การตรวจจับและรู้จำอักขระภาษาไทยในป้ายโฆษณา

วิทยานิพนธ์

ของ

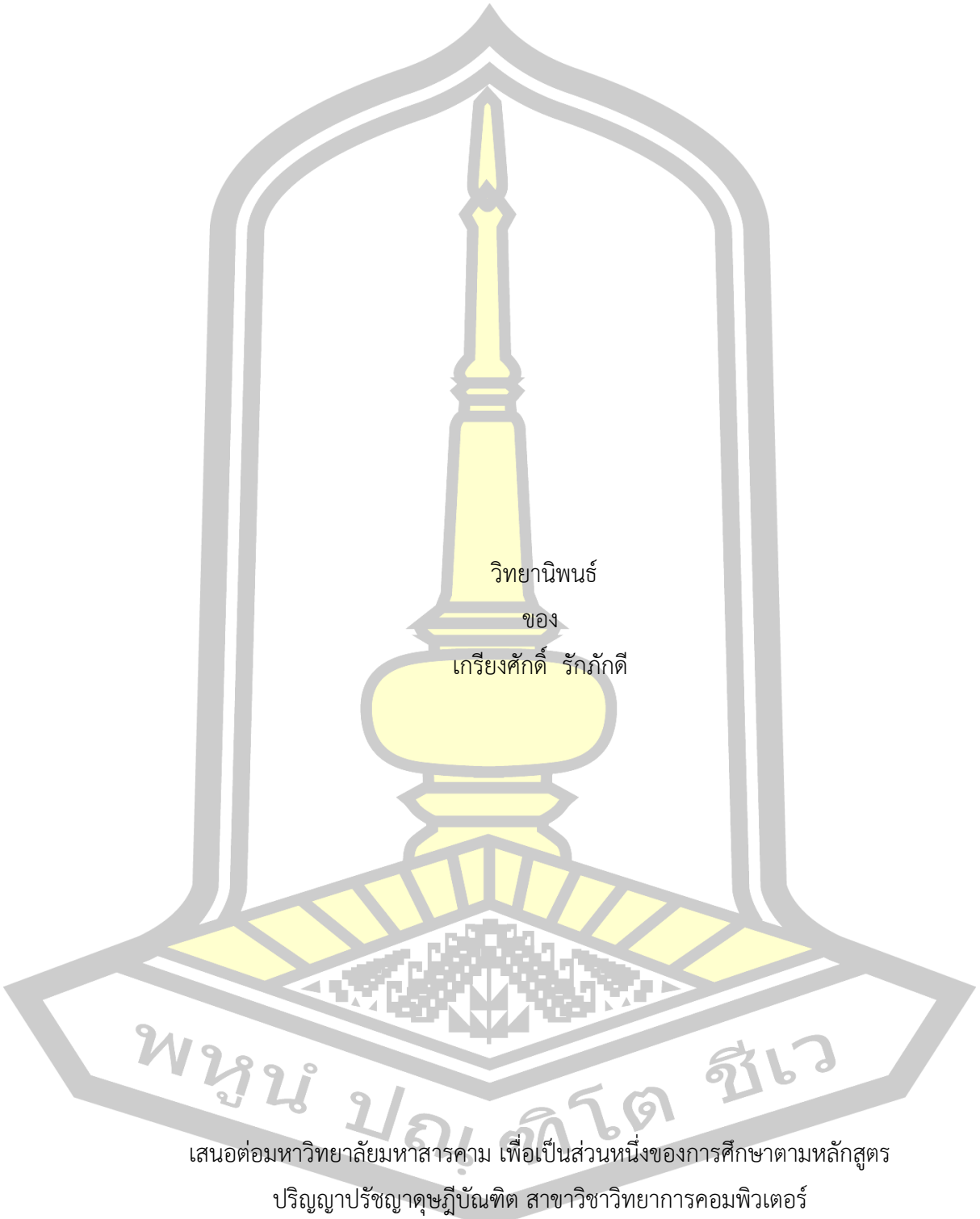
เกรียงศักดิ์ รักภักดี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

สิงหาคม 2562

การตรวจจับและรู้จำอักขระภาษาไทยในป้ายโฆษณา



วิทยานิพนธ์

ของ

เกรียงศักดิ์ รักภักดี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

สิงหาคม 2562

Thai Character Detection and Recognition in Billboard

Kriangsak Rukpukdee

A Thesis Submitted in Partial Fulfillment of Requirements

for Doctor of Philosophy (Computer Science)

August 2019

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายเกรียงศักดิ์ รักภักดี แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผศ. ดร. ธัชพงศ์ กตัญญูกุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผศ. ดร. พัฒนพงษ์ ชมภูวิเศษ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผศ. ดร. ฉัตรเกล้า เจริญผล)

.....กรรมการ
(ผศ. ดร. พนิดา ทรงรัมย์)

.....กรรมการ
(ผศ. ดร. รพีพร ช่ำชอง)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ของมหาวิทยาลัยมหาสารคาม

.....
(ผศ. ศศิธร แก้วมั่น)
คณบดีคณะวิทยาการสารสนเทศ

.....
(ผศ. ดร. กริสน์ ชัยมูล)
คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การตรวจจับและรู้จำอักขระภาษาไทยในป้ายโฆษณา
ผู้วิจัย เกรียงศักดิ์ รักภักดี
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. พัฒนพงษ์ ชมภูวิเศษ
ผู้ช่วยศาสตราจารย์ ดร. ฉัตรเกล้า เจริญผล
ปริญญา ปรัชญาดุษฎีบัณฑิต **สาขาวิชา** วิทยาการคอมพิวเตอร์
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2562

บทคัดย่อ

งานวิจัยนี้ มุ่งเน้นการพัฒนาวิธีการตรวจจับและการรู้จำอักขระภาษาไทยในป้ายโฆษณา ประกอบไปด้วย 2 กระบวนการหลัก ได้แก่ 1) การตรวจจับข้อความในภาพ งานวิจัยนี้ได้นำเสนอเทคนิค Adapted Maximally Stable Extremal Regions (AMSER) และเทคนิคการระบุตำแหน่งข้อความด้วยข้อมูลความรู้ก่อนหน้า (Prior Information) ที่อาศัยข้อมูลการเรียนรู้ก่อนหน้า เพื่อนำมาใช้ในการประมาณตำแหน่งของข้อความในภาพ และการใช้เทคนิคทางด้านการหาค่าที่เหมาะสมที่สุด (Optimization) มาใช้ในการปรับปรุงคุณภาพของการตรวจจับข้อความในภาพให้ดีขึ้น ซึ่งการวัดความแม่นยำของเทคนิคที่นำเสนอ (Precision) พบว่าวิธีการที่นำเสนอมีอัตราการตรวจจับคิดเป็นร้อยละ 86% และ 88% ตามลำดับ 2) การรู้จำข้อความในภาพ โดยภาพที่ได้จากการตรวจจับข้อความจะถูกนำมาประมวลผลเพื่อทำการแยกตัวอักขระในภาพ จากนั้นอักขระที่ได้จะถูกนำผ่านกระบวนการเรียนรู้ เพื่อให้ได้ผลลัพธ์ที่แสดงออกมาเป็นข้อความจากภาพอักขระ ซึ่งวิจัยนี้ได้มีการเปรียบเทียบกระบวนการเรียนรู้ 2 กระบวนการ ได้แก่ 1) การเรียนรู้โดยการอาศัยคุณลักษณะเด่น (Features Baesed) และ 2) การเรียนรู้ด้วยโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Networks : CNN) ซึ่งการวัดประสิทธิภาพของกระบวนการ (Accuracy) พบว่าการเรียนรู้ด้วยโครงข่ายประสาทเทียมแบบสังวัตนาการมีประสิทธิภาพมากที่สุดคิดเป็นร้อยละ 82%

คำสำคัญ : ข้อมูลความรู้ก่อนหน้า, การตรวจหาพื้นที่ข้อความในภาพ, การแบ่งส่วนข้อความ, การสกัดคุณลักษณะ, การรู้จำตัวอักขระ

TITLE Thai Character Detection and Recognition in Billboard

AUTHOR Kriangsak Rukpukdee

ADVISORS Assistant Professor Phatthanaphong Chompoowises , Ph.D.
Assistant Professor Chatklaw Jareanpon , Ph.D.

DEGREE Doctor of Philosophy **MAJOR** Computer Science

UNIVERSITY Mahasarakham University **YEAR** 2019

ABSTRACT

The research emphasizes on developing detection methods and Thai character recognition in billboards. The proposed technique consists of 2 main processes 1) Detecting text in images and 2) recognizing detected the detected texts. In the text detection process, this work applies Adapted Maximally Stable Extremal Regions (AMSER) and text positioning techniques with prior information that relies on previous learning information to estimate the location of text in images. In addition, an optimization process is carried out to improve the quality of the text detection process in images. In text recognition in images, the images obtained from the text detection and are processed to extract the characters, then the characters are fed to the learning process, in order to obtain the results of text from the character image. This research compares two learning processes, i.e. learning by using distinctive features (Features Based) and Learning by artificial neural networks (Convolutional neural networks: CNN). The evaluation results demonstrate that detecting texts achieves 86% of accuracy with non-refinement process and 88% with refinement process. Moreover, text recognition results 82% of accuracy.

Keyword : Prior Information, Text Localization, Text Segmentation, Feature Extraction, Character Recognition

กิตติกรรมประกาศ

งานวิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาระดับปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม การดำเนินการวิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยดีด้วยความอนุเคราะห์อย่างสูงยิ่งจากที่ปรึกษาหลัก ผู้ช่วยศาสตราจารย์ ดร.พัฒนพงษ์ ชมพูวิเศษ ที่ปรึกษาร่วม ผู้ช่วยศาสตราจารย์ ดร.ฉัตรเกล้า เจริญผล และคณะกรรมการผู้ทรงคุณวุฒิทั้งจากภายในและภายนอกสาขาวิชา ผู้ช่วยศาสตราจารย์ ดร.รพีพร ช่ำชอง ผู้ช่วยศาสตราจารย์ ดร.พนิดา ทรงรัมย์ และ ผู้ช่วยศาสตราจารย์ ดร.ธัชพงศ์ กัตญญูกุล ที่ได้แนะนำและให้คำปรึกษาแนวทางที่เป็นประโยชน์อย่างยิ่งต่อการแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่เป็นอย่างดีตลอดมาตั้งแต่ต้นจนสำเร็จเรียบร้อย ผู้วิจัยขอกราบขอบพระคุณไว้เป็นอย่างสูง

คุณค่าและประโยชน์จากวิทยานิพนธ์ฉบับนี้ผู้วิจัยขอมอบเป็นเครื่องบูชาคุณบิดา มารดา ที่เคารพยิ่ง ตลอดจนบูรพาจารย์และผู้มีพระคุณที่ให้การอบรมสั่งสอนให้ผู้วิจัยประสบความสำเร็จในปัจจุบันและเจริญก้าวหน้าในอนาคต

เกรียงศักดิ์ รักภักดี

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

บทที่ 1
บทนำ

1.1 หลักการและเหตุผล

ด้วยการพัฒนาเทคโนโลยีของกล้องถ่ายภาพมีอย่างต่อเนื่องและรวดเร็วไม่ว่าจะเป็นการถ่ายภาพด้วยสมาร์ทโฟน หรืออุปกรณ์พกพาอื่น ๆ ทำให้ข้อมูลภาพได้กลายมาเป็นอีกแหล่งข้อมูลหนึ่งที่มีความสำคัญและใช้งานอย่างมากกว้างขว้างนอกจากรูปภาพทั่วไป ข้อความที่มีปรากฏภายในรูปภาพถือได้ว่าเป็นส่วนประกอบที่สำคัญในการสื่อความหมายของภาพนั้น ๆ ตัวอย่างเช่น ป้ายบอกสถานที่ ป้ายชื่อร้านค้า ป้ายบอกทาง ข้อความที่ติดอยู่บนอาคาร หรือป้ายโฆษณา ฯลฯ ด้วยเหตุนี้จึงเป็นสิ่งที่สำคัญมากสำหรับมนุษย์ที่จะนำข้อมูลของข้อความที่ปรากฏภายในรูปภาพไปใช้ประโยชน์ในด้านอื่น ๆ เช่นการค้นหาข้อมูลของสถานที่จากป้ายข้อความ การอ่านข้อความที่ปรากฏอยู่ตามท้องถนนเพื่อช่วยผู้พิการทางสายตา เป็นต้น จากงานวิจัยของ Judd และคณะ [1] ได้ชี้ให้เห็นถึงความสำคัญของข้อความที่ปรากฏภายในรูปภาพที่มีผลต่อมนุษย์ โดยภายในรูปภาพที่มีข้อความและวัตถุอื่น ๆ ผู้ที่ดูรูปภาพนั้นจะมีแนวโน้มที่จะมองข้อความที่ปรากฏภายในรูปภาพมากที่สุด ซึ่งเป็นข้อพิสูจน์ที่ให้เห็นถึงความสำคัญของข้อความ ทั้งนี้วิธีการที่สามารถนำข้อความดังกล่าวมาใช้ประโยชน์ได้ก็คือ ระบบการรู้จำตัวอักษร (Optical Character Recognition) ซึ่งเป็นวิธีที่สามารถแปลภาพของข้อความจากการเขียนหรือจากการพิมพ์ ไปเป็นข้อความที่สามารถแก้ไขได้โดยเครื่องคอมพิวเตอร์

อย่างไรก็ตาม ข้อความที่ปรากฏในรูปภาพทั่วไป (Natural Scene) จะมีความแตกต่างจากข้อความในเอกสารที่ได้มาจากการสแกน (Scan) ค่อนข้างมาก เช่น ความละเอียดของภาพ สภาพแสงสว่าง ขนาดของตัวอักษร และรูปแบบของตัวอักษร ยิ่งไปกว่านั้น Shi และคณะ [2] ยังได้อธิบายถึงการสูญเสียของข้อมูลในระหว่างกระบวนการแปลงภาพให้เป็นรูปขาว-ดำ (Binarization) แทบจะไม่สามารถทำการกู้คือข้อมูลกลับมาได้ ซึ่งหมายความว่าถ้าผลของการแปลงภาพที่ไม่ดีก็จะมีโอกาสในการได้รับความถูกต้องของการรู้จำข้อความน้อยลงไปด้วย นอกจากนี้สภาพแวดล้อมในธรรมชาติก็ยังส่งผลกระทบต่อคุณภาพของภาพถ่ายโดยตรง เช่น ความคมชัด และสภาพแสง เป็นต้น ดังนั้นจึงทำให้เป็นเรื่องที่ยากที่เราจะควบคุมสิ่งแวดล้อมให้เป็นไปตามที่เราต้องการได้ซึ่งเป็นเรื่องที่ท้าทายในด้านการรู้จำเป็นอย่างมาก

ดังนั้นงานวิจัยนี้เป็นการศึกษาและการพัฒนาการรู้จำอักขระภาษาไทยโดยใช้เทคนิคข้อมูลเชิงบริบทจากภาพที่ปรากฏทั่วไป เช่น ป้ายบอกสถานที่ ป้ายชื่อร้านค้า ป้ายบอกทาง ข้อความที่ติด

อยู่บนอาคาร หรือป้ายโฆษณา ฯลฯ ซึ่งวิธีการที่นำเสนอสามารถช่วยเพิ่มประสิทธิภาพการรู้จำอักขระภาษาไทยได้อย่างถูกต้องมากยิ่งขึ้น

1.2 วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาวิธีการตรวจจับและการรู้จำอักขระภาษาไทยในป้ายโฆษณา

1.3 ความสำคัญของการวิจัย

งานวิจัยนี้ได้ดำเนินการเพื่อการแก้ไขปัญหาของการตรวจจับและการรู้จำอักขระภาษาไทยในป้ายโฆษณา โดยอาศัย การตรวจหาพื้นที่ข้อความในภาพ การแบ่งส่วนข้อความ การวิเคราะห์รูปแบบตัวอักษร การสกัดคุณลักษณะ และการรู้จำตัวอักขระ เพื่อให้เกิดประสิทธิภาพและความถูกต้องมากที่สุดและนำไปประยุกต์การใช้งานในโปรแกรมประยุกต์ที่ต้องการรู้จำข้อความ ตัวอย่างเช่น โปรแกรมประยุกต์เทคโนโลยีเสมือนจริง

1.4 ขอบเขตของการวิจัย

การดำเนินการวิจัยในครั้งนี้มีขอบเขตของการวิจัยดังนี้

1.4.1 งานวิจัยนี้จะมุ่งเน้นการพัฒนาวิธีการแก้ไขปัญหาในการตรวจจับและการรู้จำอักขระภาษาไทยในป้ายโฆษณา เช่น ปัญหาการตรวจหาพื้นที่ข้อความในภาพ ปัญหาการแบ่งส่วนข้อความ ปัญหาการวิเคราะห์รูปแบบตัวอักษร เป็นต้น

1.4.2 ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ได้แก่ ภาพถ่ายด้ายกล้องดิจิตอล (Digital Camera) และอุปกรณ์พกพา (Smartphone) ร่วมทั้งการรวบรวมภาพจากเว็บไซต์ www.google.com โดยมีการปรับขนาดภาพให้มีขนาดอยู่ที่ 480x360 พิกเซล และฐานข้อมูลภาพมาตรฐาน โดยจำแนกภาพออกเป็น 2 กลุ่มได้แก่

1) กลุ่มภาพแบบทั่วไป คือภาพป้ายร้าน ป้ายไวนิล ฯลฯ ที่มีองค์ประกอบอื่น ๆ มาปะปนในภาพ

2) กลุ่มภาพแบบเจาะจง คือภาพที่เลือกเฉพาะที่มีการมองเห็นตัวอักขระบนป้ายร้าน ป้ายไวนิล ฯลฯ ที่ชัดเจนมีองค์ประกอบอื่น ๆ ปะปนน้อย

14.3 ขนาดของตัวอักขระภาษาไทยในป้ายโฆษณาจะมีขนาดไม่เล็กจนเกินไป ตัวอย่างเช่นข้อความ "เทศบาลนครอุดรธานี" แสดงในภาพที่ 1.1

รูปที่ 1 ตัวอักษรที่เล็กจนเกินไป

1.5 นิยามศัพท์เฉพาะ

1.5.1 คอมพิวเตอร์วิทัศน์ (Computer Vision) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ โดยจุดประสงค์หลักของคอมพิวเตอร์วิทัศน์ คือการทำให้คอมพิวเตอร์สามารถเข้าใจทิวทัศน์ หรือคุณลักษณะต่าง ๆ ในภาพได้

1.5.2 การแปลงภาพให้เป็นรูปขาว -ดำ (Binarization) คือการแปลงรูปภาพให้เป็นรูปภาพขาว-ดำ โดยแต่ละจุดของภาพสามารถมีค่าที่เป็นไปได้เพียง 2 ค่าคือค่าสีขาวแทนด้วย 1 และค่าสีดำแทนค่าด้วย 0 บางครั้งอาจเรียก ภาพสองระบบ (Bi-Level) ภาพขาวดำ (Black-and-White) หรือภาพโมโนโครม (Monochome)

1.5.3 รูปภาพทั่วไป (Natural Scene) คือ ภาพรวมของพื้นที่ใดพื้นที่หนึ่ง ที่มนุษย์ รับรู้ทางสายตาในระยะห่าง อาจเป็นพื้นที่ธรรมชาติที่ประกอบด้วย แผ่นดิน น้ำ ต้นไม้ สัตว์ และสรรพสิ่งมนุษย์สร้างในสภาพอากาศหนึ่งและช่วงเวลาหนึ่ง หรือภาพรวมของเมืองหรือส่วนของเมือง

1.5.4 การตรวจจับข้อความ (Text Detection) คือ การตรวจหาพื้นที่ข้อความในภาพ โดยเป็นกระบวนการที่มีหน้าที่ในการตรวจหาส่วนที่คาดการณ์ว่าจะเป็นข้อความภายในภาพ

1.5.5 การรู้จำอักขระ (Character Recognition) คือ การสอนให้คอมพิวเตอร์รู้จักรูปแบบหรือลักษณะเฉพาะของตัวอักขระแต่ละตัว จากภาพของข้อความที่ได้จากการเขียนหรือจากการพิมพ์

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

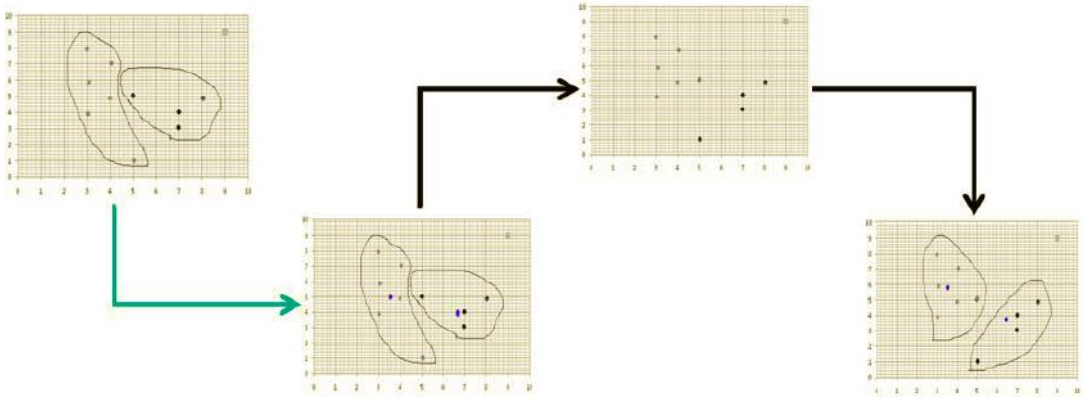
การพัฒนาอัลกอริทึมในการรู้จำอักขระภาษาไทยโดยใช้เทคนิคข้อมูลเชิงบริบท จะเกี่ยวข้องกับการตรวจหาพื้นที่ข้อความในภาพ การแบ่งส่วนภาพ และการรู้จำตัวอักขระ ซึ่งกระบวนการต่าง ๆ เหล่านี้ต้องอาศัยทฤษฎีและความรู้พื้นฐานต่าง ๆ รวมถึงเทคนิคต่าง ๆ ที่ในหลาย ๆ งานวิจัยได้นำเสนอไว้ ดังนั้นเพื่อให้การดำเนินการวิจัยในครั้งนี้สามารถบรรลุผลตามวัตถุประสงค์ที่ตั้งไว้ ซึ่งทฤษฎีและงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้จะมีรายละเอียดดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

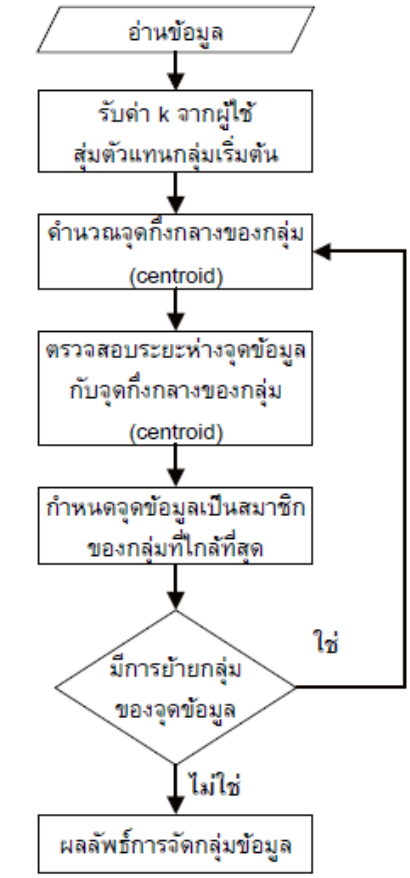
2.1.1 การเรียนรู้ของเครื่องจักร (Machine Learning)

2.1.1.1 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ใช้ข้อมูลฝึกหรือชุดตัวอย่างที่ไม่มีการใส่ฉลากให้กับข้อมูล และเรียนรู้โดยการนำข้อมูลไปผ่านกระบวนการหาความคล้ายคลึงของตัวอย่าง จนกระทั่งได้กลุ่มตัวอย่างที่จัดเป็นประเภทอย่างเหมาะสม เทคนิคประเภทนี้ได้แก่ การแบ่งกลุ่ม (clustering)

1) เทคนิคเคมีน (K-Means) [3] เป็นอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูลตามลักษณะความคล้ายคลึงกันของข้อมูล ซึ่งจะมีการแบ่งข้อมูลออกเป็น K กลุ่ม โดยจะแบ่งตามคุณสมบัติหรือคุณลักษณะประจำที่มีในแต่ละเรคคอร์ด ซึ่งจะไม่มีการระบุลักษณะของกลุ่ม แต่จะพิจารณาจากความคล้ายคลึงกันด้วยการวัดระยะทางระหว่างจุดของข้อมูลกับจุดกึ่งกลาง (Centroid) ของกลุ่มข้อมูลนั้น ๆ เรคคอร์ดที่มีความคล้ายคลึงกันจะมีระยะห่างข้อมูลน้อย และในขณะที่เรคคอร์ดที่มีความแตกต่างกันจะมีระยะห่างของข้อมูลมากกว่าดังแสดงดังรูปที่ 2



รูปที่ 2 แสดงขั้นตอนการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมเคมีน
ที่มา [3]



รูปที่ 3 ขั้นตอนการทำงานการจัดกลุ่มด้วยอัลกอริทึมเคมีน
ที่มา [3]

ขั้นตอนการทำงาน การจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมเคมีน [4]

1) ทำการรับค่า k จากผู้ใช้ และสุ่มตัวแทนของกลุ่มเริ่มต้น คำนวณจุดกึ่งกลาง (Centroid) ของกลุ่มโดยการใช้ค่าเฉลี่ยเลขคณิต (M_k วิธีคำนวณดังสมการที่ 1)

2) ตรวจสอบระยะห่างระหว่างจุดข้อมูลกับจุดกึ่งกลางของกลุ่ม ถ้าระยะห่างระหว่างจุดข้อมูลกับจุดกึ่งกลางของกลุ่มข้อมูลใดน้อยที่สุด แสดงว่ามีความคล้ายคลึงกันมากที่สุด แล้วจึงกำหนดให้จุดของข้อมูลนั้นเป็นสมาชิกของกลุ่มที่มีความคล้ายคลึงกันมากที่สุด

3) ถ้ามีการย้ายกลุ่มของจุดข้อมูล จะต้องมีการคำนวณหาจุดกึ่งกลาง (Centroid) ของแต่ละกลุ่มข้อมูลใหม่ (M_k)

4) วนทำซ้ำตามขั้นตอนที่ 2 และ 3 ไปเรื่อย ๆ จนกระทั่งไม่มีการย้ายกลุ่มของจุดข้อมูล ซึ่งจะได้กลุ่มของข้อมูลทีมีความคล้ายคลึงกันตามจำนวน k กลุ่ม

การหาจุดกึ่งกลางของกลุ่มแสดงดังสมการที่ 2.1

$$M_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ik} \qquad (2.1)$$

โดยที่ M = จุดกึ่งกลางในแต่ละ Cluster

2) การจัดกลุ่มแบบฟัซซี่ (Fuzzy C-means Clustering) [5] เป็นอัลกอริทึมที่จะยอมให้ข้อมูลที่มีอยู่ในแต่ละคลัสเตอร์สามารถซ้อนทับกันหรือซ้ำกันได้ ซึ่งจะอาศัยการกำหนดค่าการเป็นสมาชิกของข้อมูลต่อกลุ่มข้อมูลต่าง ๆ โดยการได้มาซึ่งค่าการเป็นสมาชิกนั้นส่วนหนึ่งจะมาจากการวัดระยะทางระหว่างข้อมูลและจุดกึ่งกลางของกลุ่มนั้น ๆ ซึ่งการวัดระยะทางจะมีความสำคัญเป็นอย่างมากต่อการจัดกลุ่ม ทั้งนี้วิธีการวัดระยะทางนั้นจะมีอยู่ด้วยกันหลายวิธีเช่น การวัดระยะทางแบบยูคลิเดียน (Eulidence Distance) ซึ่งการวัดระยะทางแบบยูคลิเดียนนั้นจะไม่เหมาะสมกับข้อมูลที่มีความเกี่ยวเนื่องกัน และการวัดระยะทางแบบมหาลาโนบิส (Mahalanobis Distance) นั้นจะมีความเหมาะสมกับกลุ่มข้อมูลที่มีข้อมูลโดดออกจากกลุ่ม (Outlier) รวมทั้งกลุ่มข้อมูลที่มีข้อมูลหนาแน่นต่าง ๆ เทคนิคการจัดกลุ่มของฟัซซี่จะเป็นเทคนิคที่แก้ไขข้อเสียของ K-mean ซึ่ง K-mean ไม่เหมาะกับข้อมูลที่มีความสัมพันธ์กัน (Correlation) เนื่องจากข้อมูลจะสามารถเป็นสมาชิกได้เพียงกลุ่มเดียวเท่านั้น แต่การจัดกลุ่มแบบฟัซซี่นั้น สามาชิกของกลุ่มจะมีโอกาสหรือมีค่าที่จะเป็นสามาชิกของข้อมูลระดับต่าง ๆ ในทุก ๆกลุ่ม

ขั้นตอนการทำงานของฟัซซี่มีมีน (Fuzzy C-Means) จะประกอบด้วยขั้นตอนดังต่อไปนี้

1) กำหนดกลุ่มข้อมูลที่ต้องการจัดกลุ่ม เพื่อกำหนดค่าสำหรับใช้เป็นเงื่อนไขในการให้ข้อมูลหยุดการจัดกลุ่ม (ε) กำหนดค่าฟัซซี่พารามิเตอร์ (m) ซึ่งต้องมีค่ามากกว่าหนึ่ง และกำหนดจุดกึ่งกลางการเริ่มต้นของข้อมูล

2) คำนวณหาค่าการเป็นสมาชิกของข้อมูลต่อกลุ่มข้อมูลต่าง ๆ

3) คำนวณจุดกึ่งกลางของกลุ่มข้อมูลใหม่ และตรวจสอบเงื่อนไขโดยตรวจสอบค่าการเป็นสมาชิกใหม่พร้อมกับลบค่าการเป็นสมาชิกก่อนหน้านี้

4) ถ้าเงื่อนไขเป็นจริงให้คำนวณค่าการเป็นสมาชิกและ Objective Function ถ้าเงื่อนไขเป็นเท็จ ให้คำนวณค่าการเป็นสมาชิกจากจุดศูนย์กลางล่าสุด (วนรอบ)

การคำนวณ Objective Function สามารถคำนวณได้จากสมการที่ 2.2

$$J = \sum_{i=j}^{c} \sum_{j=1}^{n} (\mu ij)^{m} d^{2}(X_{j}, Z_{i}) \qquad (2.2)$$

โดยที่ J แทน Objective Function ของขั้นตอนวิธีฟัซซี่ซีมีน

กำหนดให้เซตของข้อมูล $X = \{X_1 X_2, ..., X_n\}$

n = คือจำนวนข้อมูล

c = คือจำนวนกลุ่มข้อมูล

m = คือฟัซซี่พารามิเตอร์ที่ต้องมีค่ามากกว่า 1

μij คือค่าการเป็นสมาชิก (Membership) ของข้อมูลที่ J ในกลุ่มที่ i

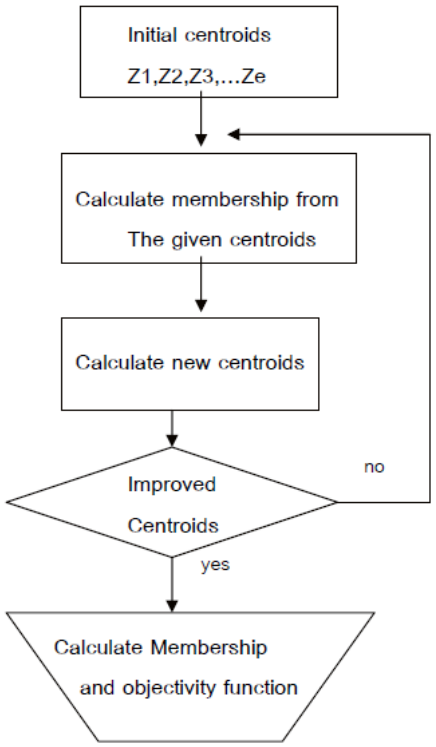
$d^{2}(X_{j}, Z_{i})$ = คือระยะทางยกกำลังสองระหว่างข้อมูล X ที่ j และจุดกึ่งกลางของข้อมูล Z กลุ่มที่ i โดย

$$Z_{i} = \frac{\sum_{j=1}^{n} (\mu ij)^{m} X_{j}}{\sum_{j=1}^{n} (\mu ij)^{m}}$$

การหาค่าการเป็นสมาชิก μij แสดงดังสมการที่ 2.3

$$\mu ij = \frac{\left[1 / d^{2}\left(X_{j} - Z_{i}\right)\right]^{1/(m-1)}}{\sum_{i=1}^{c}\left[1 / d^{2}\left(X_{j} - Z_{i}\right)\right]^{1/(m-1)}} \quad (2.3)$$

รายละเอียดการทำงานของฟัซซี่ซีมีนมีการทำงานดังนี้



รูปที่ 4 ขั้นตอนการทำงานการจัดกลุ่มด้วยอัลกอริทึมฟัซซี่ซีมีน

สำหรับการวัดระยะทางระหว่างข้อมูลและจุดกึ่งกลางของข้อมูล แบบยูคลิเดียน (Euclidean Distance) สามารถหาได้จากสมการที่ 2.4

$$ED_{ji} = \sqrt{\left(X_{j} - Z_{i}\right)\left(X_{j} - Z_{i}\right)^{T}} \quad (2.4)$$

โดย ED_{ji} แทนระยะทางแบบยูคลิเดียนระหว่างข้อมูล X ที่ j และจุกกึ่งกลางของ ข้อมูล Z กลุ่มที่ i และ T แทน Transpose Matrix

สำหรับการวัดระยะทางแบบมหาลาโนบิส (Mahalanobis Distance) นั้นเหมาะกับข้อมูลที่มีความสัมพันธ์ต่อกัน สามารถหาค่าได้จากสมการที่ 2.5

$$MD_{ji} = \sqrt{\left(X_j - Z_i\right) A^{-1} \left(X_j - Z_i\right)^T} \tag{2.5}$$

โดย MD_{ji} แทนระยะทางแบบมหาลาโนบิสระหว่างข้อมูล X ตัวที่ j และจุดศูนย์กลางข้อมูล Z กลุ่มที่ i

A คือ Variance-Covariance Matrix คำนวณจากสมการที่ 2.6

$$A = \frac{\sum_{j=1}^{n} \left(X_j - Z_i\right)^T \left(X_j - Z_i\right)}{n-1} \tag{2.6}$$

2.1.1.2 การเรียนรู้แบบมีผู้สอน (Supervised Learning) เทคนิคการเรียนรู้ของเครื่องจักรประเภทนี้ต้องใช้การเรียนรู้จากข้อมูลฝึกที่มีการใส่ฉลาก (Label) ให้กับข้อมูลฝึกไว้แล้ว เพื่อให้คอมพิวเตอร์เข้าใจรูปแบบและได้สมมติฐานเพื่อทำงานกับข้อมูลในภายหน้าได้ ตัวอย่างเทคนิคประเภทนี้ได้แก่ การเรียนรู้แบบตัวจำแนกแบบเบย์อย่างง่าย และการเรียนรู้แบบต้นไม้ตัดสินใจ เป็นต้น

1) วิธีการเพื่อนบ้านใกล้ที่สุด หรือ K-Nearest Neighbor: KNN [6] เป็นวิธีการที่คำนวณระยะห่างระหว่างข้อมูลใหม่ที่กำลังสนใจกับฐานข้อมูลเดิมที่มีการบันทึกไว้ โดยจะใช้เทคนิคยูคลิเดียน (Euclidean) จากนั้นจะเลือกข้อมูลที่มีค่าใกล้ที่สุด K อันดับแล้วทำการพิจารณาว่าข้อมูลใหม่ดังกล่าวมาว่าเป็นชนิดใด ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 กำหนดให้ L เป็นจำนวนประเภทที่เป็นไปได้ D เป็นชุดข้อมูลซึ่งประกอบไปด้วยข้อมูลและประเภทของข้อมูล

$$D = \{(v_0, c_0), (v_1, c_1), (v_2, c_2), \ldots, (v_n, c_n)\} \tag{2.7}$$

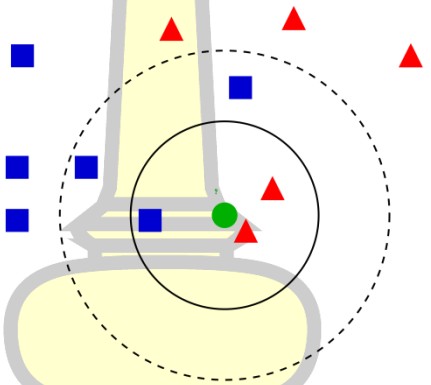
โดยที่ $v_i = \{v_{i_1}, v_{i_2}, v_{i_3}, \ldots, v_{i_n}\}$ เป็นเวคเตอร์ของข้อมูลและ $c_i \in \{1, 2, \ldots, L\}$ เป็นประเภทของข้อมูล v_i

ขั้นตอนที่ 2 ให้ $x = \{x_1, x_2, x_3, ..., x_n\}$ เป็นเวคเตอร์ข้อมูลใหม่ที่ยังไม่ทราบประเภท

ขั้นตอนที่ 3 สำหรับทุก ๆ v_i ใน D สามารถนำมาคำนวณหาค่าระยะห่างระหว่าง x และ v_i โดยใช้วิธียูคลิเดียนดังสมาการที่ 2.8

$$d(v_i, x) = \sqrt{\sum_{j=1}^{n} (v_{i_j} - x_j)^2} \quad (2.8)$$

ขั้นตอนที่ 4 กำหนดให้ประเภทของ $\boldsymbol{x}$ จากประเภทของเวคเตอร์ $\boldsymbol{v_i}$ ที่มีระยะห่างระหว่าง $\boldsymbol{x}$ น้อยที่สุด $\boldsymbol{k}$ อันดับแรกโดยเลือกประเภทที่พบบ่อยที่สุด

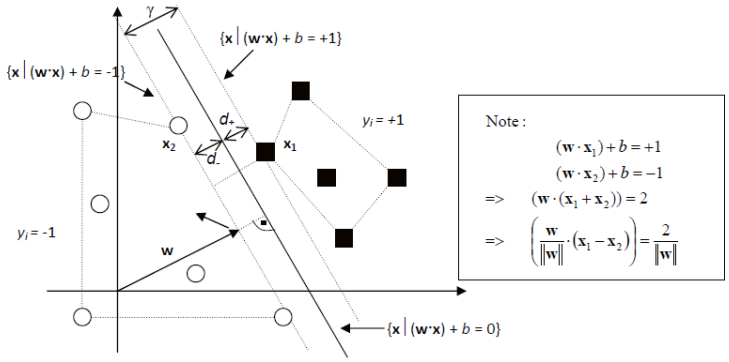


รูปที่ 5 แสดงการจำแนกประเภทด้วยวิธีเพื่อนบ้านใกล้ที่สุด

จากรูปที่ 5 แสดงถึงการจำแนกประเภทแบบ $K - NN$ โดยพิจารณาประเภทของวงกลมเมื่อค่า K มีค่าเท่ากับ 3 จะทำนายว่าวงกลมนั้นมีชนิดเป็นสามเหลี่ยม แต่ในขณะที่เพิ่มค่า K ให้มีค่าเท่ากับ 5 ทำให้จำแนกเป็นประเภทสี่เหลี่ยมโดยค่า K ที่เปลี่ยนไปทำให้ผลการจำแนกประเภทเปลี่ยนแปลงไปด้วย

2) ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine: SVM) [7] [8] เป็นเทคนิคหนึ่งที่ได้รับความนิยมเป็นอย่างแพร่หลายในงานที่มีความเกี่ยวข้องกับการจดจำรูปแบบตลอดจนการแก้ปัญหาการจัดกลุ่ม (Classification Problem) โดยหลักการของซัพพอร์ตเวคเตอร์แมชชีน คือการสร้างไฮเปอร์เพลนที่เหมาะสมบนระนาบของข้อมูลตัวอย่าง (Training Data) เพื่อทำการแบ่งกลุ่มข้อมูลที่มีความแตกต่างกัน ซึ่งการสร้างไฮเปอร์เพลนที่เหมาะสมจะนิยมกำหนดระยะห่าง

ระหว่างจุดของข้อมูลที่อยู่ใกล้กับไฮเปอร์เพลนมากที่สุดทั้งสองด้านคือ d_+ และ d_- ระยะมาร์จิ้น (Margin) γ เกิดจากระยะ $d_+ + d_-$ ทั้งนี้ไฮเปอร์เพลนที่เหมาะสมคือไฮเปอร์เพลนที่มีค่ามาร์จิ้น γ มีความกว้างมากที่สุดแสดงดังรูปที่ 6 ข้อมูลตัวอย่างที่อยู่บนขอบของมาร์จิ้น γ จะถูกเรียกว่าซัพพอร์ตเวคเตอร์ (Support Vector)



รูปที่ 6 การแบ่งกลุ่มข้อมูลตัวอย่างด้วยไฮเปอร์เพลนโดยใช้เทคนิค SVM ที่มา [9]

จากรูปที่ 6 เป็นการแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่มโดยกำหนดให้กลุ่มข้อมูลที่ใช้ในการฝึกสอน (Training Dataset) จะประกอบด้วย l ตัวอย่าง (Samples) ซึ่งจะแสดงอยู่ในรูปสมการที่ 2.9

$$\{x_k, y_k\}, k = 1, \ldots, l \qquad (2.9)$$

และ $x_k \in \Re^n, \quad y_k \in \{-1, +1\}$

โดยที่ x_k จะเป็นอินพุทเวคเตอร์

y_k เป็นคลาสของข้อมูล (Class Label)

โดยหลักการของซัพพอร์ตเวคเตอร์แมชชีน คือการสร้างโฮเปอร์เพลนที่เหมาะสมบนระนาบของข้อมูลตัวอย่าง ซึ่งไฮเปอร์เพลนดังกล่าวจะถูกกำหนดโดยพารามิเตอร์ (w, b) ดังแสดงในรูปที่ 6 โดยที่ w เป็นเวคเตอร์ที่จะตั้งฉากกับโฮเปอร์เพลนและ b จะเป็นค่าคงที่ซึ่งกำหนดตำแหน่งของเวคเตอร์ที่สัมพันธ์กับตำแหน่งดั้งเดิมในปริภูมิอินพุท (Input Space) โดยสมการของโฮเปอร์เพลนแบบเชิงเส้น

จะถูกกำหนดด้วยสมการ $(w \cdot x) + b = 0$ และเพื่อลดปัญหาในเรื่องของสเกล w และ b จะถูกกำหนดด้วยสมการ $\left|(w \cdot x) + b\right| = 1$ สำหรับจุดที่อยู่ใกล้ไฮเปอร์เพลนมากที่สุด ดังนั้นจะสามารถแสดงสมการของไฮเปอร์เพลนได้ดังสมการที่ 2.10

$$y_i \left[(w \cdot x_i) + b\right] \geq 1 \qquad \forall\, i \tag{2.10}$$

จากที่กล่าวมาข้างต้น เป็นการแบ่งกลุ่มข้อมูลด้วยไฮเปอร์เพลนแบบเชิงเส้นเท่านั้น ดังนั้นเพื่อให้สามารถแบ่งแยกกลุ่มข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (Nonlinear Dataset) จำเป็นที่จะต้องแปลงกลุ่มของข้อมูลตัวอย่างไปสู่ปริภูมิมิติที่สูงขึ้น (Higher Dimensional Space) ซึ่งจะถูกเรียกว่าปริภูมิฟีเจอร์ (Feature Space) ทั้งนี้การแปลงดังกล่าวจะถูกดำเนินการผ่านฟังก์ชั่นที่ไม่เป็นเชิงเส้น โดยสามารถแสดงสมการที่นำมาใช้ในการคำนวณค่าโฮเปอร์เพลนเพื่อแบ่งกลุ่มข้อมูลที่มีลักษณะที่ไม่เป็นเชิงเส้นได้ดังสมการที่ 2.11 [10]

- Maximize

$$w(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2.11}$$

- Subject to (1) $\sum_{i=1}^{l} \alpha_i y_i = 0$, and

(2) $0 \leq \alpha_i \leq C \qquad \forall\, i$ (2.12)

โดยตัวแปร $\alpha_i \geq 0$ จะถูกเรียกว่า Positive Lagrange Multipliers, $K(x_i, x_j)$ คือฟังก์ชั่นเคอร์แนล และ C จะเป็นค่าคงที่เพื่อใช้ในการปรับหรือชดเชยค่าที่ผิดพลาดของการฝึกสอนและความซับซ้อนของแบบจำลอง (Model Complexity) ทั้งนี้จากสมการที่ 12 จะสามารถแสดงฟังก์ชั่นเคอร์แนลที่นิยมใช้โดยทั่วไปตามสมการที่ 2.13 - 2.15

- โพลิโนเมียลดีกรี d (Polynomial of Degree d)

$$K(x, y) = (\gamma(x \cdot y) + \beta)^d \tag{2.13}$$

- Radial Basis Function (RBF)

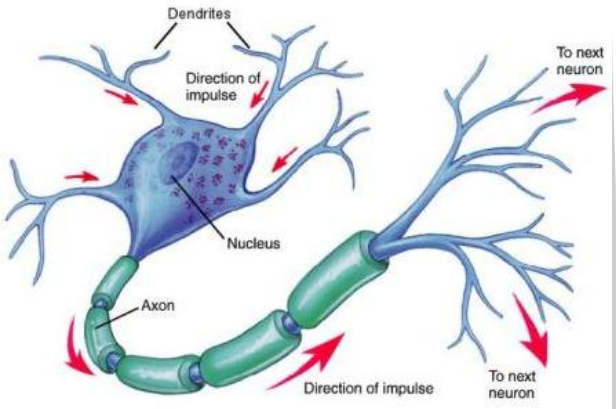
$$K(x, y) = \exp(-\gamma \left\| x - y \right\|^2) \tag{2.14}$$

-Sigmoid Function

$$K(x, y) = \tanh(\gamma(x \cdot y) + \beta) \tag{2.15}$$

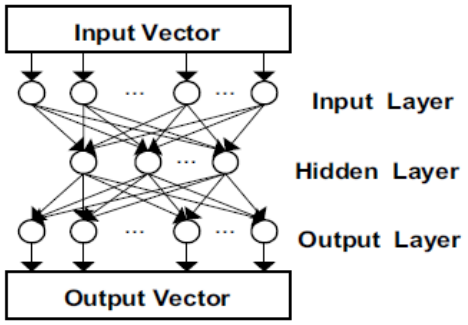
โดย γ, β และ d คือพารามิเตอร์ของเคอร์แนล (Kernel Parameters)

3) โครงข่ายประสาทเทียม (Neural Network) [11] เป็นการจำลองการทำงานบางส่วนของสมองมนุษย์ ซึ่งประกอบด้วยเซลล์ประสาท (Neuron) เป็นจำนวนมาก โดยในแต่ละเซลล์จะประกอบไปด้วยนิวเคลียส (Nucleus) ตัวเซลล์ (Cell Body) ใยประสาทนำเข้า (Dendrite) และแกนประสาทนำออก (Axon) ดังรูปที่ 7



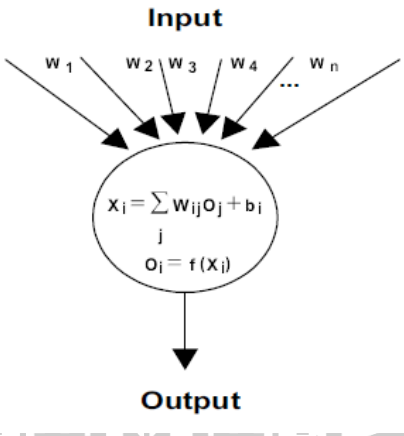
รูปที่ 7 โครงข่ายประสาทของสมองมนุษย์ ที่มา [12]

ทั้งนี้โครงข่ายประสาทเทียม [13] จะประกอบขึ้นจากเซลล์ประสาทจำนวนมากที่มีการจัดเรียงตัวกันอยู่เป็นชั้น ๆ ภายในโครงข่ายประสาท ซึ่งได้แก่ ชั้นอินพุท (Input Layer) ชั้นเอาท์พุท (Output) และชั้นซ่อน (Hidden Layer) แสดงดังรูปที่ 8 โดยแต่ละคู่ของเซลล์ประสาทที่อยู่ในชั้นติดกันจะเชื่อมต่อกันด้วยค่าถ่วงน้ำหนัก



รูปที่ 8 ลักษณะของโครงข่ายประสาทเทียม
ที่มา [13]

เซลล์ประสาททุก ๆ เซลล์ในแต่ละชั้น โดยยกเว้นในชั้นอินพุทนั้นจะทำการประมวลผลตามฟังก์ชันกระตุ้นและตามค่าถ่วงน้ำหนักที่ส่งเข้ามา จากนั้นจะสร้างเป็นเอาท์พุทเพื่อส่งเป็นอินพุทให้ในชั้นต่อไป [14] แสดงดังรูปที่ 9 ซึ่งลักษณะของการประมวลผลนั้น จะดำเนินการต่อเนื่องไปเป็นทอด ๆ จนกระทั่งถึงชั้นสุดท้ายของโครงข่ายประสาทซึ่งก็คือชั้นเอาท์พุทผลจากการประมวลผลข้อมูลของเซลล์ประสาทในชั้นนั้นจะเป็นค่าเอาท์พุทของโครงข่าย



รูปที่ 9 รูปแบบการคำนวณของโหนด
ที่มา [13]

สมการที่ 2.16 และ 2.17 ใช้ในการคำนวณหาอินพุทและเอาท์พุทของเซลล์ประสาท i ในโครงข่ายประสาท [14]

อินพุท $: X_i = \sum W_{ij} O_j + b_i$ (2.16)

เอาท์พุท $: O_i = f(X_i)$ (2.17)

โดยที่ผลรวม $\sum$ ในสมการที่ 2.16 จะครอบคลุมทุก ๆ เซลล์ประสาท j ที่อยู่ภายในชั้นก่อนหน้า การนำฟังก์ชันที่ไม่ใช่ฟังก์ชันเชิงเส้นมาคำนวณเอาท์พุทจะช่วยเพิ่มประสิทธิภาพของโครงข่ายประสาท สำหรับบางปัญหาที่มีความซับซ้อนมาก ซึ่งไม่สามารถที่จะใช้ฟังก์ชันเชิงเส้นธรรมดามาแก้ไขได้ [14] [15] ในงานวิจัยนี้จะมีการนำฟังก์ชันซิกมอยด์ (Sigmoid Function) ตามสมการที่ 2.18 เพื่อเป็นตัวที่นำมาใช้สำหรับการคำนวณเอาท์พุทของโครงข่าย

$$f(X_i) = 1/1 + \exp(-X_i) \quad (2.18)$$

โดยที่ w_{ij} ค่าถ่วงน้ำหนักที่เชื่อมระหว่างเซลล์ประสาท i และเซลล์ประสาท j

b_i ค่าตัวเลขค่าหนึ่งเรียกว่า ไบแอส (bias)

f ฟังก์ชันกระตุ้นของเซลล์ประสาท

2.1.2 ทฤษฎีคอมพิวเตอร์วิทัศน์และการประมวลผลภาพ (Computer Vision Theory and Image Processing)

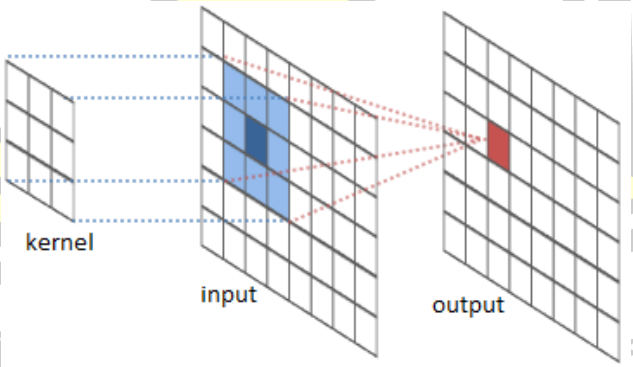
2.1.2.1 การประมวลผลภาพ (Image Processing) [16] หมายถึงการนำภาพมาประมวลผลหรือคิดคำนวณด้วยคอมพิวเตอร์ เพื่อให้ได้ข้อมูลที่ต้องการทั้งในเชิงคุณภาพและเชิงปริมาณ โดยมีขั้นตอนต่าง ๆ ที่สำคัญคือ การทำให้ภาพมีความคมชัดมากขึ้น การกำจัดสัญญาณรบกวนออกจากภาพ การแบ่งส่วนของวัตถุที่สนใจออกจากภาพ เพื่อนำภาพวัตถุที่ได้ไปวิเคราะห์หาข้อมูลเชิงปริมาณ เช่น ขนาด รูปร่าง และทิศทางการเคลื่อนตัวของวัตถุในภาพ จากนั้นเราสามารถนำข้อมูลเชิงปริมาณเหล่านี้ไปทำการวิเคราะห์ และสร้างระบบเพื่อนำมาใช้ประโยชน์ในงานด้านต่าง ๆ สำหรับพื้นฐานของการประมวลผลภาพเบื้องต้นที่สำคัญจะมีเนื้อหาดังต่อไปนี้

1) การปรับปรุงคุณภาพของภาพ (Image Enhancement) คือ การปรับเปลี่ยนคุณสมบัติทางกายภาพของภาพเพื่อให้เหมาะสำหรับการประมวลผลภาพในขั้นตอนที่สูงขึ้น เช่น การกำจัดสัญญาณรบกวน (Noise) ออกจากภาพ หรือเพื่อการเพิ่มหรือลดความคมชัดของภาพ การปรับปรุงคุณภาพจำเป็นอย่างยิ่งที่ต้องอาศัยเทคนิคทางด้านการประมวลผลภาพระดับล่าง ดังนั้นในหัวข้อนี้จะอธิบายถึงเทคนิคทางด้านการประมวลผลภาพที่ใช้สำหรับการปรับปรุงคุณภาพของภาพ

(1) การคอนโวลูชัน (Convolution) ในการประมวลผลภาพ การกรองข้อมูลภาพ (Image Filtering) เป็นเทคนิคสำหรับการปรับปรุงภาพ ตัวอย่างเช่น เราสามารถกรองภาพเพื่อเน้นคุณสมบัติบางอย่าง หรือเอาคุณสมบัติอื่น ๆ ออกจากภาพ การประมวลผลภาพด้วยการกรองข้อมูลภาพนั้นจะทำให้ภาพนวลขึ้น คมชัดขึ้น และปรับปรุงขอบภาพให้ดีขึ้น การกรองข้อมูลภาพ คือ การทำงานกับพื้นที่ใกล้เคียง ซึ่งค่าของพิกเซลใดก็ตามในภาพเอาต์พุต จะถูกกำหนดโดยการใช้วิธีการบางอย่างกับค่าของพิกเซลในพื้นที่ใกล้เคียงของพิกเซลอินพุตที่สอดคล้องกัน ซึ่งจะมีการใช้วิธีการประมวลผลภาพนี้จะเรียกว่าการคอนโวลูชัน ซึ่งการคอนโวลูชัน [17] คือ การกระทำกันระหว่างภาพ (Image) กับมาส์ค (Mask) หรืออาจจะเรียกว่า เทมเพลต (Template) หน้าต่าง (Window) รวมทั้งเคอร์เนล (Kernel) ก็ได้ โดยมาส์คคือ แมทริกซ์ขนาด $n \times m$ ของชุดตัวเลขที่จะนำไปซ้อนทับภาพที่ตำแหน่งต่าง ๆ เพื่อการหาผลลัพธ์ของการคอนโวลูชัน ถ้ากำหนดให้มาส์ค $M\ (i, j)$ เป็นหน้าขนาด $n \times m$ และภาพ $F\ (x, y)$ ต้นฉบับมีขนาด $n \times m$ การคอนโวลูชันระหว่างมาร์คกับภาพสามารถแสดงได้ดังสมการที่ 2.19

$$G\ (x, y) = M\ * F\ = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} M\ (i, j) \cdot F\ (x - i, y - j) \qquad (2.19)$$

โดยที่ $G\ (x, y)$ คือภาพผลลัพธ์ที่ได้จากการคอนโวโลชันที่จุดพิกัด (x, y) ใด ๆ



รูปที่ 10 การคอนโวลูชัน

2) การแยกส่วนของข้อมูลภาพ (Image Segmentation)

(1) การทำเทรชโฮลด์ (Thresholding) ขั้นตอนการประมวลผลภาพโดยทั่ว ๆ ไป [18] ส่วนใหญ่จะเริ่มจากการกรองภาพหรือการปรับปรุงภาพด้วยวิธีการต่าง ๆ แล้วนำภาพนั้นมาทำการแปลงให้เป็นภาพสีเทา โดยการแยกรอยต่อของวัตถุและพื้นหลังของวัตถุให้ออกจากกัน ซึ่งวิธีการที่ง่ายที่สุดของการแยกส่วนของข้อมูลภาพนี้จะเรียกว่า การทำเทรชโฮลด์ (Thresholding) ทั้งนี้การทำเทรชโฮลด์จะเป็นขั้นตอนง่าย ๆ ในการแยกส่วนของภาพ จากการนำภาพสีมาทำการแปลงให้เป็นภาพสีเทา แล้วนำภาพสีเทามาทำให้เป็นภาพไบนารีหรือภาพภาพขาวดำ ความสำคัญของกระบวนการทำเทรชโฮลด์คือ การเลือกค่าของเทรชโฮลด์ที่เหมาะสม ค่าของเทรชโฮลด์จะขึ้นอยู่กับงานและภาพที่ได้ ซึ่งในกระบวนการทำเทรชโฮลต์จะสามารถที่จะกำหนดค่าเทรชโฮลด์ได้โดยการเลือกค่าพิกเซลได้อยู่ 2 วิธี และวิธีที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ วิธีการออสุ (Otsu's Method)

