การระบุตำแหน่งข้อความภาษาไทยในภาพถ่ายฉากธรรมชาติ

นายธนานพ กอบชัยสวัสดิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557

THAI TEXT LOCALIZATION IN NATURAL SCENE IMAGES

Mr. Thananop Kobchaisawat

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2014

หัวข้อวิทยานิพนธ์	การระบุตำแหน่งข้อความภาษาไทยในภาพถ่ายฉากธรรมชาติ
โดย	นายธนานพ กอบชัยสวัสดิ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. ธนารัตน์ ชลิดาพงศ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร. บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ศาสตราจารย์ ดร. บุญเสริม กิจศิริกุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร. ธนารัตน์ ชลิดาพงศ์)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ)

.....กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร. กันต์พงษ์ วรรัตน์ปัญญา)

ธนานพ กอบชัยสวัสดิ์ : การระบุตำแหน่งข้อความภาษาไทยในภาพถ่ายฉากธรรมชาติ (THAI TEXT LOCALIZATION IN NATURAL SCENE IMAGES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. ธนารัตน์ ชลิดาพงศ์, 94 หน้า.

งานวิจัยนี้นำเสนอวิธีการระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติ โดยใช้เทคนิคการเรียนรู้ของเครื่องจักรที่สามารถเรียนรู้ตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน ร่วมกับเทคนิคการประมวลผลภาพ สามารถทำงานได้อย่างอัตโนมัติโดยไม่ต้องอาศัยการกำหนดค่าเริ่มต้นก่อนการทำงาน

ขั้นตอนวิธีที่นำเสนอนั้นประกอบไปด้วย 4 ขั้นตอนหลักคือ การประมวลผลก่อน การสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ การรวมผลจากตัวจับข้อความ และการประมวลผลภายหลัง ขั้นตอนการประมวลก่อนนั้น จะสร้างภาพนำเข้าหลายขนาดและปรับปรุงคุณภาพของภาพนำเข้า ให้มีความเหมาะสมก่อนนำไปสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ ซึ่งในขั้นตอนนี้จะใช้ตัวจำแนกประเภทที่ได้จากการเรียนรู้ของเครื่องจักรที่สามารถเรียนรู้ตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน ในการจำแนกระหว่างบริเวณที่เป็นข้อความและไม่ใช่ข้อความในแต่ละขนาดของภาพนำเข้า หลังจากนั้นจะทำการรวมแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความในแต่ละขนาดของภาพนำเข้า ซึ่งนำไปสู่ขั้นตอนสุดท้ายที่จะนำแผนภาพนี้ไปสร้างสมมุติฐานของบรรทัดของบริเวณที่มีข้อความและทำการประมวลผลภายหลัง โดยมีขั้นตอนที่ถูกออกแบบสำหรับการวิเคราะห์ข้อความภาษาไทย เพื่อสร้างกรอบของบริเวณที่มีข้อความปรากฏอยู่ในภาพนำเข้า

ผลการทดลองบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษ ชุดข้อมูลทดสอบภาษาไทยและภาษาอังกฤษจากการแข่งขัน BEST 2015 : การแข่งขันสุดยอดการหาตำแหน่งข้อความบนภาพถ่าย ที่จัดขึ้นโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ และชุดข้อมูลผสมระหว่างภาษาไทยและภาษาอังกฤษที่จัดหาโดยผู้วิจัยพบว่า ขั้นตอนวิธีที่ทำการเสนอนี้สามารถทำงานได้อย่างมีประสิทธิภาพ สามารถระบุตำแหน่งข้อความในหลายสถานการณ์ เช่น ภาพที่มีปรากฏการมัวจากการเคลื่อนที่ ข้อความที่ปรากฏในภาพมีความหลากหลายทั้งในแง่ของรูปแบบของอักษรและการจัดเรียง ข้อความ และข้อความที่มีฉากหลังที่ซับซ้อน จากผลการทดลองบนชุดข้อมูลทดสอบที่เสนอทั้งหมดนั้น จะได้ค่าความเที่ยงเฉลี่ย 73% ค่ารีคอลเฉลี่ย 70% และค่าการวัดเอฟเฉลี่ย 72%

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2557

5570544421 : MAJOR COMPUTER ENGINEERING

KEYWORDS: THAI TEXT LOCALIZATION / CONVOLUTIONAL NEURAL NETWORK

THANANOP KOBCHAISAWAT: THAI TEXT LOCALIZATION IN NATURAL SCENE IMAGES. ADVISOR: ASST. PROF. DR. THANARAT CHALIDABHONGSE, 94 pp.

This research proposes a method to locate Thai natural scene text. The method uses a trainable feature extractor machine learning and image processing techniques. The system works automatically without prior configuation.

The proposed method is consisted of 4 main steps, preprocessing, text confidence maps construction, text confidence map merging and postprocessing. In the preprocessing step, the multi-scaled input images are constructed together with image enhancement to improve quality of input image. The text confidence map construction step uses the trained text detector to classify between text and non-text areas. Then text confidence map for each scaled input image is built and merged to acquire a final input image text confidence map. Finally, in the postprocessing stage, the estimated text lines are calculated collaborating with Thai text analysis to generate final text bounding boxes.

The experimental results on standard English test datasets, BEST 2015 : Text Location Detection Contest dataset and our mixed Thai-English dataset show that our proposed method can locate natural scene text in many scenarios. For examples, texts in blur images, multiple text orientations, various text styles, text with the effects of perspective distortion, and texts on complex background. From the experimental results on selected datasets, we get average precision 73%, average recall 70% and average f-measure 72%.

Department: Computer Engineering Student's Signature .....

Field of Study: Computer Engineering Advisor's Signature .....

Academic Year: 2014

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ เกิดขึ้นได้จากการชี้แนะแนวทางในการดำเนินงานวิจัยและการศึกษาพัฒนางานวิจัยร่วมกับ ผศ. ดร. ธนารัตน์ ชลิดาพงศ์ อาจารย์ที่ปรึกษาวิทยาพนธ์ จึงขอขอบพระคุณอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณคณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ เจ้าหน้าที่ทุกท่าน และเพื่อนๆในห้องปฏิบัติการคอมพิวเตอร์กราฟิกและวิทยาการภาพทุกท่าน ที่คอยสนับสนุนการทำงานให้เป็นไปได้อย่างราบรื่น รวมถึงบิดา มารดา และครูอาจารย์ทุกท่านที่ได้ให้การศึกษาและอบรมเป็นอย่างดีเสมอมา

สารบัญ

หน้า

หน้า

หน้า

สารบัญรูปภาพ

สารบัญตาราง

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การสื่อสารเป็นสิ่งจำเป็นสิ่งหนึ่งสำหรับการดำรงชีวิตของมนุษย์ในปัจจุบัน ซึ่งรูปแบบของการสื่อสารที่พบในชีวิตประจำวันนอกเหนือจากการสื่อสารด้วยเสียงแล้ว ยังมีการสื่อสารจากภาพและข้อความต่างๆ ซึ่งเราสามารถพบข้อความปรากฏอยู่ทั่วไป ทั้งบนท้องถนน ภายในอาคาร และสื่อต่างๆ ดังแสดงตัวอย่างในรูปที่ 1-1 ข้อความเหล่านี้เป็นสิ่งสำคัญที่จะสื่อสาร และให้ข้อมูลกับผู้อ่าน ซึ่งถ้าเราสามารถสกัดและบอกได้ว่าข้อความในภาพประกอบด้วยตัวอักษรตัวใดได้ จะสามารถนำไปใช้ประโยชน์กับงานวิจัยทางด้านอื่นๆ ได้อีกมาก

รูปที่ 1-1 ตัวอย่างข้อความทั่วไปที่ปรากฏในชีวิตประจำวัน ซึ่งมีข้อความต่างๆปรากฏอยู่ปะปนอยู่โดยทั่วไป

การแปลภาษาอัตโนมัติจากภาพ เป็นงานหนึ่งที่สามารถใช้ประโยชน์จากการอ่านข้อความในภาพได้ ตัวอย่างเช่น การช่วยแปลภาษาของป้ายในสถานที่ท่องเที่ยวหรือตามเส้นทางการเดินทางให้เป็นภาษาที่นักท่องเที่ยวสามารถเข้าใจได้ซึ่งเพิ่มความสะดวกและช่วยเหลือให้นักท่องเที่ยวเดินทางได้ง่ายขึ้น

การวิจัยทางด้านหุ่นยนต์ก็เป็นอีกด้านหนึ่ง ที่สามารถใช้ข้อมูลจากการอ่านข้อความในภาพที่หุ่นยนต์เห็นนั้น มาช่วยให้หุ่นยนต์สามารถตัดสินใจโต้ตอบกับสภาพแวดล้อมได้หลากหลายยิ่งขึ้น ยกตัวอย่างเช่น งานวิจัยทางด้านการสร้างระบบขับรถอัตโนมัติแบบไร้คนขับ ซึ่งต้องอาศัยความสามารถในการอ่านป้าย และเครื่องหมายจราจร เพื่อตัดสินใจว่าจะควบควบคุมทิศทางและความเร็วในการควบคุมรถอย่างไร

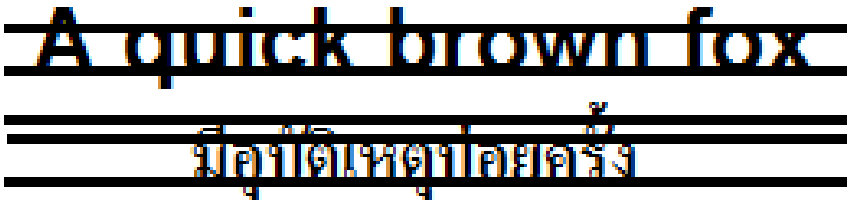
งานวิจัยด้านความเป็นจริงเสมือน (augmented reality) ก็สามารถใช้ข้อมูลที่ได้จากอ่านข้อความนั้น มาช่วยเพิ่มข้อมูลของสิ่งต่างๆที่ปรากฏภายในความจริงเสมือนที่สร้างขึ้น ยกตัวอย่างเช่น Google Glass ซึ่งเป็นผลิตภัณฑ์จากงานวิจัยด้านดังกล่าว ที่สามารถใช้ข้อมูลภายนอก เช่นข้อความที่พบจากการมองด้วย Google Glass ร่วมกับการโปรแกรมต่างๆเพื่อสร้างความเป็นจริงเสมือนที่อ้างอิงจากข้อความที่พบจากการมองเห็นได้ เป็นต้น

เทคโนโลยีด้านการช่วยเหลือคนพิการ ก็สามารถใช้ประโยชน์จากการอ่านข้อความในภาพได้เช่นกัน ซึ่งในปัจจุบันมีผู้พิการทางด้านการมองเห็นเป็นจำนวนมาก ยกตัวอย่างเช่น ผู้มีสายตาเลือนราง ที่สามารถมองเห็นได้บ้างแต่ไม่ค่อยชัดเจน ทำให้การใช้ชีวิตประจำวันที่ต้องอาศัยการอ่านข้อความจากสิ่งต่างๆนั้นทำได้ค่อนข้างลำบาก การสร้างเครื่องช่วยอ่านที่สามารถอ่านข้อความที่พบในชีวิตประจำวันได้ จะช่วยพัฒนาคุณภาพชีวิตของผู้พิการกลุ่มนี้ให้ดีขึ้น

แต่ในปัจจุบัน งานวิจัยด้านการรู้จำตัวอักษรภาษาไทยนั้น มุ่งเน้นในด้านการรู้จำตัวอักษรจากเอกสารที่ได้จากเครื่องสแกนภาพและการรู้จำตัวอักษรภาษาไทยที่ได้จากการเขียน ซึ่งภาพเอกสารที่ได้จากการสแกนภาพนั้นมีความแตกต่างกับภาพของตัวอักษรและข้อความที่ปรากฏอยู่ในภาพถ่ายฉากธรรมชาติ ทั้งในแง่สภาพของแสง ตำแหน่งของข้อความ รูปแบบของตัวอักษร การมัวของตัวอักษร ฉากหลังที่มีความซับซ้อนและมีการบดบังเกิดขึ้น ซึ่งได้แสดงตัวอย่างของภาพเหล่านี้เทียบกับภาพที่ได้จากการสแกนเอกสารในรูปที่ 1-2 อีกทั้งโครงสร้างของภาษาไทยที่ประกอบด้วยสระและวรรณยุกต์ที่มีระดับแตกต่างจากภาษาอื่นๆ ปัญหาเหล่านี้เป็นอุปสรรคที่ทำให้วิธีการรู้จำตัวอักษรที่ใช้กับเอกสารที่ได้จากเครื่องสแกนภาพนั้น ทำงานได้ผลไม่ดีนักกับอักษรที่ได้จากภาพถ่ายฉากธรรมชาติ

รูปที่ 1-2 เปรียบเทียบระหว่างรูปข้อความที่ได้จากการสแกนเอกสารกับ
รูปข้อความในภาพถ่ายในฉากธรรมชาติ

จากปัญหาเหล่านี้ เมื่อทำการสืบค้นงานวิจัยที่เกี่ยวข้องกับการรู้จำข้อความบนภาพถ่ายฉากธรรมชาติภาษาอังกฤษและเมื่อทำการเปรียบเทียบระหว่างภาพเอกสารที่ได้จากการสแกนกับภาพข้อความในภาพถ่ายฉากธรรมชาติแล้วจะเห็นได้ว่า ข้อความในภาพถ่ายฉากธรรมชาตินั้นจะอยู่ในฉาก (scene) ที่มีความซับซ้อนสูงกว่ามาก ทั้งในแง่องค์ประกอบและการรบกวนต่างๆ และการระบุตำแหน่งข้อความในภาพสำหรับภาษาไทยนั้น มีความยากมากกว่าภาษาอังกฤษเนื่องจากโครงสร้างของภาษาไทยที่ประกอบด้วยสระและวรรณยุกต์หลายระดับดังแสดงรูปที่ 1-3 ทำให้วิธีการระบุตำแหน่งข้อความบนภาพถ่ายทั่วไปสำหรับภาษาอังกฤษนั้นทำงานได้ไม่ดีนักกับภาษาไทย เนื่องจากอาจจะสามารถระบุตำแหน่งของตัวอักษรได้ดี แต่การระบุตำแหน่งของสระและวรรณยุกต์ที่ไม่ได้อยู่ในระดับเดียวกับตัวอักษรอาจจะทำได้ไม่ดีนัก



รูปที่ 1-3 เปรียบเทียบความแตกต่างของระดับอักษร สระและวรรณยุกต์
ระหว่างภาษาอังกฤษซึ่งทุกตัวอักษรอยู่บนบรรทัดเดียวกัน
กับภาษาไทยที่สระและวรรณยุกต์นั้นมีการเรียงตัวอยู่ในหลายระดับ

1.2 วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาระบบที่สามารถระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติ ที่เห็นข้อความได้ชัดเจนและมีความคมชัด โดยระบบที่สร้างจะสามารถทำงานได้โดยอัตโนมัติ

1.3 ขอบเขตของงานวิจัย

- ภาพนำเข้าต้องมีขนาดอย่างน้อย 480x320 จุดภาพ
- ข้อความที่ปรากฏในภาพมีการจัดเรียงอยู่ในแนวนอน
- ข้อความที่ปรากฏในภาพมีความเปรียบต่างระหว่างฉากหลังและตัวอักษรชัดเจน
- ข้อความที่ปรากฏในภาพไม่มีการเบลอที่เกิดจากการเคลื่อนไหว
- ข้อความที่ปรากฏในภาพต้องมีขนาดอย่างน้อย 32 จุดภาพ
- การประเมินประสิทธิภาพในการระบุตำแหน่งข้อความจะใช้วิธีการประเมินที่ใช้ในการแข่ง ICDAR 2003 Robust Reading Competition และการคำนวณพื้นที่ซ้อนทับ

1.4 วิธีการดำเนินงานวิจัย

ในงานวิจัยนี้ ใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค (Convolutional Neural Network) ซึ่งเป็นเทคนิคการเรียนรู้ของเครื่องจักรที่สามารถเรียนรู้ตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน ดังนั้นขั้นตอนการดำเนินงานจะเริ่มจากการสร้างชุดข้อมูลสอน โดยผู้วิจัยได้ทำการรวบรวมชุดข้อมูลภาพที่เป็นชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษ สำหรับการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ และสังเคราะห์ชุดข้อมูลสอนตัวอักษรภาษาไทย โดยทำการปรับแต่งให้มีความใกล้เคียงกับตัวอักษรที่พบในภาพถ่ายฉากธรรมชาติ แล้วนำชุดข้อมูลสอนที่ได้ มาสร้างตัวจำแนกประเภทที่สามารถเรียนรู้ตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน ระหว่างส่วนภาพที่เป็นข้อความและไม่ใช่ข้อความ ที่จะนำไปใช้เพื่อสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ ซึ่งผลที่ได้จะถูกนำมาวิเคราะห์ร่วมกับ คำตอบของชุดข้อมูลทดสอบ (Ground Truth) เพื่อสร้างขั้นตอนวิธีในการระบุตำแหน่งข้อความจากแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

ในการทดสอบระบบ ผู้วิจัยใช้ชุดข้อมูลทดสอบมาตรฐาน ICDAR 2003, ICDAR2010, ชุดข้อมูลทดสอบจากการแข่งขัน BEST 2015 : Text Location Detection Contest และชุดข้อมูลทดสอบที่จัดเตรียมโดยผู้วิจัย โดยใช้หลักเกณฑ์ในการประเมินประสิทธิภาพในการระบุตำแหน่งข้อความตามวิธีที่การประเมินที่ใช้ใน ICDAR 2003 Robust Reading Competition หลักเกณฑ์การ

ประเมินผลที่ใช้ในการแข่งขัน BEST 2015 : Text Location Detection Contest และการคำนวณพื้นที่ซ้อนทับ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- ได้ระบบที่สามารถวิเคราะห์หาตำแหน่งข้อความภาษาไทยในภาพถ่ายฉากธรรมชาติได้
- เป็นต้นแบบในการสร้างและพัฒนาขั้นตอนวิธีในการระบุตำแหน่งข้อความภาษาไทยบนภาพถ่ายให้มีความแม่นยำมากขึ้น
- เพื่อช่วยในการเพิ่มความแม่นยำในการรู้จำตัวอักษรบนภาพถ่ายที่ต้องอาศัยขั้นตอนการระบุตำแหน่งข้อความจากภาพ
- สามารถนำระบบที่สร้าง ไปทำการต่อยอดสำหรับงานวิจัยด้านอื่นๆได้ เช่นการรู้จำตัวอักษรจากภาพถ่ายในฉากธรรมชาติ ซึ่งสามารถนำไปใช้ในงานการแปลภาษาอัตโนมัติจากภาพถ่ายฉากธรรมชาติ การอ่านข้อความในภาพสำหรับหุ่นยนต์หรือระบบความเป็นจริงเสมือน และการช่วยเหลือผู้พิการทางด้านสายตา

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

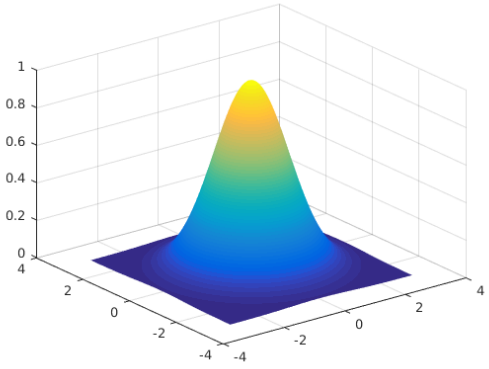
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 กระบวนการทำให้มัวแบบเกาส์เซียน (Gaussian Blur)

กระบวนการทำให้มัวแบบเกาส์เซียน (Gaussian Blur) เป็นกระบวนการทางการประมวลผลภาพอย่างหนึ่งที่ทำให้ภาพมีความมัว (blur) มากขึ้น สามารถที่จะช่วยลดสัญญาณรบกวน (noise) ที่ปรากฏอยู่ภายในภาพได้ กระบวนการนี้แตกต่างจากการทำให้ภาพมีความมัวแบบปกติ เนื่องจากใช้การแจกแจงแบบเกาส์เซียนในการสร้างเคอร์เนล ซึ่งเป็นไปตามสมการ

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.1}$$

เมื่อกำหนดให้ $G(x,y)$ เป็นเคอร์เนลที่ได้และ σ คือค่าเบี่ยงเบนมาตรฐาน จากสมการดังกล่าว จะสามารถสร้างกราฟการกระจายของการแจกแจงแบบเกาส์เซียนที่มีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 ได้ดังรูปที่ 2-1 และเคอร์เนลสำหรับกระบวนการทำให้มัวแบบเกาส์เซียนขนาด 5x5 จุดภาพ ดังรูปที่ 2-2



รูปที่ 2-1 กราฟการกระจายของการแจกแจงแบบเกาส์เซียน
ที่มีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1

$$\frac{1}{273} \cdot \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

รูปที่ 2-2 ตัวอย่างเคอร์เนลสำหรับกระบวนการทำให้มัวแบบเกาส์เซียน ขนาด 5x5 จุดภาพที่มีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1

ในวิทยานิพนธ์นี้จะใช้กระบวนการทำให้มัวแบบเกาส์เซียน ในการสร้างชุดข้อมูลสอนตัวอักษรภาษาไทย จากชุดแบบอักษร จึงต้องมีการปรับภาพให้มีความใกล้เคียงกับตัวอักษรในภาพถ่ายฉากธรรมชาติ

2.1.2 อันชาร์ป มาส์กกิง (Unsharp Masking)

อันชาร์ป มาส์กกิง (Unsharp Masking) เป็นกระบวนการทางการประมวลผลภาพอย่างหนึ่งที่ทำให้ขอบ (edge) ของวัตถุในภาพมีความคม (sharpen) มากขึ้น โดยในขั้นตอนแรก จะทำการแยกส่วนที่เป็นองค์ประกอบความถี่สูง ซึ่งก็คือขอบของวัตถุในภาพ ดังสมการ

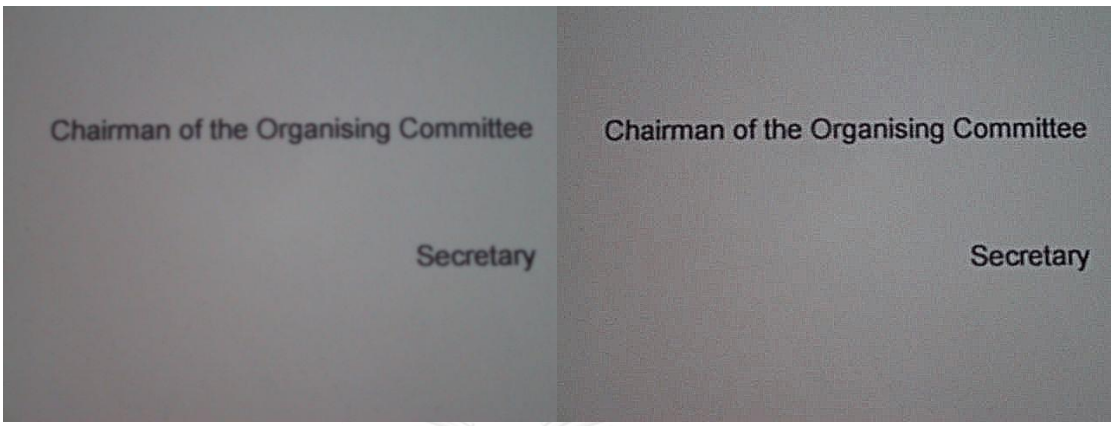
$$\bar{f}(x,y) = f(x,y) - \bar{b}(x,y) \tag{2.2}$$

เมื่อกำหนดให้ $f(x,y)$ เป็นภาพนำเข้า $\bar{b}(x,y)$ เป็นภาพที่ผ่านกระบวนการทำให้ภาพมีความมัว (blur) และ $\bar{f}(x,y)$ คือองค์ประกอบความถี่สูงของภาพนำเข้า จากนั้นจะทำกระบวนการ อันชาร์ป มาส์กกิง ดังสมการ

$$g(x,y) = f(x,y) + k * \bar{f}(x,y) \tag{2.3}$$

เมื่อกำหนดให้ $f(x,y)$ เป็นภาพนำเข้า $\bar{f}(x,y)$ คือองค์ประกอบความถี่สูงของภาพนำเข้า และ k คือค่าพารามิเตอร์ที่ค่าความคม ซึ่งโดยส่วนใหญ่แล้ว k จะอยู่ในช่วง 0.3 – 0.7 ซึ่งในวิทยานิพนธ์นี้นำขั้นตอนดังกล่าวมาใช้ในกระบวนการประมวลผลก่อน เพื่อทำการปรับปรุงคุณภาพ

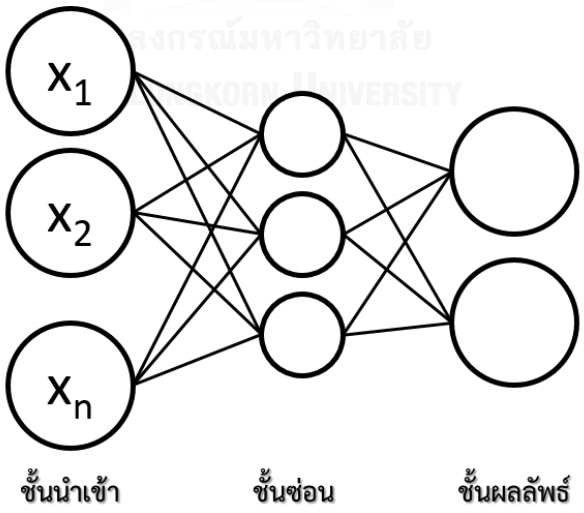
ของภาพนำเข้าให้ตัวอักษรที่ปรากฏภายในภาพมีความคมมากขึ้น ดังแสดงตัวอย่างของภาพที่ผ่านกระบวนการอันชาร์ป มาส์กกิง ในรูปที่ 2-3



รูปที่ 2-3 ภาพเปรียบเทียบระหว่างภาพนำเข้าและภาพที่ผ่านกระบวนการ Unsharp Masking

2.1.3 คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค (Convolutional Neural Network)

นิวรอลเน็ตเวิร์ค (Neural Network) จะประกอบด้วย 3 ส่วนได้แก่ ชั้นนำเข้า (Input Layer) ซึ่งข้อมูลนำเข้าเป็นแถวลำดับ 1 มิติ ของคุณลักษณะสำคัญ (feature) ที่ได้จากตัวสกัดคุณลักษณะสำคัญที่เลือกใช้ ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) เชื่อมต่อกัน โดยมีโครงสร้างดังรูปที่ 2-4



รูปที่ 2-4 โครงสร้างของนิวรอลเน็ตเวิร์ค โดยทั่วไป ประกอบด้วยชั้นนำเข้า ชั้นซ่อน และชั้นผลลัพธ์

