

การแปลผลและการบรรยายรูปภาพตารางสำหรับผู้พิการทางสายตา Tabular Image Translation and Description for Visually Impaired People

วงศ์ยศ เกิดศรี อังคนา จันทร์รุ่งอุทัย และ โชติรัตน์ รัตนามัทธนะ
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330
อีเมล: Wongyos.K@student.chula.ac.th, {g49ach, ann}@cp.eng.chula.ac.th

บทคัดย่อ

ปัจจุบันผู้พิการทางสายตาสามารถอ่านเอกสารอิเล็กทรอนิกส์บนเครื่องคอมพิวเตอร์ได้โดยใช้โปรแกรมอ่านหน้าจอ ซึ่งอ่านออกเสียงได้เฉพาะข้อมูลที่เป็นตัวอักษรเท่านั้น ไม่สามารถอ่านข้อมูลที่เป็นรูปภาพได้ ทำให้ผู้พิการทางสายตาไม่อาจทราบถึงรายละเอียดของข้อมูลรูปภาพเหล่านั้น โดยปกติแล้วรูปภาพส่วนใหญ่จะถูกบรรยายให้ผู้พิการทางสายตาเข้าใจผ่านการใช้มือสัมผัส โดยใช้เครื่องพิมพ์อักษรเบรลล์หรือเครื่องแสดงผลเบรลล์ในการแสดงรายละเอียดของรูปภาพ แต่วิธีการดังกล่าวมีค่าใช้จ่ายสูง ใช้เวลานาน และไม่สะดวกกับการใช้งานงานวิจัยนี้จึงได้นำเสนอวิธีการแปลผลและบรรยายรูปภาพแบบใหม่ด้วยการแปลงรูปภาพให้เป็นข้อความบรรยายแล้วใช้โปรแกรมอ่านหน้าจออ่านข้อความเหล่านั้นขึ้นมาแทน โดยใช้หลักการของการประมวลผลภาพดิจิทัลร่วมกับการรู้จำตัวอักษรในการประมวลผล ซึ่งใช้ตัวอย่างรูปภาพตารางเป็นกรณีศึกษา

คำสำคัญ การแปลความหมายรูปภาพ การบรรยายรูปภาพ การประมวลผลภาพ ผู้พิการทางสายตา โปรแกรมอ่านหน้าจอ เทคโนโลยีอำนวยความสะดวก

Abstract

Nowadays, visually impaired people can read electronic documents on the computer by using some screen reader software. Nevertheless, these software products only work on the textual parts, but they are unable to read nor describe images. Consequently, the visually impaired people may not understand the detail of the images. The task of translating and describing images is usually achieved by printing the images with the Braille printer or showing them via the Braille display; as a result, the visually impaired people can touch the images by their hands. However, the process is very time consuming, costly, and inconvenient. This paper proposes a new concept based on Digital Image Processing and the Optical Character Recognition (OCR) techniques for describing and translating the images into textual description which is readable from the screen reader software. We demonstrate the utility of our approach by testing on a set of tabular images.

Key Words: Image Translation, Image Description, Image Processing, Visually Impaired People, Screen Reader, Assistive Technology

1. บทนำ

ในปัจจุบันเทคโนโลยีอำนวยความสะดวก หรือเทคโนโลยีเพื่อช่วยเหลือ (Assistive Technology: AT) ได้กลายเป็นเทคโนโลยีสำคัญสำหรับผู้พิการ ที่จะคอยช่วยเหลือและเติมเต็มให้ผู้พิการสามารถทำงานและใช้ชีวิตประจำวันได้อย่างปกติสุขเช่นเดียวกับคนทั่วไป อย่างไรก็ตาม ยังมีประเด็นที่ท้าทายอีกมากที่ต้องการการพัฒนาอย่างต่อเนื่อง ซึ่งหนึ่งในนั้นคือ การแปลงผลและการบรรยายรูปภาพให้กับผู้พิการทางสายตา

การแปลงผลและการบรรยายรูปภาพให้กับผู้พิการทางสายตานั้น เป็นเรื่องที่มีการวิจัยและพัฒนาอย่างต่อเนื่องมาโดยตลอด ซึ่งเริ่มจากการแปลงรูปภาพจากหนังสือหรือจากแผ่นกระดาษ ให้เป็นรูปภาพที่ประกอบด้วยสัญลักษณ์ของเบรลล์ โดยใช้หลักการพิมพ์ภาพด้วยเครื่องพิมพ์เบรลล์ (Braille printer) หรือใช้การแสดงผลภาพออกจากเครื่องแสดงผลเบรลล์ (Braille display) เพื่อให้ผู้พิการทางสายตาใช้มือสัมผัสและเข้าใจถึงลักษณะต่างๆ ของรูปภาพได้ แต่วิธีการดังกล่าวต้องใช้เวลาในการสร้างแผนภาพเบรลล์ มีค่าใช้จ่ายสูง และไม่สะดวกในการใช้งานจริง

รูปภาพส่วนใหญ่ในปัจจุบันถูกจัดเก็บไว้ในรูปแบบของรูปภาพดิจิทัลซึ่งอาจปรากฏบนเอกสารอิเล็กทรอนิกส์ต่างๆ ในเครื่องคอมพิวเตอร์ ซึ่งโดยทั่วไปผู้พิการทางสายตาสามารถเข้าถึงเอกสารเหล่านี้ได้โดยใช้โปรแกรมอ่านหน้าจอ (Screen reader) [1] เช่น JAWS [2] Apple VoiceOver [3] Adobe Read Out Loud [4] พีพีเอตาทิพ (PPA Tatip) [5] เป็นต้น อย่างไรก็ตามโปรแกรมเหล่านี้สามารถอ่านออกเสียงได้เฉพาะในส่วนของข้อความหรือตัวหนังสือ ไม่สามารถอ่านออกเสียงและแสดงรายละเอียดของรูปภาพที่ปรากฏอยู่บนหน้าจอ ทำให้ผู้พิการทางสายตาไม่สามารถเข้าใจ และทราบถึงความหมายของรูปภาพเหล่านั้นได้ ข้อจำกัดดังกล่าวจึงเป็นจุดเริ่มต้นของงานวิจัยนี้ในการแปลงผลและบรรยายรูปภาพให้กับผู้พิการทางสายตาได้เข้าใจถึงความหมายและรายละเอียดต่างๆ ของรูปภาพ อย่างไรก็ตามรูปภาพที่ปรากฏอยู่บนเอกสาร