วิธีที่ 1 พิกเซลสูงกว่าเทรชโฮลด์ ถ้าค่าของพิกเซลมีค่ามากกว่าค่าของเทรชโฮลด์ (วัตถุมีความสว่างกว่าพื้นหลัง) พิกเซลภาพจะมีค่าเป็น 255 (สีขาว) และพิกเซลพื้นหลังจะมีค่าเป็น 0 (สีดำ) จะเรียกว่า พิกเซลสูงกว่าเทรชโฮลด์ดังรูปที่ 11

รูปที่ 11 ภาพขาวดำที่พิกเซลสูงกว่าเทรชโฮลด์ (วัตถุมีความสว่างกว่าพื้นหลัง)

วิธีที่ 2 พิกเซลต่ำกว่าเทรชโฮลด์ ซึ่งตรงกันข้ามกับพิกเซลสูงกว่าเทรชโฮลด์ พิกเซลภาพจะมีค่าเป็น 0 และพิกเซลพื้นหลังมีค่าเป็น 255 ซึ่งพิกเซลที่ตำกว่าเทรชโฮลด์ โดยปกติค่าพิกเซลของวัตถุจะมีค่าเท่ากับ 1 (สีขาว) ในขณะที่พิกเซลของพื้นหลังจะมีค่าเท่ากับ 0 (สีดำ) และสุดท้ายภาพ ไบนารีจะถูกสร้างขึ้นโดยพิกเซลสีขาวหรือสีดำ ซึ่งจะขึ้นอยู่กับระดับของพิกเซลดังรูปที่ 12

รูปที่ 12 ภาพขาวดำที่พิกเซลตำกว่าเทรชโฮลด์ (วัตถุจะมีสีดำและพื้นหลังเป็นสีขาว)

วิธีที่ 3 วิธีการออสุ Otsu's Method [19] ได้ถูกนำเสนอวิธีการที่ใช้ในการหาค่าเทรชโฮลด์ ของรูปภาพแบบอัตโนมัติจากค่าของฮิสโตแกรมของรูปภาพ ซึ่งสามารถพิจารณาได้จากค่าพิกเซลของรูปภาพระดับของสีเทา L ระดับคือ $[1,2,\ldots,L]$ จำนวนของพิกเซลที่ระดับ i โดยกำหนดให้เป็น n_i และจำนวนของพิกเซลทั้งหมดเป็น n เมื่อ $n = n_1 + n_2 + \ldots + n_L$ จากนั้นจะสมมติให้พิกเซลทั้งหมดสามารถจำแนกได้เป็น 2 กลุ่ม (Class) คือ C_0 และ C_1 โดยแทนพิกเซลที่ระดับ $[1,\ldots,k]$ และ $[k+1,\ldots,L]$ ตามลำดับ ซึ่งวิธีการนี้ขึ้นอยู่กับค่าที่ใช้ในการแยกแยะ (Discriminant Criterion) ทั้งนี้ค่าดังกล่าวสามารถจะนำมาคำนวณหาได้จากอัตราส่วนของค่าความแปรปรวนระหว่างกลุ่ม (Between-Class Variance) กับค่าความแปรปรวนรวม (Total Variance) ดังสมการที่ 2.20

$$\sigma_B^2 = \omega_0\omega_1(\mu_1 - \mu_0)^2 \qquad (2.20)$$

$$\sigma_T^2 = \sum_{i=1}^{L} (i - \mu_T)^2 Pi$$

$$\eta = \frac{\sigma_B^2}{\sigma_T^2}$$

เมื่อ

σ_B^2 เป็นค่าความแปรปรวนระหว่างกลุ่ม

σ_T^2 เป็นค่าความแปรปรวนรวมของทุก ๆ ระดับ

Pi เป็นค่าการกระจายของความน่าจะเป็น $= \frac{n_i}{N}$

ω_0 เป็นความน่าจะเป็นของการเกิดกลุ่ม $= \sum_{i=1}^{k} P_i$

ω_1 เป็นความน่าจะเป็นของระดับเฉลี่ยของกลุ่ม $= 1 - \omega_0$

μ_0 เป็นโมเมนต์สะสม (Cumulative) ลำดับศูนย์ของฮิสโตแกรมจนถึงระดับที่ $k = \frac{\sum_{i=1}^{k} P_i}{\omega_0}$

μ_1 เป็นโมเมนต์สะสมลำดับที่ 1 ของฮิสโตแกรมจนถึงระดับที่ $k = \frac{\sum_{i=1}^{k} P_i}{\omega_1}$

μ_T เป็นระดับเฉลี่ยรวมของรูปภาพเพิ่ม $= \sum_{i=1}^{L} iP_i$

η เป็นเกณฑ์ที่ใช้ในการแยกแยะ

ค่าเทรชโฮลด์ที่เหมาะสมของภาพขึ้นอยู่กับ σ_B^2 ที่มีค่ามากที่สุดซึ่งสมมูลกับ η ซึ่งเป็นตัวเพิ่มความสามารถในการแยกแยะกลุ่มของรูปภาพ Grayscale ดังรูปที่ 13

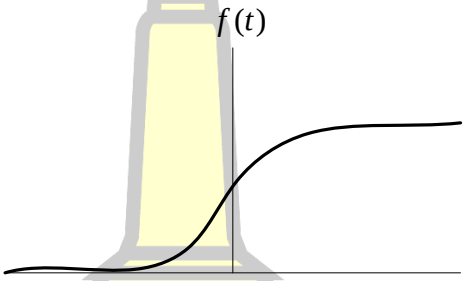
รูป A *รูป B*

รูปที่ 13 รูป A เป็นรูปภาพเดิมก่อนผ่านกระบวนการ รูป B เป็นรูปที่ผ่านกระบวนการ Otsu's method
ที่มา [20]

(2) การหาขอบของวัตถุ (Edge Detection) [11] คือจุดของภาพที่มีการเปลี่ยนแปลงระดับของความเข้มสีอย่างรวดเร็ว เช่น การเปลี่ยนจากจุดภาพสีดำเป็นสีขาวหรือจากจุดภาพสีขาวเป็นสีดำ โดยขอบของวัตถุนี้จะมีคุณสมบัติเฉพาะของแต่ละรูปภาพ ซึ่งการที่จะนำเอาภาพไปประมวลผลนั้นอาจจำเป็นที่จะต้องหาขอบของวัตถุทั้งหมดในภาพ เพราะจะเป็นการประหยัดพื้นที่ในหน่วยความจำและสามารถนำไปประมวลผลได้รวดเร็วขึ้น ขอบของวัตถุจะมีลักษณะพิเศษ (Special Feature) [21] อย่างหนึ่งของรูปภาพต่าง ๆ ซึ่งหมายถึงรอยต่อระหว่างส่วนที่เป็นพื้นที่

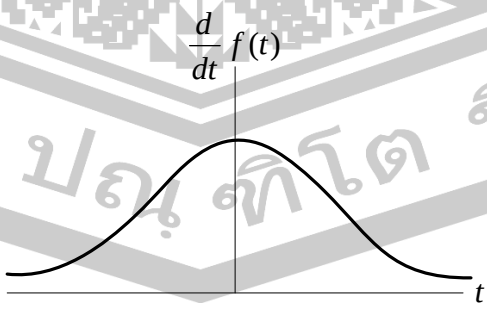
(Region) ที่มีความแตกต่างกันในรูปภาพ การหาขอบของวัตถุเป็นอีกหนึ่งวิธีในการนำเอาคุณลักษณะพิเศษเฉพาะของรูปภาพนั้น ๆ ออกมาได้โดยมีทฤษฎีต่าง ๆ มากมาย ซึ่งได้แบ่งออกเป็น 2 กลุ่มหลัก คือ วิธีเกรเดียนต์ (Gradient Method) และวิธีลาปลาเซียน (Laplacian Method) โดยทั้งสองวิธีจะมีหลักการที่แตกต่างกันโดยสามารถอธิบายพอสังเขปได้ดังนี้ [22] วิธีเกรเดียนต์จะตรวจหาขอบของวัตถุได้โดยการพิจารณาจากจุดสูงสุด และต่ำสุดของภาพที่ได้นำไปผ่านกระบวนการหาอนุพันธ์อันดับหนึ่ง ในขณะที่วิธีลาปเซียนจะค้นหาขอบภาพโดยอาศัยจุดผ่านศูนย์ของภาพที่ได้จากการนำไปผ่านกระบวนการหาอนุพันธ์อันดับสอง

พิจารณาตัวอย่างสัญญาณหนึ่งมิติ $f(t)$ ดังต่อไปนี้ [22]



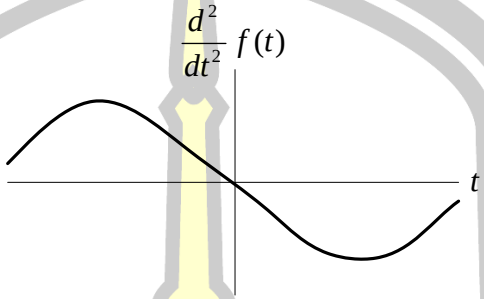
รูปที่ 14 รูปสัญญาณตัวอย่าง $f(t)$

จากรูปที่ 14 เมื่อนำมาดำเนินการหาอนุพันธ์อันดับหนึ่งของสัญญาณนี้จะได้ผลลัพธ์เป็นดังรูปที่ 15 จะสังเกตได้ว่าอนุพันธ์ที่ได้มีค่าสูงสุด ณ ที่จุดกึ่งกลางของรูป ดังนั้น ณ ตำแหน่งดังกล่าวสามารถนำมาใช้ในการระบุตำแหน่งของขอบภาพได้ แนวทางดังกล่าวนี้เป็นหลักการพื้นฐานของวิธีเกรเดียนต์



รูปที่ 15 อนุพันธ์อันดับหนึ่งของรูปสัญญาณตัวอย่างพื้นฐานของวิธีเกรเดียนต์

และเมื่อดำเนินการหาอนุพันธ์อันดับสองของสัญญาณนี้จะได้ผลลัพธ์เป็นดังรูปที่ 16 ซึ่งจะสังเกตได้ว่า ณ จุดที่ได้ผลลัพธ์ที่มีค่าเท่ากับศูนย์นั้นเป็นจุดที่สามารถบ่งบอกขอบภาพได้ และแนวทางดังกล่าวนี้เป็นหลักการพื้นฐานของวิธีลาปลาเชียน



รูปที่ 16 อนุพันธ์อันดับสองของรูปสัญญาณตัวอย่างพื้นฐานของวิธีลาปลาเชียน

แนวคิดเบื้องต้นดังกล่าวมานี้สามารถนำไปประยุกต์ใช้กับการหาขอบของวัตถุได้ ซึ่งจะได้กล่าวถึงโดยละเอียดต่อไป ทั้งนี้เนื้อหาในลำดับต่อไปนี้จะขอนำเสนอเฉพาะเทคนิคการหาขอบของวัตถุที่ใช้วิธีเกรเดียนต์เท่านั้น

1. วิธีเกรเดียนต์ (Gradient Method) เทคนิคการหาขอบของวัตถุตามวิธีเกรเดียนต์ (Gradient) จะเริ่มต้นมาจากนิยามทางคณิตศาสตร์ดังสมการที่ 2.21

$$gradient = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \tag{2.21}$$

โดยที่ f แทนภาพต้นฉบับที่ต้องการจะคำนวณหาค่าเกรเดียนต์

$\frac{\partial f}{\partial x}$ แทนค่าอนุพันธ์ของ f เทียบกับ x

$\frac{\partial f}{\partial y}$ แทนค่าอนุพันธ์ของ f เทียบกับ y

ค่าเกรเดียนต์ที่คำนวณได้นี้จะสามารถนำมาใช้ในการหาขอบของวัตถุได้ แต่ข้อมูลภาพจะมีคุณลักษณะที่เป็นจุด ๆ แบบไม่ต่อเนื่องกัน ดังนั้นการหาค่าอนุพันธ์กับข้อมูลนี้จะเป็นการนำค่าพิกเซลที่อยู่ต่อเนื่องกันมาหาค่าความแตกต่างระหว่างกันแทน ซึ่งอาจกล่าวได้ว่า $\frac{\partial f}{\partial x}$ เป็นค่าความแตกต่าง

ของพิกเซลในแนวนอนหรือแนวแกน x ส่วน $\frac{\partial f}{\partial y}$ ก็จะเป็นค่าความแตกต่างของพิกเซลในแนวตั้งหรือแนวแกน y โดยการคำนวณด้วยโปรแกรมคอมพิวเตอร์มักจะอาศัยมาสก์ (Mask) แทนยก ตัวอย่างเช่น การหาค่าอนุพันธ์ในแนวนอนจะนำมาสก์ $[1 \quad -1]$ ไปผ่านกระบวนการคอนโวลูชันกับภาพต้นฉบับ ในที่นี้จะขออธิบายตัวอย่างพอสังเขปดังต่อไปนี้

สมมุติให้

$$f = \begin{bmatrix} 10 & 10 & 10 & 10 \\ 10 & 80 & 80 & 10 \\ 10 & 80 & 80 & 10 \\ 10 & 10 & 10 & 10 \end{bmatrix} \; and \; mask\left(\frac{\partial f}{\partial x}\right) = [1 \quad -1]$$

เมื่อนำเมทริกซ์ทั้งสองมาดำเนินการหาคอนโวลูชันซึ่งจะได้ผลดังนี้

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 70 & 0 & -70 \\ 0 & 70 & 0 & -70 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

จากตัวอย่างจะแสดงให้เห็นว่า การหาคอนโวลูชันคือการนำมาสก์ไปพลิก 180 องศา แล้วจึงครอบลงบนพิกเซลที่ต้องการคำนวณ แล้วนำค่าตัวเลขของมาสก์แต่ละตัวไปคูณกับค่าของพิกเซลในภาพที่มีอยู่ในตำแหน่งที่ตรงกัน และเมื่อนำผลคูณที่ได้มาบวกรวมกันจะได้เป็นค่าคอนโวลูชันสำหรับพิกเซลนั้น ในกรณีที่การคำนวณกับพิกเซลที่อยู่ติดกับขอบของเมทริกซ์ภาพทำให้ค่าของมาสก์บางส่วนตกอยู่นอกเมทริกซ์ภาพ ให้แทนค่าในตำแหน่งขาดหายไปด้วยค่าของพิกเซลของของวัตถุที่อยู่ใกล้ที่สุด ทั้งนี้ในการหาค่าอนุพันธ์ในแนวตั้งสามารถทำได้ในทำนองเดียวกัน ซึ่งจะแตกต่างกันที่มาสก์ที่นำมาใช้ในการคำนวณดังนี้

$$f = \begin{bmatrix} 10 & 10 & 10 & 10 \\ 10 & 80 & 80 & 10 \\ 10 & 80 & 80 & 10 \\ 10 & 10 & 10 & 10 \end{bmatrix} \; and \; mask\left(\frac{\partial f}{\partial y}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

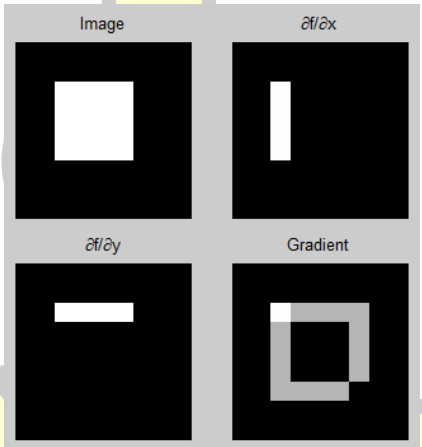
เมื่อนำเมทริกซ์ทั้งสองมาหาคอนโวลูชันจะได้ผลลัพธ์ดังนี้

$$\frac{\partial f}{\partial y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 70 & 70 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -70 & -70 & 0 \end{bmatrix}$$

และเมื่อได้ค่าทั้งอนุพันธ์ในแนวนอนและในแนวตั้ง ซึ่งก็จะสามารถหาค่าเกรเดียนต์ได้ดังนี้

$$gradient = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 70\sqrt{2} & 70 & 70 \\ 0 & 70 & 0 & 70 \\ 0 & 70 & 70 & 0 \end{bmatrix}$$

ซึ่งค่าเกรเดียนต์ที่ได้นี้สามารถนำมาใช้ในการระบุถึงขอบของวัตถุได้ดังรูปที่ 17 เป็นตัวอย่างการหาขอบของวัตถุสีเหลี่ยมสีขาวขนาด 4x4 พิกเซล บนพื้นสีดำขนาด 9x9 พิกเซล



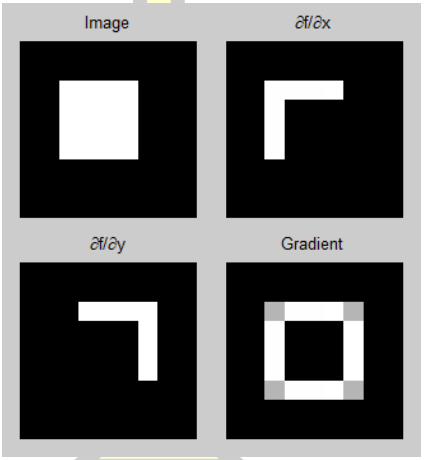
รูปที่ 17 ผลการใช้มาสก์อย่างง่ายเพื่อหาขอบของวัตถุสี่เหลี่ยมสีขาวขนาด 4x4 พิกเซล บนพื้นสีดำขนาด 9x9 พิกเซล

จากรูปที่ 17 จะสังเกตเห็นว่ามุมล่างด้านขวามือของเส้นขอบขาดหายไป โดยปัญหาดังกล่าวสามารถแก้ไขได้โดยการดัดแปลงมาสก์เพียงเล็กน้อยก็จะสามารถช่วยทำให้ได้เส้นขอบที่สมบูรณ์ขึ้น ซึ่งมาสก์ดังกล่าวได้รับการเสนอขึ้นโดย L.G. Roberts ในราวปี ค.ศ. 1965 โดยมีขนาดมาสก์ 2x2 พิกเซล

1.1 โรเบิร์ตครอส (Roberts Cross Operator)

$$mask\left(\frac{\partial f}{\partial x}\right)=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \qquad mask\left(\frac{\partial f}{\partial y}\right)=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$$

ผลลัพธ์ที่ได้จาการใช้มาสก์ของโรเบิร์ตจะมีผลที่ดีขึ้นกว่าการใช้มากส์แบบแรกดังรูปที่ 18



รูปที่ 18 ผลการหาขอบภาพด้วยมากส์ของโรเบิร์ตครอส

เส้นขอบสี่เหลี่ยมที่ได้มีความสมบูรณ์มากขึ้น และจะเห็นว่าตำแหน่งของเส้นขอบจะเลื่อนไปจากตำแหน่งที่ควรจะเป็นเล็กน้อย ซึ่งมาสก์ขนาด 2x2 มักจะให้ผลที่ไม่ดีนักในสภาพที่มีสัญญาณรบกวนในภาพมาก ด้วยเหตุนี้จึงได้มีการพัฒนามาสก์ที่มีขนาดใหญ่ขึ้นเพื่อลดผลกระทบจากสัญญาณรบกวนนี้ ซึ่งที่ผ่านมามีการนำเสนอมาสก์ในรูปแบบต่าง ๆ ดังตัวอย่างของมาสก์ที่ได้รับความนิยมได้แก่ [22] การหาขอบของวัตถุตามแบบของพรีวิตต์ (Prewitt) การหาขอบของวัตถุตามแบบของโซเบล (Sobel) การหาขอบของวัตถุตามแบบของโรบินสัน (Robinson) และการหาขอบของวัตถุตามแบบของเคิร์ส์ช (Kirsch)

1.2 พรีวิตต์ (Prewitt Operator)

$$mask\left(\frac{\partial f}{\partial x}\right)=\begin{bmatrix}1 & 0 & -1\\ 1 & 0 & -1\\ 1 & 0 & -1\end{bmatrix} \qquad mask\left(\frac{\partial f}{\partial y}\right)=\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix}$$

1.3 โซเบล (Sobel Operator)

$$mask\left(\frac{\partial f}{\partial x}\right)=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix} \qquad mask\left(\frac{\partial f}{\partial y}\right)=\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix}$$

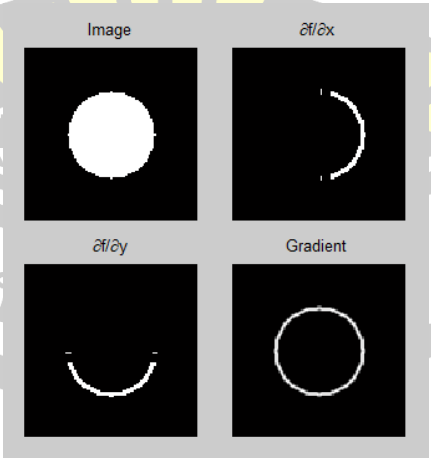
1.4 โรบินสัน (Robinson Operator)

$$mask\left(\frac{\partial f}{\partial x}\right)=\begin{bmatrix}-1 & 1 & 1\\ -1 & -2 & 1\\ -1 & 1 & 1\end{bmatrix} \qquad mask\left(\frac{\partial f}{\partial y}\right)=\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & 1\\ -1 & -1 & -1\end{bmatrix}$$

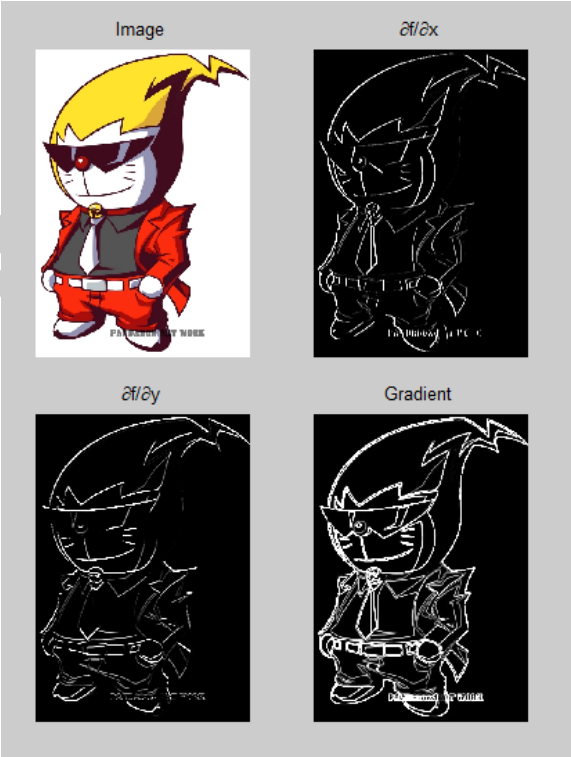
1.5 เคิรส์ซ (Kirsch Operator)

$$mask\left(\frac{\partial f}{\partial x}\right)=\begin{bmatrix}-5 & 3 & 3\\ -5 & 0 & 3\\ -5 & 3 & 3\end{bmatrix} \qquad mask\left(\frac{\partial f}{\partial y}\right)=\begin{bmatrix}3 & 3 & 3\\ 3 & 0 & 3\\ -5 & -5 & -5\end{bmatrix}$$

มาสก์ทั้งสี่แบบนี้มีขนาดเท่ากับ 3x3 ซึ่งมีขนาดใหญ่กว่ามาสก์ของโรเบิร์ตครอส จึงสามารถช่วยลดผลกระทบของสัญญาณรบกวนให้ต่ำลงได้ ดังตัวอย่างการหาขอบของวัตถุขาวดำด้วยวิธีของโซลเบล ดังรูปที่ 19 และการหาขอบของวัตถุสีด้วยวิธีของโซเบลดังรูปที่ 20 ทั้งนี้ยังมีอีกหนึ่งวิธีการในการหาขอบภาพคือวิธีแคนนี Canny

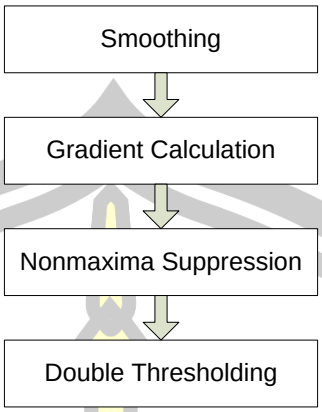


รูปที่ 19 ผลการหาขอบภาพขาวดำด้วยวิธีโซเบล



รูปที่ 20 ผลการหาขอบภาพสีด้วยวิธีโซเบล

2. วิธีการแคนนี (Canny Edge Detection) [21, 23] ได้มีการพัฒนาโดยนาย J.F. Canny โดยการนำ Mask ในลักษณะของเกาส์เซียนมาใช้ และปรับขนาดของตารางเมตริกให้ใหญ่กว่า $3x3$ โดยในขั้นตอนแรกจะเป็นการลดสัญญาณรบกวนด้วยวิธีการใช้การกรองแบบเกาส์เซียน ซึ่งขั้นตอนที่สองจะเป็นการคำนวณหาขนาด (Magnitude) และทิศทาง (Orientation) ของเกรเดียนต์โดยการหานุพันธ์อันดับที่หนึ่ง ขั้นตอนที่สามคือการปรับขอบให้มีความบางลงให้เหลือเพียง 1 พิกเซล และขั้นตอนสุดท้าย คือการระบุการเชื่อมโยงของพิกเซลแต่ละพิกเซลในภาพ ขั้นตอนของอัลกอริทึม Canny Edge Detection นั้นจะประกอบด้วยขั้นตอนต่าง ๆ ดังรูปที่ 21



รูปที่ 21 แสดงขั้นตอน Canny Edge Detection
ที่มา [21]

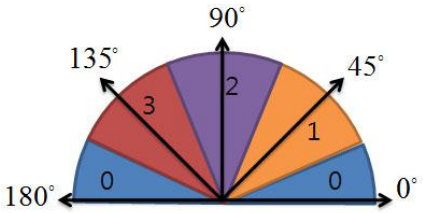
2.1. ขั้นตอนการลดสัญญาณรบกวน (Smoothing) เป็นขั้นตอนแรกซึ่งเป็นการลดสัญญาณรบกวนในภาพด้วยตัวกรองเกาส์เซียน (Gaussian Filter) โดยใช้เคอร์แนลคอนโวลูชันของตัวกรองเกาส์เซียน กับค่าเบี่ยงเบนมาตรฐานของ $\sigma = 1.4$ จะได้ดังสมการที่ 2.22

$$B = \frac{1}{159} \cdot \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \tag{2.22}$$

2.2 ขั้นตอนการคำนวณหาค่าเกรเดียนต์ (Gradient Calculation) หลังจากที่ทำการลดสัญญาณรบกวนในภาพด้วยวิธีการ Smoothing เรียบร้อยแล้วขั้นต่อมาคือการคำนวณหาค่า เกรเดียนต์ของแต่ละพิกเซลซึ่งสามารถใช้ตารางเคอร์แนลคอนโวลูชันของ Roberts Corss หรือ Sobel ในการคำนวณหาค่าเกรเดียนต์เวกเตอร์ในแนวนอนหรือ แกน x และเวกเตอร์ในแนวตั้งหรือ แกน y และคำนวณหาขนาดของเกรเดียนต์ดังสมการที่ 22 เมื่อได้ค่าขนาดของเกรเดียนต์สามารถคำนวณได้ดังสมการที่ 2.23

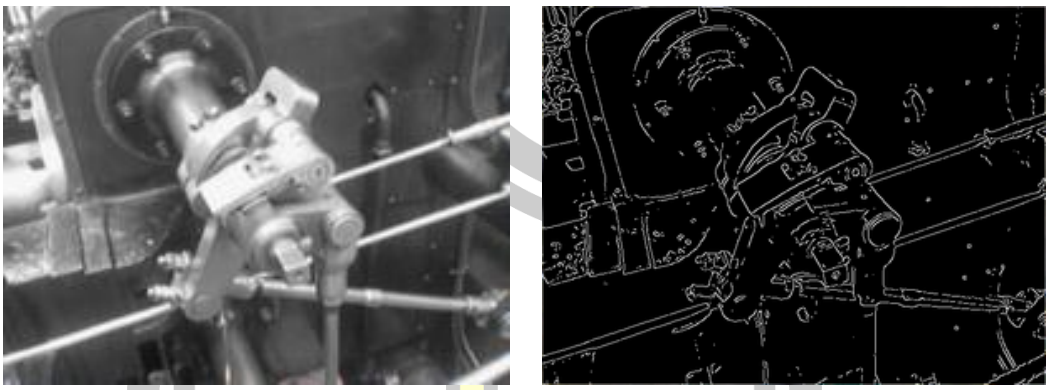
$$\theta = \arctan\left(\frac{|G_y|}{|G_x|}\right) \tag{2.23}$$

2.3 ขั้นตอนการกำจัดจุดภาพที่ไม่ใช่ขอบที่มีค่าสูงสุด (Nomaxima Supperession) ขั้นตอนนี้ใช้หน้าต่างย่อยขนาด 3 x 3 โดยจุดภาพที่ต้องการพิจารณาอยู่ตรงกลาง (X) และทิศทางของเกรเดียนต์แบ่งเป็น 4 ส่วนแสดงด้วยค่า 0 - 3 แทนทิศทาง 4 ทิศทาง โดยให้จุดศูนย์กลางของวงกลมเป็นจุดกลางภาพ รูปที่ 22 ในการพิจารณาทิศทางใดจะทำการเลือกจากค่าทิศทางเกรเดียนต์ที่อยู่จุดศูนย์กลาง (X) ของหน้าต่างย่อยแล้วนำไปเปรียบเทียบกับมุมที่แสดงในวงกลม จากนั้นทำการตรวจสอบค่าเกรเดียนต์ของจุดกลาง (X) ของหน้าต่างย่อยมีค่ามากกว่า 2 ค่าที่อยู่ในทิศทางที่พิจารณาหรือไม่ ถ้ามากกว่าให้คงค่าเดิมไว้ ถ้าน้อยกว่าค่าใดค่าหนึ่งจะทำการปรับค่าที่อยู่จุดศูนย์กลางเป็น 0



รูปที่ 22 แสดงการแบ่งค่าทิศทางของเกรเดียนต์

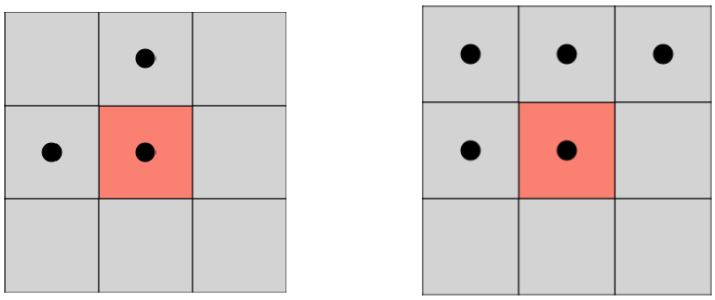
2.4 ขั้นตอนการแบ่งข้อมูลโดยใช้ค่าขีดแบ่ง (Double Thresholding) ขั้นตอนนี้จะทำเพื่อที่จะหาจุดเชื่อมโยงของขอบภาพ ซึ่งจะทำให้มีความชัดเจนเพิ่มมากขึ้นโดยการใช้เส้นขีดแบ่ง 2 ค่าคือ Hight Threshold (T_1) และ Low Threshold (T_2) ถ้าขนาดของเกรเดียนต์น้อยกว่า T_1 จะกำหนดค่าให้เป็น 0 ว่าไม่ใช่ขอบของวัตถุ ถ้ามีขนาดมากกว่า T_2 จะกำหนดค่าให้เป็น 1 ถือว่าเป็นขอบของวัตถุ ถ้าขนาดอยู่ระหว่าง T_1 และ T_2 ให้พิจารณาจุดภาพที่มีค่ามากกว่า T_1 เป็นจุดข้างเคียงมีค่าไม่มากกว่า T_1 ให้ทำการปรับค่าเป็น 0 (ไม่ใช่ขอบของวัตถุ)



รูป A รูป B

รูปที่ 23 รูป A คือภาพ Grayscale รูป B คือผ่านกระบวนการหาขอบด้วยวิธีของ Canny
ที่มา [24]

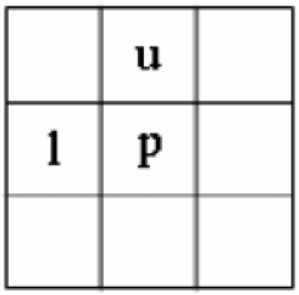
(3) การระบุส่วนเชื่อมต่อ (Connected Component Analysis) [25] เป็นกระบวนการที่นำมาวิเคราะห์ส่วนประกอบหรือบริเวณที่มีการเชื่อมต่อกันและนำมาใช้ในการแยกแยะบริเวณที่มีความแตกต่างกัน โดยวิธีนี้จะเป็นการกำหนดหมายเลขให้กับส่วนที่มีการเชื่อมต่อกัน ซึ่งการตรวจหาบริเวณที่มีการเชื่อมต่อกันของรูปภาพขาวดำ รูปภาพสีนั้น วิธีการโดยทั่วไปจะนำไปใช้กับรูปภาพขาวดำที่ได้ผ่านการประมวลผลมาแล้วมาทำการวิเคราะห์หาการเชื่อมต่อกันดังรูปที่ 24



รูปที่ 24 แบบจุด 4 จุดเชื่อมต่อกัน และแบบจุด 8 จุดที่เชื่อมต่อกัน
ที่มา [26]

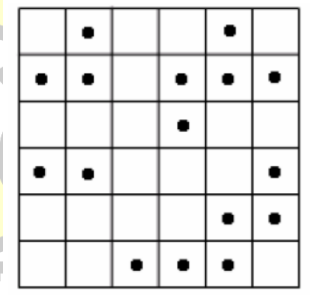
จากรูปที่ 24 จุดที่อยู่รอบ ๆ จุดที่อยู่ตำแหน่งกึ่งกลางคือ เพื่อนบ้านซึ่งมีทั้งแบบจุด 4 จุดที่เชื่อมต่อกันและแบบจุด 8 จุดที่เชื่อมต่อกันโดยเพื่อนบ้านที่มีการเชื่อมต่อกันก็คือขอบนั้นเอง ทั้งนี้การกำหนดหมายเลขจะมีด้วยกันอยู่ 2 แบบคือแบบจุด 4 จุดเชื่อมต่อกันและแบบจุด 8 จุดเชื่อมต่อกัน ซึ่งแต่ละแบบจะมีข้อที่แตกต่างกันคือแบบจุด 4 จุดจะเลือกจุดที่จะนำมาเชื่อมต่อ

เฉพาะจุดบน ล่าง ซ้าย และขวาเท่านั้น แต่แบบจุด 8 จุดจะรวมมุมทแยงจากตำแหน่งกึ่งกลางอีก 4 จุดด้วย



รูปที่ 25 ภาพอักษรตำแหน่งจุดภาพ
ที่มา [26]

จากรูปที่ 25 กำหนดให้พิกเซล p แทนจุดของรูปภาพที่กำลังพิจารณา พิกเซล u แทนจุดภาพที่อยู่ตำแหน่งเหนือจุดพิกเซล p และจุดพิกเซล l แทนจุดภาพที่อยู่ตำแหน่งทางด้านซ้ายของจุด p ซึ่งจะเริ่มกระบวนการจากซ้ายไปขวาและบนลงล่าง จากนั้นทำการกำหนดหมายเลขตามขั้นตอนดังต่อไปนี้

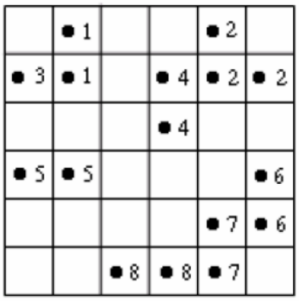


รูปที่ 26 ตัวอย่างจุดภาพและตำแหน่ง
ที่มา [26]

ขั้นตอนที่ 1 จากรูปที่ 26 ถ้าจุด p ไม่ใช่จุดภาพให้เลื่อนจุดในตำแหน่งถัดไป ถ้า p เป็นจุดภาพให้ตรวจสอบสถานะของจุด u และจุด l ถ้าไม่มีจุดใดเป็นจุดภาพให้ กำหนดหมายเลขขึ้นมาใหม่ให้กับจุด p ถ้ามี 1 จุดเป็นจุดภาพให้นำหมายเลขของจุดนั้นมากำหนดให้จุด p

แต่ถ้ามีมากกว่า 1 จุดเป็นจุดภาพสามารถนำหมายเลขของจุดใดก็ได้มากำหนดให้จุด p โดยถือว่าทุกหมายเลขมีค่าเท่าเทียมกัน

ขั้นตอนที่ 2 เมื่อสิ้นสุดกระบวนการในข้อตอนที่ 1 จุดภาพทุก ๆ จุดจะมีหมายเลขกำกับดังรูปที่ 27 แต่บางหมายเลขจะมีค่าเท่าเทียมกัน ให้รวมกลุ่มหมายเลขที่มีค่าเท่าเทียมกันดังรูปที่ 28 จากนั้นให้กำหนดหมายเลขในแต่ละกลุ่ม

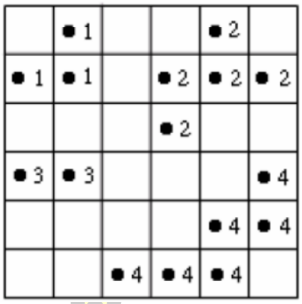


รูปที่ 27 รูปหมายเลขของแต่ละพิกเซลตามขั้นตอนที่ 1
ที่มา [26]

Set ID	Equivalent Labels
1	1,3
2	2,4
3	5
4	6,7,8

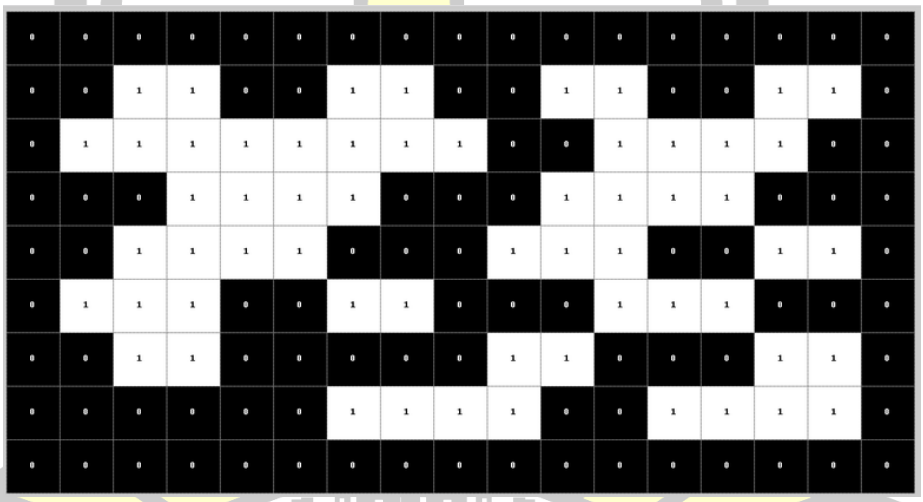
รูปที่ 28 กลุ่มรวมที่มีหมายเลขเทียบเท่ากัน
ที่มา [26]

ขั้นตอนที่ 3 นำหมายเลขของแต่ละกลุ่มจากขั้นตอนที่ 2 ไปแทนหมายเลขของจุดภาพที่อยู่ในกลุ่มเดียวกันดังรูปที่ 29

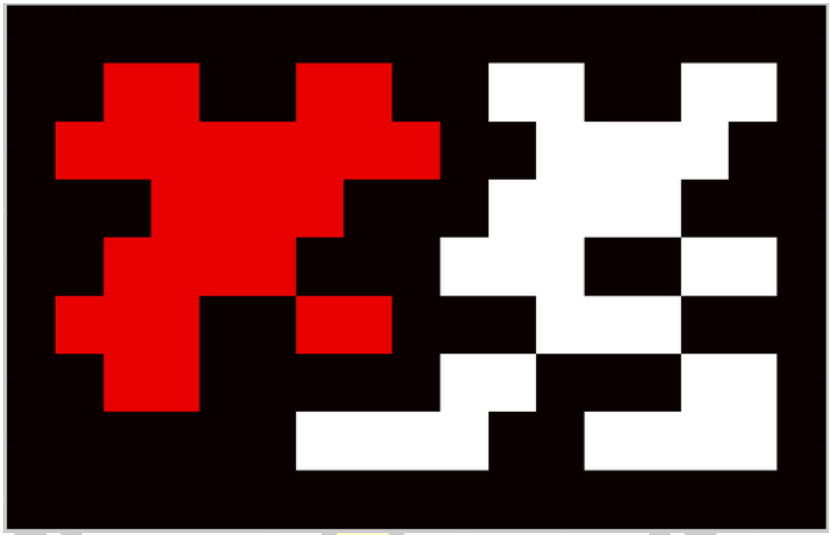


รูปที่ 29 หมายเลขของแต่ละจุดภาพตามขั้นตอนที่ 3
ที่มา [26]

จากผลลัพธ์ที่ได้จากรูปที่ 29 จะเห็นได้ว่าแต่ละบริเวณจะมีหมายเลขกำกับไม่ซ้ำกัน ดังนั้นหมายเลขสูงสุดก็คือจำนวนบริเวณทั้งหมดที่เชื่อมต่อกันดังรูปที่ 30 และรูปที่ 31 แสดงตัวอย่างของการหาจุดเชื่อมต่อ



รูปที่ 30 รูปแสดงตารางตารางเมตริกซ์
ที่มา [26]

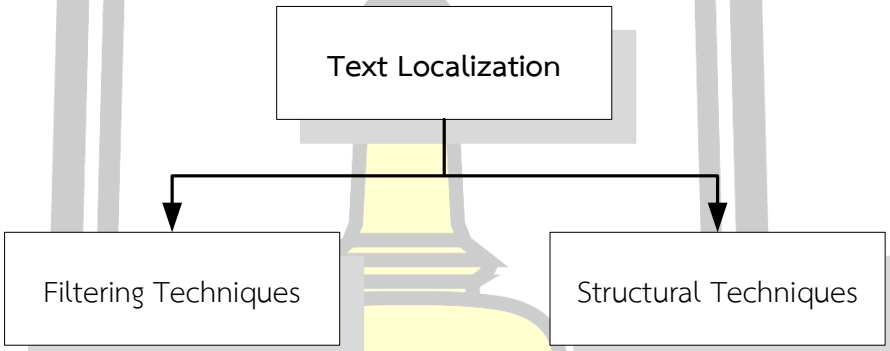


รูปที่ 31 ผลลัพธ์ของการผ่านกระบวนการ Connected-Component
ที่มา [26]

2.1.2.2 ทฤษฎีคอมพิวเตอร์วิทัศน์ (Computer Vision Theory) [27] เป็นอีกสาขาหนึ่งของวิทยาการคอมพิวเตอร์ ว่าด้วยเรื่องเกี่ยวกับการดึงข้อมูลสารสนเทศจากรูปภาพ หรือวีดีทัศน์ โดยเครื่องมือที่นำมาใช้ในงานคอมพิวเตอร์วิทัศน์ได้แก่ คณิตศาสตร์โดยเฉพาะเรขาคณิต พิชคณิตเชิงเส้น สถิติและการวิจัยดำเนินงาน (การหาค่าที่เหมาะที่สุด) และการวิเคราะห์เชิงฟังก์ชัน ซึ่งเครื่องมือเหล่านี้ใช้ในการสร้างขั้นตอนวิธีหรือขั้นตอนวิธีในการแยกส่วนภาพ และการจัดกลุ่มภาพเพื่อให้คอมพิวเตอร์สามารถเข้าใจทัศนียภาพ หรือคุณลักษณะต่าง ๆ ในภาพ ทั้งนี้เป้าหมายโดยทั่ว ๆ ไปของคอมพิวเตอร์วิทัศน์จะได้แก่ การตรวจหาวัตถุ การตัดแบ่งขอบเขต และการรู้จำวัตถุที่ต้องการภายในภาพ เพื่อที่จะให้บรรลุเป้าหมายเหล่านี้ระบบคอมพิวเตอร์วิทัศน์ จะต้องใช้กระบวนการต่าง ๆ เช่น การรู้จำแบบ การเรียนรู้เชิงสถิติ เรขาคณิตเชิงภาพฉาย การประมวลผลภาพ ทฤษฎีกราฟ และอื่น ๆ เป็นต้น ในที่นี้จะมุ่งเน้นในเรื่องของการตรวจหาพื้นที่ข้อความในภาพ การแบ่งส่วนข้อความในภาพ และการรู้จำตัวอักขระดังรูปที่ 32 ซึ่งเป็นประเด็นหลักในการดำเนินการวิจัยในครั้งนี้โดยจะมีรายละเอียดดังต่อไปนี้

รูปที่ 32 ประเด็นการดำเนินงานวิจัยหลัก

1) การตรวจหาพื้นที่ข้อความในภาพ (Text Localization) เป็นกระบวนการที่มีหน้าที่ในการตรวจหาส่วนที่คาดการณ์ว่าจะเป็นข้อความภายในภาพ โดยเทคนิคสำหรับการตรวจหาพื้นที่ข้อความในภาพนั้นปัจจุบันมีด้วยกันอยู่หลายเทคนิคด้วยกันซึ่งสามารถจัดกลุ่มได้เป็น 2 เทคนิค ดังรูปที่ 33



รูปที่ 33 เทคนิคสำหรับการตรวจหาข้อความในภาพ

(1) เทคนิคการกรองข้อมูลภาพ (Filtering Techniques) [28] คือการนำภาพไปผ่านตัวกรองสัญญาณเพื่อให้ได้ภาพผลลัพธ์ออกมา ภาพผลลัพธ์ที่ได้นั้นจะมีคุณสมบัติที่แตกต่างไปจากภาพเริ่มต้น โดยวัตถุประสงค์หลักของการกรองข้อมูลภาพจะเป็นการเน้น (Enhance) หรือการลดทอน (Attenuate) คุณสมบัติบางประการของภาพ เพื่อให้ได้ภาพที่มีคุณสมบัติตามที่ต้องการ การกรองข้อมูลภาพคือ การประมวลผลภาพอย่างหนึ่งที่จำเป็นมาก เนื่องจากในการนำไปใช้งานจริงภาพที่ได้มามักจะมีสัญญาณรบกวน หรือสัญญาณที่ไม่พึงประสงค์อื่นปะปนอยู่ด้วย การกรองข้อมูลภาพสามารถปรับปรุงให้ภาพมีคุณสมบัติที่ดีขึ้น เหมาะสำหรับการนำไปประมวลผลในขั้นต่อไป ซึ่งในที่นี้จะขอกล่าวถึงเทคนิคที่ได้รับความนิยมในการตรวจหาข้อความในภาพ

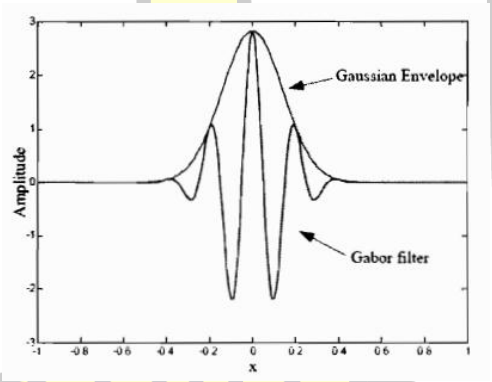
1. ตัวกรองกาบอร์ (Gabor Filter) [29] หรือเป็นเทคนิคการแปลงเวฟเล็ต (Wevelet Transform) รูปแบบหนึ่งถูกนำเสนอขึ้นเป็นครั้งแรกในปี ค.ศ. 1946 โดยชาวเยอรมัน เทคนิคดังกล่าวเป็นเทคนิคที่มีชื่อเสียงและได้รับการยอมรับอย่างกว้างขวางจนได้รับรางวัลโนเบลด้าน

ฮอโลกราฟิประดิษฐ์ (Inventing Holography) ในเวลาต่อมาตัวกรองกาบอร์สามารถนำมาพิจารณาได้เป็น 2 แบบคือ แบบที่หนึ่งพิจารณาในเชิงเวลา (Spatial Domain) มองภาพว่าเป็นที่รวมของพิกเซลต่าง ๆ แต่ละพิกเซลมีระยะห่างจากจุดเริ่มต้นแตกต่างกัน และแบบที่สองจะพิจารณาในเชิงความถี่ (Frequency Domain) มองภาพว่าเป็นผลรวมของสัญญาณรูปไซน์ที่มีอยู่อย่างไร้ขอบเขตจำกัด ตัวกรองกาบอร์เป็นแบนด์พาสฟิลเตอร์ดังแสดงในรูปที่ 34 ซึ่งสามารถอธิบายได้ด้วยฟังก์ชันผลตอบสนองอิมพัลส์ (Impulse Response Function: IRF) โดยได้จากฟังก์ชันเกาส์เซียนมอดูเลตกับสัญญาณรูปคลื่นไซน์ฟังก์ชันกาบอร์ 2 มิติ สามารถพิจารณาได้ตามสมการที่ 2.24

$$h(x, y : \phi, f) = \exp\left\{-\frac{1}{2}\left[\frac{x_\phi^2}{\delta_x^2} + \frac{y_\phi^2}{\delta_y^2}\right]\right\}\cos(2\pi f x_\phi) \qquad (2.24)$$

โดยที่ f คือ ความถี่มอดูเลต (Modulation Frequency) ของฟังก์ชันกาบอร์

δ_x, δ_y คือ ส่วนเบี่ยงเบนมาตรฐานของ Gaussian Envelope ตามแนวแกน x และ y



รูปที่ 34 ตัวกรองกาบอร์ในโดเมนเวลา
ที่มา [29]

ตัวกรองกาบอร์เป็นตัวกรองแบบแบนด์พาสฟิลเตอร์ ที่ถูกสร้างขึ้นโดยใช้เทคนิคการสร้างตัวกรองในโดเมนความถี่แบ่งเป็น 2 องค์ประกอบ ได้แก่องค์ประกอบทางรัศมีและองค์ประกอบทางมุม ดังแสดงในรูปที่ 35

$$G(u, v) = \exp\left\{-\frac{1}{2}\left[\frac{(u - u_o)^2}{\delta_u^2} + \frac{v^2}{\delta_v^2}\right]\right\} \qquad (2.25)$$

$$\delta_u = \frac{1}{2\pi\delta_x} \tag{2.26}$$

$$\delta_v = \frac{1}{2\pi\delta_y} \tag{2.27}$$

โดยที่ $G(u,v)$ คือตัวกรองกาบอร์ใน Spatial Frequency Plane

u_o คือค่าความถี่ศูนย์กลางตัวกรอง

องค์ประกอบทางรัศมี (Radial Component) + องค์ประกอบทางมุม (Angular Component) = ตัวกรอง (Filter)

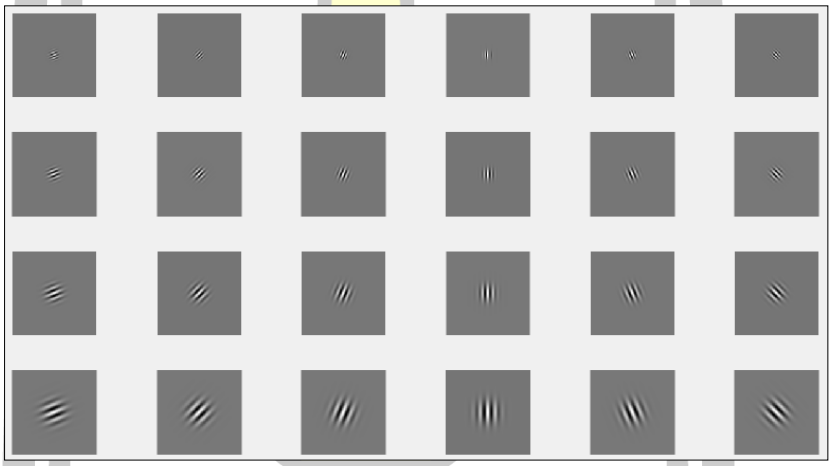
รูปที่ 35 ตัวกรองกาบอร์ใน Spatial Frequency Plane
ที่มา [29]

เมื่อสร้างตัวกรองกาบอร์ได้แล้วนำมาคูณกับภาพที่ผ่านการแปลงฟูเรียร์และนำภาพที่ได้ไปแปลงฟูเรียร์กลับจะเป็นเมตริกซ์จำนวนเชิงซ้อน โดยแต่ละจุดภาพประกอบด้วยส่วนที่เป็นจริง $\mathrm{Re}(x)$ และส่วนจินตภาพ $\mathrm{Im}(x)$ ค่าขนาดของกาบอร์ A_n ของเมตริกซ์จำนวนเชิงซ้อน ดังสมการที่ 2.28

$$A_n = \sqrt{\mathrm{Re}(x) + \mathrm{Im}(x)} \tag{2.28}$$

สำหรับตัวกรองกาบอร์เพื่อการใช้งานในการวิเคราะห์พื้นผิว [30] การพิจารณาเพื่อเลือกตัวกรองกาบอร์เพื่อการนำไปใช้ในงานการวิเคราะห์พื้นผิวโดยทั่วไปสามารถพิจารณาแบ่งออกได้ 2 แบบคือ พิจารณาการตรวจจับด้วยวิธีมัลติชันแนลกาบอร์ฟิลเตอร์ริง และพิจารณาโดยการหาค่าพารามิเตอร์ที่เหมาะสมของตัวกรองกาบอร์

1.1 การพิจารณาใช้งานตัวกรองกาบอร์ด้วยวิธีมัลติชันแนลกาบอร์ฟิลเตอร์ริง หรือวิธีการที่ไม่มีผู้ฝึกสอน (Unsupervised Method) สามารถที่จะทำได้โดยการกำหนดค่าสเกล (M) และการปรับทิศทาง (N) เพื่อสร้างแบงก์แต่ละแบงก์ของตัวกรองกาบอร์ จากนั้นจะนำแต่ละแบงก์ของตัวกรองกาบอร์ที่ได้ไปทำการคอนโวลูชัน กับภาพเพื่อหาตัวกรองที่สามารถให้ค่าคุณลักษณะเพื่อการจำแนกคุณลักษณะของพื้นผิวที่สนใจที่ดีที่สุดต่อไป ตัวอย่างเช่น กำหนดให้ M เท่ากับ 4 และ N เท่ากับ 6 พิจารณาแบงก์ของตัวกรองกาบอร์ (Gabor Filter Bank) ที่มีองค์ประกอบเป็นส่วนจินตภาพได้ดังรูปที่ 36 โดยมีจำนวนฟิลเตอร์แบงก์ทั้งหมดเท่ากับ 24 ตัว $(M \times N)$ ที่จะถูกนำไปคำนวณค่าคุณลักษณะเพื่อหาตัวกรองตัวที่ให้ค่าคุณลักษณะเพื่อหาตัวกรองตัวที่ให้ค่าคุณลักษณะที่ดีที่สุดจากทั้งหมด 24 ตัวกรอง ซึ่งการพิจารณาความแตกต่างของจำนวนฟิลเตอร์แบงก์เมื่อกำหนดค่าที่แตกต่างกันในโดนเมนความถี่ได้ดังรูปที่ 36



รูปที่ 36 รูปร่างที่เป็นองค์ประกอบส่วนจินตภาพของตัวกรองกาบอร์ ในโดเมนเวลา: M =1 - 4 และ N = 1 - 6

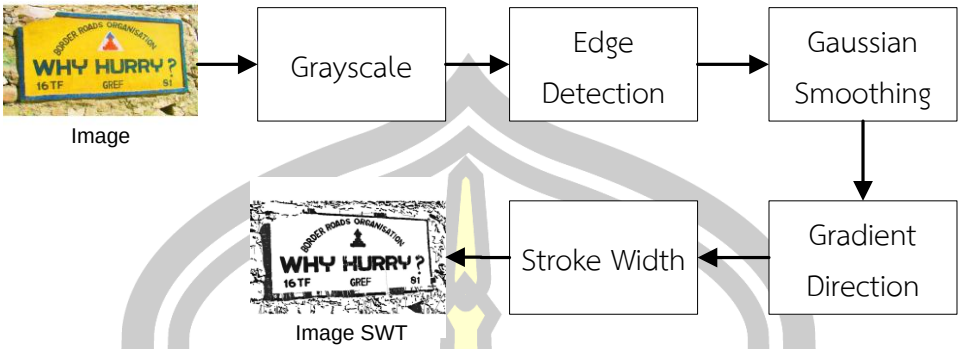
จะเห็นได้ว่าวิธีการใช้งานตัวกรองกาบอร์นี้ ต้องการการกำหนดค่าการสเกลและการปรับทิศทางที่เหมาะสมเพื่อสร้างเป็นกลุ่มของตัวกรองกาบอร์ ซึ่งปัญหาคือ จะกำหนดค่าการสเกลและการปรับทิศทางเท่าไหร่ที่จะครอบคลุมผลเฉลยหรือคำตอบที่เหมาะสมที่สุดของปัญหาที่ต้องการได้ โดย วิไลลักษณ์ คิดสร้าง [30] ได้อธิบายถึงปัญหาว่าตัวกรองกาบอร์นี้ยังขาดความยืดหยุ่นและยากต่อการใช้งาน ยิ่งไปกว่านั้นวิธีการดังกล่าวยังต้องสร้างข้อมูลในการประมวลผลมากมายผ่านกลุ่มของตัวกรอง จึงส่งผลต่อการจำแนก การรู้จำ และเวลารวมเป็นอย่างมาก

