

บทที่ 4

การทดลองและผล

4.1 การเตรียมข้อมูล

ข้อมูลตัวเลขไทยที่นำมาทดลองการรู้จำเป็นตัวเลขไทยจากตัวเลข 0 ถึง ๙ ที่ได้มาจาก font ตัวอักษรไทยบนวินโดว เวอร์ชัน 3.1 ไทยเอดิชัน โดย font ที่นำมาใช้มี 5 fonts คือ BrowalliaUPC, CordiaUPC, DilleniaUPC, EucrosiaUPC, และ FreesiaUPC ซึ่งการที่พิจารณาเลือกใช้ทั้ง 5 fonts นี้เนื่องจากแบบตัวเลขในแต่ละ font จะเป็นแบบตัวพิมพ์ปกติซึ่งไม่มีการทำลวดลายในแบบตัวเลข และ ในแต่ละ font มีความแตกต่างกันในเรื่องของส่วนโค้งของลายเส้นและความหนาของลายเส้น โดยในแต่ละ font จะใช้ตัวเลขไทยตั้งแต่ ๐ ถึง ๙ ที่มีขนาดแตกต่างกันจำนวน 5 ขนาด คือ 18, 20, 22, 24, และ 26 ดังตารางที่ 4.1 มาเป็นต้นแบบข้อมูลที่จะใช้ทำการทดสอบ ซึ่งจะรวมเป็นรูปแบบของตัวเลขไทยที่มีขนาดแตกต่างกันจำนวน 250 ตัว จากนั้นทำการพิมพ์ font ตัวเลขดังกล่าวทั้งหมดออกทางเครื่องพิมพ์เลเซอร์แล้วนำมาสแกนโดยใช้ DFI handy scanner HS - 3000 ที่ความละเอียด 300 dpi เพื่อเก็บเป็นไฟล์ข้อมูลภาพโดยเก็บในไฟล์นามสกุล .PCX และ .BMP โดยแต่ละตัวจะทำการสแกน 3 ครั้ง ซึ่งจะรวมเป็นข้อมูลที่ใช้ในการทดสอบทั้งหมด 750 ตัว

4.2 การวิเคราะห์หาต้นแบบ

การวิเคราะห์หาต้นแบบเพื่อจัดเก็บไว้สำหรับการเปรียบเทียบ จะทำโดยการนำเอาข้อมูลภาพในส่วนของ font แบบ CordiaUPC และ EucrosiaUPC ที่สแกนเข้ามาชุดแรกซึ่งมีทั้งหมด 100 ตัว นำมาผ่านกระบวนการต่าง ๆ จนกระทั่งได้ primitive ของทุกตัว แล้วนำมาวิเคราะห์หา primitive ร่วมในแต่ละตัวเลข แล้วจัดเก็บเป็นต้นแบบโดยแต่ละตัวเลขไทยจะจัดเก็บเพียงหนึ่งต้นแบบต่อหนึ่งตัวเลขเท่านั้น ซึ่งจะได้เป็นต้นแบบทั้งหมด 10 ชุด และไว้ใช้สำหรับการรู้จำต่อไป

ตารางที่ 4.1 รูปแบบตัวเลขไทยที่ใช้ในการทดสอบ

font	ขนาด	รูปแบบตัวเลขไทย
BrowalliaUPC	18	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	20	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	22	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	24	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	26	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
CordiaUPC	18	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	20	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	22	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	24	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	26	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
DilleniaUPC	18	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	20	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	22	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	24	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	26	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
EucrosiaUPC	18	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	20	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	22	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	24	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	26	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

ตารางที่ 4.1(ต่อ) รูปแบบตัวเลขไทยที่ใช้ในการทดสอบ

font	ขนาด	รูปแบบตัวเลขไทย
FreesiaUPC	18	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	20	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	22	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	24	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
	26	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

4.3 การทดสอบและผล

การทดสอบการรู้จำตัวเลขไทยของโปรแกรมที่พัฒนาขึ้นนี้จะกระทำบนเครื่องไมโครคอมพิวเตอร์ CPU 80486DX-33 โดยการทดสอบจะมีการแบ่งชุดของต้นแบบที่ได้จัดเก็บไว้ ออกเป็นกลุ่มเพื่อใช้สำหรับการคำนวณหา distance กับข้อมูลภาพที่นำมาทดสอบการรู้จำ โดยอาศัยข้อมูลที่ได้มาในระหว่างกระบวนการในช่วงแรก คือ ตำแหน่งจุดเริ่มต้นของการตรวจสอบโครงร่าง และ จำนวนลูปที่ปรากฏในโครงร่างตัวอักษร ซึ่งสามารถจัดเป็นกลุ่มได้ ดังตารางที่ 4.2