อิเล็กทรอนิกส์ในปัจจุบันนี้มีหลายประเภท เช่น ภาพวาด ภาพเขียน ภาพถ่าย ภาพกราฟิกส์ เป็นต้น ซึ่งแต่ละประเภทมีโครงสร้างและลักษณะของภาพที่แตกต่างกัน จึงทำให้มีวิธีการแปลงผลและบรรยายภาพที่แตกต่างกัน รวมทั้งต้องใช้วิธีที่ซับซ้อนในการประมวลผล ซึ่งยังเป็นเรื่องยากในการคิดหาวิธีที่มีประสิทธิภาพในการทำงาน งานวิจัยนี้จึงเริ่มต้นด้วยวิธีการประมวลผลที่ไม่ซับซ้อนเกินไปโดยเลือกรูปภาพตาราง (Tabular Image) เป็นกรณีศึกษา ซึ่งใช้หลักการทำงานร่วมระหว่างโปรแกรมแปลความหมายรูปภาพตารางให้อยู่ในรูปของข้อความบรรยาย และโปรแกรมอ่านหน้าจอในการอ่านออกเสียงข้อความที่ได้ให้กับผู้พิการทางสายตาฟัง

เหตุผลที่แปลความหมายรูปภาพให้อยู่ในรูปแบบของข้อความบรรยาย แต่ไม่แปลให้อยู่ในรูปแบบของเสียงพูดเนื่องด้วย โดยปกติแล้วโปรแกรมอ่านหน้าจอสามารถอ่านออกเสียงข้อความที่ปรากฏอยู่บนจอภาพได้ในทุกกรณี ดังนั้นการแปลให้อยู่ในรูปแบบของข้อความก็เพียงพอแล้ว เพื่อให้โปรแกรมอ่านหน้าจอ นำข้อความที่ได้ขึ้นไปประมวลผลและอ่านออกเสียงต่อไป ซึ่งจะช่วยลดความซับซ้อนของระบบลง และลดค่าใช้จ่ายในการพัฒนาระบบได้ โดยวิธีการทำงานดังกล่าวใช้กระบวนการประมวลผลภาพดิจิทัล (Digital Image Processing) และกระบวนการรู้จำตัวอักษร (Optical Character Recognition: OCR) เข้ามาประมวลผล

2. งานวิจัยที่เกี่ยวข้อง

การแปลงผลและการบรรยายรูปภาพให้กับผู้พิการทางสายตาเป็นเรื่องที่ได้รับความสนใจ โดยมีงานวิจัยจำนวนหนึ่งที่พยายามหาวิธีการและแนวทางในการแปลงผลและบรรยายรูปภาพให้อยู่ในรูปแบบต่างๆ โดยเมื่อปี ค.ศ. 2001 Hesham M. Kame และคณะ [6] ได้นำเสนอโปรแกรมวาดภาพไอซีทูดี (IC2D) ที่ช่วยให้ผู้พิการทางสายตาสามารถเข้าถึงรูปภาพได้ด้วยวิธีการติดฉลาก (Labeling Method) ร่วมกับการจัดกลุ่มของจุด เส้นตรง และรูปเรขาคณิต ซึ่ง

โปรแกรมนี้ช่วยให้ผู้พิการทางสายตาสามารถสื่อสารกับคนปกติโดยการวาดและการอ่านรูปภาพที่วาดด้วยโปรแกรมไอซีทูดี้ได้ แต่ยังมีข้อจำกัดสำหรับการอ่านรูปภาพที่ต้องการรายละเอียดสูงหรือรูปภาพที่ใช้สำหรับการอธิบายผล เช่น กราฟ ตาราง แผนผัง เป็นต้น ต่อมาในปี ค.ศ. 2004 R. Iglesias และคณะ [7] ได้เสนอวิธีการใหม่สำหรับการเข้าถึงรูปภาพกราฟิกส์คอมพิวเตอร์สามมิติด้วยการสัมผัสอุปกรณ์แสดงผลพร้อมกับการฟังเสียง ซึ่งวิธีการดังกล่าวทำให้ผู้พิการทางสายตาสามารถเข้าใจความหมายของรูปภาพได้ดีและเร็วยิ่งขึ้น เช่นเดียวกับงานวิจัยของ Stephen E. Krufka และคณะ [8] ที่ได้นำเสนอหลักการของการทำกราฟิกส์เวกเตอร์แบบปรับขนาดได้ (Scalable Vector Graphics) ในการสร้างรูปภาพแบบสัมผัส (Tactile Image) เพื่อให้ผู้พิการทางสายตาได้สัมผัสและเข้าใจถึงรายละเอียดของรูปภาพได้ สอดคล้องกับงานวิจัยของ Richard E. Ladner และคณะ [9] ที่พยายามสร้างรูปแบบการบรรยายรูปภาพด้วยการสัมผัสแบบอัตโนมัติซึ่งทำให้ผู้พิการทางสายตาสามารถเข้าใจรูปภาพที่กำลังเข้าถึง ณ เวลานั้นได้ดียิ่งขึ้น อีกทั้ง David K. McGookin และคณะ [10] ได้ออกแบบและพัฒนาเครื่องมือบรรยายกราฟแท่งหรือชาวด์บาร์ (SoundBar) เพื่ออ่านออกเสียงผลลัพธ์ ค่าระดับ และตำแหน่งของกราฟให้ผู้พิการทางสายตาฟังได้ อย่างไรก็ตามงานวิจัยเหล่านี้ [7-10] ต้องใช้อุปกรณ์และเครื่องมือเสริมในการทำงานซึ่งเป็นอุปกรณ์เฉพาะที่มีข้อจำกัดในการใช้งานและมีราคาแพง ทำให้ผู้พิการทางสายตาส่วนใหญ่ไม่มีโอกาสได้ใช้งานอุปกรณ์เหล่านี้ เมื่อปี ค.ศ. 2005 Steve Murphy [11] จากบริษัทไอบีเอ็ม (IBM) ได้มีแนวคิดที่จะเพิ่มความสามารถของผู้พิการทางสายตาในการเข้าถึงรูปภาพในหน้าเว็บเพจหรือเอกสารเอชทีเอ็มแอล (HTML) ด้วยการสร้างป้ายระบุหรือแท็ก (Tag) พิเศษเพื่อช่วยในการบรรยายรูปภาพที่ประกอบอยู่ในหน้าเว็บเพจ ซึ่งทำให้ผู้พิการทางสายตาสามารถเข้าใจรายละเอียดของรูปภาพได้มากยิ่งขึ้น แต่สามารถประมวลผลกับรูปภาพที่ปรากฏอยู่บนหน้าเว็บเพจได้เพียงอย่างเดียวเท่านั้น