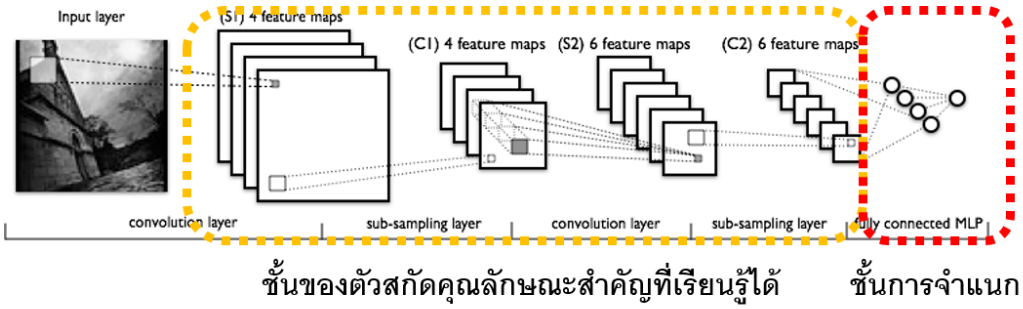
โดยทั่วไปการใช้นิวรอลเน็ตเวิร์คเพื่อเป็นตัวจำแนกประเภทนั้น จะใช้กับข้อมูลนำเข้าที่เป็นข้อมูลแถวลำดับ 1 มิติ เช่น การใช้เพื่อการพยากรณ์ราคาทองคำ ก็อาจจะใช้คุณลักษณะสำคัญเป็นราคาทองคำย้อนหลัง 30 วันในรูปแบบของแถวลำดับ 1 มิติที่มีจำนวนส่วนย่อย (element) เท่ากับ 30 ตัว หรือการพยากรณ์ปริมาณน้ำฝนก็อาจจะใช้คุณลักษณะสำคัญเป็นปริมาณน้ำฝนและความชื้นสัมพัทธ์ในช่วง 15 วันที่ผ่านมา ก็จะใช้คุณลักษณะสำคัญเป็นแถวลำดับ 1 มิติที่มีจำนวนส่วนย่อยเท่ากับ 30 ตัวที่ประกอบด้วยปริมาณน้ำฝนและความชื้นสัมพัทธ์ จากตัวอย่างดังกล่าวจะเห็นได้ว่าคุณลักษณะสำคัญที่ใช้นั้นจะเป็นข้อมูลแถวลำดับ 1 มิติทั้งสิ้น

ในการนำนิวรอลเน็ตเวิร์ค มาใช้ในการจำแนกประเภทข้อมูล 2 มิติ ซึ่งในวิทยานิพนธ์นี้คือภาพนั้น อาจจะใช้คุณลักษณะสำคัญเป็นจุดภาพ (pixel) ของภาพนำเข้า หรืออาศัยตัวสกัดคุณลักษณะสำคัญ (Feature Extractor) ในการสกัดคุณลักษณะสำคัญจากภาพที่เป็นข้อมูล 2 มิติให้อยู่ในรูปของข้อมูลแถวลำดับ 1 มิติ ซึ่งในงานทางด้านคอมพิวเตอร์วิชัน (Computer Vision) นั้นมีตัวสกัดคุณลักษณะสำคัญที่เป็นที่นิยม ตัวอย่างเช่น

- Harr-Like Features
- Histogram of Oriented Gradients (HOG)
- Speeded Up Robust Features (SURF)
- Scale-Invariant Feature Transform (SIFT)
- Maximally Stable Extremal Regions (MSERs)

ตัวสกัดคุณลักษณะสำคัญดังกล่าว ได้ถูกออกแบบมาสำหรับใช้ในงานที่แตกต่างกัน ตัวอย่างเช่น Histogram of Oriented Gradients (HOG) ได้ถูกออกแบบมาสำหรับงานการตรวจจับมนุษย์ในภาพ หรือ Harr-Like Features ซึ่งถูกออกแบบมาเพื่อใช้ในงานทางด้านการตรวจจับใบหน้า

คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค มีความแตกต่างจากนิวรอลเน็ตเวิร์คคือ คอนโวลูชันนอลนิวรอลเน็ตเวิร์คนั้น มีข้อมูลนำเข้าเป็นภาพ ในรูปแบบของแถวลำดับ และมีโครงสร้างซึ่งประกอบด้วย 2 ส่วนหลักได้แก่ ชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ (Trainable Feature Extractor Layer) และ ชั้นของการจำแนก (Classification Layer) ดังแสดงในรูปที่ 2-5



รูปที่ 2-5 โครงสร้างของคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค[1]
ซึ่งประกอบด้วยชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ และชั้นการจำแนก

ชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ (Trainable Feature Extractor Layer) ประกอบไปด้วย 3 ชั้นหลักคือ ชั้นคอนโวลูชัน (Convolution), ชั้นของฟังก์ชันไม่เชิงเส้น (Nonlinear Function) และชั้นการสุ่มตัวอย่าง (Subsampling) และยังมีชั้นอื่นๆเช่น ชั้นดรอปเอ้าท์ (Dropout)

คอนโวลูชันนอล นิวรอลเน็ตเวิร์คจะมีการจัดเรียงภายในชั้นของตัวสกัดคุณลักษณะสำคัญได้ในหลายรูปแบบ โดยปกติจะมีการจัดเรียงชุดตัวสกัดคุณลักษณะสำคัญ ประกอบด้วยชั้นคอนโวลูชัน ชั้นของฟังก์ชันไม่เชิงเส้นและชั้นการสุ่มตัวอย่าง ดังแสดงในรูปที่ 2-6 โดยแต่ละชุดตัวสกัดคุณลักษณะสำคัญ สามารถเรียงต่อกันได้ขึ้นอยู่กับการเลือกใช้ของผู้ใช้งาน

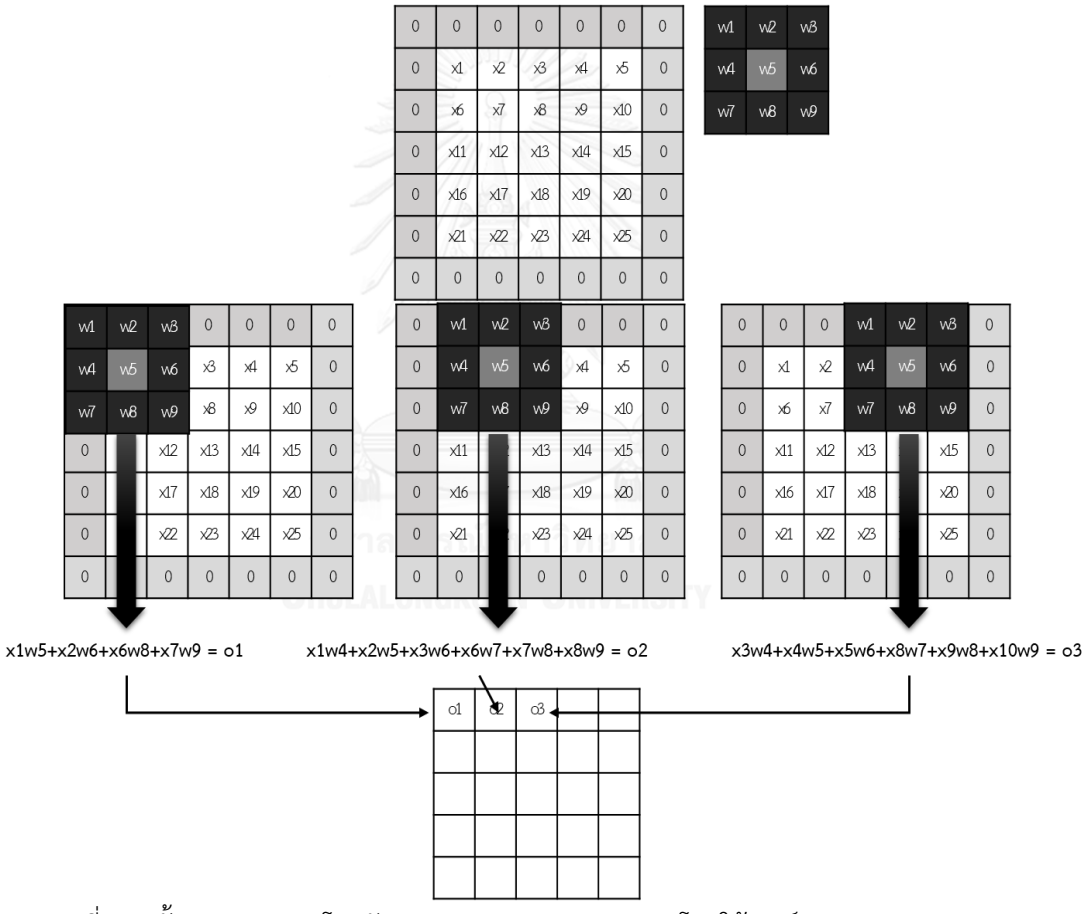
(ก)

(ข)

รูปที่ 2-6 (ก) ตัวอย่างการจัดเรียงภายในชุดตัวสกัดคุณลักษณะสำคัญ
(ข) ตัวอย่างการจัดเรียงภายในชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้
ที่มีชุดตัวสกัดคุณลักษณะสำคัญจำนวน n ชุด

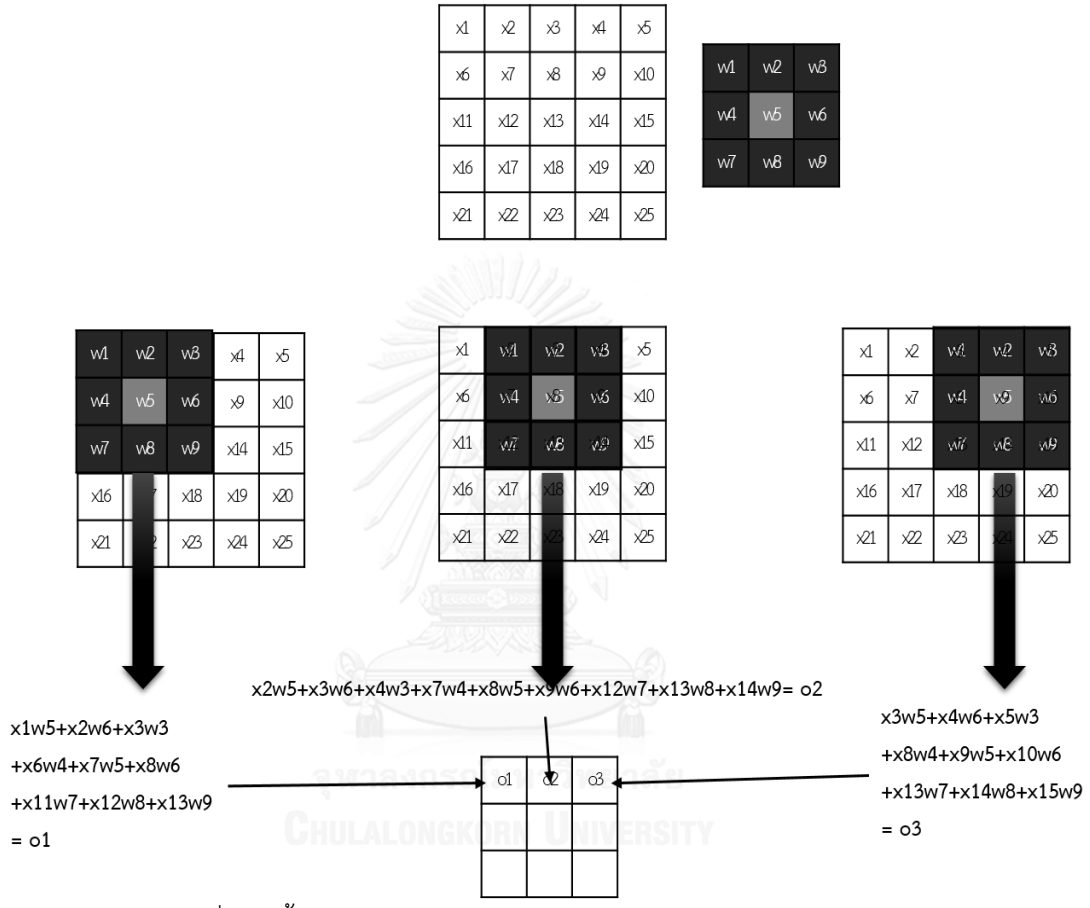
[1] ที่มา : http://deeplearning.net/tutorial/lenet.html

ในชั้นคอนโวลูชัน จะทำการคอนโวลูชันแบบ valid ซึ่งมีความแตกต่างจากการคอนโวลูชันบนภาพ ที่โดยปกติแล้ว ขนาดของผลลัพธ์ที่ได้จะมีขนาดเท่ากับขนาดของภาพนำเข้า โดยมีการเพิ่มค่าจุดภาพ (padding) ที่บริเวณขอบภาพ ดังแสดงในรูปที่ 2-7 แต่สำหรับการคอนโวลูชันแบบ valid ผลลัพธ์ที่ได้จะมีขนาดเล็กกว่าภาพนำเข้าเนื่องจากไม่มีการเพิ่มค่าจุดภาพ ตัวอย่างเช่นภาพขนาด $n \times n$ จุดภาพทำการคอนโวลูชันแบบ valid กับเคอร์เนลขนาด $m \times m$ จุดภาพ จะได้ผลลัพธ์ขนาด $(n-m+1) \times (n-m+1)$ จุดภาพ โดยได้แสดงขั้นตอนวิธีการคอนโวลูชันแบบ valid ในรูปที่ 2-8



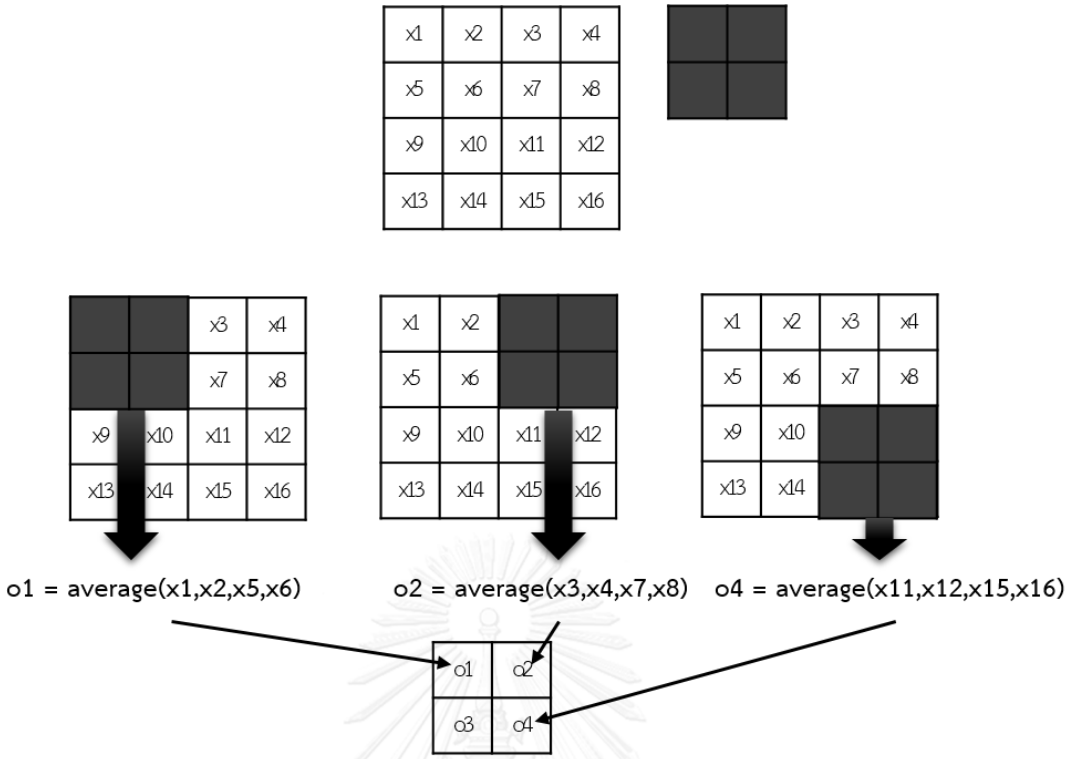
รูปที่ 2-7 ขั้นตอนการคอนโวลูชันบนภาพขนาด 5x5 จุดภาพ โดยใช้เคอร์เนลขนาด 3x3 จุดภาพ และให้ผลลัพธ์ขนาดเท่ากับภาพนำเข้า

เคอร์เนลที่ใช้ในของชั้นคอนโวลูชัน จะถูกสุ่มในตอนตั้งต้นและจะถูกปรับให้มีความเหมาะสมกับชุดข้อมูลสอนโดยขั้นตอนวิธีแบคพรอพพาเกชั่น (backpropagation) และผลที่ได้จากชั้นคอนโวลูชันนั้น จะเรียกว่า feature maps ซึ่งจะมีจำนวนเท่ากับเคอร์เนล



รูปที่ 2-8 ขั้นตอนการคอนโวลูชันแบบ valid บนภาพขนาด 5x5 จุดภาพ
โดยใช้เคอร์เนลขนาด 3x3 จุดภาพ

การทำงานของชั้นการสุ่มตัวอย่างนั้น เพื่อลดจำนวนตัวแปรและลดความซ้ำซ้อนของข้อมูลเนื่องจากผลที่ได้จากชั้นคอนโวลูชัน ซึ่งสำหรับคอนโวลูชัน นิวรอลเน็ตเวิร์คนั้น จะมีวิธีการสุ่มตัวอย่างที่เป็นที่นิยมใช้กัน 2 แบบคือ การหาค่าสูงสุด และการหาค่าเฉลี่ย โดยชั้นของการสุ่มตัวอย่างนั้น จะมีวิธีการเลื่อน window แตกต่างจากชั้นคอนโวลูชันคือ โดยส่วนใหญ่จะไม่มีการทับกันของ window เกิดขึ้น วิธีการทำงานของชั้นการสุ่มตัวอย่างนั้นแสดงดังรูปที่ 2-9



รูปที่ 2-9 ขั้นตอนการสุ่มตัวอย่างบนภาพขนาด 4x4 จุดภาพ
โดยใช้เคอร์เนลการสุ่มตัวอย่างแบบหาค่าเฉลี่ยขนาด 2x2 จุดภาพ

โดยส่วนใหญ่แล้วสำหรับคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค ชั้นคอนโวลูชันและชั้นการสุ่มตัวอย่าง มักจะต่อด้วยชั้นของฟังก์ชันไม่เชิงเส้น ทั้งนี้มีสาเหตุเนื่องจากจากข้อมูลที่พบได้ในธรรมชาตินั้นโดยส่วนใหญ่จะอยู่ในรูปแบบของข้อมูลที่ไม่เป็นเชิงเส้น ซึ่งในวิทยานิพนธ์นี้ ได้เลือกใช้ฟังก์ชันไม่เชิงเส้นเรคติไฟเออร์ (Rectifier Linear) ซึ่งเป็นฟังก์ชันที่ถูกนิยามตามสมการ 2.4 เมื่อกำหนดให้ $\boldsymbol{x}$ เป็นเวคเตอร์ของข้อมูลนำเข้า

$$f(x) = \max(0, x) \tag{2.4}$$

ในส่วนต่อไปนั้นจะแสดงวิธีการทำงานของคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค เมื่อนำไปใช้งานในการจำแนกประเภทของภาพ โดยยกตัวอย่างโครงสร้างของชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ดังตารางที่ 2-1

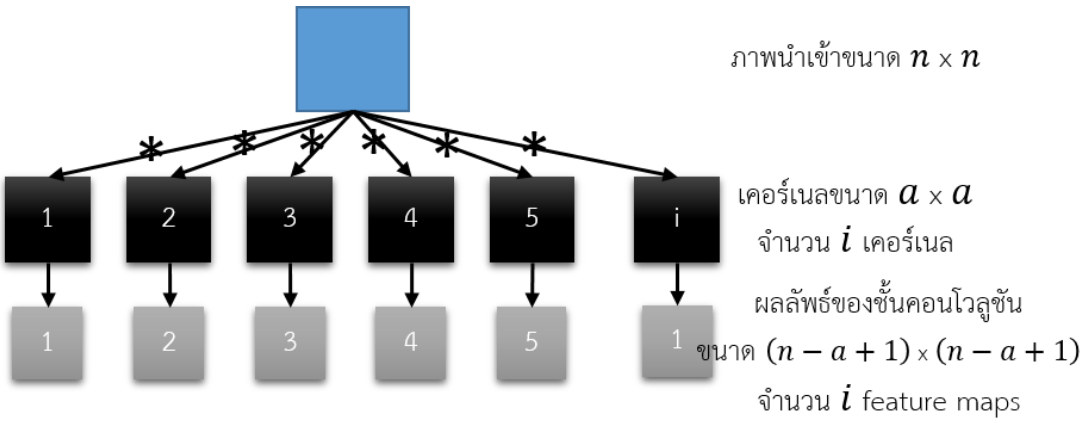
ตารางที่ 2-1 ตัวอย่างโครงสร้างของชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้

ชั้นที่	ชนิดของชั้น	ขนาดข้อมูลนำเข้า	ขนาดของเคอร์เนล	จำนวนเคอร์เนล
0	ชั้นนำเข้า	$n \times n$	-	-
1	ชั้นคอนโวลูชัน	$n \times n$	$a \times a$	$1 \times j$
2	ชั้นฟังก์ชันไม่เชิงเส้น	$(n-a+1) \times (n-a+1) \times j$	-	-
3	ชั้นการสุ่มตัวอย่าง	$(n-a+1) \times (n-a+1) \times j$	$p \times p$	1
4	ชั้นคอนโวลูชัน	$\left(\frac{n-a+1}{p}\right) \times \left(\frac{n-a+1}{p}\right) \times j$	$b \times b$	$j \times k$
5	ชั้นฟังก์ชันไม่เชิงเส้น	$\left(\left(\frac{n-a+1}{p}\right)-b+1\right) \times \left(\left(\frac{n-a+1}{p}\right)-b+1\right) \times k$	-	-
6	ชั้นการสุ่มตัวอย่าง	$\left(\left(\frac{n-a+1}{p}\right)-b+1\right) \times \left(\left(\frac{n-a+1}{p}\right)-b+1\right) \times k$	$q \times q$	1
ชั้นการจำแนก		$\frac{\left(\left(\frac{n-a+1}{p}\right)-b+1\right)}{q} \times \frac{\left(\left(\frac{n-a+1}{p}\right)-b+1\right)}{q} \times k$	-	-

ในชั้นที่ 1 ซึ่งเป็นชั้นคอนโวลูชัน มีข้อมูลนำเข้าเป็นภาพขนาด $n \times n$ จุดภาพนั้น จะผ่านกระบวนการคอนโวลูชันกับเคอร์เนลในชั้นดังกล่าว จำนวน j เคอร์เนลดังสมการ

$$y_i = x * \emptyset_i \; ; \; \forall i \in \{1,2,3,\dots,j\} \tag{2.5}$$

เมื่อกำหนดให้ y_i เป็น feature map ที่ได้ x และ $\emptyset_i$ คือภาพนำเข้าและเคอร์เนลของชั้นของคอนโวลูชัน ชั้นที่ 1 ซึ่งมีจำนวน j เคอร์เนลตามลำดับ

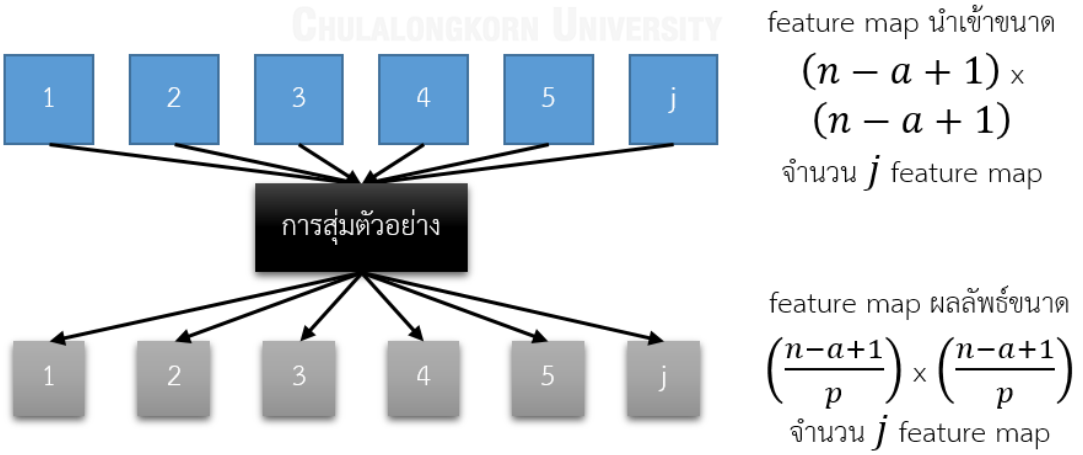


รูปที่ 2-10 กระบวนการทำงานของชั้นคอนโวลูชัน (ชั้นที่ 1)

ในชั้นถัดมาจะเป็นชั้นของฟังก์ชันไม่เชิงเส้น (ชั้นที่ 2) จะใช้ฟังก์ชันไม่เชิงเส้นที่เลือกกับทุกตำแหน่งบน feature map ในวิทยานิพนธ์นี้ได้เลือกใช้ฟังก์ชันไม่เชิงเส้นเรคติไฟเออร์ และในขั้นต่อมาจะเข้าสู่ชั้นการสุ่มตัวอย่าง (ชั้นที่ 3) เพื่อลดความซ้ำซ้อนของข้อมูล โดยเป็นไปดังสมการ

$$y_i = x_i \cdot \emptyset \; ; \; \forall i \; \in \{1,2,3, \dots, j\} \tag{2.6}$$

เมื่อกำหนดให้ y_i เป็น feature map ที่ได้จากขั้นตอนการสุ่มตัวอย่าง x_i เป็น feature map ที่ได้จากชั้นของฟังก์ชันไม่เชิงเส้นเส้น (ชั้นที่ 2) $\emptyset$ เป็นเคอร์เนลของการสุ่มตัวอย่าง และนิยามให้ $\cdot$ เป็นตัวดำเนินการสุ่มตัวอย่าง

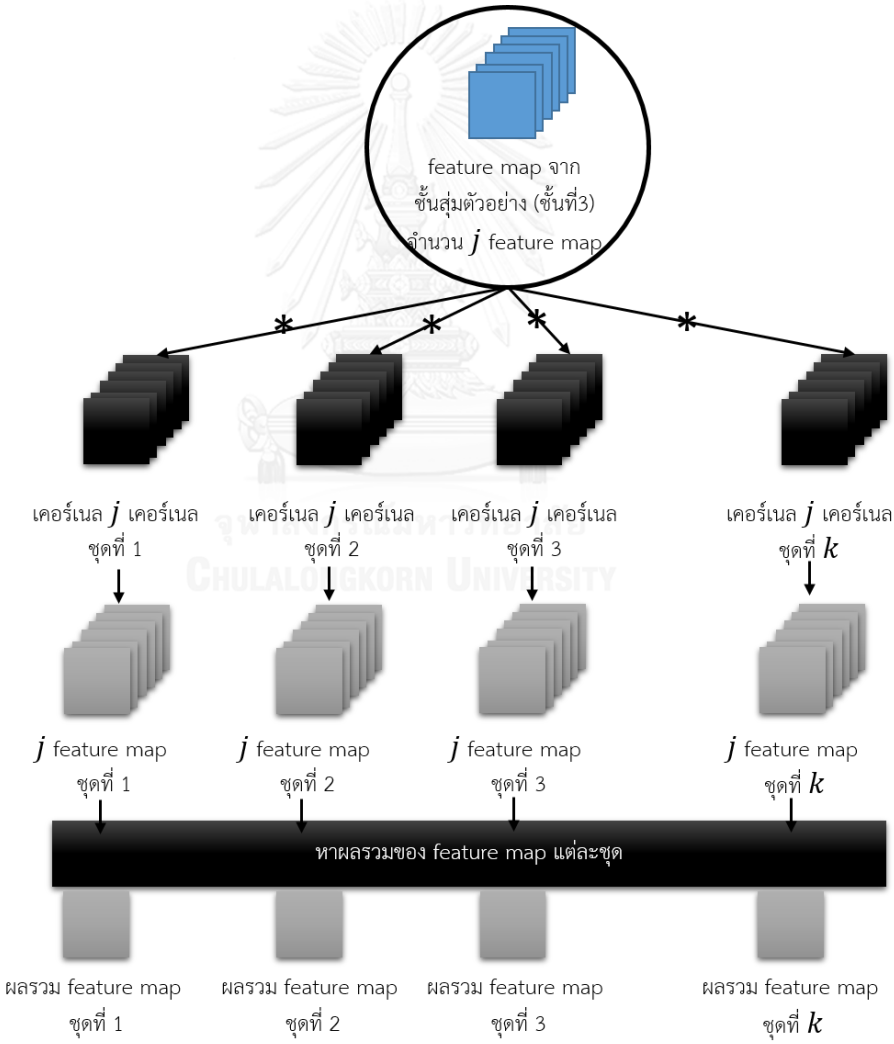


รูปที่ 2-11 กระบวนการทำงานของชั้นการสุ่มตัวอย่าง (ชั้นที่ 3)