1.2 การพิจารณาใช้งานโดยการหาค่าพารามิเตอร์ที่เหมาะสมของตัวกรองกาบอร์หรือวิธีการที่มีผู้ฝึกสอน (Supervised Method) สามารถทำได้โดยไม่จำเป็นต้องกำหนดค่าการสเกลและการปรับทิศทาง วิธีการนี้จะช่วยลดข้อด้อยของวิธีการแรกได้ด้วยการปรับค่าพารามิเตอร์ให้เข้ากับพื้นหลังของลายผิวที่ต้องการ โดยไม่มีการจำกัดค่าการสเกลและการปรับทิศทาง รวมถึงไม่ต้องคำนวณหาพารามิเตอร์ที่เหมาะสมจากทุกการสเกลและการปรับทิศทาง

(2) Structural Techniques เป็นเทคนิคที่จะพิจารณาหรือวิเคราะห์ถึงโครงสร้างของวัตถุที่มีอยู่ในภาพ โดยเทคนิคที่นิยมใช้ในกลุ่มนี้มีด้วยกันอยู่หลายวิธีได้แก่

1. กระบวนการตรวจหาขอบของภาพ (Edge Detection) คือ กระบวนการที่ตรวจหาเส้นขอบของวัตถุภายในภาพและนำเส้นขอบของวัตถุนั้นมาพิจารณาหาสิ่งที่ต้องการในกระบวนการทัดไป ทั้งนี้กระบวนการตรวจหาขอบของภาพสามารถดำเนินการได้โดยการนำภาพต้นฉบับไปผ่านการคำนวณด้วยวิธีการคอนโวลูชันด้วยมาสก์ต่าง ๆ ซึ่งผลลัพธ์ที่ได้นั้นจะถูกเรียกว่า Edge map ในงานวิจัยของ Ezaki และคณะ [31] ได้มีการประยุกต์ใช้กระบวนการนี้ในการตรวจหาขอบของภาพด้วยการนำ Sobel Operator ที่มีขนาด 3x3 พิกเซล มาใช้ในการตรวจหาเส้นขอบของตัวอักขระที่มีอยู่ในภาพ อีกทั้งในงานวิจัยของ Liu และคณะ [32] มีการปรับปรุงวิธีการให้สามารถดำเนินการตรวจหาเส้นขอบของตัวอักขระได้ดียิ่งขึ้น ด้วยการสร้าง Edge map ในทิศทางที่แตกต่างกันได้แก่ ทิศทางที่ 0 องศา 45 องศา 90 องศาและ 135 องศา นอกจากการนำ Sobel operator มาใช้ในการตรวจหาเส้นขอบของตัวอักขระแล้ว ในงานวิจัยของ Phan และคณะ [33] ได้มีการนำ Laplacian Operator ขนาด 3x3 พิกเซลมาใช้ในการตรวจหาเส้นขอบของตัวอักขระเช่นเดียวกัน ทั้งนี้กระบวนการตรวจหาขอบของภาพนั้นเป็นกระบวนการที่ใช้ในการตรวจสอบโครงสร้างของวัตถุที่มีอยู่ในภาพเพียงหยาบ ๆ เท่านั้น การที่จะทราบถึงบริเวณหรือพื้นที่ของข้อความที่แทรกอยู่ภายในภาพนั้นจำเป็นต้องอาศัยกระบวนการที่มีการนำข้อมูลจาก Edge Map ไปพิจารณาหรือวิเคราะห์ต่อไป

2. กระบวนการ Stroke Width Transform (SWT) เป็นกระบวนการในพิจารณาจากความกว้างของวัตถุที่ปรากฏอยู่ภายในภาพ โดยจะคำนวณจากขอบของวัตถุด้านหนึ่งไปยังอีกด้านหนึ่ง และทำการบันทึกค่าความกว้างที่ได้เก็บไว้ในแต่ละพิกเซล เทคนิคนี้ได้มีการนำมาประยุกต์ใช้ในหลายงายวิจัย เช่น งานวิจัยของ Epshtein และคณะ [34] งานวิจัยของ Subramanian และคณะ [35] หรืองานวิจัยของ Jung และคณะ [36] ในการตรวจหาข้อความที่อยู่ภายในภาพทั้งนี้ในงานวิจัยต่าง ๆ ได้ให้แนวคิดว่า ข้อความหรือตัวอักขระจะมีขนาดความกว้างที่เท่ากัน ซึ่งขั้นตอนในการดำเนินการดังรูปที่ 37



รูปที่ 37 กระบวนการ The Stroke Width Transform

จากรูปที่ 37 จะประกอบไปด้วยกระบวนการทั้งหมด 5 ขั้นตอนโดยจะขออธิบายดังนี้ ขั้นตอนแรก คือการทำภาพเป็นสีเทา (Grayscale) เป็นโหมดสีที่มีการไล่เฉดสีของสีเทา โดยจะมีระดับความเข้มของสีเทา คือ 0-255 (8 bit) ภาพ Grayscale เกิดจากการแปลงภาพสี RGB มาเป็นภาพ Grayscale ขั้นตอนที่สอง คือการตรวจหาขอบภาพ (Edge Detection) การตรวจจับขอบภาพเพื่อหาเส้นขอบรอบรูปของวัตถุภายในภาพ เป็นขั้นตอนพื้นฐานหนึ่งในการประมวลผลภาพ การวิเคราะห์ภาพ การจดจำรูปแบบภาพ เทคนิคการมองเห็นของเครื่องจักร (Machine Vision) และการมองเห็นของคอมพิวเตอร์ (Computer vision) โดยเฉพาะอย่างยิ่งการตรวจสอบหาคุณสมบัติและการแยกคุณลักษณะ ซึ่งภาพที่นำเข้ามาประมวลผลจะมีคุณภาพที่แตกต่างกัน เช่น ความคมชัดของภาพ โดยเทคนิคการหาขอบภาพจะมีด้วยกันอยู่หลายวิธี ซึ่งในเทคนิคนี้จะใช้การตรวจจับขอบภาพด้วยวิธีแคนนี (Canny Method) ขั้นตอนที่สาม คือการกรองแบบเกาส์เซียน (Gaussian Smoothing) เป็นการกรองโดยการใช้ฟังก์ชั่นแบบเกาส์เซียน (Gaussian Function) ผลที่ออกมาภาพจะมีความเบลอ ซึ่งเป็นการลดสัญญาณรบกวน (Noise) และลดรายละเอียดของรูปภาพ โดยวิธีการนี้จะเหมือนกับการทำคอนวูโลชัน (Convolution) ซึ่งจะมีผลในการลดองค์ประกอบที่มีความถี่สูง ซึ่งการตัวกรองนี้จะยอมให้ความถี่ต่ำผ่านได้ (Low-pass Filter) ขั้นตอนที่สี่ คือการหาทิศทางของภาพด้วยเกรเดียนต์ (Gradient Direction) ซึ่งการเกรเดียนต์ภาพจะสามารถใช้ในการดึงข้อมูลออกจากรูปภาพได้ ภาพเกรเดียนต์จะถูกสร้างขึ้นจากภาพต้นฉบับ ทั้งนี้โดยทั่วไปจะได้จากการกรอง คอนวูโลชันที่เป็นวิธีหนึ่งที่ง่ายที่สุด ผลที่ได้จะเป็นทิศทางเกรเดียนต์ของภาพในแนวแกน X และแกน Y โดยการไล่ระดับสีของภาพ สุดท้ายขั้นตอน Stroke Width ในขั้นตอนนี้จะนำข้อมูลของรูปภาพซึ่งได้แก่ ข้อมูลพิกเซลขอบของรูปภาพ (Edge Image) ข้อมูลพิกเซลรูปภาพเกรเดียนต์แกน X และข้อมูลพิกเซลรูปภาพเกรเดียนต์แกน Y นำมาคำนวณหาขนาดความกว้างของวัตถุด้วยสมการที่ 2.29

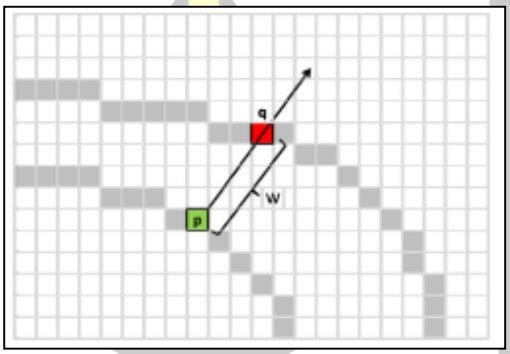
$$r = p + n \cdot dp, n > 0 \quad (2.29)$$

โดยที่

r เป็นตำแหน่งของพิกเซลทัดไป

dp เป็นทิศทางการตั้งฉากกับแนวของแต่ละพิกเซล p ซึ่งได้จากการหาเกรเดียนต์แกน X และ เกรเดียนต์แกน Y

กระทำซ้ำสมการที่ 2.29 จนกระทั่งพบขอบของพิกเซล q ดังรูปที่ 38



รูปที่ 38 รูปภาพแสดงการหาทิศทางของขอบในพิกเซล p และ q

ทั้งนี้ยังสามารถสามารถจะพิจารณาทิศทางการได้จาก d_q จากพิกเซล q โดยที่ d_q จะได้จากทิศทางของพิกเซล q ซึ่งได้จากการหา เกรเดียนต์แกน X และ เกรเดียนต์แกน Y หลังจากนั้นนำค่าที่ได้จากตำแหน่ง p และตำแหน่ง q มาคำนวณหาค่า w ดังสมการที่ 2.30 และแทนที่ค่าที่ได้ลงในแต่ละพิกเซล โดยผลที่ได้จากการหา SWT ดังรูปที่ 39

$$w = \sqrt{(q_x - p_x)^2 + (q_y - p_y)^2} \quad (2.30)$$

โดยที่

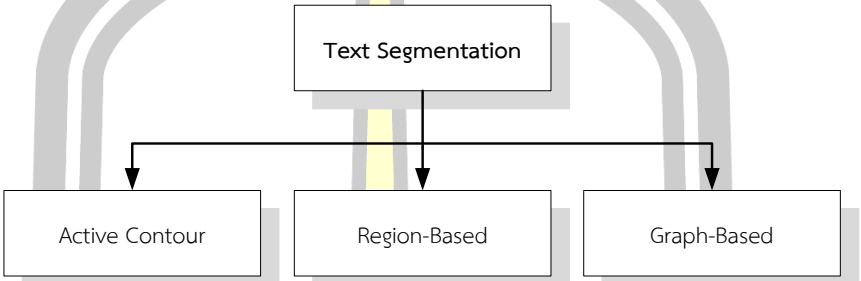
w = เป็นค่าขนาดความกว้างของเส้น $\|\overrightarrow{p-q}\|$

รูปที่ 39 รูป (A) เป็นรูปภาพขอบแคนนี รูป (B) เป็นรูปภาพที่ได้จาการแปลงเกรเดียนต์ในแนวแกน X รูป (C) เป็นรูปภาพที่ได้จาการแปลงเกรเดียนต์ในแนวแกน X และ (E) เป็นรูปภาพ Stroke Width Transform (SWT)

ในขั้นตอนของกระบวนการ Stroke Width ได้มีงานวิจัยของ Subramanian และคณะ [35] ที่มีแนวคิดที่แตกต่างออกไปในการคำนวณความกว้างทำการค้นหาภาพในแนวนอนเพื่อจับคู่ของพิกเซลกับพิกเซลฝั่งตรงข้าม (จุดพิกเซลที่เป็นขอบภาพจากฝั่งหนึ่งไปอีกฝั่งหนึ่ง) ซึ่งเป็นวิธีการที่สามารถตรวจสอบข้อความที่อยู่ใกล้ในแนวนอนเท่านั้น เทคนิค SWT นี้สามารถช่วยให้เราสามารถนำไปใช้งานได้ในหลาย ๆ ภาษา และรูปแบบตัวอักขระที่มีความหลากหลาย พร้อมทั้งการตรวจหาข้อความด้วยวิธีนี้ยังสามารถตรวจหาความที่มีโค้งของบรรทัดข้อความได้ดี

2) การแบ่งส่วนข้อความ (Text Segmentation) เป็นการแบ่งส่วนของข้อมูลภาพออกเป็นส่วนย่อย ๆ โดยที่แต่ละส่วนจะมีพื้นที่ที่ติดต่อกัน ซึ่งแต่ละส่วนจะอาจจะเป็นวัตถุที่อยู่ในภาพ การแบ่งส่วนภาพจะเสร็จได้เมื่อวัตถุที่ต้องการถูกแบ่งออกได้อย่างสมบูรณ์ ผลลัพธ์ที่ได้จากการแบ่งส่วนจะเป็นตัวชี้วัดความสำเร็จในขั้นตอนการวิเคราะห์ภาพ ดังนั้นการแบ่งส่วนข้อความ

ในภาพ (Text Segmentation) จึงเป็นกระบวนการที่ดำเนินการหลังจากเราได้พื้นที่ที่คาดว่าเป็นข้อความภายในภาพแล้ว ซึ่งกระบวนการนี้จะทำหน้าที่ในการแบ่งส่วนของพื้นที่ที่คาดว่าเป็นข้อความออกเป็นส่วน ๆ เพื่อให้ได้พื้นที่ที่คาดว่าเป็นตัวอักขระ โดยเทคนิคสำหรับการแบ่งส่วนข้อความในภาพนั้นปัจจุบันมีด้วยกันอยู่หลายเทคนิคด้วยกันซึ่งสามารถจัดกลุ่มได้เป็น 3 เทคนิคดังรูปที่ 40



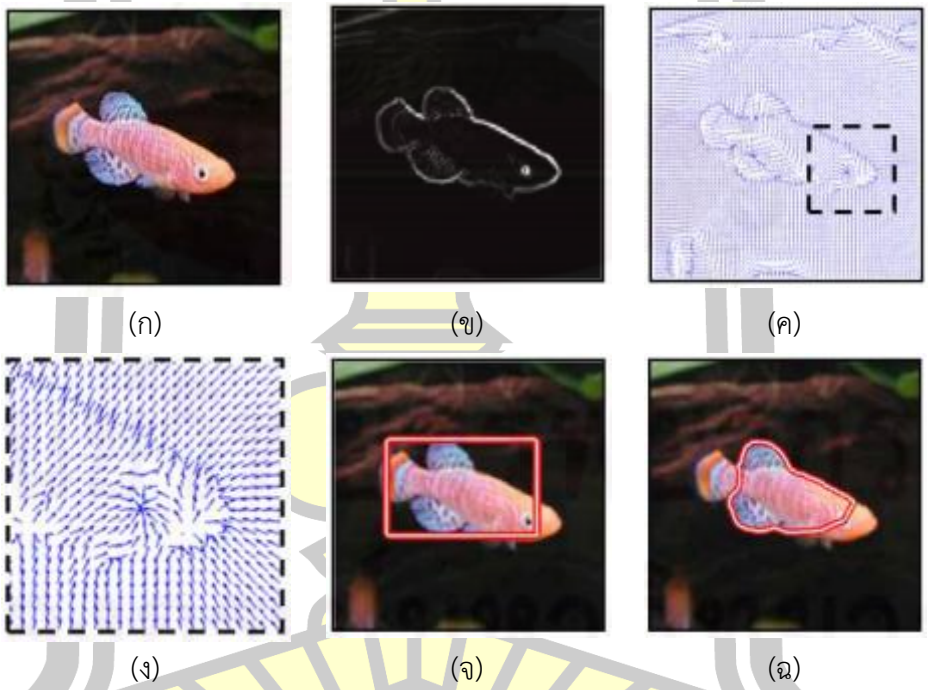
รูปที่ 40 เทคนิคสำหรับการแบ่งส่วนข้อความ

(1) วิธีการแอ็กทีฟคอนทัวร์ (Active Contour)

1. การแอ็กทีฟคอนทัวร์หรือที่เรียกว่า สเน็ก (Snake) [37] เนื่องจากเป็นวิธีการที่มีความยืดหยุ่นมากในการพัฒนาและออกแบบ และสามารถนำไปประยุกต์ ใช้กับภาพในงานด้านต่าง ๆ ได้ดีอีกด้วยซึ่ง วิธีแอ็กทีฟคอนทัวร์นี้มีหลักการเบื้องต้น คือ เราจะทำการปล่อยเส้นคอนทัวร์ (Contour) ลงไปบนภาพที่ต้องการแบ่งส่วน จากนั้นเส้นคอนทัวร์จะค่อย ๆ เคลื่อนที่และเปลี่ยนรูปร่างไปยังวัตถุ (Object) ที่เราต้องการในภาพจนกระทั้งได้วัตถุที่ต้องการออกมา ทั้งนี้ในการควบคุมการเคลื่อนที่และการเปลี่ยนรูปร่างของเส้นคอนทัวร์นั้น สามารถทำได้โดยการอาศัยแรง 2 รูปแบบคือ แรงภายในคอนทัวร์ (Internal Force) และแรงภายนอกคอนทัวร์ (External Force) ซึ่งแรงภายในคอนทัวร์จะทำหน้าที่ควบคุมความราบเรียบของคอนทัวร์ในขณะเคลื่อนที่ ดังนั้นจึงถูกเรียกอีกชื่อหนึ่งว่า แรงราบเรียบ (Smoothing Force) ส่วนแรงภายนอกนั้นจะทำหน้าที่ในการเปลี่ยนรูปร่างและขับเคลื่อนคอนทัวร์ไปยังวัตถุที่เราต้องการในภาพ จึงถูกเรียกว่า แรงหลัก (Main Force) แอ็กทีฟคอนทัวร์สามารถแบ่งออกเป็น 2 ประเภทได้แก่

1.1 แอ็กทีฟคอนทัวร์แบบใช้ขอบ (Edge-Based) จะใช้ข้อมูลสารสนเทศของขอบ (Edge Information) ของภาพอินพุตในการคำนวณแรง นั่นคือค่าเกรเดียนต์ของภาพ (Image Gradient) ทำหน้าที่เป็นตัวบอกให้คอนทัวร์รู้ว่าต้องเคลื่อนที่ไปในทิศทางใด เพื่อให้คอนทัวร์วิ่งไปยังวัตถุที่ต้องการในภาพ แอ็กทีฟคอนทัวร์แบบใช้ขอบสามารถแบ่งออกเป็น 2 ประเภทได้แก่

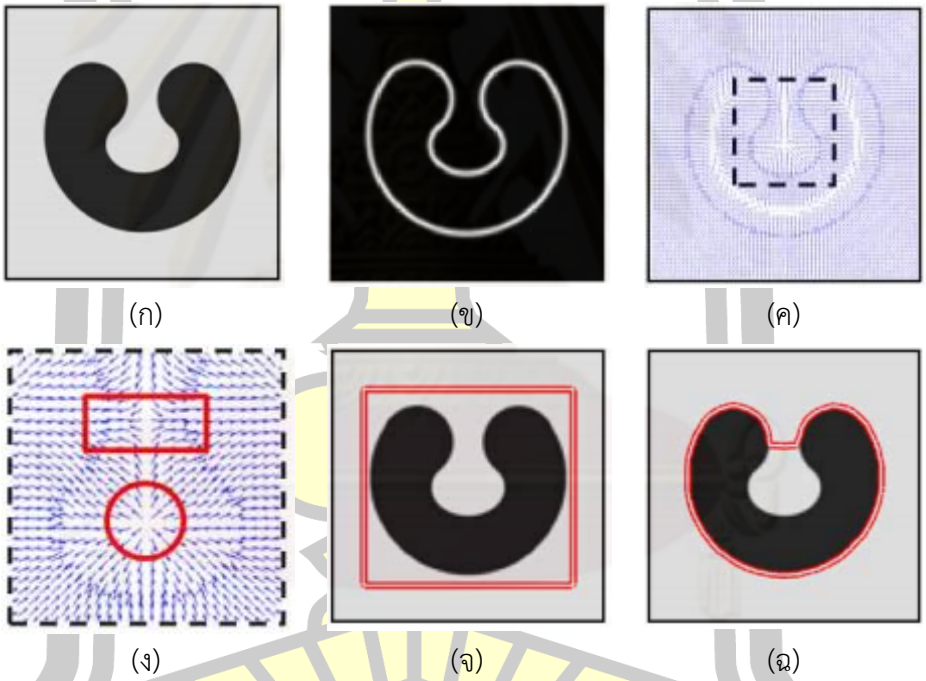
1.1.1 แบบใช้สนามเวกเตอร์ (Vector Field) สนามเวกเตอร์ คือสนามที่ประกอบไปด้วยลูกศรจำนวนมาก (มีจำนวนเท่ากับขนาดของภาพอินพุต) และมีทิศทางที่ชี้ไปยังขอบของวัตถุที่แสดงในภาพขอบ โดยลูกศรเหล่านี้เป็นตัวบอกทางให้คอนทัวร์รู้ว่าจะต้องเคลื่อนที่ไปในทิศทางใดและด้วยความเร็วเท่าใด (ซึ่งความเร็วจะขึ้นอยู่กับขนาดของลูกศร) เพื่อมุ่งไปยังขอบของวัตถุที่ต้องการได้ อย่างไรก็ตามแอ็กทีฟคอนทัวร์แบบใช้สนามเวกเตอร์ยังคงมีปัญหาหลายประการได้แก่ จุดจม (Sink Point) ซึ่งเกิดจากส่วนใดส่วนหนึ่งของวัตถุหรือพื้นหลังในภาพที่ทำให้มีค่าเกรเดียนต์สูง ๆ เกิดขึ้นในภาพขอบและส่งผลให้เกิดจุดรวมตัวกันของลูกศรขึ้นในสนามเวกเตอร์ ดังรูปที่ 41



รูปที่ 41 การแบ่งส่วนภาพปลา (ก) ภาพปลา (ข) ภาพขอบ (ค) ภาพสนามเวกเตอร์ (ง) ภาพสนามเวกเตอร์ ณ บริเวณตาปลา (จ) คอนทัวร์เริ่มต้น (ฉ) ผลการแบ่งส่วนภาพที่ได้ ที่มา [37]

ปัญหาต่อมาคือ จุดอานม้า (Saddle Point) และจุดหยุดนิ่ง (Stationary Point) [38-40] แสดงในรูปที่ 41 (ก) เป็นภาพรูปตัวอย่างที่บริเวณตรงกลางมีลักษณะคล้ายอ่าว และรูป (ข) คือภาพขอบ เมื่อใช้วิธีการที่นำเสนอโดย Li และ Acton [41] ในการคำนวณแรง จะทำให้ได้สนามเวกเตอร์ (สนามที่มีลูกศรจำนวนมาก ที่ทำหน้าที่ชี้ทางให้กับคอนทัวร์ในขณะเคลื่อนที่เพื่อวิ่งไปยังขอบของวัตถุ) แสดงใน

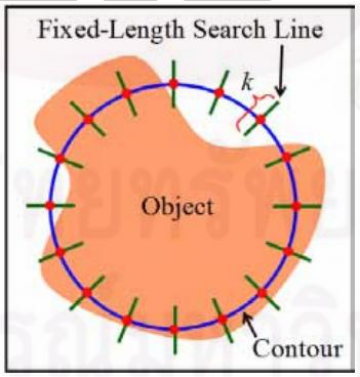
รูป (ค) และเมื่อพิจารณาเฉพาะบริเวณพื้นที่ตรงกลางของภาพในรูป (ง) จะเห็นได้ว่าลูกศรที่อยู่บริเวณอ่าวมีทิศทางชี้ออกไปโดยรอบ ซึ่งจะถูกเรียกจุดนี้ว่าจุดหยุดนิ่ง ดังแสดงภายในภาพวงกลม และจุดหยุดนิ่งนี้เป็นจุดที่คอนทัวร์ไม่สามารถเข้าถึงและเคลื่อนที่ข้ามผ่านไปได้ และจากการเกิดจุดหยุดนิ่งนี้เองก็ส่งผลให้เกิดจุดอานม้าขึ้นบริเวณคอขวดของตัวยู ดังแสดงในกรอบสี่เหลี่ยม ซึ่งจุดอานม้านี้จะทำตัวเปรียบเสมือนเป็นกำแพงที่จะขัดขวางไม่ให้คอนทัวร์เคลื่อนที่เข้าไปภายในอ่าวของรูปตัวยูได้ และเมื่อได้ทดลองวางคอนทัวร์เริ่มต้นไว้ภายนอกวัตถุแสดงในรูป (จ) และผลการแบ่งส่วนภาพที่ได้ในรูป (ฉ) ซึ่งจะเห็นได้ว่าคอนทัวร์เคลื่อนที่ไปติดกับจุดอานม้า จึงไม่สามารถแบ่งส่วนวัตถุรูปตัวอยู่ที่ต้องการได้อย่างสมบูรณ์



รูปที่ 42 ปัญหาจุดอานม้าและจุดหยุดนิ่ง (ก) ภาพรูปตัวยูที่บริเวณตรงกลางมีลักษณะคล้ายอ่าว (ข) ภาพขอบ (ค) ภาพสนามเวกเตอร์ที่ได้ (ง) จุดอ้านม้า (แสดงในสีเหลี่ยม) และขุดหยุดนิ่ง (แสดงในวงกลม) (จ) คอนทัวร์เริ่มต้น (ฉ) ผลการแบ่งส่วนภาพที่ได้
ที่มา [37]

1.1.2 แบบใช้เส้นค้าหา (Search Line) คือเส้นที่อยู่บนเส้นคอนทัวร์มีทิศทางตั้งฉากกับเส้นคอนทัวร์ และมีความยาวเท่ากับ K พิกเซลเท่ากันทุกเส้น ซึ่งความยาวของเส้นค้นหาจะไม่มีการเปลี่ยนแปลงตลอดการเคลื่อนที่ของคอนทัวร์ จึงเรียกวิธีนี้ว่าเส้นคนหาความ

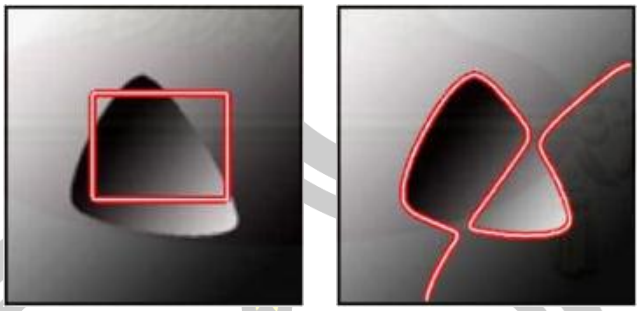
ยาวคงที่ (Fixed-Length Search Line Method) มีหน้าที่ในการทำให้รู้ว่าเส้นคอนทัวร์ต้องวิ่งไปในทิศทางใดเพื่อไปยังขอบเขตของวัตถุที่ต้องการดังรูปที่ 43



รูปที่ 43 วิธีเส้นค้นหาความยาวคงที่
ที่มา [37]

1.2 แอ็กทีฟคอนทัวร์แบบใช้บริเวณ (Region-Based) จะใช้ข้อมูลสารสนเทศบริเวณ (Regional Information) ของภาพ ในการควบคุมการเคลื่อนที่ของคอนทัวร์ โดยอาศัยข้อมูลสารสนเทศบริเวณที่แตกต่างกันระหว่างวัตถุ (Object) ที่เราต้องการพื้นหลัง (Background) ซึ่งสามารถแบ่งชนิดของแอ็กทิฟคอนทัวร์แบบบริเวณออกได้เป็น 2 ประเภทคือ

1.2.1 ข้อสนเทศบริเวณครอบคลุม (Goble Regional Information) เป็นการเลือกใช้ค่าทางสถิติต่าง ๆ มาคำนวณจากค่าความเข้ม (Intensity) ของภาพอินพุตในการบอกทางให้กับคอนทัวร์รู้ว่าต้องเคลื่อนที่และเปลี่ยนรูปร่างอย่างไร เพื่อที่จะแบ่งภาพออกได้เป็นสองส่วน คือส่วนที่เป็นวัตถุและส่วนที่เป็นพื้นหลัง โดยการใช้เส้นคอนทัวร์เป็นตัวแบ่งเขต ซึ่งภายในเส้นคอนทัวร์จะเป็นวัตถุที่ต้องการและภายนอกเส้นคอนทัวร์จะเป็นพื้นหลัง ตัวอย่างค่าทางสถิติที่นำมาใช้ เช่น ค่าเฉลี่ย (Average) [42] ค่าความแปรปรวน (Variance) [43] และฮิสโทแกรม (Histoogram) [44] เป็นต้น ในกรณีที่วัตถุและพื้นหลังเป็นเนื้อผสม (Heterogeneous Texture) ดังแสดงในรูปที่ 44

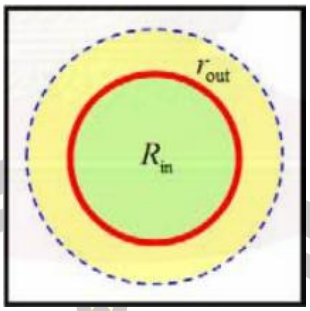


(ก) (ข)

รูปที่ 44 การแบ่งส่วนภาพในกรณีที่ทั้งวัตถุและพื้นหลังเป็นเนื้อผสมโดย (ก) คอนทัวร์เริ่มต้น (ข) ผลการแบ่งส่วนภาพที่ได้
ที่มา [37]

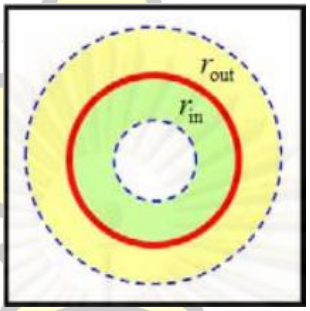
การใช้แอ็กทิฟคอนทัวร์แบบใช้บริเวณที่กล่าวมาข้างต้นจะไม่เหมาะสม เนื่องจากวิธีการต่าง ๆ เหล่านั้นจะใช้ข้อสนเทศบริเวณของทั้งภาพหรือเรียกว่าแบบครอบคลุม ดังนั้นเมื่อทดลองแบ่งส่วนภาพที่มีลักษณะเป็นเนื้อผสมนี้โดยการวางคอนทัวร์เริ่มต้นไว้ดังภาพ (ก) ทำให้ได้ผลการแบ่งส่วนภาพแสดงในภาพ (ข) ซึ่งจะเห็นได้ว่าคอนทัวร์จะพยายามแบ่งภาพออกเป็นสองบริเวณที่มีสีที่แตกต่างกันอย่างชัดเจน คือ บริเวณที่เป็นสีอ่อนและบริเวณที่เป็นสีเข้ม ทำให้ได้ผลการแบ่งส่วนภาพที่ไม่ถูกต้อง ซึ่งแท้จริงแล้ววัตถุที่ต้องการคือวัตถุที่มีรูปร่างคล้ายสามเหลี่ยมที่อยู่ตรงบริเวณกลางภาพ

1.2.2 ข้อสนเทศบริเวณท้องถิ่น (Local Regional Information) จากข้อจำกัดของวิธีการแอ็กทีฟคอนทัวร์แบบใช้ข้อสนเทศบริเวณครอบคลุม ที่ไม่สามารถแบ่งส่วนภาพที่มีวัตถุหรือพื้นหลังหรือทั้งสองอย่างเป็นเนื้อผสมได้ จึงได้มีหลาย ๆ งานวิจัยได้มีการปรับปรุงและพัฒนาวิธีการแอ็กทีฟคอนทัวร์แบบใช้ข้อสนเทศบริเวณท้องถิ่นขึ้นมา ด้วยการเลือกใช้บริเวณของภาพเพียงแค่บางส่วน ซึ่งไม่ได้มีการใช้บริเวณทั้งหมดของภาพ เช่น แบบที่ 1 นำเสนอโดย Mille และ Cohen [45] ใช้ค่าความเข้มเฉลี่ยที่อยู่ภายในคอนทัวร์และส่วนที่อยู่ภายนอกคอนทัวร์เฉพาะในแถบที่ขยายออกไปจากคอนทัวร์ในช่วงที่กำหนดเท่านั้น ซึ่งเรียกว่าแถบนอก (Outer Band) ดังรูปที่ 45



รูปที่ 45 การใช้ขอบสนเทศบริเวณที่อยู่ภายในคอนทัวร์และแถบนอก
ที่มา [37]

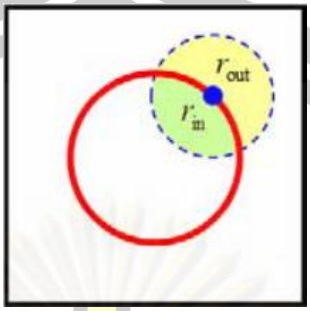
จากวิธีการนี้สามารถแบ่งส่วนภาพในกรณีที่วัตถุเป็นเนื้อเดียวกันวางตัวอยู่บนพื้นหลังเป็นเนื้อผสมได้ แต่ไม่สามารถแบ่งส่วนภาพในกรณีที่วัตถุเป็นเนื้อผสมได้ แบบที่ 2 นำเสนอโดย Ronfard [46] เป็นการใช้ข้อสนเทศบริเวณท้องถิ่นของภาพทั้งที่อยู่ในแถบใน (Inner Band) และแถบนอก (Outer Band) ของคอนทัวร์ดังรูปที่ 46



รูปที่ 46 การใช้ข้อสนเทศบริเวณที่อยู่ภายในแถบในและแถบนอก
ที่มา [37]

วิธีการนี้สามารถแบ่งส่วนภาพได้ในกรณีที่วัตถุเป็นเนื้อผสมที่บริเวณใกล้เคียงกับขอบของวัตถุทั้งหมดที่อยู่ด้านในเป็นเนื้อเดียวกัน และบริเวณใกล้เคียงกับขอบวัตถุทั้งหมดที่อยู่ด้านนอกต้องเป็นเนื้อเดียวกันด้วยเช่นกัน แบบที่ 3 นำเสนอโดย Lankton และ Tannenbaum [47] วิธีการนี้เรียกว่า LRAC (Localizing Region-base Active Contour) คือบนเส้นคอนทัวร์จะมีจุดศูนย์กลางของวงกลมและในวงกลมจะมีการแบ่งเป็นแถบในและแถบนอก ซึ่งในรูปจะนำเสนอเพียงจุดเดียว แต่ความเป็นจริงนั้นจะมีวงกลมหลายจุดบนเส้นคอนทัวร์ อีกทั้งยังมีรัศมีที่มีขนาดเท่ากันทุกวง และมีการ

คำนวณค่าความเข้มเฉลี่ยของตัวเอง ทำให้จุดต่าง ๆ บนคอนทัวร์มีความเป็นอิสระในการเคลื่อนที่เข้าหาวัตถุดังรูปที่ 47



รูปที่ 47 การใช้ข้อสนเทศบริเวณที่อยู่ภายในวงกลมของแต่ละจุดบนคอนทัวร์
ที่มา [37]

วิธีการนี้สามารถแบ่งภาพในกรณีที่บริเวณใกล้ ๆ กับขอบของวัตถุมีหลายเฉดสีได้ เนื่องจากวงกลมมีรัศมีที่มีขนาดเท่ากันทุกวง จึงส่งผลให้วิธีการนี้มีข้อจำกัดคล้ายกับวิธีการแอ็กทิฟคอนทัวร์แบบใช้เส้นค้นหาความยาวคงที่ นั่นคือปัญหาในเรื่องการหาค่ารัศมีที่เหมาะสมกับตำแหน่งขอบคอนทัวร์เริ่มต้น ซึ่งหากกำหนดค่ารัศมีของวงกลมเล็กเกินไป คอนทัวร์จะมีชวงการเคลื่อนที่เข้าหาวัตถุที่แคบ ทำให้คอนทัวร์ไม่สามารถเคลื่อนที่ไปยังวัตถุที่ต้องการได้อย่างสมบูรณ์

แอ็กทิฟคอนทัวร์แบบดั้งเดิม (Traditional Active Contour:TAC) [48] เป็น แอ็กทิฟคอนทัวร์แบบแรกที่ถูกนำเสนอขั้นมาโดย Kass และคณะ [49] วิธีการนี้เป็นแอ็กทิฟคอนทัวร์แบบใช้ขอบเนื่องจากมีการนำค่าเกรเดียนต์ของขอบภาพ (Edge Map) ในการนำมาคำนวณแรงสำหรับขับเคลื่อนคอนทัวร์ซึ่งส่งผลให้วิธีการนี้มีข้อจำกัดหลายประการ เช่น ไม่สามารถเคลื่อนที่เข้าไปในส่วนโค้งหรือส่วนเว้ามาก ๆ ของวัตถุได้ และไม่สามารถเกาะติดกับขอบที่ไม่ชัดเจนได้ พร้อมทั้งยังไม่ทนทานต่อสัญญาณรบกวน และมีช่วงการเคลื่อนที่เข้าหาวัตถุ (Capture Range) ที่จำกัด ทำให้ในการวางตำแหน่งเริ่มต้น (Initial Position) ของคอนทัวร์นั้นจำเป็นต้องวางใกล้กับวัตถุที่เราต้องการ จึงจะทำให้คอนทัวร์สามารถเคลื่อนที่ไปยังวัตถุที่เราต้องการได้ ซึ่งปัญหานี้เกิดจากบริเวณที่ไกลจากขอบของวัตถุ ค่าเกรเดียนต์จะมีค่าน้อยมาก ทำให้เมื่อส่วนใดส่วนหนึ่งของคอนทัวร์ตกอยู่ภายในบริเวณเหล่านี้ คอนทัวร์จะไม่มีแรงขับเคลื่อนไปยังขอบของวัตถุได้

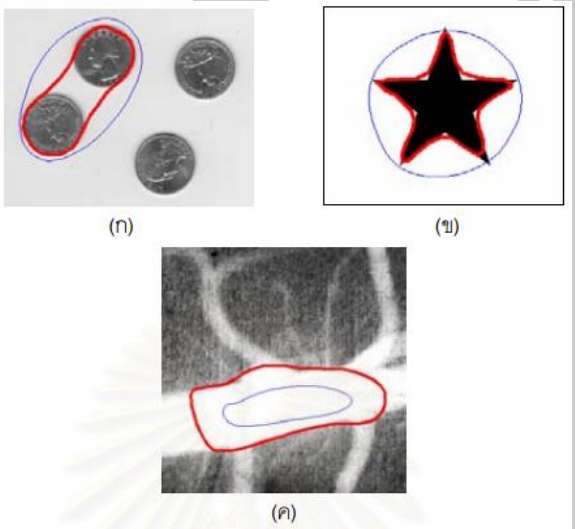
2. วิธีเลเวลเซต (Level Set Method) [50] หลังจาก Kass และคณะได้นำเสนอ Active Contour Models หรือ Snake วิธีการนี้ก็ยังคงมีปัญหาคือ เส้นคอนทัวร์เริ่มต้น (Initial Contour) จำเป็นต้องอยู่ใกล้ขอบภาพ ทำให้ผู้ใช้ต้องกำหนดเส้นคอนทัวร์เริ่มต้นใหม่ทุกครั้ง

ในการแบ่งส่วนภาพเพื่อให้ได้ผลการแบ่งส่วนที่ถูกต้อง ทั้งนี้นอกจากปัญหาดังกล่าวมาข้างต้นยังคงมีปัญหาอื่น ๆ จากงานวิจัยของ Cohen และคณะ [51] งานวิจัยของ Malladi และคณะ [52] และงานวิจัยของ Caselles และคณะ [53] อีกดังนี้

1) ในกรณีที่มีวัตถุที่สนใจหลายชิ้น การใช้เส้นคอนทัวร์ปิดเพียงเส้นเดียวไม่สามารถแบ่งส่วนวัตถุที่สนใจได้ทั้งหมดดังรูปที่ 48 (ก)

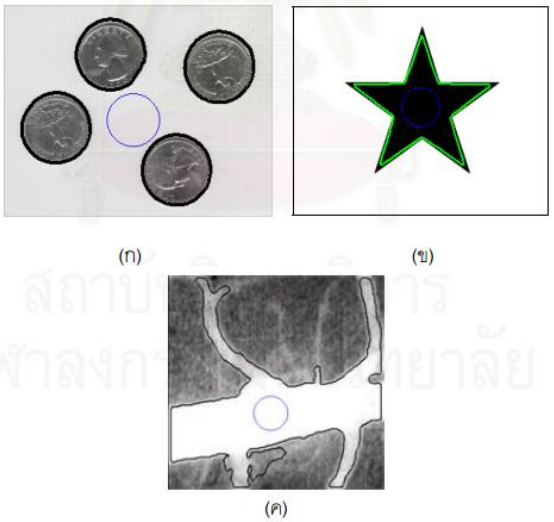
2) ไม่สามารถแบ่งส่วนภาพได้อย่างถูกต้องในกรณีที่วัตถุที่สนใจมีมุมแหลมดังรูปที่ 48 (ข)

3) เส้นคอนทัวร์ไม่สามารถยื่นออกตามบริเวณที่มีการยื่นออกเป็นกิ่งแง่ง เช่น ภาพของเส้นเลือดในรูปที่ 48 (ง)



รูปที่ 48 ปัญหาของการแบ่งส่วนภาพโดยใช้ Active Contour ตามวิธีการของ Kass ที่มา [50]

ต่อมาได้มีนักวิจัยหลายกลุ่ม ได้แก่งานวิจัยของ Chan และคณะ [42] งานวิจัยของ Zhang และคณะ [54] และงานวิจัยของ Lankton และคณะ [47] ที่นำเสนอเทคนิคเลเวลเซตเข้ามาช่วยในการแก้สมการพลังงานของ Kass ซึ่งเทคนิคเลเวลเซต มีข้อดีคือสามารถแก้ไขปัญหาทั้งสามประการที่เกิดขึ้นกับวิธีการของ Kass ได้ดังแสดงในรูปที่ 49



รูปที่ 49 การแบ่งส่วนภาพโดยวิธีการเลเวลเซต
ที่มา [50]

แต่อย่างไรก็ตาม Active Contour มันจะเกิดปัญหาในการแสดงเส้นโค้งแบ่งส่วนซึ่งมีดังนี้

1) ในการเคลื่อนที่ของเส้นโค้งอย่างมีเสถียรภาพจำเป็นต้องใช้ Time Step ที่น้อย

2) ในบริเวณที่มีความโค้งสูงจุดที่ใช้แสดงเส้นโค้งมักจะกระจุกตัวรวมกันทำให้การคำนวณไม่มีเสถียรภาพ จำเป็นต้องมีการปรับพารามิเตอร์ใหม่ซึ่งทำให้เส้นโค้งที่ได้มีความคลาดเคลื่อนไปจากขอบจริงของวัตถุ

3) จะเกิดปัญหาในการสร้างเส้นโค้งหลังจากที่ที่เส้นโค้งมีการแยกตัวหรือรวมตัวกัน

4) ไม่สามารถแสดงความเป็นเหลี่ยมมุมได้อย่างถูกต้อง

ซึ่งต่อมางานวิจัยของ Osher และ Sethian [55] ได้นำเสนอวิธีการ Level Set เพื่อแก้ปัญหาในการเคลื่อนที่ของเส้นโค้ง โดยกำหนดนิยามการแสดงเส้นโค้งขึ้นมาใหม่ ซึ่งวิธีการนี้จะไม่เกิดปัญหาดังที่ได้กล่าวมาข้างต้น ทั้งนี้วิธีการเลเวลเซตมีข้อดีดังนี้

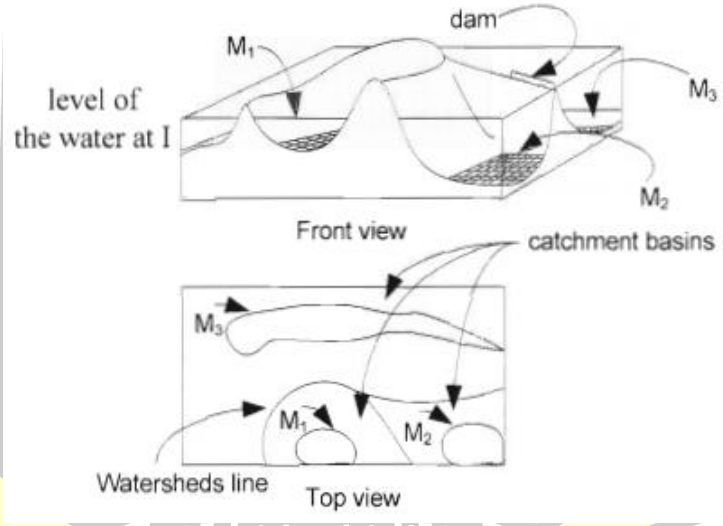
1) การเคลื่อนที่ของเส้นโค้งเปลี่ยนจากการปรับพิกัดโดยตรงมาเป็นการปรับ Amplitude ที่พิกัดต่าง ๆ ทำให้การคำนวณมีความเสถียรภาพมากกว่า

2) ไม่เกิดปัญหาในกรณีที่เส้นโค้งจำเป็นต้องมีการรวมตัวหรือแยกตัว

3) เส้นโค้งสามารถแสดงความเป็นเหลียมมุมได้ดีกว่า

4) สามารถนำไปใช้กับข้อมูลบนพิกัดที่มีขนาดเท่าใดก็ได้

(2) Region-Based Segmentation เป็นวิธีการแยกองค์ประกอบของภาพโดยพิจารณาจากตำแหน่งของพิกเซลและความเหมือนกันของคุณสมบัติภายในพื้นที่ โดยการพิจารณาถ้าพิกเซลที่อยู่ติดกันและมีคุณสมบัติเหมือนกันจะถูกจัดให้อยู่ในกลุ่มเดียวกัน ซึ่งข้อดีของวิธีการนี้จะทำให้ได้พื้นที่ที่ต่อเนื่องกัน วิธีการที่นิยมนำมาใช้ในกลุ่มนี้ คือกระบวนการวอเตอร์เชด (Watersheds) ซึ่งมีงานวิจัยของ Cui และคณะ [56] งานวิจัยของ Pinto และคณะ [57] รวมทั้งงานวิจัยของ Amankwah และ Aldrich [58] ได้นำวิธีการนี้มาประยุกต์ใช้ในการแบ่งส่วนภาพ โดยแนวคิดของกระบวนการวอเตอร์เชด [59] เป็นการนำภาพมาพิจารณาในรูปแบบ 3 มิติ คือมีตำแหน่งพิกัด (แกน x และแกน y) และระดับเทาที่ตำแหน่งพิกัดนั้น ๆ ซึ่งภาพที่ปรากฏจะแสดงลักษณะที่มีความคล้ายคลึงกับแผนภูมิประเทศดังแสดงในรูปที่ 50



รูปที่ 50 ตัวอย่างภาพที่มีลักษณะเป็นแผนภูมิประเทศ
ที่มา [60]

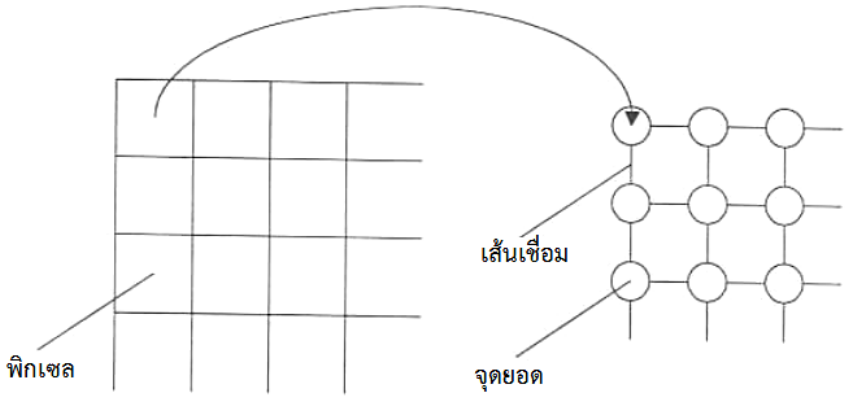
จากรูปที่ 50 บริเวณ M_1 M_2 และ M_3 ซึ่งถูกแรงเงาจะแสดงถึงพื้นที่ระดับต่ำ (Minima) เมื่อนำมาเปรียบเทียบกับบริเวณพื้นที่รอบ ๆ ข้างบริเวณพื้นที่ M_2 จะมีระดับพื้นที่น้อยที่สุด บริเวณพื้นที่ M_3 จะมีระดับความสูงมากกว่า M_2 ส่วนบริเวณพื้นที่ M_1 จะมีพื้นที่สูงที่สุด ทั้งนี้หากสมมุติให้ผิดพลาด! ไม่พบแหล่งการอ้างอิง เป็นวัตถุที่สนใจ เราจะนำวัตถุนั้นมาทำการเจาะรูเล็ก ๆ ในบริเวณ M_1 M_2

และ M_3 อย่างละหนึ่งจุด โดยรูที่เจาะนั้นจะต้องทะลุไปถึงด้านล่างสุดของวัตถุ ต่อมาจึงเอาวัตถุที่ได้เจาะรูแล้วไปวางลงในสระน้ำ วัตถุนี้จะค่อย ๆ จมลงไปด้วยความเร็วคงที่ น้ำจากสระน้ำจะค่อย ๆ ไหลผ่านรูที่ถูกเจาะไว้ โดยน้ำจะไหลเข้าไปในพื้นที่ M_2 ก่อนเนื่องจากเป็นจุดที่อยู่ต่ำที่สุด ในขณะที่วัตถุจมลงไปก็จะคล้ายกับว่าระดับน้ำที่ท่วมวัตถุนั้นสูงขึ้น เมื่อระดับน้ำสูงจนถึงระดับความสูง I จะมีบางส่วนของผ่านไหลผ่านเข้ามายังรูที่เจาะไว้ของ M_2 จะล้นข้ามผ่านไปยัง M_3 ซึ่งจะต้องหลีกเลี่ยงการที่น้ำไหลข้ามผ่านมานี้โดยการสร้างเขื่อนกั้นน้ำเอาไว้ ที่ระดับความสูงของน้ำที่ระดับอื่น ถ้าหากเกิดเหตุการณ์เช่นนี้อีกกับระดับความสูงของน้ำที่ระดับ I ก็ให้สร้างเขื่อนกั้นน้ำขึ้นมากั้นเช่นเดิม กระบวนการทำให้จมจะดำเนินการอย่างนี้ไปเรื่อย ๆ จนผลสุดท้ายของกระบวนการที่เกิดขึ้นคือ เราจะสามารถแยกพื้นที่ออกได้เป็นสามส่วน โดยจะมีเขื่อนกั้นน้ำเป็นตัวแบ่งแยกพื้นที่ซึ่งเราจะเรียกเส้นเขตที่ถูกสร้างขึ้นเป็นเขื่อนกันน้ำนี้ว่าเส้นวอเตอร์เชด (Watershed Line)

(3) ทฤษฎีกราฟ (Graph Theory) [61] ได้ถูกนำมาประยุกต์ใช้ในการแบ่งส่วนภาพ ซึ่งเป็นวิธีการที่มีการนำเอาข้อมูลส่วนใหญ่ของภาพมาใช้เป็นเกณฑ์ในการตัดสินใจในการแยกหรือรวมแต่ละส่วนย่อย ๆ เข้าด้วยกัน ดังนั้นการแบ่งส่วนภาพโดยใช้ทฤษฎีกราฟจึงสามารถให้ภาพที่มีความกลมกลืนกันในแต่ละพื้นที่ย่อย ตามลักษณะโครงสร้างของภาพเดิมและยังมีคุณสมบัติเด่น คือสามารถที่จะกำหนดจำนวนพื้นที่ย่อยได้ตามต้องการ โดยวิธีการประยุกต์ใช้ทฤษฎีกราฟนี้จำเป็นต้องมีการจัดการข้อมูลของภาพให้อยู่ในลักษณะของกราฟ (Mapping Image Onto Graph) เสียก่อน Morris และ Lee [62, 63] ได้อธิบายวิธีการจัดการข้อมูลนี้ด้วยการพิจารณาถึงความเข้มของค่าระดับสีเทาของ แต่ละพิกเซล จะถูกแปลงไปเป็นค่าจุดยอดของกราฟในแต่ละจุด ทั้งนี้ถ้าภาพที่มีความละเอียด 256 x 256 พิกเซล ก็จะถูกแปลงเป็นกราฟที่มีจุดยอด 256 x 256 จุดด้วยเช่นกัน ดังนั้นเมื่อให้ความเข้มของค่าระดับสีเทาของจุดที่ตำแหน่ง (x, y) คือ $f(x, y)$ แล้วค่าน้ำหนักในจุดยอดของกราฟ V_i จะเป็น

$$V_i = f(x, y) \tag{2.31}$$

โดยที่ x, y จะถูกแปลงให้เป็น i ในลักษณะจุดต่อจุด (One-to-One Mapping) จากนั้นจะทำการหาค่าน้ำหนักของตัวเชื่อมในแต่ละจุดยอดต่าง ๆ ซึ่งในที่นี้จะให้การเชื่อมต่อแบบ 4 ทิศ ทางกระทำกับจุดยอดของกราฟเฉพาะจุดยอดที่อยู่ใกล้กันที่สุดเท่านั้น เพื่อเป็นการลดขั้นตอนและขนาดของข้อมูลที่จะเกิดขึ้นขณะประมวลผลดังรูปที่ 51 แสดงการแปลงข้อมูลภาพไปเป็นกราฟ

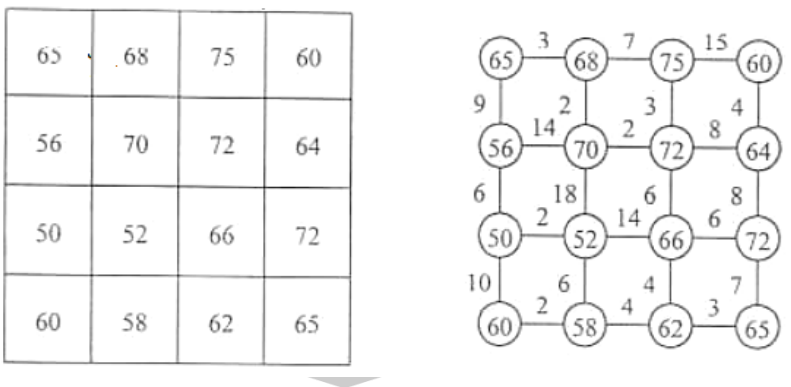


รูปที่ 51 การจัดข้อมูลภาพให้อยู่ในลักษณะกราฟ
ที่มา [61]

ค่าน้ำหนักของเส้นเชื่อมจะหาได้จากค่าสัมบูรณ์ ของค่าความแตกต่างระหว่างจุดยอดของกราฟที่อยู่ติดกัน ซึ่งเป็นการวัดความเหมือนหรือความใกล้เคียงกันของความเข้มระหว่างพิกเซลที่มีเส้นเชื่อมต่อกัน โดยกำหนดให้ $e_{i,j}$ เป็นความเข้มของเส้นเชื่อมหาได้จากสมการที่ 2.32

$$e_{i,j} = \left| v_i - v_j \right| \tag{2.32}$$

จากสมการที่ 31 และ 32 สามารถนำมาสร้างกราฟที่มีค่าจุดยอดและค่าน้ำหนักเส้นเชื่อมแสดงเป็นตัวอย่างได้ดังรูปที่ 52



รูปที่ 52 ค่าจุดยอดและน้ำหนักเส้นเชื่อมของกราฟที่ขนาดภาพ 4 x 4 พิกเซล
ที่มา [61]

ในขั้นต่อมาหลังจากที่เราได้กราฟแล้วจะเป็นการแบ่งส่วนภาพด้วย การหาซ็อตเตสท์สแพนนิ่งทรี (Shortest Spanning Tree : SST) [62, 63] ซึ่งเป็นการเปลี่ยนเส้นเชื่อมที่มีค่าน้ำหนักน้อยที่สุดให้มีค่าเป็นศูนย์ และทำการเฉลี่ยค่าน้ำหนักของจุดยอดคู่ที่ถูกเปลี่ยนค่าเส้นเชื่อมให้มีค่าเท่ากัน กระทำการเปลี่ยนแปลงค่าเส้นเชื่อม และค่าเฉลี่ยค่าน้ำหนักจุดยอดต่อไป จนกระทั่งค่าน้ำหนักของจุดยอดของทั้งภาพมีค่าเท่ากันเส้นเชื่อมที่ยังคงอยู่จะเรียกว่า ซ็อตเตสท์สแพนนิ่งทรี ซึ่งรายละเอียดของวิธีการมีดังต่อไปนี้

1. จัดเรียงลำดับค่าน้ำหนักเส้นเชื่อมของกราฟจากค่าต่ำไปหาค่าสูง
2. ตรวจหาเส้นเชื่อมที่มีค่าน้ำหนักน้อยที่สุด
3. เปลี่ยนค่าน้ำหนักเส้นเชื่อมจุดยอดที่ตรวจพบให้มีค่าเป็นศูนย์
4. เฉลี่ยค่าน้ำหนักของกลุ่มจุดยอดระหว่างเส้นเชื่อม ดังสมการที่ 2.33

