักษร r และ f ที่ติดกันอยู่สามารถแยกออกได้ ดังแสดงในรูปที่ 4-7



รูปที่ 4-7 แสดงตัวอักษรภาษาอังกฤษที่ติดกันที่สามารถแยกได้

ตัวอักษร r และ f ที่ติดกันอยู่สามารถแยกออกได้ เนื่องจากจำนวน pixel ที่ติดกันอยู่ระหว่างตัวอักษรจะมีจำนวน 1 pixel แต่ตัวอักษร r และ m ที่ติดกันอยู่นั้นไม่สามารถแยกได้ เนื่องจากจำนวน pixel ที่ติดกันอยู่มีจำนวน 2 pixel ซึ่งถ้าต้องการแยกตัวอักษร r และ m ออกจากกัน ต้องกำหนดค่า error ไว้ที่ 8 เปอร์เซนต์ แต่ตัวอักษร m และ n จะเกิดการผิดพลาดจากการทำ segment ที่เกิดจากการกำหนดค่า error สูงเกินไป ดังแสดงในรูปที่ 4-8



รูปที่ 4-8 แสดงตัวอักษรภาษาอังกฤษที่ติดกันที่ไม่สามารถแยกได้

จากตัวอย่างในรูปที่ 4-9 เป็นตัวอักษรภาษาอังกฤษตัวเอียง จะเห็นว่าตัวอักษร r และ f มีจำนวน pixel ที่ติดกันอยู่ระหว่างตัวอักษรเป็นจำนวน 2 pixel ซึ่งถ้ากำหนดค่า error ไว้ที่ 5 เปอร์เซนต์ จะสามารถแยกตัวอักษร r และ f ออกจากกันได้ แต่เนื่องจากเป็นอักษรตัวเอียง ทำให้ที่ฐานของตัวอักษร f อยู่ในแนวเดียวกับส่วนที่ติดกันของตัวอักษร ทำให้ไม่สามารถแยกตัวอักษร r และ f ออกจากกันได้



รูปที่ 4-9 แสดงตัวอย่างของตัวอักษรภาษาอังกฤษตัวเอียงที่ใช้ในการทดลอง

ตัวอย่างในรูปที่ 4-10 เป็นส่วนหนึ่งของหน้าเอกสารภาษาไทยที่ใช้ในการทดลอง ซึ่งได้จากการ scan จากตัวอย่างมีตัวอักษร ฐ ที่ติดกันอยู่ ซึ่งต้องใช้ histogram ช่วยในการแยกตัวอักษรที่ติดกันอยู่ออกจากกัน



รูปที่ 4-10 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ใช้ในการทดลอง

เมื่อดังค่า error ไว้ที่ 4 เปอร์เซนต์ ตัวอักษร ฐ ที่ติดกันอยู่สามารถแยกออกได้ เนื่องจากจำนวน pixel ที่ติดกันอยู่ระหว่างตัวอักษรมีจำนวน 1 pixel ดังแสดงในรูปที่ 4-11



รูปที่ 4-11 แสดงตัวอักษรภาษาไทยที่ติดกันที่สามารถแยกได้

การกำหนดค่า error ไว้ที่ 10 เปอร์เซนต์ ตัวอักษร ฐ จะเกิดการผิดพลาดจากการทำ segment ที่เกิดจากการกำหนดค่า error สูงเกินไป ดังแสดงในรูปที่ 4-12



รูปที่ 4-12 แสดงตัวอย่างการกำหนดค่า Error ที่สูงเกินไป

ตัวอย่างในรูปที่ 4-13 เป็นส่วนหนึ่งของหน้าเอกสารภาษาไทยที่ใช้ในการทดลอง ซึ่งได้จากการ scan จากตัวอย่างมีตัวอักษร ปี่ ที่ติดกันอยู่ ดังแสดงที่ในรูปข้างล่าง

ปี่

รูปที่ 4-13 แสดงตัวอย่างของตัวอักษรภาษาไทยที่ใช้ในการทดลอง

ตัวอักษร ปี่ ที่ติดกันอยู่ไม่สามารถแยกออกได้ เนื่องจากการแยกตัวอักษร ปี่ ออกจากกัน จะได้ผลออกมาดังแสดงในรูปที่ 4-14 ซึ่งต้องกำหนดค่า error ไว้ที่ 10 เปอร์เซนต์ แต่ตัวอักษร ปี่ จะเกิดการผิดพลาดจากการทำ segment ดังที่แสดงในรูปข้างล่าง

ปี่

รูปที่ 4-14 แสดงตัวอักษรภาษาไทยที่ติดกันที่ไม่สามารถแยกได้

การใช้ histogram ในการแยกตัวอักษรที่ติดกัน สามารถใช้กับตัวอักษรภาษาอังกฤษได้ผลมากกว่าตัวอักษรภาษาไทย เนื่องจากตัวอักษรภาษาอังกฤษ จะมีการติดกันของตัวอักษรในแนวตั้งเป็นส่วนใหญ่ ซึ่งทำให้สามารถแยกตัวอักษรที่ติดกันได้เป็นบางส่วน เนื่องจากทั้งภาษาไทย และภาษาอังกฤษ ตัวอักษรจะมีความสูงมากกว่าความกว้าง ทำให้ที่ค่า error เดียวกัน จำนวน pixel ที่จะถูกตัดในแนวตั้งจึงมีมากกว่าแนวนอน

จากการทดลอง ส่วนสำคัญที่มีผลทำให้ผลลัพธ์ในการทำ segment ตัวอักษรภาษาไทย และตัวอักษรภาษาอังกฤษนั้นมีความถูกต้อง คือ

1. ความคมชัดของเอกสารที่นำมาใช้ในการทำ segment ตัวอักษร เอกสารที่มีความคมชัดสูงกว่า ย่อมให้ผลลัพธ์ในการทำ segment ที่ถูกต้องมากกว่า