สำหรับชั้นที่ 4 ซึ่งเป็นชั้นคอนโวลูชันนั้น จะนำ feature map ที่ได้จากชั้นที่ 3 ซึ่งขนาด $\left(\frac{n-a+1}{p}\right)$ x $\left(\frac{n-a+1}{p}\right)$ จุดภาพจำนวน j feature map ที่ได้จากชั้นที่ 3 นั้นมาเข้าสู่กระบวนการของชั้นคอนโวลูชันชั้นที่ 4 กับเคอร์เนล ขนาด b x b จำนวน j x k เคอร์เนล ดังสมการ

$$y_i = \sum_{\delta=1}^{j}(x_\delta * \emptyset_{\delta i}) \; ; \; \forall i \; \in \{1,2,3,\ldots,k\} \tag{2.6}$$

เมื่อกำหนดให้ y_i เป็นผลลัพธ์ของ feature map ที่ได้จากชั้นคอนโวลูชัน (ชั้นที่ 4) x_i เป็น feature map นำเข้าที่ได้จากชั้นการสุ่มตัวอย่าง (ชั้นที่ 3) $\emptyset$ เป็นเคอร์เนลของชั้นคอนโวลูชัน (ชั้นที่ 4) โดยสามารถอธิบายได้ด้วยรูปที่ 2-12



รูปที่ 2-12 กระบวนการทำงานของชั้นคอนโวลูชัน (ชั้นที่ 4)

ในชั้นถัดมาจะเป็นชั้นของฟังก์ชันไม่เชิงเส้น (ชั้นที่ 5) และชั้นการสุ่มตัวอย่าง (ชั้นที่ 6) ซึ่งจะใช้กระบวนการในลักษณะเดียวกับชั้นที่ 3 และ 4 หลังจากนั้น ผลลัพธ์ที่ได้จะถูกแปลงเป็นแถวลำดับแบบ 1 มิติเพื่อส่งเข้าสู่ชั้นการจำแนกต่อไป

สำหรับคอนโวลูชันนอล นิวรอลเน็ตเวิร์คนั้น ในบางกรณีจะมีชั้นที่แตกต่างจากชั้นหลักที่ใช้ ซึ่งในวิทยานิพนธ์นี้ ได้ใช้ชั้นดรอปเอ้าท์ (Dropout) เป็นชั้นที่จะทำการสุ่มเพื่อทิ้งค่าคุณลักษณะสำคัญบางตัว โดยมีการกำหนด ค่าพารามิเตอร์ความน่าจะเป็นในการเลือกที่จะทิ้งค่าคุณลักษณะ (Dropout Rate) จากงานวิจัยของ Hinton และคณะ [1] ได้ทำการทดสอบแล้วพบว่า ชั้นดรอปเอ้าท์สามารถช่วยลดปรากฏการณ์โอเวอร์ฟิต (overfit) ซึ่งเป็นปรากฏการณ์ที่ นิวรอลเน็ตเวิร์คให้ผลลัพธ์ที่แม่นยำมากบนชุดข้อมูลสอน แต่ให้ผลลัพธ์ที่ไม่ดีนักบนชุดข้อมูลทดสอบ

เมื่อผ่านชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้แล้ว คุณลักษณะสำคัญทั้งหมดที่ได้จะถูกนำไปเป็นข้อมูลนำเข้าของชั้นการจำแนก ซึ่งสามารถเลือกได้ว่าจะใช้ขั้นตอนวิธีใดในการจำแนกกลุ่มของคุณลักษณะสำคัญที่กัดได้ โดยอาจจะเป็นนิวรอลเน็ตเวิร์คทั่วไปที่ประกอบด้วย ชั้นนำเข้า ชั้นฮิดเด่น และชั้นนำออก เพื่อให้ผลลัพธ์การจำแนกของคอนโวลูชันนอล นิวรอลเน็ตเวิร์คในวิทยานิพนธ์นี้ได้ใช้ชั้นของการจำแนกเป็น ชั้นซอฟต์แมกซ์ (Softmax Layer) โดยไปเป็นตามสมการ

$$f(x) = \frac{e^x}{\sum e^x} \tag{2.7}$$

เมื่อ $\boldsymbol{x}$ เป็นเวคเตอร์ของคุณลักษณะสำคัญ ที่ได้จากชั้นของตัวสกัดคุณลักษณะสำคัญ โดยจะให้ผลลัพธ์ที่อยู่ในช่วง [0,1] ซึ่งเป็นค่าความน่าจะเป็นของแต่ละคุณลักษณะสำคัญที่สกัดได้

2.2 งานวิจัยที่เกี่ยวข้อง

การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาตินั้น มีความแตกต่างจากการระบุตำแหน่งข้อความในภาพถ่ายเอกสารที่ได้จากการเครื่องสแกนภาพ เนื่องจากภาพถ่ายฉากธรรมชาตินั้น มีความซับซ้อนและหลากหลาย ในแง่ขององค์ประกอบต่างๆ เช่น สภาพของแสง เงา การถูกบดบัง และองค์ประกอบอื่นๆในภาพ ต่างจากภาพถ่ายเอกสารที่โดยส่วนใหญ่แล้วมีความเป็นระเบียบ ข้อความถูกพิมพ์หรือเขียนบนเส้นบรรทัด มีความคมชัด และสภาพแสงค่อนข้างคงที่ จากการสำรวจวรรณกรรมของ Ye และ Doermann [2] ได้แบ่งวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติออกเป็น 2 วิธีหลักๆ คือ วิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน (Connected Component Based Approach) และวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาส่วนภาพ (Region / Sliding Windows Based Approach)

2.2.1 การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน (Connected Component Based Approach)

การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน อาศัยคุณสมบัติของตัวอักษรโดยมองตัวอักษรเป็นองค์ประกอบที่เชื่อมต่อกัน (connected component) และใช้คุณสมบัติท้องถิ่น ตัวอย่างเช่น สี ความกว้างของลายเส้น (stroke-width) ความหนาแน่นของขอบ (edge-intensity) และคุณสมบัติอื่นๆ ซึ่งคุณสมบัติเหล่านี้ จะถูกนำมาวิเคราะห์เพื่อระบุตำแหน่งของข้อความในภาพถ่ายฉากธรรมชาติ งานวิจัยที่มีความน่าสนใจและจัดอยู่ในประเภทนี้ได้แก่

Subramanian และคณะ [3] ได้เสนอวิธีการเพื่อระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ โดยสกัดลายเส้น (stroke) จากแต่ละแถวของภาพนำเข้า แล้วนำมาวิเคราะห์หาความสัมพันธ์ระหว่างลายเส้นของแต่ละแถวเพื่อสร้างบริเวณส่วนภาพที่มีข้อความจากภาพนำเข้า

Epshtein และคณะ [4] ได้เสนอวิธีการแปลงความกว้างลายเส้น (stroke width transform) ในการระบุตำแหน่งของข้อความในภาพถ่ายฉากธรรมชาติ ภาพนำเข้าจะถูกนำไปหาแผนภาพขอบ (edge map) เพื่อนำเข้าสู่การแปลงลายเส้น ซึ่งเป็นกระบวนการที่ตัดสินใจว่าจุดภาพ (pixel) จากแผนภาพขอบที่สนใจนั้นควรจะเป็นตัวอักษรหรือไม่เมื่อเทียบกับคุณสมบัติท้องถิ่น ได้แก่ ทิศทางและความห่างกับจุดภาพที่ใกล้เคียง หลังจากนั้นการวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน (connected component analysis) จะถูกนำมาใช้เพื่อรวมผลลัพธ์จากการแปลงความ

กว้างลายเส้นเป็นตัวอักษร ตัวอักษรเหล่านั้นจะถูกรวมเข้ากันเป็นข้อความ โดยอาศัยคุณสมบัติของข้อความ ที่มีความกว้างลายเส้น ความกว้างของตัวอักษรและช่องว่างระหว่างตัวอักษรที่มีลักษณะคล้ายคลึงกัน แล้วจึงรวมบริเวณส่วนภาพที่น่าจะเป็นข้อความที่มีการทับซ้อนกัน โดยอาศัยคุณสมบัติของข้อความที่จะมีทิศทางในลักษณะเดียวกันเข้าด้วยกัน

Karaoglu และคณะ [5] นำผลที่ได้จากการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันซึ่งจะได้องค์ประกอบที่เชื่อมต่อกันที่น่าจะเป็นข้อความ จากภาพที่ผ่านการประมวลผลก่อน มาทำการสกัดคุณลักษณะสำคัญ 3 ประเภทคือ คุณลักษณะสำคัญทางเรขาคณิต ความสม่ำเสมอของรูปร่าง (shape regularity) และคุณลักษณะสำคัญเพิ่มเติมที่ขึ้นกับมุม ซึ่งคุณสมบัติเหล่านี้จะถูกนำไปใช้กับการเรียนรู้ขั้นตอนวิธีต้นไม้แบบสุ่ม (random forest) โดยขั้นตอนวิธีดังกล่าวได้ทำการเรียนรู้จากชุดข้อมูลสอนมาตรฐาน ICDAR 2003 ที่ทำการรวบรวมโดย Lucas และคณะ [6] เมื่อได้บริเวณส่วนภาพที่น่าจะเป็นตัวอักษรจากขั้นตอนวิธีที่กล่าวมาแล้ว บริเวณส่วนภาพดังกล่าวจะถูกนำมารวมเข้าเป็นข้อความโดยอาศัยคุณสมบัติของระยะห่าง ความกว้าง ความยาว และมุมระหว่างจุดศูนย์กลาง เพื่อรวมตัวอักษรเหล่านั้นเป็นข้อความ

Huang และ Ma [7] ได้เสนอวิธีการตรวจจับและระบุตำแหน่งข้อความในภาพเคลื่อนไหว โดยสร้างแผนภาพลายเส้นจากภาพนำเข้าที่ผ่านตัวกรองแบบ Log-Gabor ซึ่งสามารถกรองภาพพื้นหลังบางส่วนได้ จากนั้นทำการวิเคราะห์แผนภาพลายเส้น โดยอาศัยสมมุติฐานของบริเวณส่วนภาพที่เป็นข้อความ และบรรทัดของข้อความจะมีลายเส้นมากกว่าบริเวณอื่น โดยทำการวิเคราะห์ทีละแถวของแผนภาพลายเส้น เพื่อระบุตำแหน่งบรรทัดของข้อความ แล้วจึงวิเคราะห์บรรทัดของข้อความที่ได้โดยการตรวจจับมุมโดยขั้นตอนวิธีของฮาร์ริส (Harris Corner Detection) ร่วมกับการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันเพื่อระบุตำแหน่งของข้อความในภาพ

Neumann และ Matas [8] [9] ได้เสนอวิธีการใช้ Maximally Stable Extremal Regions (MSERs) ในการระบุตำแหน่งข้อความในภาพถ่ายโดยทำการสกัด MSERs จากภาพนำเข้า แล้วนำบริเวณส่วนภาพที่ได้ มาทำการสกัดหาคุณลักษณะสำคัญที่ไม่ขึ้นกับสเกลของภาพเช่น อัตราส่วนของภาพ (aspect ratio) ความสูงของส่วนที่สนใจแบบสัมพัทธ์ (relative segment height) และความสอดคล้องของสีตัวอักษร แล้วนำคุณสมบัติที่ได้ไปจำแนกด้วยตัวจำแนกประเภทซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine : SVM) ที่ใช้เคอร์เนลฟังก์ชั่นแบบรัศมีฐานหลัก (radial basis function) และใช้การวิเคราะห์เพื่อสร้างสมมุติฐานบรรทัดของข้อความจากคุณลักษณะสำคัญได้แก่ ความกว้างและความสูงของตัวอักษร สีของตัวอักษร อัตราส่วนของตัวอักษร ความกว้างของลายเส้นและความสูงจากระยะบรรทัด ร่วมกับข้อมูลของแต่ละองค์ประกอบที่ได้จากการสกัด MSER เพื่อสร้างเส้นทางที่เป็นไปได้ของการรวมแต่ละตัวอักษรเข้าด้วยกันเป็นข้อความ

ได้มีงานวิจัยที่ใช้ MSERs ในการหาบริเวณที่น่าจะเป็นตัวอักษรเช่นเดียวกันโดย Yin และคณะ [10] นำภาพนำเข้ามาสกัด MSERs เพื่อหาบริเวณที่น่าจะเป็นตัวอักษร จากนั้นจะสร้างบริเวณส่วนภาพที่น่าจะเป็นข้อความโดยอาศัยการรวมบริเวณส่วนภาพที่น่าจะเป็นตัวอักษรเข้าด้วยกัน โดยอาศัยกฎเกณฑ์ทางเรขาคณิตร่วมกับคุณสมบัติของตัวอักษร คือตัวอักษรภายในข้อความเดียวกันน่าจะมีความคล้ายคลึงกันในแง่ของความกว้างของลายเส้น ความกว้างและความสูงของตัวอักษร ตำแหน่งและสีของตัวอักษร จากนั้นบริเวณส่วนภาพที่น่าจะเป็นข้อความที่ได้จะถูกนำไปจำแนกโดยตัวจำแนกประเภทอดาบูสต์ (Adaboost) เพื่อทำการจำแนกอีกครั้งหนึ่งว่าบริเวณที่น่าจะเป็นข้อความเหล่านั้นเป็นข้อความหรือไม่ เพื่อสร้างเซตของบริเวณส่วนภาพที่เป็นข้อความจากภาพนำเข้า

Neumann และ Matas [9] ได้ใช้การค้นหาลายเส้นเพื่อระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ ภาพนำเข้าจะถูกสกัดลายเส้นในทิศทางต่างๆ โดยการหาภาพฉายของเกรเดียนท์และนำไปผ่านตัวกรองภาพทำให้ได้บริเวณที่น่าจะเป็นลายเส้นของตัวอักษร จากนั้นจะทำการรวมบริเวณลายเส้นเหล่านั้นที่มีการทับซ้อนกันเข้าด้วยกันเป็นบริเวณที่มีข้อความปรากฏอยู่

Iqbal และคณะ [11] ได้นำเสนอวิธีการใช้ MSERs ร่วมกับเครือข่ายการให้คะแนนแบบเบย์เซียน (Bayesian Network Score) บริเวณที่น่าจะเป็นตัวอักษรจะถูกเลือกโดยคุณลักษณะสำคัญ MSERs ซึ่งจะถูกคัดกรองโดยกระบวนการทำภาพไบนารีแบบปรับตัวได้ (Adaptive Binarization) หลังจากนั้นจะถูกนำมาสกัดคุณลักษณะสำคัญทางตำแหน่ง เพื่อนำไปวิเคราะห์การรวมกลุ่มกันของตัวอักษรโดยใช้ตัวจำแนกประเภทเครือข่ายการให้คะแนนแบบเบย์เซียน และนำไปตัดสินโดยเครือข่ายการให้คะแนนแบบเบย์เซียนที่มีหน้าที่ตัดสินกลุ่มของข้อความว่าเป็นข้อความจริงหรือไม่

งานวิจัยการระบุตำแหน่งข้อความในภาพถ่ายโดยพิจารณาองค์ประกอบที่เชื่อมต่อกันตามที่ได้สำรวจและสรุปข้างต้นนั้น พบว่าเป็นวิธีเพื่อระบุตำแหน่งข้อความภาษาอังกฤษจากภาพเป็นหลัก ซึ่งเมื่อพิจารณาจากคุณลักษณะสำคัญต่างๆที่ใช้จำแนกและคัดกรองความเป็นตัวอักษรนั้นจะเห็นได้ว่าคุณลักษณะเหล่านั้นอาจจะทำงานได้ดีกับภาษาอังกฤษแต่ไม่ดีกับภาษาไทย เนื่องจากความต่างของระดับสระและวรรณยุกต์ที่มีมากกว่าภาษาอังกฤษ รวมถึงวิธีการในการรวมตัวอักษรเป็นข้อความที่มีทั้งสระและวรรณยุกต์

สำหรับงานวิจัยที่เกี่ยวข้องกับการระบุตำแหน่งข้อความภาษาไทยในภาพถ่ายนั้น พบงานวิจัยดังนี้

Jirattitichareon และ Chalidabhongse [12] ได้เสนอวิธีการตรวจจับและแบ่งส่วนข้อความบนภาพป้ายที่มีคุณภาพต่ำโดยมีสมมุติฐานดังนี้

- ข้อความจะมีความเปรียบต่างทั้งในแง่ของสีและความเข้มค่อนข้างมากเมื่อเทียบกับพื้นหลัง
- แต่ละตัวอักษรประกอบด้วยพื้นที่ที่มีความต่อเนื่องกันจำนวนหนึ่ง
- ตัวอักษรในสภาพแวดล้อมเดียวกันจะมีขนาดและความหนาแน่นใกล้เคียงกัน แต่อาจจะมีสีแตกต่างกัน
- ตัวอักษรในสภาพแวดล้อมเดียวกันจะมีพื้นหลังที่คล้ายคลึงกัน

ภาพนำเข้าจะถูกนำไปประมวลผลก่อนและนำไปผ่านกระบวนการหา LOG (Laplacian of Gaussian) เพื่อสร้างแผนภาพขอบ แล้วใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันเพื่อสร้างบริเวณส่วนภาพที่น่าจะมีข้อความอยู่ และคัดกรองบริเวณส่วนภาพที่ได้ ด้วยการหาอัตราส่วนของตัวอักษรเพื่อทำการคัดกรองส่วนภาพที่ไม่ใช่ข้อความ จากนั้นจึงทำการวิเคราะห์การจัดเรียงของตัวอักษรตามหลักของภาษาไทยเพื่อหาสระและวรรณยุกต์ และใช้ GMM (Gaussian Mixture Model) เพื่อแยกพื้นหน้า (foreground) , พื้นหลัง (background) ออกจากกัน และแบ่งส่วนข้อความในแบบจำลองสีที่เลือก อย่างไรก็ตาม งานวิจัยนี้มุ่งเน้นที่การตรวจจับและแบ่งส่วนข้อความบนภาพป้ายซึ่งมีความหลากหลายของภาพน้อยกว่าภาพถ่าย จึงทำให้ขั้นตอนวิธีในงานวิจัยนี้ อาจจะได้ผลไม่ดีนักบนภาพถ่ายฉากธรรมชาติที่มีความซับซ้อนมากกว่า

งานวิจัยของ Woraratpanya และคณะ [13] ได้เสนอวิธีการระบุตำแหน่งข้อความภาษาไทยโดยใช้สมมุติฐานในการพิจารณาการเป็นข้อความในภาพถ่ายฉากธรรมชาติ คล้ายคลึงกับงานของ Jirattitichareon และ Chalidabhongse [12] ในงานวิจัยนี้ ได้นำภาพนำเข้ามาหาแผนภาพขอบและใช้กระบวนการ Fast Boundary Clustering ซึ่งเป็นการสกัดเวคเตอร์ของคุณลักษณะที่ประกอบด้วยค่าสีสูงสุด ค่าสีต่ำสุดและค่าตำแหน่ง คุณลักษณะเหล่านี้สกัดได้จากแผนภาพขอบของแต่ละวัตถุ จากนั้นจัดกลุ่มเวคเตอร์คุณลักษณะที่ได้ด้วยวิธีการจัดกลุ่มแบบk-mean โดยแบ่งออกเป็น 5 กลุ่ม แล้วจึงทำการวิเคราะห์ส่วนประกอบที่เชื่อมต่อกันเพื่อกำจัดส่วนประกอบที่ไม่ใช่ตัวอักษร และสร้างบริเวณส่วนภาพที่มีข้อความขึ้นภายใต้กฎที่ได้กำหนดไว้ อย่างไรก็ตาม เนื่องจากในงานวิจัยนี้ใช้ข้อสมมุติฐานในการพิจารณาการเป็นข้อความที่คล้ายคลึงกับงานวิจัยของ Jirattitichareon และ

Chalidabhongse [12] ซึ่งค่อนข้างเจาะจงกับภาพที่เป็นป้าย จึงอาจจะได้ผลไม่ดีนักเมื่อทำงานบนภาพถ่ายฉากธรรมชาติที่มีความซับซ้อนขององค์ประกอบในภาพค่อนข้างมาก

ในเวลาต่อมา Woraratpanya และคณะ [14] ได้ปรับปรุงวิธีการระบุตำแหน่งข้อความภาษาไทย โดยการใช้กระบวนการ Adaptive Boundary Clustering (ABC) ซึ่งเป็นการสกัดเวคเตอร์ของคุณลักษณะที่ประกอบด้วยค่าสีสูงสุด ค่าสีต่ำสุดและค่าตำแหน่ง คุณลักษณะเหล่านี้สกัดได้จากแผนภาพขอบของแต่ละวัตถุ คุณลักษณะสีที่ได้จากแต่ละวัตถุจะถูกควอนไทซ์ (quantize) เป็น 8 ค่าสีที่กำหนดและแปลงเป็นฮิสโทแกรม จากนั้นจัดกลุ่มเวคเตอร์คุณลักษณะที่ได้ด้วยวิธีการจัดกลุ่มแบบ k-mean โดยแบ่งออกเป็น 5 กลุ่ม แล้วจึงทำการวิเคราะห์ส่วนประกอบที่เชื่อมต่อกันเพื่อกำจัดส่วนประกอบที่ไม่ใช่ตัวอักษร และสร้างบริเวณส่วนภาพที่มีข้อความขึ้นภายใต้กฎที่ได้กำหนดไว้

จากวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ โดยพิจารณาองค์ประกอบที่เชื่อมต่อกันที่ได้เสนอไปนั้น จะสามารถสรุปคุณลักษณะสำคัญ และตัวจำแนกประเภทที่ใช้ได้ดังตารางที่ 2-2

ตารางที่ 2-2 วิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ โดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน

	เสนอโดย	**คุณลักษณะสำคัญที่ใช้**	**ตัวจำแนกประเภทที่ใช้**
1	Subramanian และคณะ [3]	ลายเส้นและสี	การวิเคราะห์คุณลักษณะสำคัญ
2	Jirattitichareon และ Chalidabhongse [12]	ขอบและสี	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันและการวิเคราะห์โมเดลเกาส์เซียน
3	Epshtein และคณะ [4]	การแปลงลายเส้น	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
4	Karaoglu และคณะ [5]	การวิเคราะห์เชิงเรขาคณิตรูปร่างและคุณลักษณะขอบ	ขั้นตอนวิธีต้นไม้แบบสุ่ม
5	Huang และ Ma [7]	การวิเคราะห์ความหยาบและค่าทางสถิติ	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
6	Neumann และ Matas [8]	MSERs และค่าทางสถิติ	ซัพพอร์ตเวคเตอร์แมชชีน
7	Neumann และ Matas [15]	MSERs และค่าทางสถิติ	ซัพพอร์ตเวคเตอร์แมชชีน
8	Yin และคณะ [10]	MSERs ฮิสโทแกรม การเชื่อมต่อกันของขอบและค่าทางสถิติ	อดาบูสต์
9	Iqbal และคณะ [11]	MSERs และคุณลักษณะสำคัญทางตำแหน่ง	เครือข่ายการให้คะแนนแบบเบย์เซียน
10	Woraratpanya และคณะ [13]	ขอบ ตำแหน่งขององค์ประกอบที่เชื่อมต่อกันและความสอดคล้องของสี	Fast Boundary Clustering และการวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
11	Woraratpanya และคณะ [14]	ขอบ ตำแหน่งขององค์ประกอบที่เชื่อมต่อกันและความสอดคล้องของสี	Adaptive Boundary Clustering และการวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน

2.2.2 การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพ (Region / Window Based Approach)

การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพ จะอาศัยเทคนิคลักษณะที่คล้ายคลึงกับการตรวจจับวัตถุ โดยอาศัยการทำ sliding window เพื่อหาบริเวณที่เป็นตัวอักษรในภาพ และนำ window ที่ได้ไปผ่านตัวจำแนกประเภทเพื่อตัดสินว่าบริเวณนั้นเป็นข้อความหรือไม่ จากนั้นจึงสร้างแผนภาพความเชื่อมั่นของบริเวณส่วนภาพที่น่าจะมีตัวอักษรปรากฏอยู่ แล้วนำแผนภาพความเชื่อมั่นนั้น ไปผ่านกระบวนการประมวลผลภายหลังเพื่อสร้างบริเวณส่วนภาพที่มีข้อความปรากฏอยู่ งานวิจัยในประเภทนี้ สามารถแบ่งตามลักษณะของตัวจำแนกประเภทที่ใช้ในการตัดสินว่าเป็นข้อความหรือไม่ ได้ 2 ประเภทคือ ประเภทที่ต้องอาศัยตัวสกัดคุณลักษณะสำคัญที่สร้างด้วยมนุษย์ และประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน

2.2.2.1 การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพประเภทที่ต้องอาศัยตัวสกัดคุณลักษณะสำคัญที่สร้างด้วยมนุษย์

การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพนั้น มีกระบวนการเพื่อจำแนกว่า บริเวณส่วนภาพที่สนใจเป็นข้อความหรือไม่ โดยประเภทตัวจำแนกที่เป็นที่นิยมได้แก่ อดาบูสต์ (Adaboost) ซัพพอร์ตเวคเตอร์แมชชีนและนิวรอลเน็ตเวิร์ค (Neural Network) ซึ่งจำเป็นต้องอาศัยตัวสกัดคุณลักษณะสำคัญ เพื่อเป็นการบ่งชี้หรืออธิบายสิ่งที่ต้องการให้ตัวจำแนกประเภทตัดสินใจ โดยงานวิจัยที่มีความน่าสนใจและใช้ตัวสกัดคุณลักษณะสำคัญที่สร้างโดยมนุษย์นั้นสรุปได้ดังต่อไปนี้

Chen และ Yuille [16] ได้สร้างตัวตรวจจับข้อความโดยใช้ตัวจำแนกประเภทแบบอดาบูสต์ ซึ่งอาศัยชุดตัวสกัดคุณลักษณะสำคัญที่สร้างขึ้นจาก คุณลักษณะสำคัญทางสถิติ คุณลักษณะสำคัญจากฮิสโทแกรม และคุณสมบัติจากการเชื่อมต่อของขอบ งานวิจัยนี้ได้นำตัวสกัดคุณลักษณะสำคัญไปสกัดจากชุดข้อมูลสอน ที่ประกอบด้วยส่วนภาพขนาดเล็ก (patch) ของบริเวณที่เป็นข้อความและไม่ใช่ข้อความ และนำคุณลักษณะที่ได้มาสอนตัวตรวจจับข้อความแบบอดาบูสต์ ตัวตรวจจับข้อความที่ได้นั้นจะถูกนำมาเป็นตัวจำแนกประเภทในกระบวนการทำ sliding window กับภาพนำเข้าเพื่อสร้างบริเวณส่วนภาพที่มีข้อความในภาพถ่ายฉากธรรมชาติ

Gllavata และคณะ [17] ได้เสนอวิธีการตรวจจับข้อความในภาพ โดยอาศัยสัมประสิทธิ์เวฟเลตความถี่สูง ภาพนำเข้าจะถูกนำไปผ่านการแปลงเวฟเลต (Wavelet Transform) แล้วนำสัมประสิทธิ์ในความถี่สูงย่อย HH, HL และ LH มาสร้างเวคเตอร์ของคุณลักษณะสำคัญ โดยทำ

sliding window เพื่อหาค่าเบี่ยงเบนมาตรฐานของฮิสโทแกรม จากนั้นจึงใช้วิธีการจัดกลุ่มแบบ k-mean เพื่อแยกระหว่างข้อความและฉากหลัง และใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันกับส่วนที่เป็นข้อความเพื่อสร้างพิกัดของบริเวณส่วนภาพที่มีข้อความ