จากงานวิจัยที่กล่าวมาข้างต้นยังมีข้อจำกัดคือ ใช้เวลานาน เสียค่าใช้จ่ายสูง และไม่สะดวกต่อการใช้งาน จึงทำให้เกิดงานวิจัยนี้ขึ้น

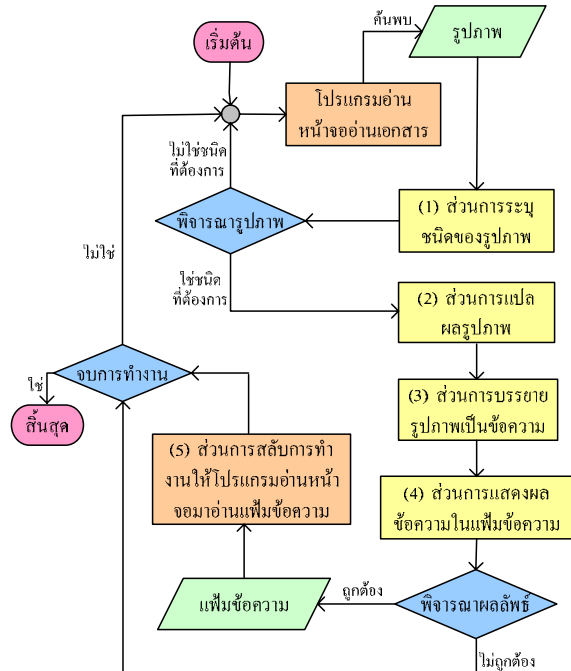
3. สถาปัตยกรรมของระบบ

โดยปกติแล้วเมื่อโปรแกรมอ่านหน้าจอ [2-5] อ่านพบรูปภาพในเอกสารอิเล็กทรอนิกส์ ก็จะข้ามการทำงานในส่วนของรูปภาพนั้นไป หรืออาจแจ้งให้ผู้ใช้ทราบแต่เพียงว่า ณ ตำแหน่งที่อ่านในเวลานั้นคือรูปภาพ แต่ไม่สามารถอธิบายหรือบรรยายถึงรายละเอียดภายในรูปภาพเหล่านั้นออกมาได้ งานวิจัยนี้จึงมีแนวคิดที่จะออกแบบโปรแกรมย่อยสำหรับแปลผลและบรรยายรูปภาพเหล่านั้น เพื่อให้ผู้พิการทางสายตาได้อ่านและสามารถเข้าใจถึงรายละเอียดของเอกสารซึ่งมีรูปภาพเป็นองค์ประกอบได้ดียิ่งขึ้น

ตามโครงสร้างและสถาปัตยกรรมของงานวิจัยนี้ เมื่อโปรแกรมอ่านหน้าจออ่านพบรูปภาพจากเอกสารอิเล็กทรอนิกส์ใดๆ จะทำการสลับการทำงานมาที่โปรแกรมย่อยที่ทำหน้าที่แปลผลและบรรยายรูปภาพเพื่อประมวลผลกับรูปภาพที่ตรวจพบ แล้วจึงย้อนกลับมาทำงานกับเอกสารอิเล็กทรอนิกส์เดิมอีกครั้งหนึ่ง โดยขั้นตอนของสถาปัตยกรรมระบบได้แสดงไว้ตามรูปที่ 1 ดังรายละเอียดต่อไปนี้

- (1) ส่วนการระบุชนิดของรูปภาพ ทำหน้าที่บอกชนิดของรูปภาพที่ตรวจพบว่าเป็นรูปภาพชนิดใด เช่น ตาราง กราฟ แผนภูมิ แผนผัง หรือภาพวาด เป็นต้น เพื่อตัดสินใจว่าการแปลผลสามารถทำต่อไปได้หรือไม่
- (2) ส่วนการแปลผลรูปภาพ ทำหน้าที่แปลผลรูปภาพตามอัลกอริทึมสำหรับรูปภาพแต่ละชนิด
- (3) ส่วนการบรรยายรูปภาพเป็นข้อความ ทำหน้าที่บรรยายรูปภาพที่ได้จากส่วนของขั้นตอนที่ 2 ออกมาเป็นตัวหนังสือหรือข้อความบรรยาย
- (4) ส่วนการแสดงผลข้อความในแฟ้มข้อความ จะนำข้อความที่ได้มาแสดงผลในแฟ้มข้อความ (Text File) ใหม่ที่อยู่ภายนอกเอกสารเดิม