2. ค่าความละเอียดในการ scan การใช้ค่าความละเอียดในการ scan ที่มีค่าความละเอียดสูง ทำให้ข้อมูลของตัวอักษรที่ได้มีความชัดเจนมากขึ้น ซึ่งมีผลทำให้การ segment ตัวอักษรทำได้ง่ายขึ้น และผลลัพธ์มีความถูกต้องมากขึ้น แต่เวลาที่ใช้ในการ segment ตัวอักษรก็จะมากขึ้นด้วย

3. ลักษณะของ font ที่ใช้ในการทำ segment ตัวอักษร เนื่องจาก font แต่ละแบบมีลักษณะเฉพาะต่างกัน ทำให้ผลลัพธ์ที่ได้ในการทำ segment ตัวอักษรข้อความเดียวกันนั้นมีความถูกต้องไม่เท่ากัน

4. ขนาดของ font ซึ่งก็มีผลทำให้การ segment ตัวอักษรนั้นมีเปอร์เซ็นต์ความถูกต้องสูงขึ้น เนื่องจากขนาดของตัวอักษรที่ใหญ่ขึ้น ก็จะมีขนาดของช่องว่างระหว่างตัวอักษรที่ใหญ่ขึ้นด้วย

5. วิธีในการทำ segment ตัวอักษร วิธีที่ใช้ในการ segment บางวิธีไม่สามารถแยกตัวอักษรที่มีการ over lap กันของ pixel ได้ ซึ่งทำให้ความถูกต้องในการทำ segment ต่ำกว่าวิธีที่สามารถแยกการ over lap กันของ pixel ได้



บทที่ 5

บทสรุป

การทำงานของขั้นตอนในการทำ Segment ตัวอักษรภาษาไทย และการทำ Segment ตัวอักษรภาษาอังกฤษ ข้อมูลที่นำมาเป็นอินพุต คือหน้าเอกสารทั่วไป ที่ผ่านการ Scan เพื่อให้เครื่องคอมพิวเตอร์สามารถนำมาทำการประมวลผลได้ โดยข้อมูลอินพุตที่นำมา Scan จะถูกจัดเก็บไว้เป็นไฟล์ข้อมูลภาพ ซึ่งในการทดลองได้ทำการ Scan ที่ค่าความละเอียด 300 dpi ข้อมูลที่ใช้มีลักษณะของ Font ต่างกัน 3 Font คือ Angsana New , Browallia New และ Crodia New และมีลักษณะของตัวอักษร 3 แบบคือ ตัวอักษรปกติ ตัวเอียง และตัวหนา

5.1 ส่วนประกอบสำคัญที่ทำให้การ Segment มีความถูกต้องสูง

การประมวลผลภาพเพื่อทำการ Segment ตัวอักษร ส่วนสำคัญที่มีผลทำให้ผลลัพธ์ของการทำ Segment ตัวอักษรนั้นมีความถูกต้อง คือ

1. ความคมชัดของเอกสารที่นำมาใช้ในการทำ Segment การใช้เอกสารที่มีความคมชัดสูง ย่อมทำให้ผลลัพธ์ที่ได้จากการทำ Segment ตัวอักษรมีความถูกต้องสูงด้วย
2. ค่าความละเอียดในการ Scan มีผลกับผลลัพธ์ที่ได้พอสมควร การ Scan ที่ค่าความละเอียดสูงกว่าจะให้ผลลัพธ์ที่ดีกว่า หรือมีความถูกต้องมากกว่า
3. ลักษณะของ Font ก็เป็นส่วนประกอบสำคัญ ที่มีผลทำให้ผลลัพธ์ที่ได้มีความถูกต้อง เนื่องจากลักษณะของ Font แต่ละแบบแตกต่างกัน เช่น Font บางชนิดมีช่องว่างระหว่างตัวอักษรมาก เมื่อนำมา Scan และทำการ Segment ก็จะได้ผลลัพธ์ที่มีความถูกต้องสูง
4. ขนาดของ Font Font ที่มีขนาดใหญ่จะมีช่องว่างระหว่างตัวอักษรที่กว้างมากกว่า Font ที่มีขนาดเล็ก ซึ่งทำให้การทำ Segment ง่ายขึ้น
5. วิธีในการทำ Segment การทำ Segment ในแต่ละวิธีให้ผลลัพธ์ที่แตกต่างกัน ควรเลือกวิธีที่ให้ผลลัพธ์ที่มีความถูกต้องมากที่สุด

5.2 ข้อเสนอแนะ

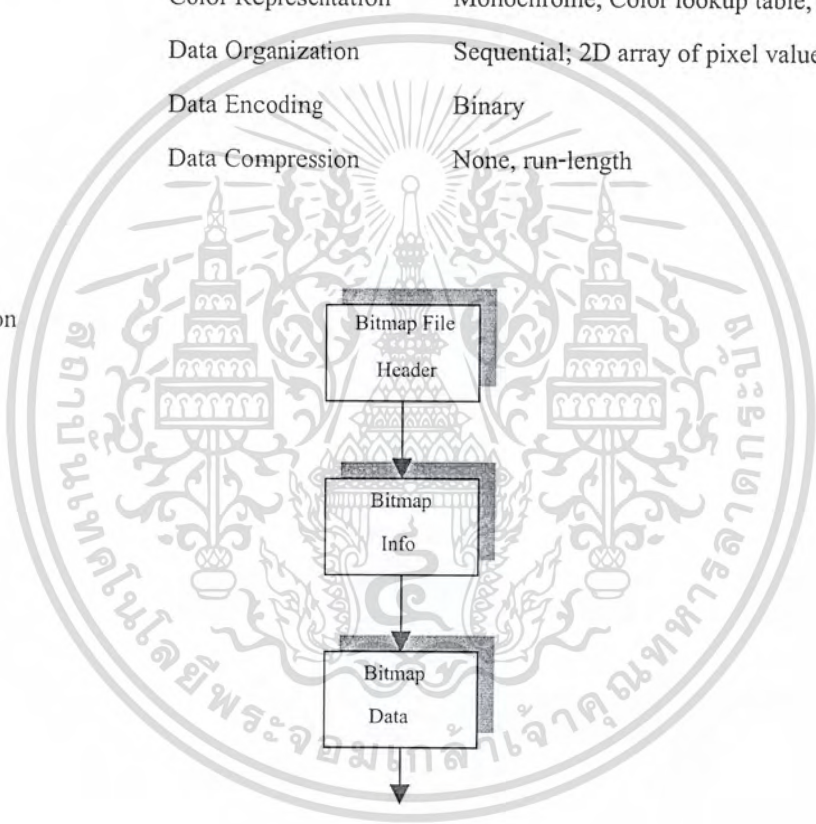
การ Segment ตัวอักษรภาษาไทย และภาษาอังกฤษ สามารถนำไปใช้ได้จริง แต่ข้อมูลภาพที่นำมาใช้ในการวิเคราะห์ และประมวลผล เป็นลักษณะของข้อมูลภาพขาวดำเท่านั้น หากสามารถจัดการกับข้อมูลภาพที่มีลักษณะเป็น Gray level หรือข้อมูลภาพที่มีลักษณะเป็นสีได้ จะเป็นการเพิ่มประสิทธิภาพในการทำงาน การ Segment ตัวอักษรยังไม่สามารถแยกตัวอักษรที่ติดกันได้ทั้งหมด หากสามารถเพิ่มวิธีการที่จะสามารถแยกตัวอักษรที่ติดกัน ให้แยกออกจากกันได้ทั้งหมด จะทำให้ผลลัพธ์ที่ได้มีความถูกต้องสูงขึ้น