จากงานวิจัยของ Hanif และคณะ [18] ได้สร้างตัวตรวจจับข้อความโดยใช้ตัวจำแนกประเภทแบบอดาบูสต์เช่นเดียวกัน แต่ใช้ตัวสกัดคุณลักษณะสำคัญที่นำมาสอนตัวจำแนกประเภทต่างจากงานวิจัยของ Chen และ Yuille [16] ในงานวิจัยนี้ได้ใช้ชุดตัวสกัดคุณลักษณะสำคัญที่สร้างจากคุณลักษณะสำคัญความแตกต่างของค่าเฉลี่ย (Mean Difference Feature : MDF) ค่าเบี่ยงเบนมาตรฐาน และตัวสกัดคุณลักษณะสำคัญ HOG (Histogram of Oriented Gradient) โดยใช้ชุดข้อมูลสอน ICDAR 2003 ตัวตรวจจับข้อความที่ได้จะถูกนำไปใช้บนภาพนำเข้าหลายขนาด เพื่อสร้างแผนภาพของบริเวณส่วนภาพที่มีข้อความ ซึ่งจะถูกรวมเข้าด้วยกัน โดยอาศัยคุณลักษณะสำคัญความหนาแน่นของขอบ จากนั้นจะใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันเพื่อสร้างบริเวณส่วนภาพที่มีข้อความ

Pan และคณะ [19] ได้สร้างตัวตรวจจับข้อความโดยใช้ตัวจำแนกประเภทแบบอดาบูสต์เช่นเดียวกับงานวิจัยที่กล่าวมาข้างต้น ในงานวิจัยนี้ใช้ชุดตัวสกัดคุณลักษณะสำคัญที่สร้างจากตัวสกัดคุณลักษณะสำคัญ HOG และตัวสกัดคุณลักษณะสำคัญ msLBP (Multi-scale Local Binary Pattern) แล้วใช้ตัวตรวจจับข้อความที่ได้กับภาพนำเข้าหลายขนาดเพื่อสร้างบริเวณส่วนภาพที่มีข้อความ แล้วจึงรวมบริเวณที่มีการทับซ้อนกันเข้าด้วยกัน หลังจากนั้นจึงใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันร่วมกับทฤษฎีสนามสุ่มแบบมาคอฟ (Markov Random Field : MRF) ที่ได้ค่าพารามิเตอร์จากการเรียนรู้ด้วยนิวรอลเน็ตเวิร์ค เพื่อสกัดเฉพาะตัวอักษรจากบริเวณส่วนภาพที่ได้

Hanif และ Prevost [20] ได้เสนองานวิจัยในลักษณะคล้ายคลึงกับงานในปี 2008 ตัวตรวจจับข้อความที่ใช้นั้นได้ใช้ตัวจำแนกประเภทอดาบูสต์แบบซับซ้อน (Complexity Adaboost) และใช้ตัวสกัดคุณลักษณะสำคัญเช่นเดียวกับงานวิจัยเดิม ในขั้นตอนของการรวมบริเวณส่วนภาพน่าจะมีข้อความเข้าด้วยกันนั้น จะใช้นิวรอลเน็ตเวิร์คร่วมกับชุดตัวสกัดคุณลักษณะสำคัญได้แก่ คุณลักษณะสำคัญขอบ คุณลักษณะสำคัญเกรเดียนท์ และคุณสมบัติสำคัญทางพื้นผิว โดยคุณลักษณะสำคัญเหล่านี้จะสกัดจากบริเวณส่วนภาพที่น่าจะมีข้อความที่ผ่านการวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน เมื่อผ่านกระบวนการเหล่านี้จะได้บริเวณส่วนภาพที่มีข้อความปรากฏอยู่

ในระยะเวลาใกล้เคียงกัน Pan และคณะ [21] [22] ได้ใช้ตัวตรวจจับข้อความโดยใช้ตัวจำแนกประเภทแบบ วอลด์บูสต์ (Waldboost) ที่เสนอโดย Sochman และ Matas [23] ในงานวิจัยนี้ใช้ตัวสกัดคุณลักษณะสำคัญ HOG ตัวตรวจจับข้อความที่ได้ จะถูกนำไปใช้กับภาพนำเข้าหลายขนาดเพื่อสร้างแผนภาพความมั่นใจของบริเวณที่น่าจะเป็นข้อความ ซึ่งจะถูกนำไปสร้างภาพไบนารี่ด้วย

กระบวนการ Niblack's binarization เพื่อนำไปสู่การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันร่วมกับทฤษฎีสนามสุ่มแบบมีเงื่อนไข (Conditional Random Field) เพื่อสร้างบริเวณส่วนภาพที่มีข้อความปรากฏอยู่

Pan และคณะ [24] ได้เสนอวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติอีกวิธีหนึ่ง โดยแบ่งเป็น 2 ขั้นตอนคือ การตรวจจับข้อความแบบหยาบ (Coarse Text Detection) และการระบุตำแหน่งข้อความแบบละเอียด (Fine Text Localization) ในขั้นตอนการตรวจจับข้อความแบบหยาบนั้น จะใช้ตัวตรวจจับข้อความโดยใช้ตัวจำแนกประเภทแบบวอลด์บูสต์สร้างแผนภาพบริเวณส่วนภาพที่น่าจะเป็นข้อความ ซึ่งจะถูกนำไปผ่านกระบวนการประมวลภายหลังเพื่อสร้างบรรทัดของข้อความของขั้นตอนการตรวจจับข้อความแบบหยาบ จากนั้นจะเข้าสู่ขั้นตอนการระบุตำแหน่งข้อความแบบละเอียด ที่จะนำผลบรรทัดของข้อความจากขั้นตอนการตรวจจับข้อความแบบหยาบ มาสกัดคุณลักษณะสำคัญในแต่ละบริเวณส่วนภาพ โดยใช้ชุดตัวสกัดคุณลักษณะสำคัญที่ประกอบด้วย ตัวสกัดคุณลักษณะสำคัญ HOG ตัวสกัดคุณลักษณะสำคัญ LBP (Local Binary Pattern) คุณลักษณะสำคัญจากการแปลงดีสครีตโคไซน์ (Discrete Cosine Transform : DCT) คุณลักษณะสำคัญจากตัวกรองกาบอร์ (Gabor Filter) และคุณลักษณะสำคัญจากการแปลงเวฟเลต แล้วนำคุณลักษณะสำคัญที่สกัดได้ ไปทำการจำแนกด้วยตัวจำแนกประเภทพหุนาม (Polynomial Classifier) ซึ่งบริเวณส่วนภาพที่ได้รับการจำแนกว่าเป็นข้อความนั้น จะถูกนำไปรวมเพื่อสร้างเป็นบรรทัดของข้อความ และจะถูกตรวจสอบอีกครั้งด้วยการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันเพื่อกำจัดส่วนที่ไม่ใช่ข้อความ และรวมบริเวณส่วนภาพที่เป็นข้อความเข้าด้วยกัน

Bouman และคณะ [25] ได้เสนอวิธีการตรวจจับระบุตำแหน่งข้อความบนป้ายโดยวิธีที่มีความซับซ้อนต่ำ ภาพนำเข้าจะถูกแบ่งเป็นตารางเพื่อหา seed point ของบริเวณส่วนภาพที่มีความคล้ายกัน ทั้งนี้มีสาเหตุเนื่องมาจากสมมุติฐานของป้ายที่ได้ระบุไว้ในงานวิจัยว่า ในบริเวณของภายในป้าย จะเป็นบริเวณส่วนภาพที่มีความคล้ายคลึงกันมาก จากนั้น seed point ที่ได้จะไปผ่านกระบวนการเพื่อหาบริเวณที่มีความคล้ายกันด้วยการ growing seed point นั้น เมื่อได้บริเวณที่มีความคล้ายกันแล้ว จะถูกนำไปวิเคราะห์เพื่อหาบริเวณที่เป็นพื้นหลังของป้ายโดยใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน เพื่อแยกระหว่างข้อความบนป้ายและพื้นหลัง

มีการเสนอวิธีการตรวจจับข้อความบนภาพถ่ายฉากธรรมชาติโดยใช้การวิเคราะห์จากพื้นที่โดดเด่น (Salient Region) โดย Meng และ Song [26] ได้ใช้สมมุติฐานที่ว่าบริเวณส่วนภาพที่มีข้อความจะเป็นจุดที่คนสนใจมอง และมีความโดดเด่นมากกว่าบริเวณอื่นๆ จึงได้สกัดชุดคุณลักษณะสำคัญความโดดเด่นจากภาพ ซึ่งประกอบด้วย คุณลักษณะสำคัญหลายขนาดของความเปรียบต่างสี การกระจายตัวของสี ฮิสโทแกรมของความหนาแน่นของขอบ และความเหมือนกันของลายเส้น จากนั้นจึงหาพื้นที่โดดเด่น โดยใช้ทฤษฎีสนามสุ่มแบบมีเงื่อนไขร่วมกับคุณลักษณะสำคัญที่

สกัดได้ พื้นที่โดดเด่นที่ได้จะถูกนำไปวิเคราะห์องค์ประกอบที่เชื่อมต่อกันและสกัดคุณลักษณะสำคัญ ได้แก่ อัตราส่วนขององค์ประกอบที่เชื่อมต่อกัน ขนาดขององค์ประกอบที่เชื่อมต่อกัน ความหยาบของคอนทัวร์ (contour roughness) ค่า compactness ความเหมือนกันของลายเส้นและขนาดของลายเส้นซึ่งคุณลักษณะสำคัญเหล่านี้จะถูกใช้เพื่อจำแนกโดยตัวจำแนกประเภทซัพพอร์ตเวคเตอร์แมชชีน ว่าองค์ประกอบที่เชื่อมต่อกันส่วนนั้นเป็นบริเวณส่วนภาพที่มีข้อความหรือไม่

ในงานวิจัยของ Mishra และคณะ [27] ได้เสนอวิธีการระบุตำแหน่งข้อความบนภาพถ่ายฉากธรรมชาติโดยใช้ตัวสกัดคุณลักษณะสำคัญ HOG ร่วมกับตัวจำแนกประเภทซัพพอร์ตเวคเตอร์แมชชีน โดยทำ sliding window บนภาพนำเข้าหลายขนาด แต่ละ window นั้นจะถูกจำแนกเป็นข้อความหรือไม่ใช่ข้อความ ซึ่ง window ที่เป็นข้อความนั้นจะถูกคัดกรองจากค่าความดี (goodness score) ที่คำนวณจากค่าความเชื่อมั่นของการเป็นข้อความที่ได้จากตัวจำแนกประเภท อัตราส่วนของ window ร่วมกับค่าเฉลี่ยและค่าความแปรปรวนของอัตราส่วนของข้อความจากชุดข้อมูลสอน widow ที่ผ่านกระบวนการคัดกรองนั้น จะถูกนำไปผ่านขั้นตอนวิธี Non-Maximum Suppression (NMS) เพื่อแก้ปัญหาการทับซ้อนกันของ window ซึ่ง window ที่มีค่าความเชื่อมั่นในการเป็นข้อความต่ำจะถูกกำจัดออกไปและให้ผลตอบเป็นบริเวณส่วนภาพที่มีข้อความ

Bo และคณะ [28] ได้เสนอวิธีการใช้คุณลักษณะสำคัญสหสัมพันธ์แบบท้องถิ่นของเกรเดียนท์ (gradient local correlation feature) โดยใช้การวิเคราะห์จากแผนภาพขอบที่ได้จากขั้นตอนวิธีของ Canny [29] โดยมีสมมุติฐานว่า ในแต่ละส่วนภาพของบริเวณที่เป็นข้อความนั้น จะมีการกระจายความหนาแน่นของขอบ และการกระจายของความสอดคล้องของลายเส้น เป็นลักษณะการกระจายแบบเกาส์เซียน หลังจากนั้น จะทำการสกัดแต่ละตัวอักษรโดยใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันในแต่ละบริเวณที่เป็นข้อความและนำไปสกัดคุณลักษณะสำคัญ HOG และทำการคัดกรองว่าเป็นตัวอักษรหรือไม่ อีกครั้งหนึ่งด้วยตัวจำแนกประเภทซัพพอร์ตเวคเตอร์แมชชีน แต่ละตัวอักษรที่ผ่านการคัดกรองนั้นจะถูกรวมเข้าเป็นบริเวณของข้อความโดยใช้คุณลักษณะสำคัญสี และตำแหน่งข้องแต่ละตัวอักษร

จากวิธีการเหล่านี้ เราสามารถตั้งข้อสังเกตได้ในลักษณะเดียวกันกับการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยใช้องค์ประกอบที่เชื่อมต่อกันคือ ตัวสกัดคุณลักษณะสำคัญที่ใช้นั้นอาจจะไม่เหมาะสมกับภาษาไทย เนื่องจากความต่างของระดับสระและวรรณยุกต์ที่มีมากกว่าภาษาอังกฤษ ทำให้อาจจะไม่สามารถระบุตำแหน่งข้อความภาษาไทยได้อย่างถูกต้อง โดยอาจจะสามารถระบุตัวอักษรที่มีลักษณะคล้ายกันทั้งในภาษาไทยและภาษาอังกฤษ แต่อาจจะไม่สามารถระบุตำแหน่งสระและวรรณยุกต์ได้ดีเป็นต้น

วิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพ ประเภทที่ต้องอาศัยตัวสกัดคุณลักษณะสำคัญที่สร้างด้วยมนุษย์นั้น จะสามารถสรุปคุณลักษณะสำคัญ ตัวจำแนกประเภทและวิธีในการรวมองค์ประกอบเข้าด้วยกันที่ใช้ ได้ดังตารางที่ 2-3

ตารางที่ 2-3 ขั้นตอนวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพประเภทที่ต้องอาศัยตัวสกัดคุณลักษณะสำคัญที่สร้างด้วยมนุษย์

	เสนอโดย	**คุณลักษณะสำคัญที่ใช้**	**ตัวจำแนกประเภทที่ใช้**	**วิธีในการรวมองค์ประกอบ**
1	Chen และ Yuille [16]	ค่าทางสถิติ ฮิสโทแกรม และการเชื่อมต่อกันของขอบ	อดาบูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
2	Gllavata และคณะ [17]	ค่าสัมประสิทธิ์เวฟเลตความถี่สูง	ขั้นตอนวิธีการจัดกลุ่มแบบ k-mean	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
3	Hanif และคณะ [18]	ความแตกต่างของค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานและ HOG	อดาบูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
4	Pan และคณะ [19]	LBP และ HOG	อดาบูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน โดยวิเคราะห์ร่วมกับทฤษฎีสนามสุ่มแบบมาคอฟ
5	Hanif และ Prevost [20]	ความแตกต่างของค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานและ HOG	อดาบูสต์แบบซับซ้อน	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันและนิวรอลเน็ตเวิร์ค
6	Pan และคณะ [21]	HOG	วอลด์บูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
7	Pan และคณะ [22]	HOG	วอลด์บูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน

8	Pan และคณะ [24]	เกรเดียนต์, คุณลักษณะขอบ, การแปลงดิสครีทโคไซน์, HOG และ LBP	วอลด์บูสต์	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน
9	Meng และ Song [26]	ชุดคุณลักษณะสำคัญความโดดเด่น	ซัพพอร์ตเวคเตอร์แมชชีน	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันร่วมกับทฤษฎีสนามสุ่มแบบมีเงื่อนไข
10	Mishra และคณะ [27]	HOG	ซัพพอร์ตเวคเตอร์แมชชีน	Non-Maximum Suppression (NMS)
11	Bo และคณะ [28]	สหสัมพันธ์แบบท้องถิ่นของเกรเดียนท์ ขอบ และ HOG	ซัพพอร์ตเวคเตอร์แมชชีน	การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน

2.2.2.2 การระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน

วิธีการการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอนและวิธีที่ใช้ตัวสกัดคุณลักษณะสำคัญที่สร้างด้วยมนุษย์นั้นมีข้อแตกต่างกันคือ ตัวสกัดคุณลักษณะสำคัญในวิธีที่ได้เสนอไปก่อนหน้าจะถูกสร้างขึ้นโดยอาศัยลักษณะเฉพาะของบริเวณที่มีข้อความ ตัวอย่างเช่นความหนาของลายเส้นความคล้ายคลึงกันของสีหรือองค์ประกอบที่เชื่อมต่อกัน แต่วิธีที่สร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอนนั้น จะเรียนรู้ตัวสกัดคุณลักษณะสำคัญจากชุดข้อมูลสอนได้เอง โดยไม่ต้องอาศัยมนุษย์เป็นคนกำหนดสิ่งที่จะบ่งชี้ถึงคุณลักษณะสำคัญของข้อความ โดยงานวิจัยการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ ที่สร้างตัวสกัดคุณลักษณะสำคัญจากชุดข้อมูลสอน ที่น่าสนใจได้แก่

Coates และคณะ [30] ได้เสนอวิธีการตรวจจับข้อความและรู้จำตัวอักษรบนภาพถ่ายฉากธรรมชาติ โดยใช้การเรียนรู้คุณลักษณะสำคัญจากชุดข้อมูลสอนแบบ unsupervised (unsupervised feature learning) เริ่มจากการเก็บชุดข้อมูลสอนจากส่วนภาพขนาดเล็ก ที่เป็นข้อความและไม่ใช่ข้อความขนาด 8x8 จุดภาพ ไปผ่านกระบวนการประมวลผลก่อน และใช้วิธีการจัดกลุ่มแบบ k-mean เพื่อเรียนรู้เซตของเวคเตอร์ที่จะนำมาเป็นพจนานุกรม ที่จะนำไปใช้สกัดคุณลักษณะสำคัญจากข้อมูลนำเข้า เมื่อเสร็จสิ้นกระบวนการเรียนรู้ตัวสกัดคุณลักษณะสำคัญแล้ว จะนำผลที่ได้ไปใช้เป็นตัวสกัดคุณลักษณะสำคัญสำหรับตัวจำแนกประเภทซัพพอร์ตเวคเตอร์แมชชีน ซึ่งจะใช้เป็นตัวจำแนกประเภทระหว่างข้อความและไม่ใช่ข้อความ โดยเรียนรู้จากชุดข้อมูลสอนมาตรฐาน ICDAR 2003 ซึ่งในงานวิจัยนี้ ภาพนำเข้าผ่านกระบวนการ sliding window โดยใช้ window ขนาด 32x32 จุดภาพนำไปผ่านตัวจำแนกประเภทที่สร้างขึ้น เพื่อสร้างแผนภาพความมั่นใจของส่วนภาพที่มีข้อความ

ต่อมาได้มีการเสนอวิธีการตรวจจับข้อความและรู้จำตัวอักษรบนภาพถ่ายฉากธรรมชาติ โดยใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค โดย Wang และคณะ [31] ในงานวิจัยนี้ ตัวสกัดคุณลักษณะสำคัญในชั้นแรกสำหรับคอนโวลูชันนอล นิวรอลเน็ตเวิร์คที่ใช้ตรวจจับข้อความนั้น ใช้วิธีการเรียนรู้ในลักษณะเดียวกับงานวิจัยของ Coates และคณะ [30] โครงสร้างของคอนโวลูชันนอล นิวรอลเน็ตเวิร์คที่ใช้นั้นประกอบด้วย ชั้นนำเข้าซึ่งเป็นภาพขนาด 32x32 จุดภาพ และชั้นคอนโวลูชัน1 ซึ่งเป็นตัวกรองขนาด 8x8 จุดภาพ เป็นตัวกรองที่ได้จากกระบวนการเรียนรู้ที่ได้กล่าวไป ตามด้วยการหาค่าเฉลี่ยด้วยตัวกรองขนาด 5x5 จุดภาพ จากนั้นจะตามด้วยชั้นคอนโวลูชันที่2 ซึ่งเป็นตัวกรองขนาด 2x2 จุดภาพและหาค่าเฉลี่ยด้วยตัวกรองขนาด 2x2 อีกครั้งหนึ่ง หลังจากนั้น คุณลักษณะสำคัญ

ทั้งหมดที่ได้จะถูกเชื่อมต่อเข้ากับชั้นการจำแนก ซึ่งจะจำแนกคุณลักษณะที่ได้เป็น 2 ประเภทคือ เป็นข้อความหรือไม่ใช่ข้อความ กระบวนการเหล่านี้จะถูกทำบนภาพหลายขนาดร่วมกับการทำ sliding window บนภาพเหล่านั้นเพื่อสร้างแผนภาพความมั่นใจของบริเวณส่วนภาพที่มีข้อความ ซึ่งจะนำไปผ่านกระบวนการประมวลผลภายหลังเพื่อระบุตำแหน่งของบริเวณส่วนภาพที่มีข้อความ

งานวิจัยทั้ง 2 งานวิจัยนี้เป็นการสร้างตัวสกัดคุณลักษณะสำคัญจากชุดข้อมูลสอน โดยสามารถสรุปขั้นตอนวิธีในการเรียนรู้คุณลักษณะสำคัญ และตัวจำแนกประเภทที่ใช้ได้ดังตารางที่ 2-4

ตารางที่ 2-4 ขั้นตอนวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน

	เสนอโดย	**วิธีการสร้าง ตัวสกัดคุณลักษณะสำคัญ**	**ตัวจำแนกประเภท**
1	Coates และคณะ [30]	ขั้นตอนวิธีการจัดกลุ่มแบบ k-mean	ซัพพอร์ตเวคเตอร์แมชชีน
2	Wang และคณะ [31]	ขั้นตอนวิธีการจัดกลุ่มแบบ k-mean และคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค	คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค

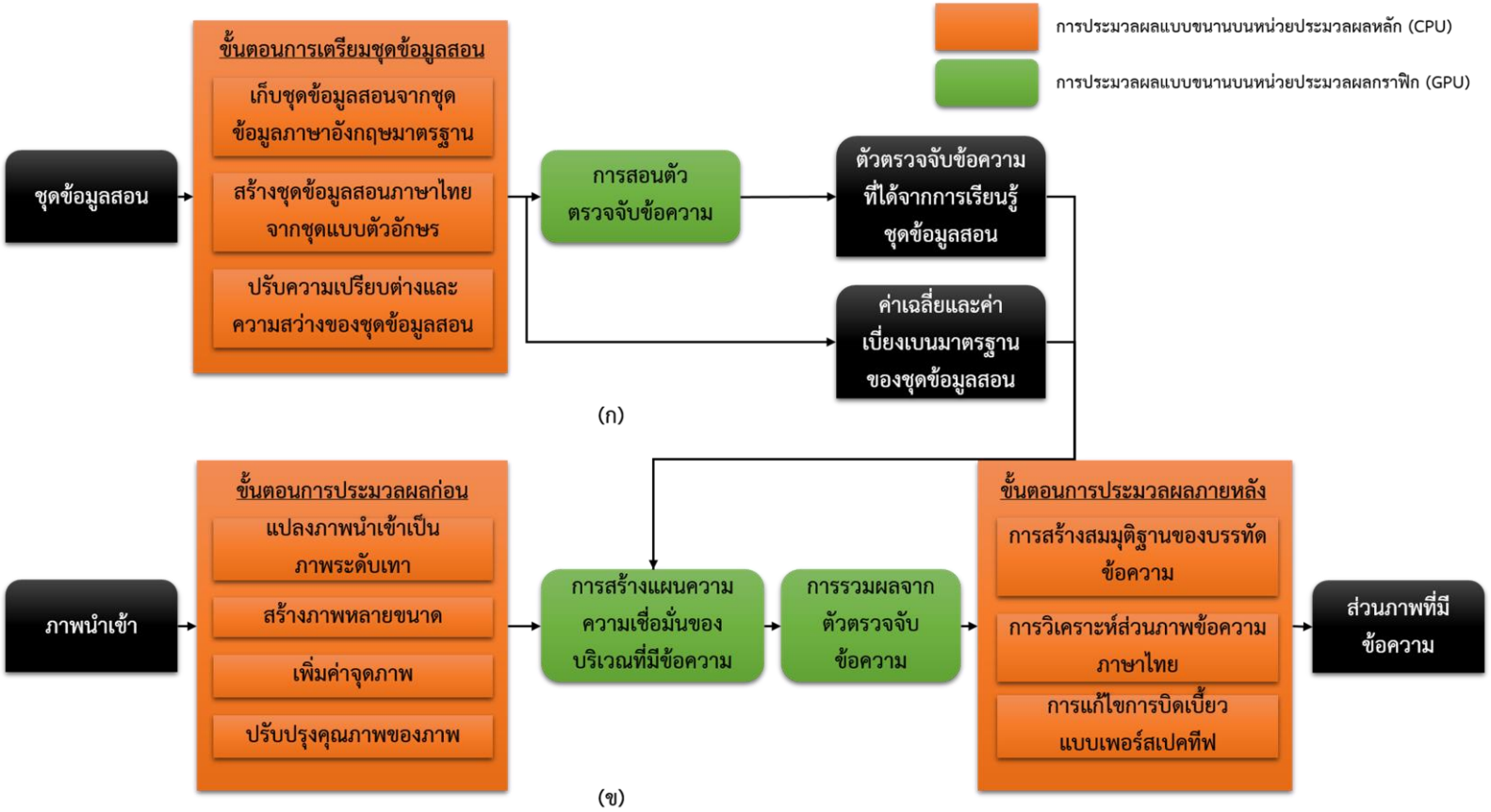
จากขั้นตอนวิธี ในการระบุตำแหน่งข้อความบนภาพถ่ายฉากธรรมชาติที่ได้สำรวจวรรณกรรมนั้น ผู้วิจัยได้เลือกใช้วิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพ ประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน เนื่องจากความแตกต่างระหว่างโครงสร้างของภาษาไทยที่มีตัวอักษร วรรณยุกต์และสระ เรียงตัวในหลายระดับและมีขนาดไม่เท่ากัน ทำให้วิธีการวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติพิจารณาองค์ประกอบที่เชื่อมต่อกันใช้กับภาษาไทยได้ไม่ดีนัก เพราะสระและวรรณยุกต์ที่มีขนาดเล็ก จะถูกคัดกรองเป็นสัญญาณรบกวน ด้วยเงื่อนไขการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันที่ได้ถูกออกแบบสำหรับภาษาอังกฤษ และการสกัดองค์ประกอบที่เชื่อมต่อกันบนภาพที่มีปรากฏการมัว และมีสัญญาณรบกวนอาจทำได้ไม่ดีนัก ซึ่งสำหรับวิธีการวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาบริเวณส่วนภาพ ประเภทที่สามารถสร้างตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอนนั้นอาจจะสามารถเรียนรู้คุณลักษณะบางประการ ที่มีความเหมาะสมกับภาษาไทยจากชุดข้อมูลสอนที่มีตัวอักษรภาษาไทยร่วมอยู่ได้ ทำให้ได้ผลลัพธ์ของการระบุตำแหน่งข้อความภาษาไทยบนภาพถ่ายฉากธรรมชาติมีความแม่นยำยิ่งขึ้น

บทที่ 3
ขั้นตอนวิธีที่เสนอ

ขั้นตอนวิธีที่เสนอในวิทยานิพนธ์ฉบับนี้ ประกอบด้วย 2 ขั้นตอนหลักคือ

1. ขั้นตอนการสร้างตัวตรวจจับข้อความ เพื่อสร้างตัวจำแนกประเภทของส่วนภาพที่เป็นข้อความและไม่ใช่ข้อความ จากชุดข้อมูลสอนมาตรฐาน และชุดข้อมูลสอนที่ผู้วิจัยได้จัดเตรียมไว้

2. ขั้นตอนที่ใช้ในการประมวลผลจริง เพื่อสกัดส่วนภาพที่มีข้อความ เป็นขั้นตอนที่นำตัวตรวจจับข้อความที่สร้างนั้นมาใช้งานจริงบนภาพนำเข้า ซึ่งในขั้นตอนนี้ประกอบด้วย ขั้นตอนการประมวลผลก่อน ขั้นตอนการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ ขั้นตอนการรวมผลจากตัวตรวจจับข้อความและขั้นตอนการประมวลผลภายหลัง ซึ่งขั้นตอนการประมวลจริงนั้นจะทำการประมวลผลแบบขนานบนหน่วยประมวลผลกราฟิก (Graphic Processing Unit : GPU) โดยใช้ CUDA (Compute Unified Device Architecture) ร่วมกับการประมวลผลแบบขนานบนหน่วยประมวลผลกลาง (Central Processing Unit : CPU) โดยใช้ OpenMP โดยได้แสดงภาพรวมของขั้นตอนวิธีที่เสนอนั้น ในรูปที่ 3-1