(5) ส่วนการสลับการทำงานให้โปรแกรมอ่านหน้าจอมาอ่านเพิ่มข้อความ เป็นการสลับการทำงานของโปรแกรมอ่านหน้าจอเพื่ออ่านเพิ่มข้อความแล้วจึงสลับกลับไปอ่านยังเพิ่มเอกสารเดิม



รูปที่ 1 สถาปัตยกรรมแนวคิดของระบบการแปลผลและการบรรยายรูปภาพสำหรับผู้พิการทางสายตา

งานวิจัยนี้ได้เลือกรูปภาพตารางซึ่งมีโครงสร้างที่ไม่ซับซ้อนมาใช้เป็นกรณีศึกษาและรูปภาพทดสอบ โดยเจาะจงในรายละเอียดการทำงานของขั้นตอนที่ (1) (2) และ (3) เป็นหลัก ซึ่งขั้นตอน วิธีการ และผลลัพธ์จากการประมวลผลต่างๆ ได้แสดงไว้ในหัวข้อถัดไป

4. วิธีการแปลผลและบรรยายรูปภาพ

งานวิจัยนี้ใช้เทคนิคการประมวลผลภาพดิจิทัล และการรู้จำตัวอักษรเพื่อใช้แปลผลและบรรยายรูปภาพ เนื่องจากรูปภาพมีหลายชนิด โดยแต่ละชนิดมีลักษณะเฉพาะที่แตกต่างกัน จึงมีรายละเอียดของขั้นตอนการแปลผลและบรรยายรูปภาพที่แตกต่างกัน งานวิจัยนี้จึงเลือกเฉพาะรูปภาพตารางเพื่อใช้เป็นกรณีศึกษา เนื่องจากรูปภาพตารางเป็นรูปภาพพื้นฐานที่มีการใช้งานกันอย่างกว้างขวาง และมี

โครงสร้างที่ไม่ซับซ้อนในการประมวลผล โดยขั้นตอนการแปลผลและการบรรยายรูปภาพสามารถแสดงรายละเอียดได้ดังต่อไปนี้



ก.

Table 1 Expense List		
NO	Expense	
	Item	Price
1	Book	200
2	Pencil	50
3	Snack	70
4	Red Cap	150

ข.

Table 1 Expense List		
NO	Expense	
	Item	Price
1	Book	200
2	Pencil	50
3	Snack	70
4	Red Cap	150

ค.

รูปที่ 2 ตัวอย่างรูปภาพตาราง

- รายละเอียดของรูปภาพตาราง ซึ่งประกอบด้วย ความกว้างและความสูงของภาพ ความกว้างและความสูงจริงของภาพตาราง และความกว้างและความสูงของตาราง
- รูปภาพตารางต้นฉบับ (ก่อนการประมวลผล)
- รูปภาพตารางที่ผ่านการหาเส้นขอบ และระบุขอบเขตของรูปภาพ (หลังการประมวลผล)

4.1 การนิยามรูปภาพตาราง

ตาราง ประกอบด้วยแถว (Rows) และหลัก (Columns) ที่มีการจัดเรียงกันอย่างเป็นระเบียบ [12] สำหรับการนิยามรูปภาพตาราง งานวิจัยนี้ได้มีการกำหนดรายละเอียดและคุณลักษณะเฉพาะต่างๆ ของรูปภาพตารางดังนี้

- เป็นเพิ่มรูปภาพ เช่น เพิ่มนามสกุล .gif .jpg .bmp .png เป็นต้น
- เป็นรูปตารางที่ประกอบด้วยแถวและหลัก
- ใช้เส้นตรงในการแบ่งแต่ละแถวและหลัก
- ไม่เป็นรูปภาพที่มีหนึ่งแถวและหนึ่งหลัก

- ช่องแต่ละช่องในรูปภาพตาราง เรียกว่าเซลล์ซึ่งสามารถเชื่อมต่อกันได้

4.2 โครงสร้างของรูปภาพตาราง

การศึกษาโครงสร้างของรูปภาพตารางเป็นการวิเคราะห์ถึงชนิดของรูปภาพตารางและลักษณะเฉพาะของรูปภาพตาราง โดยใช้นิยามจากหัวข้อ 4.1 ซึ่งโครงสร้างของรูปภาพตารางนั้นประกอบด้วย (1) ชื่อตารางซึ่งอาจปรากฏอยู่ส่วนบนสุดหรือล่างสุดของรูปภาพตาราง และ (2) ตัวตารางซึ่งเป็นส่วนที่เหลือทั้งหมดของรูปภาพตาราง

ส่วนของตัวตารางนั้นจะประกอบไปด้วยเซลล์ต่างๆ โดยที่เซลล์แต่ละเซลล์นั้นเป็นรูปสี่เหลี่ยมที่ถูกแบ่งด้วยเส้นตรงซึ่งมีจุดพิคคตำแหน่งของมุมทั้งสี่ของเซลล์ และอาจมีตัวหนังสืออยู่ภายในเซลล์เหล่านั้น ดังแสดงตัวอย่างในรูปที่ 2ก.