ภาคผนวก ก
โครงสร้างไฟล์รูปภาพ

ไฟล์รูปภาพชนิด BMP

Acronym	BitMap format	
Version	Version 1	
Overview	Major Type of Data	2D raster
	Color Representation	Monochrome, Color lookup table, RGB
	Data Organization	Sequential; 2D array of pixel values
	Data Encoding	Binary
	Data Compression	None, run-length

Data -
Organization



Bitmap File Header

All data values are in LSB first format

Byte#	Data	Details
1-2	File type	Must be ASCII text “BM”
3-6	Size of the file	In double words (32-bit integers)
7-10	Reserved for future use	Must be zero
11-14	Byte offset to bitmap data	Offset from the Bitmap File Header

Bitmap Information

Byte#	Data	Details
1-4	Number of bytes in header	Currently 40 bytes
5-8	Width of bitmap	In pixels
9-12	Height of bitmap	In pixels
13-14	Number of color planes	Must be set to 1
15-16	Number of bits per pixel	Valid choices are 1, 4, 8, 24;if not 24, Determines the size of the palette
17-20	Type of compression	0: No compression; 1: run length (8 bits per pixel); 2: run length (4 bits per Pixel
21-24	Size of image	In bytes
25-28	Horizontal resolution	In pixels/meter
29-32	Vertical resolution	In pixels/meter
33-36	Number of color indexes Used by the bitmap	Zero indicates all colors are important
37-40	Number of color indexes Important for displaying bitmap	A value of zero indicates all colors are important
41	Blue color value	Beginning color palette (entry 0)
42	Green color value	
43	Red color value	
44	Reserved for future use	Must be zero
...	... Remaining color palette entries	4 bytes per palette entry

Bitmap Data

The pixels are stored by rows, left to right within each row. *The rows are stored bottom to top*; the origin of the bitmap is the lower-left corner. The bits of each pixel are packed into bytes and each scan line is padded with zeros, if necessary, to be aligned with a 32-bit word boundary.

File Name Suffix	WINDOWS, DOS, OS/2	.BMP
------------------	--------------------	------



ภาคผนวก ข

ข้อมูลภาษาไทย และภาษาอังกฤษที่ใช้ในการทดสอบ

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อ
 เสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระ
 พระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลาย
 นานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้าง
 ความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราช
 โองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม”
 โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และ
 แต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ
 ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้
 พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคงยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความ เป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อ เสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า “เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม” โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมายบนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อ
 เสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปร
 มินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการใน
 การวางรากฐานการพัฒนาประเทศให้มีความมั่นคง ยั่งยืน เพื่อเสริมสร้างความผาสุก
 ในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า "เรา
 จะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม" โดยได้ทรงคิดค้น
 กระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้า
 หมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง
 ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความ
 เป็นไทย

The Pentium III processor family's significant performance
 improvement over previous Intel Architecture processors is based on the
 seamless combination of the P6 microarchitecture, Internet Streaming SIMD
 Extensions, Intel MMX media enhancement technology, and a processor
 serial number. The result is higher software performance plus headroom for
 applications that take advantage of the Pentium III processor.

เป็นที่ประจักษ์แล้วว่า ตลอดระยะเวลา ๕๐ ปีที่ผ่านมา จำเดิมแต่เมื่อเสด็จขึ้นครองราชสมบัติสืบต่อจากสมเด็จพระบรมเชษฐาธิราช พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชได้ทรงบำเพ็ญพระราชกรณียกิจหลากหลายนานัปการในการวางรากฐานการพัฒนาประเทศให้มีความมั่นคงยั่งยืน เพื่อเสริมสร้างความผาสุกในทุกทางให้แก่ประชาชนชาวไทยอย่างต่อเนื่อง สมดังพระปฐมบรมราชโองการว่า "เราจะครองแผ่นดินโดยธรรม เพื่อประโยชน์สุขแห่งมหาชนชาวสยาม" โดยได้ทรงคิดค้นกระบวนการพัฒนาในรูปแบบต่าง ๆ ที่เหมาะสมกับแต่ละพื้นที่และแต่ละกลุ่มเป้าหมาย บนรากฐานของการเรียนรู้ การพึ่งตนเอง การมีจิตสำนึกรับผิดชอบ ต่อตนเอง ครอบครัว ชุมชน และสังคม และการมีส่วนร่วมในการพัฒนาภายใต้พื้นฐานของความเป็นไทย

The Pentium III processor family's significant performance improvement over previous Intel Architecture processors is based on the seamless combination of the P6 microarchitecture, Internet Streaming SIMD Extensions, Intel MMX media enhancement technology, and a processor serial number. The result is higher software performance plus headroom for applications that take advantage of the Pentium III processor.