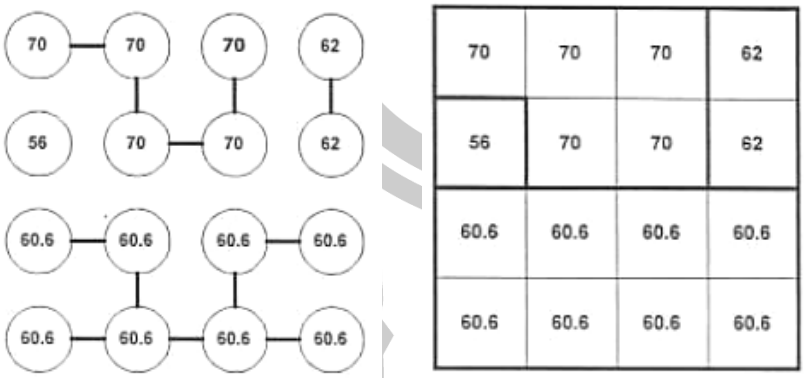
$$Vertex_{mean} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{2.33}$$

เมื่อ N คือ จำนวนจุดยอดที่ถูกปรับค่าน้ำหนักเส้นเชื่อม

x_i คือ ค่าน้ำหนักจุดยอดของกราฟแต่ละจุดที่ถูกปรับค่าน้ำหนักจุดยอด

5. เมื่อเชื่อมเส้นและเฉลี่ยค่าน้ำหนักจุดยอดแล้วเกิดวนรอบให้ตัดเส้นเชื่อมที่ทำให้เกิดวนรอบออก
6. ทำซ้ำขั้นตอนที่ 1 จนกระทั่งค่าน้ำหนักของจุดยอดมีเพียงค่าเดียว

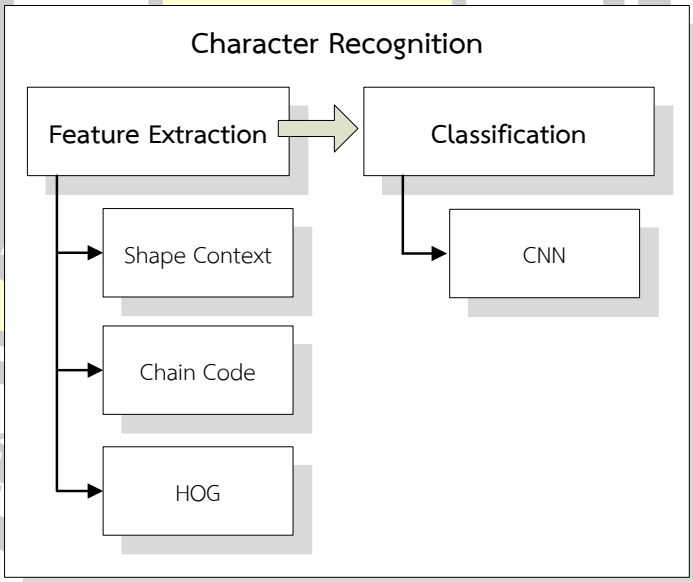
หลังจากที่ได้ซ็อตเตสต์สแพนนิ่งทรีแล้ว ก็ทำการแบ่งส่วนภาพโดยการตัดเส้นเชื่อมที่มีค่าสูงสุด นั้นก็คือตัวเชื่อมที่ถูกเชื่อมเข้าไปตอนท้ายสุด และรองลงไป ถ้าต้องการส่วนของภาพ N ส่วนจะต้องตัดเส้นเชื่อมซ็อตเตสต์สแพนนิ่งทรี N-1 ครั้ง



(ก) (ข)

รูปที่ 53 แสดงการตัดเส้นเชื่อมของช็อตเตสต์สแพนนิ่งทรี
ที่มา [64]

3) การรู้จำตัวอักขระ (Character Recognition) เป็นกระบวนการที่ดำเนินการต่อจากกระบวนการแบ่งส่วนข้อความในภาพ ซึ่งกระบวนการนี้จะมีขั้นตอนการดำเนินงานโดยทั่วไปอยู่ 2 ขั้นตอนหลักคือ การสกัดคุณลักษณะ (Feature Extraction) และ การจำแนกประเภท (Classification) โดยในแต่ละขั้นตอนจะมีเทคนิคสำหรับขั้นตอนนั้น ๆ ทั้งนี้สามารถแสดงได้ดังรูปที่ 54



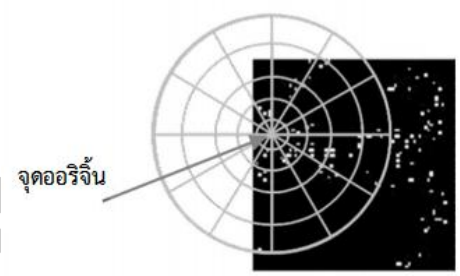
รูปที่ 54 เทคนิคในแต่ละขั้นตอนของกระบวนการรู้จำ

(1) การสกัดคุณลักษณะ (Feature Extraction) คือกระบวนการหรือวิธีการอย่างใดอย่างหนึ่งที่นำมาใช้ในการพิจารณาถึงคุณลักษณะเด่นของภาพ เช่น คุณลักษณะทางพื้นผิว คุณลักษณะทางรูปร่าง (พื้นที่, เส้นรอบวง) หรือคุณลักษณะทางสี เป็นต้น การสกัดคุณลักษณะออกจากภาพเป็นขั้นตอนที่สำคัญที่จะนำไปสู่การดึงข้อมูลออกมาจากภาพ โดยวิธีการที่นิยมใช้ในการสกัดคุณลักษณะของตัวอักขระออกจากภาพมีดังต่อไปนี้

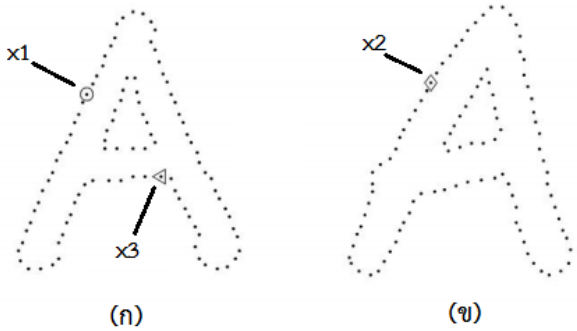
1. Shape Context จากงานวิจัยของ Srisuk และคณะ [65] และงานวิจัยของ Belongie และคณะ [66] มีการแนะนำวิธีการรู้จำวัตถุด้วยการวัดความคล้ายของวัตถุซึ่งที่เรียกว่า Shape Context ซึ่งเป็นวิธีการที่จะพิจารณาถึงรูปร่างของวัตถุที่แสดงอยู่บนภาพ โดยวิธีการนี้จะดำเนินการหลังจากที่ได้ขอบภาพแล้ว Shape Context คือการกำหนดเซตของจุดบนภาพแทนด้วย $P = \{ p_1, p_2, ..., p_n, \}, p_i \in \square^2$ ทำให้สามารถคำนวณค่าของล็อกโพล่าฮิสโตรแกรม (Log-Polar Histogram หรือ Log-Polar) ซึ่งแสดงความสัมพันธ์ระหว่างจุดใด ๆ บนภาพกับจุดอื่นจำนวน $n-1$ บนส่วนที่เป็นขอบวัตถุดังสมการที่ 2.34

$$h_i(k) = \#\{q \neq p_i \,|\, (q - p_i) \in bin(k)\} \tag{2.34}$$

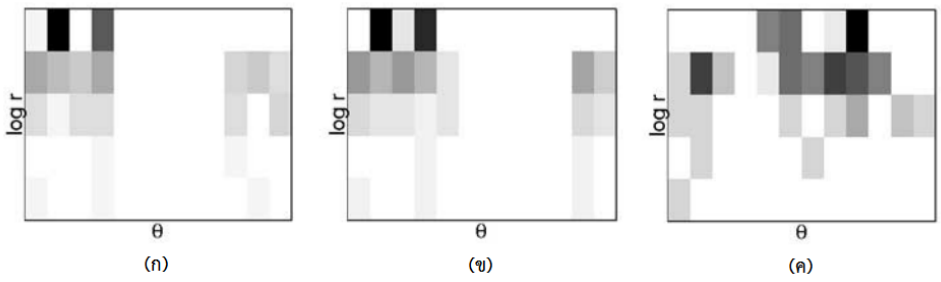
กำหนดให้ค่าฮิสโตรแกรม h_i คือ Shape Context ของ p_i ซึ่งค่า Shape Context แต่ละค่า คือค่าล็อกโพล่าฮิสโตรแกรมของพิกัดของชุดข้อมูลซึ่งถูกวัดโดยอ้างอิงจากการสุ่มจุดออริจิ้น (Origin) โดยทำการนับจำนวนจุดที่พบอยู่ภายในแต่ละช่อง (bin-k) ลักษณะของต้นแบบล็อกโพล่าฮิสโตรแกรมจะเป็นช่องวงกลมทั้งหมด 5 วงแต่ละวงจะถูกแบ่งออกเป็น 12 ส่วน ๆ ละ 30 องศาแสดงดังรูปที่ 55 และสำหรับรูปที่ 56 (ก)(ข) และ (ค) คือ Shape Context ที่อธิบายถึงความหนาแน่นหรือความสัมพันธ์กันระหว่างจุดอ้างอิงกับจุดอื่น ๆ ซึ่งสามารถสังเกตได้ว่าล็อกโพล่าฮิสโตรแกรมนั้นจะมีความอ่อนไหวต่อจุดที่อยู่ใกล้กับจุดอ้างอิงที่ใกล้เคียงกันกับรูปร่างของวัตถุที่มีความคล้ายกัน ซึ่งจะสามารถเปรียบเทียบได้ดังรูปที่ 57 (ก) และ (ข) ที่มีจุดอ้างอิงที่เดียวกันรูปร่างของวัตถุที่เหมือนกันหรือคล้ายกันจะมีค่าล็อกโพล่าฮิสโตรแกรมที่มีคุณลักษณะที่คล้ายกัน แต่ในทางกลับกันหากมีการกำหนดจุดอ้างอิงต่างกันจะทำให้ภาพของค่าล็อกโพล่าฮิสโตรแกรมก็จะแตกต่างกันโดยสิ้นเชิงดังรูปที่ 57 (ค)



รูปที่ 55 ล็อกโพล่าฮิสโตรแกรม
ที่มา [67]



รูปที่ 56 (ก) พิกัดของตัวอักขระ A แบบแรก (ข) พิกัดของตัวอักขระ A แบบที่สอง
ที่มา [66]



รูปที่ 57 (ก) ค่า Shape Context ของอักขระ A ณ จุดอ้างอิง X1 (ข) ค่า Shape Context ณ จุดอ้างอิง X2 และ (ค) ค่า Shape Context ณ จุดอ้างอิง X3
ที่มา [66]

การทำล็อกโพล่าฮิสโตรแกรมเพื่อใช้ในการนับจำนวนจุดภาพที่อยู่ในแต่ละช่องของล็อกโพล่าฮิสโตรแกรม ซึ่งมีทั้งหมด 60 ช่อง (5 วง x 12 ช่อง) ที่จะได้ข้อมูลของแต่ละจุด โดยวงที่ 1 นับจากวงในสุดไล่มาข้างนอกสุด และช่องที่ 1 นับจากองศาที่ 0 ไล่มาองศาที่ 360 (ตามเข็มนาฬิกา) จะได้ข้อมูลของจุด X1 ดังรูปที่ 58

วงที่ ↓ ช่องที่ →

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	2	0	0	4	0	2	1	5	1	1	0	0
3	0	0	0	3	0	1	6	0	0	3	0	0
4	0	0	0	1	3	0	7	3	0	0	0	0
5	0	0	0	0	1	1	5	0	0	1	0	0

รูปที่ 58 ตัวอย่างล็อกโพล่าฮิสโตรแกรม
ที่มา [67]

จากรูปที่ 58 เป็นการหาค่าล็อกโพล่าฮิสโตรแกรมของจุดภาพ X1 ในรูปที่ 56 ถ้ามีการสุ่มจุดภาพ 100 จุดก็จะได้ข้อมูลดังรูปที่ 58 ทั้งหมด 100 ชุด และในการพิจารณาความคล้ายคลึงกันของจุดภาพสามารถทำได้ด้วยการกำหนดให้ p_i เป็นกลุ่มของจุดที่อยู่บนรูปร่างซึ่งได้จากการสุ่มจากเส้นขอบภาพ และ q_j เป็นกลุ่มของจุดที่อยู่บนรูปร่างที่สอง และให้ $C_{ij} = C(p_i, q_j)$ ที่แทนด้วยระยะห่างของจุดสองจุด ดังนั้นสามารถคำนวณความเหมือนของ Shape Context ดังสมการที่ 2.35

$$C_{ij} \equiv C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{k}\frac{\left[h_i(k) - h_j(k)\right]^2}{h_i(k) + h_j(k)} \qquad (2.35)$$

เมื่อกำหนดให้

$h_i(k)$ คือค่าที่ปรับแต่ง (Normalized) ฮิสโตรแกรมของช่อง (bin) ที่ k ที่จุด p_i

$h_j(k)$ คือค่าที่ปรับแต่ง (Normalized) ฮิสโตรแกรมของช่อง (bin) ที่ k ที่จุด p_j

ในการพิจารณาความเหมือนของจุดจะทำการคำนวณกับทุก ๆ จุดระหว่างรูปร่างที่หนึ่งและที่สองเพื่อหาค่าต่ำที่สุดซึ่งจะเป็นการบ่งบอกว่าจุดภาพในรูปร่างที่หนึ่งมีความคล้ายกับจุดใดในรูปร่างที่สองแบบจุดต่อจุด และทำการจับคู่ระหว่างจุดของรูปร่างทั้งสองดังรูปที่ 59

รูปที่ 59 ผลของความคล้ายคลึงกันระหว่างรูปร่างที่หนึ่งและที่สอง

ทั้งนี้ในการประเมินค่าความเหมือนของรูปร่างที่จะพิจารณาว่ารูปร่างทั้งสอรูปร่างเหมือนกันหรือไม่จะใช้การประเมินแบบ Thin Plate Spline (TPS) [66] โดยความเหมือน ระหว่างรูปร่าง p และ q จะประเมินได้จากการหาค่าผลรวมที่ดีที่สุดของจุดใน Shape Context ดังสมการที่ 2.36

$$D_{sc} = \frac{1}{n}\sum_{p \in P} \arg\min c(p, T(q)) + \frac{1}{n}\sum_{p \in Q} \arg\min c(p, T(q)) \qquad (2.36)$$

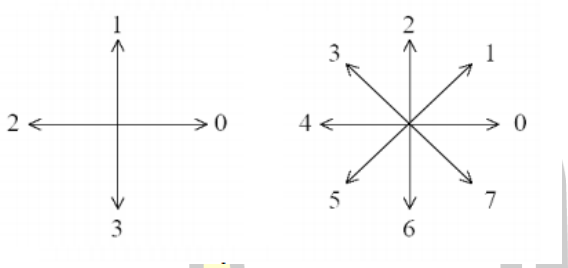
$T(.)$ แทนด้วยการประเมินค่าแบบ TPS โดยคำนวณได้จากสมการที่ 37 และ 38

$$T(x, y) = (f_x(x, y), f_y(x, y)) \qquad (37)$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} W_i U(\|(x_i, y_i) - (x, y)\|) \qquad (38)$$

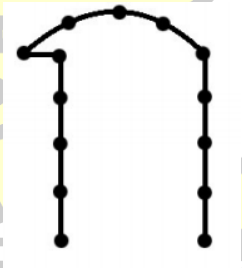
2. Chain Code คือกระบวนการที่แสดงรูปร่างของวัตถุ โดยงานวิจัยของ Siriboon และคณะ [68] งานวิจัยของ Panggabean และคณะ [69] หรืองานวิจัยของ Siddiqi และ Vincent [70] ได้มีการนำ Chain Code มาประยุกต์ใช้ในการอธิบายคุณลักษณะของตัวอักขระในเทอมของลำดับของจุดที่มีความต่อเนื่องกันไปเป็นตัวอักขระ โดยอาศัยการเปลี่ยนแปลงทิศทางของ

พิกเซลว่าจะเป็นไปในทิศทางใด ซึ่งโดยทั่วไปจะเป็นแบบ 4 ทิศทาง หรือ 8 ทิศทาง ดังรูปที่ 60 การหาทิศทางจะหาในทิศทางที่ทวนเข็มนาฬิกา



รูปที่ 60 ทิศทางในการกำหนด Chain code
ที่มา [71]

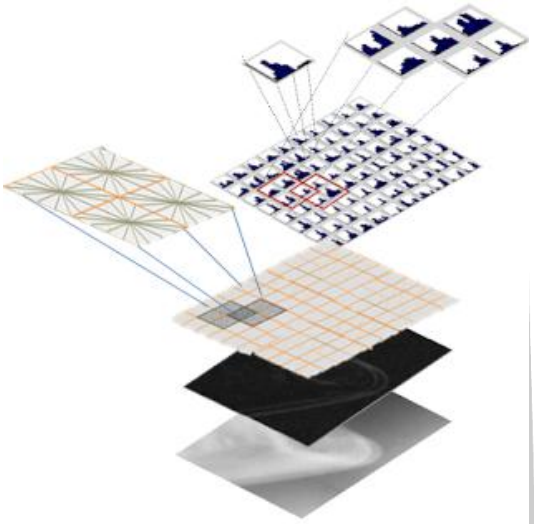
ซึ่งก่อนที่จะทำ Chain Code ของภาพใด ๆ จะต้องทำการหาขอบของภาพก่อน (Edge Detection) ทั้งนี้จะสามารถใช้การหาขอบแบบใดก็ได้ เช่น Robert Prewitt หรือ Canny เป็นต้น หลังจากทำการหาขอบของภาพแล้วให้พิจารณาเฉพาะพิกเซลที่เป็นเส้นขอบเท่านั้นในการทำ Chain Code ตัวอย่างของตัวอักขระที่ถูกแปลงเป็น Chain Code ของตัวอักขระ "ก" แบบ 8 ทิศทางจะได้รหัสคือ 2 2 2 2 2 2 4 1 1 7 7 6 6 6 6 6 ดังรูปที่ 61 ซึ่งรหัสตัวเลขนี้คือคุณลักษณะของตัวอักขระ "ก"



รูปที่ 61 การอ่านจุดพิกเซลจากตัวอักขระ ก
ที่มา [71]

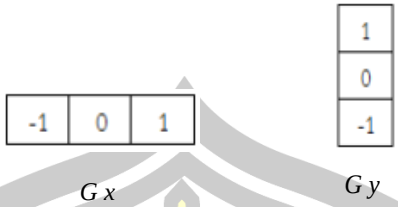
นอกจากการอธิบายคุณลักษณะเป็นรหัสตัวเลขที่ได้จากการหาทิศทางด้วย Chain Code แล้ว ยังมีการนำ Chain Code มาประยุกต์ใช้ในการอธิบายคุณลักษณะ [70] ทิศทางด้วยการเก็บทิศทางที่ได้จาก Chain Code ในรูปแบบ Histogram ซึ่งเป็นวิธีการอธิบายคุณลักษณะที่น่าสนใจ

3. Histograms of Oriented Gradients: HOG [72] คือการดึงลักษณะเด่นของภาพโดยใช้การอธิบายภาพด้วยค่าความถี่ของทิศทางตามค่าเกรเดียนต์ ซึ่งเป็นวิธีการดึงลักษณะของรูปร่างภายในภาพโดยใช้การกระจายตัวของความเข้มเกรเดียนต์ หรือทิศทางของเส้นขอบโดยใช้วิธีการวิธีการแบ่งภาพออกเป็นเซลล์ (Cell) เล็ก ๆ ซึ่งในแต่ละเซลล์นั้นจะประกอบด้วยทิศทางของค่าเกรเดียนต์ ที่ถูกเก็บไว้ในรูปแบบของค่าฮิสโตรแกรมที่จะอธิบายคุณลักษณะของวัตถุที่อยู่ในแต่ละเซลได้ และเพื่อเพิ่มประสิทธิภาพของความถูกต้องสามารถนำค่าฮิสโตรแกรมมาทำการนอร์มอลไลซ์ (Normalize) ด้วยการคำนวณตัวชี้วัดค่าความเข้มจากโอเวอร์แลป (Overlap) ของเซลภายในบล็อก (Block) เพื่อลดผลกระทบจากการเปลี่ยนแปลงของแสงและเงาให้น้อยที่สุด โดยกระบวนการสกัดคุณลักษณะของ HOG ประกอบไปด้วย 4 ขั้นตอนดังนี้



รูปที่ 62 แสดงกระบวนการสกัดคุณลักษณะของ HOG

ขั้นตอนที่ 1 การคำนวณทิศทางของค่าเกรเดียนต์ (Gradient Computation) โดยการคำนวณหาทิศทางของค่าเกรเดียนต์นี้สามารถหาได้จากการคำนวณจากค่าเวกเตอร์ในแนวแกน x และเวกเตอร์ในแนวแกน y ซึ่งการคำนวณหาทิศทางของค่าเกรเดียนต์สามารถกระทำได้จากการทำคอนโวลูชันกับเคอร์แนลดังรูปที่ 63



รูปที่ 63 เคอร์แนลฟิวเตอร์ในแนวแกน x และแกน y

จากนั้นจะทำการหาค่า Magnitude ของค่าเกรเดียนต์โดยสามารถหาได้จากสมการที่ 2.39

$$|G| = \sqrt{G_x^2 + G_y^2} \tag{2.39}$$

โดยกำหนดให้

$|G|$ คือค่าเกรเดียนต์

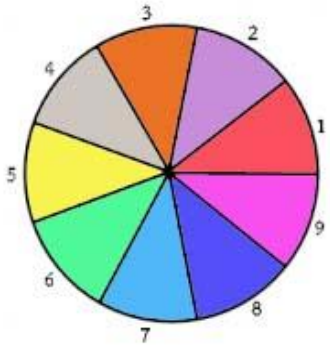
G_x^2 คือการทำอนุพันธ์อันดับที่หนึ่งในแนวแกน x

G_y^2 คือการทำอนุพันธ์อันดับที่หนึ่งในแนวแกน y

หลังจากการหาค่า Magntude จะน่าที่ได้มาคำนวณหาทิศทางของค่าเกรเดียนต์ดังสมการที่ 2.40

$$\theta = \arctan\frac{G_x}{G_y} \tag{2.40}$$

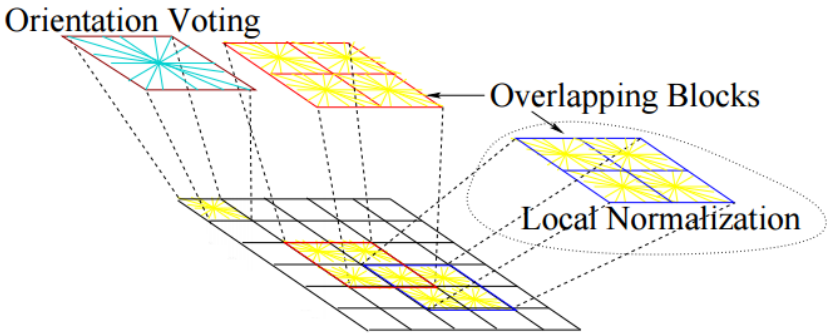
ขั้นตอนที่ 2 การเก็บทิศทางลง Bin (Orientation Binning) ขั้นตอนนี้จะทำให้ทราบถึงค่าน้ำหนักของทิศทางในแต่ละเซลล์ โดยการนำค่าที่ได้จากการหาทิศทางของค่าเกรเดียนต์มาเก็บลงไปในถัง ซึ่งจะทำการแบ่งเซลล์และแต่ละพิกเซลที่อยู่ในแต่ละเซลล์จะเป็นทิศทางของค่าเกรเดียนต์ ทั้งนี้แต่ละเซลล์จะถูกนำมาสร้างช่องฮิสโตรแกรมสำหรับเก็บทิศทาง เป็น 0-180 องศา หรือ 0-360 องศา ซึ่งจะมีช่องฮิสโตรแกรมจำนวน 9 ช่อง (Bin) ดังรูปที่ 64



จำนวนถัง (Bin)	ช่วงขององศา
1	0-40
2	40-80
3	80-120
4	120-160
5	160-200
6	200-240
7	240-280
8	280-320
9	320-360

รูปที่ 64 แสดงการกำหนดถังกับทิศทาง 0-360 องศา

ขั้นตอนที่ 3 การอธิบายคุณลักษณะของบล็อก (Descriptor Blocks) สำหรับการจัดการเกี่ยวกับการเปลี่ยนแปลงความสว่างของแสง และความคมชัดของค่าเกรเดียนต์จะต้อง เป็นบริเวณปกติ ซึ่งจะต้องมีการจัดกลุ่มของเซลล์เข้าไว้ด้วยกันเป็นกลุ่มของเซลล์ขนาดใหญ่ ที่มีการเชื่อมต่อกันเป็นบล็อก (Blocks) ทั้งนี้ในการอธิบายคุณลักษณะของ HOG นี้จะอธิบายคุณลักษณะในรูปของเวกเตอร์ของส่วนประกอบของเซลล์จากบล็อกทั้งหมด โดยปกติบล็อกเหล่านี้มักจะมีการซ้อนทับกัน ดังนั้นจะหมายความได้ว่าแต่ละเซลล์มีส่วนที่ซ้อนทับกันมากกว่าหนึ่งครั้งจนสิ้นสุดการอธิบายดังรูปที่ 65 บล็อกที่ใช้เป็นรูปสี่เหลี่ยมช่องตารางแสดงแทนโดยใช้สามค่าพารามิเตอร์ได้แก่ 1) จำนวนของเซลล์ต่อบล็อก 2) จำนวนพิกเซลต่อเซลล์ และ 3) จำนวนช่องต่อฮิสโตรแกรมเซลล์



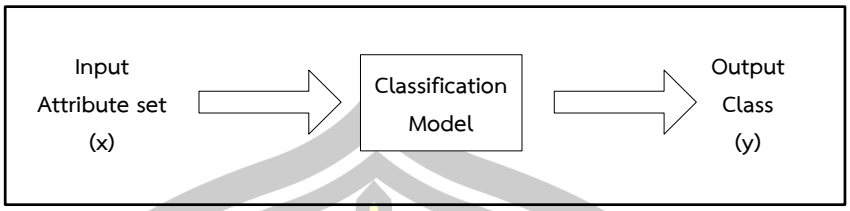
รูปที่ 65 แสดงการซ้อนทับกันของบล็อก

ขั้นตอนที่ 4 การทำนอร์มอลไลซ์บล็อก (Block Normalization) ในขั้นตอนการทำนอร์มอลไลซ์บล็อกมีวิธีการที่แตกต่างที่จะกำหนดให้ v เป็นเวกเตอร์ที่ไม่ปกติ (Non-Normalized) ที่มีการเก็บค่าฮิสโตรแกรมทั้งหมดในบล็อกที่กำหนด $\|vk\|$ เป็น k-norm สำหรับค่า $k = 1, 2$ และ e จะเป็นค่าคงที่ขนาดเล็ก ซึ่งเป็นค่าที่จะไม่มีอิทธิผลต่อผลลัพธ์ จากนั้นค่านอร์มอลไลซ์สามารถหาได้จากสมการที่ 41

$$L1-norm: f = \frac{v}{\sqrt{\|v\|_1 + e}} \tag{41}$$

(2) การจำแนกประเภท (Classification) [73] เป็นกระบวนการประเภทที่ต้องมีการเรียนรู้ (Supervised Learning) ซึ่งจำเป็นที่จะต้องมีชุดข้อมูลสอน หรือชุดข้อมูลตัวอย่าง (Training Set) ซึ่งการจำแนกประเภทข้อมูลนั้นจะจำแนกข้อมูลออกจากกันเป็นหมวดหมู่ (Categories or Class) โดยดูจากคุณลักษณะ (Attributes) ของข้อมูลในชุดสอน โดยชุดข้อมูลสอนที่นำมาใช้สอนนั้นจะต้องมีลักษณะเป็นชุดของข้อมูล (Set of Records) ซึ่งแต่ละเรคคอร์ดนั้นจะเรียกอีกอย่างหนึ่งได้ว่า Instance ที่จะประกอบด้วยข้อมูล 2 ส่วนคือ x เป็นกลุ่มของ Attributes ที่จะอธิบายถึงคุณลักษณะของข้อมูลและ y เป็น Attributes พิเศษที่กำหนดว่าข้อมูล x นั้นเป็นหมวดหมู่หรือคลาสอะไร

ทั้งนี้การจำแนกประเภทคือการเรียนรู้ถึงคุณลักษณะจากชุดของ Attributes x และทำการสร้างโมเดลจากคลาสที่มีการกำหนดไว้ล่วงหน้าแล้ว (Attributes y) ซึ่งโมเดลที่ได้จะสามารถนำไปจำแนกชุกข้อมูลที่ไม่รู้คลาสได้ดังรูปที่ 66

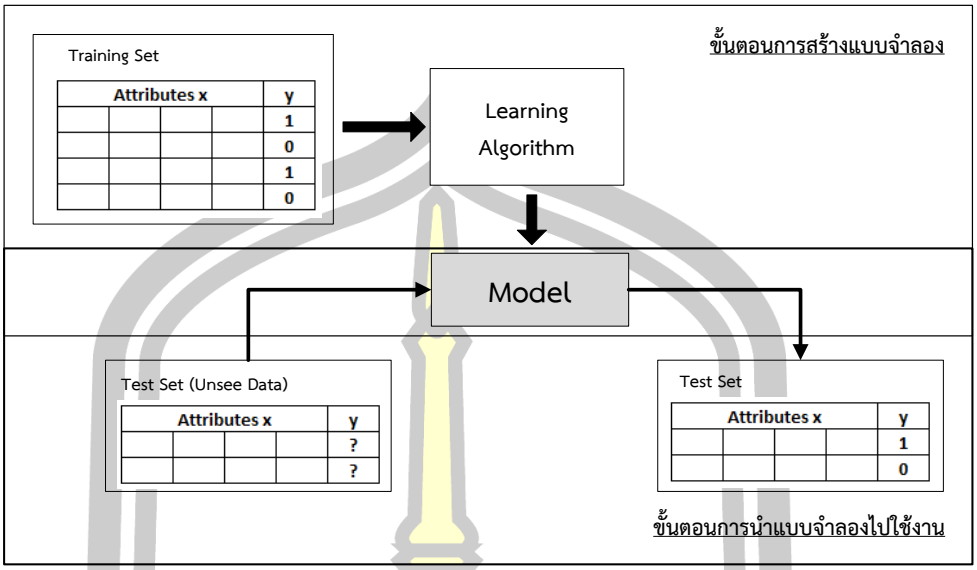


รูปที่ 66 กระบวนการจำแนกประเภท

การจำแนกประเภทสามารถแบ่งออกได้เป็น 2 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างแบบจำลอง (Model Construction) หรือ Descriptive Modeling เป็นการนำชุดข้อมูลตัวอย่างที่มีการกำหนดคลาสไว้ล่วงหน้านำไปสร้างแบบจำลอง ซึ่งแบบจำลองนั้นสามารถสร้างได้จากอัลกอริธึมต่าง ๆ เช่น เพื่อนบ้านใกล้ที่สุด ซัพพอร์ตเวคเตอร์แมชชีน หรือ โครงข่ายประสาทเทียม เป็นต้น

ขั้นตอนที่ 2 การนำแบบจำลองไปใช้ (Model Usage) หรือ Predictive Modeling เป็นการนำข้อมูลที่ไม่ทราบคลาสมาป้อนให้กับแบบจำลอง ซึ่งแบบจำลองจะให้คลาสคำตอบ ทั้งนี้ก่อนที่จำนำแบบจำลองนั้นไปใช้ได้ อาจจะนำข้อมูลที่ทราบคลาสมาป้อนให้กับแบบจำลองเพื่อเป็นการทดสอบความแม่นยำ (Accuracy) หรือความถูกต้องของแบบจำลองเสียก่อน โดยจะมีการวัดค่าออกมาเป็นค่าประสิทธิภาพ โดยค่าประสิทธิภาพที่มักจะนำมาใช้ได้นั้นได้แก่ ค่าร้อยละของความถูกต้อง ดังรูปที่ 67

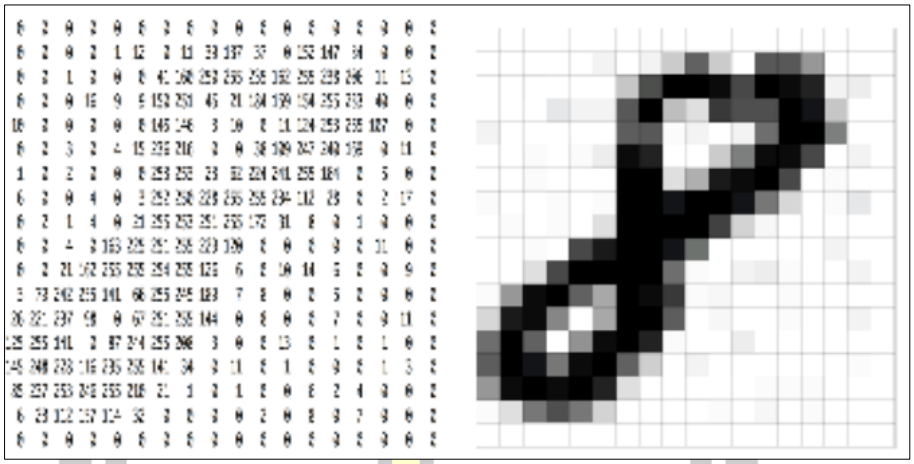


รูปที่ 67 ขั้นตอนของการจำแนกประเภท
ที่มา [73]

จากรูปที่ 67 ข้อมูลชุดสอน (Training Set) จะถูกส่งเข้ามายังอัลกอริทึมประเภทเรียนรู้ (Learning Algorithm) เพื่อที่จะสร้างเป็นแบบจำลอง (Model) โดยที่ข้อมูลชุดสอนนั้นจะมีการระบุคลาส (Attribute y) ว่าคุณลักษณะแบบไหนที่จะอยู่ในคลาสใด ซึ่งเมื่อสร้างแบบจำลองเสร็จแล้ว แบบจำลองที่ได้จะสามารถนำไปจำแนกประเภท (ระบุคลาส) ให้กับข้อมูลที่ไม่รู้จักคลาสได้ โดยสามารถทำสอบแบบจำลองได้โดยการป้อนข้อมูลที่ทราบคลาสอยู่ก่อนแล้วให้แบบจำลองเพื่อทดสอบว่าแบบจำลองนั้นมีความแม่นยำมากน้อยเพียงใด แต่อย่างไรก็ตามข้อมูลที่นำมาทดสอบนั้นจะต้องไม่ใช่ข้อมูลในชุดสอน ซึ่งจะทำให้เกิดการ Over Fitting หรือแบบจำลองสามารถตอบได้ถูกต้องเพราะได้ผ่านการเรียนรู้ในคุณลักษณะแบบนั้นมาแล้ว ซึ่งจะไม่สามารถวัดประสิทธิภาพที่แท้จริงของแบบจำลองได้

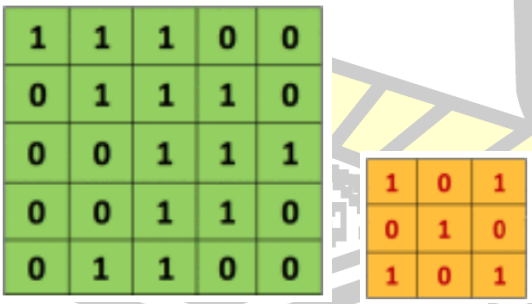
(3) โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional neural networks) [74]

ความรู้เกี่ยวกับการเรียนรู้เชิงลึก (Deep Learning) นั้นเป็นแนวคิดในการให้คอมพิวเตอร์สามารถเรียนรู้และเข้าใจข้อมูลที่ได้รับ โดยจะมีสถาปัตยกรรมการเรียนรู้ข้อมูลของคอมพิวเตอร์มากมายโดยจะกล่าวถึงโครงข่ายประสาทเทียมแบบสังวัตนาการ ที่นำมาใช้เพื่อการสกัดคุณลักษณะเด่น จนทำให้ทราบถึงผลลัพธ์หรือวัตถุที่ตรวจจับ ซึ่งการทำงานของโครงข่ายประสาทเทียมแบบสังวัตนาการนั้นมี 4 กระบวนการหลักด้วยกันได้แก่



รูปที่ 68 ตัวอย่างการแทนค่าพิกเซลลงบนรูปที่รับเข้ามา

1. คอนโวลูชั่น (Convolution) เป็นกระบวนการที่ทำเพื่อที่จะสกัดเอาคุณลักษณะที่สำคัญจากภาพออกมา โดยการใช้ค่าพิกเซล โดยค่าพิกเซลจะได้มาจากการมองของกล้องถ่ายภาพทั่วไปนั้นจะมีด้วยกันสามแชนแนล (Channel) โดยแบ่งเป็นสี ได้แก่ สีแดง น้ำเงิน และเขียว โดยแต่ละจุดสามารถแทนค่าด้วยตัวเลขเพื่อบอกความเข้มของสีนั้น ๆ โดยมีค่าตั้งแต่ 0 ถึง 255 จากความเข้มน้อยไปหามาก โดยในการทำภาพขาวดำแชนแนลของภาพนั้นจะมีเพียงหนึ่งแชนแนลเท่านั้นคือแชนแนลของสีดำซึ่งค่าของตัวเลข 0 นั้นคือสีขาวไล่ไปจนถึง 255 ซึ่งเป็นสีดำสนิทตามรูปที่ 68 ซึ่งในแต่ละชุดจะทำการคำนวณแล้วเก็บไว้ตัวอย่างการทำงานของคอนโวลูชั่นจะมีดังรูปที่ 69

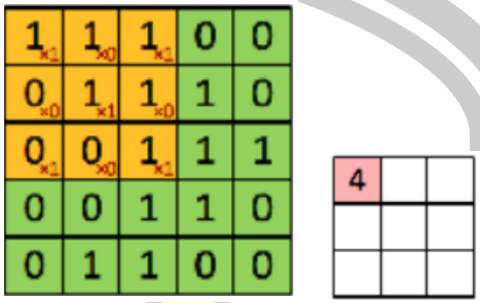


(ก) (ข)

รูปที่ 69 จำลองเมทริกซ์ที่ได้จากรูปที่รับเข้ามาและเมทริกซ์ตัวกรองค่า

ในแต่ละรูปจะมีชุดค่าเมทริกซ์ที่ต่างกัน จากรูปที่ 69 เป็นรูปภาพขนาด 5x5 พิกเซล ที่เป็นภาพขาวดำ โดยที่นี้จะกำหนดค่าให้ค่า 0 คือพิกเซลสีขาวและ 1 คือพิกเซลสีดำ และมีการกำหนดเมทริกซ์อีกหนึ่ง

ชุดขึ้นมา โดยเราจะให้เมทริกซ์ชุดนี้เป้นตัวกรองค่าไปเก็บไว้ในเมทริกซ์ชุดที่เล็กกว่า ซึ่งจะเรียกเมทริกซ์ชุดนี้ว่าตัวกรองค่า (Filter) เคอเนล (Kernel) หรือตัวตรวจจับคุณลักษณะสำคัญ (Feature Detector) ดังที่ปรากฏในรูปที่ 70



(ก) (ข)

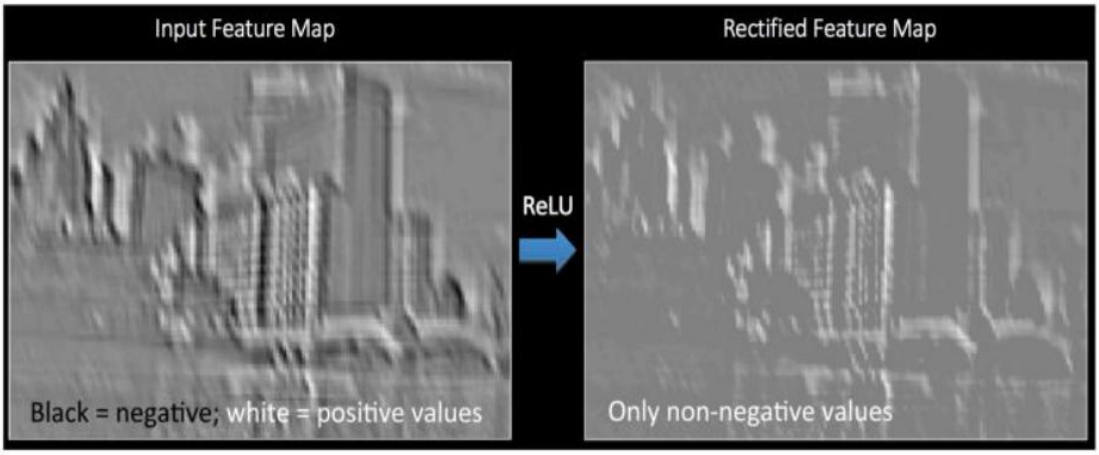
รูปที่ 70 การทำงานของตัวกรองค่าและการกำหนดเมทริกชุดใหม่หรือฟีเจอร์แมพ

เมทริกซ์ที่ทำหน้าที่เป็นตัวกรองค่าจะเคลื่อนไปทั่วภาพและทำการคูณค่าเก็บไว้ในเมทริกซ์ชุดใหม่ดังรูป 70 ซึ่งจะเรียกเมทริกซ์ชุดใหม่นี้ว่าคอนโวลฟีเจอร์ (Convolved Feature) หรือ ฟีเจอร์แมพ (Feature Map) ดังที่ปรากฏในรูปที่ 71 โดยผลลัพธ์ที่ได้เมื่อรูปผ่านการทำคอนโวลูชั่นมีตัวอย่างดังรูปที่ 72

รูปที่ 71 ภาพขาวดำดั้งเดิมเมื่อผ่านการทำคอนโวลูชั่นกลายเป็นฟีเจอร์แมพ

2. การขจัดความเป็นเชิงเส้น (ReLU) หลังจากการทำกระบวนการคอนโวลูชั่นรูปภาพและได้ฟีเจอร์แมพมาแล้ว เราจะนำฟีเจอร์แมพมาปรับแต่งให้ไม่เป็นลักษณะเชิงเส้น

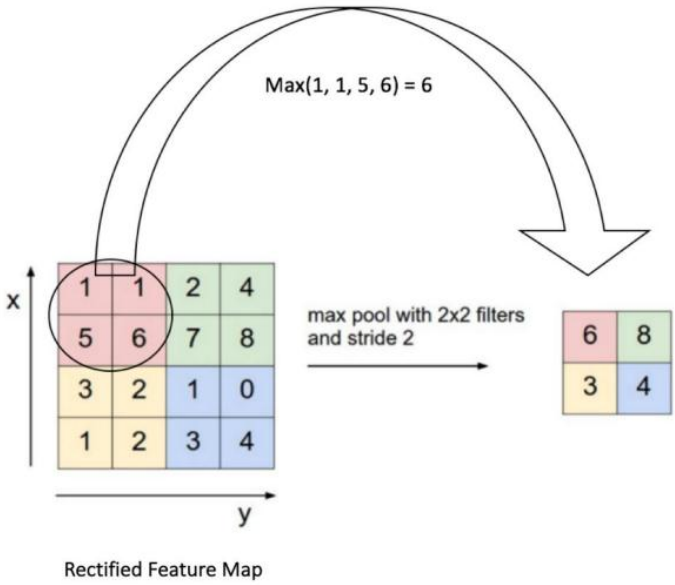
ด้วยวิธีการ ReLU เป็นการแทนที่ผลของค่าพิกเซลที่มีค่าเป็นเชิงลบในฟีเจอร์แมพด้วยค่า 0 ซึ่งจุดประสงค์ของการทำ ReLU นั้นเพื่อให้โครงข่ายประสาทเทียมแบบสังวัตนาการ สามารถเรียนรู้ข้อมูลที่ไม่เป็นเชิงเส้นจากภาพ เมื่อนำภาพเข้าสู่กระบวนการที่เป็นฟีเจอร์แมพเข้ามาทำ ReLU โดยสีดำในภาพเป็นค่าเชิงลบและสีขาวในภาพเป็นค่าเชิงบวก และเมื่อทำ ReLU ค่าที่ได้จะเหลือเพียงค่าที่เป็นเชิงบวกเท่านั้นดังรูป 72



รูปที่ 72 ตัวอย่างผลลัพธ์หลังการทำ ReLU

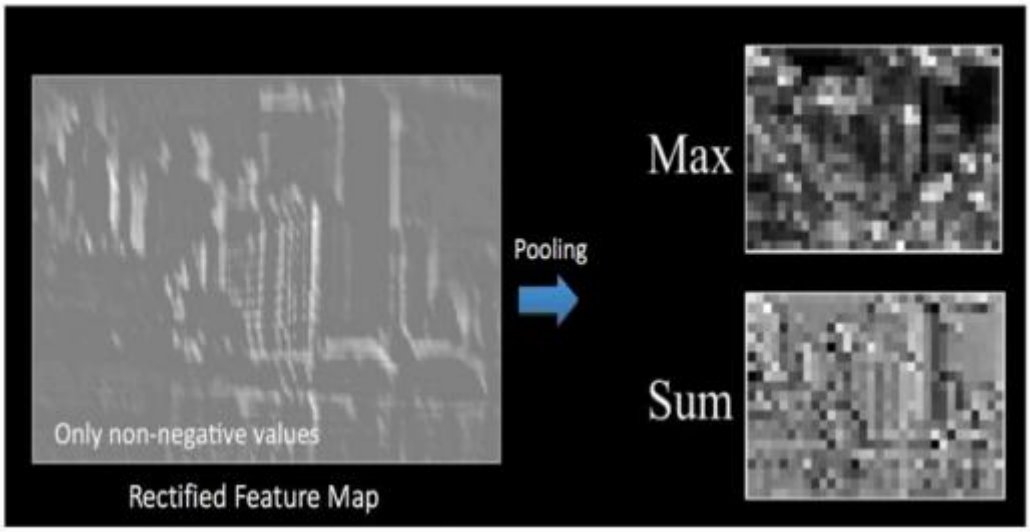
3. การพูลลิ่ง (Pooling) จะช่วยในการลดมิติของฟีเจอร์แมพลงแต่ยังคงรักษาข้อมูลที่สำคัญไว้ การพูลลิ่งสามารถที่จะจำแนกเป็นประเภทต่าง ๆ ได้เช่น พูลลิ่งด้วยค่าสูงสุด (Max Pooling) ค่าเฉลี่ย (Average Pooling) ผลรวมหรืออื่น ๆ โดยการพูลลิงนั้นจะทำให้ได้ผลลัพธ์ที่มีขนาดเล็กลงและสามารถจัดการได้ง่ายขึ้น นอกจากนี้ยังทำให้ลดจำนวนพารามิเตอร์และการคำนวณที่เกิดความจำเป็นในโครงข่าย

ในกรณีที่ต้องการทำพูลลิ่งด้วยค่าสูงสุด จะมีการกำหนดหน้าต่างหนึ่งขึ้นมาใหม่ ตัวอย่างในที่นี้สมมุติให้หน้าต่างมีขนาด 2x2 และหน้าต่างนี้จะทำการเคลื่อนที่ไปทีละ 2 พิกเซลไปจนทั่วเมทริกซ์ของฟีเจอร์แมพเพื่อทำการเก็บค่าที่สูงที่สุดในทุก ๆ 2 พิกเซลตามรูป 73



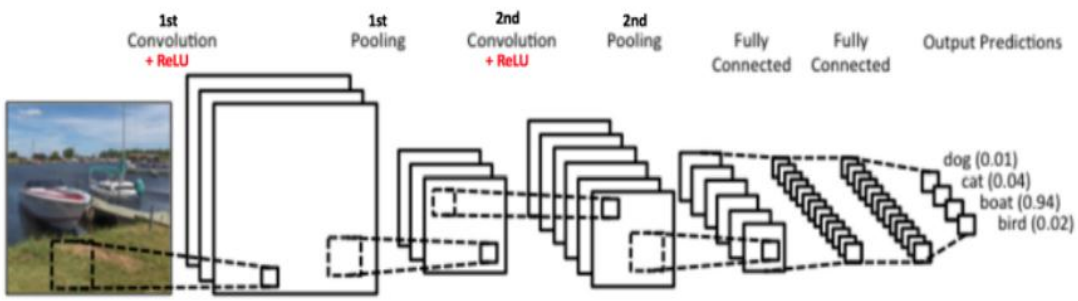
รูปที่ 73 การพูลลิ่งค่ามากที่สุด

นอกจากนี้ การพลูลิ่งจะทำตามจำนวนแชนแนลของภาพซึ่งในกรณีภาพสีเราจะได้รับผลลัพธ์ออกมาถึงสามผลลัพธ์ เมื่อฟีเจอร์แมพที่ผ่านการทำ ReLU เข้ามาทำการพูลลิ่งจะมีผลลัพธ์ดังรูป 74



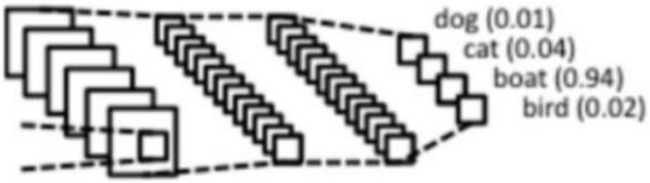
รูปที่ 74 ภาพผลลัพธ์ที่ได้หลังการทำพูลลิ่ง

4. การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer)



รูปที่ 75 ภาพผังการทำงานของระบบโครงข่ายประสาทเทียมแบบสังวัตนาการ

กระบวนการคอโวลูชั่น ReLU และการพูลลิ่งกระบวนการทั้งสามกระบวนการนี้ต้องมีการทำซ้ำจนกว่าจะมีการเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer) ผลลัพธ์จากการทำคอนโวลูชั่นและพูลลิ่งนั้น ทำให้ได้คุณลักษณะเด่นในระดับสูง (High-Level Features) ของรูปที่รับเข้ามาจุดประสงค์ของการทำให้เชื่อมต่อกันแต่ละชั้นโดยสมบูรณ์นั้นเพื่อ นำคุณลักษณะเด่นไปทำการคัดกรองรูปที่รับเข้ามาให้อยู่ในรูปของคลาส (Classes)



รูปที่ 76 การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์

โดยผลลัพธ์ที่ได้จะแสดงค่าความมั่นใจ (Confident) ออกมา ตัวอย่างของการคัดกรองที่มีการเทรนข้อมูลไว้สีประเภท เมื่อรูปภาพจะถูกนำเข้ามาสู่กระบวนการทั้งหมดจะแสดงดังรูปที่ 76 จากการใช้การตรวจจับวัตถุโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการซึ่งผลลัพธ์ที่ได้คือค่าความมั่นใจและตำแหน่งของวัตถุ

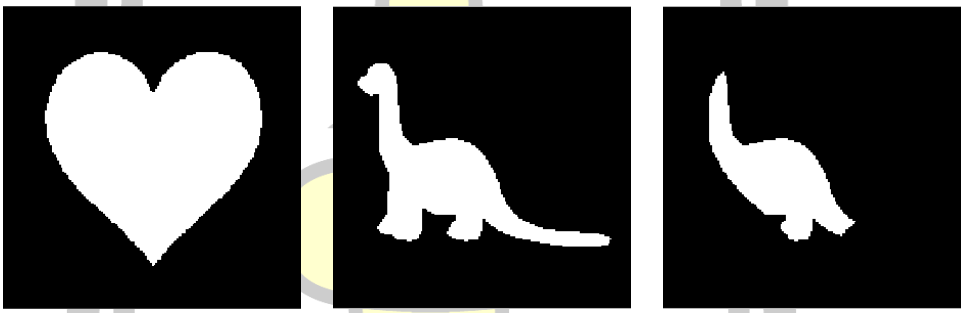
2.1.3 การวัดประสิทธิภาพ

ในงานวิจัยนี้ได้แบ่งการวัดประสิทธิภาพเป็น 2 ส่วนดังนี้

2.1.3.1 เกณฑ์ในการประเมินผลการแบ่งส่วนภาพ จะใช้การวัดค่าความคล้ายเชิงพื้นที่ (Area Similarity) [48] ซึ่งค่าความคล้ายเชิงพื้นที่ S_{area} สามารถคำนวณได้จากสมการที่ 2.42

$$S_{area} = \frac{2n(A_1 \wedge A_2)}{n(A_1) + n(A_2)} \tag{2.42}$$

โดยที่ A_1 คือภาพขาวดำ (Binary Image) ของผลการแบ่งส่วนภาพด้วยมือหรือภาพ Ground truth และภาพ A_2 คือภาพขาวดำของผลการแบ่งส่วนภาพ ซึ่งในที่นี้จะกำหนดให้วัตถุที่ได้จากการแบ่งส่วนภาพเป็นสีขาวมีค่าความเข้มเท่ากับ 1 และกำหนดให้พื้นหลังเป็นสีดำมีค่าความเข้มเท่ากับ 0 $n(A)$ คือจำนวนพิกเซลที่เป็นสีขาวของภาพ A และ $\wedge$ คือตัวดำเนินการ "And"



(ก) (ข) (ค)

รูปที่ 77 ตัวอย่างการคำนวณค่าความคล้ายเชิงพื้นที่ (ก) ภาพขาวดำ A_1 (ข) ภาพขาวดำ A_2 (ค) ภาพขาวดำ $A_1 \wedge A_2$

ค่าความคล้ายเชิงพื้นที่ S_{area} จะมีค่าอยู่ในช่วง 0 ถึง 1 กล่าวคือ ถ้าภาพ A_1 และ A_2 เป็นภาพเดียวกันค่า S_{area} ที่คำนวณได้จะมีค่ามากที่สุดคือ 1 ตัวอย่างการคำนวณค่าความคล้ายเชิงพื้นที่ โดยที่รูป (ก) เป็นภาพขาวดำ A_1 ที่มีขนาด 200x200 พิกเซล โดยมีวัตถุสีขาวในภาพเป็นรูปหัวใจ ซึ่งมีจำนวนพิกเซลที่เป็นสีขาวทั้งหมด $n(A_1)$ เท่ากับ 14,108 พิกเซลและรูป (ข) เป็นรูปภาพไดโนเสาร์ A_2 มีจำนวนพิกเซลที่เป็นสีขาวทั้งหมด $n(A_2)$ เท่ากับ 6,217 พิกเซลและรูป (ค) เป็นภาพที่ได้จาก $A_1 \wedge A_2$ เท่ากับ 4,315 พิกเซล ดังนั้นสามารถคำนวณหาค่า S_{area} ได้ดังนี้

$$S_{area} = \frac{2n(A_1 \wedge A_2)}{n(A_1) + n(A_2)} = \frac{2x4,315}{14,108 + 6,217} = \frac{8,630}{20,325} = 0.42$$

ค่า S_{area} เท่ากับ 0.42 ที่คำนวณได้นี้สามารถบอกได้ว่าวัตถุสีขาวในภาพ A_1 และ A_2 มีความคล้ายคลึงกันในเชิงพื้นที่เท่ากับร้อยละ 42

2.1.3.2 การวัดประสิทธิภาพของโมเดลมี 2 ขั้นตอน [75] ดังต่อไปนี้

1) การใช้เทคนิค Cross-Validation เป็นการนำชุดข้อมูลมาใช้สำหรับสอนและทดสอบโดยแบ่งข้อมูลส่วนหนึ่งไว้สำหรับสอนและอีกส่วนไว้สำหรับทดสอบโดยการทำ Cross-Validation มีหลายวิธีเช่น Holdout Method และ K-fold Cross Validation ดังนี้

(1) วิธี Holdout Method มีวิธีการคือชุดข้อมูลจะแบ่งออกเป็น 2 ชุดคือ ชุดข้อมูลสอน (Training Set) และชุดข้อมูลทดสอบ (Testing Set) ข้อดีคือ สามารถประเมินชุดข้อมูลจำนวนมากได้ในเวลาที่ไม่นาน

(2) วิธี K-Fold Cross Validation เป็นวิธีที่ปรับปรุงมาจากวิธีการแรก โดยจะนำข้อมูลมาแบ่งออกเป็นส่วนย่อย ๆ แล้วกำหนดส่วนย่อยจำนวนหนึ่งเป็นชุดข้อมูลสอนและมีส่วนย่อยเพียงส่วนเดียวที่กำหนดเป็นชุดข้อมูลทดสอบตามหลักการของวิธีการแรก ทำการกำหนดชุดข้อมูลสอนและชุดข้อมูลทดสอบในลักษณะเดียวกันแต่จะสลับกลุ่มกันไปเรื่อย ๆ จนชุดข้อมูลทดสอบนั้นรอบคลุมในทุกกลุ่มของข้อมูลเป็นจำนวน k ครั้งข้อเสียคือในการสอนต้องประมวลผลใหม่ทั้งหมด k ครั้งตามจำนวนที่แบ่งกลุ่ม

2) การใช้มาตรวัดประสิทธิภาพของโมเดลจำแนกประเภทข้อมูล (Confusion Matrix) โดยทั่วไปแล้วการวัดประสิทธิภาพที่นิยมใช้กันในงานวิจัยและการทำงานต่าง ๆ จะมีด้วยกันอยู่ 4 ค่าคือ [76]

(1) Precision เป็นการวัดความแม่นยำของวิธีการ โดยจะพิจารณาแยกทีละคลาส

(2) Recall เป็นการวัดความถูกต้องของวิธีการ โดยจะพิจารณาแยกทีละคลาส

(3) F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของวิธีการ โดยจะพิจารณาแยกทีละคลาส