4.3 การแปลผลรูปภาพตาราง

กระบวนการแปลผลรูปภาพตารางเป็นการนำรายละเอียดของโครงสร้างรูปภาพตาราง มาพิจารณาเพื่อตรวจสอบว่ารูปภาพใดบ้างที่มีลักษณะเป็นรูปภาพตาราง และไม่ใช่อรูปภาพตาราง ซึ่งมีขั้นตอนดังต่อไปนี้

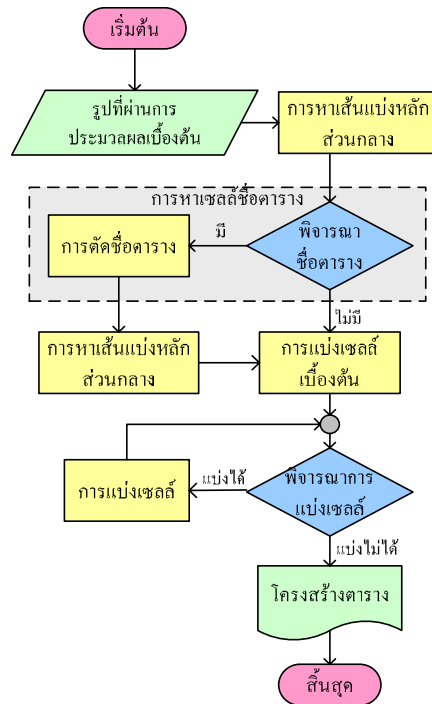
4.3.1 การประมวลผลเบื้องต้น

โดยทั่วไปรูปภาพตารางจะมีส่วนของพื้นที่ว่างบริเวณขอบของรูปภาพก่อนถึงเส้นขอบของรูปตารางที่แท้จริง ซึ่งขั้นตอนนี้เป็นการหาเส้นขอบและระบุขอบเขตของรูปภาพตารางที่แท้จริง โดยใช้เทคนิคการประมวลผลภาพดิจิทัล และแปลงให้เป็นภาพขาวดำเพื่อลดทรัพยากรของระบบ และเพื่อง่ายต่อการคำนวณเกี่ยวกับจำนวนจุดภาพ (Pixel) การหาขอบเขตของรูปภาพตารางที่แท้จริงทำได้โดยการหาขอบเขตสี่เหลี่ยมที่เล็กที่สุดที่สามารถบรรจุจุดภาพสีดำไว้ได้ทั้งหมด ผลลัพธ์ของขั้นตอนนี้แสดงไว้ในรูปที่ 2ก.

4.3.2 การหาโครงสร้างของตาราง

การหาโครงสร้างของตารางจะเริ่มต้นจากการหาส่วนของชื่อตาราง และส่วนของเซลล์ที่เป็นเนื้อหาภายในตาราง ขนาดแถวคูณหลัก โดยใช้วิธีการหาเส้นตรงที่ยาวเพียงพอ

ในแนวนอนและแนวตั้ง เรียกว่าเส้นแบ่งหลักส่วนกลาง เพื่อแบ่งแถวและหลักในเบื้องต้นก่อน หลังจากนั้นแต่ละเซลล์จะถูกพิจารณาว่ามีเส้นแบ่งหลักเฉพาะที่ซึ่งสามารถแบ่งเซลล์นั้นออกได้อีกหรือไม่ โปรแกรมจะแบ่งจนกว่าไม่มีเซลล์ใดสามารถแบ่งได้อีก ขั้นตอนการทำงานแสดงดังรูปที่ 3 ซึ่งรายละเอียดของขั้นตอนย่อยต่างๆ มีดังนี้



รูปที่ 3 ผังงานของการค้นหาโครงสร้างตาราง

(1) การหาเส้นแบ่งหลักส่วนกลาง ทำได้โดยการคำนวณอัตราส่วนของผลรวมของสี (กำหนดให้ค่าสีดำเป็น 1 สีขาวเป็น 0) ในแถวเดียวกันเทียบกับความกว้างของภาพแล้วเลือกเก็บตำแหน่งแถวที่มีค่าอัตราส่วนนี้มากกว่าค่าอัตราส่วนขั้นต่ำของเส้นหลัก (α) สำหรับค่าตำแหน่งของหลักก็สามารถคำนวณได้ในทำนองเดียวกัน กำหนดเซตของตำแหน่งแถว (R) และเซตของตำแหน่งหลัก (C) ดังสมการที่ (1) และ (2) ตามลำดับ

$$R = \left\{ r \mid 1 \leq r \leq h, \frac{\sum_{c=1}^w B(r, c)}{w} \geq \alpha \right\} \quad (1)$$

$$C = \left\{ c \mid 1 \leq c \leq w, \frac{\sum_{r=1}^h B(r, c)}{h} \geq \alpha \right\} \quad (2)$$

เมื่อ h และ w คือความสูงและความกว้างของรูปภาพ $B(r, c)$ คือค่าสีในจุดภาพแถวที่ r หลักที่ c และค่าสัดส่วนขั้นต่ำของเส้นหลัก (α) นี้จะไม่เกิน 1.00 และไม่ควรมีน้อยกว่า 0.90 เพื่อให้มั่นใจได้ว่า ตำแหน่งที่เลือกมาเป็นเส้นตรง ไม่ใช่ส่วนของตัวอักษรที่มีความสูงใกล้เคียงกับความสูงของเซลล์ที่อักษรนั้นอยู่

สมาชิกในแต่ละเซตจะถูกนำมาเรียงลำดับ และหาผลต่างของกลุ่มตัวเลขที่ติดกัน กลุ่มที่มีผลต่างมากกว่าค่าขั้นต่ำของเซลล์ (β) หรือค่าจำนวนจุดภาพขั้นต่ำของความกว้างและความยาวของเซลล์ จะถูกเลือกไปใช้ในการแบ่งเซลล์เบื้องต้น