ภาคผนวก ค
โปรแกรมการทำงานของระบบ

File name : ai.m

```
function ai(action)
more off;
echo off;
clc;

if nargin<1,
    action = 'initialize';
end;

if strcmp(action,'initialize'),
    oldFigNumber = watchon;
    figNumber = figure(...
        'name',' Segmentation For Thai & English Document ',...
        'numbertitle','off',...
        'menubar','none',...
        'visible','off');
    axes(...
        'units','normalized',...
        'position',[0.08 0.10 0.70 0.80],...
        'visible','off');

    axis([0 1 0 1]);

    bottom = 0.05;
    left = 0.825;
    btnWid = 0.15;
    btnHt = 0.09;
    spacing = 0.05;
    frmBorder = 0.02;
```

```

h = uicontrol(...
    'style','frame',...
    'units','normalized',...
    'position',[left-frmBorder bottom-frmBorder btnWid+2*frmBorder 0.9+2*frmBorder],...
    'backgroundcolor',[0.50 0.50 0.50]);

```

```

labelStr = 'Load Image';
callbackStr = '[pwd,image_file]=load_image';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+8.25*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

labelStr = 'Filter';
callbackStr = '[pwd,image_file]=fil_image(pwd,image_file)';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+7.00*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

labelStr = 'Segment 1';
callbackStr = '[pwd,count_file]=seg_image1(pwd,image_file)';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+5.75*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

labelStr = 'Segment 2';
callbackStr = '[pwd,count_file]=seg_image3(pwd,image_file)';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+4.50*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

labelStr = 'Segment 3';
callbackStr = '[pwd,count_file]=seg_image2(pwd,image_file)';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+3.25*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

labelStr = 'Apply Gray';
callbackStr = 'apply_g(pwd,image_file)';
infoHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom+2.00*btnHt+spacing btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

```

```

% labelStr = 'Apply Color';
% callbackStr = 'apply_c(pwd,image_file)';
% infoHndl = uicontrol(...
%     'style','push',...
%     'units','normalized',...
%     'position',[left bottom+0.75*btnHt+spacing btnWid btnHt],...
%     'string',labelStr,...

```

```

% 'call',callbackStr);

% labelStr = 'Recognize';
% callbackStr = 'recog_image(pwd,count_file)';
% infoHndl = uicontrol(...
% 'style','push',...
% 'units','normalized',...
% 'position',[left bottom+1.75*btnHt+spacing btnWid btnHt],...
% 'string',labelStr,...
% 'call',callbackStr);

```

```

labelStr = 'Close';
callbackStr = 'close(gcf)';
closeHndl = uicontrol(...
    'style','push',...
    'units','normalized',...
    'position',[left bottom btnWid btnHt],...
    'string',labelStr,...
    'call',callbackStr);

watchoff(oldFigNumber);
figure(figNumber);
end

```



File name : load_image.m

```
function [pwd,image_file]=load_image;
clear;
more off;
echo off;
clc;

[image_file,pwd] = uigetfile('*.bmp','Open Image File');

if (image_file == 0)
    if (pwd == 0)
        return;
    end
end

pwd_image_file = strcat(pwd,image_file);

info = imfinfo(pwd_image_file);
[x,map] = imread(pwd_image_file);
image(x);
colormap(map);
```

File name : fil_image.m

```
function [pwd,image_file]=fil_image(pwd,image_file);
more off;
echo off;
clc;

pwd_image_file = strcat(pwd,image_file);
info = imfinfo(pwd_image_file);
[m,map]=imread(pwd_image_file);
m = double(m);
m = char(m);

for h=1:info.Height
    for w=1:info.Width
        if m(h,w)==0
            if (m(h,w-1)~=0)&(m(h,w+1)~=0)&...
                (m(h+1,w-1)~=0)&(m(h+1,w+1)~=0)&(m(h+1,w)~=0)&...
                (m(h-1,w-1)~=0)&(m(h-1,w+1)~=0)&(m(h-1,w)~=0)
                m(h,w) = 1;
            end
        end
    end
end

m = double(m);
m = uint8(m);
image(m);
colormap(map);

image_file = 'fil.bmp';
save_pwd_image_file = strcat(pwd,image_file);
imwrite(m,map,save_pwd_image_file,'bmp');
```

File name : seg_image1.m

```
function [pwd,count_file]=seg_image1(pwd,image_file);

more off;
echo off;
clc;

pwd_image_file = strcat(pwd,image_file);
info = imfinfo(pwd_image_file);
[imf,map] = imread(pwd_image_file);

[err_row] = min_row(imf,1,info.Width,1,info.Height);
[err_col] = min_col(imf,1,info.Width,1,info.Height);
err_row = num2str(err_row);
err_col = num2str(err_col);

error = inputdlg({'Defind error in row matrix','Defind error in column'},....
    'Defind error',1,{err_row,err_col});
error = char(error);
error = str2num(error);
err_row = error(1);
err_col = error(2);

[row_1] = find_row(imf,1,info.Width,1,info.Height,err_row);

maxs = size(row_1);
maxs = maxs(2);
maxs = maxs-1;
count_file = 0;

for a_1=1:maxs

    im = imf(row_1(a_1):row_1(a_1+1),:);
    ih = row_1(a_1+1)-row_1(a_1)+1;
```

```

image(im);
colormap(map);
drawnow;
pause;

[col_1] = find_col(im,1,info.Width,1,ih,err_col);
maxs = size(col_1);
maxs = maxs(2);
maxs = maxs-1;

for b_1=1:maxs

    picture_1 = im(:,col_1(b_1):col_1(b_1+1));

    image(picture_1);
    colormap(map);
    drawnow;
    pause;

% picture_1 = double(picture_1);
% [picture_1] = size_image(picture_1,h_1,w_1);

% picture_1 = uint8(picture_1);
% image(picture_1);
% colormap(map);
% drawnow;

question = questdlg('You want save image','Image Processing',...
    'Yes','No','No');

switch question
case 'Yes',
    [save_file,save_pwd] = uiputfile('*.bmp','Save Image File');
    if (save_file ~= 0)