(4) Accuracy เป็นการวัดความถูกต้องของวิธีการ โดยจะพิจารณารวมกันทุกคลาส

ซึ่งการที่จะวัดประสิทธิภาพของวิธีการได้นั้นจำเป็นต้องมีการสร้างตาราง Confusion Matrix ก่อนโดยตาราง Confusion Matrix นั้นคือตารางแบบจัตุรัสที่มีจำนวนแถวเท่ากับจำนวนของคอลัมน์และเท่ากับจำนวนของคลาส เช่นในการตรวจจับข้อความเราจะสนใจอยู่ 2 คลาสคือ คลาสที่เป็นข้อความ และคลาสที่ไม่ใช่ข้อความ ฉะนั้นตาราง Confusion Matrix ที่จะสร้างเป็นตารางจะมีขนาด 2x2 ดังในตาราง 1 โดยข้อมูลด้านคอลัมน์คือ คลาสที่อยู่ในข้อมูลเทรนนิงดาต้า (Actual) และข้อมูลในแนวแถว คือคลาสที่วิธีการนั้นทำนายออกมาได้ (Predicated)

ตาราง 1 ตาราง Confusion Matrix ของข้อมูลการตรวจจับข้อความ

Predicated/Actual	Text	Not Text
Text	TP	FP
Not Text	FN	TN

ในตาราง 1 ค่าที่แสดงในช่องต่าง ๆ ของตารางจะประกอบไปด้วย

1. TP (True Positive) คือ สิ่งที่อัลกอริทึมทำนายว่าจริง และคนบอกว่ามันจริง
2. TN (True Negative) คือ สิ่งที่อัลกอริทึมทำนายว่าเท็จ และคนบอกว่ามันเท็จ
3. FP (False Positive) คือ สิ่งที่อัลกอริทึมทำนายว่าจริง และคนบอกว่ามันเท็จ
4. FN (False Negative) คือ สิ่งที่อัลกอริทึมทำนายว่าเท็จ แต่คนบอกว่ามันจริง

หลังจากได้ตาราง Confusion matrix ของแต่ละอัลกอริทึมแล้วขั้นต่อไปจะเป็นการคำนวณหาค่า Precision Recall F-measure และ Accuracy ในแต่ละวิธีดังสมการต่อไปนี้

สมการคำนวณหาค่า Precision ดังสมการที่ 2.43

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{2.43}$$

สมการคำนวณหาค่า Recall ดังสมการที่ 44

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{2.44}$$

สมการคำนวณหาค่า F-measure ดังสมการที่ 2.45

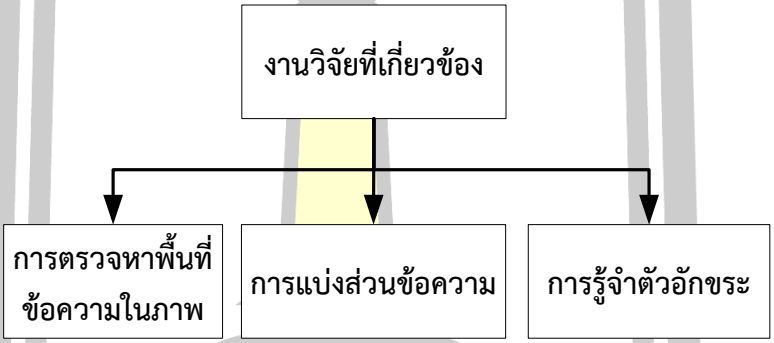
$$F - measure = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (2.45)$$

สมการคำนวณหาค่า Accuracy ดังสมการที่ 2.46

$$Acucuracy = \frac{TP + TN}{(TP + FP + FN + TN)} \quad (2.46)$$

2.2 งานวิจัยที่เกี่ยวข้อง

ในการดำเนินการวิจัยครั้งนี้มีงานวิจัยที่ต้องศึกษาหลายส่วนโดยแบ่งเป็นส่วนดังนี้



รูปที่ 78 การจัดกลุ่มงานวิจัยที่เกี่ยวข้อง

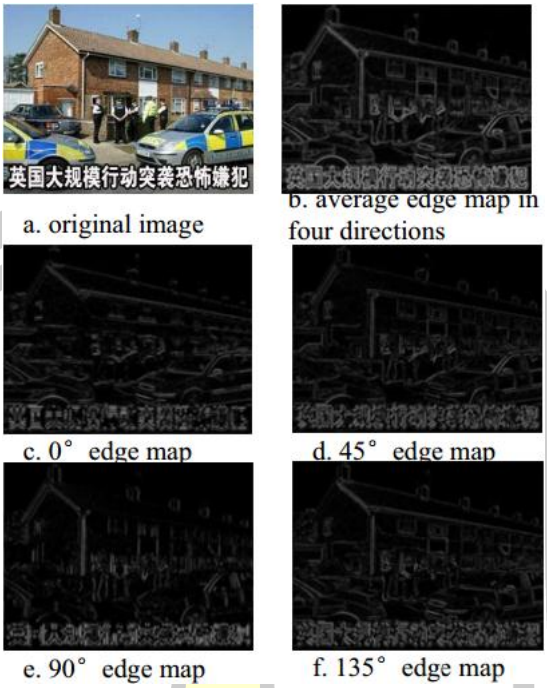
2.2.1 การตรวจหาพื้นที่ข้อความในภาพ (Text Localization)
งานวิจัยที่ทำการศึกษาเกี่ยวกับการตรวจหาพื้นที่ข้อความในภาพ อาทิ เช่น

ปี ค.ศ. 2004 Chen และคณะ[77] ได้สรุปวิธีการตรวจหาข้อความที่มีอยู่ในวิดีโอ ซึ่งสามารถจำแนกวิธีการได้ 3 ประเภทคือ วิธีการหาองค์ประกอบที่เชื่อมต่อกัน (Connected Component or Region-based) วิธีการจำแนกพื้นผิว (Texture Classification Methods) และการตรวจหาเส้นขอบ (Edge Detection Methods) ซึ่ง Chen และคณะ ได้กล่าวถึงวิธีการเหล่านี้ว่าวิธีการทั้งหมดที่ได้กล่าวมานี้ไม่สามารถนำมาประยุกต์ใช้ได้กับข้อมูลของภาพวิดีโอได้ แต่สามารถนำไปใช้ได้กับข้อมูลรูปภาพทั่ว ๆ ไปได้ และ Chen และคณะยังได้ระบุถึงปัญหาของการตรวจหาข้อความที่มีความละเอียดของภาพต่ำ (Low-Resolution) โดย Chen และคณะได้นำเสนอวิธีการแก้ปัญหานี้ด้วยการคำนวณค่า Histograms ของภาพระดับสีเทา (Grayscale) ซึ่งจะค่า Histograms นำมาเปรียบเทียบกันระหว่างเฟรมแต่ละเฟรมที่ติดกัน และในการแบ่งส่วนของขอบข้อความนั้นได้นำ

วิธีการของโซเบล (Sobel Operator) มาใช้ในการหาขอบภาพในแนวนอน และแนวตั้งเพื่อใช้ในการค้นหาพื้นที่ ๆ เป็นข้อความ ผลจากการทดลองของ Chen และคณะจะให้อัตราความถูกต้องประมาณร้อยละ 85 ในกรณีที่ข้อมูลวีดีโอมีความละเอียดต่ำทั่ว ๆ ไป แต่ยังมีอุปสรรคหลักที่เกิดขึ้นในงานนี้ คือจำนวนของความผิดพลาดในการตรวจหา ที่เกิดขึ้นในขั้นตอนการแบ่งส่วนข้อความ ซึ่งจะส่งผลต่อความถูกต้องในกระบวนการดำเนินงาน

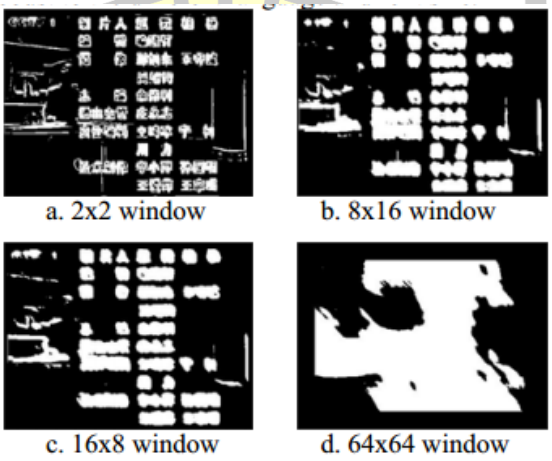
ปี ค.ศ. 2004 Ezaki และคณะ[31] ได้นำเสนอระบบที่อ่านข้อความที่พบในฉากธรรมชาติโดยมีวัตถุประสงค์เพื่อให้ความช่วยเหลือแก่ผู้พิการทางสายตา โดยวิธีการที่นำเสนอจะประกอบไปด้วย 4 ขั้นตอนหลักได้แก่ ขั้นตอนแรกการสกัดตัวอักษรออกจากขอบภาพ ขั้นตอนนี้เป็นการใช้ Sobel Operator กับช่องสีแต่ละช่องสีของภาพ RGB หลังจากนั้นจะนำภาพ Edge Map ของแต่ละช่องสีมารวมกัน โดยจะเลือกเฉพาะพิกเซลที่สอดคล้องกันในแต่ละพิกเซล ขั้นตอนต่อมาจะนำภาพ Edge map มาแปลงเป็นภาพไบนารี ด้วยวิธีการของ Otsu's ขั้นตอนที่สอง หลังจากที่ได้ภาพไบนารีแล้วจะดำเนินการสลับค่าของภาพไบนารี คือจากค่าพิกเซลมีค่าเป็น 1 จะถูกเปลี่ยนให้มีค่าเป็น 0 และค่าที่เป็น 0 จะถูกเปลี่ยนให้มีค่าเป็น 1 ขั้นตอนที่สาม เป็นการสกัดตัวอักษรโดยใช้คุณสมบัติของสี ขั้นตอนนี้จะเป็นการใช้วิธีการของ Otsu's กับช่องสีแต่ละช่องสีของภาพ RGB ซึ่งสุดท้ายจะได้ภาพไบนารี ขั้นตอนสุดท้ายจะเป็นการนำภาพไบนารีที่ได้จากขั้นตอนที่สอง และขั้นตอนที่สามมาหาพื้นที่ที่เป็นข้อความด้วยวิธีการ Connected Component Analysis ผลการทดลองที่ได้ยังไม่เพียงพอสำหรับการนำไปใช้งานจริง ซึ่งทำให้การทำงานในอนาคตของเขาจะมุ่งเน้นไปที่วิธีการใหม่สำหรับการแยกตัวอักษรขนาดเล็กที่มีความแม่นยำสูง

ปี ค.ศ. 2005 Liu และคณะ[32] นำเสนอการตรวจหาข้อความที่อยู่ในภาพด้วยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) จากคุณสมบัติของขอบภาพ วิธีการที่นำเสนอจะมีขั้นตอนดังต่อไปนี้ ขั้นตอนการตรวจหาขอบภาพ เป็นขั้นตอนในการสร้าง Edge Map ทั้งหมด 4 ภาพ โดยแต่ละภาพจะมีทิศทางที่แตกต่างกันได้แก่ ทิศทางที่ 0 องศา 45 องศา 90 องศาและ 135 องศา ดังรูปที่ 79 ขั้นตอนการสกัดคุณลักษณะขอบภาพ โดยการสกัดคุณลักษณะของขอบภาพจะมีการหาคุณลักษณะทั้งหมด 24 คุณลักษณะโดยจะแบ่งเป็นทิศทางละ 6 คุณลักษณะ โดยการหาคุณลักษณะจะสร้างจากการทำ Sliding Window โดยกำหนดขนาด w x h Pixels ซึ่ง w, คือความกว้าง และ h คือความสูง (ในการทดลองจะมีการกำหนด w และ h หลายขนาดโดยผลการทดลองที่ดีที่สุด w = 16 h = 8) ขั้นตอนการตรวจหาข้อความ ในขั้นตอนนี้จะใช้ K-Means Algorithm ในการจำแนกคุณลักษณะออกเป็น 2 กลุ่มคือพื้นหลังและข้อความ



รูปที่ 79 แสดง Edge maps ของการตรวจหาขอบภาพ

จากผลการทดลองการตรวจหาข้อความมีการกำหนดขนาดของ w x h หลายค่าได้แก่ 2x2, 8x16, 16x8, 64x64 ดังรูปที่ 80 ซึ่งผลการทดลองแสดงให้เห็นว่าหากกำหนดค่าที่น้อยเกินไปหรือมากเกินไป จะทำให้การตรวจหาข้อความจะผิดพลาดได้ ซึ่งจะส่งผลต่อการดำเนินการในขั้นตอนต่อไปในการระบุข้อความ ผลของค่าการระลึกคิดเป็นร้อยละ 81.5 และค่าความแม่นยำคิดเป็นร้อยละ 78.3

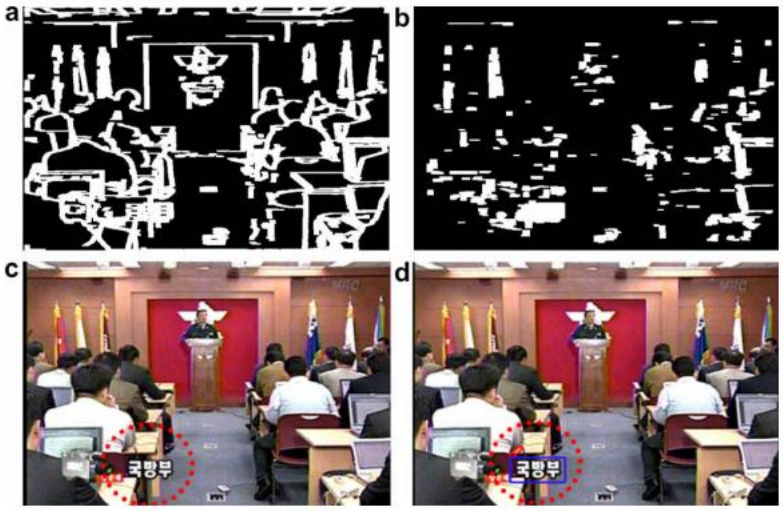


รูปที่ 80 ผลการทดลองด้วยหน้าต่างแบบหลายขนาด

ปี ค.ศ. 2007 Yi และคณะ[78] ได้นำเสนอวิธีการในการตรวจหาข้อความและสกัดข้อความจากคุณสมบัติของสีโดย Yi และคณะได้ให้เหตุผลว่า วิธีการจากงานวิจัยต่าง ๆ ที่ผ่านมาไม่มีการนำคุณสมบัติของสีมาพิจารณา ซึ่งคุณสมบัติของสีนั้นจะเป็นประโยชน์ และมีประสิทธิภาพในการขจัดสัญญาณรบกวน (Noise) ออกจากภาพได้ วิธีการของ Yi และคณะส่วนใหญ่จะอยู่ในการจัดกลุ่มสีที่ใช้ในภาพ โดยการดำเนินงานจะแบ่งออกเป็น 2 ช่วง ซึ่งช่วงแรกจะเป็นการตรวจหาข้อความ ทีจะพิจารณาจาก 2 คุณสมบัติร่วมกัน คือ สีที่เหมือนกัน ทั้งนี้ข้อความที่เป็นข้อความเดียวกันโดยส่วนใหญ่จะมีความเป็นไปได้ที่จะใช้สีที่เหมือนกัน และความคมชัดของขอบ ช่วงที่สองเป็นการสกัดข้อความ จะเป็นการพิจารณาถึงความแตกต่างระหว่างสี กับพื้นหลังในรูปภาพ ในช่วงนี้จะสามารถเลือกปรับเปลี่ยนระนาบสีที่ค่อนข้างดีที่สุดตามความแตกต่างที่ตรงกันข้ามข้อความหมู่ระนาบสีทุกสีสำหรับการสกัดข้อความ พร้อมทั้งมีการกำจัดสัญญาณรบกวน (Noise Removal) ซึ่งเป็นการพิจารณาความแตกต่างระหว่างสีของข้อความและพื้นหลังในภาพ พร้อมทั้งการจัดกลุ่มสีที่ใช้จะใช้ในการลบสัญญาณรบกวนภายในภาพ ซึ่งส่งผลให้การรู้จำข้อความมีประสิทธิภาพมากขึ้น ผลการทดลองได้มีการประเมินประสิทธิภาพของการตรวจหาข้อความ ด้วยการเปรียบเทียบผลการดำเนินงานกับวิธีการของ Lyu's และคณะ[79] ซึ่งผลการทดลองแสดงให้เห็นว่าวิธีการของ Yi และคณะให้ผลที่ดีกว่าด้วยค่าความแม่นยำ (Precision) ที่ 0.719 ค่าความระลึก (Recall) ที่ 0.623 และ f-measure ที่ 0.643

ปี ค.ศ. 2009 Jung และคณะ[36] ได้อธิบายว่า โดยทั่วไปแล้วการสกัดข้อมูลตัวอักษรจากวิดีโอจะประกอบไปด้วย 3 ขั้นตอนที่สำคัญคือ การระบุตำแหน่งข้อความ (Text Localization) คือ การจำกัดพื้นที่ของข้อความโดยใช้กรอบสี่เหลี่ยม การแบ่งส่วนข้อความ (Text Segmentation) คือ การแบ่งส่วนข้อความที่ถูกต้องเพื่อการนำไปคำนวณหาส่วนที่เป็นข้อความในภาพ สุดท้ายการรู้จำข้อความ (Text Recognition) คือการแปลงภาพข้อความให้เป็นข้อความธรรมดา อีกทั้งวิธีการที่มีอยู่ของการตรวจหาข้อความสามารถแบ่งได้เป็น 4 ประเภทคือ Connected Component Analysis (CCA)-Based Method Edge-Based Method Corner-Based Method และTexture-Based Method ซึ่งจะแตกต่างไปจาก Chen และคณะ[77] ที่ได้นำเสนอไปก่อนหน้านี้ซึ่ง Jung และคณะต้องการคิดวิธีที่แตกต่างจาการใช้คุณสมบัติของ ขอบ มุม และพื้นผิว สำหรับการระบุพื้นที่ที่เป็นข้อความ ด้วยการนำเสนอวิธีการใช้ Stroke Filter ร่วมกับการประยุกต์ใช้ SVM Classifier ซึ่งมีความสามารถในการตรวจหาข้อความ และสามารถในการสกัดส่วนที่ไม่ใช่ข้อความออกจากส่วนของ Text Candidate โดยวิธีการประกอบด้วย 3 ขั้นตอนที่สำคัญได้แก่ กระบวนการตรวจหาข้อความ จะเป็นการคำนวณด้วยวิธีการ Stroke Filter และการเชื่อมต่อกลุ่มของ Stroke ด้วยการใช้ Morphologic Filter ซึ่ง Morphologic Filter จะเป็นการสร้าง Binary Stroke Map จากนั้นจะเป็นการคำนวณส่วนประกอบที่เชื่อมต่อกันด้วยวิธีการ Connected Components (CCs) จาก Binary

Stroke Map ซึ่งการวิเคราะห์ CCA จะใช้จุดเชื่อมต่อแบบ 4 ช่อง โดยจะไม่ใช้แบบ 8 ช่องเพื่อหลีกเลี่ยงการเชื่อมต่อที่มากเกินไป ทั้งนี้เปรียบเทียบการกรองระหว่าง Canny Filter และ Stroke Filter ในรูปที่ 81 ซึ่งจากรูปจะเห็นได้ว่าวิธีการ Canny Filter จะมีความซับซ้อนมากว่าวิธีการ Stroke Filter



รูปที่ 81 แสดงผลลัพธ์บางส่วน (a) ผล Binary Map ของวิธีการ Canny Filter (b) ผล Binary Map ของวิธีการ Stroke Filter (c) ผลการระบุตำแหน่งข้อความของวิธีการ Canny Filter (d) ผลการระบุตำแหน่งข้อความของวิธีการ Stroke Filter

ขั้นตอนต่อมาคือการตรวจสอบข้อความ ขั้นตอนนี้เป็นการนำกระบวนการ Support Vector Machine (SVM) มาเป็นตัวจำแนกพื้นที่ ๆ เป็นข้อความ โดยเลือกใช้ Radial Basis Function (RBF) เป็นฟังก์ชันในการจำแนกของ SVM ทั้งนี้ผลการคำนวณค่าเฉลี่ยที่ได้มีค่ามากกว่าค่า Threshold ที่กำหนดไว้จะถือได้ว่าพื้นที่นั้นเป็นพื้นที่ที่เป็นข้อความ ซึ่งค่า Threshold จะกำหนดในงานวิจัยนี้คือ 0.3 ขั้นตอนสุดท้ายคือการปรับแต่งบรรทัดข้อความ โดยส่วนใหญ่ผลลัพธ์ที่ได้จากการกำหนดของเขตของบรรทัดข้อความจะมีผลที่ดี แต่มีบางส่วนที่การกำหนดขอบเขตไม่ถูกต้อง ดังรูปที่ 82 เช่นการกำหนดขอบเขตมากเกินไป น้อยเกินไป หรือการตัดขอบเขตไม่ถูกต้อง จากปัญหาเหล่านี้ จึงได้มีการนำ SVM มาจำแนกคุณสมบัติของสี เพื่อเป็นการปรับแต่งขอบเขตของข้อความให้ดีขึ้นผลดังรูปที่ 83

รูปที่ 82 ผลการกำหนดขอบเขตข้อความ (a) การกำหนดขอบเขตมากเกินไป (b) การกำหนดขอบเขตน้อยเกินไป (c) การตัดขอบเขตไม่ถูกต้อง

รูปที่ 83 แสดงผลของการปรับแต่งบรรทัดข้อความ

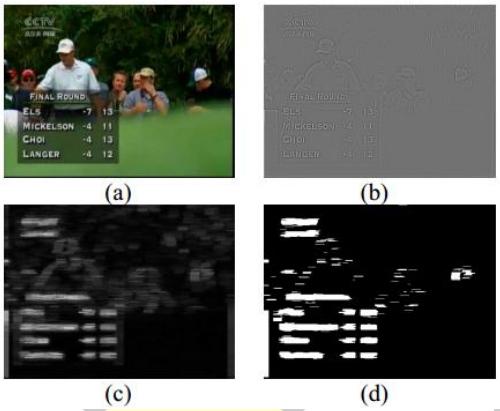
ผลการทดลองได้มีการเปรียบเทียบประสิทธิภาพกับวิธีการ Canny Filter ซึ่งวิธีการของ Jung และคณะจะได้ผลดีกว่าในแง่มุมของอัตราความระลึกอยู่ที่ 97.0 อัตราความแม่นยำอยู่ที่ 42.4 แต่อัตราความเร็วจะอยู่ที่ 0.111 ซึ่งจะด้อยกว่าวิธีการ Canny Filter

ปี ค.ศ. 2009 Phan และคณะ[33] ได้นำเสนอวิธีการตรวจหาข้อความที่มีประสิทธิภาพซึ่งขึ้นอยู่กับตัวดำเนินการลาปาเซียน (Laplacian) โดยกระบวนการจะเริ่มจากการนำรูปภาพมาแปลงให้เป็นภาพเทา (Grayscale) และนำไปกรองด้วย Laplacian Mask ขนาด 3x3 พิกเซล เพื่อนำไปตรวจสอบความไม่ต่อเนื่องกันในแนวนอน แนวตั้ง ทแยงซ้าย และทแยงขาวดังรูปที่ 84 ซึ่งจะได้ภาพ Laplacian-Filtered

1	1	1
1	-8	1
1	1	1

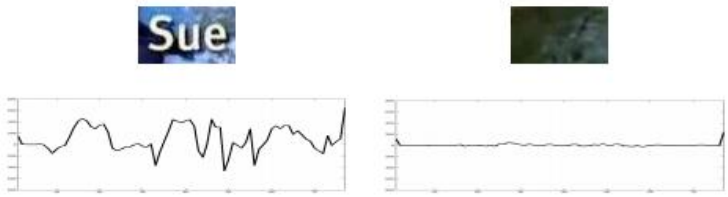
รูปที่ 84 Laplacian Mask ขนาด 3 × 3 พิกเซล

เนื่องจากภาพ Laplacian-Filtered ที่ได้จะประกอบไปด้วยค่า Positive และค่า Negative ซึ่งค่าเหล่านี้จะมีความสอดคล้องกับส่วนที่เป็นข้อความและพื้นหลังซึ่งจะใช้ Maximum Gradient Difference (MGD) ในการกำหนดความแตกต่างระหว่างค่าที่มากที่สุด และค่าที่น้อยที่สุด ของภาพ Laplacian filtered ดังรูปที่ 85



รูปที่ 85 ขั้นตอนการตรวจหาข้อความ (a) ภาพดั้งเดิม (b) ภาพ Laplacian Filtered (c) Maximum Gradient Difference Map (d) การจัดกลุ่มข้อความ

ทั้งนี้พื้นที่ที่เป็นข้อความจะมีค่า MGD มากกว่าพื้นที่ที่ไม่ใช่ข้อความ เพราะค่า Positive มีค่าที่สูงกว่าค่า Negative ดังรูปที่ 86



รูปที่ 86 ตัวอย่างของข้อความและพื้นที่ที่ไม่ใช่ข้อความ

นอกจากนี้ยังมีการแปลงค่า (Normalize) ของ MGD map ให้อยู่ในช่วง 0 - 1 และใช้อัลกอริทึม K-means ในการจัดกลุ่มพิกเซลออกเป็น 2 กลุ่มคือ กลุ่มที่เป็นข้อความและไม่ใช่ข้อความ ผลการทดลองแสดงให้เห็นว่ามีประสิทธิภาพที่ดีกว่าวิธีการ Edge-Based และ Gradient-Based ด้วยอัตราการตรวจหา Detection Rate (DR) คิดเป็นร้อยละ 93.3 และ False Positive Rate (FPR) คิดเป็นร้อยละ 7.9

ปี ค.ศ. 2010 Epshtein และคณะ[34] ได้มีความต้องการที่จะพัฒนาระบบ OCR ที่มีประสิทธิภาพที่สามารถตรวจหาข้อความในภาพที่มีพื้นหลังที่ซับซ้อนได้ Epshtein อธิบายถึงความแตกต่างระหว่างภาพสแกนเอกสาร กับภาพข้อความที่อยู่ในฉากธรรมชาติว่า ภาพสแกนเอกสารจะเป็นภาพที่สามารถนำไปผ่านกระบวนการรู้จำได้ง่ายกว่าภาพข้อความที่อยู่ในฉากธรรมชาติมากกว่า โดยภาพข้อความที่อยู่ในฉากธรรมชาตินั้นจะมีความหลากหลายมากกว่าเช่น การรบกวนของสี ภาพที่เบลอ และความละเอียดของภาพเป็นต้น ซึ่งถือเป็นความท้าทายอย่างมาก การดำเนินงานจะค่อนข้างมีความคิดที่คล้ายคลึงกันกับงานวิจัยของ Subramanian และคณะ[35] ในการตรวจหาความกว้างของตัวอักษร โดยวิธีการที่นำเสนอของ Subramanian และคณะจะทำการค้นหาภาพในแนวนอนเพื่อจับคู่ของพิกเซลกับพิกเซลฝั่งตรงข้าม (จุดพิกเซลที่เป็นขอบภาพจากฝั่งหนึ่งไปอีกฝั่งหนึ่ง) ซึ่งเป็นวิธีการที่สามารถตรวจสอบข้อความที่อยู่ใกล้ในแนวนอนเท่านั้น วิธีการของ Epshtein เป็นการพัฒนาวิธีการที่เรียกว่า Stroke Width Transform (SWT) ที่เป็นกระบวนการในการตรวจหาค่าความกว้างของวัตถุที่ปรากฏอยู่ภายในภาพ โดยจะคำนวณจากขอบของวัตถุด้านหนึ่งไปยังอีกด้านหนึ่งและทำการบันทึกค่าความกว้างที่ได้เก็บไว้ในแต่ละพิกเซล ซึ่งผลการทดลองจะมีการเปรียบเทียบประสิทธิภาพการทำงานของอัลกอริทึมกับวิธีการของ Subramanian และคณะโดยอัตราความระลึกคำคิดเป็นร้อยละ 79.04 และอัตราความแม่นยำคิดเป็นร้อยละ 79.59 ซึ่งมีประสิทธิภาพดีกว่าผลจากวิธีการขอ Subramanian และคณะวิธีนี้ช่วยให้เราสามารถนำไปใช้ ได้ในหลาย ๆ ภาษาและแบบตัวอักษร พร้อมทั้งการตรวจหาวิธีนี้สามารถตรวจหาความโค้งของบรรทัดข้อความได้ดี

ปี ค.ศ. 2010 Anoual และคณะ[80] เนื่องจากความแปรปรวนของข้อความที่อยู่ในรูปภาพซึ่งมาจากปัญหาหลาย ๆ ด้าน เช่นพื้นหลังที่ซับซ้อน รูปแบบของตัวอักษร และขนาดของตัวอักษร จึงส่งผลให้การสกัดข้อความจากภาพเป็นเรื่องที่ยากมาก เพื่อแก้ปัญหาที่ได้กล่าวมา Anoual และคณะได้มีการนำเสนอระบบที่มีประสิทธิภาพ สำหรับการตรวจหาข้อความในภาพด้วย ขั้นแรกการตรวจหาขอบภาพ จะใช้เทคนิค Canny Edge Detection ในการตรวจหาขอบภาพ ดังรูปที่ 87

รูปที่ 87 ผลลัพธ์ของวิธีการหาขอบภาพด้วย Canny

ขั้นตอนต่อมาเป็นเทคนิคที่ใช้ในการแยกแยะพื้นที่ข้อความและพื้นที่ไม่ใช่ข้อความ โดยการพิจารณาพื้นที่จากความแตกต่างของขอบที่อยู่ในภาพ ซึ่งพื้นที่ ๆ เป็นข้อความเส้นรูปร่าง (Contour) จะมีลักษณะที่เชื่อมต่อกันหรือบรรจบกัน ซึ่งแนวคิดของ Anoual และคณะจะมีอยู่ว่าตัวอักษรจะมีเส้นรูปร่างที่เชื่อมต่อกันจากจุดหนึ่งไปอีกจุดหนึ่ง โดยจะถือว่าส่วนนั้นคือข้อความดังรูปที่ 88

รูปที่ 88 ผลของการเลือกเส้นรูปร่างที่ปิด

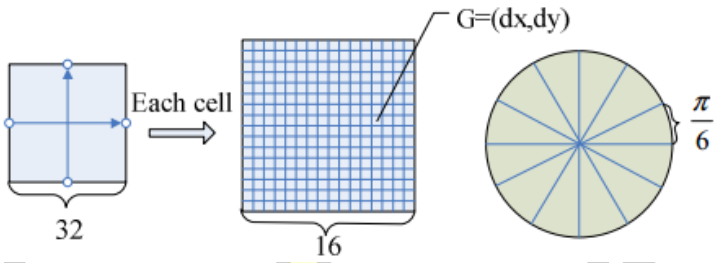
ขั้นตอนสุดท้ายเป็นการวิเคราะห์ลักษณะพื้นผิดจากพื้นที่ของเส้นรูปร่างเพื่อเป็นการหาพื้นที่ที่เป็นข้อความจริง ๆ ด้วยการนำวิธีการ Gradient Magnitude มาใช้กับพื้นที่ที่ตรวจพบในขั้นตอนก่อนหน้านี้ เพื่อเป็นการประเมินประสิทธิภาพของวิธีการนี้ ได้มีเปรียบเทียบกับวิธีการของ Ezaki และคณะ[31] ผลของการทดลองแสดงให้เห็นว่าอัตราความแม่นยำที่ได้รับอยู่ที่ 0.95 สำหรับอัตราความระลึกเท่ากับ 0.89 และ F- Measure ที่มีค่าเท่ากับ 0.92

ปี ค.ศ. 2010 Ma และคณะ[81] ได้นำเสนอวิธีการที่มีประสิทธิภาพสำหรับการตรวจหาข้อความในภาพ โดยอัลกอริทึมจะอยู่บนพื้นฐานของการตรวจหาของของภาพ และกระบวนการ Connected Component ขั้นตอนในการดำเนินงานจะแบ่งออกเป็น 4 ขั้นตอน ขั้นตอนแรกเป็นขั้นตอนการตรวจหาของของภาพ ที่จะดำเนินการโดยการใช้ตัวดำเนินการ 2 ตัวได้แก่ Canny Operator และ Binary Gradient Map ซึ่งการที่เลือก Canny เพราะตัวดำเนินการนี้สามารถที่จะตรวจหาขอบภาพได้ดีที่สุดในรูปภาพที่ซับซ้อน สำหรับ Gradient Map จะดำเนินการด้วย Sobel Edge Operator ซึ่งจะนำมาใช้กับช่องสี RGB ทั้ง 3 ช่องและในแต่ละช่องสีจะมีการหาค่าทิศทางในแนวนอน แนวตั้ง ทแยงซ้าย และทแยงขวา แล้วนำมาหาค่ามากที่สุดในแต่ละทิศทางดังรูปที่ 89

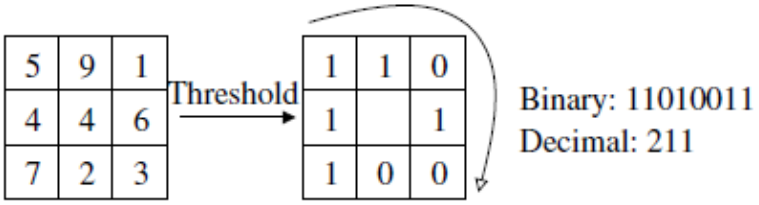
รูปที่ 89 (a) ภาพต้นฉบับ (b) Canny Edge Image (c) Binary Gradient Image (d) ผลของ Edge Detection

ขั้นตอนที่สองเป็นการสกัดคุณลักษณะของภาพ ซึ่งงานวิจัยนี้จะมีการใช้ 2 อัลกอริทึมในการสกัดคุณลักษณะได้แก่ Histogram of Gradient (HOG) และ Local Binary Pattern (LBP) โดยมีการดำเนินการดังนี้ HOG จะมีการแบ่งพื้นที่ออกเป็น 32x32 เซลในแต่ละเซลจะมีขนาด 16x16 พิกเซล โดยได้กำหนดการคำนวณหาทิศทางเกรเดียนต์ทีละ 2x2 เซลและถังทิศทางเกรเดียนต์ได้กำหนดไว้ทั้งหมด 12 ทิศทางซึ่งจะได้คุณลักษณะจากวิธีการของ HOG ทั้งหมด 48 คุณลักษณะ (12*4) = 48 สำหรับการหาคุณลักษณะด้วย LBP การคำนวณด้วย LBP จะถูกนำมาคำนวณบนรูปภาพเกรเดียนต์

ซึ่ง LBP จะเปรียบเทียบค่าระหว่างพิกเซลที่อยู่ตรงกลางกับพิกเซลที่อยู่รอบ ๆ ขนาด 3x3 พิกเซล ซึ่งวิธีการนี้จะได้ทั้งหมด 59 คุณลักษณะ



รูปที่ 90 ภาพประกอบของคุณสมบัติ HOG



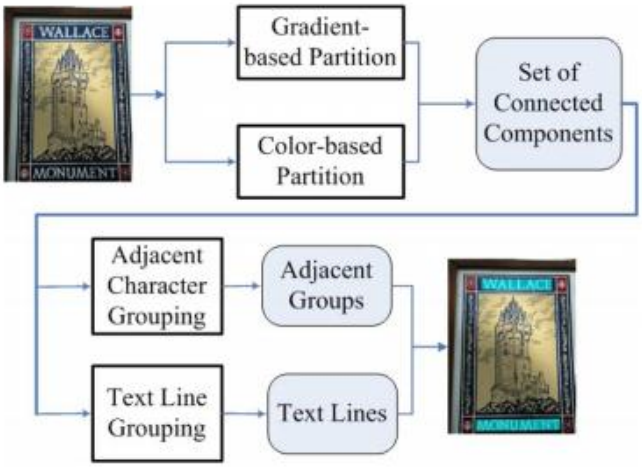
รูปที่ 91 ตัวอย่างการคำนวณด้วยวิธีการ local binary pattern (LBP)

ขั้นตอนที่สามเป็นขั้นตอนการจำแนกและการแบ่งกลุ่ม ด้วยชุดข้อมูลที่ในงานวิจัยนี้จะเป็นชุดข้อมูลของ ICDAR 2003 ที่ถูกนำมาปรับใหม่โดยการทำ Sliding Window เป็นบล็อกขนาด 32x32 พิกเซล และมีการซ้อนทับกันในแต่ละบล็อก ซึ่งทำให้ได้ 10510 บล็อกที่เป็นข้อความ และ 12797 บล็อกที่ไม่ใช่ข้อความ เพื่อจะนำมาเป็นชุดฝึกและชุดทดสอบสำหรับกระบวนการ SVM ที่นำมาใช้ในการจำแนกส่วนที่เป็นบล็อกข้อความ ขั้นตอนสุดท้ายเป็นการวิเคราะห์กลุ่มของบล็อกข้อความซึ่งสามารถพิจารณาได้จากระยะห่างระหว่างบล็อกข้อความที่มีระยะน้อยกว่า 50% จะถูกจัดให้อยู่กลุ่มเดียวกัน ผลการทดลองจะแสดงให้เห็นถึงประสิทธิภาพที่ดีกว่าวิธีการของ Ezaki และคณะ[31] ด้วยอัตราความแม่นยำคิดเป็นร้อยละ 67 และอัตราความระลึกคิดเป็นร้อยละ 72

ปี ค.ศ. 2011 Yi และคณะ[82] เนื่องจากความท้าทายในการค้นหาตำแหน่งของข้อความในภาพที่มีพื้นหลังที่ซับซ้อน ซึ่งการแสดงตัวอักษรในภาพปกติจะมีการแสดงข้อความไม่ได้เป็นแนวตรง พร้อมทั้งยังฝังตัวอยู่ในพื้นหลังที่มีความซับซ้อนดังรูปที่ 92

รูปที่ 92 ตัวอย่างของข้อความในภาพฉากธรรมชาติ

จากความท้าทายของการค้นหาตำแหน่งของข้อความ Yi และคณะจึงได้นำเสนอกรอบแนวคิดสำหรับวิธีแก้ไขปัญหาดังกล่าวดังรูปที่ 93

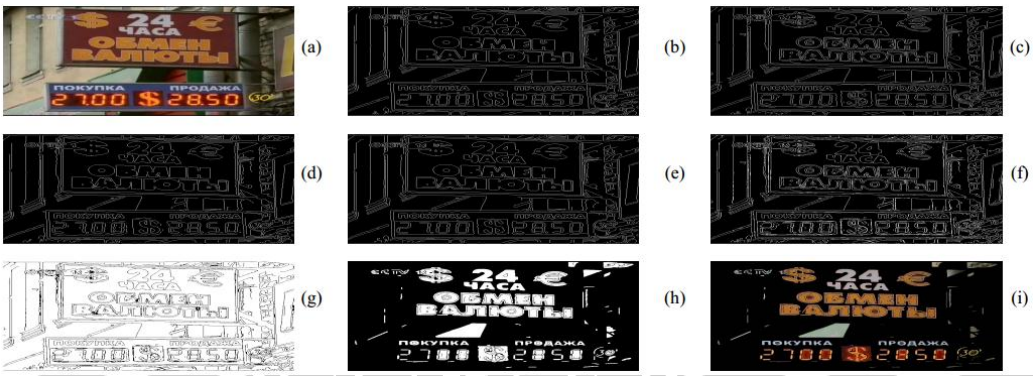


รูปที่ 93 ผังงานของกรอบแนวคิดในการตรวจหาตัวอักษร

จากผังงานของงานวิจัยนี้ ซึ่งประกอบไปด้วย 2 ขั้นตอนหลัก ขั้นตอนแรกจะมีการกำหนด Image Partition เพื่อที่จะค้นหาข้อความ โดยขึ้นกับคุณลักษณะของ Gradient-Based Method และ Color-Based Method หลังจากนั้นเป็นการลบส่วนที่เชื่อมต่อกันที่ไม่ได้เป็นข้อความ ขั้นตอนที่สอง การจัดกลุ่มตัวอักษร จะจัดกลุ่มตามโครงสร้างที่ใช้ร่วมกัน เช่น ขนาดตัวอักษร ระยะห่างระหว่างตัวอักษรสองตัวที่อยู่ใกล้เคียงกัน และการวางแนวของตัวอักษร ซึ่งในขั้นตอนนี้ได้นำเสนอวิธีการวิเคราะห์โครงสร้างของข้อความ ออกเป็น 2 วิธีคือ Adjacent Character Grouping คือการจัดกลุ่มตัวอักษรที่อยู่ติดกัน และ Text Line Grouping Method คือการจัดกลุ่มบรรทัดข้อความ ผลการ

ทดลองแสดงให้เห็นว่าการทดลองด้วยวิธี Color-Based Partition จะได้ประสิทธิภาพดีกว่าการทดลองด้วยวิธี Gradient-Based Partition แต่จะใช้เวลามากขึ้นในการตรวจหาข้อความในแต่ละ Partition สี

ปี ค.ศ. 2012 Huang และคณะ[83] ได้อธิบายถึงข้อความที่ปรากฏอยู่ในภาพที่สามารถแบ่งได้เป็น 2 ประเภทคือข้อความที่ถูกนำมาซ้อนทับ และข้อความที่ฝังตัวอยู่ในฉากธรรมชาติ ทั้งนี้การตรวจหาข้อความ การระบุตำแหน่งข้อความ และการสกัดข้อความ สำหรับข้อความที่ฝังตัวอยู่ในฉากธรรมชาติเป็นเรื่องที่ทำได้ยากกว่า ข้อความที่ถูกนำมาซ้อนทับ เนื่องมาจากแสงที่ส่องเข้ามา ไม่สม่ำเสมอกัน ซึ่งแสงจะมีผลต่อการตรวจหาข้อความในฉากธรรมชาติเป็นอย่างมาก แต่แสงไม่สามารถส่งผลกระทบต่อคุณลักษณะของขอบภาพได้ แต่พื้นหลังที่ซับซ้อนนั้นจะมีผลกระทบต่อคุณลักษณะของขอบภาพ จากปัญหาที่ท้าทาย Huang จึงได้พัฒนาวิธีการตรวจหาข้อความที่มีประสิทธิภาพ ที่ทนต่อความเปลี่ยนแปลงของแสง และพื้นหลังที่มีความซับซ้อน โดยวิธีการที่นำเสนอกระบวนการดังต่อไปนี้ ขั้นตอนแรกเป็นขั้นตอนการตรวจหาข้อความจากคุณสมบัติของขอบภาพ ด้วยการใช้วิธีการของ Canny ซึ่งมีความสามารถในการปรับเปลี่ยนตัวเองในสภาพแวดล้อมต่าง ๆ ได้ดี ค่าที่ได้รับจะมีความเหมาะสมกับการรับรู้คุณลักษณะของขอบดังรูปที่ 94

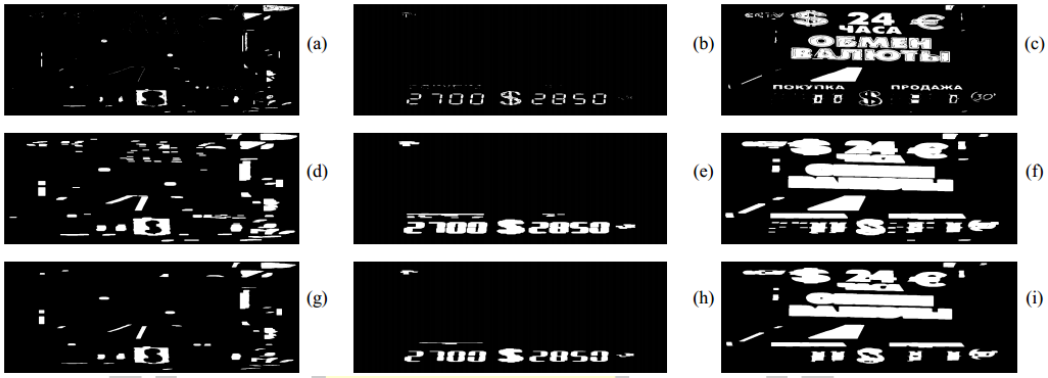


รูปที่ 94 การตรวจหาข้อความที่ปรากฏขึ้นในภาพโดยใช้คุณลักษณะของขอบ

จากรูป (a) เป็นรูปภาพต้นฉบับ (b) เป็นผลการตรวจหาขอบจากภาพสีเทา (d)(e)(f) เป็นผลของการตรวจหาขอบจากภาพแต่ละช่องสัญญาณ RGB (c) เป็นผลของการตรวจหาขอบจากภาพ RGB ทั้ง 3 ช่องสัญญาณ เมื่อทำการเปรียบเทียบรูป (b) กับ (c) จะพบว่า รูป (c) จะมีรายละเอียดของขอบที่มีความสมบูรณ์มากกว่ารูป (b) ซึ่งจะเหมาะสำหรับการตรวจหาข้อความที่ปรากฏในภาพด้วยแสดงที่มีความสว่างที่ไม่สม่ำเสมอ เมื่อได้ Edge Map (EM) แล้วขั้นตอนมาคือการดำเนินการด้วยวิธีการ

Connected Component Analysis (CCA) จากรูป (g) พื้นที่สีขาวเป็นพื้นที่ ๆ มีการเชื่อต่อกัน ซึ่งพื้นที่สีขาวนี้จะเป็นพื้นที่ ๆ คาดว่าจะเป็นข้อความ รูป (h) เป็นพื้นที่ ๆ เป็นข้อความและจะมีการลบพื้นที่ ๆ ไม่ใช่ข้อความให้มากที่สุด และรูป (i) เป็นผลจากการตรวจหาข้อความจากคุณลักษณะของขอบซึ่งเป็นการตรวจหาพื้นที่ ๆ เป็นข้อความแบบหยาบ ๆ ขั้นตอนที่สองเป็นขั้นตอนของการตรวจหาข้อความจากคุณสมบัติของสี หลังจากที่ได้ดำเนินการตรวจหาข้อความตามลักษณะของขอบแบบหยาบ ๆ ภายในพื้นที่ ๆ เป็นข้อความจะมีการใช้สีที่คล้ายคลึงกัน ดังนั้นจึงใช้การจัดกลุ่มสีเพื่อหาพื้นที่ ๆ เป็นข้อความได้อย่างถูกต้องด้วย K-Means Clustering โดยจะมีการกำหนดค่า K เท่ากับ 3 คือเป็นการแบ่งกลุ่มสีออกเป็น 3 กลุ่ม ดัง

รูปที่ 95



รูปที่ 95 การตรวจหาข้อความด้วยการจัดกลุ่มสี

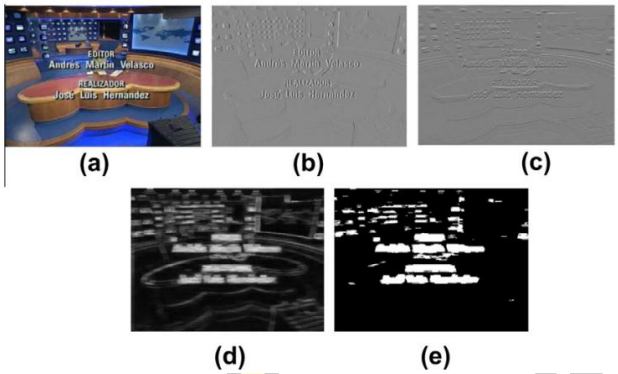
ขั้นตอนที่สามเป็นขั้นตอนการระบุตำแหน่งของบรรทัดข้อความ หลังจากที่ได้พื้นที่ ๆ เป็นข้อความแล้วจะมีการใช้ Morphological Operation ในการเชื่อมต่อตัวอักษรที่อยู่ใกล้เคียงกัน ซึ่งในขณะเดียวกัน Morphological Operation จะกำจัดสัญญาณรบกวนออกจากพื้นหลังออกไปเล็กน้อย ดังรายละเอียดวิธีการดังต่อไปนี้ 1) ในผลลัพธ์ของการจัดกลุ่มสีจะมีการขยายตัว Morphological Operation ออกทางด้านความกว้างและความสูง โดยทั่วไปแล้วการจัดเรียงตัวของตัวอักษรภาษาอังกฤษจะเป็นการจัดเรียงตัวในแนวนอน ซึ่งในภาษาจีนจะมีการจัดเรียงตัวในแนวตั้ง ดังรูป (d)(e)(f) ซึ่งเป็นผลของการขยายตัวออก 2) จะเป็นการดำเนินการกัดเซาะ Morphological Operation ออกเล็กน้อยโดยการคำนวณค่าความกว้างและค่าความสูง ซึ่งจะสามารถลบพื้นหลังออกได้เล็กน้อย ดังรูป (g)(h)(i) เป็นผลของการกัดเซาะ Morphological Operation ขั้นตอนสุดท้ายเป็นการระบุพื้นที่ข้อความด้วยวิธีการ SVM หลังจากการตรวจหาข้อความในภาพ ซึ่งจะได้พื้นที่ ๆ คาดว่า

เป็นข้อความแล้วจะมีการใช้ SVM Classifier ในการฝึกเพื่อให้สามารถจำแนกความแตกต่างระหว่างพื้นที่ ๆ เป็นข้อความ และไม่ใช่ข้อความดังรูปที่ 96

รูปที่ 96 การตรวจหาข้อความบนพื้นฐานของ SVM จากรูป (a) เป็นผลของการตรวจหาข้อความต้นฉบับ และ (b) เป็นผลลัพธ์ของการตรวจหาข้อความที่ได้รับการฝึกด้วย SVM

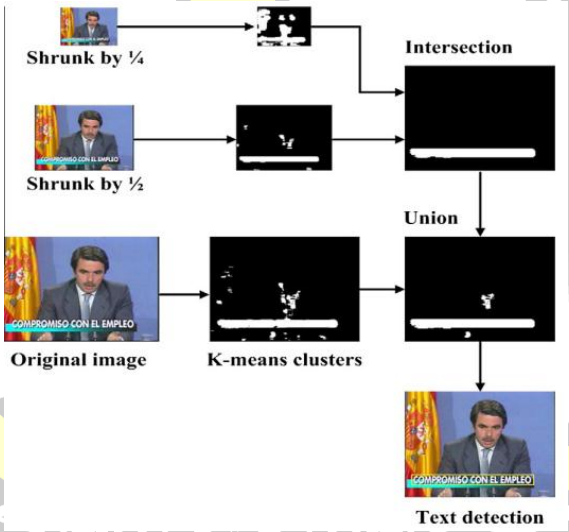
ผลการทดลองได้มีการเปรียบเทียบประสิทธิภาพของอัลกอริทึมกับวิธีการของ Kim และคณะ[84] ซึ่งได้นำเสนอวิธีการในการสกัดข้อความในฉากธรรมชาติ ด้วยการใช้คุณสมบัติของภาพในระดับต่ำ (Low-Level) และการตรวจหาพื้นที่ด้วยคุณสมบัติความกว่างของข้อความ อย่างไรก็ตามวิธีที่ที่นำเสนอของ Kim และคณะนี้ยังมีความยากลำบากในการตรวจหาตัวอักษรที่มีขนาดใหญ่เกินไปหรือผลกระทบจากการเปลี่ยนแปลงของแสงสว่างที่รุนแรง ผลการเปรียบเทียบแสดงให้เห็นประสิทธิภาพที่ดีกว่า สำหรับการตรวจหาข้อความด้วย สี ขนาดตัวอักษร และการวางแนวข้อความที่ต่างกัน

ปี ค.ศ. 2012 Wei และคณะ[85] การเปลี่ยนแปลง ในพื้นหลังที่มีความซับซ้อน ขนาดและสีตัวอักษร การสร้างพื้นที่ในการตรวจหาข้อความ ในภาพวิดีโอ ซึ่งเป็นเรื่องที่ยาก Wei และคณะได้นำเสนอแผนงานพีระมิด (Pyramidal Scheme) ในการตรวจหาข้อความในภาพวิดีโอ โดยประกอบด้วย 3 ขั้นตอนหลักขั้นตอนแรกเป็นขั้นตอนการปรับขนาดของภาพระดับเทา (Grayscale) ซึ่งจะลดขนาดลงจากเดิม 1/2 และ 1/4 จากนั้นจะมีการคำนวณค่าเกรเดียนต์ในแนวตั้ง เกรเดียนต์ในแนวนอน และการดำเนินการ Maximum Gradient Difference (MGD) หลังจากนั้นใช้วิธีการจัดกลุ่มด้วย K-Means clustering ซึ่งจะแยก MGD ออกเป็น 2 กลุ่มคือ กลุ่มที่เป็นข้อความ และไม่ใช่ข้อความแสดงดังรูปที่ 97



รูปที่ 97 การตรวจหาพื้นที่ของข้อความ (a) ภาพต้นฉบับ (b) ภาพเกรเดียนต์ในแนวนอน (c) ภาพเกรเดียนต์ในแนวตั้ง (d) ภาพ MGD Map (e) การจัดกลุ่มข้อความ

หลังจากกระบวนการ K-Means Clustering เสร็จสิ้นแล้วก็จะปรับขนาดของภาพให้เป็นขนาดเท่าเดิม แล้วนำทั้งสามภาพมารวมกันดังรูปที่ 98



รูปที่ 98 การรวมกันของการจัดกลุ่ม K-Means

ขั้นตอนที่สอง เป็นการปรับแต่งพื้นที่ข้อความ จากการจัดกลุ่มด้วย K- Means โดยการใช้วิธีการ Connected Component แบบ 4 ช่องเพื่อเป็นการกำจัดพื้นที่ที่มีขนาดเล็กเกินไปที่จะสามารถเป็นพื้นที่ข้อความได้ จากนั้นจะใช้ Connected Component เชื่อมต่อพื้นที่โดยการใช้ SVM ที่สร้างขึ้นจากภาพระดับเทา ที่ได้จากภาพที่นำเข้ามาสุดท้ายการกำหนดขอบเขต จะปรับแต่งวิธีการในการค้า

หารายละเอียดจากค่า Threshold ขั้นตอนสุดท้ายเป็นการระบุข้อความ โดยการใช้คุณสมบัติทางเลขาคณิตและคุณสมบัติของพื้นผิวของข้อวาม ซึ่งมีการนำกระบวนการ SVM มาใช้ในขั้นตอนนี้เพื่อเพิ่มความถูกต้องมากขึ้น ผลการทดลองได้มีการเปรียบเทียบประสิทธิภาพกับวิธีการ Edge-Based ที่นำเสนอโดย Liu และคณะ[32] วิธีการ Uniform-Colored ที่นำเสนอโดย Mariano และคณะ[86] วิธีการ Laplcian ที่นำเสนอโดย Phan และคณะ[33] ซึ่งประสิทธิภาพของวิธีการที่นำเสนอนี้มีประสิทธิภาพที่ดีกว่าสามวิธีที่ทำการเปรียบเทียบด้วยอัตราการตรวจหา Detection Rate (DR) คิดเป็นร้อยละ 95.1 อัตราความแม่นยำ Precision Rate (PR) คิดเป็นร้อยละ 89.6 และอัตราการตรวจหาที่ผิดพลาด Misdetection Rate (MDR) คิดเป็นร้อยละ 5.2

ปี ค.ศ. 2012 Sharma และคณะ[87] ได้นำเสนอวิธีการตรวจจับข้อความที่มีการวางตัวในแนวนอน และไม่ได้วางตัวในแนวนอน โดยวิธีการที่นำเสนอประกอบด้วยขั้นตอนดังต่อไปนี้ ขั้นตอนแรกเป็นการนำทิศทางเกรเดียนต์และคุณสมบัติของค่า Magnitude มาพิจารณาร่วมกันเพื่อที่จะระบุพิกเซลของข้อความโดดเด่นจาก Sobel Edge Map ขั้นตอนที่สองเป็นขั้นตอนการขจัดส่วนที่คาดว่าไม่ใช่ข้อความโดยการใช้วิธีการ Connected Component Analysis เพื่อหา Word Patches ซึ่ง Word Patches คือ ส่วนประกอบของข้อความที่อยู่ใกล้เคียงกัน ขั้นตอนที่สามเป็นขั้นตอนการหาทิศทางของบรรทัดข้อความจาก Word Patches Word Patches จะขยายตัวไปในทิศทางเดียวกันในกลุ่มของ Sobel Edge Map ขั้นตอนที่สี่เป็นขั้นตอนการจัดกลุ่มข้อความโดยขึ้นอยู่กับทิศทางของบรรทัดข้อความในขั้นตอนที่สามเพื่อที่จะเรียกคืนข้อมูลตัวอักษรที่ขาดหายไป ซึ่งในบางครั้งในการหาทิศทางของข้อความ อาจจะมีการรวมบรรทัดข้อความสองบรรทัดเข้าด้วยกัน เนื่องจากพื้นที่ระหว่างบรรทัดข้อความมีน้อยเกินไป ขั้นตอนสุดท้ายเป็นขั้นตอนการจัดกลุ่ม เฟรมข้อความที่วางตัวในแนวนอน และเฟรมข้อความที่ไม่ได้วางตัวในแนวนอน เพื่อแก้ปัญหาของขั้นตอนที่สี่ ซึ่งการศึกษาครั้งนี้ได้มีการเปรียบเทียบวิธีการของ Zhou และคณะ[88] ที่ได้แนะนำวิธีการตรวจหาข้อความทั้งในแนวนอนและแนวตั้งในภาพวิดีโอ ผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอมีประสิทธิภาพที่ดีกว่าวิธีการของ Zhou และคณะ