รูปที่ 3-1 ภาพรวมของวิธีที่เสนอ ประกอบด้วย 2 ขั้นตอนหลัก
(ก) ขั้นตอนการสร้างตัวตรวจจับข้อความ
(ข) ขั้นตอนการประมวลผลจริงเพื่อสกัดส่วนภาพที่มีข้อความ

3.1 ขั้นตอนการสร้างตัวตรวจจับข้อความ

เนื่องจากในวิทยานิพนธ์นี้ ได้เลือกใช้ตัวตรวจจับข้อความที่สร้างจากขั้นตอนวิธีการเรียนรู้ของเครื่องจักรที่สามารถเรียนรู้ตัวสกัดคุณลักษณะสำคัญได้จากชุดข้อมูลสอน โดยได้เลือกใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค ซึ่งต้องมีการเตรียมชุดข้อมูลสอน และต้องมีการหาโครงสร้างที่เหมาะสมที่ให้ความแม่นยำในการจำแนกระหว่างส่วนภาพที่เป็นข้อความ และไม่ใช่ข้อความ โดยผ่านขั้นตอนการสอน และทดสอบบนชุดข้อมูลสอนและชุดข้อมูลทดสอบที่ได้จัดเตรียมไว้

3.1.1 ขั้นตอนการเตรียมชุดข้อมูลสอน

ที่มาของชุดข้อมูลสอนและทดสอบที่ใช้ในวิทยานิพนธ์นี้จะมาจาก 2 แหล่งคือ ชุดข้อมูลมาตรฐานภาษาอังกฤษที่ได้จากชุดข้อมูล ICDAR 2003, ICDAR 2011, SVT (Street View Text), ImageNet และ Char74k และชุดข้อมูลภาษาไทยที่ผู้วิจัยรวบรวมและสร้างขึ้นเอง

สำหรับการสร้างชุดข้อมูลสอนภาษาอังกฤษในชุดข้อมูล ICDAR 2003, ICDAR 2011, SVT (Street View Text) ได้มีการจัดเตรียมคำตอบ (Ground truth) ของบริเวณที่มีข้อความปรากฏอยู่ในรูปแบบของพิกัดจุดมุมซ้ายบน ความกว้างและความยาวของสี่เหลี่ยมของบริเวณที่มีข้อความปรากฏอยู่ โดยได้แสดงรูปตัวอย่างและบริเวณของส่วนภาพที่มีข้อความที่อยู่ในคำตอบของชุดข้อมูล ในรูปที่ 3-2

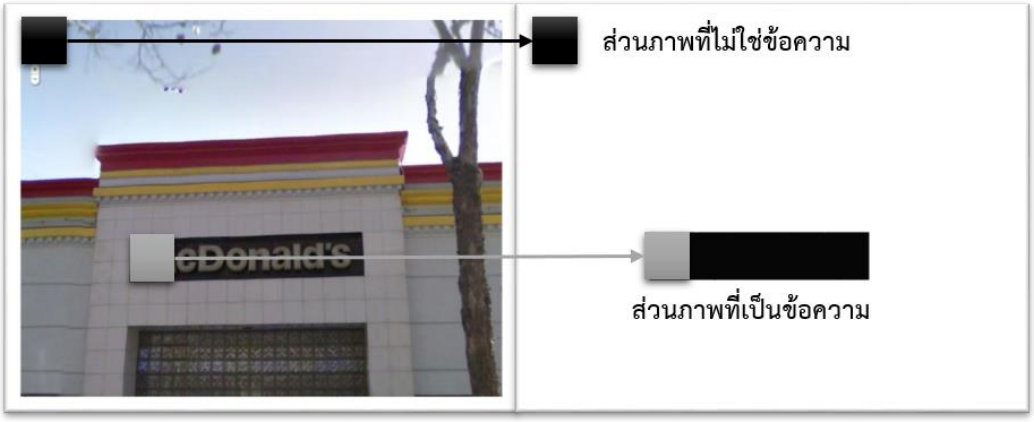
รูปที่ 3-2 ภาพตัวอย่างของชุดข้อมูลมาตรฐานภาษาอังกฤษ
จากชุดข้อมูลมาตรฐานภาษาอังกฤษ ICDAR 2003, ICDAR2011 และ SVT ตามลำดับ

จากข้อมูลพิกัดจุดมุมซ้ายบน ความกว้างและความยาวของสี่เหลี่ยมของบริเวณที่มีข้อความปรากฏอยู่ที่ได้จากคำตอบของชุดข้อมูลสอน จะทำการสร้างมาส์กของบริเวณที่มีข้อความปรากฏอยู่ โดยกำหนดให้ บริเวณที่เป็นข้อความนั้นแทนด้วยสีขาว (ค่าของจุดภาพ = 255) และบริเวณที่ไม่ใช่ข้อความแทนด้วยสีดำ (ค่าของจุดภาพ = 0) ดังแสดงในรูปที่ 3-3

(ก) (ข)

รูปที่ 3-3 (ก) ภาพและข้อมูลคำตอบของบริเวณที่เป็นส่วนภาพ
(ข) มาส์กของบริเวณที่มีข้อความปรากฏอยู่

หลังจากนั้น จะทำการเก็บชุดข้อมูลสอนโดยการใช้ window ขนาด 32x32 จุดภาพบนภาพนำเข้าจากชุดข้อมูลมาตรฐาน โดยกำหนดให้ window ที่มีการซ้อนทับของพื้นที่มากกว่า 80% เมื่อเทียบกับบริเวณที่มีข้อความจะถูกจัดเก็บเป็นส่วนภาพที่เป็นข้อความ ในขณะที่ window ที่มีการซ้อนทับของพื้นที่น้อยกว่า 80% จะถูกจัดเก็บเป็นส่วนภาพที่ไม่ใช่ข้อความ ดังแสดงในรูปที่ 3-4



รูปที่ 3-4 การจำแนกประเภทส่วนภาพ
(ก) ตัวอย่างภาพจากชุดข้อมูลมาตรฐาน
(ข) มาส์กคำตอบของบริเวณที่มีข้อความ

ชุดข้อมูล Char74k นั้นจะอยู่ในรูปแบบของตัวอักษรจากภาพถ่ายฉากธรรมชาติที่ถูกตัดแบ่ง (segment) มาเรียบร้อยแล้ว โดยมีตัวอย่างแสดงในรูปที่ 3-5 ดังนั้นจะทำการปรับภาพให้อยู่ในขนาด 32x32 จุดภาพ เช่นเดียวกับส่วนภาพที่เป็นข้อความอื่นๆ

รูปที่ 3-5 ตัวอย่างชุดข้อมูลมาตรฐาน Char74k

สำหรับการสร้างชุดข้อมูลสอนข้อความภาษาไทยนั้น เนื่องจากในขณะที่ดำเนินงานวิจัยนั้นยังไม่พบชุดข้อมูลมาตรฐานของข้อความภาษาไทยในภาพถ่ายฉากธรรมชาติ ดังนั้นชุดข้อมูลสอนในส่วนของภาษาไทยนั้นจะประกอบด้วย 2 ส่วนคือ ชุดข้อมูลที่เก็บจากภาพถ่ายที่ผู้วิจัยได้ทำการจัดเก็บจากภาพถ่ายฉากธรรมชาติ และชุดข้อมูลที่ได้ทำการสังเคราะห์ภาพข้อความ ให้มีความใกล้เคียงกับข้อความในภาพถ่ายฉากธรรมชาติจากชุดรูปแบบตัวอักษร โดยได้แสดงตัวอย่างดังรูปที่ 3-6

(ก) (ข)

รูปที่ 3-6 ตัวอย่างชุดข้อมูลสอนภาษาไทย
(ก) ชุดข้อมูลสอนข้อความภาษาไทยจากภาพถ่ายฉายธรรมชาติ
(ข) ชุดข้อมูลสอนภาษาไทยที่สังเคราะห์จากชุดรูปแบบตัวอักษร

เมื่อทำการจัดเตรียมชุดข้อมูลสอนทั้งชุดข้อมูลสอนภาษาไทยและภาษาอังกฤษดังขั้นตอนที่กล่าวมา จะได้ส่วนภาพของบริเวณที่เป็นข้อความและไม่เป็นข้อความ ประเภทละ 2,500,000 ส่วนภาพ หลังจากนั้นนำชุดข้อมูลสอนไปแปลงเป็นภาพระดับเทา และนำไปผ่านกระบวนการปรับความเปรียบต่างและความสว่าง (Contrast and Brightness Normalization) ซึ่งเป็นไปตามสมการ

$$g(x,y) = \frac{f(x,y)-\mu(x,y)}{\sigma(x,y)} \tag{3.1}$$

กำหนดให้ $g(x,y), f(x,y), \mu(x,y)$ และ $\sigma(x,y)$ คือ ส่วนภาพผลลัพธ์ของกระบวนการปรับความเปรียบต่างและความสว่าง, ส่วนภาพนำเข้า, ค่าเฉลี่ยของส่วนภาพทั้งชุดข้อมูลสอน และค่าเบี่ยงเบนมาตรฐานของส่วนภาพจากทั้งชุดข้อมูลสอน ชุดข้อมูลสอนที่ผ่านกระบวนการนี้แล้วจะถูกสุ่มเพื่อนำไปใช้ในการสอนตัวตรวจจับข้อความ ซึ่งทางผู้วิจัยได้ทำการสุ่มส่วนภาพที่ใช้ในการสอนตัวตรวจจับข้อความ ทั้งส่วนภาพที่เป็นข้อความและส่วนภาพที่ไม่ใช่ข้อความ ประเภทละ 500,000 ส่วนภาพเพื่อเป็นชุดข้อมูลสอน และประเภท 250,000 ส่วนภาพเพื่อเป็นชุดข้อมูลทดสอบ

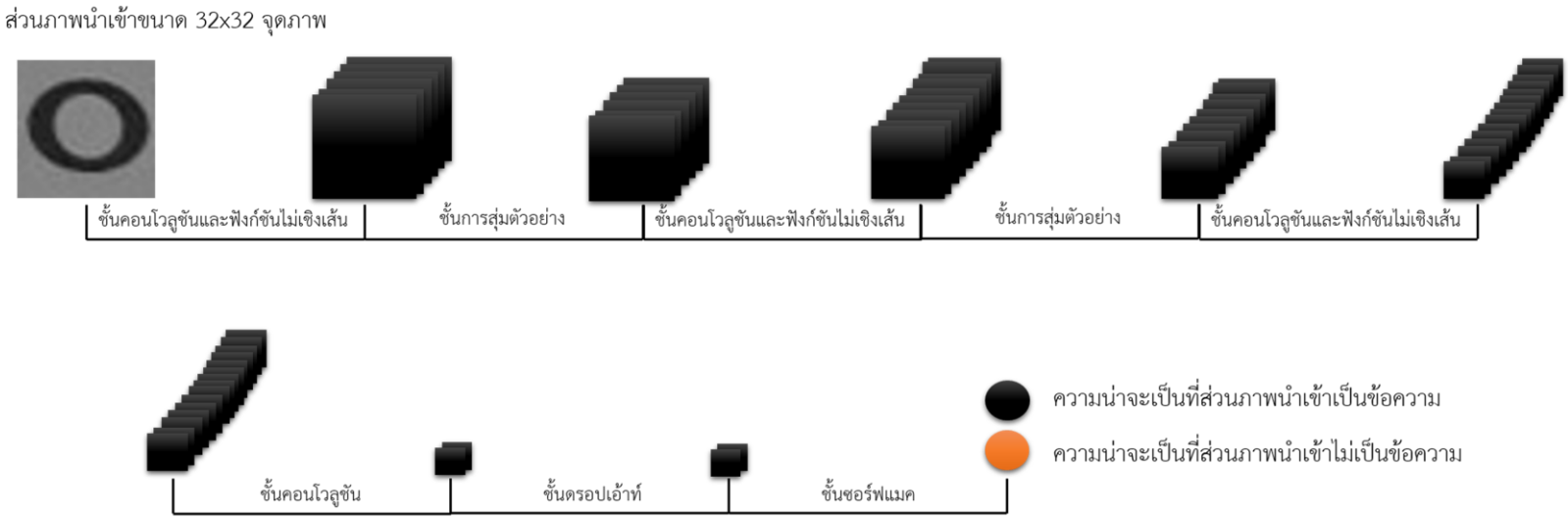
โดยขั้นตอนการสอนตัวตรวจจับข้อความนั้น จะทำสอนโดยใช้หน่วยประมวลผลกราฟิก NVIDIA GeForce GTX780Ti และใช้ Caffe [32] ซึ่งเป็นเฟรมเวิร์คการเรียนรู้ของเครื่องจักร

3.1.2 ขั้นตอนการสอนตัวตรวจจับข้อความ

ในขั้นตอนนี้จะนำชุดข้อมูลสอนที่ผ่านกระบวนการจัดเตรียมในหัวข้อ 3.1.1 มาใช้ในการสอนตัวตรวจจับข้อความ ซึ่งผู้วิจัยได้เลือกใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค เป็นตัวตรวจจับข้อความ ซึ่งต้องมีการทดลองเพื่อหาโครงสร้างและค่าพารามิเตอร์ที่เหมาะสม เพื่อให้ได้ผลของการจำแนกระหว่างส่วนภาพที่เป็นข้อความและไม่ใช่ข้อความให้แม่นยำที่สุด โดยโครงสร้างของคอนโวลูชันนอล นิวรอลเน็ตเวิร์คนั้น จากการสำรวจวรรณกรรมในหัวข้อที่เกี่ยวข้องกับการใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์คเพื่อการจำแนกประเภทภาพแล้วพบว่า จะมีการจัดเรียงโครงสร้างเป็นชุดของชั้นคอนโวลูชัน ชั้นฟังก์ชันไม่เชิงเส้นและชั้นสุ่มตัวอย่าง โดยมีจำนวนชุดขึ้นอยู่กับขนาดของชุดข้อมูลที่นำมาใช้กับคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค และสำหรับค่าพารามิเตอร์ของแต่ละชั้นที่ใช้หาได้จากการทดลอง ซึ่งจากการทดลองแล้วพบว่าโครงสร้างและค่าพารามิเตอร์ที่ใช้ ดังตารางที่ 3-1 ให้ผลลัพธ์ในการจำแนกระหว่างส่วนภาพที่เป็นข้อความและส่วนภาพที่ไม่ใช่ข้อความดีที่สุด และได้แสดงภาพโครงสร้างที่ใช้ในรูปที่ 3-7

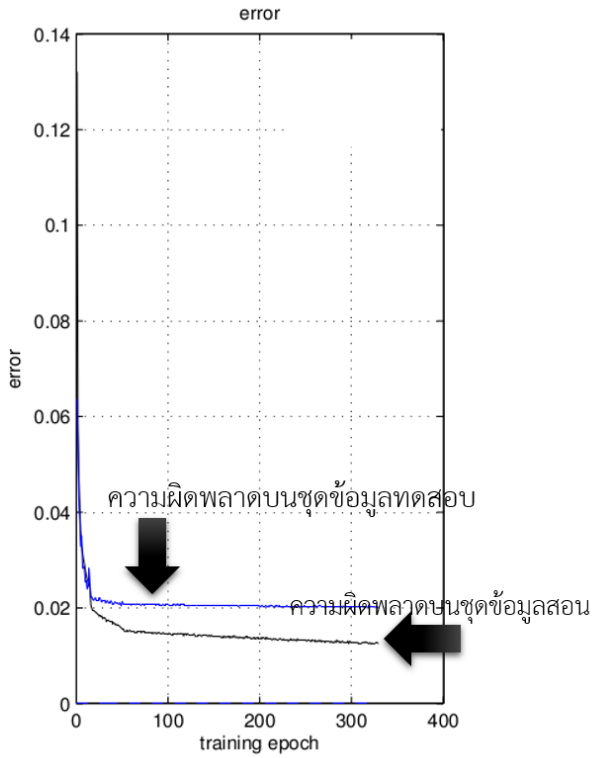
ตารางที่ 3-1 โครงสร้างของตัวตรวจจับข้อความที่ใช้

ชั้นที่	ประเภทชั้น	ขนาดของเคอร์เนล	ขนาดของผลลัพธ์
	นำเข้า	-	32x32
1	ชั้นคอนโวลูชัน	5x5x1x16	28x28x16
2	ชั้นฟังก์ชันไม่เชิงเส้นเรคติไฟเออร์	-	28x28x16
3	ชั้นการสุ่มตัวอย่างแบบมากสุด	2x2	14x14x16
4	ชั้นคอนโวลูชัน	5x5x16x64	10x10x64
5	ชั้นฟังก์ชันไม่เชิงเส้นเรคติไฟเออร์	-	10x10x64
6	ชั้นการสุ่มตัวอย่างแบบมากสุด	2x2	5x5x64
7	ชั้นคอนโวลูชัน	5x5x64x512	1x1x512
8	ชั้นฟังก์ชันไม่เชิงเส้นเรคติไฟเออร์	-	1x1x512
9	ชั้นคอนโวลูชัน	1x1x512x2	1x1x2
10	ชั้นดรอปเอ้าท์	ค่าพารามิเตอร์ = 0.5	1x1x2
11	ชั้นซอฟต์แมกซ์	-	1x1x2



รูปที่ 3-7 โครงสร้างของตัวตรวจจับข้อความที่ใช้ในวิทยานิพนธ์

จากการทดสอบตัวตรวจจับข้อความโดยใช้โครงสร้างที่เลือกบนชุดข้อมูลสอนและชุดข้อมูลทดสอบ โดยทำการสอนจำนวน 325 รอบ จะได้ความแม่นยำอยู่ที่ 99.2% บนชุดข้อมูลสอนจำนวน 500,000 ส่วนภาพ และ 98% บนชุดข้อมูลทดสอบจำนวน 250,000 ส่วนภาพ โดยผลความผิดพลาดของการสอนจะแสดงในรูปที่ 3-8



รูปที่ 3-8 ผลการทดสอบความผิดพลาดของตัวตรวจจับข้อความที่ใช้ในวิทยานิพนธ์

การสอนและทดสอบตัวตรวจจับข้อความนั้นจะใช้ ขั้นตอนวิธีแบคพรอพพาเกชั่นในการปรับแก้ค่าพารามิเตอร์ของแต่ละชั้น และทดสอบบนหน่วยประมวลผลกราฟิก โดยได้แสดงการเปรียบเทียบเวลาที่ใช้ในการสอนตัวตรวจจับข้อความ 1 รอบ เมื่อทดสอบบนหน่วยประมวลผลหลัก ทดสอบบนหน่วยประมวลผลหลักแบบขนาน และการรันบนหน่วยประมวลผลกราฟิก ดังตารางที่ 3-2

ตารางที่ 3-2 เปรียบเทียบเวลาที่ใช้ในการสอน 1 รอบบนสภาพแวดล้อมการประมวลผลที่แตกต่างกัน

สอนตัวตรวจจับข้อความบน	**เวลาในการสอนต่อ 1 รอบ (นาที)**	**เปรียบเทียบความเร็วในการสอนกับการรันบนหน่วยประมวลผลหลัก (เท่า)**
หน่วยประมวลหลัก	24	1
หน่วยประมวลผลหลักแบบขนาน	10.5	2.29
หน่วยประมวลผลกราฟิก	1.4	17.14

3.2 ขั้นตอนการประมวลผลจริงเพื่อสกัดส่วนภาพที่มีข้อความ

3.2.1 ขั้นตอนการประมวลผลก่อน

เนื่องจากภาพถ่ายฉากธรรมชาติที่จะนำมาระบุตำแหน่งของข้อความที่ปรากฏในภาพนั้นมีหลากหลายรูปแบบ ดังนั้นจึงต้องมีการปรับปรุงคุณภาพของภาพให้มีความเหมาะสมก่อนจะเข้าสู่เข้าตอนถัดไป ภาพนำเข้าจะถูกแปลงเป็นภาพระดับเทาเพื่อลดความซับซ้อนในการประมวลผลและจากผลการทดสอบพบว่า เมื่อเปรียบเทียบความแม่นยำในการระบุตำแหน่งข้อความระหว่างการใช้ภาพสีและภาพระดับเทานั้น ให้ความแม่นยำที่ใกล้เคียงกัน ดังนั้นในวิทยานิพนธ์นี้จึงทำการแปลงภาพสีนำเข้าเป็นภาพระดับเทาตามสมการ 3.2

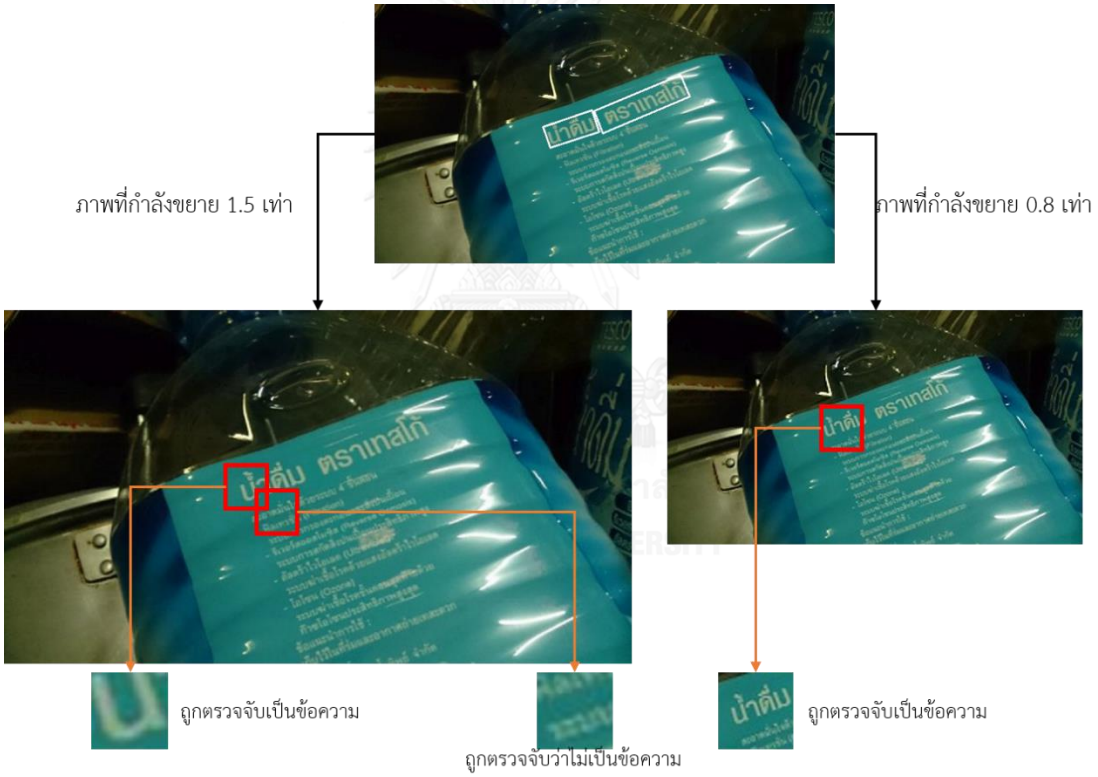
$$f(x,y) = 0.2126 * R(x,y) + 0.7152 * G(x,y) + 0.0722 * B(x,y) \qquad (3.2)$$

เมื่อ $R(x,y), G(x,y)$ และ $B(x,y)$ คือค่าสีแดง เขียวและน้ำเงินของแต่ละจุดภาพที่ตำแหน่ง (x,y) และ $f(x,y)$ คือภาพระดับเทา โดยมีสาเหตุที่ให้ค่าน้ำหนักของสีเขียว มากกว่าสีอื่นๆทั้งนี้เนื่องจาก ในการรับรู้สีของมนุษย์ด้วยการมองเห็นนั้น ตาของมนุษย์จะตอบสนองกับสีเขียวมากที่สุดและตอบสนองน้อยที่สุดกับสีน้ำเงิน

เนื่องจากตัวตรวจจับข้อความที่สร้างนั้น มีขนาดของภาพนำเข้ากำหนดอยู่ที่ 32x32 จุดภาพ เพื่อให้ตัวตรวจจับข้อความสามารถที่จะตรวจจับข้อความจากภาพถ่ายฉากธรรมชาติที่มีขนาดไม่คงที่ได้ ในขั้นตอนต่อมาจะทำการสร้างภาพหลายขนาด (Multiscale Image) จากภาพนำเข้าต้นฉบับ โดย

กำหนดให้กำลังขยาย (scaling) มีค่าเป็น1.5,1.2,1.1,1.0,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1 เท่าของภาพนำเข้าโดยใช้ขั้นตอนวิธีไบคิวบิค (bicubic)

สาเหตุที่ภาพหลายขนาดที่สร้างนั้น มีกำลังขยายที่ไม่สมมาตร ทั้งนี้มีสาเหตุเนื่องจาก ในการที่จะตรวจจับข้อความที่มีตัวอักษรขนาดใหญ่ เช่นมีตัวอักษรขนาดใหญ่เพียงตัวเดียวปรากฏในภาพ ตัวอักษรนี้จะถูกตัวตรวจจับข้อความตรวจพบว่าเป็นข้อความที่ภาพหลายขนาดกำลังขยายน้อย ในขณะที่ตัวอักษรที่มีขนาดเล็กมากๆนั้น เมื่อนำไปผ่านกระบวนการขยาย ให้มีขนาดใหญ่ขึ้นด้วยขั้นตอนวิธีไบคิวบิคจะทำให้เกิดการปรากฏการมัวขึ้น ซึ่งจากการทดลองแล้วพบว่า โดยส่วนมากตัวตรวจจับข้อความจะไม่สามารถที่จะตรวจจับข้อความบริเวณดังกล่าวได้ ดังแสดงการเปรียบเทียบส่วนภาพที่ได้จากการสร้างภาพหลายขนาด ที่กำลังขยาย 1.5 เท่าและ 0.8 เท่า ในรูปที่ 3-9

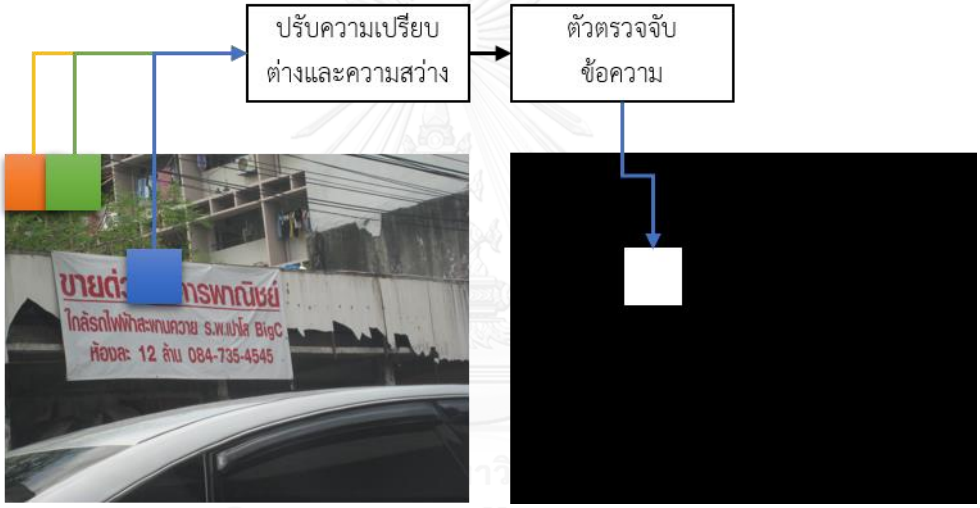


รูปที่ 3-9 การเปรียบเทียบบริเวณส่วนภาพที่ได้จากการสร้างภาพหลายขนาด ที่กำลังขยาย 1.5 เท่า 0.8 เท่า และผลลัพธ์การจำแนกประเภทส่วนภาพ