```

```

if (save_pwd ~= 0)
    save_image = strcat(save_pwd,save_file);
    imwrite(picture_1,map,save_image,'bmp');
end
end
end

question = questdlg('You want running program','Image process',...
    'Yes','No','No');

switch question
    case 'No',
        return;
    end
end

question = questdlg('The end of line , You want running program','Image Processing',...
    'Yes','No','No');

switch question
    case 'No',
        return;
    end
end

questdlg('The end of image','Image Processing','OK','OK');

```

File name : seg_image2.m

```
function [pwd,count_file]=seg_image2(pwd,image_file);
more off;
echo off;
clc;

pwd_image_file = strcat(pwd,image_file);
info = imfinfo(pwd_image_file);
[imf,map] = imread(pwd_image_file);

[row_1] = row(imf,1,info.Width,1,info.Height);

maxs = size(row_1);
maxs = maxs(2);
maxs = maxs-1;
count_file = 0;

for a_1=1:maxs

    im = imf(row_1(a_1):row_1(a_1+1),:);
    ih = row_1(a_1+1)-row_1(a_1)+1;

    image(im);
    colormap(map);
    drawnow;
    pause;

    im = double(im);

    for pixel_x=1:info.Width
        have = all(im(:,pixel_x));
        if have==0
            for pixel_y=1:ih
```

```

if im(pixel_y,pixel_x)==0
    im(pixel_y,pixel_x) = 1;
    [im,bound_y,bound_x] = segment...
    (im,pixel_y,pixel_x,1,info.Width,1,ih);

```

```

maxs = size(bound_x);
maxs = maxs(2);

```

```

max_x = bound_x(1);
min_x = bound_x(1);
max_y = bound_y(1);
min_y = bound_y(1);

```

```

for m=2:maxs
    if max_x < bound_x(m)
        max_x = bound_x(m);
    end
    if max_y < bound_y(m)
        max_y = bound_y(m);
    end
    if min_x > bound_x(m)
        min_x = bound_x(m);
    end
    if min_y > bound_y(m)
        min_y = bound_y(m);
    end
end

```

```

h_l = max_y-min_y+1;
w_l = max_x-min_x+1;

```

```

picture_1 = ones(h_l,w_l);

```

```

for a=1:maxs

```

```

picture_1(bound_y(a)-min_y+1,bound_x(a)-min_x+1) = 0;
end

picture_1 = uint8(picture_1);
image(picture_1);
colormap(map);
drawnow;
pause;

question = questdlg('You want save image','Image Processing',...
    'Yes','No','No');
switch question
case 'Yes',
    [save_file,save_pwd] = uiputfile('*.bmp','Save Image File');
    if (save_file ~= 0)
        if (save_pwd ~= 0)
            save_image = streat(save_pwd,save_file);
            imwrite(picture_1,map,save_image,'bmp');
        end
    end
end

question = questdlg('You want run histogram','Image Processing',...
    'Yes','No','No');
switch question
case 'Yes',
    back(picture_1,h_1,w_1,map);
end

question = questdlg('You want running program','Image Processing',...
    'Yes','No','No');
switch question
case 'No',
    return;

```

```

end

% picture_2 = double(image);
% [picture_2] = size_image(image,h,w);

% picture_2 = uint8(image);
% image(image);
% colormap(map);
% drawnow;

end
end
end
end

question = questdlg('The end of line , You want running program','Image Processing',...
    'Yes','No','No');
switch question
case 'No',
    return;
end
end
questdlg('The end of image','Image Processing','OK','OK');

```

File name : seg_image3.m

```
function [pwd,count_file]=seg_image3(pwd,image_file);

more off;
echo off;
clc;

pwd_image_file = strcat(pwd,image_file);
info = imfinfo(pwd_image_file);
[imf,map]= imread(pwd_image_file);

[row_1] = row(imf,1,info.Width,1,info.Height);

maxs = size(row_1);
maxs = maxs(2);
maxs = maxs-1;
count_file = 0;

for a_1=1:maxs

    iw = info.Width+2;
    ih = row_1(a_1+1)-row_1(a_1)+3;
    im = ones(ih,iw);
    im(2:ih-1,2:iw-1) = imf(row_1(a_1):row_1(a_1+1),:);

    im = uint8(im);
    image(im);
    colormap(map);
    drawnow;
    pause;

    im = double(im);

    for pixel_x=1:iw
```

```

have = all(im(:,pixel_x));
if have==0
    for pixel_y=1:ih
        if im(pixel_y,pixel_x)==0

            [bound_x,bound_y] = boundary(im,pixel_y,pixel_x);

            maxs = size(bound_x);
            maxs = maxs(2);

            max_x = bound_x(1);
            min_x = bound_x(1);
            max_y = bound_y(1);
            min_y = bound_y(1);

            for m=2:maxs
                if max_x < bound_x(m)
                    max_x = bound_x(m);
                end
                if max_y < bound_y(m)
                    max_y = bound_y(m);
                end
                if min_x > bound_x(m)
                    min_x = bound_x(m);
                end
                if min_y > bound_y(m)
                    min_y = bound_y(m);
                end
            end

            h_1 = max_y-min_y+1;
            w_1 = max_x-min_x+1;

            picture_1 = ones(h_1,w_1);