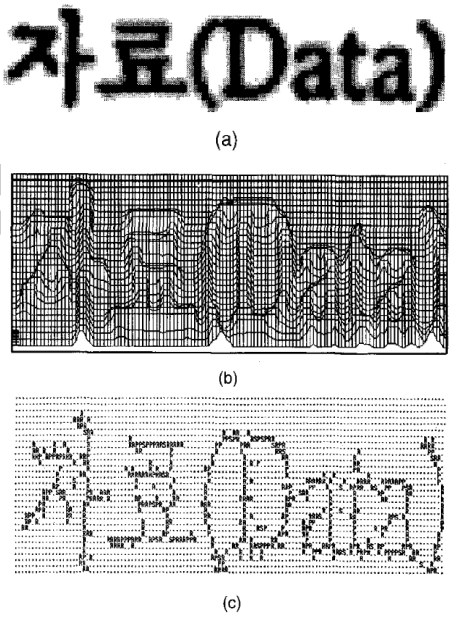
ปี ค.ศ. 2014 Pise และคณะ[89] การตรวจหาข้อความในฉากธรรมชาติเป็นปัญหาที่มีความท้าทายเพราะการเปลี่ยนปลงในรูปแบบของตัวอักษร ขนาดของตัวอักษร การวางแนวตัวอักษร พื้นหลังที่ซับซ้อน และแสงที่ไม่สม่ำเสมอกัน เพื่อแก้ปัญหาเหล่านี้ Pise และคณะได้นำเสนอวิธีการในการใช้คุณสมบัติของ Histogram of Oriented Gradients (HOG) ซึ่งมีกระบวนการดังนี้ กระบวนการก่อนการประมวลผล (Pre-Processing) จะเริ่มจากการดำเนินการแปลงภาพต้นฉบับให้เป็นภาพระดับเทา (Grayscale) และมีการใช้ HOG เพื่อที่จะในการอธิบายคุณลักษณะของภาพ ซึ่ง

วัตถุประสงค์ของการตรวจหาพื้นที่ที่เป็นข้อความไม่ได้ที่จะเป็นการหาตำแหน่งข้อความที่ถูกต้อง แต่จะเป็นการประเมินความน่าจะเป็นถึงตำแหน่งของข้อความ การกำหนดรายละเอียดของ HOG จะนำภาพระดับเทามาแปลงเป็นภาพเกรเดียนต์ด้วยการใช้ Sobel Operator และกำหนดขนาดของหน้าต่างเลื่อน (Sliding Window) ขนาด 16 x 16 เพื่อคำนวณหาทิศทางของเกรเดียนต์ และถังทิศทางเกรเดียนต์ได้กำหนดไว้ทั้งหมด 4 ทิศทาง ต่อมาเป็นการแบ่งส่วนของภาพด้วยวิธีการ Niblack's [90] ซึ่งถูกนำเสนอในปี ค.ศ. 1986 ด้วยวิธีการหาค่าเธรสโฮลต์จากพื้นที่เล็ก ๆ ในรูปแบบของหน้าต่างการประมวลผลแบบเลื่อน (Sliding Window) เพื่อที่จะแบ่งส่วนของภาพระดับเทา ขั้นตอนการสกัดข้อความจะดำเนินการโดยการหาค่าเฉลี่ยของความกว้างและความสูงของข้อความจากค่าของพิกเซลที่มีความคล้ายกัน เพื่อสกัดส่วนที่เป็นข้อความและไม่ใช่ข้อความ

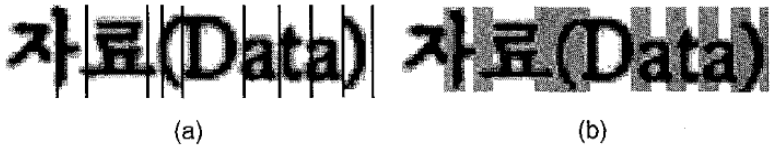
จากการศึกษางานวิจัยที่ผ่านมาพบว่า ปัญหาที่เกิดขึ้นของการตรวจหาพื้นที่ข้อความในภาพ ได้แก่ พื้นหลังที่มีความซับซ้อน ผลกระทบที่เกิดจากความสว่างของแสงที่ไม่เท่ากัน หรือแม่กระทั้งตัวอักษรที่ปรากฏอยู่ในภาพ เช่น รูปแบบตัวอักษร ขนาดตัวอักษร หรือการวางแนวของตัวอักษร เป็นต้น ล้วนส่งผลให้กับประสิทธิภาพของการตรวจหาพื้นที่ข้อความในภาพลดลงทั้งสิ้น

2.2.2 การแบ่งส่วนข้อความ (Text Segmentation)
งานวิจัยที่ทำการศึกษาเกี่ยวกับการแบ่งส่วนข้อความ อาทิ เช่น

ปี ค.ศ. 1996 Lee และคณะ [91] ได้นำเสนอการแก้ปัญหาการตัดแยกตัวอักษรในภาพที่มีส่วนที่สัมผัสและซ้อนทับกัน โดย Lee และคณะได้อธิบายว่าในข้อความที่ถูกพิมพ์ส่วนมากนั้น จะมีแนวโน้มที่ตัวอักษรจะมีส่วนที่ติดกัน สัมผัสกัน หรือซ้อนทับกัน ซึ่งปัญหานี้คือปัจจัยที่สำคัญที่ก่อให้เกิดข้อผิดพลาดในการแบ่งส่วนตัวอักษร การแก้ปัญหาตัวอักษรที่ซ้อนทับกันนั้น Lee และคณะได้นำเสนอวิธีการประยุกต์ใช้คุณสมบัติของ Topographic ซึ่งจะให้ค่าที่ดีที่สุด Topographic จะมีลักษณะเป็นพารามิเตอร์ (Parameter) แบบกระจายพื้นที่ (Distributed Parameter) โดยจะมีความแตกต่างและแปรผันไปตามตำแหน่งและลักษณะของพื้นที่ และการเปลี่ยนแปลงความเข้มข้นในภาพระดับสีเทา (Grayscale) กระบวนการตัดแยกตัวอักษรจะประกอบไปด้วย 3 ขั้นตอนคือ การกำหนดพื้นที่ในการตัดแยกตัวอักษร การค้นหาเส้นการตัดแยกตัวอักษรที่ไม่เป็นเส้นตรงโดยอัลกอริทึม (Algorithm) การค้นหากราฟแบบหลายระยะ และการยืนยันการตัดแยกตัวอักษรที่ไม่เป็นเส้นตรง แสดงดังรูปที่ 99 จะแสดงรูปร่างของ Topographic และคุณลักษณะของภาพ Grayscale และรูปที่ 100 จะรูปแสดงพื้นที่ก่อนการตัดแยกตัวอักษรและพื้นที่การตัดแยกตัวอักษร ซึ่งวิธีการดังกล่าวสามารถนำมาประยุกต์ใช้ในการตัดแยกตัวอักษรที่ปะปนกับพื้นหลังได้



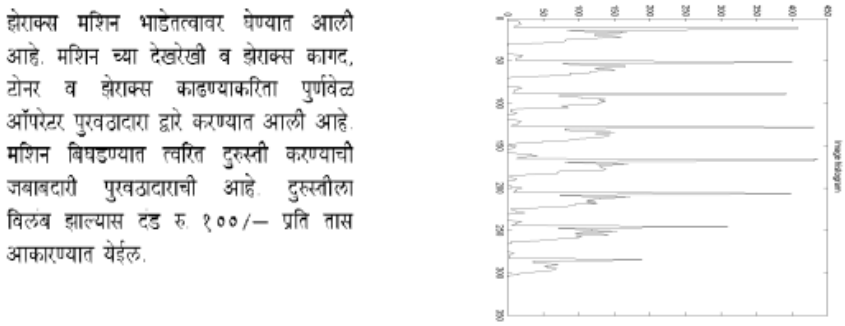
รูปที่ 99 รูปแสดงรูปร่างของ Topographic และคุณลักษณะของภาพ Grayscale, a) รูปภาพของ Grayscale, b) รูปร่างของ 30 Topographic, c) การสกัดคุณลักษณะของ Topographic



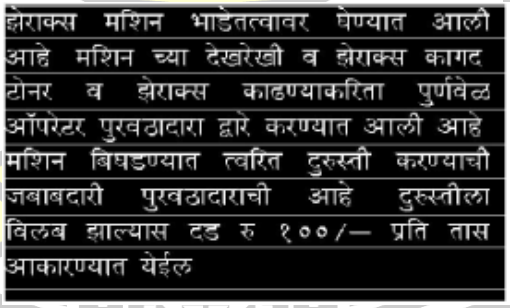
รูปที่ 100 รูปแสดงพื้นที่ก่อนการตัดแยกตัวอักษรและพื้นที่การตัดแยกตัวอักษร a) ผลลัพธ์ก่อนการตัดแยก, b) พื้นที่การตัดแยกตัวอักษรที่ได้รับจาก a

จากผลการทดลองของ Lee และคณะทำให้ทราบถึงเทคนิควิธีการในการตัดแยกตัวอักษร ซึ่งเป็นวิธีการการที่แก้ไขข้อผิดพลาดที่เกิดขึ้นจากกระบวนการ Binarization ในหลาย ๆ กรณี วิธีที่มีประสิทธิภาพมาก สำหรับการแบ่งส่วน และการรับรู้ ที่มีตัวอักษรที่มีการสัมผัส หรืออักษรที่ซ้อนทับกัน แต่ในงานวิจัยนี้ยังเกิดความผิดพลาดในการแบ่งส่วนข้อความซึ่งเกิดขึ้นจากกระบวนการ Binarization เช่น ความเสียหายที่เกิดขึ้นในตัวอักษร

ปี ค.ศ. 2011 Dongre และคณะ[92] ได้นำเสนอวิธีการแก้ปัญหาของการแบ่งส่วนที่เหมาะสมของคำในอักษร Devnagari ซึ่งจะประกอบไปด้วย สระ พยัญชนะ และสัญลักษณ์ต่าง ๆ ด้วยใช้ค่าฮิสโตแกรม (Histogram) วิธีการที่นำเสนอมีดังต่อไปนี้ กระบวนการก่อนการประมวลผล (Pre Processing) เป็นกระบวนการในการเตรียมภาพโดยการแปลงภาพให้เป็นภาพไบนารี โดยกำหนดให้ส่วนที่มีวัตถุจะมีค่าเป็น 1 และส่วนที่เป็นพื้นหลังให้มีค่าเป็น 0 กระบวนการต่อมาคือวิธีการแบ่งส่วนที่มีการดำเนินการด้วย 3 ขั้นตอนหลักคือ การแบ่งส่วนบรรทัดเป็นขั้นตอนในการแบ่งส่วนบรรทัดข้อความ ซึ่งจะใช้ค่าฮิสโตแกรม ในแนวนอนในการตัดแบ่งบรรทัด ซึ่งค่าฮิสโตแกรมนี้จะเป็นค่าที่ได้จากการนับจำนวนพิกเซลที่มีค่าเท่ากับ 1 ดังรูปที่ 101 และ รูปที่ 102



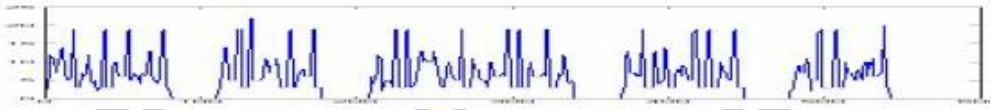
รูปที่ 101 รูปซ้ายคือภาพเอกสารอักษร Devnagari รูปขวาคือรูปฮิสโตแกรม



รูปที่ 102 ผลการแบ่งส่วนบรรทัดข้อความ

ขั้นตอนต่อมาการแบ่งส่วนคำเป็นขั้นตอนในการแบ่งส่วนคำ ซึ่งจะใช้ค่าฮิสโตแกรมในแนวตั้ง โดยวิธีการจะดำเนินการเหมือนขั้นตอนแรกดังรูปที่ 103 - รูปที่ 105

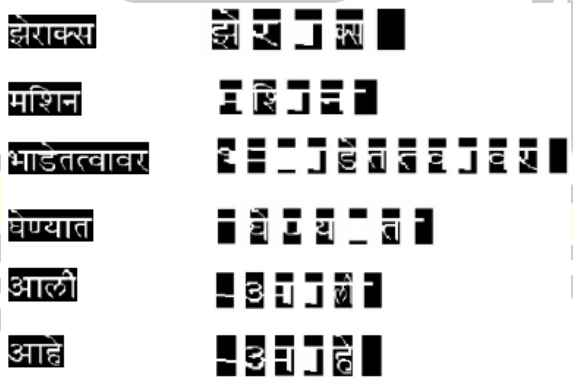
รูปที่ 103 ภาพบรรทัดข้อความ



รูปที่ 104 ฮิสโตแกรมของคำ

รูปที่ 105 ผลการแบ่งส่วนคำ

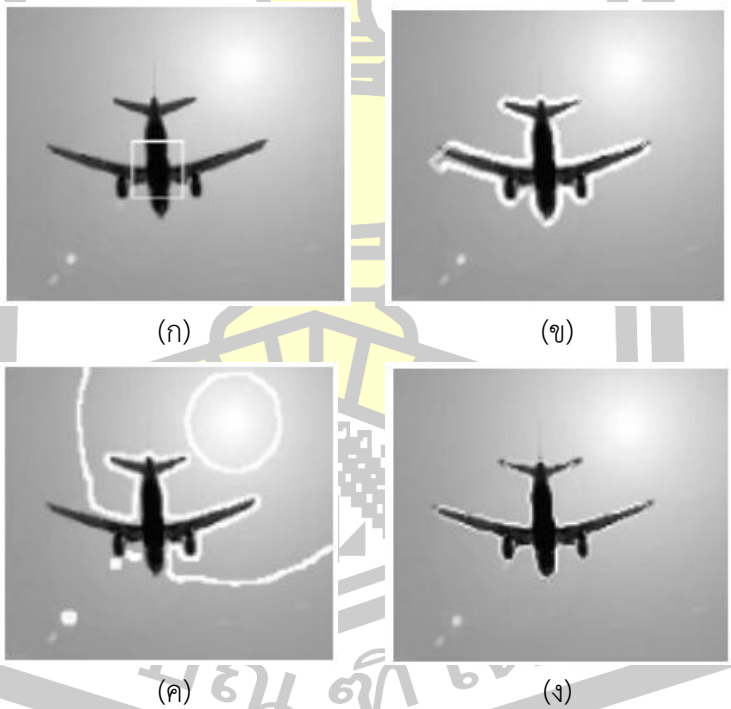
ขั้นตอนสุดท้ายการแบ่งส่วนตัวอักษรเป็นขั้นตอนในการแบ่งส่วนตัวอักษร ซึ่งจะดำเนินการเหมือนกับวิธีการ แบ่งส่วนคำ แต่จะมีการปรับเปลี่ยนเล็กน้อยโดยจะใช้ฟังก์ชัน bwmorph ในโปรแกรม Matlab ในการแปลงภาพ ก่อนแล้วดำเนินการนับพิกเซลในแต่ละคอลัมน์ ดังรูปที่ 106



รูปที่ 106 รูปซ้ายคือพื้นที่ที่สนใจ รูปขวาคือการแบ่งส่วนตัวอักษร

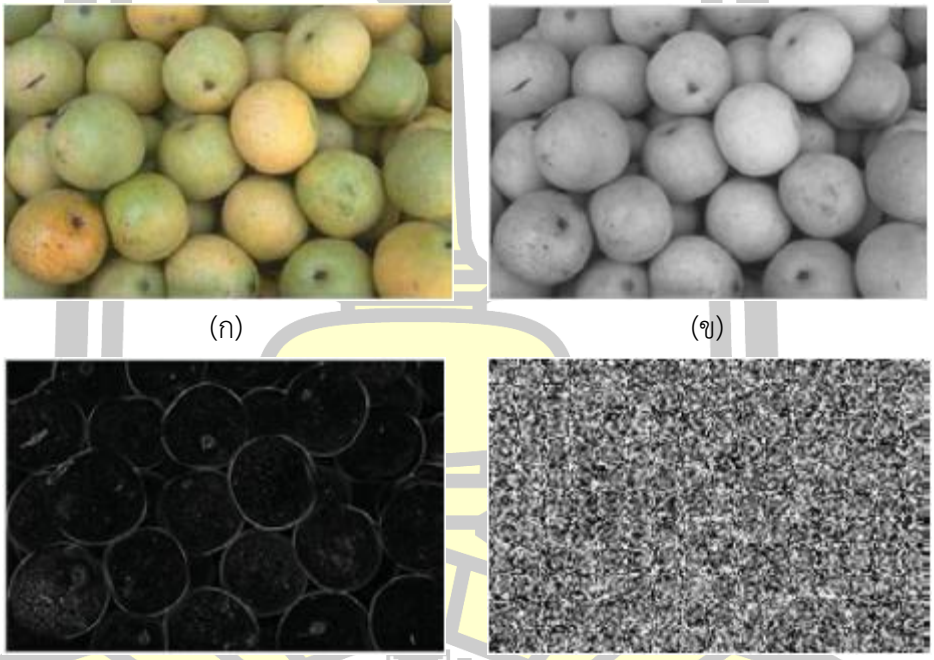
ผลการทดลองวิธีนี้มีประสิทธิภาพมากในการแบ่งส่วนบรรทัดข้อความ ส่วนการแบ่งส่วนคำจะมีความถูกต้องลดน้อยลงหากมีตัวเลขปะบนมาในข้อความซึ่งความถูกต้องจะมีความถูกต้องคิดเป็นร้อยละ 91

ปี ค.ศ. 2012 Xu และคณะ [93] ได้นำเสนอการพัฒนาการแบ่งส่วนภาพด้วยการทำงานร่วมกันของวิธีการ Fast Level Set กับ C-V Model โดย Fast Level Set ได้ถูกนำเสนอโดย Shi's [94] ซึ่งเป็นการพัฒนาวิธีการ Level Set แบบเดิมด้วยการปรับปรุงจุดขอบเขต (Boundary Point) 2 จุดได้แก่ จุดขอบเขตภายนอก และจุดขอบเขตภายใน ด้วยการแก้ไขสมการบางค่าเพื่อเพิ่มความเร็วในการคำนวณ ผลลัพธ์คือเส้นแบ่งของเขตภายนอก และภายในวัตถุ และวิธีการ C-V Model ได้ถูกนำเสนอโดย Chan and Vese [42] เป็นการทำงานร่วมกันระหว่าง Level Set แบบเดิมและ Mumford-Shah Model ที่ถูกนำเสนอโดย Mumford และ Shah [95] วิธีการ C-V Model สามารถตรวจหาขอบเขตของวัตถุได้ โดยไม่จำเป็นต้องกำหนดค่าเกรเดียนต์ ซึ่งเป็นการแก้ไขปัญหาของการหดตัวและเคลื่อนที่ของเส้นโค้งที่ไม่หยุดบนขอบเขตที่ต้องการ ซึ่งในงานวิจัยนี้การนำ C-V Model มาใช้เพื่อเป็นการกำหนดความเร็วของฟังก์ชั่นและลดความซับซ้อนของวิธีการ Fast Level Set ด้วยการปรับปรุงให้มีการใช้จุดขอบเขตเพียงแค่จุดเดียว ผลการทดลองแสดงให้เห็นว่าวิธีการนี้มีความแม่นยำและรวดเร็วมากกว่าวิธีการ C-V Model และวิธีการ Level Set ของ Shi's ดังรูปที่ 107



รูปที่ 107 ผลการแบ่งส่วนภาพ (ก) ภาพเส้นเริ่มต้น (ข) ผลจากวิธีการของ Shi's (ค) ผลจากวิธีการ C-V Model (ง) ผลจากวิธีการที่นำเสนอ
ที่มา [93]

ปี ค.ศ. 2014 Cui และคณะ [56] ได้นำเสนอการแบ่งส่วนภาพด้วยการปรับปรุงวิธีการ Watershed เพื่อลดปัญหาของการแบ่งส่วนภาพที่เกิดขึ้นจากวิธีการ Watershed แบบเดิม โดยวิธีการดำเนินงานมีดังนี้ จากรูปที่ 108 (ก) คือการนำภาพต้นฉบับมาผ่านการแปลงเป็นภาพระดับเทา ดังรูปที่ 108 (ข) (Grayscale) และดำเนินการใช้ Sobel Masks ในการคำนวณหาเส้นขอบของวัตถุในภาพพร้อมทั้งคำนวณหาค่า Gradient Magnitude ดังรูปที่ 108 (ค) จากค่า Gradient Magnitude ที่ได้มี 2 ค่าคือ 1) ค่าเกรเดียนต์สูงคือบริเวณที่เป็นเส้นขอบของวัตถุ และ 2) ค่าเกรเดียนต์ต่ำคือบริเวณที่เป็นส่วนภายในของวัตถุ ถ้าดำเนินการแบ่งส่วนภาพด้วยการใช้วิธีการ Watershed ด้วยค่า Gradient Magnitude ผลที่ได้มักจะเกิดการ Over Segmentation ดังรูปที่ 108 (ง)



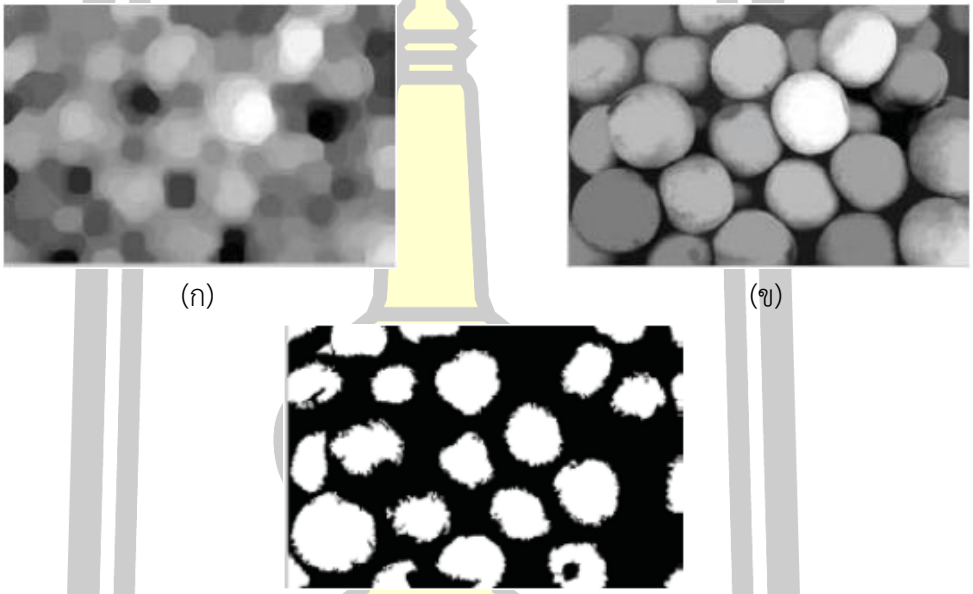
(ก) (ข)

(ค) (ง)

รูปที่ 108 (ก) ภาพต้นฉบับ (ข) ภาพระดับเทา (ค) ภาพ Gradient Magnitude และ (ง) ภาพ Gradient Watershed
ที่มา [56]

ดังนั้น Cui และคณะจึงมีแนวคิดที่จะพัฒนาการทำเครื่องหมายเพื่อระบุวัตถุที่ซึ่งอยู่บริเวณด้านหน้าของภาพ และกำหนดบริเวณที่เป็นพื้นหลังของวัตถุ เพื่อให้วิธีการแบ่งส่วนมีประสิทธิภาพมากกว่าเดิมด้วยการปรับปรุงวิธีการคำนวณให้ดีขึ้น ซึ่งจะเรียกเทคนิคนี้ว่า Opening และ Closing ในการ

ปรับปรุงภาพให้ดีขึ้น ทั้งนี้วิธีการ Opening คือการคำนวณด้วยวิธีการ Erosion และตามด้วยวิธีการ Dilation และวิธีการ Closing คือการคำนวณด้วยวิธีการ Dilation และตามด้วยวิธีการ Erosion โดยทั้งสองวิธีการนี้สามารถช่วยลดรายละเอียดบางอย่างที่ไม่ต้องการของภาพที่มีขนาดเล็กได้ และ Closing นั้นสามารถที่จะเติมเต็มในส่วนที่เป็นช่องว่างที่มีขนาดเล็กในภาพได้ จากรูปที่ 109 (ก) และรูปที่ 109 (ข) แสดงให้เห็นถึงผลลัพธ์ที่เกิดจากวิธีการ Opening และ Closing ที่ได้ถูกพัฒนาขึ้นมาใหม่ ซึ่งมีประสิทธิภาพการในการกำจัดส่วนที่ไม่ต้องการที่มีขนาดเล็กภายในภาพได้ดีมากกว่า Opening และ Closing แบบมาตรฐาน ขั้นตอนต่อมาเป็นการคำนวณหาพื้นที่ที่มีค่าสูงสุดจากภาพรูปที่ 109 (ข) เพื่อจะได้ Foreground Markers ที่ดีที่สุดดังรูปที่ 109 (ค)

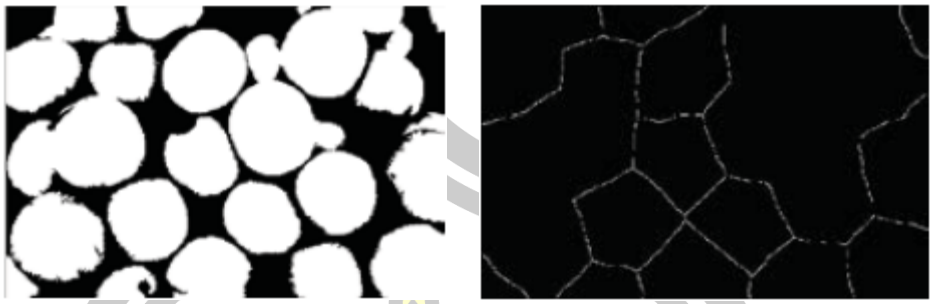


(ก) (ข)

(ค)

รูปที่ 109 (ก) Opening และ Closing แบบเดิม (ข) Opening และ Closing ที่พัฒนาใหม่ และ(ค) Foreground Markers
ที่มา [56]

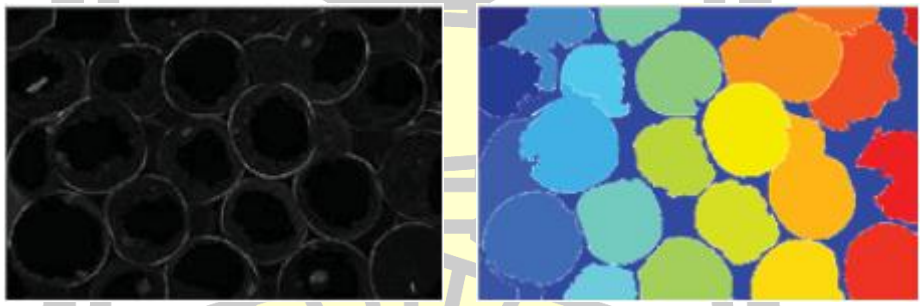
ต่อมาในการคำนวณหา Background Markers ดังรูปที่ 110 (ก) โดยดำเนินการด้วยการใช้เทคนิคการหาค่า เทรชโฮลด์ (Thresholding) ซึ่ง Background นั้นจะอยู่ในบริเวณที่ให้ค่าพิกเซลเป็นสีดำ แต่ Cui และคณะได้ให้แนวคิดว่าไม่ต้องการให้ Background Markers มีส่วนที่ติดกับเส้นขอบของวัตถุมากเกินไป จึงนำวิธีการ Skeleton มาใช้เพื่อหาบริเวณของ Backgound ดังรูปที่ 110 (ข)



(ก) (ข)

รูปที่ 110 (ก) ภาพการแบ่งส่วนด้วยเทรชโฮลด์ และ(ข) Background Markers
ที่มา [56]

ถัดมาเป็นการปรับปรุงภาพ ด้วยการกำหนดบริเวณที่มีค่าต่ำสุด และบริเวณที่มีค่าสูงสุด ซึ่งจะทำให้เกิด Foreground Markers และ Background Markers ดังรูปที่ 111 (ก) สุดท้ายคือการคำนวณด้วยวิธีการ Watershed จากรูปที่ 111 (ข) วิธีการนี้แสดงให้เห็นว่าในการปรับปรุงส่วนภาพจะได้ผลที่ดี ในการลดการ Over Segmentation



(ก) (ข)

รูปที่ 111 (ก) โครงสร้างของภาพที่พัฒนาขึ้นใหม่ (ข) ภาพผลลัพธ์จากวิธีการของ Cui และคณะ
ที่มา [56]

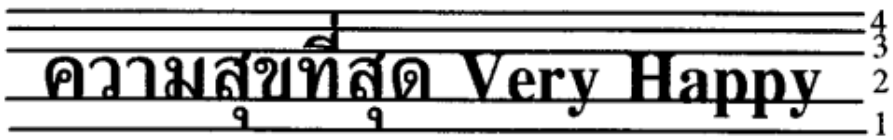
จากการศึกษางานวิจัยที่ผ่านมาพบว่า ปัญหาที่เกิดขึ้นในการแบ่งส่วนข้อความในภาพคือรูปแบบตัวอักษรที่ใช้ เช่น รูปแบบตัวอักษรภาษาจีน รูปแบบตัวอักษรภาษา Devnagari หรือแม้กระทั้งภาษาไทย โดยรูปแบบตัวอักษรแต่ละภาษานั้นมีเอกลักษณ์เป็นของตัวเองดังนั้นจึงเป็นเรื่องยากที่จะใช้วิธีการใดวิธีการหนึ่งในการแบ่งส่วนข้อความในภาพให้เหมาะกับทุก ๆ รูปแบบภาษา ทั้งนี้

ทางผู้วิจัยจึงได้ศึกษางานวิจัยเพิ่มเติมในการแบ่งส่วนภาพด้วยวิธีการแอ็กทีฟคอนทัวร์ เพื่อศึกษาความเป็นไปได้ของวิธีการดังกล่าวจากปัญหาที่พบ

2.2.3 การรู้จำตัวอักขระ (Character Recognition)

งานวิจัยที่ทำการศึกษาเกี่ยวกับการรู้จำตัวอักขระ อาทิ เช่น

ปี ค.ศ. 1999 Tanprasert และSae-Tang [96] เนื่องจากการรู้จำตัวอักษรในภาษาไทยนั้นมีความนิยมในการทำวิจัยเป็นอย่างมากภายในประเทศไทย โดยมีหลากหลายบริษัทที่ได้มีการให้บริการที่เรียกว่า "Thai OCR" แต่ยังไม่มีบริษัทใดเลยที่สามารถรักษารูปแบบของตัวอักษรให้เหมือนต้นฉบับได้ เช่น ตัวหนา ตัวเอียง ตัวหนาผสมตัวเอียง เป็นต้น จากปัญหาดังกล่าวงานวิจัยของ Tanprasert และSae-Tang จึงได้นำเสนอเทคนิคการรักษารูปแบบตัวอักษรด้วยการประมวลผลด้วยโครงข่ายประสาทเทียม ในการรู้จำภาษาไทยค่อนข้างจะมีปัญหาที่ซับซ้อนมาก สาเหตุที่ทำให้ภาษาไทยมีความยากกว่าภาษาอื่นเกิดขึ้นจาก 1) ระดับของประโยคในภาษาไทยมีหลายระดับดังรูปที่ 112 2) ตัวอักษรในภาษาไทยมีความคล้ายคลึงกันมาก 3) ไม่มีช่องว่างระหว่างตัวอักษรในภาษาไทย จึงทำให้เกิดความยากลำบากในกระบวนการสกัดตัวอักษร (Segmentation)



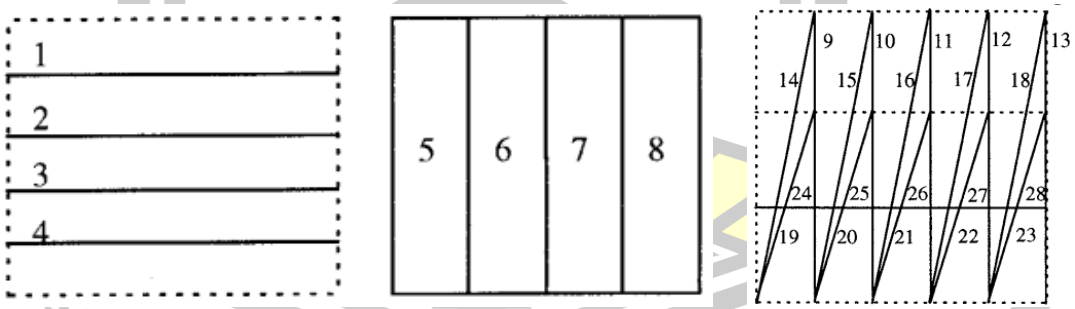
รูปที่ 112 ระดับของตัวอักษรในภาษาไทย
ที่มา [96]

วิธีการที่นำเสนอมีกระบวนการดังนี้ ขั้นตอนการเตรียมข้อมูลจะประกอบไปด้วย การแปลงข้อมูลเป็นไบนารี่ (Binary) โดยกำหนดให้พิกเซลที่มีสีขาวมีค่าเป็น 0 และพิกเซลที่มีสีดำมีค่าเป็น 1 ในงานวิจัยนี้จะพิจารณาตัวอักษรที่อยู่ในระดับที่ 2 เท่านั้น อีกทั้งในงานวิจัยนี้ยังได้กำหนดกลุ่มให้กับตัวอักษร โดยจะจำแนกออกเป็น 3 กลุ่มตามจำนวนขาของตัวอักษร คือ กลุ่มของตัวอักษรที่มี 1 ขา 2 ขา และมากกว่า 2 ขา ดังรูปที่ 113 ซึ่งการแบ่งกลุ่มให้กับตัวอักษรนี้จะช่วยลดความซับซ้อนให้กับกระบวนการจัดกลุ่มของโครงข่ายประสาทเทียมได้

Number of Members	Number of Legs	Member of the Group
13	1	ร เ ไ ไ ง จ ฐ ธ ว ฯ า ๆ
42	2	ก ข ค ฉ ช ฎ ฏ ฑ ถ ท น บ ป ผ ฝ ภ ม ย ฤ ล ฦ ศ ษ ส ห ฟ อ ฮ ฆ ง จ ฐ ด ต ธ พ ฟ ว ฯ า ๆ
9	>2	ณ ญ ฒ ฌ ฆ ด ต พ ฟ

รูปที่ 113 ตัวอักษรที่เป็นสมาชิกใน 3 กลุ่ม
ที่มา [96]

ขั้นตอนต่อมาจะเป็นการปรับขนาดของภาพไบนารีให้เหลือ 32x32 พิกเซลเพื่อเป็นการลดความซับซ้อนและสร้างความสมดุลให้กับภาพที่มีขนาดที่แตกต่างกัน รวมถึงความละเอียดของภาพที่แตกต่างกัน ขั้นตอนถัดมาเป็นการสกัดคุณลักษณะ โดยในการสกัดคุณลักษณะนี้จะเป็นการนับจำนวนพิกเซลที่ตกอยู่ในช่องของแม่แบบนั้น ๆ โดยจะแบ่งแม่แบบ (Template) ออกเป็น 3 รูปแบบ ซึ่งแบบแรกจะเป็นการสกัดคุณลักษณะความหนาบางของตัวอักษรในแนวนอน แบบที่สองเป็นการสกัดความหนาบางของตัวอักษรในแนวตั้ง และสุดท้ายเป็นการหาความแตกต่างระหว่างตัวอักษรที่มีลักษณะหางยาว หรือหางสั้น ดังรูปที่ 114 จากการสกัดคุณลักษณะทำให้ได้คุณลักษณะทั้งหมด 28 คุณลักษณะ



รูปที่ 114 แม่แบบ (Template) ทั้ง 3 รูปแบบ
ที่มา [96]

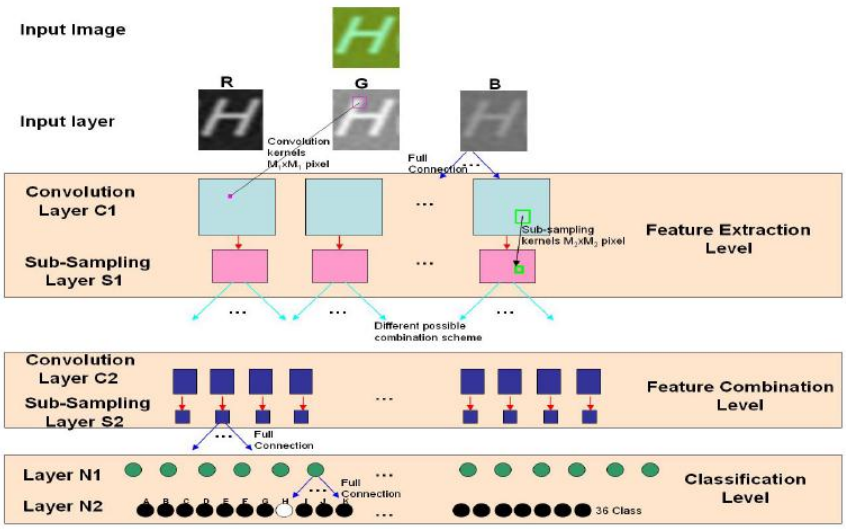
ขั้นตอนสุดท้ายเป็นการนำคุณลักษณะที่ได้ทั้งหมดมาผ่านกระบวนการเรียนรู้ด้วยอัลกอริทึมโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back-Propagation) เพื่อใช้ในการจำแนกประเภทของตัวอักษร

จากผลการทดลองมีอัตราการรู้จำเฉลี่ยอยู่ที่ร้อยละ 93.31 ซึ่งเป็นการยืนยันได้ว่าเทคนิคที่นำเสนอสามารถรักษารูปแบบตัวอักษรภาษาไทยได้อย่างมีประสิทธิภาพ

ปี ค.ศ. 1999 Sawaki และคณะ [97] ได้นำเสนอวิธีการรู้จำตัวอักษรจากภาพหนังสือที่อยู่บนชั้นวางหนังสือ ความต้องการของงานวิจัยนี้คือ การทราบถึงตัวอักษรที่อยู่บนสันหนังสือ ซึ่งปัญหาที่เกิดขึ้นนั้นเกิดจากรอยพับ หรือรอยขีดข่วนของสันหนังสือ ทำให้ตัวอักษรที่อยู่บนสันหนังสือนั้นไม่ชัดเจน โดยงานวิจัยนี้ได้นำเสนอวิธีการจับคู่ (Matching) กับแม่แบบ (Templates) ด้วยการหาความใกล้เคียงกันของตัวอักษร ซึ่งแม่แบบที่นำมาใช้นี้จะถูกเรียกว่า Context-Base Image Templates แม่แบบนี้ถูกนำมาใช้เพื่อ เพิ่มความถูกต้องของการรู้จำตัวอักษร ทั้งนี้ตัวแม่แบบจะไม่เก็บเฉพาะข้อมูลตัวอักษรเพียงอย่างเดียวเท่านั้นแต่ จะเก็บข้อมูลของตัวอักษรที่อยู่ใกล้เคียงด้วย กระบวนการรู้จำในงานวิจัยนี้ดำเนินการโดยการแปลงภาพให้เป็นภาพระดับเทา และแปลงเป็นภาพไบนารี่ ในการกำหนดขอบเขตของหนังสือจะพิจารณาจากเส้นสีดำที่เกิดขึ้นจากเงา โดยเส้นสีดำนี้เกิดขึ้นจากพื้นที่ระหว่างหนังสือ ซึ่งสามารถพิจารณาได้จากความถี่ของสีที่เกิดขึ้น (การเปลี่ยนแปลงของสีจากสีดำ ไปสีขาว จากสี่ขาว ไปสีดำ) เส้นสีดำนี้จะถูกเรียกว่าบรรทัดข้อความ (Text-Line) หลังจากนั้นจะสร้างหน้าต่าง (Window) ให้ดำเนินการวิ่งไปตามแนวของบรรทัดข้อความ ทั้งนี้ข้อมูลที่ได้ในหน้าต่างอาจจะมีบางส่วนของตัวอักษรที่อยู่ใกล้เคียงติดมาด้วยทั้งด้านหน้าและด้านหลัง ขั้นต่อมาดำเนินการวัดความเหมือนของข้อมูลในหน้าต่างที่ได้มาเปรียบเทียบกับแม่แบบที่เก็บไว้ด้วยวิธีการจับคู่ ซึ่งแม่แบบที่นำมาเปรียบเทียบจะมี 2 แบบคือ 1) Single-Templates คือแม่แบบที่เก็บข้อมูลตัวอักษรเพียงหนึ่งตัวเท่านั้น และ 2) Multiple-Templates คือแม่แบบที่เก็บข้อมูลตัวอักษรที่อยู่ใกล้เคียงที่ปะปนมาด้วย วิธีการจับคู่จะดำเนินการโดยการนับจำนวนพิกเซลสีดำที่อยู่ในหน้าต่างเปรียบเทียบกับจำนวนพิกเซลสีดำที่อยู่ในแม่แบบ จากการทดลองแสดงให้เห็นว่าวิธีที่นำเสนอประสบความสำเร็จในอัตราการรู้จำคิดเป็นร้อยละ 96.3 สำหรับ Multiple-Templates และอัตราการรู้จำสำหรับ Single-Templates คิดเป็นร้อยละ 88.4

ปี ค.ศ. 2007 Saidane และคณะ[98] ได้นำเสนอวิธีการพัฒนาการรับรู้แบบอัตโนมัติสำหรับการสกัดอักษรข้อความสี จากรูปภาพ ซึ่งจะมีประสิทธิภาพมากต่อ การบิดเบือนที่รุนแรง การซ้อนทับของพื้นหลัง และ ความละเอียดต่ำ โดย Saidane และคณะได้ให้ข้อสังเกตว่า วิธีการที่ผ่านมาก่อนหน้าจะเป็นวิธการที่จะต้องมีขั้นตอนก่อนการประมวลผล (Pre-processing) ซึ่งเป็นขั้นตอนที่สำคัญในบางงานที่ผ่านมา ดังนั้นวิธีการที่ Saidane และคณะนำเสนอนั้นจะเป็นวิธีการรู้จำตัวอักษร

แบบอัตโนมัติ สำหรับภาพข้อความที่มีฉากสีธรรมชาติ โดยอยู่บนพื้นฐานของการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งจะไม่มีการประยุกต์ใช้กระบวนการ Pre-Processing ดังรูปที่ 115

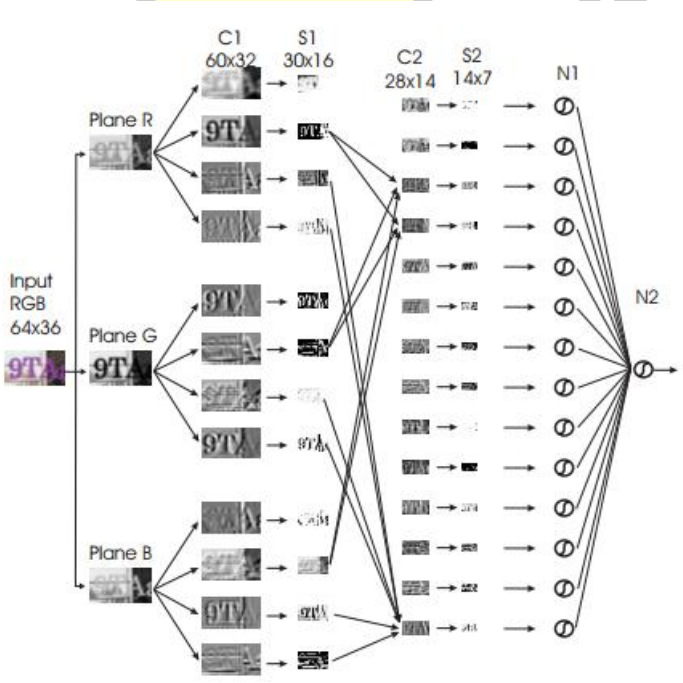


รูปที่ 115 รูปสถาปัตยกรรมของโครงข่ายประสาทเทียม

วิธีการดำเนินงานที่นำเสนอจะเป็นประยุกต์ใช้โครงข่ายประสาทเทียม (Convolutional Neural Networks) ในการเรียนรู้รูปแบบของตัวอักษรด้วยการป้อนภาพที่เป็นภาพสี RGB และทำการ Normalized ภาพให้มีค่าสีอยู่ในช่วง [-1, 1] พร้อมทั้งทำการแยกแยะความแตกต่างออกเป็นเป็นสามชั้น (Layer) ได้แก่ชั้นที่ 1 เป็นชั้นของการสกัดคุณลักษณะ โดยจะอาศัยชั้น C1 และชั้น S1 ซึ่งชั้น C1 จะสกัดคุณลักษณะเฉพาะบางอย่างเช่น Edges Corners และ End Points ของแต่ละช่องสี และชั้น S1 จะเป็นผลของค่าเฉลี่ยของชั้น C1 ที่อยู่ก่อนหน้า ชั้นที่ 2 เป็นชั้นของการนำคุณลักษณะต่าง ๆ มารวมกัน โดยจะอาศัยชั้น C2 และชั้น S2 โดยชั้น C2 จะช่วยในการสกัดข้อมูลที่มีความซับซ้อนมากด้วยการนำผลจากชั้น S1 ที่มีคุณลักษณะที่แตกต่างกันนำมารวมกัน และชั้น S2 จะเป็นผลของค่าเฉลี่ยของชั้น C2 ที่อยู่ก่อนหน้า และชั้นสุดท้ายชั้นที่ 3 เป็นชั้นของการจำแนก ซึ่งจะอาศัยชั้น N1 และชั้น N2 ด้วยการพิจารณารูปแบบตัวอักษรและตัวเลขที่นำเข้ามาทั้งหมด 62 กลุ่ม (Class) โดยจะจำแนกเป็น ตัวอักษรพิมพ์เล็ก 26 ตัว ตัวอักษรตัวพิมพ์ใหญ่ 26 ตัว และตัวเลขอีก 10 ตัว ผลการทดลองมีการเปรียบเทียบกับเทคนิคการรับรู้ตัวอักษรแบบอื่น ๆ ด้วยการใช้ข้อมูลของ ICDAR 2003 Public Samples Dataset ในการเปรียบเทียบประสิทธิภาพ ซึ่งผลการทดลองแสดงให้ว่า อัตราการรับรู้เฉลี่ยอยู่ที่ร้อยละ 84.53 ถึง 93.47 ของรูปภาพปกติ และร้อยละ 67.86 สำหรับภาพที่ผิดปกติ

ปี ค.ศ. 2007 Zhou และคณะ [99] ได้นำเสนองานวิจัยในการเรียนรู้จำตัวอักษรภาษาญี่ปุ่นร่วมกับการวิเคราะห์บริทบทางเรขาคณิต (Geometric Context) โดยปัญหาในการรู้จำภาษาญี่ปุ่นนี้คือ ตัวอักษรนั้นไม่มีการแยกออกจากกันอย่างเห็นได้ชัดเนื่องจาก ความแปรปรวนของขนาดและระยะหว่างตัวอักษร ซึ่งในการรู้จำตัวอักษรที่มีประสิทธิภาพได้นั้นจำเป็นต้องอาศัยการแบ่งส่วน (Segmentation) ที่ดี บริทบทางเรขาคณิตจะมีการพิจารณาได้แก่ ความสูง ความกว้าง ผลรวมของช่องว่างภายในกรอบ Bounding Box รากที่สองของพื้นที่กรอบ Bounding Box ความยาวของเส้นทะแยงมุมกรอบ Bounding Box และความสูงเฉลี่ย ซึ่งสามารถช่วยในการแบ่งส่วนตัวอักษร ผลการทดลองแสดงให้เห็นว่าการนำบริทบทางเรขาคณิตมาใช้ร่วมกับการรู้จำตัวอักษรสามารถเพิ่มประสิทธิภาพในการแบ่งส่วน และการรู้จำตัวอักษรได้อย่างมีประสิทธิภาพ

ปี ค.ศ. 2008 Delakis และคณะ[100] แนะนำแนวทางในการประยุกต์กระบวนการโครงข่ายประสาทเทียม โดยระบบการตรวจหาข้อความในภาพของเขานั้น สามารถเรียนรู้การสกัดข้อความแบบอัตโนมัติด้วยตัวมันเอง นอกจากนี้ระบบของเขาไม่ได้เรียนรู้เพียงการตรวจหาข้อความเพียงอย่างเดียว แต่ยังมีความสามารถในการปฏิเสธพื้นที่ ๆ ไม่ใช่ข้อความได้ อีกทั้งเทคนิคการตรวจหาบรรทัดข้อความในทางแนวนอน (Horizontal) ที่อยู่ในรูปภาพได้อีกด้วย ดังรูปที่ 116

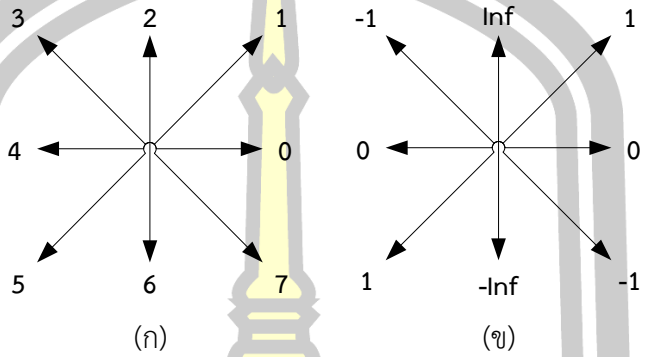


รูปที่ 116 การประยุกต์ใช้โครงข่ายประสาทเทียมของ Delakis

วิธีการดำเนินงานที่นำเสนอจะเป็นประยุกต์ใช้โครงข่ายประสาทเทียม ด้วยการรับข้อมูลภาพสี RGB ขนาด 64 x 32 พิกเซล และภาพจะถูกแยกออกเป็น 3 ช่องสีคือช่องสี R ช่องสี G และช่องสี B โดยมีรายละเอียดแต่ละชั้นดังนี้ ชั้น C1 จะประกอบไปด้วย 12 ชุดข้อมูลซึ่งจะถูกเรียกว่า Feature maps โดยแต่และ Maps จะถูกนำมาผ่านกระบวนการ Convolution ด้วย Mask ขนาด 5x5 ตามด้วย Sub sampling คือชั้น S1 ชั้นนี้ถูกสร้างขึ้นเพื่อเพิ่มประสิทธิภาพให้กับโครงข่ายประสาทเทียมในการป้อนข้อมูลที่มีความผิดปกติ ต่อมาในชั้นที่ C2 จะมี 14 คุณลักษณะ โดยจะมีการเชื่อมต่อคุณลักษณะจากชั้น S1 เอาท์พุทจะเป็นการผสมกันระหว่าง คุณลักษณะของ Maps ที่จะช่วยในการสกัดข้อมูลที่ซับซ้อนมากขึ้น การดำเนินงานของชั้น S2 และชั้น C2 จะคล้ายกับชั้น S1 และชั้น C1 ตามลำดับมีความแตกต่างเพียงอย่างเดียวคือกระบวนการ Convolution ด้วย Mask ขนาด 3 × 3 ในชั้น C2 จากงานวิจัยนี้ทำให้ทราบถึงเทคนิควิธีการที่มีความแม่นยำในการจัดการบรรทัดของข้อความโดยการฝึก (Train) โครงข่ายประสาทเทียม ซึ่งสามารถประยุกต์นำเทคนิควิธีการนี้มาใช้กับรูปภาพที่มีพื้นหลังที่ซับซ้อนและสภาพแวดล้อมอิสระ

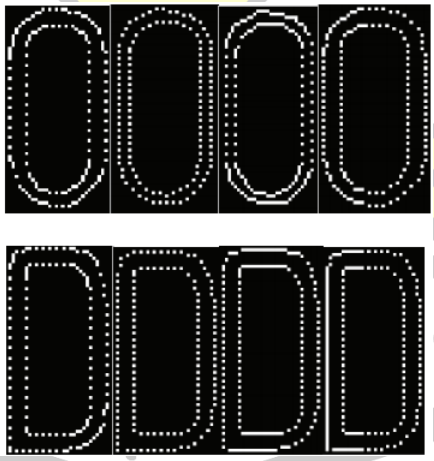
ปี ค.ศ. 2014 Mammeri และคณะ [101] ได้นำเสนอการรู้จำข้อความที่อยู่บนแผ่นป้ายจราจรเพื่อเป็นต้นแบบสำหรับการพัฒนาระบบการขนส่งอัจฉริยะ โดยมีวัตถุประสงค์ในการสนับสนุนคนขับรถยนต์เมื่อพบข้อความที่สำคัญบนแผ่นป้ายจราจร Mammeri และคณะได้อธิบายถึงปัญหาที่เกิดขึ้น ซึ่งมีความท้าทายในการประมวลผลภาพที่ได้จากรถยนต์ขณะเคลื่อนที่ดังนี้ 1) ปัญหาของรูปแบบ ขนาด สี และการวางแนวของตัวอักษร 2) ปัญหาของกล้องติดรถ ที่จะได้รับข้อมูลภาพสั่นไหว ซึ่งได้รับผลกระทบจากพื้นผิวถนนที่ขรุขระ 3) ปัญหาของข้อความที่มีขนาดเล็กและพร่ามัว และ 4) ปัญหาจากความสว่างของแสงที่มีต่อข้อความ ซึ่งระบบการรู้จำในงานวิจัยจะการดำเนิน 1) ขั้นตอนการตรวจจับ (Detection) ขั้นตอนนี้จะมีการใช้ Histogram of Oriented gradients (HOG) ในการสกัดคุณลักษณะ และนำไปผ่านการฝึกสอนด้วย Support Vector Machine (SVM) โดยการกำหนดการจำแนกกลุ่มของภาพที่ได้ออกเป็น 2 กลุ่มคือ ภาพที่เป็นแผ่นป้ายสัญญาณ หรือไม่ใช่แผ่นป้ายสัญญาณ 2) ขั้นตอนการกรองภาพ (Filtering) จะใช้ Gaussian filler ซึ่งเป็นการปรับแต่งภาพเพื่อลดผลกระทบที่มีต่อภาพแผ่นป้าย 3) ขั้นตอนการรู้จำ (Recognition) ในการรู้จำในงานวิจัยนี้ให้นำโปรแกรม Tesseract มาช่วยในการรู้จำ ในการประเมินเพื่อวัดประสิทธิภาพของวิธีการแบ่งออกเป็น 2 ประเภทคือ 1) กลุ่มตัวเลข เช่น แผ่นป้ายจำกัดความเร็ว 2) กลุ่มแผ่นป้ายข้อความ เช่น ป้ายหยุด ผลงานวิจัยพบว่าการนำโปรแกรม Tesseract มาใช้ร่วมกับขั้นตอนที่แนะนำสามารถใช้งานได้อย่างมีประสิทธิภาพในการอ่านข้อความจากแผ่นป้ายจราจร

ปี ค.ศ. 2014 Yuzhe และคณะ [102] ได้นำเสนอการรู้จำอักขณะบนแผ่นป้ายทะเบียนด้วยบริบทรูปร่าง งานวิจัยนี้ได้มีการปรับปรุงวิธีการ Freeman Chain Code ให้มีประสิทธิภาพมากขึ้นและเข้าใจได้ง่ายขึ้น โดยการปรับค่าทิศทางของ Freeman Chain Code ใหม่ดังรูปที่ 117



รูปที่ 117 (ก) ค่าทิศทางของ Freeman Chain Code แบบเดิม (ข) ค่าทิศทางของ Freeman Chain Code แบบใหม่

ภาพที่นำเข้ามาใช้ในงานวิจัยจะมีการกำหนดขนาดอยู่ที่ 32x16 พิกเซลต่อหนึ่งตัวอักษร กระบวนการดำเนินงานวิจัยจะเริ่มจากการแปลงภาพตัวอักษรด้วยการหาขอบของภาพ (Edge Detection) แล้วแปลงขอบของภาพด้วย Freeman Chain Code แบบใหม่ที่ได้พัฒนาขึ้นดังรูปที่ 118



รูปที่ 118 ผลการแปลงภาพด้วยวิธีการ Freeman Chain Code แบบใหม่ ที่มา [102]

ขั้นตอนต่อมาเป็นการหาค่าล็อกโพล่าฮิสโตรแกรม (Log-Polar Histogram หรือ Log-Polar) เพื่อกำหนดคุณลักษณะของตัวอักษร ขั้นสุดท้ายเป็นการจับคู่ตัวอักษรด้วยค่าที่ได้จากโพล่าฮิสโตรแกรม โดยการพิจารณาค่าความเหมือนที่มีค่าน้อยที่สุด โดยผลการทดลองพบว่าวิธีการนี้สามารถระบุตัวอักษรที่มีความคล้ายคลึงกันได้อย่างมีประสิทธิภาพในระดับหนึ่งเท่านั้น