หลังจากนั้นจะทำการเพิ่มค่าจุดภาพ (padding) โดยให้ค่าน้ำหนักเท่ากับ 0 ขนาด 32 จุดภาพในทุกด้านของภาพหลายขนาด หลังจากนั้นจะนำไปผ่านกระบวนการปรับปรุงคุณภาพโดยใช้ขั้นตอนวิธี การทำให้มัวแบบเกาส์เซียนและ Unsharp Masking เพื่อลดสัญญาณรบกวนที่ปรากฏในภาพดังที่กล่าวในหัวข้อ 2.1.2

3.2.2 ขั้นตอนการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

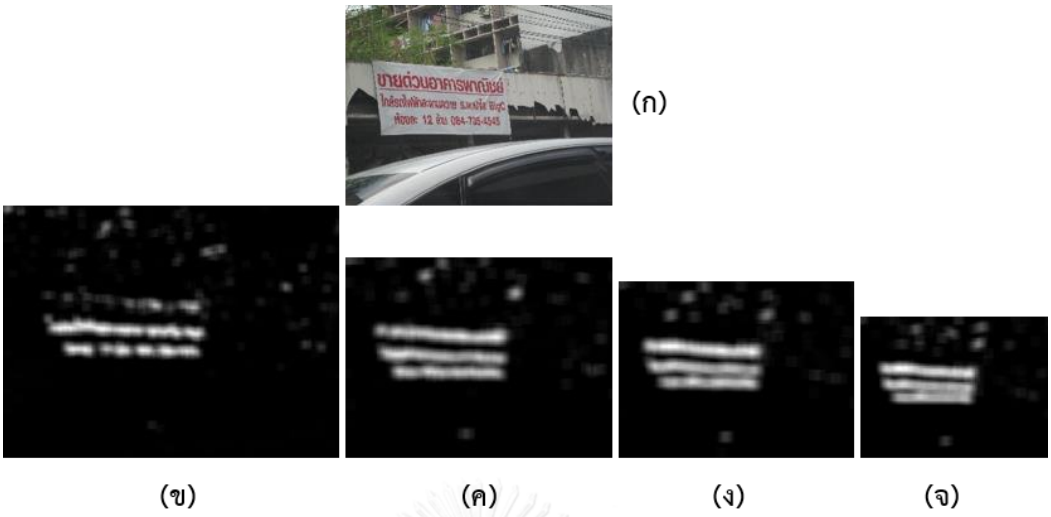
ขั้นตอนนี้จะนำตัวตรวจข้อความที่ถูกสร้างในหัวข้อ 3.1.2 มาใช้ในการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความปรากฏอยู่ โดยในแต่ละภาพหลายขนาดที่ได้จากขั้นตอนการประมวลก่อน จะใช้ขั้นตอนวิธี sliding window ดังแสดงในรูปที่ 3-10 แต่ละส่วนภาพจะถูกนำไปผ่านกระบวนการปรับความเปรียบต่างและความสว่างดังสมการที่ 3.1 โดยยังคงใช้ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานที่คำนวณได้จากชุดข้อมูลสอน ก่อนจะถูกจำแนกโดยตัวตรวจจับข้อความ เพื่อทำการจำแนกว่าเป็นส่วนภาพที่เป็นข้อความหรือไม่ใช่ข้อความ โดยตัวตรวจจับข้อความจะให้ผลในรูปแบบค่าความน่าจะเป็น ซึ่งจะถูกนำไปสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความของแต่ภาพหลายขนาด



รูปที่ 3-10 การสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

ขั้นตอนวิธี sliding window นั้น จะใช้ window ขนาด 32x32 จุดภาพซึ่งมีขนาดเท่ากับภาพนำเข้าของตัวตรวจจับข้อความ และเลื่อน window ทีละ 4 จุดภาพและทำการประมวลผลทั้งหมดบนหน่วยประมวลผลกราฟิก

การสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความนั้น ขั้นตอนแรกจะสร้างภาพที่มีขนาดเท่ากับภาพหลายขนาดและมีค่าจุดภาพทุกจุดเท่ากับ 0 เมื่อ window ใดที่ถูกจำแนกว่าเป็นส่วนภาพที่มีข้อความด้วยความน่าจะเป็นมากกว่า 0.7 หรือมีความน่าจะเป็นที่ส่วนภาพนั้นจะเป็นข้อความมากกว่า 70% แล้ว ส่วนภาพของแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความในบริเวณเดียวกันกับ window นั้นๆจะถูกทบค่าความน่าจะเป็นที่ได้บนบริเวณดังกล่าว ตัวอย่างของภาพนำเข้าและแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ ในแต่ละภาพหลายขนาดนั้น ได้แสดงในรูปที่ 3-11



รูปที่ 3-11 (ก) ภาพนำเข้า
(ข) – (จ) แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่กำลังขยาย 1.5, 1.0, 0.8, 0.5 เท่าของภาพนำเข้า

3.2.3 ขั้นตอนการรวมผลจากตัวตรวจจับข้อความ

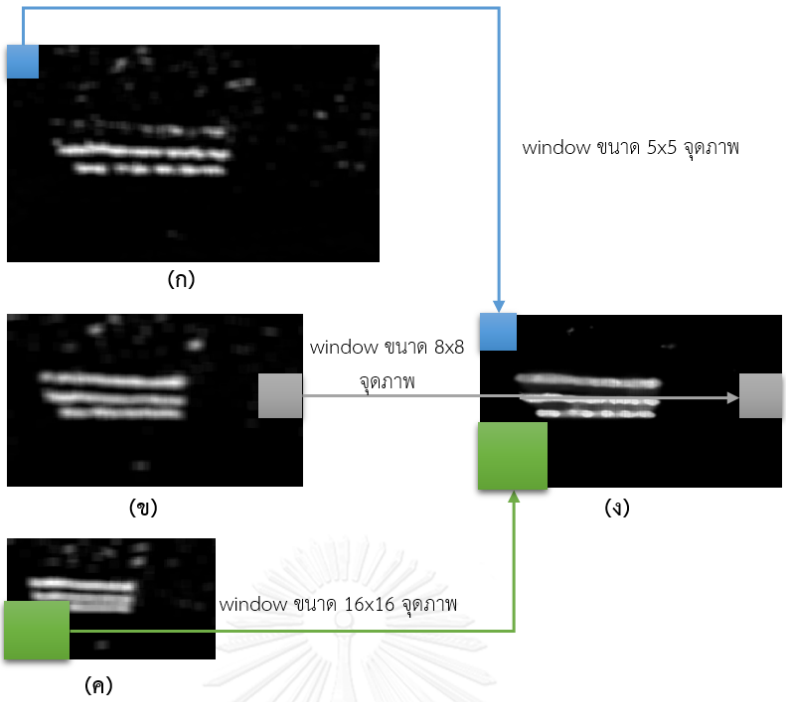
ขั้นตอนนี้จะใช้แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความของภาพหลายขนาดที่ได้จากขั้นตอน 3.2.2 เพื่อสร้างผลรวมของแผนภาพความมั่นใจที่มีขนาดเท่ากับภาพนำเข้าต้นฉบับ โดยใช้วิธีการกวาด window ขนาด n x n จุดภาพ เป็นไปตามตารางที่ 3-3 ซึ่งขนาดของ n จะสอดคล้องกับกำลังขยาย ดังสมการ 3.3 โดยกำหนดให้ S คือกำลังขยาย

$$n = \frac{8}{s} \tag{3.3}$$

ตารางที่ 3-3 ขนาดของ window ที่ใช้ในการกวาดแต่ละขนาดของแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

กำลังขยาย	**ขนาดของ window (จุดภาพ)**
1.5	5x5
1.2	7x7
1.1	7x7
1.0	8x8
0.9	9x9
0.8	10x10
0.7	11x11
0.6	13x13
0.5	16x16
0.4	20x20
0.3	27x27
0.2	40x40
0.1	80x80

ในแต่ละแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความของภาพหลายขนาด window ที่มีค่าเฉลี่ย $\boldsymbol{\tau}$ โดยในวิทยานิพนธ์นี้ ได้กำหนดให้ $\boldsymbol{\tau}$ = 15 ซึ่งเป็นค่าที่หาได้จากการค้นหาค่าพารามิเตอร์แบบกริด (grid search) จากผลลัพธ์ของบริเวณที่มีข้อความของชุดข้อมูลสอน แต่ละ window จะถูกจับคู่กับบริเวณที่เหมาะสมในแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่มีขนาดเท่ากับภาพนำเข้า ซึ่งแต่ละ window จะถูกปรับขนาดให้มีความเหมาะสมกับขนาดของภาพนำเข้าโดยเทียบกับขนาดของแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ ดังแสดงตัวอย่างในรูปที่ 3-12



รูปที่ 3-12 การรวมผลจากตัวตรวจจับข้อความ
(ก) แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่กำลังขยาย 1.5 เท่าของภาพนำเข้า
(ข) แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่ขนาดภาพนำเข้า
(ค) แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่กำลังขยาย 0.5 เท่าของภาพนำเข้า
(ง) ผลของการรวมแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

3.2.4 ขั้นตอนการประมวลผลภายหลัง

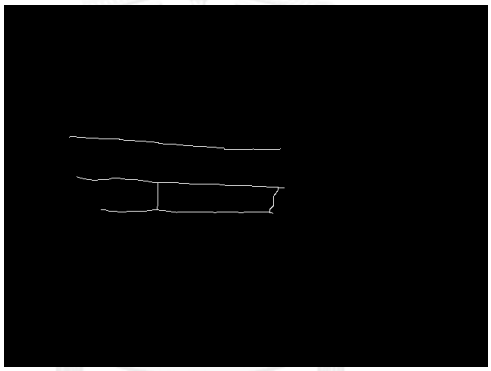
ขั้นตอนนี้จะใช้ผลจากการรวมแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ มาใช้ในการสร้างสมมุติฐานของบรรทัดข้อความ เพื่อนำไปผ่านขั้นตอนการวิเคราะห์รูปแบบของข้อความภาษาไทย

3.2.4.1 ขั้นตอนการสร้างสมมุติฐานของบรรทัดข้อความ

จากแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่สร้างได้จากกระบวนการในขั้นตอน 3.2.3 จะถูกนำไปผ่านกระบวนการกรองโดยค่าขีดแบ่ง ซึ่งหาได้จากการค้นหาค่าพารามิเตอร์แบบกริด (grid search) จากผลลัพธ์ของบริเวณที่มีข้อความของชุดข้อมูลสอน และนำผลที่ได้ไปผ่านเทคนิคมอร์โฟโลยี (morphology) แบบขยายตัว (dilate) และปิด (closing) เพื่อจัดการกับบริเวณของข้อความความ ที่บางส่วนที่มีการเชื่อมติดกัน ซึ่งกระบวนการมอร์โฟโลยี จะใช้เคอร์เนลสี่เหลี่ยมจัตุรัสขนาด 5x5 จุดภาพ

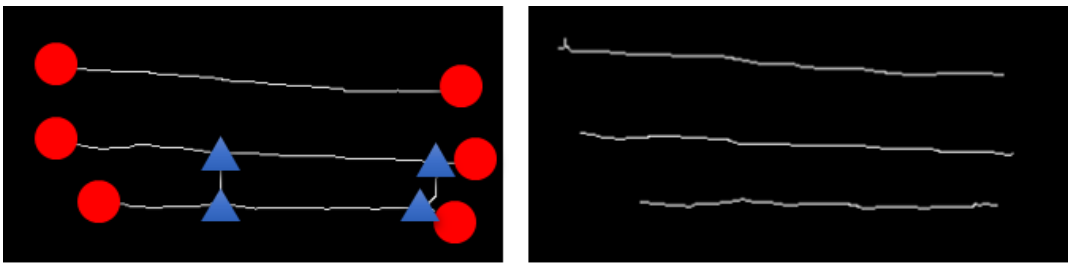
รูปที่ 3-13 ผลที่ได้จากการกระบวนการกรองค่าขีดแบ่งและมอร์โฟโลยี
บนแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่ผ่านกระบวนการกรองโดยค่าขีดแบ่งและเทคนิคมอร์โฟโลยี จะถูกนำไปผ่านกระบวนการทำให้บาง (thinning) เพื่อหาสมมุติฐานของบริเวณที่มีบรรทัดของข้อความปรากฏอยู่ ทำให้สามารถประมาณบรรทัดของบริเวณที่มีข้อความ



รูปที่ 3-14 สมมุติฐานของบรรทัดที่ได้จากกระบวนการทำให้บาง

เนื่องจากในบางครั้ง ถ้าข้อความที่ปรากฏในภาพถ่ายฉากธรรมชาติมีส่วนที่ติดกันหรืออยู่ใกล้กันมาก สมมุติฐานบรรทัดของข้อความจะมีโอกาสที่จะเชื่อมถึงกันได้ ทำให้การจัดกลุ่มของข้อความผิดเพี้ยนไป ดังนั้นจึงต้องนำไปผ่านกระบวนการเพื่อตรวจสอบบรรทัดของข้อความอีกครั้งหนึ่ง โดยการหาจุดเริ่มต้น, จุดปลาย (สัญลักษณ์วงกลม) และจุดที่เป็นจุดเชื่อมต่อของเส้นบรรทัด (สัญลักษณ์สามเหลี่ยม) จากนั้นจึงทำการคำนวณค่ามุมระว่างกลุ่มจุดเหล่านั้น คู่ของจุดเริ่มต้นและจุดปลาย ที่ให้ค่ามุมน้อยที่สุดเมื่อเทียบกับคู่อื่นๆ กลุ่มจุดที่อยู่ในกลุ่มเหล่านั้นจะถูกนับว่าเป็นบรรทัดของข้อความที่ถูกสร้างขึ้นใหม่ ดังแสดงในรูปที่ 3-15



รูปที่ 3-15 การตรวจสอบการเชื่อมต่อกันของเส้นบรรทัดและผลลัพธ์ที่ได้

เมื่อได้บรรทัดของข้อความแล้ว จะทำการสร้างบริเวณส่วนภาพที่มีข้อความปรากฏอยู่จากเส้นบรรทัดที่ได้ โดยใช้กระบวนการ sliding window บนแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ โดยใช้ window ขนาด 8x8 จุดภาพ window ที่มีค่าเฉลี่ยมากกว่าค่าขีดแบ่งที่กำหนดไว้จะถูกจัดกลุ่มเข้ากับเส้นบรรทัดที่ได้โดยทำการวัดระยะทางแบบยูคลิด (Euclidean Distance) ระหว่างจุดกึ่งกลางของ window กับจุดที่ใกล้ที่สุดของแต่ละเส้นบรรทัดของข้อความ

รูปที่ 3-16 ผลลัพธ์ของส่วนภาพที่มีข้อความที่ได้ก่อนการวิเคราะห์ส่วนภาพข้อความภาษาไทย

3.2.4.2 ขั้นตอนการวิเคราะห์ส่วนภาพข้อความภาษาไทย

เนื่องจากผลที่ได้จากขั้นตอน 3.2.4.1 นั้นในบางครั้งจะไม่สามารถเก็บสระและวรรณยุกต์บางส่วนได้ ทั้งนี้มีสาเหตุเนื่องจากสระและวรรณยุกต์นั้น มีขนาดค่อนข้างเล็กเมื่อเทียบกับตัวอักษร ทำให้ถูกมองว่าเป็นสัญญาณรบกวน ดังนั้นจึงต้องมีการวิเคราะห์เพิ่มเติมในส่วนของสระและวรรณยุกต์ที่หายไป

รูปที่ 3-17 ตัวอย่างส่วนภาพจากขั้นตอนที่ 3.5.1

ในขั้นตอนแรกจะทำการหาขอบของแต่ละส่วนภาพที่เป็นข้อความ โดยใช้ขั้นตอนวิธีของ Canny [29] ร่วมกับการตัดแบ่งส่วนวัตถุ แล้วทำการประมาณความสูงของข้อความโดยใช้พิกัดจุดกึ่งกลางของกรอบแต่ละตัวอักษรทั้งด้านล่างและบน ร่วมกับการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) เพื่อหาเส้นบรรทัดโดยประมาณ

รูปที่ 3-18 การวิเคราะห์หาเส้นบรรทัดโดยประมาณ

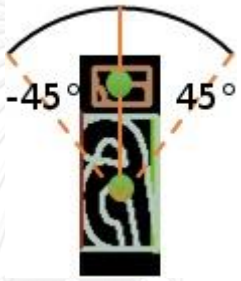
หลังจากนั้นจะทำการวิเคราะห์ลักษณะของข้อความบรรทัดภาษาไทยแล้ว จะทำการเพิ่มระยะทั้งด้านบนและล่างของเส้นบรรทัด จากการวิเคราะห์จากชุดข้อมูลสอนพบว่าพบว่าโดยส่วนใหญ่แล้ววรรณยุกต์และสระที่ไม่ได้อยู่ในบรรทัดหลัก จะมีระยะห่างจากกรอบของแต่ละตัวอักษรอยู่ในระยะ 30 - 50% ของความสูงตัวอักษร ดังนั้นจะแบ่งบรรทัดของข้อความออกเป็น 3 ส่วนคือ ส่วนบน ส่วนบรรทัดหลัก และส่วนล่างดังรูปที่ 3-19

รูปที่ 3-19 การวิเคราะห์โครงสร้างของข้อความ ตามลักษณะบรรทัดที่ได้

เมื่อวิเคราะห์เพื่อแบ่งแต่ละวัตถุที่ได้ทำการตัดแบ่ง ออกเป็นประเภทตามการแบ่งบรรทัดของข้อความ แล้วจะทำการวิเคราะห์ความเชื่อมโยงระหว่างสระ/วรรณยุกต์กับแต่ละตัวอักษรว่าอยู่ระยะที่เหมาะสมหรือไม่ โดยยกตัวอย่างการวิเคราะห์ความเชื่อมโยงระหว่างตัวอักษรและสระ/วรรณยุกต์กับตัวอักษรที่ได้ทำการตัดแบ่งในรูปที่ 3-20

รูปที่ 3-20 ตัวอักษรที่จะทำการวิเคราะห์ความเชื่อมโยงระหว่างตัวอักษรกับสระและวรรณยุกต์

จากการวิเคราะห์ตัวอักษร โดยใช้สมมุติฐานที่ได้จากชุดข้อมูลสอนที่ว่า สระและวรรณยุกต์ที่อยู่ในบรรทัดหลักจะวางตัวอยู่ในช่วง -45 – 45 องศา เมื่อเทียบกับจุดศูนย์กลางมวลของตัวอักษรนั้นๆ โดยวัดระยะทางแบบยูคลิดที่ใกล้ที่สุดระหว่างส่วนประกอบที่อยู่ในบรรทัดหลัก กับส่วนประกอบที่อยู่ในบรรทัดบนและล่าง เพื่อทำการจับคู่ระหว่างส่วนประกอบเหล่านั้น โดยได้ทำการยกตัวอย่างการวิเคราะห์ตัวอักษร “ล” และไม้โท ดังแสดงในรูปที่ 3-21



รูปที่ 3-21 การวิเคราะห์ร่วมกับระหว่างตัวอักษรกับวรรณยุกต์

โดยถ้ามีส่วนประกอบที่อยู่ในบรรทัดบนและล่างมากกว่า 70% อยู่ในตำแหน่งที่เหมาะสมตามเงื่อนไขที่ได้กำหนดไว้แล้ว จะให้ผลส่วนภาพของข้อความตามที่ได้ทำการวิเคราะห์ไปข้างต้น แต่ถ้าหากส่วนประกอบที่อยู่ในบรรทัดบนและล่างมีน้อยกว่า 70% อยู่ในตำแหน่งที่เหมาะสมจะให้ผลลัพธ์ของส่วนภาพก่อนทำจะการวิเคราะห์ส่วนภาพข้อความภาษาไทย

3.2.4.3 ขั้นตอนการแก้ไขการบิดเบี้ยวแบบเพอร์สเปคทีฟ

เนื่องจากในบางกรณีจะเกิดการบิดเบี้ยวของภาพแบบเพอร์สเปคทีฟ (Perspective Distortion) จากมุมในการถ่ายภาพ ซึ่งมีผลทำให้บางส่วนของภาพมีการบิดเบี้ยวและอาจจะส่งผลต่อประสิทธิภาพการทำงาน หากต้องการนำส่วนภาพของข้อความที่ระบุตำแหน่งได้ไปทำการรู้จำตัวอักษรต่อไป ในขั้นตอนนี้จะทำการแก้ไขการบิดเบี้ยวดังกล่าวจากผลในขั้นตอนที่ 3.2.4.2 ซึ่งแสดงในรูปที่ 3-22 สำหรับขั้นตอนวิธีในการแก้ไขการบิดเบี้ยวของภาพแบบเพอร์สเปคทีฟนั้น จากข้อมูลพิกัดมุมของภาพ 4 จุดที่ทราบจากส่วนภาพที่ได้ในขั้นตอนที่ 3.2.4.2 นั้น จะทำการสร้างเมทริกซ์การแปลง จากพิกัดมุมทั้ง 4 ของภาพ ไปยังพิกัดมุมทั้ง 4 ของสี่เหลี่ยมที่สามารถบรรจุส่วนภาพของ

ข้อความได้ โดยได้แสดงในรูปที่ 3-23 เมื่อกำหนดให้สัญลักษณ์วงกลมแทนพิกัดมุมทั้ง 4 ของส่วนภาพที่ต้องการแก้ไข และสัญลักษณ์วงกลมแทนพิกัดมุมทั้ง 4 ของสี่เหลี่ยมที่สามารถบรรจุส่วนภาพของข้อความที่ต้องการแก้ไขได้

รูปที่ 3-22 ส่วนภาพของข้อความก่อน การแก้ไขการบิดเบี้ยวของภาพแบบเปอร์สเปคทีฟ

รูปที่ 3-23 พิกัดมุมที่ใช้ในขั้นตอนการแก้ไขการบิดเบี้ยวของภาพแบบเปอร์สเปคทีฟ

รูปที่ 3-24 ผลลัพธ์ของการแก้การบิดเบี้ยวแบบเปอร์สเปคทีฟ

บทที่ 4
การทดลองและผลการทดลอง

การทดสอบขั้นตอนวิธีที่ได้เสนอนั้น ได้ทำการทดสอบบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษได้แก่ ชุดข้อมูลทดสอบ ICDAR 2003, ICDAR 2011 และทำการทดสอบบนชุดข้อมูลทดสอบภาษาไทยที่จัดเตรียมโดยผู้วิจัย และชุดข้อมูลทดสอบจากการแข่งขัน BEST 2015 : Text Location Detection Contest ทำการทดสอบโดยใช้เครื่องคอมพิวเตอร์ที่มีรายละเอียดดังต่อไปนี้

หน่วยประมวลผลหลัก (CPU)	: Intel Core i5 4430
แรม (RAM)	: 16 GB
หน่วยประมวลผลกราฟิก (GPU)	: NVIDIA GeForce GTX 780Ti
ระบบปฏิบัติการ (OS)	: Ubuntu 15.04 (x64)
คอมไพเลอร์	: gcc 4.9.2
ไลบรารี่ (Library)	: OpenCV 3.0, NVIDIA CUDA 7.0, NVIDIA cuDNN

4.1 ขั้นตอนวิธีที่ใช้ในการทดสอบ

วิทยานิพนธ์ฉบับนี้ใช้เกณฑ์ในการประเมินผลที่ได้จากขั้นตอนวิธีที่เสนอ โดยใช้การประเมินผลสำหรับแต่ละชุดข้อมูลทดสอบ ดังวิธีต่อไปนี้

4.1.1 การประเมินผลด้วยเกณฑ์การทดสอบของ International Conference on Document analysis and Recognition 2003 (ICDAR 2003)

การประเมินด้วยเกณฑ์ของ ICDAR นั้นจะอ้างอิงจากงานวิจัยของ Lucas และคณะ [6] เนื่องจากการระบุตำแหน่งของข้อความของชุดคำตอบของชุดข้อมูลมาตรฐานนั้นทำโดยมนุษย์ ซึ่งอาจมีความคลาดเคลื่อนแตกต่างกัน ดังนั้นจึงได้มีการกำหนดค่าที่ใช้ในการวัดความแม่นยำของการระบุตำแหน่งข้อความ โดยมีการนิยามดังต่อไปนี้

$$precision\ (p) = \frac{c}{|E|} \tag{4.1}$$

ความเที่ยง (Precision) สามารถคำนวณได้จาก อัตราส่วนระหว่างจำนวนส่วนภาพของข้อความที่ถูกต้องที่ได้จากขั้นตอนวิธี (c) และจำนวนของส่วนภาพทั้งหมดที่หาได้จากขั้นตอนวิธี $(|E|)$

$$recall\ (r) = \frac{c}{|T|} \tag{4.2}$$

ค่ารีคอล (Recall) สามารถคำนวณได้จากอัตราส่วนระหว่างจำนวนส่วนภาพของข้อความที่ถูกต้องที่ได้จากขั้นตอนวิธี (c) และจำนวนของส่วนภาพทั้งหมดจากคำตอบของชุดข้อมูลทดสอบ $(|T|)$

แต่อย่างไรก็ตาม เนื่องจากปัญหาการระบุตำแหน่งข้อความในภาพนั้น การที่จะให้กรอบของข้อความที่ได้จากขั้นตอนวิธี ไม่คลาดเคลื่อนเลยเมื่อเทียบกับคำตอบของชุดข้อมูลทดสอบที่ระบุตำแหน่งโดยมนุษย์นั้นเป็นไปได้ยาก ดังนั้นจึงได้มีการนิยามการจับคู่ที่ถูกต้อง m_p คืออัตราส่วนระหว่าง พื้นที่ซ้อนทับของกรอบสี่เหลี่ยมที่ครอบคลุมส่วนภาพของข้อความจากขั้นตอนวิธีที่ทดสอบ และกรอบสี่เหลี่ยมที่ครอบคลุมส่วนภาพของข้อความจากคำตอบของชุดข้อมูลทดสอบ และได้มีการนิยามการจับคู่ที่ดีที่สุด (best match) $m(r, R)$ สำหรับกรอบสี่เหลี่ยมที่สามารถครอบคลุมส่วนภาพของข้อความที่ได้จากขั้นตอนวิธี (r) เมื่อทำการเปรียบเทียบกับเซตของกรอบสี่เหลี่ยมที่สามารถครอบคลุมส่วนภาพของข้อความจากคำตอบของชุดข้อมูลทดสอบ (R) ดังสมการ

$$m(r, R) = \max m_p(r, r') \mid r' \in R \tag{4.3}$$

ดังนั้นจะสามารถนิยาม ความเที่ยง ค่ารีคอล และค่าการวัดเอฟ (f) โดยใช้เงื่อนไขการจับคู่ที่ดีที่สุด ได้ดังสมการ

$$p' = \frac{\sum_{r_e \in E} m(r_e, T)}{|E|} \tag{4.4}$$

$$r' = \frac{\sum_{r_t \in T} m(r_t, E)}{|T|} \tag{4.5}$$

$$f = \frac{1}{\frac{\propto}{p'} + \frac{1-\propto}{r'}} \tag{4.6}$$

เมื่อกำหนดให้ $\propto$ เป็นค่าน้ำหนักระหว่าง ความเที่ยง และค่ารีคอล โดยปกติแล้วจะกำหนดให้ $\propto = 0.5$

4.1.2 การประเมินผลด้วยเกณฑ์การทดสอบจากการแข่งขัน BEST 2015 : Text Location Detection Contest