```

```

for a=1:maxs
    picture_1(bound_y(a)-min_y+1,bound_x(a)-min_x+1) = 0;
    im(bound_y(a),bound_x(a)) = 1;
end

for a=1:h_1
    for b=1:w_1
        if im(min_y+a-1,min_x+b-1)==0
            [copy] = copy_image(picture_1,h_1,w_1,a,b);
            if copy==1
                picture_1(a,b) = 0;
                im(min_y+a-1,min_x+b-1) = 1;
            end
        end
    end
end

picture_1 = uint8(picture_1);
image(picture_1);
colormap(map);
drawnow;
pause;

question = questdlg('You want save image','Image Processing',...
    'Yes','No','No');

switch question
    case 'Yes',
        [save_file,save_pwd] = uiputfile('*.bmp','Save Image File');
        if (save_file ~= 0)
            if (save_pwd ~= 0)
                save_image = strcat(save_pwd,save_file);
                imwrite(picture_1,map,save_image,'bmp');
            end
        end
    end
end

```

```

        end
    end

    question = questdlg('You want run histogram','Image Processing',...
        'Yes','No','No');

    switch question
        case 'Yes',
            back(picture_1,h_1,w_1,map);
    end

    question = questdlg('You want running program','Image Processing',...
        'Yes','No','No');

    switch question
        case 'No',
            return;
    end

    % picture_2 = double(picture_2);
    % [picture_2] = size_image(picture_2,h_2,w_2);

    % picture_2 = uint8(picture_2);
    % image(picture_2);
    % colormap(map);
    % drawnow;

    end
end
end
end

question = questdlg('The end of line , You want running program','Image Processing',...
    'Yes','No','No');

switch question
    case 'No',

```

```

        return;
    end
end
questdlg('The end of image','Image Processing','OK','OK');

```



File name : apply_g.m

```
function apply_g(pwd,image_file);

more off;
echo off;
clc;

level = inputdlg({'Defind level of gray image'},'Defind level',1,'4');
level = char(level);
level = str2num(level);

hold = round(256/level);

pwd_image_file = strcat(pwd,image_file);
info = imfinfo(pwd_image_file);
[imf,map] = imread(pwd_image_file);
imf = double(imf);

for tone=1:level
    if tone==1
        tone_up = hold+1;
        tone_low = 1;
    else
        tone_up = (tone*hold)-1;
        tone_low = (tone*hold)-hold;
    end

    picture = zeros(info.Height,info.Width);

    for y=1:info.Height
        for x=1:info.Width
            if (imf(y,x)>=tone_low)&(imf(y,x)<=tone_up)
                picture(y,x) = imf(y,x);
                imf(y,x) = 0;
            end
        end
    end
end
```

```
[imf,picture] = segment_g...  
(imf,picture,tone_up,tone_low,y,x,info.Width,info.Height);  
end  
end  
end  
  
picture = uint8(picture);  
image(picture);  
colormap(map);  
drawnow;  
end  
questdlg('The end of image','Image Processing','OK','OK');
```



File name : row.m

```
function [out]=row(in,xstart,xend,ystart,yend);
```

```
th = 0;
```

```
error = round((xend-xstart+1)*(th/100));
```

```
in = double(in);
```

```
in = char(in);
```

```
for y=ystart:yend
```

```
    temp = find(in(y,xstart:xend)==0);
```

```
    temp = size(temp);
```

```
    graph(y) = temp(2);
```

```
end
```

```
count = 0;
```

```
ecount = 0;
```

```
temp = error;
```

```
if graph(ystart)>temp
```

```
    count = 1;
```

```
    out(1) = ystart;
```

```
else
```

```
    ecount = 1;
```

```
    merror(1) = ystart;
```

```
    if graph(ystart)<temp
```

```
        temp = graph(ystart);
```

```
    end
```

```
end
```

```
for y=ystart+1:yend
```

```
    if graph(y)==temp
```

```
        ecount = ecount+1;
```

```
        merror(ecount) = y;
```

```
    elseif graph(y)<temp
```

```

    ecount = 1;
    merror(1) = y;
    temp = graph(y);
elseif (ecount~=0)&(graph(y)>error)
    count = count+1;
    out(count) = merror(round(ecount/2));
    temp = error;
    ecount = 0;
end
end

if ecount~=0
    count = count+1;
    out(count) = merror(round(ecount/2));
else
    count = count+1;
    out(count) = yend;
end
end

```



File name : min_row.m

```
function [out]=min_row(in,xstart,xend,ystart,yend);
```

```
in = double(in);
```

```
for y=ystart:yend
```

```
    temp = find(in(y,xstart:xend)==0);
```

```
    temp = size(temp);
```

```
    graph(y) = temp(2);
```

```
end
```

```
out = graph(ystart);
```

```
for y=ystart+1:yend
```

```
    if graph(y)<out
```

```
        out = graph(y);
```

```
    end
```

```
end
```

```
out = round((out/(xend-xstart+1))*100);
```



File name : find_row.m

```
function [out]=find_row(in,xstart,xend,ystart,yend,th);
```

```
error = round((xend-xstart+1)*(th/100));
```

```
in = double(in);
```

```
for y=ystart:yend
```

```
    temp = find(in(y,xstart:xend)==0);
```

```
    temp = size(temp);
```

```
    graph(y) = temp(2);
```

```
end
```

```
count = 0;
```

```
ecount = 0;
```

```
temp = error;
```

```
if graph(ystart)>temp
```

```
    count = 1;
```

```
    out(1) = ystart;
```

```
else
```

```
    ecount = 1;
```

```
    merror(1) = ystart;
```

```
    if graph(ystart)<temp
```

```
        temp = graph(ystart);
```

```
    end
```

```
end
```

```
for y=ystart+1:yend
```

```
    if graph(y)==temp
```

```
        ecount = ecount+1;
```

```
        merror(ecount) = y;
```

```
    elseif graph(y)<temp
```

```
        ecount = 1;
```

```
        merror(1) = y;
```



```

    temp = graph(y);
elseif (ecount~=0)&(graph(y)>error)
    count = count+1;
    out(count) = merror(round(ecount/2));
    temp = error;
    ecount = 0;
end
end

```

```

if ecount~=0
    count = count+1;
    out(count) = merror(round(ecount/2));
else
    count = count+1;
    out(count) = yend;
end