(2) การหาชื่อตาราง และการแบ่งเซลล์เบื้องต้น ชื่อตารางจะอยู่ได้ทั้งบริเวณเหนือหรือใต้ของตาราง ซึ่งสามารถหาได้โดยพิจารณาขอบเขตของรูปภาพที่ตำแหน่งบนสุดหรือล่างสุดที่ไม่ใช่เส้นตรง (ตำแหน่งของเส้นตรงได้คำนวณไว้แล้วจากสมการที่ (1)) โดยที่ชื่อตารางนี้จะถูกเก็บเป็นส่วนหนึ่งของโครงสร้างตาราง และขณะเดียวกันก็ทำการลดขนาดของภาพลงเป็นเพียงส่วนของตารางที่แท้จริง จากนั้นหาเส้นแบ่งหลักส่วนกลางใหม่อีกครั้งเพื่อบอกแถวและหลักของตาราง และเก็บข้อมูลของเซลล์เริ่มต้นในอาร์เรย์ของเซลล์

(3) การตรวจสอบเซลล์ ในขั้นตอนนี้จะทำการตรวจสอบทุกๆ เซลล์ว่าสามารถแบ่งออกได้หรือไม่จากเซลล์ทางด้านซ้ายไปขวา และจากด้านบนลงล่าง โดยใช้หลักการเกี่ยวกับการหาเส้นแบ่งหลักส่วนกลาง แต่ในที่นี้จะเป็นการหาเส้นแบ่งหลักเฉพาะเซลล์แทน หากพบว่าเซลล์ใดสามารถแบ่งได้ หรือมีจำนวนแถวหรือหลักที่มากกว่า 1 แล้ว จะแบ่งเซลล์เหล่านั้นอีกครั้งจนกว่าจะไม่มีเซลล์ใดที่สามารถแบ่งได้อีก ซึ่งถือว่าสิ้นสุดในขั้นตอนนี้

(4) การแบ่งเซลล์ ขั้นตอนนี้จะทำงานเมื่อขั้นตอนการตรวจสอบเซลล์ตรวจพบว่ามีเซลล์ที่สามารถแบ่งได้ กรณีการแบ่งเซลล์เพิ่มจำนวนแถว เซลล์ทั้งหมดที่อยู่ด้านใต้เซลล์ที่พิจารณาจะถูกเลื่อนลงตามจำนวนแถวที่เพิ่ม และแบ่งเซลล์ที่พิจารณาออก โดยการใส่ข้อมูลในโครงสร้าง

ตารางตามขอบเขตที่ได้จากการหาเส้นแบ่งเฉพาะที่ ส่วนเซลล์อื่นๆ ที่เคยอยู่ในแถวเดียวกันจะแบ่งเป็นสองด้าน คือเซลล์ทางด้านซ้ายซึ่งเคยถูกตรวจสอบมาก่อนหน้านี้แล้วที่ไม่สามารถแบ่งได้ซึ่งจะถูกคัดลอกมาทั้งหมดเพื่อใส่ให้กับเซลล์ที่เพิ่มทุกๆ แถว ส่วนเซลล์ทางด้านขวานั้นจะถูกตรวจสอบก่อนว่าแบ่งแถวได้หรือไม่ ถ้าแบ่งไม่ได้ก็จะทำเช่นเดียวกับเซลล์ทางซ้าย แต่ถ้าแบ่งได้ก็จะแบ่งตามจำนวนเดียวกับเซลล์ที่กำลังพิจารณา

ตัวอย่างรูปตารางต้นฉบับดังแสดงในรูปที่ 2x. นั้น เมื่อผ่านขั้นตอนการหาโครงสร้างของตารางเสร็จสมบูรณ์แล้วจะได้เป็นอาร์เรย์ของเซลล์ตารางดังแสดงในรูปที่ 4

Table 1 Expense List

Table 1 Expense List		
Title		
NO	Expense	Expense
Cell-1-1	Cell-1-2	Cell-1-3
NO	Item	Price
Cell-2-1	Cell-2-2	Cell-2-3
1	Book	200
Cell-3-1	Cell-3-2	Cell-3-3
2	Pencil	50
Cell-4-1	Cell-4-2	Cell-4-3
3	Snack	70
Cell-5-1	Cell-5-2	Cell-5-3
4	Red Cap	150
Cell-6-1	Cell-6-2	Cell-6-3

รูปที่ 4 ตัวอย่างอาร์เรย์ของเซลล์จากรูปตารางต้นฉบับของรูปที่ 2x. ที่ผ่านการหาโครงสร้างตาราง

4.3.3 การรู้จำตัวอักษร

เมื่อได้โครงสร้างของตารางที่สามารถอ้างอิงถึงตำแหน่งของแต่ละเซลล์ในรูปภาพตารางที่รับเข้ามาและถูกแบ่งออกเป็นเซลล์แต่ละเซลล์เรียบร้อยแล้ว ขั้นตอนนี้จะเป็นการระบุการรู้จำตัวอักษร ซึ่งงานวิจัยนี้ได้เลือกใช้โปรแกรมสำเร็จรูป SimpleOCR [13] ในการประมวลผลและเป็นเพียงขั้นตอนอย่างง่ายเท่านั้น โดยเพียงต้องการที่จะเลือกตัวอย่างของอัลกอริทึมรู้จำตัวอักษรในการทดสอบรูปภาพของเซลล์แต่ละเซลล์ที่ได้จากขั้นตอนที่ 4.3.2 ว่า

สามารถรู้จำตัวอักษรได้หรือไม่ ซึ่งความถูกต้องของการรู้จำตัวอักษรจะขึ้นอยู่กับกระบวนการทำงานของอัลกอริทึมรู้จำตัวอักษรที่นำมาใช้