ชุดข้อมูลทดสอบที่ใช้ในการแข่งขัน BEST 2015 : Text Location Detection Contest นั้นมีความแตกต่างจากชุดข้อมูลอื่นๆ โดยอ้างอิงจากงานวิจัยของ Wolf และ Jolion [33] เนื่องจากข้อความที่ปรากฏในภาพนั้นไม่ได้อยู่ในแนวนอนเพียงอย่างเดียว โดยมีลักษณะข้อความทั้งรูปแบบที่มีการเอียงเนื่องจากมุมของการถ่ายร่วมอยู่ด้วย โดยมีการวัดความถูกต้อง 2 รูปแบบคือ

4.1.2.1 การวัดความถูกต้องระดับจุดภาพ (Pixel Level Detection)

การวัดความถูกต้องรูปแบบนี้ จะใช้การตัดสินโดยมีข้อมูลนำเข้าคือ มาส์ก (mask) ของบริเวณที่เป็นข้อความที่ได้จากขั้นตอนวิธีที่ออกแบบ และชุดข้อมูลทดสอบ ซึ่งมาส์กนั้นอาจจะไม่ได้มีรูปร่างเป็นสี่เหลี่ยม โดยได้แสดงตัวอย่างของกรอบที่ระบุตำแหน่งข้อความและมาส์กจากชุดข้อมูลทดสอบดังแสดงในรูปที่ 4-1

(ก)

(ข) (ค)

รูปที่ 4-1 (ก) ตัวอย่างข้อมูลนำเข้าจากชุดข้อมูลทดสอบ BEST 2015
(ข) มาส์กของคำตอบบริเวณที่เป็นข้อความ จากชุดข้อมูลทดสอบ
(ค) มาส์กของคำตอบบริเวณที่เป็นข้อความที่ได้จากขั้นตอนวิธีที่เสนอ

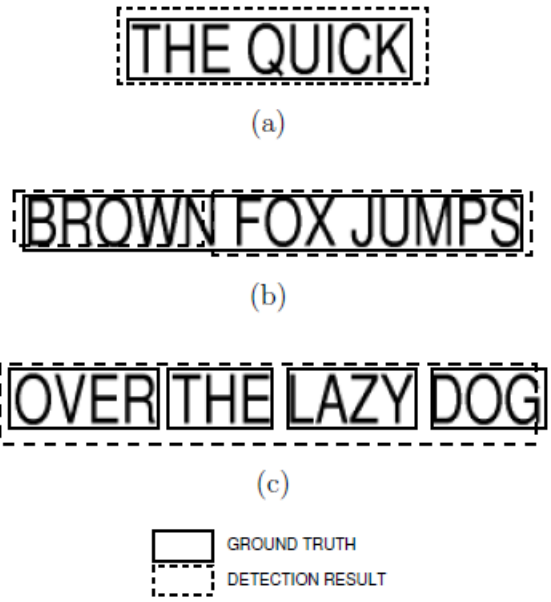
ในการวัดความถูกต้องระดับจุดภาพนั้น ได้มีการนิยามค่าสำหรับการวัดประสิทธิภาพดังต่อไปนี้

$$precision_{pixel} = \frac{\text{จำนวน } \boldsymbol{pixel} \text{ ที่ถูกต้อง}}{\text{จำนวน } \boldsymbol{pixel} \text{ ของข้อความที่ได้จากขั้นตอนวิธี}} \quad (4.7)$$

$$recall_{pixel} = \frac{\text{จำนวน } \boldsymbol{pixel} \text{ ที่ถูกต้อง}}{\text{จำนวน } \boldsymbol{pixel} \text{ ของข้อความจากชุดข้อมูลสอน}} \quad (4.8)$$

4.1.2.2 การวัดความถูกต้องระดับบล็อบ (Blob Level Detection)

ในการวัดความถูกต้องระดับบล็อบสำหรับการแข่งขัน BEST 2015 : Text Location Detection Contest นั้น จะคล้ายกับการวัดความถูกต้องด้วยการประเมินด้วยเกณฑ์ของ ICDAR ในหัวข้อที่ 4.1.1 แต่จะปรับเปลี่ยนหลักการในการจับคู่ที่ดีที่สุด เมื่อเปรียบเทียบระหว่างคำตอบของชุดข้อมูลสอนและคำตอบที่ได้จากขั้นตอนวิธีที่ออกแบบ ดังต่อไปนี้



รูปที่ 4-2 ประเภทของการจับคู่ที่ดีที่สุดจากงานวิจัยของ Wolf และ Jolion [33]

- การจับคู่แบบหนึ่งต่อหนึ่งบล็อบ (one-to-one) ดังแสดงในรูปที่ 4-2 (a)
- การจับคู่แบบหนึ่งบล็อบต่อหลายบล็อบแบบมีการแบ่ง (one-to-many : a split) ดังแสดงในรูปที่ 4-2 (b)
- การจับคู่แบบหนึ่งบล็อบต่อหลายบล็อบแบบมีการรวม (one-to-many : a merge) ดังแสดงในรูปที่ 4-2 (c)

ฟังก์ชันการจับคู่ที่ดีที่สุดได้มีการนิยามดังสมการ

$$best\_match(r,r') = \begin{cases} 1 & \text{เมื่อ } r \text{ มีการจับคู่ที่ตรงกับ } r' \text{ จากชุดข้อมูลสอนเพียง } 1 \text{ บล็อบ} \\ 0 & \text{เมื่อ } r \text{ มีการจับคู่ที่ไม่ตรงกับ } r' \text{ จากชุดข้อมูลสอน} \\ f_{sc}(k) & \text{เมื่อ } r \text{ มีการจับคู่ที่ตรงกับ } r' \text{ จากชุดข้อมูลสอนมากกว่า } 1 \text{ บล็อบ} \end{cases} \quad (4.9)$$

เมื่อกำหนดให้ r และ r' เป็นบล็อบที่ได้จากขั้นตอนวิธีที่เสนอและชุดข้อมูลสอน ตามลำดับ และ $f_{sc}(k)$ คือฟังก์ชันถ่วงน้ำหนักเมื่อมีการจับคู่มากกว่า 1 บล็อบ โดยในการแข่งขันนี้ ได้มีการนิยามฟังก์ชันดังสมการ 4.10 และกำหนดให้ k = 0.8 และการจับคู่ที่ถูกต้องนั้นจะเกิดขึ้นเมื่อมีการซ้อนทับกันของพื้นที่กรอบที่ได้จากขั้นตอนวิธีที่ทดลอง กับกรอบของคำตอบบริเวณที่เป็นข้อความมากกว่า 70%

$$f_{sc}(k) = \frac{1}{1+\ln(k)} \quad (4.10)$$

4.1.3 การประเมินผลด้วยเกณฑ์การวัดพื้นที่ซ้อนทับ

การประเมินด้วยเกณฑ์การวัดพื้นที่ซ้อนทับ จะใช้ค่าในการวัดประสิทธิภาพของ ขั้นตอนวิธีที่เสนอในลักษณะเดียวกับสมการที่ 4.1 และ 4.2 โดยจะกำหนดเงื่อนไขของ ส่วนภาพของข้อความที่ถูกต้องที่ได้จากขั้นตอนวิธี จะต้องมีการซ้อนทับกับส่วนภาพของข้อความจากชุดข้อมูลทดสอบมากกว่า 80% จึงจะนับว่าส่วนภาพนั้นถูกต้อง

4.2 ผลการทดสอบบนชุดข้อมูลทดสอบ ICDAR 2003

สำหรับการทดสอบบนชุดข้อมูลทดสอบ ICDAR 2003 นั้นจะใช้วิธีการประเมินผลตามหัวข้อ 4. 1.1 และ เนื่องจากคำตอบของชุดข้อมูลทดสอบ ICDAR 2003 นั้นจะอยู่ในรูปแบบพิกัดของสี่เหลี่ยมที่ระบุตำแหน่งของบริเวณข้อความ ระบุจุดมุมซ้ายบนของสี่เหลี่ยม ความยาวและความกว้างของสี่เหลี่ยม โดยผลที่ได้จากการทดสอบขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบด้วยเกณฑ์การประเมินของ ICDAR นั้นได้ผลดังตารางที่ 4-1 และได้ค่าความเที่ยง 0.72 และค่ารีคอล 0.63 โดยใช้การประเมินพื้นที่ซ้อนทับ

ตารางที่ 4-1 ผลการทดลองบนชุดข้อมูลทดสอบ ICDAR 2003 เมื่อเปรียบเทียบกับขั้นตอนวิธีอื่นๆ

ขั้นตอนวิธีโดย	**ความเที่ยง**	**ค่ารีคอล**	**ค่าการวัดเอฟ**
Ashida และคณะ [6]	0.55	0.46	0.5
Becker [34]	0.62	0.67	0.62
Epshtein และคณะ [4]	0.73	0.60	0.66
ขั้นตอนวิธีที่เสนอ	***0.75***	***0.68***	***0.71***

สำหรับผลการทดสอบที่ได้นำมาเปรียบเทียบประสิทธิภาพนั้น จะนำผลมาจากการงานวิจัยที่ชนะจากการประกวดอันดับ 1 – 3 จากสรุปผลจากการแข่งขัน ICDAR 2003 : Robust Reading Competition [6] โดยแต่ขั้นตอนวิธีที่นำมาทำการเปรียบเทียบนั้น มีรายละเอียดดังต่อไปนี้

Ashida และคณะ[6] ได้เสนอขั้นตอนวิธีในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติโดยวิธีการพิจารณาองค์ประกอบที่เชื่อมต่อกัน มีขั้นตอนในการทำงาน 3 ขั้นตอนคือ การจัดกลุ่มสีในปริภูมิสีแบบ LUV เพื่อทำการแยกระหว่างฉากหลังและฉากหน้า ผลที่ได้จะเป็นจุดภาพที่น่าจะเป็นตัวอักษร หลังจากนั้นจะเข้าสู่ขั้นตอนการจัดกลุ่มตัวอักษรที่ได้เข้าด้วยกัน โดยการพิจารณาคุณลักษณะสำคัญทางเรขาคณิต ตำแหน่งของแต่ตัวอักษรที่สกัดได้และสีของแต่ละตัวอักษร เพื่อทำการรวมตัวอักษรในบริบทเป็นข้อความ และหลังจากนั้นจะสกัดคุณลักษณะสหสัมพันธ์ของตำแหน่งและขนาดของแต่ละตัวอักษร ความเรียบของคอนทัวร์ (smoothness contour line) และอัตราส่วนของมุมมอง (aspect ratio)เพื่อเป็นคุณสำคัญในการจำแนกแต่ละบริเวณข้อความในขั้นตอนสุดท้าย โดยใช้ซัพพอร์ตเวคเตอร์แมชชีนเป็นตัวจำแนกประเภท ซึ่งถูกสอนโดยชุดข้อมูลสอนที่เตรียมโดย Ashida และคณะจำนวน 13,289 ส่วนภาพร่วมกับชุดข้อมูลสอน ICDAR 2003

Becker [34] ได้เสนอขั้นตอนวิธีในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติโดยวิธีการพิจารณาองค์ประกอบที่เชื่อมต่อกัน ซึ่งในงานวิธีนี้ได้ใช้เทคนิคทำไบนาไรเซชันแบบปรับตัวได้ (adaptive binarization) เพื่อแยกองค์ประกอบที่เป็นตัวอักษรออกจากฉากหลัง หลังจากนั้นจะรวมแต่ละตัวอักษรเข้าด้วยกันด้วยการพิจารณาองค์ประกอบที่เชื่อมต่อกันร่วมกับคุณลักษณะสำคัญทางเรขาคณิต

Epshtein และคณะ [4] ได้เสนอวิธีการแปลงความกว้างลายเส้น (stroke width transform) ในการระบุตำแหน่งของข้อความในภาพถ่ายฉากธรรมชาติ ภาพนำเข้าจะถูกนำไปหาแผนภาพขอบ (edge map) เพื่อนำเข้าสู่การแปลงลายเส้น ซึ่งเป็นกระบวนการที่ตัดสินใจว่า จุดภาพ (pixel) จากแผนภาพขอบที่สนใจนั้นควรจะเป็นตัวอักษรหรือไม่เมื่อเทียบกับคุณสมบัติท้องถิ่น ได้แก่ ทิศทางและความห่างกับจุดภาพที่ใกล้เคียง หลังจากนั้นการวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน จะถูกนำมาใช้เพื่อรวมผลลัพธ์จากการแปลงความกว้างลายเส้นเป็นตัวอักษร ตัวอักษรเหล่านั้นจะถูกรวมเข้ากันเป็นข้อความ โดยอาศัยคุณสมบัติของข้อความ ที่มีความกว้างลายเส้น ความกว้างของตัวอักษรและช่องว่างระหว่างตัวอักษรที่มีลักษณะคล้ายคลึงกัน แล้วจึงรวมบริเวณส่วนภาพที่น่าจะเป็นข้อความที่มีการทับซ้อนกัน โดยอาศัยคุณสมบัติของข้อความที่จะมีทิศทางในลักษณะเดียวกันเข้าด้วยกัน

จากขั้นตอนวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติที่ได้มาเปรียบเทียบบนชุดข้อมูลทดสอบ ICDAR 2003 นั้น จะมีทั้งขั้นตอนวิธีที่ต้องใช้ชุดข้อมูลสอนร่วมด้วย และขั้นตอนที่ใช้การวิเคราะห์ความเป็นข้อความเพียงอย่างเดียว ซึ่งในแต่ละขั้นตอนวิธีนั้นจะมีความแตกต่างกับขั้นตอนที่เสนอในวิทยานิพนธ์นี้ ทั้งในแง่ของลักษณะของวิธีที่ใช้และชุดข้อมูลสอนที่แตกต่างกัน

รูปที่ 4-3 ตัวอย่างผลการทดลองขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR 2003

4.3 ผลการทดสอบบนชุดข้อมูลทดสอบ ICDAR 2011

สำหรับการทดสอบบนชุดข้อมูลทดสอบ ICDAR 2011 นั้นจะทำการทดลองในลักษณะเดียวกับการทดลองบนชุดข้อมูลทดสอบ ICDAR 2003 ในหัวข้อที่ 4.2 โดยผลที่ได้จากการทดสอบขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR 2011 ด้วยเกณฑ์การประเมินของ ICDAR นั้นได้ผลดังแสดงในตารางที่ 4-2 และได้ค่าความเที่ยง 0.62 และค่ารีคอล 0.70 โดยใช้การประเมินพื้นที่ซ้อนทับ

ตารางที่ 4-2 ผลการทดลองบนชุดข้อมูลทดสอบ ICDAR 2011 เมื่อเปรียบเทียบกับขั้นตอนวิธีอื่นๆ

ขั้นตอนวิธีโดย	**ความเที่ยง**	**ค่ารีคอล**	**ค่าการวัดเอฟ**
Yi และ Tian [35]	0.71	0.67	0.62
Epshtein และคณะ [4]	0.60	0.73	0.66
Neumann และ Matas [9]	0.66	0.79	0.72
ขั้นตอนวิธีที่เสนอ	***0.66***	***0.72***	***0.69***

สำหรับผลการทดสอบที่ได้นำมาเปรียบเทียบประสิทธิภาพนั้น จะนำผลมาจากการงานวิจัยที่ชนะจากการประกวดอันดับ 1 – 3 จากสรุปผลจากการแข่งขัน ICDAR 2011 : Robust Reading Competition Challenge 2 : Reading Text in Scene Images [36] โดยแต่ขั้นตอนวิธีที่นำมาทำการเปรียบเทียบนั้น มีรายละเอียดดังต่อไปนี้

Yi และ Tian [35] ได้เสนอขั้นตอนวิธีในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติ โดยวิธีการพิจารณาองค์ประกอบที่เชื่อมต่อกัน ในขั้นตอนแรกจะสร้างแผนภาพไบนารี จากการพิจารณาคุณลักษณะสำคัญสีและเกรเดียนท์จากแผนภาพเกรเดียนท์ที่ได้จากภาพนำเข้า ซึ่งแผนภาพไบนารีที่ได้จะแสดงถึงบริเวณที่มีข้อความปรากฏอยู่ หลังจากนั้นจะทำการสร้างกรอบข้อความ โดยการพิจารณาคุณลักษณะสำคัญที่มีความเกี่ยวเนื่องกันในแต่ละข้อความ ได้แก่ ขนาด สี ระยะระหว่างตัวอักษร และการเรียงตัวของตัวอักษร

Neumann และ Matas [9] ได้ใช้การค้นหาลายเส้นเพื่อระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติ ภาพนำเข้าจะถูกสกัดลายเส้นในทิศทางต่างๆ โดยการหาภาพฉายของเกรเดียนท์และนำไปผ่านตัวกรองภาพทำให้ได้บริเวณที่น่าจะเป็นลายเส้นของตัวอักษร จากนั้นจะทำการรวมบริเวณลายเส้นเหล่านั้นที่มีการทับซ้อนกันเข้าด้วยกันเป็นบริเวณที่มีข้อความปรากฏอยู่

จากขั้นตอนวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติที่ได้นำมาเปรียบเทียบบนชุดข้อมูลทดสอบ ICDAR 2011 ซึ่งเป็นชุดข้อมูลทดสอบภาษาอังกฤษนั้น เป็นขั้นตอนวิธีการระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน ในแต่ละขั้นตอนวิธีนั้นจะมีความแตกต่างกับขั้นตอนที่เสนอในวิทยานิพนธ์นี้ ทำให้ได้ผลลัพธ์ของการระบุตำแหน่งข้อความออกมาแตกต่างกัน ซึ่งแตกต่างกับขั้นตอนวิธีที่ได้เสนอในวิทยานิพนธ์ ที่เป็นขั้นตอนวิธีระบุตำแหน่งข้อความในภาพถ่ายฉากธรรมชาติโดยพิจารณาส่วนภาพ ที่ต้องอาศัยชุดข้อมูลสอนและมีขั้นตอนวิธีในการสร้างกรอบของบริเวณที่มีข้อความที่แตกต่างกัน

รูปที่ 4-4 ตัวอย่างผลการทดลองขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR 2011

สำหรับชุดข้อมูลทดสอบ ICDAR2011 ได้มีการเปลี่ยนแปลงกฎเกณฑ์ในการระบุแหน่งข้อความของคำตอบ จากเดิมที่จะแบ่งในลักษณะของคำ (word) จะทำการแบ่งในลักษณะของประโยค คำที่อยู่ใกล้กันมากจะถูกรวมเป็นกรอบของข้อความเพียงกรอบเดียว แต่ในขั้นตอนวิธีที่ได้ออกแบบนั้นไม่ได้ทำการรวมคำที่อยู่ใกล้กันให้เป็นกรอบข้อความเดียวกัน ดังนั้นจึงได้ค่ารีคอลต่ำกว่าขั้นตอนวิธีของ Neumann และ Matas [9]

4.4 ผลการทดสอบบนชุดข้อมูลทดสอบ BEST 2015

การทดสอบขั้นตอนวิธีที่เสนอสำหรับชุดข้อมูลทดสอบ BEST 2015 นั้น จะใช้เกณฑ์การประเมินผลในหัวข้อ 4.1.2 เช่นเดียวกับในการแข่งขัน ทั้งนี้เนื่องจากข้อความที่ปรากฏในชุดภาพทดสอบสำหรับการแข่งขันนั้น มีข้อความที่ไม่ได้เรียงในแนวนอนปรากฏอยู่ด้วย โดยผลที่ได้จากการทดสอบขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบนั้น ได้ผลดังตารางที่ 4-3 และได้ค่าความเที่ยง 0.63 และค่ารีคอล 0.72 เมื่อใช้เกณฑ์การประเมินพื้นที่ซ้อนทับ

ตารางที่ 4-3 ผลการทดลองบนชุดข้อมูลทดสอบ BEST 2015 โดยขั้นตอนวิธีในการแข่งขัน

ขั้นตอนวิธีโดย	**ความเที่ยงระดับจุดภาพ**	**ค่ารีคอลระดับจุดภาพ**	**ค่าการวัดเอฟระดับจุดภาพ**	**ความเที่ยงระดับบล็อบ**	**ค่ารีคอลระดับบล็อบ**	**ค่าการวัดเอฟระดับบล็อบ**
ขั้นตอนวิธีที่ได้ลำดับที่ 8	0.25	0.67	0.30	0.12	0.72	0.18
ขั้นตอนวิธีที่ได้ลำดับที่ 7	0.51	0.51	0.47	0.08	0.6	0.13
ขั้นตอนวิธีที่ได้ลำดับที่ 6	0.40	0.47	0.35	0.28	0.39	0.29
ขั้นตอนวิธีที่ได้ลำดับที่ 5	0.25	0.75	0.31	0.28	0.52	0.33
ขั้นตอนวิธีที่ได้ลำดับที่ 4	0.38	0.57	0.42	0.59	0.30	0.35
ขั้นตอนวิธีที่ได้ลำดับที่ 3	0.45	0.57	0.45	0.45	0.46	0.35
ขั้นตอนวิธีที่ได้ลำดับที่ 2	0.44	0.74	0.51	0.38	0.65	0.44
ขั้นตอนวิธีที่เสนอ (ลำดับที่ 1)	***0.54***	***0.81***	***0.65***	***0.78***	***0.69***	***0.72***

โดยในการแข่งขัน BEST 2015 นั้นได้มีการกำหนดวิธีการประเมินผลประสิทธิภาพของขั้นตอนวิธีที่เสนอโดยใช้เกณฑ์การประเมินดังที่ได้อธิบายในหัวข้อ 4.1.2 และสามารถใช้ชุดข้อมูลสอนใดก็ได้ในการสร้างขั้นตอนวิธี ซึ่งขั้นตอนวิธีในลำดับที่ 2, 3 และลำดับที่ 5 - 8 นั้นจะเป็นขั้นตอนวิธีการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติประเภทที่พิจารณาองค์ประกอบที่เชื่อมต่อกัน โดยมีการใช้คุณลักษณะสำคัญต่างๆเช่น MSERs ลายเส้นของตัวอักษรและแผนภาพขอบ ร่วมกับการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันเพื่อสร้างบรรทัดของข้อความ สำหรับขั้นตอนวิธีในลำดับที่ 4 จะเป็นขั้นตอนวิธีการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติ ประเภทที่พิจารณาบริเวณส่วนภาพโดยใช้ขั้นตอนวิธีที่ค่อนข้างคล้ายกับวิธีที่เสนอในวิทยานิพนธ์นี้ในการหาบริเวณส่วนภาพที่เป็นข้อความ แต่ในขั้นตอนการรวมส่วนภาพที่เป็นตัวอักษรเข้าด้วยกันเป็นข้อความนั้น จะใช้ขั้นตอนวิธีไบนาไรเซชันร่วมกับการวิเคราะห์ค่าสี

รูปที่ 4-5 ตัวอย่างผลการทดลองขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ BEST2015

เนื่องจากขั้นตอนวิธีการระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติที่ได้เสนอวิทยานิพนธ์นี้ ได้มีการวิเคราะห์เพิ่มเติมสำหรับลักษณะบางประการที่มีเฉพาะในภาษาไทย ดังนั้นเพื่อเปรียบเทียบประสิทธิภาพที่แท้จริงในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติ กับวิธีอื่นๆ ที่ออกแบบมาสำหรับภาษาอังกฤษ จึงได้ทำการทดสอบขั้นตอนวิธีที่เสนอในวิทยานิพนธ์นี้ โดยไม่รวมการประมวลผลภายหลังซึ่งมีการวิเคราะห์คุณลักษณะบางประการที่มีในภาษาไทย โดยได้ผลแสดงดังตารางที่ 4-4

ตารางที่ 4-4 ผลการทดลองบนชุดข้อมูลทดสอบ BEST 2015 ระหว่างขั้นตอนวิธีที่เสนอกับขั้นตอนวิธีอื่นๆ

ขั้นตอนวิธีโดย	**ความเที่ยงระดับจุดภาพ**	**ค่ารีคอลระดับจุดภาพ**	**ค่าการวัดเอฟระดับจุดภาพ**	**ความเที่ยงระดับบล็อบ**	**ค่ารีคอลระดับบล็อบ**	**ค่าการวัดเอฟระดับบล็อบ**
Epshtein และคณะ [4]	0.57	0.5	0.54	0.55	0.59	0.57
Wang และคณะ [31]	0.61	0.57	0.59	0.59	0.61	0.60
ขั้นตอนวิธีที่เสนอ (ไม่มีการประมวลผลภายหลัง)	***0.65***	***0.60***	***0.63***	***0.74***	***0.66***	***0.7***
ขั้นตอนวิธีที่เสนอ (มีการประมวลผลภายหลัง)	***0.68***	***0.62***	***0.65***	***0.75***	***0.69***	***0.72***

4.5 ผลการทดสอบบนชุดข้อมูลทดสอบโดยผู้วิจัย

สำหรับการทดสอบบนชุดข้อมูลทดสอบโดยผู้วิจัยนั้น จะใช้เกณฑ์การประเมินประสิทธิภาพของ ICDAR 2003 และเพื่อให้ผลการทดสอบที่ได้มีความสอดคล้องกับผลการทดสอบบนชุดข้อมูลทดสอบ BEST 2015 เมื่อทำการเปรียบเทียบกับขั้นตอนวิธีอื่นๆ ที่ได้ออกแบบมาสำหรับการระบุตำแหน่งข้อความภาษาอังกฤษ จึงได้ทำการทดสอบในลักษณะเดียวกันกับการทดสอบบนชุดข้อมูลทดสอบ BEST 2015 โดยไม่รวมการประมวลผลภายหลังซึ่งมีการวิเคราะห์คุณลักษณะบางประการที่มีในภาษาไทยผล เพื่อให้ได้ผลของที่ได้จากการทดสอบขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบโดยผู้วิจัยนั้น ได้ผลดังนั้นได้แสดงในตารางที่ 4-5 และได้ค่าความเที่ยง 0.70 และค่ารีคอล 0.68 เมื่อใช้เกณฑ์ใช้การประเมินพื้นที่ซ้อนทับ

ตารางที่ 4-5 ผลการทดลองบนชุดข้อมูลทดสอบโดยผู้วิจัย

ขั้นตอนวิธีโดย	**ความเที่ยง**	**ค่ารีคอล**	**ค่าการวัดเอฟ**
Epshtein และคณะ [4]	0.68	0.65	0.66
Wang และคณะ [31]	0.72	0.65	0.68
ขั้นตอนวิธีที่เสนอ (มีการประมวลผลภายหลัง)	***0.77***	***0.71***	***0.74***
ขั้นตอนวิธีที่เสนอ (ไม่มีการประมวลผลภายหลัง)	***0.73***	***0.69***	***0.71***

รูปที่ 4-6 ตัวอย่างผลการทดลองขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบโดยผู้วิจัย

4.6 ผลการทดสอบการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติในรูปแบบต่างๆ

ขั้นตอนวิธีที่เสนอนี้ สามารถระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติที่มีการเรียงตัวในหลายรูปแบบได้ ผลที่ได้จากการทดสอบขั้นตอนวิธีที่ออกแบบบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษพบว่า สามารถระบุตำแหน่งข้อความในภาพที่มีความซับซ้อนได้ เช่น ภาพที่พื้นหลังมีความซับซ้อน, มีรูปแบบตัวอักษรที่หลากหลาย, ข้อความที่ปรากฏในภาพมีการเรียงตัวไม่อยู่ในแนวนอน, ภาพมีปรากฏการมัวจากการเคลื่อนที่ขณะถ่ายภาพ ทำให้ข้อความไม่ชัดเจน ดังแสดงในรูปที่ 4-7 – 4-10 ตามลำดับ

รูปที่ 4-7 ผลการระบุตำแหน่งข้อความบนภาพถ่ายฉากธรรมชาติที่มีความซับซ้อนของพื้นหลังสูง

รูปที่ 4-8 ผลการระบุตำแหน่งข้อความบนภาพถ่ายที่มีรูปแบบตัวอักษรข้อความที่หลากหลาย

รูปที่ 4-9 ผลการระบุตำแหน่งข้อความ
บนภาพถ่ายทั่วไปที่ข้อความที่ปรากฏในภาพไม่ได้เรียงตัวอยู่ในแนวนอน