```



File name : min_col.m

```
function [out]=min_col(in,xstart,xend,ystart,yend);
```

```
in = double(in);
```

```
for x=xstart:xend
```

```
    temp = find(in(ystart:yend,x)==0);
```

```
    temp = size(temp);
```

```
    graph(x) = temp(1);
```

```
end
```

```
out = graph(xstart);
```

```
for x=xstart+1:xend
```

```
    if graph(x)<out
```

```
        out = graph(x);
```

```
    end
```

```
end
```

```
out = round((out/(yend-ystart+1))*100);
```



File name : find_col.m

```
function [out]=find_col(in,xstart,xend,ystart,yend,th);
```

```
error = round((yend-ystart+1)*(th/100));
```

```
in = double(in);
```

```
for x=xstart:xend
```

```
    temp = find(in(ystart:yend,x)==0);
```

```
    temp = size(temp);
```

```
    graph(x) = temp(1);
```

```
end
```

```
count = 0;
```

```
ecount = 0;
```

```
temp = error;
```

```
if graph(xstart)>temp
```

```
    count = 1;
```

```
    out(1) = xstart;
```

```
else
```

```
    ecount = 1;
```

```
    merror(1) = xstart;
```

```
    if graph(xstart)<temp
```

```
        temp = graph(xstart);
```

```
    end
```

```
end
```

```
for x=xstart+1:xend
```

```
    if graph(x)==temp
```

```
        ecount = ecount+1;
```

```
        merror(ecount) = x;
```

```
    elseif graph(x)<temp
```

```
        ecount = 1;
```

```
        merror(1) = x;
```

```

    temp = graph(x);
elseif (ecount~=0)&(graph(x)>error)
    count = count+1;
    out(count) = merror(round(ecount/2));
    temp = error;
    ecount = 0;
end
end

```

```

if ecount~=0
    count = count+1;
    out(count) = merror(round(ecount/2));
else
    count = count+1;
    out(count) = xend;
end

```



File name : boundary.m

```
function [bound_x,bound_y]=boundary(picture,y,x);
```

```
count = 1;
```

```
bound_x(count) = x;
```

```
bound_y(count) = y;
```

```
tx1 = x;
```

```
ty1 = y;
```

```
tx2 = x;
```

```
ty2 = y-1;
```

```
while (ty2~=y)|(ty1~=y)|(tx2~=x)|(tx1~=x-1)
```

```
if picture(ty2,tx2)==0
```

```
count = count+1;
```

```
bound_x(count) = tx2;
```

```
bound_y(count) = ty2;
```

```
end
```

```
if (tx1<tx2)&(ty1==ty2) type = 1; end
```

```
if (tx1>tx2)&(ty1==ty2) type = 2; end
```

```
if (tx1==tx2)&(ty1<ty2) type = 3; end
```

```
if (tx1==tx2)&(ty1>ty2) type = 4; end
```

```
switch type
```

```
case 1,
```

```
if picture(ty2,tx2)==0
```

```
tx1 = tx2;
```

```
ty1 = ty2;
```

```
ty2 = ty2-1;
```

```
elseif picture(ty2,tx2)~=0
```

```
tx1 = tx2;
```

```

    ty1 = ty2;
    ty2 = ty2+1;
end

```

case 2,

```

    if picture(ty2,tx2)==0
        tx1 = tx2;
        ty1 = ty2;
        ty2 = ty2+1;
    elseif picture(ty2,tx2)~=0
        tx1 = tx2;
        ty1 = ty2;
        ty2 = ty2-1;
    end

```

case 3,

```

    if picture(ty2,tx2)==0
        tx1 = tx2;
        tx2 = tx2+1;
        ty1 = ty2;
    elseif picture(ty2,tx2)~=0
        tx1 = tx2;
        tx2 = tx2-1;
        ty1 = ty2;
    end

```

case 4,

```

    if picture(ty2,tx2)==0
        tx1 = tx2;
        tx2 = tx2-1;
        ty1 = ty2;
    elseif picture(ty2,tx2)~=0
        tx1 = tx2;
        tx2 = tx2+1;
    end

```

```
    ty1 = ty2;  
  end  
end  
end
```



File name : copy_image.m

```
function [copy]=copy_image(picture,h,w,y,x);
```

```
copy = 0;
```

```
for i=x+1:w
```

```
    if picture(y,i)==0
```

```
        for j=y+1:h
```

```
            if picture(j,x)==0
```

```
                for k=1:x-1
```

```
                    if picture(y,x-k)==0
```

```
                        for l=1:y-1
```

```
                            if picture(y-l,x)==0
```

```
                                copy = 1;
```

```
                                return;
```

```
                            end
```

```
                        end
```

```
                    end
```

```
                end
```

```
            end
```

```
        end
```

```
    end
```

```
end
```

File name : segment.m

```
function [seg_pic,pixel_y,pixel_x]=segment(seg_pic,y,x,min_x,max_x,min_y,max_y);
```

```
check = 0;
```

```
count = 1;
```

```
max_size = 1;
```

```
pixel_x(1) = x;
```

```
pixel_y(1) = y;
```

```
for temp_y=y-1:y+1
```

```
    for temp_x=x-1:x+1
```

```
        if (temp_y>=min_y)&(temp_y<=max_y)&...
```

```
            (temp_x>=min_x)&(temp_x<=max_x)
```

```
            if seg_pic(temp_y,temp_x)==0
```

```
                seg_pic(temp_y,temp_x) = 1;
```

```
                check = check+1;
```

```
                max_size = max_size+1;
```

```
                pixel_y(max_size) = temp_y;
```

```
                pixel_x(max_size) = temp_x;
```

```
            end
```

```
        end
```

```
    end
```

```
end
```

```
while check>0
```

```
    count = count+1;
```

```
    check = check-1;
```

```
    for temp_y=pixel_y(count)-1:pixel_y(count)+1
```

```
        for temp_x=pixel_x(count)-1:pixel_x(count)+1
```

```

if (temp_y>=min_y)&(temp_y<=max_y)&...
    (temp_x>=min_x)&(temp_x<=max_x)

if seg_pic(temp_y,temp_x)==0
    seg_pic(temp_y,temp_x) = 1;
    check = check+1;
    max_size = max_size+1;
    pixel_y(max_size) = temp_y;
    pixel_x(max_size) = temp_x;
end
end
end
end
end