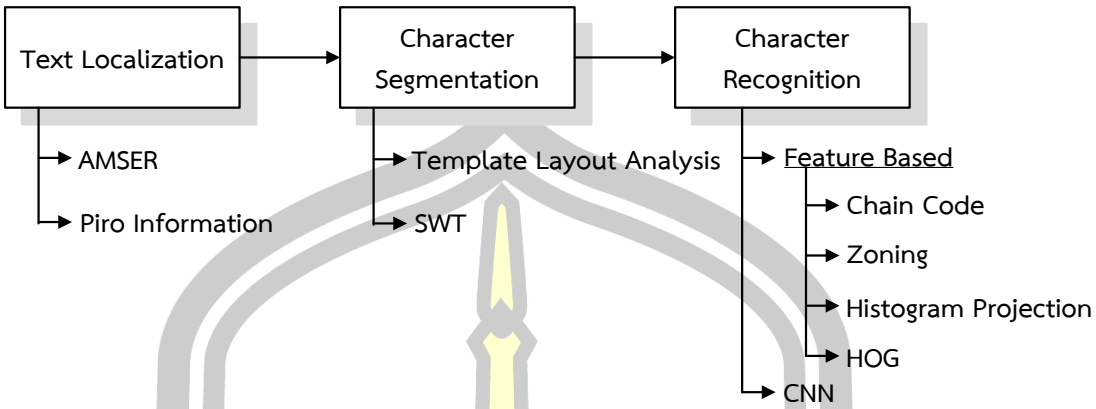
จากการศึกษางานวิจัยที่ผ่านมาพบว่า ปัญหาที่เกิดขึ้นในการรู้จำตัวอักขระ เป็นการเลือกวิธีการในการสกัดคุณลักษณะและเลือกคุณลักษณะที่เหมาะสมกับข้อมูล ปริมาณของข้อมูลที่ใช้ในการฝึกสอน รวมถึงอัลกอริทึมที่ใช้ในการรู้จำ

บทที่ 3
การตรวจจับข้อความในภาพ

การตรวจจับและรู้จำป้ายข้อความในภาพ (Text Detection and Recognition) ประกอบไปด้วยขั้นตอนการทำงาน 2 ขั้นตอนการทำงานหลัก ได้แก่ การระบุตำแหน่งของข้อความในภาพ (Text Detection) และการรู้จำข้อความที่ตรวจจับได้ (Text Recognition) การตรวจจับและรู้จำข้อความในภาพถือได้มีความท้าทายอย่างมาก เนื่องจากข้อมูลภาพนั้นมีความหลากหลายและไม่มีรูปแบบที่ตายตัว โดยความท้าทายของการตรวจจับและรู้จำข้อความในภาพนั้นมีดังต่อไปนี้

- ความหลากหลายของภาพเนื่องจากเป็นภาพทั่วไป (Natural Scene Images)
- มีความไม่แน่นอนของตำแหน่งข้อความในภาพ (Spatial Location Variation)
- มีความหลากหลายของข้อความในภาพ รูปแบบอักษร สี และแสง (Morphology Variation)
- มีความหลากหลายของการจัดเรียงข้อความ (Layout Variation)
- มีความหลากหลายของขนาดของข้อความ (Scale Variation)

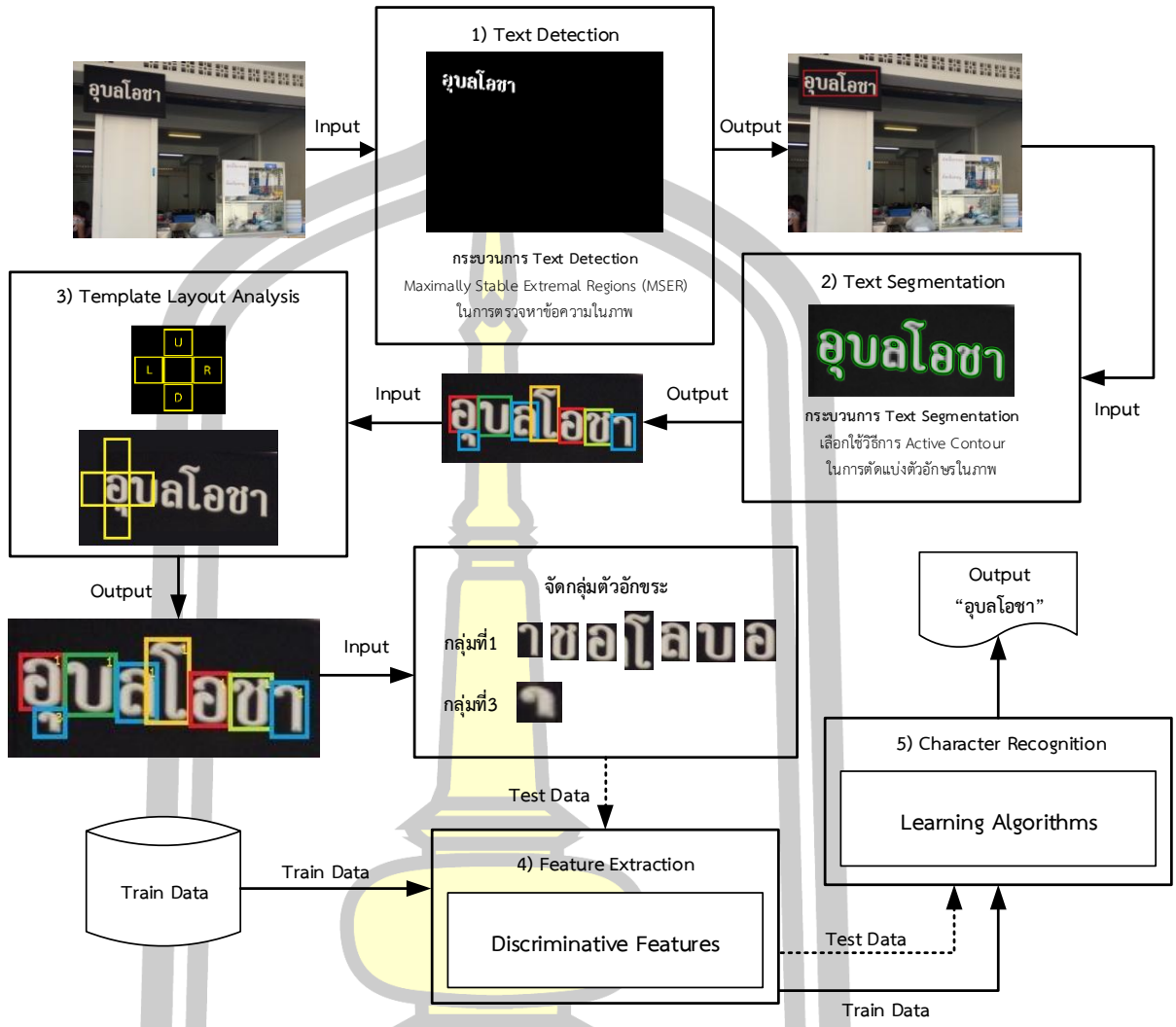
จะเห็นได้ว่าความท้าทายของการตรวจจับและรู้จำข้อความในภาพนั้นส่วนใหญ่จะมาจากข้อมูลที่ไม่มีความแน่นอน (Free From Data) ดังนั้นเทคนิคที่ใช้ในการตรวจจับข้อความทั่วไปอาจจะให้ผลลัพธ์ที่ไม่ดีเท่าที่ควร นอกจากนั้นรูปแบบของอักขระที่ปรากฏอยู่ในภาพทำให้การรู้จำอักขระในข้อความนั้นเป็นได้ยาก การตรวจจับและรู้จำข้อความนั้นจำเป็นอย่างยิ่งที่จะต้องใช้เทคนิคทางด้านการประมวลผลภาพและการเรียนรู้ของเครื่อง (Machine Learning Techniques) มาใช้ในการประมวลผล ปริญญานิพนธ์มีวิธีการดำเนินงานและการทดลองในภาพรวมแสดงดังรูปที่ 119 พร้อมทั้งมีการออกแบบวิธีสำหรับการตรวจจับและรู้จำข้อความในภาพ ดังต่อไปนี้ 1. การตรวจหาพื้นที่ข้อความในภาพ (Text Localization) 2. การแบ่งส่วนข้อความ (Text Segmentation) 3. การวิเคราะห์การจัดเรียงตัวอักษร (Layout Analysis) 4. การสกัดคุณลักษณะ (Feature Extraction) และ 5. การรู้จำตัวอักขระ (Character Recognition) แสดงดังรูปที่ 120



รูปที่ 119 ภาพรวมกระบวนการดำเนินงาน

ในบทนี้จะนำเสนอวิธีการในการตรวนจับข้อความในภาพทั่วไป โดยการตรวจจับข้อความในภาพนั้นถือเป็นกระบวนการเริ่มต้นในการวิจัยในครั้งนี้ และการตรวจจับข้อความในภาพนั้นจะมีผลกับการรู้จำข้อความในภาพ (ซึ่งเป็นกระบวนการถัดไปและจะอธิบายในบทที่ 4) ในปริญญานิพนธ์นี้จะนำเสนอเทคนิคที่นิยมใช้สำหรับการตรวจจับข้อความในภาพ และจะนำเสนอเทคนิคใหม่ที่อาศัยข้อมูลเรียนรู้ก่อนหน้า (Prior Knowledge) มาใช้ในการประมาณตำแหน่งของข้อความในภาพ รวมถึงการทำการตรวจจับข้อความในภาพโดยใช้เทคนิคการแยกส่วนภาพหลายระดับ และการใช้เทคนิคทางด้านการหาค่าที่เหมาะสมที่สุด (Optimization) มาใช้ในการปรับปรุงคุณภาพของการตรวจจับข้อความในภาพ

เนื้อหาในบทที่จะมีดังต่อไปนี้ 3.1 จะทำการอธิบายถึงข้อมูลที่ใช้ในการศึกษาในครั้งนี้ 3.2 จะอธิบายถึงวิธีการที่ใช้ในการสำหรับจับข้อความในภาพ 3.3 การทดลองและผลการทดลองซึ่งจะมีการเปรียบเทียบกับเทคนิคที่นิยมใช้ในการตรวจจับข้อความในภาพ 3.4 อภิปรายและสรุปผล



รูปที่ 120 กระบวนการตรวจจับและการรู้จำอักขระ

3.1 การจัดเตรียมชุดข้อมูลที่ใช้ในการวิจัย

ในงานวิจัยนี้ใช้ข้อมูลภาพจากการถ่ายด้วยกล้องดิจิตอล (Digital Camera) อุปกรณ์พกพา (Smartphone) และจากข้อมูลอินเตอร์เน็ต รวมทั้งรวมทั้งสิ้นจำนวน 1200 ภาพ โดยขนาดของภาพให้มีขนาดที่ต่างกัน ภาพที่เก็บรวบรวมจะประกอบไปด้วย ภาพแบบทั่วไป เป็นภาพป้ายร้าน ภาพป้ายไวนิล ที่จะมีองค์ประกอบอื่น ๆ มาปะปนจำนวน 1150 ภาพ ดังตัวอย่างรูปที่ 121 และกลุ่มภาพแบบเจาะจง ที่มีการเลือกภาพเฉพาะภาพที่มองเห็นตัวอักขระป้ายร้าน ป้ายไวนิล ชัดเจนมีองค์ประกอบอื่น ๆ ปะปนน้อยจำนวน 50 ภาพ ดังตัวอย่างรูปที่ 122

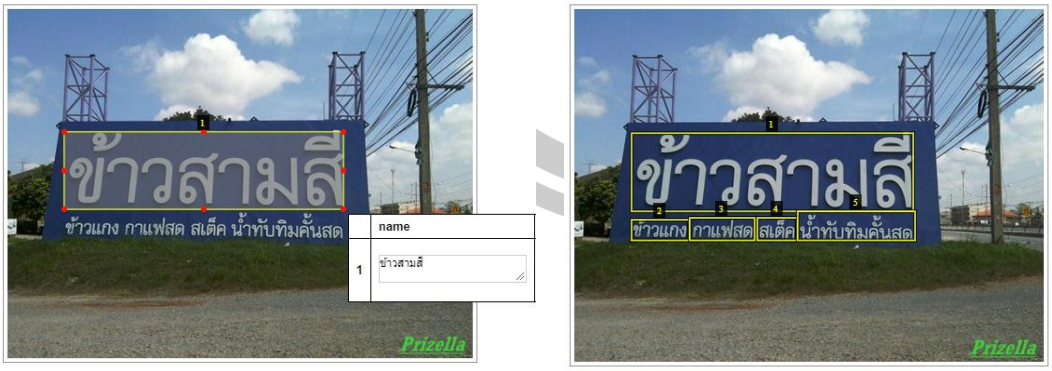
(ก) (ข)

รูปที่ 121 ตัวอย่างข้อมูลภาพกลุ่มภาพแบบทั่วไป

(ก) (ข)

รูปที่ 122 ตัวอย่างข้อมูลภาพกลุ่มภาพแบบเจาะจง

ผลเฉลยของข้อมูล (Ground Truth Images) จะถูกเตรียมโดยการระบุตำแหน่งของข้อความในภาพและข้อความ สำหรับข้อความในแต่ละตำแหน่งในภาพนั้นจะทำการบันทึกตำแหน่งคู่อันดับ (x,y) บนระบบคาร์ทีเซียน (Cartesian Coordinate System) และความกว้าง (Width) และความยาว (Height) ของพื้นที่ข้อความที่แสดงด้วยกรอบสี่เหลี่ยม (Bounding Boxes) การเตรียมภาพผลเฉลยแสดงดังรูปที่ 123



รูปที่ 123 ตัวอย่างการสร้างภาพผลเฉลยสำหรับการตรวจจับและรู้จำข้อความในภาพ

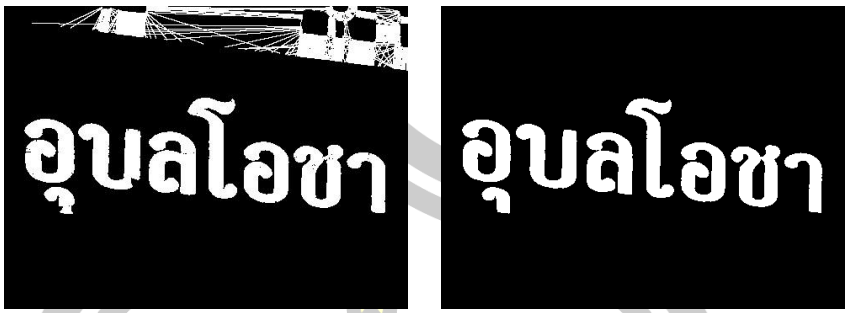
3.2 การตรวจหาพื้นที่ข้อความในภาพ (Text Localization)

ในหัวข้อนี้จะอธิบายวิธีการในตรวจจับหรือระบุตำแหน่งของข้อความในภาพโดยจะแบ่งเนื้อหาในหัวข้อนี้ออกเป็น 2 ส่วน ได้แก่ 1. Adapted Maximally Stable Extremal Regions (AMSER) และ 2. เทคนิคการระบุตำแหน่งข้อความด้วยข้อมูลความรู้ก่อนหน้าแบบหลายระดับ

3.2.1 Adapted Maximally Stable Extremal Regions (AMSER)

Maximally Stable Extremal Regions (MSER) เป็นวิธีการที่มีประสิทธิภาพในการตรวจหาพื้นที่ข้อความในภาพหนึ่งที่มีการรายงานในวรรณกรรม โดยอาศัยหลักการของพื้นที่ที่มีความเป็นสุดขีด (Extrema Regions) ซึ่งอาจจะเป็นส่วนพื้นที่คุณสมบัติรวมแบบมากที่สุดหรือน้อยที่สุดก็ได้ ตัวอย่างของผลลัพธ์ของเทคนิค MSER และเทคนิค Stroke Width Transform (SWT) แสดงดังรูปที่ 124

(ก) (ข)



(ค) (ง)

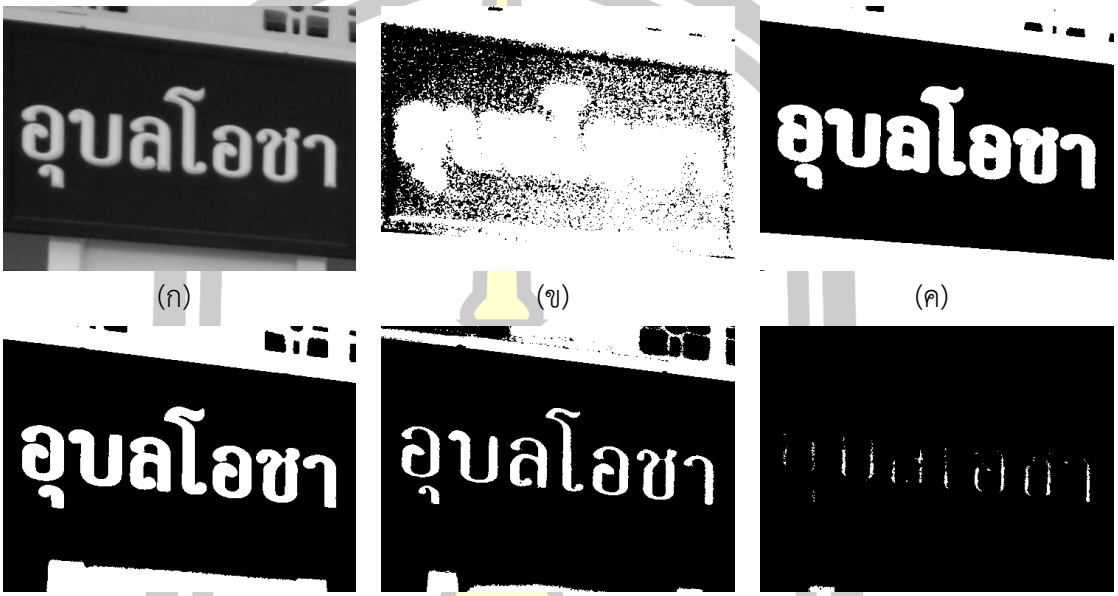
รูปที่ 124 ภาพการเปรียบเทียบประสิทธิภาพการตรวจหาพื้นที่ข้อความในภาพ (ก) ภาพต้นฉบับ (ข) Dhanushka Method (ค) ภาพ SWT และ(ง) ภาพ MSER

จากรูปที่ 124 พื้นที่ที่มีโอกาสเป็นกลุ่มของข้อความจะถูกกำหนดให้มีค่าของพิกเซลเท่ากับ 1 (สีขาว) และพื้นที่ที่ไม่ใช่ข้อความจะถูกกำหนดให้มีค่าพิกเซลเท่ากับ 0 (สีดำ) การเปรียบเทียบประสิทธิภาพการตรวจหาพื้นที่ข้อความในภาพ พบว่าวิธีการ MSER ให้ผลที่ดีที่สุด ขั้นตอนการดำเนินการวิธีการ MSER ประกอบไปด้วย 4 ขั้นตอนสำคัญได้แก่ 1) กำหนดระดับค่าเธรชโชลด์ที่เหมาะสม 2) การคำนวณพื้นที่ที่มีส่วนเชื่อมต่อกัน 3) การคำนวณหาค่า Extremal Region และ4) การคัดเลือกพื้นที่ที่เหมาะสม ขั้นตอนแสดงดังรูปที่ 125

รูปที่ 125 ขั้นตอนการดำเนินการวิธีการ Maximally Stable Extremal Regions (MSER)

1) กำหนดค่าเธรชโชลด์ที่เหมาะสมในการแบ่งภาพ เป็นขั้นตอนที่เริ่มจากการแปลงภาพต้นฉบับเป็นภาพระดับเทา (Grayscale) และทำการไล่ระดับค่าเธรชโชลด์อยู่ในช่วง 0 - 255 ระดับ ซึ่งในการไล่ระดับแต่ละช่วงนั้นมีการแปลงภาพระดับเทาให้เป็นภาพไบนารี (Binary Image) การแปลงภาพ

ไบนารีจะมีการอ้างอิงค่าเธรชโชลด์ของระดับนั้น ๆ โดยการกำหนดค่าพิกเซลที่มีค่ามากกว่าค่าเธรชโชลด์ที่กำหนดให้มีค่าเท่ากับ 1 (สีขาว) ส่วนพิกเซลที่มีค่าน้อยกว่าค่าเธรชโชลด์ที่กำหนดให้มีค่าเท่ากับ 0 (สีดำ) ดังรูปที่ 126



(ก) (ข) (ค)

(ง) (จ) (ฉ)

รูปที่ 126 ภาพแสดงการไล่ระดับค่าเธรชโชลด์

จากรูปที่ 126 คือภาพตัวอย่างการไล่ระดับค่าเธรชโชลด์ เช่น (ข) ค่าเธรชโชลด์เท่ากับ 25 (ค) ค่าเธรชโชลด์เท่ากับ 95 (ง) ค่าเธรชโชลด์เท่ากับ 125 (จ) ค่าเธรชโชลด์เท่ากับ 175 และ(ฉ) ค่าเธรชโชลด์เท่ากับ 200 ตามลำดับ เมื่อได้ภาพไบนารีในแต่ระดับค่าเธรชโชลด์แล้ว นำภาพไบนารีทั้งหมดมาคำนวณหาพื้นที่ที่มีส่วนเชื่อมต่อกันในขั้นตอนต่อไป

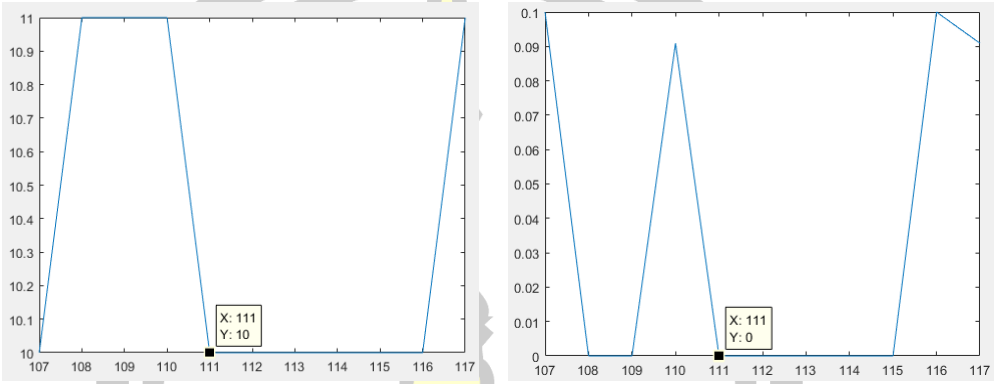
2) การคำนวณพื้นที่ที่มีส่วนเชื่อมต่อกัน เป็นขั้นตอนที่คำนวณพื้นที่ที่มีส่วนเชื่อมต่อกันจากภาพไบนารีแต่ละภาพด้วยวิธีการ Connected Components Analysis ซึ่งวิธีการคำนวณได้กล่าวไปแล้วในบทที่ 2

3) คำนวณหาค่า Extremal Region สามารถคำนวณได้จาก สมการที่ 3.1

$$\upsilon(R_i) = \frac{\left|R_{i+\Delta} - R_i\right|}{\left|R_i\right|} \qquad (3.1)$$

โดยที่ R_i คือจำนวนของ Connected Components ในระดับค่าเธรชโชลด์ที่ i และ $R_{i+\Delta}$ คือจำนวนของ Connected Components ในระดับค่าเธรชโชลด์ถัดจากค่า i ซึ่ง Δ เป็น

พารามิเตอร์ในการกำหนดระยะห่างระหว่างค่าเธรชโชลด์ การพิจารณาพื้นที่ที่มีความคงทนมากที่สุด สามารถพิจารณาได้จากจุดต่ำสุดของค่า Connected Components และค่า Extremal Region ดังรูปที่ 127



(ก) (ข)

รูปที่ 127 กราฟแสดงค่าเธรชโชลด์ที่เหมาะสม

จากรูปที่ 127 (ก) แกน X คือค่าเธรชโชลด์ แกน Y คือจำนวน Connected Components และ (ข) แกน X คือค่าเธรชโชลด์ แกน Y คือค่า Extremal Region เมื่อพิจารณาจุดต่ำสุดของทั้งสองค่าพบว่า ค่าเธรชโชลด์ที่เหมาะสมที่สุดคือค่า 111 เมื่อนำค่าเธรชโชลด์นี้ มาใช้ในการแปลงภาพไบนารี ผลลัพธ์แสดงดังรูปที่ 128

รูปที่ 128 ภาพการแปลงภาพระดับเทาเป็นภาพไบนารีด้วยค่าเธรชโชลด์ที่เหมาะสม

4) การคัดเลือกพื้นที่ที่เหมาะสม คือขั้นตอนที่เลือกเฉพาะพื้นที่ที่สนใจเท่านั้น พื้นที่ที่งานวิจัยนี้สนใจคือพื้นที่ที่คาดว่าเป็นข้อความ เมื่อพิจารณารูปที่ 128 จะพบว่าการแปลงภาพไบนารี

ด้วยค่าเธรชโชลด์ที่ได้จากขั้นตอนก่อนหน้า ยังคงมีพื้นที่หรือบริเวณที่ไม่ใช่ข้อความปรากฏอยู่ ดังนั้น การตัดหรือลบส่วนที่ไม่ใช่ข้อความออกจากภาพไบนารี สามารถกระทำได้ด้วยการกำหนดขนาดของพื้นที่ที่ไม่ต้องการ ซึ่งมีด้วยกันอยู่ 2 ขนาดได้แก่ 1) พื้นที่ที่มีขนาดใหญ่เกินไป และ2) พื้นที่ที่มีขนาดเล็กเกินไป เมื่อกำหนดขนาดของพื้นที่ที่ไม่ต้องการได้แล้ว ผลลัพธ์แสดงดังรูปที่ 129

รูปที่ 129 ภาพแสดงการตัดพื้นที่ด้วยการกำหนดขนาดของพื้นที่ที่ไม่ต้องการ

จากรูปที่ 129 คือภาพที่ได้จากการตัดหรือลบพื้นที่ด้วยการกำหนดขนาดของพื้นที่ที่ไม่ต้องการ หากพิจารณาจะพบว่าการดำเนินการด้วยวิธีการ Maximally Stable Extremal Regions สามารถเลือกพื้นที่ที่คาดว่าเป็นข้อความได้อย่างมีประสิทธิภาพ แต่อย่างไรก็ตามวิธีการนี้ก็ยังประสบปัญหาหลายประการได้แก่ 1) การกำหนดค่าพารามิเตอร์ที่เหมาะสม ซึ่งวิธีการนี้ไม่สามารถใช้ค่าพารามิเตอร์มาตรฐานกับทุกภาพได้ 2) การกำหนดขนาดของพื้นที่ที่ไม่ต้องการ ซึ่งวิธีการนี้ไม่สามารถกำหนดขนาดของพื้นที่ที่ไม่ต้องการแบบตายตัวได้ เนื่องจากหากมีการกำหนดขนาดพื้นที่ที่ไม่ต้องการให้มีขนาดใหญ่หรือขนาดเล็กจนเกินไป อาจจะส่งผลให้พื้นที่ที่สนใจถูกลบออกไปด้วย หรือพื้นที่ที่ไม่ต้องการอาจถูกลบออกไปไม่หมด และ3) ภาพที่มีพื้นหลังที่ความซับซ้อนหรือภาพที่มีสีของตัวอักษรที่ความมืดมากจนเกินไป (พื้นหลังมีความสว่างมากกว่าตัวอักษร) ซึ่งวิธีการนี้เหมาะกับภาพที่มีตัวอักษรที่มีความสว่าง (พื้นหลังมีความสว่างน้อยกว่าตัวอักษร) และพื้นหลังที่ไม่ซับซ้อนมากนัก ดังรูปที่ 130

(ก) (ข) (ค)

รูปที่ 130 ภาพผลลัพธ์ของวิธีการ Maximally Stable Extremal Regions ที่นำมาใช้กับภาพที่มีพื้นหลังซับซ้อน (ก) ภาพต้นฉบับ (ข) ภาพ MSER ที่ใช้ค่าพารามิเตอร์มาตรฐาน และ(ค) ภาพ MSER ที่มีการกำหนดค่าพารามิเตอร์ด้วยมือ

อย่างที่ได้กล่าวมาข้างต้น MSER นั้นเป็นเทคนิคที่อาศัยผลลัพธ์ของการกำหนดกลุ่มของพิกเซลที่จะเป็นข้อความในภาพ (Groups of Candidates) จะเห็นได้ว่าจากรูปที่ 130 MSER ค่อนข้างมีความไม่แน่นอนเมื่อมีการปรับเปลี่ยนค่าพารามิเตอร์ของ MSER (Parameter Sensitive) ดังนั้นเพื่อเป็นปรับปรุงการทำงานของ MSER ในปริญญานิพนธ์นี้จึงประยุกต์ใช้เทคนิคการค้นหาที่ดีที่สุด มาใช้เพื่อเป็นการปรับปรุงผลลัพธ์ของการทำงานของ MSER โดยเรียกว่า Adaptive MSER (AMSER) โดยขั้นตอนการทำงานจะการหาค่าเหมาะสม (Optimization) เพื่อหาค่าเทรชโฮลด์ที่ต่ำสุดและมากสุดที่เหมาะสม กำหนดให้ T_{min} เป็นค่าเทรชโฮลด์ต่ำสุด และ T_{max} เป็นค่าเทรชโฮลด์ที่มากสุดที่เป็นไปได้ โดย $T_{min}, T_{max} \in [0,1]$ เราจะมี f เป็นฟังก์ชันที่ประเมินค่าผลลัพธ์ โดย $f \in [0,\infty]$ โดยจะทำการประเมินผลลัพธ์ของฟังก์ชันหรือฟังก์ชันจุดประสงค์ (Objective Function) $f: I^D \rightarrow P$ ซึ่งเป็นฟังก์ชันในแปลงจากภาพ I ไปเป็นให้มีค่าน้อยที่สุด โดยการปรับค่า T_{min} และ T_{max}

วิธีการ Adaptive MSER
อินพุต : ภาพ เอาท์พุต : ภาพที่มีการ 1. กำหนดค่า T 2. ประมวลผลเทคนิค MSER ด้วยค่า T 2.1 ทำการประเมินค่า $r_t = \mathcal{V}(R_T)$ เพื่อหาค่า Extreme Region (สมการ 3.1) 2.1.1 ทำการประเมินค่า Smoothness ของ R_T โดยพิจารณาจากค่าความแปรปรวนของพื้นที่ 2.1.2 ทำการประเมินค่า $p(\text{text})$ 2.1.3 คำนวณค่าความเป็นข้อความ r_t 3. ทำการเพิ่มค่า T เพื่อทำการสร้าง ΔR 4. ทำการเลือกชุด R ที่ได้ค่า r มากที่สุด

$p(\text{text})$ จะได้จากค่าความน่าจะของพื้นที่ที่จะเป็นข้อความซึ่งพิจารณาจากการประมาณค่าด้วยกฎของ Bayes จาก

$$p(\text{text}|R) \propto p(R|\text{text})p(\text{text}) \tag{3.2}$$

โดย $p(R|\text{text})$ คือ ค่าความน่าจะเป็นที่ได้จากฟังก์ชันความน่าจะเป็นร่วม (Likelihood Function) และ $p(\text{text})$ คือความน่าจะเป็นก่อนหน้า (Prior Probability) ที่ได้จากการเรียนรู้จากจากข้อมูลจะพบว่าสัดส่วนของพิกเซลที่เป็นข้อความนั้นจะน้อยกว่าส่วนของพิกเซลที่ไม่ใช่ข้อความจากชุดข้อมูลสำหรับเรียนรู้ค่าความน่าจะเป็นก่อนหน้าของ สำหรับ$(R|\text{text})$ จะพิจารณาจากตำแหน่งของข้อความในภาพ โดยได้จากชุดข้อมูลฝึกฝน

เพื่อปรับปรุงผลลัพธ์หลังจากมีการระบุตำแหน่งของข้อความในภาพ จึงมีการพยายามตัดส่วนของพื้นที่ไม่ใช่ข้อความออก โดยขั้นตอนนี้จะได้คำนวณจากข้อสังเกตคือ ส่วนพื้นที่ที่เป็นข้อความจะอยู่ใกล้กันกับส่วนของข้อความ ดังนั้นการปรับปรุงผลลัพธ์การระบุตำแหน่งข้อความจะอาศัยค่าความเป็นข้อความของพื้นที่ (ที่ได้จาก AMSER) และค่าความสัมพันธ์เป็นเชิงพื้นที่ (Spatial Probability) ของพื้นที่ที่ใกล้เคียงกัน เพื่อนำมาคำนวณค่าความสัมพันธ์ระหว่างพื้นที่ที่ได้ โดยจะคำนวณผ่านฟังก์ชันพลังงาน E

ที่ประกอบไปด้วยค่าพลังย่อยที่แสดงถึงความสัมพันธ์ระหว่างพื้นที่ (E_{Bin}) และค่าพลังงานย่อยของค่าความเป็นข้อความของพื้นที่ E_{sing} ดังนี้

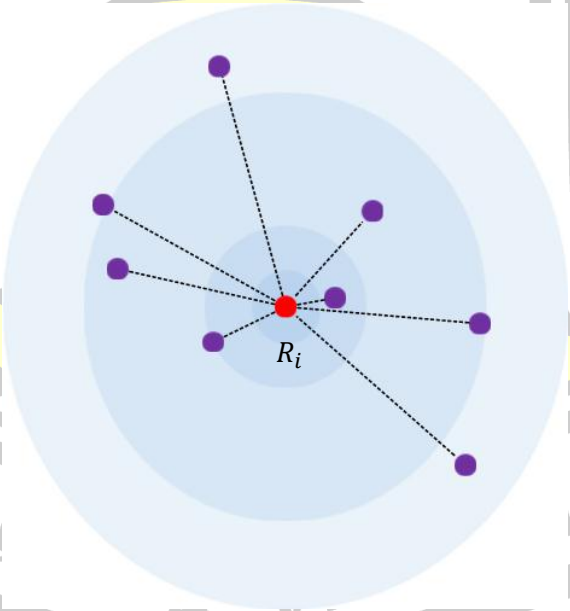
$$E = E_{Bin} + E_{sing} \tag{3.3}$$

$$E_{bin} = \sum_{\{j \in \omega_i\}} -\log(N(R_i, R_j)) \tag{3.4}$$

โดย $N(R_i, R_j)$ เป็นฟังก์ชันที่ประเมินค่าน้ำหนักระหว่าง R_i และ R_j ใด ๆ ในเซตของพื้นที่ข้างเคียง ω ของ R_i โดยค่าน้ำหนักจะคำนวณโดยใช้ฟังก์ชันแบบ Radial โดยใช้ฟังก์เกาเซียน (Gaussian function) ดังนี้

$$N(R_i, R_j|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\frac{(R_i - R_j)^2}{2\sigma^2})} \tag{3.5}$$

การแทนค่าน้ำหนักแสดงดังตัวอย่างรูปที่ 121



รูปที่ 131 แสดงตัวอย่างของความสัมพันธ์ของพื้นที่ R_i และ R_j

จากรูปที่ 131 จะมีจำนวน R_j ในเซต ω_i จำนวน 8 พื้นที่ และค่าน้ำหนักระหว่าง R_i กับทุก ๆ พื้นที่ ใน ω_i จะคำนวณจากสมการ 3.5 โดยพื้นที่อยู่ห่างจาก R_i จะมีค่าน้ำหนักที่น้อยลงซึ่งจะสอดคล้องกับข้อสังเกตที่กล่าวมาข้างต้น

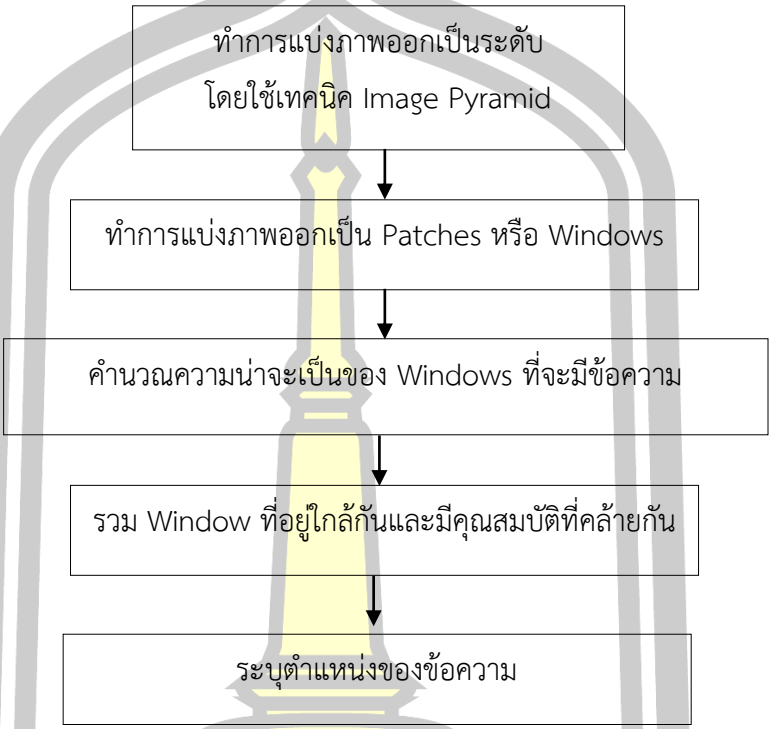
สำหรับ E_{sing} จะพิจารณาจากค่าความเป็นข้อความจากค่า r_i ของพื้นที่ R_i และเนื่องจากในงานนี้จะค่าพลังงานที่น้อยที่สุด ค่า E_{sing} จึงสามารถคำนวณได้ดังนี้

$$E_{sing} = -\log(r) \tag{3.6}$$

ในการหาค่าพลังงานที่ดีที่สุดจะใช้วิธีการ Graph-Cut ในการประมวลผล

3.2.1 การระบุตำแหน่งข้อความด้วยข้อมูลความรู้ก่อนหน้า

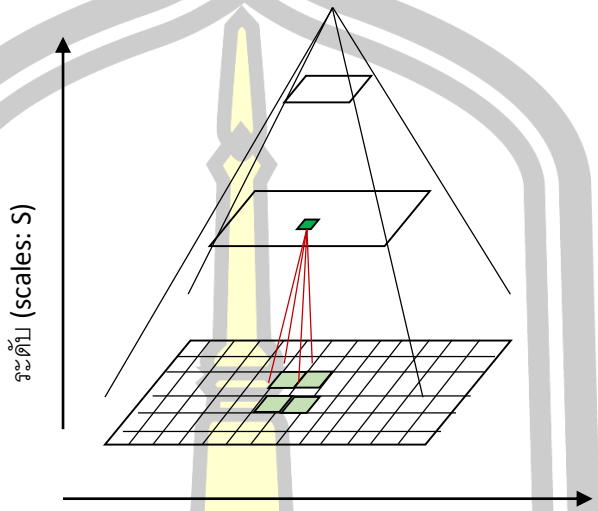
นอกจากวิธีการตรวจจับและระบุตำแหน่งข้อความในภาพด้วยวิธีการอาศัยการเชื่อมต่อของพิกเซลเป็นหลัก (Connected Component) ในปริญญานิพนธ์นี้ยังได้นำเสนอเทคนิคการระบุตำแหน่งข้อความในภาพโดยใช้ข้อมูลก่อนหน้า (Prior Information) ในการระบุตำแหน่งของข้อความในภาพ โดยจะทำการแบ่งภาพออกเป็นส่วน (Patches หรือ Windows) โดยเป็นแบบที่ไม่ซ้อนทับกัน จากนั้นจะทำการคำนวณหาค่าความน่าจะเป็นที่ Window ที่มีส่วนของข้อความอยู่ โดยการแบ่ง Window นี้จะพิจารณาโดยข้อความที่อยู่ในภาพนั้นจะต้องอยู่ใน Window ใด Window หนึ่ง ดังนั้นเพื่อให้การคำนวณค่าความรู้ก่อนหน้าที่จะทำมาใช้ในการระบุตำแหน่งของข้อความได้อย่างถูกต้องในงานนี้จึงทำการกำหนดขนาดของ Window โดยการแบ่งภาพออกเป็น 9 ส่วน (3x3) นอกจากนั้นจะมีการนำข้อมูลก่อนหน้า นำมาช่วยพิจารณา ซึ่งจะประมวลผลในรูปแบบหลายระดับ (Multi-Scaling) Window ที่มีความน่าจะเป็นที่จะมีข้อความอยู่จะนำมาเพื่อพิจารณาในการรวม Window ที่ใกล้เคียงกันและมีความน่าจะเป็นที่มีข้อความเพื่อสร้างเป็นพื้นที่ที่เป็นข้อความในภาพ ภาพรวมของการประมวลผลแสดงดังรูปที่ 132



รูปที่ 132 แสดงภาพรวมของเทคนิคในการตรวจจับข้อความในภาพโดยใช้ข้อมูลก่อนหน้า (Prior Information)

ในหัวข้อนี้จะอธิบายวิธีการที่ใช้ในการตรวจจับข้อความในภาพโดยใช้ข้อมูลก่อนหน้า ซึ่งมีขั้นตอนดังต่อไปนี้

1. การแบ่งภาพออกเป็นหลายระดับ ภาพนำเข้า I จะถูกนำมาแบ่งให้เป็นระดับ (Scale: S_i โดย i คือ ระดับของภาพ) ในงานนี้จะใช้การแบ่งระดับภาพด้วยเทคนิค Image Pyramid ซึ่งจะใช้การแบ่งด้วยค่าน้ำหนักเกาเซียน (Gaussian) การแบ่งภาพด้วยเทคนิค Image Pyramid แสดงดังรูปที่ 133



รูปที่ 133 แสดงภาพ I ที่มีการแบ่งด้วย Image Pyramid

สำหรับแต่ละระดับของภาพ S_i จะมีการคำนวณ

$$G_{i+1}(x,y) = REDUCE(G_i(x,y)) \tag{3.7}$$

โดยที่ $REDUCE(.)$ เป็น low-pass filter โดยใช้ Gaussian Filter

$$REDUCE\big(G_i(x,y)\big) = I \otimes K \tag{3.8}$$

เมื่อ K เป็น Mask แบบเกาเซียน โดยภาพ I จะถูกนำมา Convolute ด้วย K ซึ่งมีพารามิเตอร์ 1 ตัว ได้แก่ ค่าความแปรปรวน (σ) โดยในงานนี้กำหนดให้เป็น 0.2

2. ทำการแบ่งภาพในแต่ละระดับออกเป็น Window โดยขนาดของ Window นั้นจะถูกกำหนดด้วยขนาด (m คือจำนวนแถว และ n คือจำนวนหลัก) ในแต่ละ window

3. ทำการสร้างแผนที่ความน่าจะเป็นเพื่อเป็นตัวกำหนดตำแหน่งของข้อความในภาพ โดยการสร้างแผนที่ความน่าจะเป็นนี้จะถูกสร้างขึ้นจากข้อมูลก่อนหน้า (Prior Information) ในงานนี้

มีการกำหนดข้อมูลก่อนหน้า 2 แบบด้วยกัน คือ 1) การใช้จุดโฟกัสกลางในภาพ (Focal Points) และ 2) การสร้างจากกลุ่มของชุดฝึกฝน

การสร้างแผนที่ความน่าจะเป็นโดยการใช้จุดโฟกัสกลางในภาพนั้น จะอยู่บนสมมุติฐานว่าข้อความป้ายที่ปรากฏอยู่ในภาพนั้นโดย “ทั่วไป” (Hard Assumption) แล้วจะไม่อยู่ส่วนที่เป็นขอบหรือมุมของภาพ โดยใช้ Distribution กำหนดให้ $\theta = \{\theta_1, \dots, \theta_n\}$ เป็นตัวแปรที่สนใจขนาด n ตัวแปร จะมีพารามิเตอร์ $\alpha = \{\alpha_1, \dots, \alpha_n\}$ ที่ทำให้สามารถคำนวณการแจกแจงความน่าจะเป็นของตัวแปรได้ดังต่อไปนี้

$$p(\theta) = \frac{1}{\beta(\alpha)} \prod_{i=1}^{n} \theta_i^{\alpha_i - 1} I(\theta \in S) \tag{3.9}$$

เมื่อ I เป็น Identify function และ S เป็น Probability Simplex โดย

$$S = \{x \in R^n : x \geq 0, \sum x_i = 1\}) \tag{3.10}$$

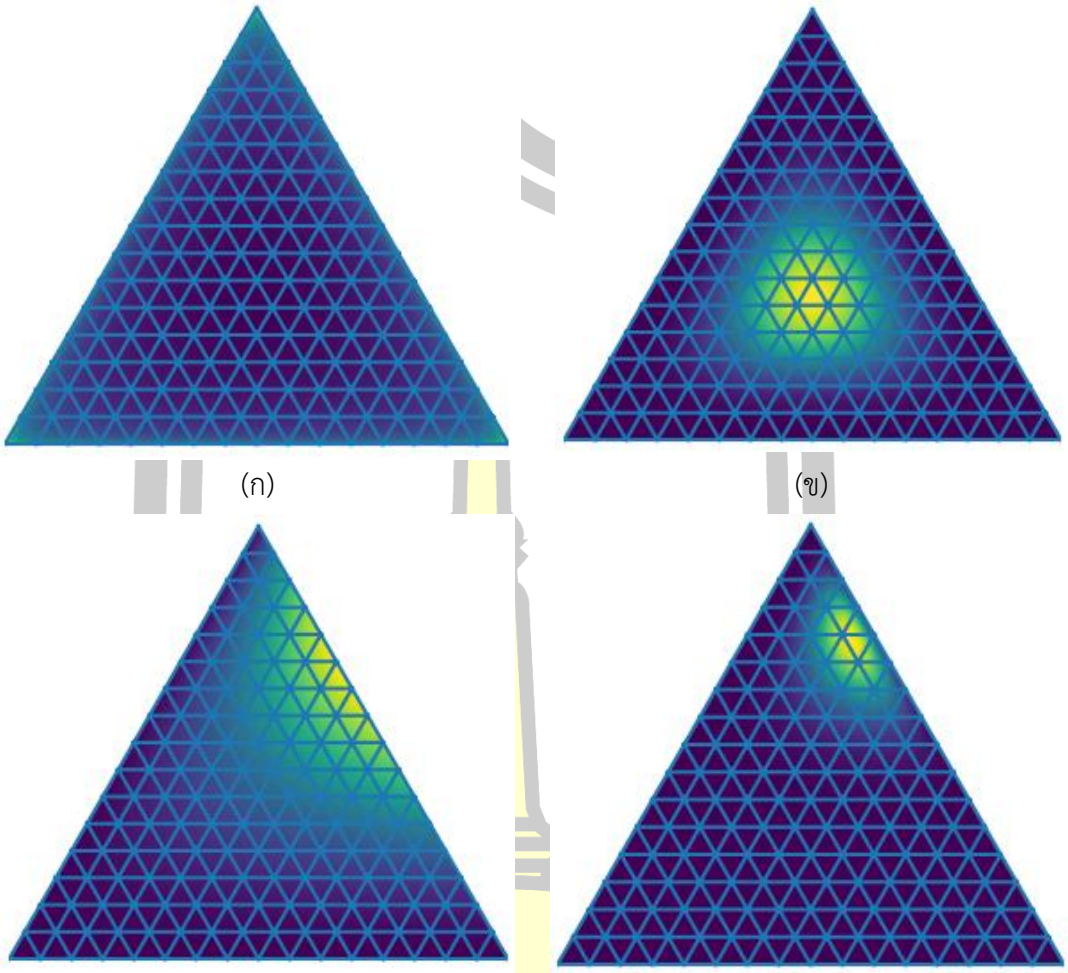
และ $\frac{1}{\beta(\alpha)}$ เป็นฟังก์ชัน nominator

$$\frac{1}{\beta(\alpha)} = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n} \alpha_i} \tag{3.11}$$

โดย

$$\alpha_0 = \sum_{i=1}^{n} \alpha_i \tag{3.12}$$

ตัวอย่างของการแจกแจงแบบ Dirichlet แสดงดังรูปที่ 134



(ก) (ข)

(ค) (ง)

รูปที่ 134 แสดงการแจกแจงแบบ Dirichlet

4. ทำการระบุตำแหน่งของความโดยการการฝึกฝนข้อมูลของแต่ละ Window ในการใช้เอกลักษณ์ของแต่ละ window โดยใช้เทคนิค Convolutional Neural Network (CNN) โดยมีขั้นตอนดังนี้

การฝึกฝนเรียนรู้ window ที่เป็นข้อความ
อินพุต : ภาพในแต่ละ window และผลเฉลย เอาท์พุต : ตัวแบบที่ระบุความเป็นข้อความของ window 1. ทำการสร้าง CNN 1.1 จำนวน Convolution Layer = 5 1.2 Max Pooling แบบ 2x2 stride 1 1.3 Activation function : Relu 2. ทำการสร้าง Fully Connected Network 3. Loss : Logistic Loss 4. ใช้ SGD ในการประเมินค่าที่ดีที่สุด

. ขั้นตอนนี้จะทำการทำนาย Window (w) ในภาพ I การทำนายจะใช้ตัวแบบที่ทำการเรียนรู้จากเทคนิค CNN ซึ่งทำการประเมินค่าฟังก์ชันเพื่อได้ความน่าจะเป็น (Posterior Probability) ของ window ที่จะเป็นข้อความและไม่ใช่ข้อความ $p(w|\theta)$ เมื่อ θ คือ ตัวแบบที่ได้จากการฝึกฝนด้วย CNN

6. ทำการสร้างรวม Window ซึ่งการรวม Window นี้จะใช้หลักการที่คล้ายกับเทคนิคการปรับปรุงผลลัพธ์ที่อธิบายในหัวข้อที่ผ่านมาโดยใช้ฟังก์ชันพลังงานและนำเอาข้อมูลก่อนหน้า (Prior Information) เข้ามาใช้ ดังนี้

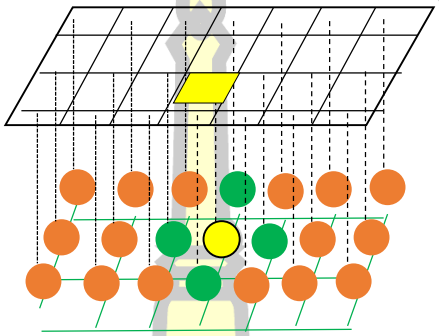
$$E_{sing} = -\log(p(w|\theta) * p(w_R)) \tag{3.13}$$

เนื่องจากต้องการค่าที่เหมาะสมน้อยที่สุด (Minimization) จึงนำค่าความน่าจะเป็นที่ได้ดำเนินการด้วย $-\log(.)$ โดย $p(w|\theta)$ เป็นความน่าจะเป็นของ Window ที่จะเป็นข้อความและไม่ใช่ข้อความ และ $p(w_R)$ เป็นค่าความน่าจะเป็นก่อนหน้าที่ได้จาก Dirichlet

พลังงานย่อยสำหรับประเมินค่าความสัมพันธ์ระหว่าง Window คำนวณได้ดังนี้

$$E_{bin} = \sum_{\{j \in \omega_i\}} V(w_i, w_j) \tag{3.14}$$

ω_i เป็นเซตของ Window ข้างเคียงของ w_i โดยใช้การพิจารณา Window ข้างเคียง 4 Window ในรูปแบบ Image Grid ดังแสดงในรูปที่ 125 ฟังก์ชัน $V(.,.)$ ทำประเมินค่าความสัมพันธ์ระหว่าง window ในกรณีนี้จะไม่นำระยะทางระหว่าง window มาพิจารณา เนื่องจากภายใน Image Grid นั้น ระยะระหว่าง Window ข้างเคียงนั้นจะมีระยะทางที่เท่ากัน ดังนั้น $V(.,.)$ จึงถูกกำหนดให้เป็นค่าคงที่ m

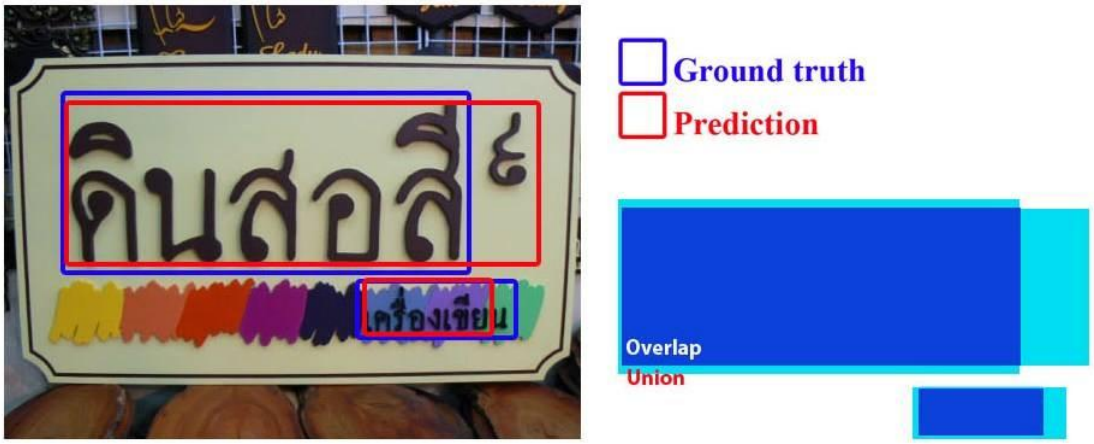


รูปที่ 135 แสดงตัวอย่างของ Window (Image Grid) ในภาพและการพิจารณา Window ข้างเคียง

เมื่อทำการกำหนดค่าพลังงานแล้ว จะทำการประเมินค่าเหมาะสมแบบต่ำสุดด้วยเทคนิค Graph-Cut เพื่อทำการปรับปรุงผลลัพธ์

3.3 การทดลองและผลการทดลอง

การวัดประสิทธิภาพจะใช้เทคนิคของพื้นที่ที่ซ้อนทับกัน (Intersection over Union: IoU) ซึ่งเป็นเทคนิคที่นิยมใช้สำหรับการวัดประสิทธิภาพของระบบการตรวจจับวัตถุในภาพ ภาพนำเข้า (I) จะถูกประมวลผลโดยการตรวจจับวัตถุในภาพ ซึ่งจะประกอบไปด้วยวัตถุที่ถูกต้อง (Positive) และวัตถุที่ตรวจจับที่ผิดพลาด (Negative) การประเมินประสิทธิภาพของ Positive Detections จะนำผลของการตรวจจับนำไปเปรียบเทียบผลเฉลย (Ground Truth) ผลเฉลยได้ทำการเตรียมโดยการระบุตำแหน่งของวัตถุด้วย Bounding Boxes (ที่ประกอบไปด้วยตำแหน่ง x y ความสูงความกว้างของ Bounding Boxes) โดยการระบุผลการตรวจจับที่ถูกต้องนั้นจะใช้อัตราการซ้อนทับของผลเฉลยและการตรวจจับวัตถุในภาพที่ได้จากเทคนิคต่าง ดังแสดงในรูปที่ 126



รูปที่ 136 แสดงการเปรียบเทียบระหว่างวัตถุ (ข้อความ) ที่ตรวจจับได้ในภาพกับภาพผลเฉลย โดยพิจารณาพื้นที่ที่ซ้อนทับกัน (IoU)

โดยค่า IoU จะคำนวณได้จากสัดส่วนของพื้นที่ซ้อนทับของวัตถุและผลเฉย กับพื้นที่ที่ได้จากการ Union ระหว่างวัตถุและผลเฉลย ดังสมการที่ 3.15

$$IoU_p = \frac{\text{พื้นที่ซ้อนทับของวัตถุและผลเฉลย}}{\text{พื้นที่ที่ได้จากการ Union ระหว่างวัตถุและผลเฉลย}} \quad (3.15)$$

เมื่อ p คืออัตราของพื้นซ้อนทับของวัตถุและผลเฉลย จากนั้นจะทำค่า IoU มาใช้ในการคำนวณอัตราการตรวจจับข้อความในภาพ (Detection rate : R_p) โดยคำนวณได้จาก

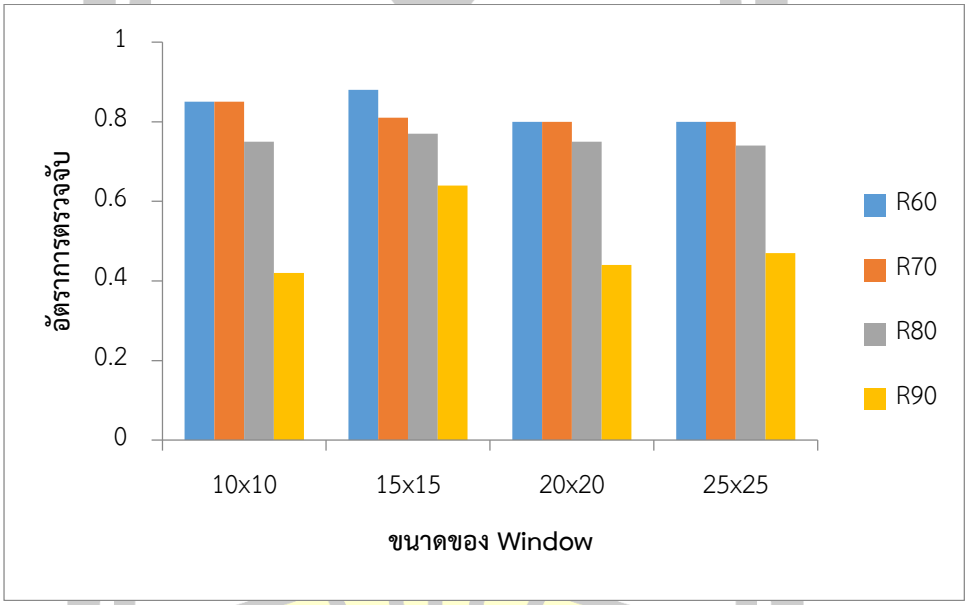
$$R_p = \frac{\sum_n ioU \text{ x}100}{n} \quad (3.16)$$

โดย Obj_{all} เป็นจำนวนของวัตถุ (ข้อความ) ทั้งหมดในภาพ ซึ่งอัตราการตรวจจับข้อความในภาพจะบ่งบอกประสิทธิภาพของเทคนิค