รูปที่ 4-10 ผลการระบุตำแหน่งข้อความบนภาพถ่ายทั่วไปที่ข้อความที่ปรากฏมีปรากฏการมัว

สำหรับการทดสอบขั้นตอนวิธีที่ได้เสนอบนชุดข้อมูลทดสอบภาษาไทยนั้น พบว่าขั้นตอนวิธีที่ได้เสนอ สามารถระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติบนชุดข้อมูลทดสอบภาษาไทยได้อย่างมีประสิทธิภาพ โดยสามารถครอบคลุมสระและวรรณยุกต์ ที่อยู่นอกเหนือจากบรรทัดหลักของข้อความหลักได้เมื่อเปรียบเทียบกับวิธีการระบุตำแหน่งข้อความบนภาพถ่ายฉากธรรมชาติวิธีอื่น

(ก) (ข)

(ค) (ง)

รูปที่ 4-11 เปรียบเทียบระหว่างผลการระบุตำแหน่งที่ได้จากขั้นตอนวิธีที่เสนอกับขั้นตอนวิธีอื่นๆ
(ก) ภาพนำเข้า
(ข) ผลที่ได้จากขั้นตอนวิธีที่เสนอ
(ค) ผลที่ได้จากขั้นตอนวิธีที่เสนอโดย Neumann และ Matas [9]
(ง) ผลที่ได้จากขั้นตอนวิธีที่เสนอโดย Wang และคณะ [31]

แต่อย่างไรก็ตามขั้นตอนวิธีที่ได้เสนอนั้นยังไม่สามารถระบุตำแหน่งข้อความบนภาพถ่ายที่ข้อความมีการเปรียบต่างน้อย เมื่อเทียบกับพื้นหลัง ข้อความที่มีขนาดเล็ก และข้อความที่มีการบดบังโดยวัตถุอื่นๆ รูปที่ 4-12 – 4-14 แสดงตัวอย่างภาพที่ไม่สามารถระบุตำแหน่งข้อความได้เนื่องจากเหตุผลดังกล่าว

รูปที่ 4-12 ข้อความบนภาพถ่ายที่มีการเปรียบต่างน้อยเมื่อเทียบกับพื้นหลัง

รูปที่ 4-13 ข้อความบนภาพถ่ายที่มีขนาดเล็กเกินไป

รูปที่ 4-14 ข้อความบนภาพถ่ายที่มีการบดบัง

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้ ได้เสนอขั้นตอนวิธีในการระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติ โดยมีขั้นตอนการทำงานที่ประกอบไปด้วย 2 ขั้นตอนหลักคือ

1. ขั้นตอนการสร้างตัวตรวจจับข้อความ เพื่อสร้างตัวตรวจจับข้อความจากชุดข้อมูลสอนที่ได้ทำการจัดเตรียมไว้ โดยการสร้างตัวตรวจจับข้อความนั้นจะใช้คอนโวลูชันนอล นิวรอลเน็ตเวิร์คเป็นตัวตรวจจับข้อความ และทำการสอนบนหน่วยประมวลผลกราฟิก

2. ขั้นตอนการประมวลผลจริงเพื่อสกัดส่วนภาพที่มีข้อความ ซึ่งในขั้นตอนนี้จะนำตัวตรวจจับข้อความที่ได้จากขั้นตอนแรก มาใช้ในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาตินำเข้า โดยในขั้นตอนนี้จะประกอบไปด้วยขั้นตอนย่อย 4 ขั้นตอน ได้แก่

- ขั้นตอนการประมวลผลก่อน

ขั้นตอนนี้จะเป็นการเตรียมภาพนำเข้าให้มีความเหมาะสมก่อนนำไปหาบริเวณที่มีข้อความ โดยนำไปผ่านกระบวนการแปลงภาพระดับเทา การสร้างภาพหลายขนาดและการปรับปรุงคุณภาพของภาพนำเข้าให้มีความเหมาะสม

- ขั้นตอนการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ

ขั้นตอนนี้จะนำภาพนำเข้าหลายขนาดมาทำการหาบริเวณที่มีข้อความปรากฏอยู่ โดยใช้ตัวตรวจจับข้อความที่ได้จากขั้นตอนการสร้างตัวตรวจจับข้อความร่วมกับเทคนิค sliding window เพื่อสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความในแต่ละภาพนำเข้าหลายขนาด

- ขั้นตอนการรวมผลจากตัวตรวจจับข้อความ

แผนภาพความเชื่อมั่นของบริเวณที่มีข้อความในแต่ละภาพนำเข้าหลายขนาดที่ได้จากขั้นตอนการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ จะถูกนำมารวมเพื่อสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่มีขนาดเท่ากับภาพนำเข้าโดยใช้เทคนิค sliding window

- ขั้นตอนการประมวลผลภายหลัง

ในขั้นตอนนี้จะนำแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความที่มีขนาดเท่ากับภาพนำเข้ามาทำการสร้างสมมุติฐานกรอบของข้อความโดยผ่านขั้นตอนการกรองค่าขีดแบ่งและมอร์โฟโลยี และทำการวิเคราะห์สมมุติฐานกรอบของข้อความที่ได้ร่วมกับคุณลักษณะที่พบในภาษาไทยเพื่อสร้างกรอบของข้อความที่เหมาะสม และในขั้นตอนสุดท้ายจะทำการแก้ไขการบิดเบี้ยวแบบเพอร์สเปคทีฟเพื่อแก้ไขผลจากการถ่ายภาพที่มุมมองต่างๆ ที่ทำให้บริเวณที่มีข้อความปรากฏอยู่นั้นมีบิดเบี้ยวไป

การทดสอบขั้นตอนวิธีที่เสนอนั้น ในวิทยานิพนธ์นี้ได้ใช้วิธีการประเมินผลด้วยเกณฑ์การทดสอบ 3 รูปแบบที่แตกต่างกันได้แก่

- เกณฑ์การทดสอบของ International Conference on Document Analysis and Recognition 2003 เป็นเกณฑ์การทดสอบที่วัดประสิทธิภาพในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติในระดับบล็อบ

- เกณฑ์การทดสอบจากการแข่งขัน BEST 2015 เป็นเกณฑ์การทดสอบที่วัดประสิทธิภาพในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติในระดับบล็อบและระดับจุดภาพ โดยมีการกำหนดเงื่อนไขเพิ่มเติมในกรณีที่มีการซ้อนทับของบล็อบของข้อความเดียวกันมากกว่า 1 บล็อบ

- เกณฑ์การวัดพื้นที่ซ้อนทับ เป็นเกณฑ์การทดสอบที่วัดประสิทธิภาพในการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติในระดับบล็อบ

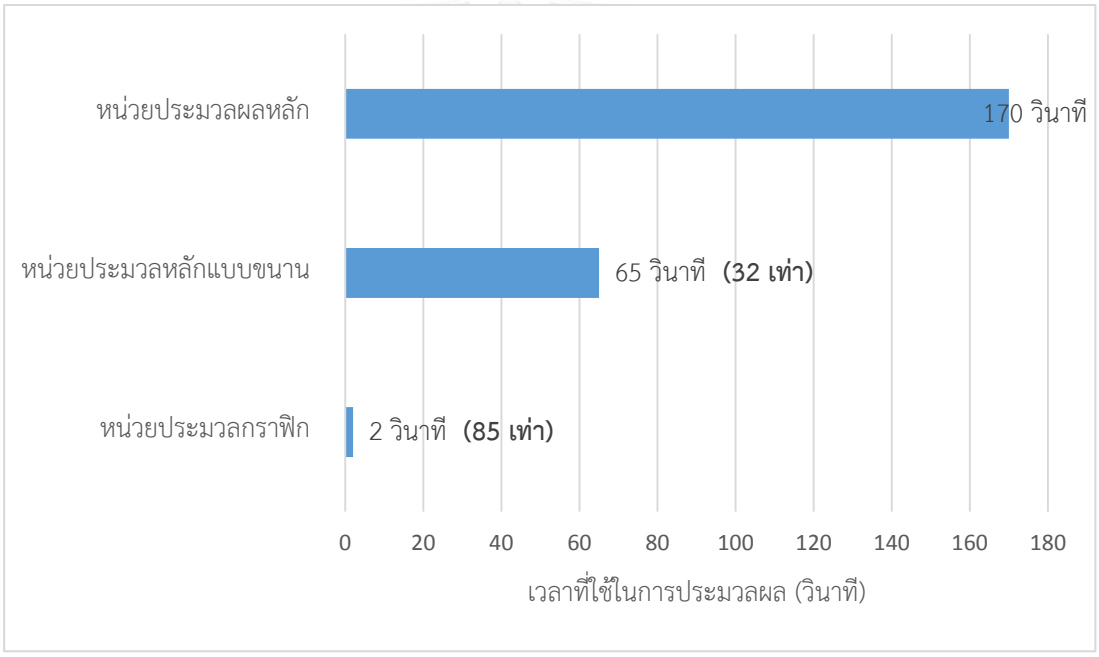
จากเกณฑ์การทดสอบดังกล่าว ในวิทยานิพนธ์นี้ได้ทำการทดสอบบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษ ICDAR 2003 และ ICDAR 2011 และเปรียบกับขั้นตอนวิธีที่ได้เสนอกับขั้นตอนวิธีการระบุตำแหน่งข้อความภาษาอังกฤษอื่นๆ โดยได้ทดสอบทั้งในกรณีที่มีและไม่มีการประมวลผลภายหลังซึ่งมีการวิเคราะห์คุณลักษณะเพิ่มเติมสำหรับภาษาไทย เพื่อเปรียบเทียบประสิทธิภาพที่แท้จริงของตัวตรวจจับข้อความที่ได้สร้างขึ้น

สำหรับการทดสอบบนชุดข้อมูลทดสอบภาษาไทย BEST 2015 และชุดข้อมูลทดสอบโดยผู้วิจัยนั้น ได้ทำการสอบในลักษณะเดียวกับการทดสอบบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษ ICDAR 2003 และ ICDAR 2011 เปรียบเทียบกับขั้นตอนวิธีอื่นๆที่เป็นขั้นตอนวิธีที่ออกแบบมาสำหรับการระบุตำแหน่งข้อความภาษาอังกฤษ

จากผลการทดลองที่ได้พบว่า ขั้นตอนวิธีที่เสนอนี้ สามารถระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติที่มีการเรียงตัวในหลายรูปแบบได้ โดยไม่ต้องกำหนดค่าเริ่มต้นในการทำงาน ผลที่ได้จากการทดสอบขั้นตอนวิธีที่ออกแบบบนชุดข้อมูลทดสอบมาตรฐานภาษาอังกฤษพบว่า สามารถระบุตำแหน่งข้อความในภาพที่มีความซับซ้อนได้ เช่น ภาพที่พื้นหลังมีความซับซ้อน, มีรูปแบบตัวอักษรที่

หลากหลาย, ข้อความที่ปรากฏในภาพมีการเรียงตัวไม่อยู่ในแนวนอน, ภาพมีปรากฏการมัวจากการเคลื่อนที่ขณะถ่ายภาพ ทำให้ข้อความไม่ชัดเจน และจากการวิเคราะห์เพิ่มเติมสำหรับคุณลักษณะที่พบในภาษาไทยนั้นทำให้สามารถระบุตำแหน่งข้อความภาษาไทยได้ดีขึ้น เมื่อเปรียบเทียบกับขั้นตอนวิธีการระบุตำแหน่งข้อความที่ออกแบบมาสำหรับภาษาอังกฤษ ในแง่ของการเก็บสระและวรรณยุกต์ที่หายไปในในวิธีอื่นๆ

เนื่องจากขั้นตอนวิธีการระบุตำแหน่งข้อความที่ได้เสนอนั้น ผู้วิจัยได้ออกแบบขั้นตอนวิธีให้ทำงานบนหน่วยประมวลผลกราฟิกเพื่อให้มีความรวดเร็วในการทำงานมากขึ้น โดยได้ทดสอบความเร็วในการทำงานบนภาพนำเข้าขนาด 640x480 จุดภาพ และได้แสดงผลเปรียบเทียบระหว่างการใช้ขั้นตอนวิธีที่เสนอบนสภาพแวดล้อมต่างๆ ดังรูปที่ 5-1



รูปที่ 5-1 เปรียบเทียบความเร็วในการประมวลผลของขั้นตอนวิธีที่เสนอบนสภาพแวดล้อมต่างๆ

จากผลการทดสอบจะเห็นได้ว่า การใช้ขั้นตอนวิธีที่ออกแบบบนสภาพแวดล้อมที่ใช้การ์ดประมวลผลกราฟิกช่วยในการประมวลผลจะช่วยเพิ่มความเร็วได้อย่างมาก ทั้งนี้มีสาเหตุเนื่องจากลักษณะงานในขั้นตอนการสร้างแผนภาพความเชื่อมั่นของบริเวณที่มีข้อความ อยู่ในลักษณะที่สามารถประมวลผลแบบขนานได้อย่างดี เนื่องจากเป็นงานที่มีขนาดเล็กแต่มีจำนวนมาก การ์ดประมวลผลกราฟิกที่มีประสิทธิภาพในการประมวลผลแบบขนานสูงจึงสามารถช่วยเพิ่มความเร็วได้เป็นอย่างดี

ในวิทยานิพนธ์นี้ ได้เสนอขั้นตอนวิธีในการระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติที่สามารถทำงานได้อย่างมีประสิทธิภาพ โดยสามารถระบุตำแหน่งข้อความได้ในหลากหลายสถานการณ์ ทั้งในภาพที่พื้นหลังมีความซับซ้อน, มีรูปแบบตัวอักษรที่หลากหลาย, ข้อความที่ปรากฏในภาพมีการเรียงตัวไม่อยู่ในแนวนอน, ภาพมีปรากฏการมัวจากการเคลื่อนที่ขณะถ่ายภาพ ทำให้ข้อความไม่ชัดเจน และสามารถเพิ่มความเร็วในการทำงานได้ โดยอาศัยการประมวลผลบนหน่วยประมวลผลกราฟิก ผลการทดสอบที่ได้เมื่อทำการทดสอบบนชุดข้อมูลทดสอบมาตราฐานภาษาอังกฤษ และชุดข้อมูลทดสอบภาษาไทยนั้นพบว่าขั้นตอนวิธีที่เสนอสามารถทำงานได้เป็นอย่างดีเมื่อเปรียบเทียบกับขั้นตอนวิธีอื่นๆ ทั้งในแง่ของความแม่นยำในการระบุตำแหน่ง และการวิเคราะห์องค์ประกอบของภาษาไทย ซึ่งขั้นตอนวิธีที่เสนอนั้นสามารถเก็บสระและวรรณยุกต์ที่ขาดหายไปในขั้นตอนวิธีการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติอื่นๆ ได้เป็นอย่างดี

เมื่อทำการเปรียบเทียบขอบเขตของวิทยานิพนธ์ที่ได้ทำการเสนอในตอนต้น กับผลที่ได้จากขั้นตอนวิธีการระบุตำแหน่งข้อความจากภาพถ่ายฉากธรรมชาติที่เสนอนั้นจะพบว่า ขั้นตอนวิธีที่เสนอสามารถนั้น สามารถตัดขอบเขตบางข้อที่เสนอในตอนต้นได้บางข้อ ซึ่งสามารถสรุปขอบเขตสุดท้ายของวิทยานิพนธ์นี้ได้ ดังต่อไปนี้

- ภาพนำเข้าต้องมีขนาดอย่างน้อย 480x320 จุดภาพ
- ข้อความที่ปรากฏในภาพมีความเปรียบต่างระหว่างฉากหลังและตัวอักษรชัดเจน ***(สามารถระบุตำแหน่งข้อความได้ในบางกรณี)***
- ข้อความที่ปรากฏในภาพต้องมีขนาดอย่างน้อย 32 จุดภาพ
- การประเมินประสิทธิภาพในการระบุตำแหน่งข้อความจะใช้วิธีการประเมินที่ใช้ในการแข่ง ICDAR 2003 Robust Reading Competition และการคำนวณพื้นที่ซ้อนทับ

5.2 ข้อเสนอแนะ

วิทยานิพนธ์นี้ สามารถนำไปปรับปรุงและพัฒนาต่อในอนาคตได้ โดยมีแนวทางดังต่อไปนี้

- การรู้จำข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติ

เนื่องจากในปัจจุบัน ยังไม่มีงานวิจัยที่สามารถรู้จำข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติได้อย่างมีประสิทธิภาพ ขั้นตอนวิธีการระบุตำแหน่งข้อความภาษาไทยจากภาพถ่ายฉากธรรมชาติที่ได้จากวิทยานิพนธ์นี้ สามารถนำไปใช้เพื่อระบุตำแหน่งข้อความได้ดีขึ้น เนื่องจากขั้นตอนวิธีที่ได้ออกแบบมีขั้นตอนวิธีที่ได้ออกแบบมาเพื่อให้เหมาะกับคุณลักษณะของภาษาไทย ซึ่งเป็นการช่วยให้ขั้นตอนวิธีการรู้จำข้อความจากภาพถ่ายฉากธรรมชาติสามารถทำงานได้อย่างมีประสิทธิภาพ

- การปรับปรุงความซับซ้อนในการประมวลผล

ขั้นตอนวิธีการระบุตำแหน่งข้อความจากภาพถ่ายฉากที่ออกแบบนั้น ยังต้องอาศัยการประมวลผลแบบขนานบนหน่วยประมวลผลกราฟิก เพื่อให้สามารถทำงานได้เร็ว ทำให้ขั้นตอนวิธีที่ออกแบบนั้นยังไม่สามารถทำงานได้เร็วเพียงพอ เมื่อทำงานบนอุปกรณ์อื่นๆที่มีประสิทธิภาพในการประมวลผลน้อย เช่น สมาร์ทโฟน (Smartphone) โดยแนวทางการพัฒนาในขั้นต่อไปนั้น อาจจะทำการระบุตำแหน่งข้อความบนภาพแบบหยาบก่อน (Course Text Detection) เพื่อหาตำแหน่งที่มีโอกาส ที่น่าจะมีข้อความปรากฏอยู่สูง แล้วจึงทำการระบุแบบตำแหน่งแบบละเอียด (Fine Text Detection) อีกครั้งหนึ่ง

รายการอ้างอิง

[1] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv e-prints,* July 2012.

[2] Q. Ye and D. Doermann, "Text Detection and Recognition in Imagery: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 37, pp. 1480-1500, 2015.

[3] K. Subramanian, P. Natarajan, M. Decerbo, and D. Castanon, "Character-Stroke Detection for Text-Localization and Extraction," In 9th International Conference on Document Analysis and Recognition, 2007. (ICDAR2007), 2007.

[4] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," In IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012), 2010.

[5] S. Karaoglu, B. Fernando, and A. Trémeau, "A Novel Algorithm for Text Detection and Localization in Natural Scene Images," In International Conference on Digital Image Computing: Techniques and Applications, 2010. (DICTA2010), 2010.

[6] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 Robust Reading Competitions," In 7th International Conference on Document Analysis and Recognition, 2003. (ICDAR2003), 2003.

[7] X. Huang and H. Ma, "Automatic Detection and Localization of Natural Scene Text in Video," In 20th International Conference on Pattern Recognition,2010 (ICPR2010), 2010.

[8] L. Neumann and J. Matas, "A Method for Text Localization and Recognition in Real-world Images," In 10th Asian Conference on Computer Vision, 2010. (ACCV2010), 2010.

[9] L. Neumann and J. Matas, "Scene Text Localization and Recognition with Oriented Stroke Detection," In IEEE International Conference on Computer Vision, 2013. (ICCV2013), 2013.

[10] X. Yin, X.-C. Yin, H.-W. Hao, and K. Iqbal, "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost," In 21st International Conference on Pattern Recognition, 2012. (ICPR2012), 2012.

[11] K. Iqbal, Y. Xu-Cheng, H. Hong-Wei, S. Asghar, and H. Ali, "Bayesian network scores based text localization in scene images," In International Joint Conference on Neural Networks, 2014. (IJCNN2014), 2014.

[12] W. Jirattitichareon and T. H. Chalidabhongse, "Automatic Detection and Segmentation of Text in Low Quality Thai Sign Images," In IEEE Asia-Pacific Conference on Circuits and Systems, 2006. (APCCAS2006), 2006.

[13] K. Woraratpanya, P. Boonchukusol, Y. Kuroki, and Y. Kato, "Improved Thai text detection from natural scenes," In International Conference on Information Technology and Electrical Engineering, 2013. (ICITEE2013), 2013.

[14] K. Woraratpanya, K. Pasupa, U. Suttapakti, P. Boonchukusol, T. Titijaroonroj, R. Hokking, *et al.*, "Text-background decomposition for thai text localization and recognition in natural scenes," In Information Technology and Electrical Engineering, 2014. (ICITEE 2014), 2014.

[15] L. Neumann and J. Matas, "Real-time scene text localization and recognition," In IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012), 2012.

[16] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," In IEEE Conference on Computer Vision and Pattern Recognition, 2004. (CVPR2004), 2004.

[17] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," In 17th International Conference on Pattern Recognition, 2004. (ICPR2004), 2004.

[18] S. M. Hanif, L. Prevost, and P. A. Negri, "A cascade detector for text detection in natural scene images," In 19th International Conference on Pattern Recognition, 2008. (ICPR2008), 2008.

[19] Y.-F. Pan, X. Hou, and C.-L. Liu, "A Robust System to Detect and Localize Texts in Natural Scene Images," In 8th IAPR International Workshop on Document Analysis Systems, 2008. (DAS2008), 2008.

[20] S. M. Hanif and L. Prevost, "Text Detection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm," In 10th International Conference on Document Analysis and Recognition, 2009. (ICDAR2009), 2009.

[21] Y.-F. Pan, X. Hou, and C.-L. Liu, "Text Localization in Natural Scene Images Based on Conditional Random Field," In 10th International Conference on Document Analysis and Recognition, 2009. (ICDAR2009), 2009.

[22] Y.-F. Pan, X. Hou, and C.-L. Liu, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images," *IEEE Transactions on Image Processing,* vol. 20, pp. 800–813, 2011 2011.

[23] J. Sochman and J. Matas, "WaldBoost - learning for time constrained sequential detection," In IEEE Conference on Computer Vision and Pattern Recognition, 2005. (CVPR2005), 2005.

[24] Y.-F. Pan, C.-L. Liu, and X. Hou, "Fast scene text localization by learning-based filtering and verification," In 17th IEEE International Conference on Image Processing, 2010. (ICIP2010), 2010.

[25] K. L. Bouman, G. Abdollahian, M. Boutin, and E. J. Delp, "A Low Complexity Sign Detection and Text Localization Method for Mobile Applications," *IEEE Transactions on Multimedia,* vol. 13, pp. 922-934, 2011 2011.

[26] Q. Meng and Y. Song, "Text Detection in Natural Scenes with Salient Region," In 10th IAPR International Workshop on Document Analysis Systems, 2012. (DAS2012), 2012.

[27] A. Mishra, K. Alahari, and C. V. Jawahar, "Top-down and bottom-up cues for scene text recognition," In IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012), 2012.

[28] B. Bo, Y. Fei, and L. Cheng Lin, "Scene Text Localization Using Gradient Local Correlation," in 12*th International Conference on Document Analysis and Recognition,* 2013. (*ICDAR*2003), 2013, pp. 1380-1384.

[29] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 679-98, Jun 1986.

[30] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, W. Tao*, et al.*, "Text Detection and Character Recognition in Scene Images with Unsupervised

Feature Learning," In 11th International Conference on Document Analysis and Recognition, 2011. (ICDAR2011), 2011.

[31] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," In 21st International Conference on Pattern Recognition,2012 (ICPR2012), 2012.

[32] Caffe | Deep Learning Framework [Online]. Available: http://caffe.berkeleyvision.org/

[33] C. Wolf and J.-M. Jolion, "Object count/Area Graphs for the Evaluation of Object Detection and Segmentation Algorithms," *International Journal on Document Analysis and Recognition,* 2006. (*IJDAR*2006*),* vol. 8, pp. 280-296, 2006.

[34] S. M. Lucas, "ICDAR 2005 text locating competition results," In 8th International Conference on Document Analysis and Recognition, 2005. (ICDAR2005), 2005.

[35] C. Yi and Y. Tian, "Text String Detection From Natural Scenes by Structure-Based Partition and Grouping," *IEEE Transactions on Image Processing,* vol. 20, pp. 2594–2605, 2011 2011.

[36] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," In 11th International Conference on Document Analysis and Recognition, 2011. (ICDAR2011), 2011.

ภาคผนวก

ภาคผนวก ก
ตัวอย่างผลการทดลองขั้นตอนวิธีที่เสนอเพิ่มเติม

รูปที่ ก-1 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR2003

รูปที่ ก-2 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR2003

รูปที่ ก-3 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR2011

รูปที่ ก-4 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ ICDAR2011

รูปที่ ก-5 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ BEST2015

รูปที่ ก-6 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบ BEST2015

รูปที่ ก-7 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบโดยผู้วิจัย

รูปที่ ก-8 ตัวอย่างผลที่ได้จากขั้นตอนวิธีที่เสนอบนชุดข้อมูลทดสอบโดยผู้วิจัย

ภาคผนวก ข
ผลงานตีพิมพ์ที่เป็นส่วนหนึ่งของวิทยานิพนธ์

- Kobchaisawat, T.; Chalidabhongse, T.H., "Thai text localization in natural scene images using Convolutional Neural Network," Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA), pp.1-7, 9-12 Dec. 2014

- Kobchaisawat, T.; Chalidabhongse, T.H., "A Method for Multi-Oriented Thai Text Localization in Natural Scene Images using Convolutional Neural Network," IEEE International Conference on Signal and Image Processing Applications (ICSIPA 2015), Oct. 2015 (To be published)

ภาคผนวก ค

ผลงานที่เป็นส่วนหนึ่งของวิทยานิพนธ์

- ได้รับรางวัลชนะเลิศ จากการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ 17 (NSC 2015) ในหัวข้อพิเศษ BEST 2015 : การแข่งขันสุดยอดการหาตำแหน่งข้อความบนภาพถ่าย (BEST 2015 : Text Location Detection Contest) ในโครงการการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย (NSC) ซึ่งจัดโดยหน่วยปฏิบัติการวิจัยเทคโนโลยีภาพ (IMG) ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

- ได้รับการพิจารณาให้ผ่านเข้ารอบชิงชนะเลิศ ในการประกวดรางวัลเจ้าฟ้าไอที โดยมูลนิธิเทคโนโลยีสารสนเทศ (Foundation for Research in Information Technology) โดยอยู่ในระหว่างการรอผลตัดสิน


Thailand
IT
Contest Festival
๒๕๕๘ • 2015
NECTEC
a member of NSTDA
สวทช
NSTDA
สสส
มหกรรมประกวดเทคโนโลยีสารสนเทศแห่งประเทศไทย ครั้งที่ ๑๔
Thailand IT Contest Festival 2015
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ
ขอมอบเกียรติบัตรนี้ ให้แก่
นายธนานพ กอบชัยสวัสดิ์
รางวัลที่ 1
จุฬาลงกรณ์มหาวิทยาลัย
โครงการ : ข้อความเอย เจ้าอยู่ไหน, ระบบระบุตำแหน่งข้อความบนภาพถ่ายทั่วไป
ประเภท BEST 2015-Text Location Detection Contest
การแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ ๑๗
(The Seventeenth National Software Contest : NSC 2015)
ให้ไว้ ณ วันศุกร์ที่ ๒๐ มีนาคม พ.ศ. ๒๕๕๘
(นายศรัณย์ สัมฤทธิ์เดชชจร)
ผู้อำนวยการ
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
(นายทวีศักดิ์ กออนันตกูล)
ผู้อำนวยการ
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

ประวัติผู้เขียนวิทยานิพนธ์

นายธนานพ กอบชัยสวัสดิ์ จบการศึกษาระดับมัธยมศึกษาจากโรงเรียนเทพศิรินทร์ และจบการศึกษาระดับปริญญาตรีจากภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2554