```



File name : back.m

```
function back(imf,height,width,map);
```

```
[err_row] = min_row(imf,1,width,1,height);
```

```
[err_col] = min_col(imf,1,width,1,height);
```

```
err_row = num2str(err_row);
```

```
err_col = num2str(err_col);
```

```
error = inputdlg({'Defind error in row matrix','Defind error in column'},...  
                'Defind error',1,{err_row,err_col});
```

```
error = char(error);
```

```
error = str2num(error);
```

```
err_row = error(1);
```

```
err_col = error(2);
```

```
[row_1] = find_row(imf,1,width,1,height,err_row);
```

```
maxs = size(row_1);
```

```
maxs = maxs(2);
```

```
maxs = maxs-1;
```

```
aaaa = maxs;
```

```
for a_1=1:maxs
```

```
im = imf(row_1(a_1):row_1(a_1+1),:);
```

```
ih = row_1(a_1+1)-row_1(a_1)+1;
```

```
[col_1] = find_col(im,1,width,1,ih,err_col);
```

```
maxs = size(col_1);
```

```
maxs = maxs(2);
```

```
maxs = maxs-1;
```

```
bbbb = maxs;
```

```
for b_1=1:maxs
```

```
    picture_1 = im(:,col_1(b_1):col_1(b_1+1));
```

```
    iw = col_1(b_1+1)-col_1(b_1)+1;
```

```
    image(picture_1);
```

```
    colormap(map);
```

```
    drawnow;
```

```
    pause;
```

```
    question = questdlg('You want save image','Image Processing',...
```

```
        'Yes','No','No');
```

```
    switch question
```

```
        case 'Yes',
```

```
            [save_file,save_pwd] = uiputfile('*.bmp','Save Image File');
```

```
            if (save_file ~= 0)
```

```
                if (save_pwd ~= 0)
```

```
                    save_image = strcat(save_pwd,save_file);
```

```
                    imwrite(picture_1,map,save_image,'bmp');
```

```
                end
```

```
            end
```

```
    end
```

```
if (aaaa>1)|(bbbb>1)
```

```
    question = questdlg('You want run histogram this picture','Image Processing',...
```

```
        'Yes','No','No');
```

```
    switch question
```

```
        case 'Yes',
```

```
            back(picture_1,ih,iw,map);
```

```
        end
```

```
    end
```

```
end
```

```
end
```

```
question = questdlg('You want run histogram all picture','Image Processing',...  
    'Yes','No','No');  
switch question  
case 'Yes',  
    back(imf,height,width,map);  
end
```



File name : segment_g.m

```
function [seg_pic,copy_pic]=segment_g...
    (seg_pic,copy_pic,tone_up,tone_low,y,x,max_x,max_y);

check = 0;
count = 1;
max_size = 1;
pixel_x(1) = x;
pixel_y(1) = y;

for temp_y=y-1:y+1
    for temp_x=x-1:x+1

        if (temp_y>=1)&(temp_y<=max_y)&...
            (temp_x>=1)&(temp_x<=max_x)

            if (seg_pic(temp_y,temp_x)>=tone_low)&...
                (seg_pic(temp_y,temp_x)<=tone_up)

                copy_pic(temp_y,temp_x) = seg_pic(temp_y,temp_x);
                seg_pic(temp_y,temp_x) = 0;

                check = check+1;
                max_size = max_size+1;
                pixel_y(max_size) = temp_y;
                pixel_x(max_size) = temp_x;
            end
        end
    end
end

while check>0
```

```

count = count+1;
check = check-1;

for temp_y=pixel_y(count)-1:pixel_y(count)+1
    for temp_x=pixel_x(count)-1:pixel_x(count)+1

        if (temp_y>=1)&(temp_y<=max_y)&...
            (temp_x>=1)&(temp_x<=max_x)

                if (seg_pic(temp_y,temp_x)>=tone_low)&...
                    (seg_pic(temp_y,temp_x)<=tone_up)

                        copy_pic(temp_y,temp_x) = seg_pic(temp_y,temp_x);
                        seg_pic(temp_y,temp_x) = 0;

                        check = check+1;
                        max_size = max_size+1;
                        pixel_y(max_size) = temp_y;
                        pixel_x(max_size) = temp_x;
                    end
                end
            end
        end
    end
end
end

```

File name : size_image.m

```
function [matrix]=size_image(front,size_h,size_w);
```

```
height = 20;
```

```
width = 20;
```

```
matrix = ones(height,width);
```

```
retio_h = size_h/height;
```

```
retio_w = size_w/width;
```

```
for h=1:height
```

```
    temp_h = round((retio_h*h)+0.5);
```

```
    if temp_h > size_h
```

```
        temp_h = size_h;
```

```
    end
```

```
    for w=1:width
```

```
        temp_w = round((retio_w*w)+0.5);
```

```
        if temp_w > size_w
```

```
            temp_w = size_w;
```

```
        end
```

```
        matrix(h,w) = front(temp_h,temp_w);
```

```
    end
```

```
end
```

บรรณานุกรม

1. ศศิกานต์ วรรณา, ศิริลักษณ์ ชุติศักดิ์แสงทอง, “การประมวลผลภาพเพื่อตัดอักขระภาษาไทยและเรียนรู้จดจำเบื้องต้น”, วิทยานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2536.
2. สมศักดิ์ วัลย์รัชต์, “การวิเคราะห์และระบุส่วนประกอบของหน้าเอกสารและการรู้จำตัวอักษร”, วิทยานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2537.
3. Rafael G. Gonzalez, Richard E. Woods, “Digital Image Processing”, Addison Wesley, 1993.
4. C. wayne Brown, Bary J. Shepherd, “Graphics File Formats”, Manning Greenwich, 1995.