4.4 การบรรยายรูปภาพตาราง

ในขั้นตอนนี้ โครงสร้างของตารางที่ได้จากขั้นตอนที่ 4.3 จะถูกบรรยายเป็นข้อความ ซึ่งใช้การบรรยายเป็นข้อความสั้นๆ เพื่อให้ผู้พิการทางสายตาสามารถทำความเข้าใจได้ง่าย และไม่เสียเวลาในการฟัง ดังได้แสดงตัวอย่างคำบรรยายไว้ในรูปที่ 5 ซึ่งข้อความในแต่ละเซลล์จะขึ้นบรรทัดใหม่ ทำให้สามารถกดปุ่มลูกศรขึ้นลงเพื่อเลือกบรรทัดที่จะฟังได้

```
6-Row-3-Column Tabular Image
Title : Table 1 Expense List
Cell-Row1-Column1 : NO
Cell-Row1-Column2 : Expense
Cell-Row1-Column3 : Expense
Cell-Row2-Column1 : NO
Cell-Row2-Column2 : Item
Cell-Row2-Column3 : Price
Cell-Row3-Column1 : 1
Cell-Row3-Column2 : Book
Cell-Row3-Column3 : 200
Cell-Row4-Column1 : 2
Cell-Row4-Column2 : Pencil
Cell-Row4-Column3 : 50
Cell-Row5-Column1 : 3
Cell-Row5-Column2 : Snack
Cell-Row5-Column3 : 70
Cell-Row6-Column1 : 4
Cell-Row6-Column2 : Red Cap
Cell-Row6-Column3 : 150
End of Table
```

รูปที่ 5 ผลลัพธ์การแสดงข้อความคำบรรยายของรูปภาพตารางให้กับผู้พิการทางสายตา

5. การทดลองและการประเมินผล

ในการประเมินผลการแปลผลและการบรรยายรูปภาพตารางนั้น งานวิจัยนี้ได้ใช้ชุดของข้อมูลรูปภาพนามสกุลต่างๆ จำนวน 50 รูปแรกที่เหมาะสมที่ได้จากการค้นหาใน Google ด้วยคำค้น “Tabular Image” โดยเกณฑ์การวัดที่ใช้ได้แก่ (1) ความถูกต้องของการระบุได้ว่าเป็นตาราง และ (2) ความถูกต้องของการแบ่งเซลล์ตาราง

ในการทดลอง มีการใส่ค่าพารามิเตอร์สองค่าคือ ค่าสัดส่วนขั้นต่ำของเส้นหลัก (α) และค่าขั้นต่ำของเซลล์ (β) เป็น 0.96 และ 10 จุดภาพ ตามลำดับ ซึ่งเป็นค่าที่ดี

ที่สุดของค่าต่างๆ ที่ใช้ในการทดสอบ โดยผลการทดสอบแสดงได้ในตารางที่ 1

ตารางที่ 1 ความถูกต้องของอัลกอริทึมของชุดรูปภาพตารางทดสอบจำนวน 50 รูป ด้วยคำค้น “Tabular Image” จาก Google

เกณฑ์การวัด	จำนวนรูปภาพ	เปอร์เซ็นต์
ความถูกต้องของการระบุได้ว่าเป็นตาราง	48	96.0%
ความถูกต้องของการแบ่งเซลล์ตาราง	43	89.6%

จากตารางที่ 1 ความถูกต้องของการระบุได้ว่าเป็นรูปตารางจากรูปภาพจำนวน 50 รูป เป็น 96.0 เปอร์เซ็นต์ และความถูกต้องของการแบ่งเซลล์ตารางจากรูปภาพจำนวน 48 รูปที่ระบุได้แล้วว่าเป็นรูปตารางเป็น 89.6 เปอร์เซ็นต์ ซึ่งเมื่อพิจารณาผลการทำงานโดยรวมแล้ว อัลกอริทึมนี้สามารถประมวลผลรูปภาพตารางได้ถูกต้องจำนวน 43 รูปจากรูปภาพทั้งหมด 50 รูป ซึ่งคิดเป็น 86.0 เปอร์เซ็นต์ โดยปัจจัยที่ทำให้เกิดความผิดพลาดสามารถวิเคราะห์ได้เป็นสองกลุ่ม ได้แก่

(1) ความผิดพลาดที่โปรแกรมระบุได้ว่าไม่ใช่รูปภาพตาราง (False Negative) เกิดจากการที่อัลกอริทึมไม่สามารถหาเส้นตรงที่บอกรอบเขตของรูปภาพหรือขอบเขตของเซลล์ได้ ซึ่งอาจเป็นเพราะการใช้สีที่ใกล้เคียงกันมากเป็นตัวแบ่งแถวหรือหลัก รูปภาพมีความไม่สมบูรณ์ รูปภาพมีการหมุน รูปภาพมีจุดรบกวน (Noise) และอาจเกิดกรณีหนึ่งแถวและหนึ่งหลักขึ้น

(2) ความผิดพลาดในการแบ่งแยกเซลล์ เกิดจากในบางเซลล์มีตัวอักษรที่มีความสูงเกือบเท่ากับความสูงของเซลล์นั้นๆ ทำให้อัลกอริทึมประมวลผลได้ว่าตัวอักษรตัวนั้นเป็นเส้นแบ่งหลักเฉพาะของเซลล์ด้วย ทำให้การแบ่งแถวและการแบ่งหลักเกิดความผิดพลาดขึ้น

6. บทสรุปและแนวทางการวิจัยในอนาคต

งานวิจัยนี้มีแนวคิดในการแปลผลและการบรรยายรูปภาพให้ออกมาเป็นตัวหนังสือหรือข้อความคำบรรยาย