ตารางที่ 4.2 การแบ่งกลุ่มในการคำนวณหา distance

จำนวนลูป	ตำแหน่งของจุดเริ่มต้น	ตัวเลข
0 - 1 ลูป	โซน 1	๐, ๒, ๖
	โซน 2	๐, ๔, ๕, ๗, ๘, ๙
	โซน 3	๑, ๓
	โซน 4	๑
2 ลูป	โซน 1 - 4	๕

ตำแหน่งจุดเริ่มต้นจะแบ่งออกเป็น 4 โซนดังรูปที่ 4.1 ส่วนจำนวนของลูปจะมีอยู่ 2 อย่างคือ มี 0 หรือ 1 ลูป และมี 2 ลูป

โซน 1	โซน 2
โซน 4	โซน 3

รูปที่ 4.1 การแบ่งโซนของข้อมูลภาพ

จากการทดสอบหาค่า distance ระหว่างข้อมูลที่นำมาวิเคราะห์หาต้นแบบทั้งหมด คือ font แบบ CordiaUPC และ EucrosiaUPC ที่สแกนเข้ามาเป็นชุดแรกกับต้นแบบที่ได้จัดสร้างขึ้น ซึ่งผลที่ได้เป็นดังแสดงตามตารางที่ ค.1 ถึง ค.10 ในภาคผนวก ค. โดยจะแบ่งแยกตามชุดของ font และขนาดของภาพตัวเลข จะเห็นว่าค่า distance ที่ได้ของข้อมูลภาพในแต่ละภาพเทียบกับต้นแบบที่ตรงกับภาพนั้นจะมีค่าอยู่ในช่วง 0 ถึง 5 และจากผลดังกล่าวเราสามารถแยกการพิจารณาผลที่ได้ในการรู้จำเป็นดังนี้

- | | |
|-------------|---|
| จำถูก | จะเป็นกรณีที่โปรแกรมสามารถทำการรู้จำภาพตัวเลขนั้นได้ถูกต้อง (ค่า distance ที่น้อยที่สุดที่ได้เป็นของต้นแบบที่ตรงกับภาพตัวเลขนั้น) |
| จำผิด | จะเป็นกรณีที่โปรแกรมจำภาพตัวเลขนั้นเป็นตัวเลขอื่น (ค่า distance ที่น้อยที่สุดที่ได้เป็นของต้นแบบที่ไม่ตรงกับภาพตัวเลขที่นำมาทดสอบ) |
| ไม่ตัดสินใจ | จะเกิดในกรณีที่โปรแกรมไม่สามารถทำการตัดสินใจได้ว่าภาพตัวเลขนั้นเป็นตัวเลขใด (ค่า distance ที่คำนวณได้เทียบต้นแบบที่มีค่าน้อยที่สุดเกิดขึ้นมากกว่า 1 ตัว) |
| จำไม่ได้ | จะเกิดเมื่อโปรแกรมไม่สามารถจำได้ว่าภาพตัวเลขนั้นเป็นตัวเลขใด (ค่า distance ที่คำนวณได้เทียบกับต้นแบบที่มีค่าน้อยที่สุดมีค่ามากกว่า 5) ซึ่งการที่กำหนดกรณีนี้เพื่อป้องกันกรณีที่ภาพข้อมูลที่นำมารู้จำเป็นภาพตัวอักษรอื่นที่ไม่ใช่ตัวเลขไทย |

ตารางที่ 4.4 ผลการทดสอบการรู้จำข้อมูลแยกตาม font

Font	BrowalliaUPC	CordiaUPC	DilleniaUPC	EucrosiaUPC	FreesiaUPC
จำนวนข้อมูล	150	150	150	150	150
จำถูก	148	148	144	147	144
จำผิด	-	-	2	-	-
ไม่ตัดสินใจ	-	2	4	3	2
จำไม่ได้	2	-	-	-	4
อัตราความถูกต้อง	98.66 %	98.66 %	96 %	98 %	96 %

จากการทดสอบสามารถสรุปผลทั้งหมดได้เป็นดังนี้

ข้อมูลภาพที่นำมาทดสอบทั้งหมด	750	ตัว
จำถูก	731	ตัว
จำผิด	2	ตัว
ไม่ตัดสินใจ	11	ตัว
จำไม่ได้	6	ตัว
อัตราความถูกต้อง	97.46	%

สำหรับเวลาที่ใช้ในการรู้จำภาพตัวเลขไทยโดยโปรแกรมที่พัฒนาขึ้นนี้ โดยเริ่มนับจากกระบวนการที่ 3.1 ในบทที่ 3 ไปจนกระทั่งได้ผลการรู้จำออกมา ซึ่งจะได้เวลาในการรู้จำเฉลี่ย 756 msec. ต่อหนึ่งภาพตัวอักษร โดยทดสอบบนเครื่องไมโครคอมพิวเตอร์ CPU 80486DX-33