เพื่อให้ผู้พิการทางสายตาสามารถใช้โปรแกรมอ่านหน้าจออ่านออกเสียง และเข้าใจถึงรายละเอียดของเอกสารอิเล็กทรอนิกส์เหล่านั้นได้ดียิ่งขึ้น โดยใช้หลักการของการประมวลผลภาพดิจิทัลร่วมกับการรู้จำตัวอักษร งานวิจัยนี้ได้เลือกเฉพาะรูปภาพตารางเป็นกรณีศึกษาเนื่องจากมีการใช้งานอย่างกว้างขวาง และไม่ซับซ้อนต่อการอธิบายและทำการทดลอง ซึ่งค้นแบบอัลกอริทึมการประมวลผลรูปภาพที่นำเสนอได้แก่ การนิยามรูปภาพตาราง การออกแบบโครงสร้างของรูปภาพตาราง การแปลผลรูปภาพตาราง การรู้จำตัวอักษร และการบรรยายรูปภาพตารางเป็นข้อความ ในการประเมินผลการทำงานสามารถแบ่งได้เป็นสองกรณีคือ ความถูกต้องที่ระบุได้ว่าเป็นรูปภาพตารางและความถูกต้องในการแบ่งแยกเซลล์

งานวิจัยนี้ต้องการนำเสนอแนวคิดใหม่ในการแปลผลและบรรยายรูปภาพ และนำเสนอค้นแบบอัลกอริทึมในการประมวลผลกับรูปภาพตาราง ซึ่งจะเป็แนวทางหนึ่งในการช่วยเหลือให้ผู้พิการทางสายตาสามารถเข้าถึงรูปภาพและเข้าใจความหมายของรูปภาพได้หลากหลายชนิดต่อไปในอนาคต

แนวทางสำหรับงานวิจัยต่อไป จะมีการนำเสนอวิธีการเพิ่มประสิทธิภาพการทำงานของอัลกอริทึมของงานวิจัยนี้ เพื่อให้สามารถใช้ประโยชน์จากการแปลผลและบรรยายรูปภาพได้ดียิ่งขึ้น เช่น สามารถประมวลผลได้กับรูปภาพชนิดต่างๆ สามารถระบุตำแหน่ง ขอบเขต และขนาดของรูปภาพในเอกสารได้ ซึ่งจะช่วยให้ผู้พิการทางสายตาสามารถเข้าถึงรายละเอียดของรูปภาพได้มากยิ่งขึ้น

7. กิตติกรรมประกาศ

ขอขอบคุณนายคณิตย์ ผามะณี ซึ่งเป็นผู้พิการทางสายตา และเป็นผู้ที่จุดประกายความคิดในการพัฒนางานวิจัยเรื่องนี้ นอกจากนี้ขอขอบคุณคณาจารย์ที่ให้ข้อเสนอแนะที่เป็นประโยชน์ซึ่งประกอบด้วย อ.ดร.โชติรัตน์ รัตนามัทธนะ อ.ดร.อดิวงส์ สุขาโต อ.ดร.พิชญคนองชัยยศ และ อ.ดร.อรรณสิทธิ์ สุรฤกษ์

8. เอกสารอ้างอิง

- [1] Wikipedia, "Screen reader," Available from: http://en.wikipedia.org/wiki/Screen_reader, Access date: July 3, 2007.
- [2] JAWS, Available from: <http://www.freedomscientific.com>, Access date: July 3, 2007.
- [3] Apple VoiceOver, Available from: <http://www.apple.com/macosx/features/voiceover/>, Access date: July 3, 2007.
- [4] Adobe Read Out Loud, Available from: <http://www.adobe.com/enterprise/accessibility/reader6/sec2.html>, Access date: July 3, 2007.
- [5] PPA Tatip, Available from: <http://www.tabod.net>, Access date: July 3, 2007.
- [6] H. M. Kamel and J. A. Landay, "The Use of Labeling to Communicate Detailed Graphics in a Non-visual Environment," In Proc. of the Conf. on Human Factors in Computing Systems, Washington, USA, 2001.
- [7] R. Iglesias, S. Casado, T. Gutierrez, J. I. Barbero, C. A. Avizzano, S. Marcheschi and M. Bergamasco, "Computer Graphics Access for Blind People through a Haptic and Audio Virtual Environment," In Proc. of 3rd IEEE Int. Workshop on Haptic, Audio and Visual Environments and Their Applications, 2004.
- [8] S. E. Krufka and K. E. Barner, "Automatic Production of Tactile Graphics from Scalable Vector Graphics," In Proc. of 7th Int. ACM SIGACCESS Conf. on Computers and Accessibility, USA, 2005.
- [9] R. E. Ladner, M. Y. Ivory, R. Rao, S. Burgstahler, D. Comden, S. Hahn, M. Renzelmann, S. Krisnandi, M. Ramasamy, B. Slabosky, A. Martin, A. Lacenski, S. Olsen and D. Groce, "Automating Tactile Graphics Translation," In Proc. of 7th Int. ACM SIGACCESS Conf. on Computers and Accessibility, USA, 2005.
- [10] D. K. McGookin and S. A. Brewster, "SoundBar: Exploiting Multiple Views in Multimodal Graph Browsing," In Proc. of 4th Nordic Conf. on Human-Computer Interaction, Oslo, Norway, 2006.
- [11] S. Murphy, "Accessibility of Graphics in Technical Documentation for the Cognitive and Visually Impaired," In Proc. of 23rd Annual Int. Conf. on Design of Communication, UK, 2005.
- [12] Wikipedia, "Basic description of table," Available from: [http://en.wikipedia.org/wiki/Table_\(information\)](http://en.wikipedia.org/wiki/Table_(information)) Access date: July 3, 2007.
- [13] Meta Enterprises, LLC, "SimpleOCR Version 3.1," Available from: <http://www.simpleOcr.com>, Access date: July 3, 2007.