ในปริญญานิพนธ์นี้ได้ทำการนำเสนอเทคนิคที่ใช้ในการตรวจจับ ด้วยข้อมูลก่อนหน้าและทำการเปรียบเทียบกับเทคนิคที่มีการวิจัยก่อนหน้านี้ เทคนิคที่ทำเสนอในปริญญานิพนธ์นี้และเทคนิคที่นำมาเปรียบเทียบจะอยู่ในรูปแบบของการทำงาน 2 ขั้นตอน (Two-Stage) ที่จะทำการหาส่วนจะเป็น

ข้อความในภาพ (Candidates หรือ Region Proposals) จากนั้นส่วนที่จะเป็นข้อความในภาพจะถูกนำไปจำแนกด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อระบุภาพที่เป็นข้อความแล้วไม่ใช่ข้อความ

จากหัวข้อที่ 3.3 ได้ทำการนำเสนอเทคนิคที่จะใช้ในการระบุตำแหน่งของข้อความในภาพนั้น ในหัวข้อนี้จะอธิบายถึงการทดลองเพื่อประเมินประสิทธิภาพของเทคนิคที่ได้นำเสนอในปริญญานิพนธ์นี้ ข้อมูลถูกเก็บรวบจำนวน 1200 โดยจะมีการแบ่งข้อมูลเพื่อใช้ในการเตรียมข้อมูลก่อนหน้าจำนวน 200 ภาพ การทดลองได้ดำเนินการกับวิธีการตรวจจับข้อความโดยใช้ข้อมูลก่อนหน้า จะมีการทดลองการใช้ widow ที่มีขนาดที่ต่างกัน ดังนั้นจึงมีการทดลองเพื่อพิจารณาขนาดของ Window ที่ดีที่สุดกับชุดข้อมูลที่ทำการรวบรวมมา โดยทำการพิจารณาค่าอัตราการตรวจจับ (Detection Rate) จากพื้นที่ที่ซ้อนทับกันกับระหว่างผลลัพธ์และผลเฉลย 60% 70% 80% และ 90% และผลลัพธ์แสดงดังรูปที่ 127



รูปที่ 137 แสดงประสิทธิภาพของการตรวจจับข้อความในภาพโดยใช้ขนาดของ Window ที่ต่างกัน

ตาราง 2 แสดงอัตราการตรวจจับด้วยเทคนิคการใช้ความน่าจะเป็นก่อนหน้า (R_{60})

Algorithm	$10x10$	$15x15$	$20x20$	$25x25$
Prior Information	0.85	**0.88**	0.80	0.80

จากการผลลัพธ์ของการทดลองข้างต้นจะพบว่าขนาดของ Window จะมีผลกับอัตราการตรวจจับข้อความโดยใช้เทคนิคข้อมูลก่อนหน้าและพบว่าขนาด Window 15x15 ให้ผลลัพธ์ที่ดีที่สุด

(พิจารณาจากทุก ๆ R_p) นอกจากนั้นยังพบว่าเมื่อมีการเพิ่มขนาดของ Window จะทำให้แนวโน้มของอัตราตรวจจับลดลง เนื่องจากขนาดของ Window ที่ใหญ่ขึ้นอาจจะมีส่วนของพื้นที่ที่เป็นทั้งข้อความและไม่ใช่ข้อความใน Window ซึ่งอาจจะส่งผลให้การตรวจจำแนก Window นั้นผิดพลาด

นอกจากนั้นในงานนี้ได้ทำการทดลองดำโดยใช้วิธีการที่นำเสนอในวิทยานิพนธ์นี้และทำการเปรียบเทียบกับเทคนิคที่ใช้ในการตรวจจับข้อความ 7 วิธี และผลลัพธ์แสดงดังตารางที่ 3

ตาราง 3 แสดงการเปรียบเทียบอัตราการตรวจจับของแต่ละวิธีการ

Algorithm	R_{60}	R_{70}	R_{80}	R_{90}
Adaptive-th	0.62	0.51	0.45	0.45
Otsu	0.66	0.57	0.50	0.33
dhanushka	0.85	0.82	0.67	0.67
Ray Smith	0.84	0.77	0.64	0.55
Huizhong Chen	0.82	0.80	0.70	0.70
SWT	0.76	0.72	0.65	0.60
MSER	0.80	0.80	0.74	0.62
AdaptiveMSER	**0.86**	**0.80**	**0.72**	**0.60**
Prior Information	**0.88**	**0.81**	**0.77**	**0.64**

การเปรียบเทียบประสิทธิภาพของการตรวจหาพื้นที่ข้อความในภาพจะพบว่า การวัดความแม่นยำของวิธีการ (Precision) วิธีการ AMSER มีอัตราการตรวจจับสูงที่สุด 0.86 (R_{60}) และมีความต่างกันแบบไม่มีนัยสำคัญ (p=0.048) การปรับปรุงผลที่ดำเนินการหลังจากขั้นตอน AMSER และขั้นตอนการจำแนก Window (ในวิธีการใช้ข้อมูลก่อนหน้า) ทำให้ส่วนของพื้นที่ที่เป็น False Positive นั้นลดลง และส่วนที่เป็นข้อความในพื้นที่ครอบคลุมข้อความในภาพได้ดีขึ้น จึงส่งผลให้อัตราการตรวจจับดีขึ้น

(ก) (ข)

(ค) (ง)

รูปที่ 138 ผลการทดลองของกระบวนการ Maximally Stable Extremal Regions (MSER)

(ก) (ข)

(ค) (ง)

รูปที่ 139 ผลการทดลองของกระบวนการ AMSER

(ก) (ข)

(ค) (ง)

รูปที่ 140 ผลการทดสอบวิธีการ Adaptive Thresholding

3.3 สรุปผล

การตรวจจับรู้จำป้ายข้อความในภาพนั้นประกอบไปด้วย 2 ขั้นตอนหลักได้แก่ การระบุตำแหน่งของข้อความในภาพ และการรู้จำข้อความที่ตรวจจับได้ โดยการรู้จำข้อความจะดีได้ต้องอาศัยกระบวนการระบุตำแหน่งข้อความที่ดีเช่นกัน ซึ่งการตรวจจับข้อความและการรู้จำข้อความนั้นได้มีกรนำเทคนิคทางด้านการประมวลผลภาพและการเรียนรู้ของเครื่องมาใช้ในการประมวลผล ซึ่งกระบวนการที่ได้ออกแบบสำหรับการตรวจจับและรู้จำข้อความในภาพจะประกอบไปด้วย 1. การตรวจหาพื้นที่ข้อความในภาพ (Text Localization) 2. การแบ่งส่วนข้อความ (Text Segmentation) 3. การวิเคราะห์การจัดเรียงตัวอักษร (Layout Analysis) 4. การสกัดคุณลักษณะ (Feature Extraction) และ 5. การรู้จำตัวอักขระ (Character Recognition) โดยการเปรียบเทียบประสิทธิภาพของการตรวจหาพื้นที่ข้อความในภาพจะพบว่า วิธีการ AMSER มีอัตราการตรวจจับที่ดีที่สุดหากเปรียบเทียบกับวิธีการอื่น ๆ

บทที่ 4
การรู้จำข้อความในภาพ

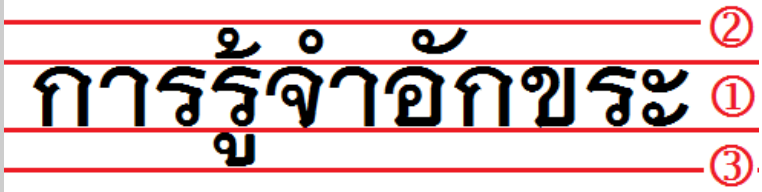
ปริญญานิพนธ์นี้มีจุดประสงค์เพื่อทำการตรวจจับข้อความและรู้จำข้อความในภาพทั่วไป ความท้าทายของตรวจจับและรู้จำข้อความในภาพนั้นคือ ความหลายของข้อมูล ดังที่ได้อธิบายในบทที่ 3 ซึ่งได้ทำเสนอขั้นตอนในการตรวจจับข้อความในภาพ ซึ่งได้นำเสนอวิธีการในการตรวจจับข้อความในภาพ 2 วิธีด้วยกัน ได้แก่ 1) ใช้เทคนิค AMSER และ 2) การใช้ข้อมูลก่อนหน้า (Prior Information) ในการระบุข้อความในภาพ ในบทนี้จะอธิบายและนำเสนอเทคนิคที่ใช้ในการรู้จำอักขระเพื่อระบุข้อความในภาพ ภาพที่ได้จากการตรวจจับข้อความ (แสดงดังรูปที่ 141) จะถูกประมวลผลเพื่อทำการแยกตัวอักขระในภาพ จากนั้นอักขระที่ทำการแยกได้จะไปรู้จำเพื่อนำมาแสดงเป็นข้อความในภาพ

รูปที่ 141 ตัวอย่างผลลัพธ์ที่ได้จากการตรวจจับข้อความในภาพ

หัวข้อในบทนี้มีดังต่อไปนี้ 4.1 การวิเคราะห์รูปแบบตัวอักษรและการจำแนกตัวอักษรจากภาพ 4.2 การสกัดเอกลักษณ์จากอักขระ 4.3 การรู้จำอักขระด้วยเครื่องจักรเรียนรู้ 4.4 การสร้างข้อความ 4.5 การทดลองและการผลการทดลอง และ 4.6 สรุปผลการทดลอง

4.1 การวิเคราะห์รูปแบบตัวอักษรและการแยกตัวอักษร (Layout Analysis and Character Segmentation)

หลังจากการแบ่งส่วนข้อความออกเป็นส่วน ๆ แล้วขั้นตอนต่อมาคือการวิเคราะห์ว่าส่วนใดเป็นส่วนของพยัญชนะ สระ หรือวรรณยุกต์ ซึ่งการวิเคราะห์นี้จะช่วยให้ขั้นตอนการรู้จำอักขระง่ายขึ้น จากการศึกษางานวิจัยที่ผ่านมาพบว่า วิทยา จิรัฐิติเจริญ [103] และรุจิพันธุ์ โกษารัตน์ [104] ได้แนะนำการแบ่งส่วนของพื้นที่ซึ่งสามารถทำการแบ่งพื้นที่ออกเป็น 3 พื้นที่ (Zone) ดังรูปที่ 142



รูปที่ 142 การแบ่งพื้นที่ในกรอบข้อความ

จากรูปที่ 132 เป็นการแบ่งพื้นที่ในการวิเคราะห์ว่าพื้นที่ใดเป็นพยัญชนะ สระ หรือวรรณยุกต์ โดยการวิเคราะห์จะพิจารณาดังนี้

1) พื้นที่ของตัวพยัญชนะ ซึ่งได้แก่ตัวพยัญชนะภาษาไทยที่ไม่มีเชิง และสระได้แก่ อะ อา โอะ โอ แอะ แอ และ เอ พื้นที่ส่วนนี้จะพิจารณาจากตำแหน่งการวางตัวอักษรและอัตราส่วนความกว้างต่อความสูงของสี่เหลี่ยมที่ล้อมรอบตัวอักษรให้ตรงกับสระ และพยัญชนะที่กำหนดไว้

2) พื้นที่ของสระและวรรณยุกต์ ได้แก่ อิ อี อึ อือ ไม้เอก ไม้โท ไม้ตรี ไม้จัตวา ตัวการันต์ และสระอำ พื้นที่ส่วนนี้จะพิจารณาจากอัตราส่วนความกว้างต่อความสูงของสี่เหลี่ยมรอบตัวอักษรให้ตรงกับสระและวรรณยุกต์ที่กำหนดไว้

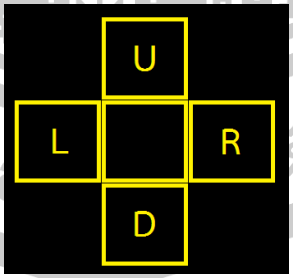
3) พื้นที่ของพยัญชะที่มีเชิง ได้แก่ ฐ ญ ฏ ฎ ฤ และ ฦ และสระได้แก่ อุ อู พื้นที่ส่วนนี้จะพิจารณาจากอัตราส่วนความกว้างต่อความสูงของสี่เหลี่ยมที่ล้อมรอบตัวอักษรให้ตรงกับสระและส่วนของพยัญชนะที่กำหนดไว้ และรวมถึงการให้ตำแหน่งการว่างตัวร่วมด้วยเพื่อทำการพิจารณาตัวอักษรที่มีเชิงของอักษรนั้นว่าเป็นตัวอักษรในพื้นที่นี้ด้วย

จากวิธีการดังกล่าวนี้จะมีความเหมาะสมในการนำไปใช้กับตัวอักษรที่มีรูปแบบที่เป็นรูปแบบมาตรฐานและมีการวางตัวอักษรในแนวตรงเท่านั้น หากมีการวางแนวไม่เป็นแนวตรงจะประสบปัญหาดังรูปที่ 143

รูปที่ 143 การแบ่งพื้นที่ในกรอบข้อความพบตัวอักษรที่มีการวางแนวเฉียง

จากรูปที่ 143 คือปัญหาที่เกิดขึ้นในการวิเคราะห์พื้นที่จะสังเกตเห็นว่า ตัวสระ "อุ" แท้ที่จริงแล้วควรจะอยู่ในพื้นที่ส่วน 3 แต่ผลในการวิเคราะห์พื้นที่อยู่ในพื้นที่ส่วนที่ 1 ซึ่งจากผลดังกล่าวจะทำให้กระบวนการรู้จำอักขระสามารถทำได้ยากขึ้น และอาจจะส่งผลให้การรู้จำผิดพลาดอีกด้วย ทั้งนี้เนื่องจากงานวิจัยฉบับนี้ข้อมูลจะอยู่ในรูปแบบของตัวอักษรไม่ได้อยู่ในรูปแบบที่เป็นรูปแบบมาตรฐาน ดังนั้นจะต้องมีการพัฒนาการวิเคราะห์รูปแบบตัวอักษรให้มีประสิทธิภาพมากขึ้นจะที่สามารถวิเคราะห์ตัวอักษรในรูปแบบต่าง ๆ ได้

ในการแก้ปัญหาดังกล่าว การพิจารณาถึงตำแหน่งของตัวอักษรว่าอยู่ในระดับใดงานวิจัยที่ผ่านมาจะนิยมใช้การวิเคราะห์รูปแบบตัวอักษร (Layout Analysis) ในการวิเคราะห์หาพื้นที่ของตัวอักษร ผู้วิจัยจึงได้คิดกระบวนการวิเคราะห์รูปแบบตัวอักษรด้วยแม่แบบ (Template Layout Analysis) ซึ่งเป็นกระบวนการวิเคราะห์ถึงตำแหน่งของตัวอักษรโดยการใช้แม่แบบ ดังรูปที่ 144 ในการวิเคราะห์หาตำแหน่งของตัวอักษรว่าเป็นตัวอักษรกลุ่มใด ซึ่งในที่นี้จะแบ่งเป็น 3 ระดับ คือ ระดับบน ระดับกลาง และระดับล่าง แทนการแบ่งพื้นที่ตัวอักษร



รูปที่ 144 ภาพแม่แบบที่ใช้ในการวิเคราะห์รูปแบบตัวอักษร

จากรูปที่ 144 U คือตำแหน่งของตัวอักษรที่อยู่ด้านบน D คือตำแหน่งตัวอักษรที่อยู่ด้านล่าง L คือตำแหน่งตัวอักษรที่อยู่ด้านซ้าย และ R คือตำแหน่งตัวอักษรขวา โดยกระบวนการวิเคราะห์จะกระทำด้วยการปรับขนาดของแม่แบบให้มีขนาดตามตัวอักษรที่ต้องการวิเคราะห์ ซึ่งขนาดของตัวอักษรสามารถทราบได้จากกระบวนการแบ่งส่วนข้อความที่ได้ดำเนินการผ่านมาแล้ว และขั้นตอนต่อมาคือการวางแม่แบบไปบนตัวอักษรที่ต้องการ ซึ่งการวิเคราะห์ระดับของตัวอักษรจะสามารถวิเคราะห์ได้ดังนี้

1) ระดับบน จะพิจารณาจากตำแหน่ง U ที่จะต้องไม่มีตัวอักษรในตำแหน่งนี้ ตำแหน่ง D จำเป็นต้องมีตัวอักษรในตำแหน่งนี้ และ ตำแหน่ง L R จะมีตัวอักษรหรือไม่มีก็ได้ ดัง

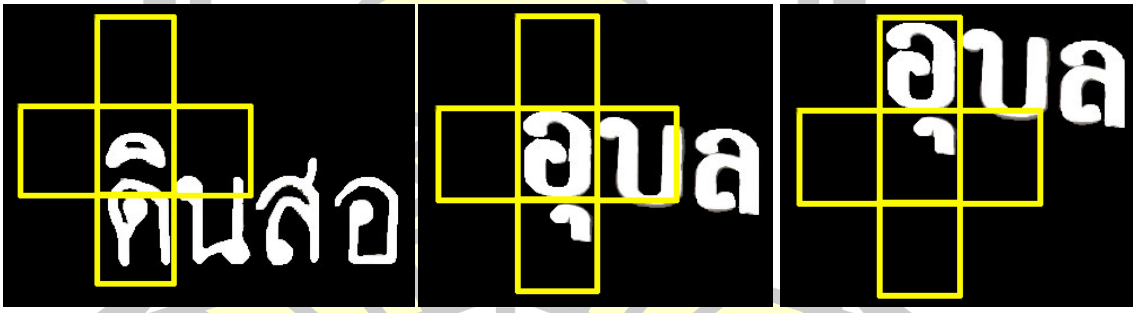
รูปที่ 145 (ก)

2) ระดับกลาง จะพิจารณาจากตำแหน่ง U D L R ว่าในตำแหน่งต่างเหล่านี้จะมีตัวอักษรหรือไม่มีก็ได้ ดัง

รูปที่ 145 (ข)

3) ระดับล่าง จะพิจารณาจากตำแหน่ง D ที่จะต้องไม่มีตัวอักษรในตำแหน่งนี้ ตำแหน่ง U จะมีต้องมีตัวอักษรอยู่ในตำแหน่งนี้ และ L R จะมีหรือไม่มีก็ได้ ดัง

รูปที่ 145 (ค)

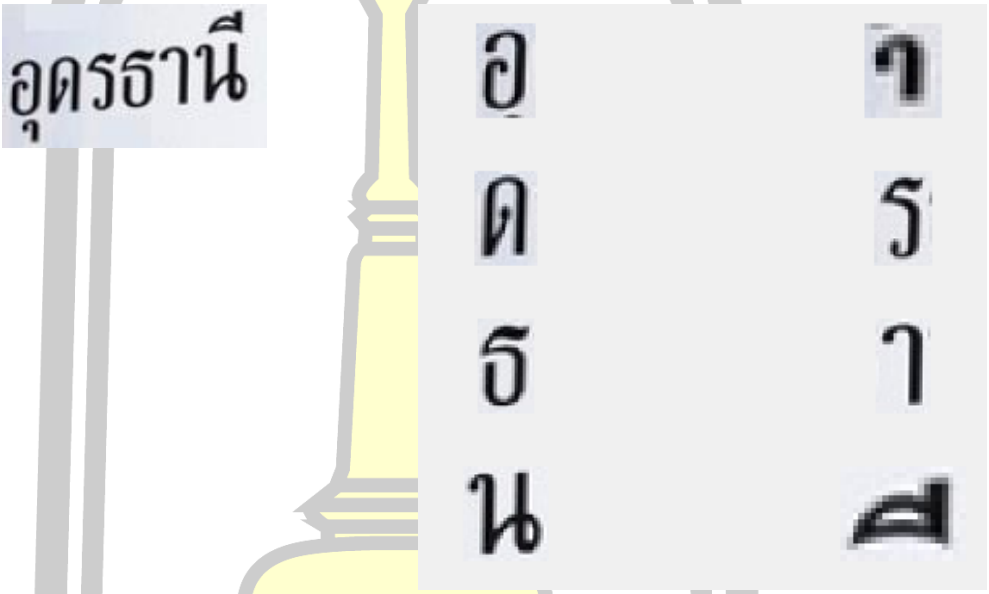


(ก) (ข) (ค)

รูปที่ 145 (ก) แสดงตำแหน่งแม่แบบในการพิจารณาตัวอักษรระดับบน (ข) แสดงตำแหน่งแม่แบบในการพิจารณาตัวอักษรระดับกลาง และ(ค) แสดงตำแหน่งแม่แบบในการพิจารณาตำแหน่งตัวอักษรระดับล่าง

ซึ่งวิธีการนี้จะมีการแบ่งระดับคล้ายกับการวิเคราะห์รูปแบบตัวอักษร (Layout Analysis) ในวิธีการที่ผ่านมา แต่วิธีการวิเคราะห์ที่ผ่านมานั้นจะเป็นการพิจารณาถึงการวางแนวเส้นแบ่งเพื่อพิจารณาถึงระดับของตัวอักษร ทำให้เกิดปัญหาจากการวางแนวของเส้นแบ่งซึ่งจะส่งผลกระทบการ

การวิเคราะห์ทั้งหมด ทั้งนี้วิธีการที่นำเสนอจะมีการพิจารณาที่ตัวอักษรแต่ละตัวด้วยการพิจารณาบริบทรอบข้าง เพื่อนำมาระบุตำแหน่งของตัวอักษรที่สนใจ และขั้นต่อไปจะเป็นการกำหนดลาเบล (Label) ให้กับตัวอักษรเพื่อเป็นการระบุว่าตัวอักษรนี้อยู่ในระดับใด ซึ่งจะเป็นการกำหนดลำดับของอักขระที่จะใช้ในการรู้จำและสร้างข้อความต่อไป ตัวอย่างของการแยกอักขระแสดงดังรูปที่ 146 –รูปที่ 149



รูปที่ 146 ตัวอย่างของการตัดอักขระจากภาพ

รูปที่ 147 ตัวอย่างของการตัดอักขระจากภาพ

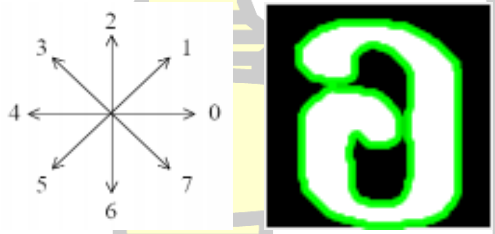
รูปที่ 148 ตัวอย่างของการตัดอักขระจากภาพ

รูปที่ 149 ตัวอย่างของการตัดอักขระจากภาพ

4.2 การสกัดคุณลักษณะจากอักขระ (Feature Extraction)

การสกัดคุณลักษณะเป็นกระบวนการที่สกัดหาข้อมูลจากภาพตัวอักษร เพื่อนำมาสร้างเป็นเวกเตอร์ (Vector) ของภาพตัวอักษรนั้น ๆ เพื่อใช้สำหรับการจำแนก โดยในงานวิจัยนี้จะใช้เทคนิค 4 วิธีการเพื่ออธิบายคุณลักษณะของภาพตัวอักษรคือ 1) การอธิบายคุณลักษณะทางรูปร่างด้วยรหัสลูกโซ่ (Chain Code) 2) การอธิบายคุณลักษณะความหนาแน่นของจุดพิกเซลด้วยวิธีการแบ่งโซน (Zoning) 3) การอธิบายคุณลักษณะความถี่โดยใช้วิธีฮิสโตแกรมโปรเจคชันในแนวตั้งและแนวนอน (Histogram Projection) และ 4) การอธิบายคุณลักษณะเด่นของภาพด้วยวิธีการ Histograms of Oriented Gradients (HOG)

1) การอธิบายคุณลักษณะทางรูปร่างด้วยรหัสลูกโซ่ (Chain Code) คือวิธีการแสดงรูปร่างของวัตถุ สามารถนำไปใช้ในการอธิบายลำดับของจุดพิกเซลที่ต่อเนื่องกันไปเป็นตัวอักษร โดยอาศัยการเปลี่ยนแปลงทิศทางของพิกเซลตามเส้นขอบว่าจะเป็นไปในทิศทางใด ซึ่งในงานวิจัยนี้จะเป็นแบบ 8 ทิศทางดังรูปที่ 150 (ก)

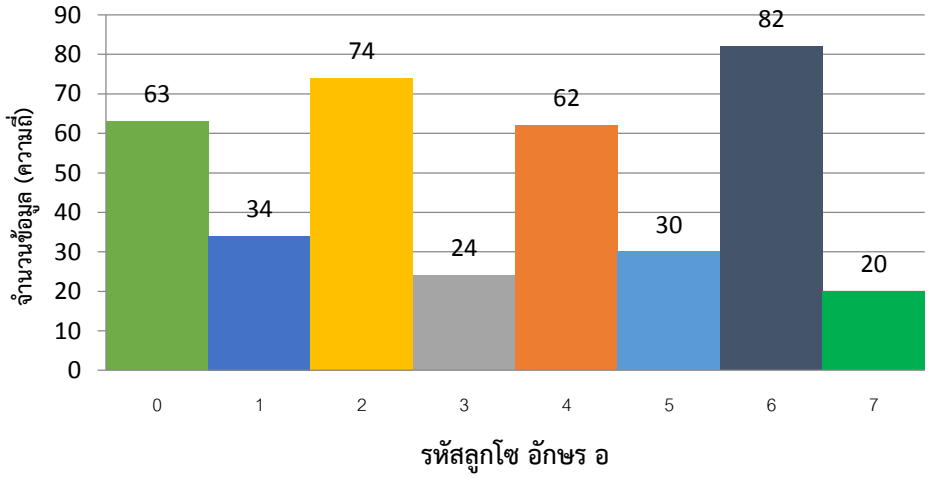


(ก) (ข)

รูปที่ 150 (ก) ทิศทางในการกำหนดรหัสลูกโซ่แบบ 8 ทิศ (ข) เส้นรอบตัวอักษรที่ทำการพิจารณารหัสลูกโซ่

จากรูปที่ 150 (ข) เป็นตัวอย่างของตัวอักษร "อ" ที่จะพิจารณารหัสลูกโซ่จากเส้นขอบรอบตัวอักษรที่มีขนาดภาพอยู่ที่ 80x80 พิกเซล โดยถูกแปลงเป็นรหัสลูกโซ่แบบ 8 ทิศจะได้รหัส "2 2 2 2 2 2 1 1 1 . . . 3 3 0" ซึ่งมีจำนวนรหัสลูกโซ่ทั้งหมด 389 รหัส ทั้งนี้ในการได้มาซึ่งรหัสลูกโซ่ในแต่ละตัวอักษรจะได้จำนวนรหัสลูกโซ่ที่ไม่เท่ากันในแต่ละตัวอักษร เช่น ตัวอักษร "ล" จะมีจำนวนรหัสลูกโซ่ทั้งหมด 319 รหัส หรือสระ "อา" มีจำนวนรหัสลูกโซทั้งหมด 214 รหัส เป็นต้น เมื่อผ่านการแปลงเป็นรหัสลูกโซ่แล้ว นำมาทำการนอร์มัลไลเซชัน (Normalization) ด้วยการแปลงรหัสลูกโซ่ที่ได้ให้อยู่ในรูปของเวกเตอร์จำนวน 8 ช่อง ซึ่งเวกเตอร์แต่ละช่องคือทิศทางในแต่ละทิศของรหัสลูกโซ่

และค่าที่เก็บในเวกเตอร์แต่ละช่องคือความถี่ของทิศทางนั้น ๆ เมื่อนำเวกเตอร์มาสร้างเป็นกราฟโดยจะเรียงลำดับตัวเลขทิศทางในระนาบแกน X และความถี่ของรหัสลูกโซ่ในแต่ละทิศทางในระนาบแกน Y ซึ่งข้อมูลของความถี่ทั้งหมดรวมจะต้องไม่มีค่าเกินจำนวนทั้งหมดของรหัสลูกโซ่ที่แปลงมาได้ เช่น ตัวอักษร "อ" จะสามารถมีจำนวนความถี่ทั้งหมดรวมกันได้ไม่เกิน 389 ดังรูปที่ 151



รูปที่ 151 กราฟแสดงความถี่ของรหัสลูกโซ่ของตัวอักษร "อ"

จากรูปที่ 151 หากพิจารณาค่าความถี่ที่อยู่ในแต่ละทิศทางจะพบว่า ค่าดังกล่าวนั้นจะมีค่าที่มากซึ่งจะส่งผลให้มีการใช้หน่วยความจำและเวลาในการประมวลผลที่มากขึ้น ดังนั้นเพื่อเป็นการลดการใช้หน่วยความจำและเวลาในการประมวลผลจำเป็นต้องมีการแปลงข้อมูลให้มีค่าเล็กลง ด้วยวิธีการ Min-Max Normalization ซึ่งเป็นการแปลงข้อมูลให้อยู่ในช่วงของค่าที่ต้องการโดยสมการที่ 4.1

$$v' = \frac{v - \min}{\max - \min}(\max' - \min') + \min' \qquad (4.1)$$

โดยกำหนดให้

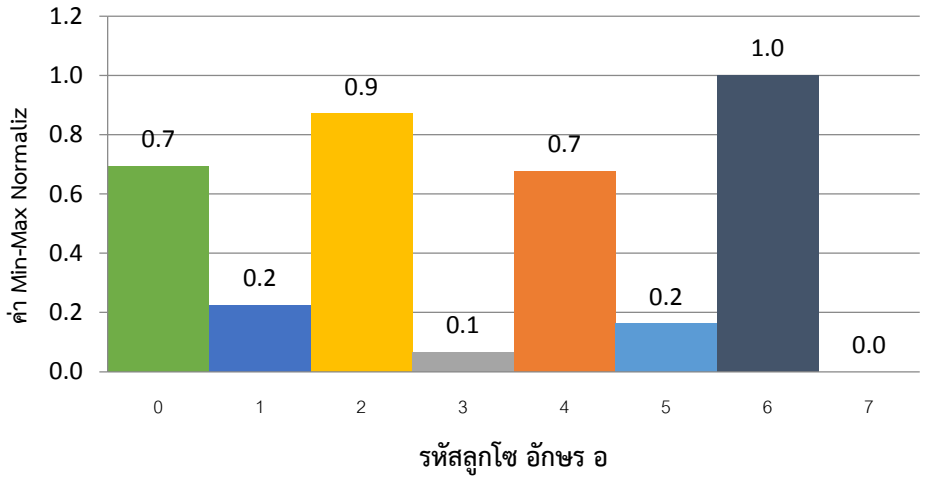
$\min$ คือค่าที่น้อยที่สุดของชุดข้อมูล

$\max$ คือค่าที่มากที่สุดของชุดข้อมูล

$\min'$ คือค่าใหม่ที่มีค่าน้อยที่สุด

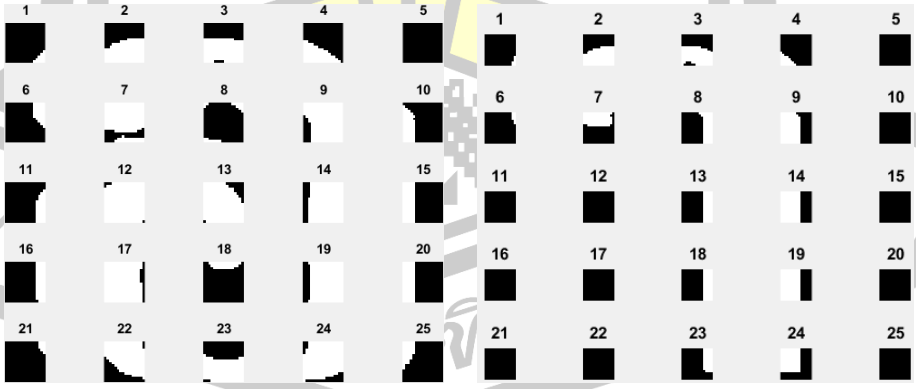
$\max'$ คือค่าใหม่ที่มีค่ามากที่สุด

ในการแปลงข้อมูลด้วยวิธีการ Min-Max Normalization ในงานวิจัยนี้จะแปลงข้อมูลให้อยู่ในช่วง 0 - 1 โดยการแปลงข้อมูลกับทุก ๆ เวกเตอร์ของรหัสลูกโซ่เมื่อนำมาสร้างเป็นกราฟจะได้ดังรูปที่ 152



รูปที่ 152 กราฟแสดงค่า Min-Max Normalization ของรหัสลูกโซ่ของตัวอักษร "อ"

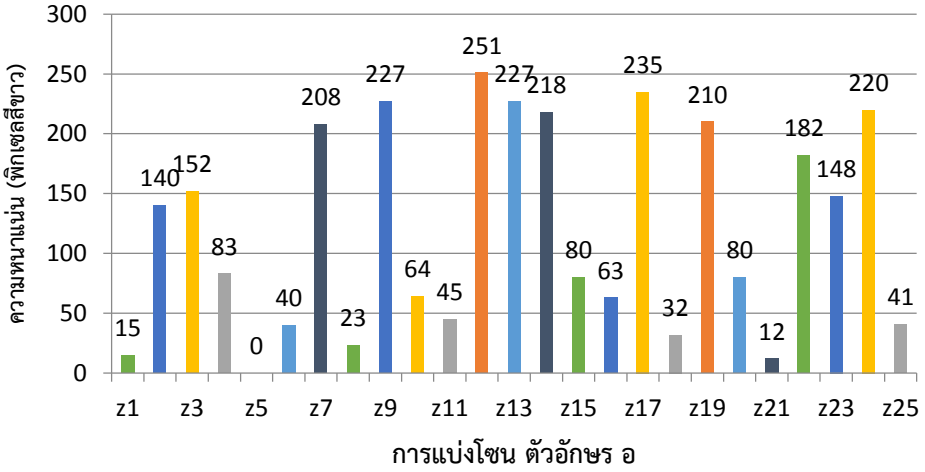
2) การอธิบายคุณลักษณะความหนาแน่นของจุดพิกเซลด้วยวิธีการแบ่งโซน (Zoning) เป็นกระบวนการแบ่งภาพออกเป็นโซนเพื่อหาความหนาแน่นของจุดพิกเซลในแต่ละโซน โดยในงานวิจัยนี้ได้ทำการแบ่งโซนออกเป็น 25 โซนแต่ละโซนจะมีขนาด 16x16 พิกเซลจากภาพต้นฉบับขนาด 80x80 พิกเซล ดัง
รูปที่ 153



(ก) (ข)

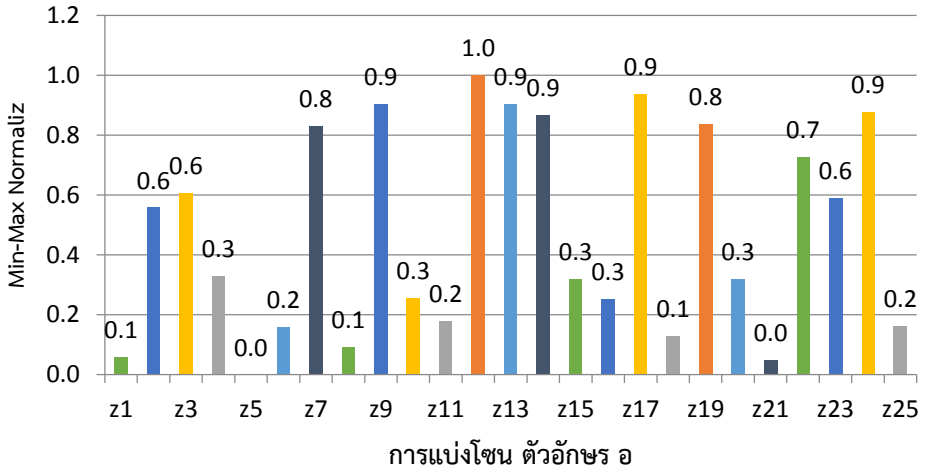
รูปที่ 153 ภาพการแบ่งโซน 25 โซน (ก) การแบ่งโซนของตัวอักษร "อ" (ข) การแบ่งโซนของตัวสระ "อา"

รูปที่ 153 (ก) คือตัวอย่างของการแบ่งโซนของตัวอักษร "อ" เป็นจำนวน 25 โซน หลังจากนั้นดำเนินการนับจำนวนพิกเซลที่มีค่าเท่ากับ 1 (สีขาว) ในแต่ละโซนซึ่งจะได้ค่าเวกเตอร์ทั้งหมด 25 ค่า เมื่อนำมาสร้างเป็นกราฟจะได้ดังรูปที่ 154



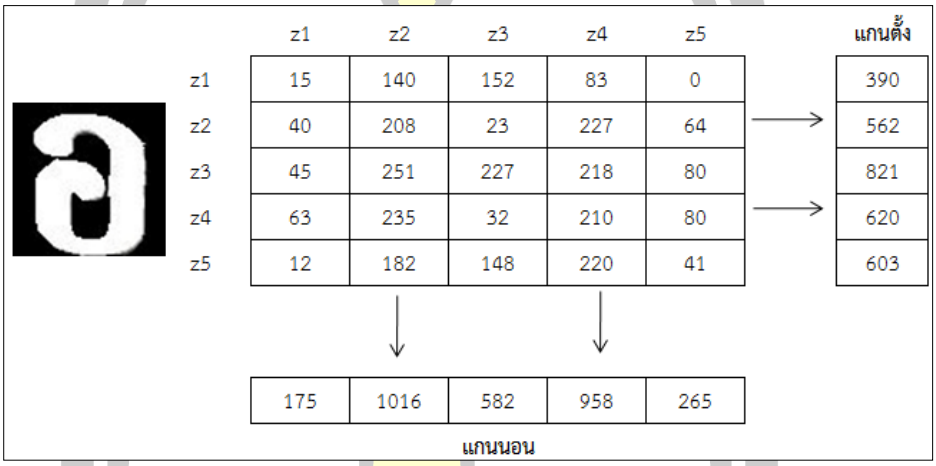
รูปที่ 154 กราฟแสดงความหนาแน่นจากการแบ่งโซนของอักษรตัว "อ"

จากรูปที่ 154 คือการแสดงค่าความหนาแน่น (พิกเซลสีขาว) ในแต่ละโซนของตัวอักษร "อ" เพื่อเป็นการลดการใช้หน่วยความจำและเวลาในการประมวลผลจำเป็นต้องมีการแปลงข้อมูลให้มีค่าเล็กลง ด้วยวิธีการ Min-Max Normalization ซึ่งจะมีค่าในแต่ละโซนอยู่ในช่วง 0 - 1 เมื่อนำมาสร้างเป็นกราฟจะได้ดังรูปที่ 155



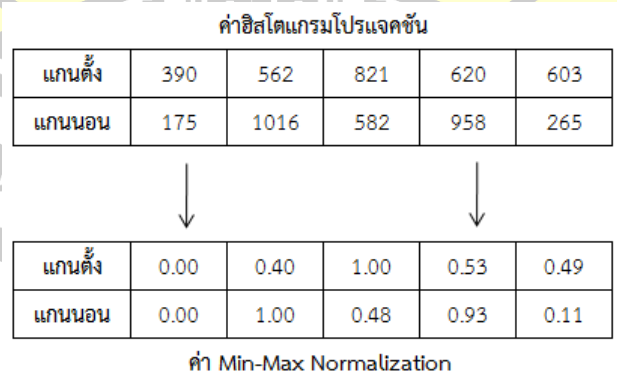
รูปที่ 155 กราฟแสดงค่า Min-Max Normalization ด้วยวิธีการแบ่งโซนของตัวอักษร "อ"

3) การอธิบายคุณลักษณะความถี่โดยใช้วิธีฮิสโตแกรมโปรเจคชันในแนวตั้งและแนวนอน (Histogram Projection) คือการทำการฉายภาพ (Projection) เพื่อหาความถี่ของข้อมูลในแต่ละระนาบที่ใช้ในการแยกกลุ่มตัวอักษรที่มีความคล้ายคลึงกันด้วยการแบ่งภาพที่มีขนาด 80x80 ออกเป็น 5 โซนขนาด 16x16 ตามแนวแกนตั้งและนอน นับจำนวนจุดพิกเซลที่มีค่าเท่ากับ 1 (สีขาว) ตามแนวแกนตั้งจนครบทุกแถว และนับจำนวนจุดพิกเซลตามแนวแกนนอนจนครบทุกแถวดังรูปที่ 156



รูปที่ 156 การหาค่าฮิสโตแกรมโปรเจคชันในแนวตั้งและแนวนอน

หลังจากที่ได้ค่าฮิสโตแกรมโปรเจคชันในแนวตั้งและแนวนอนจะพบว่าค่าฮิสโตแกรมในแต่ละแกนอยู่ในช่วงที่มีค่าสูง เพื่อเป็นการลดการใช้หน่วยความจำและเวลาในการประมวลผลจำเป็นต้องมีการแปลงข้อมูลให้มีค่าเล็กลง ด้วยวิธีการ Min-Max Normalization ซึ่งจะมีค่าในแต่ละโซนอยู่ในช่วง 0 - 1 ดังรูปที่ 157



รูปที่ 157 การแปลงค่าฮิสโตแกรมโปรเจคชันด้วยวิธีการ Min-Max Normalization

4) การอธิบายคุณลักษณะเด่นของภาพด้วยวิธีการ Histograms of Oriented Gradients (HOG) Navneet Dalal และ Bill Triggs [105] ได้คิดค้นวิธีการนี้ขึ้นมาเพื่อใช้ในการตรวจจับมนุษย์ โดยพื้นฐานของงานวิจัยจะใช้คุณลักษณะเด่นของภาพด้วยรูปร่าง ซึ่งงานวิจัยนี้ได้นำวิธีการนี้มาใช้เพื่ออธิบายรูปร่างของตัวอักษร ขั้นตอนวิธีการคำนวณด้วย HOG สามารถดำเนินการได้ดังนี้ [106]

1. ปรับขนาดรูปภาพในงานวิจัยนี้ให้มีค่าเท่ากับ 32x32 พิกเซล และคำนวณหาค่าเกรเดียนต์ของภาพด้วยการใช้สูตรในสมการที่ 4.2 และ 4.3

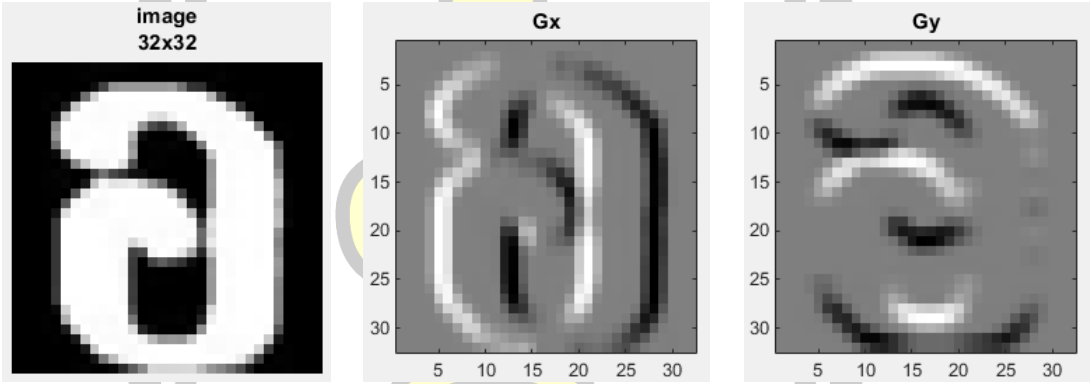
$$G_x = M_x * I = [-1 \quad 0 \quad 1] \tag{4.2}$$

$$G_y = M_y * I = [-1 \quad 0 \quad 1]^T \tag{4.3}$$

เมื่อ

I คือ ภาพระดับเทา

G_x และ G_y คือ อนุพันธ์อันดับที่ 1 ของภาพตามแนวนอนและแนวตั้งตามลำดับ



(ก) (ข) (ค)

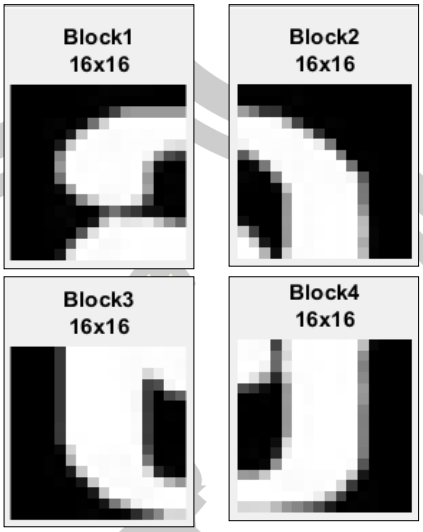
รูปที่ 158 (ก) ภาพตัวอักษรต้นฉบับ ภาพ (ข) และ (ค) ค่าเกรเดียนต์ของภาพในแนวตั้งและแนวนอน

จากรูปที่ 158 (ก) คือภาพตัวอักษรต้นฉบับที่ผ่านการปรับขนาดให้มีขนาด 32x32 พิกเซล (ข) และ (ค) คือภาพผลของการคำนวณหาค่าเกรเดียนต์ของภาพในแนวตั้งและแนวนอนตามลำดับ

2. แบ่งภาพเป็นภาพย่อยซึ่งภาพย่อยนี้จะถูกเรียกว่าบล็อก (Block) และจะมีขนาดเท่ากับ 16x16 พิกเซล ดังนั้นภาพที่มีขนาด 32x32 พิกเซลจะมีจำนวนบล็อกเท่ากับ 4 บล็อก แสดง

ดัง

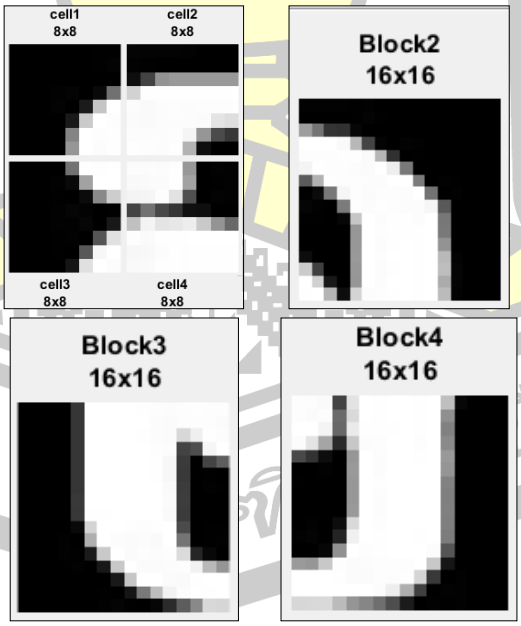
รูปที่ 159



รูปที่ 159 การแบ่งภาพย่อย 4 บล็อก

3. กำหนดจำนวนเซลล์ (Cell) ในแต่ละบล็อก โดยกำหนดให้เซลล์มีขนาด 2x2 เซลล์ ซึ่งในแต่ละเซลล์จะมีขนาดเป็น 8x8 พิกเซล ดัง

รูปที่ 160



รูปที่ 160 การกำหนดเซลล์ในบล็อก

4. หาขนาดและทิศทางของภาพโดยใช้สมการที่ 4.4 และ 4.5 ตามลำดับ

$$\left|G(x,y)_{k=i}\right| = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (4.4)$$

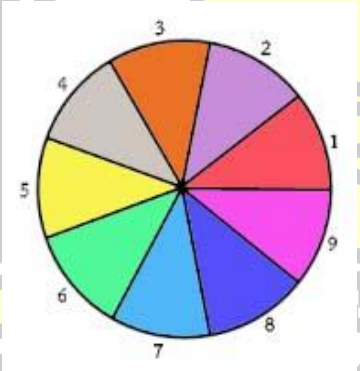
กำหนดให้

i คือ ลำดับบล็อก $1 \le i \le k$

คำนวณทิศทางแต่ละบล็อกที่ตำแหน่ง (x,y) ค่าทิศทางที่หาได้จะมีค่า $0^\circ - 360^\circ$

$$q(x,y)_{k=i} = \arctan(\frac{G_y}{G_x}) \quad (4.5)$$

5. ทิศทางของค่าเกรเดียนต์ ทั้งนี้แต่ละเซลล์จะถูกนำมาสร้างช่องฮิสโตรแกรมสำหรับเก็บทิศทาง 0-360 องศา ซึ่งจะมีช่องฮิสโตรแกรมจำนวน 9 ช่อง (Bin) ดังรูปที่ 161



จำนวนถัง (Bin)	ช่วงขององศา
1	0-40
2	40-80
3	80-120
4	120-160
5	160-200
6	200-240
7	240-280
8	280-320
9	320-360

รูปที่ 161 แสดงการกำหนดถังกับทิศทาง 0-360 องศา

การคำนวณหาผลรวมของแต่ละทิศทางที่กำหนดไว้จำนวน 9 ช่องในแต่ละเซลล์สามารถดำเนินการได้ดังสมการที่ 4.6

$$C_b = \sum_{i=1}^{n} q(x,y)_i = b \tag{4.6}$$

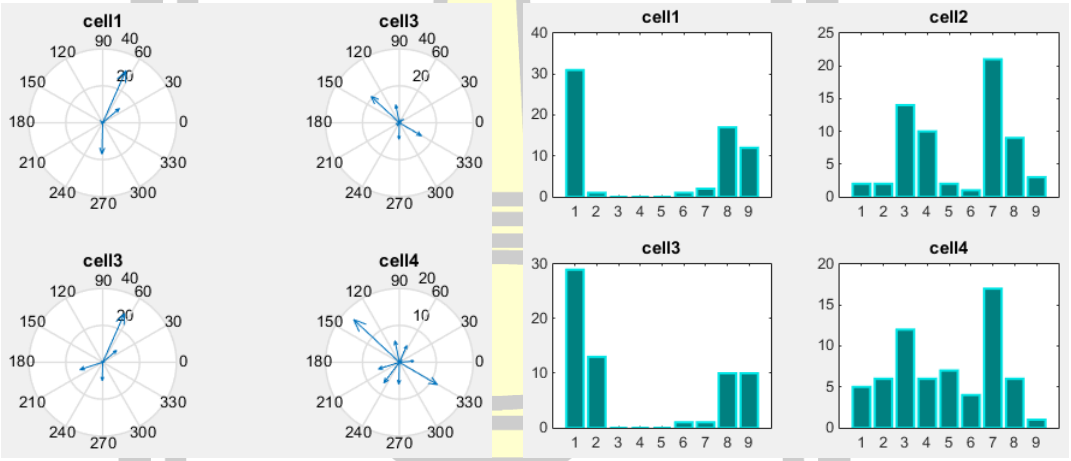
กำหนดให้

n คือ จำนวนตำแหน่ง (x,y) ในแต่ละเซลล์

b คือ ทิศทางที่พิจารณา

C_b คือ ผลรวมของแต่ละทิศทาง

จากสมการที่ 4.6 เมื่อดำเนินการคำนวณกับทุกเซลล์ในบล็อกแรก และนำมาแสดงเป็นกราฟจะได้ดังรูปที่ 152



(ก) (ข)

รูปที่ 162 (ก) กราฟแสดงทิศทาง 9 ช่อง (Bin) ในแต่ละเซลล์ของบล็อกแรก (ข) กราฟแสดงความถี่ในแต่ละทิศทางของเซลล์ในบล็อกแรก

6. คำนวณคุณลักษณะเด่นของ 9 ช่องในแต่ละเซลล์ดังสมการที่ 4.7

$$V_k = \sum_{i=1}^{n} \left(\left| G(x,y)_k \right| * C_b \right) \mid q(x,y) = b \tag{4.7}$$

โดยแต่ละบล็อกจะรวมเป็นคุณลักษณะเด่นดังสมการที่ 4.8

$$V_k = \begin{bmatrix} V_{k=1} \\ V_{k=2} \\ \dots \\ V_{k=b} \end{bmatrix} \quad (4.8)$$

7. ปรับค่าคุณลักษณะเด่นด้วยสมการที่ 4.9

$$V = \frac{V_k}{\sqrt{\|V_k\|^2 + 1}} \quad (4.9)$$

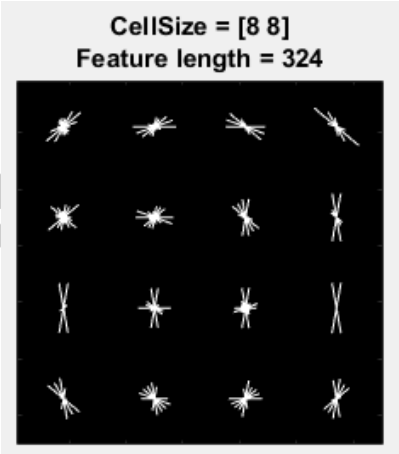
กำหนดให้

v คือ ค่าคุณลักษณะเด่นที่ปรับอย่างเหมาะสม

จากตัวอย่างการคำนวณหาค่าคุณลักษณะเด่นจะเป็นการคำนวณในบล็อกแรกเท่านั้น ดังนั้นจำเป็นต้องทำการคำนวณกับทุก ๆ เซลล์ในแต่ละบล็อก ซึ่งสุดท้ายจะได้คุณลักษณะเด่นของภาพทั้งหมด ในการหาจำนวนคุณลักษณะของ HOG จะสามารถคำนวณได้จากสมการ

จำนวนคุณลักษณะของ HOG = จำนวนบล็อก x จำนวนเซลล์ต่อบล็อก x จำนวนของช่อง (Bin)

ในงานวิจัยได้มีการกำหนดจำนวนบล็อกทั้งหมด 9 บล็อก จำนวนเซลล์ต่อบล็อก 4 เซลล์ และจำนวนทิศทาง 9 ช่อง ดังนั้นจะได้เท่ากับ 9x4x9 = 324 คุณลักษณะ ซึ่งจะได้ดังรูปที่ 163



รูปที่ 163 ภาพแสดงคุณลักษณะเด่นของตัวอักษร อ

จากการสกัดคุณลักษณะในแต่ละวิธีจะได้คุณลักษณะทั้งหมด 367 คุณลักษณะจาก การอธิบายคุณลักษณะทางรูปร่างด้วยรหัสลูกโซ่จำนวน 8 คุณลักษณะ การอธิบายคุณลักษณะความหนาแน่นของจุดพิกเซลด้วยวิธีการแบ่งโซนจำนวน 25 คุณลักษณะ การอธิบายคุณลักษณะความถี่โดยใช้วิธีฮิสโตแกรมโปรเจคชันในแนวตั้งและแนวนอนจำนวน 10 คุณลักษณะ และการอธิบายคุณลักษณะเด่นของภาพด้วยวิธีการ HOG จำนวน 324 คุณลักษณะที่ถูกนำมาใช้ในการอธิบายตัวอักขระ

4.3 การรู้จำตัวอักขระ (Character Recognition)

1) การเตรียมชุดข้อมูลชุดฝึกฝนสำหรับการรู้จำตัวอักขระ

ชุดข้อมูลชุดฝึกฝนที่ใช้สำหรับการรู้จำตัวอักขระภาษาไทย ประกอบไปด้วยรูปภาพอักขระภาษาไทยที่นำมาจัดกลุ่มได้ 85 กลุ่มอักขระ และแบ่งกลุ่มตามตำแหน่งของตัวอักขระเป็น 3 กลุ่ม ได้แก่ กลุ่มบน กลุ่มกลางและกลุ่มล่าง ดังตาราง 4 แล้วทุกกลุ่มตัวอักขระจะถูกนำมากสกัดคุณลักษณะตามกระบวนการที่กล่าวมาแล้ว เช่น รหัสลูกโซ่ (Chain Code) การแบ่งโซน (Zoning) วิธีฮิสโตแกรมโปรเจคชัน (Histogram Projection) และ Histograms of Oriented Gradients (HOG)

ตาราง 4 ตัวอักษรและตำแหน่งของตัวอักษรที่ใช้ในการจัดกลุ่มข้อมูล

ตำแหน่ง	ชนิดตัวอักขระ	จำนวน	กลุ่มอักขระภาษาไทย						
กลุ่มบน 12	วรรณยุกต์	4	่	้	๊	๋			
	สระ	7	ิ	ี	ึ	ื	ั	็	๎
	วรรคตอน	1	์						
กลุ่มกลาง 71	พยัญชนะ	44	ก	ข	ฃ	ค	ฅ	ฆ	ง
			จ	ฉ	ช	ซ	ฌ	ญ	ฎ
			ฏ	ฐ	ฑ	ฒ	ณ	ด	ต
			ถ	ท	ธ	น	บ	ป	ผ
			ฝ	พ	ฟ	ภ	ม	ย	ร
			ล	ว	ศ	ษ	ส	ห	ฬ
			อ	ฮ					
	สระ	7	ะ	า	เ	แ	โ	ใ	ไ
	ตัวเลข	20	๐	๑	๒	๓	๔	๕	๖
			๗	๘	๙				
			0	1	2	3	4	5	6
			7	8	9				
กลุ่มล่าง 2	สระ	2	ุ	ู					

2) การจำแนกประเภท

การจำแนกประเภทเป็นกระบวนการที่ทำให้คอมพิวเตอร์ทำนายข้อมูลใหม่ที่สนใจโดยนำมาเปรียบเทียบกับข้อมูลเดิมที่ได้บันทึกไว้ แล้วพิจารณาว่าข้อมูลใหม่ที่สนใจนั้นมีค่าใกล้เคียงกับข้อมูลเดิมที่บันทึกไว้หรือไม่ ซึ่งวิธีการที่ใช้ในการจำแนกประเภทในงานวิจัยนี้ได้แก่ 1) วิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor หรือ KNN) 2) ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine: SVM) และ 3) โครงข่ายประสาทเทียม (Neural Network) ในงานวิจัยนี้จะดำเนินการทดสอบการจำแนกประเภทแต่ละวิธีการเพื่อหาวิธีการที่มีประสิทธิภาพมากที่สุดต่อไป

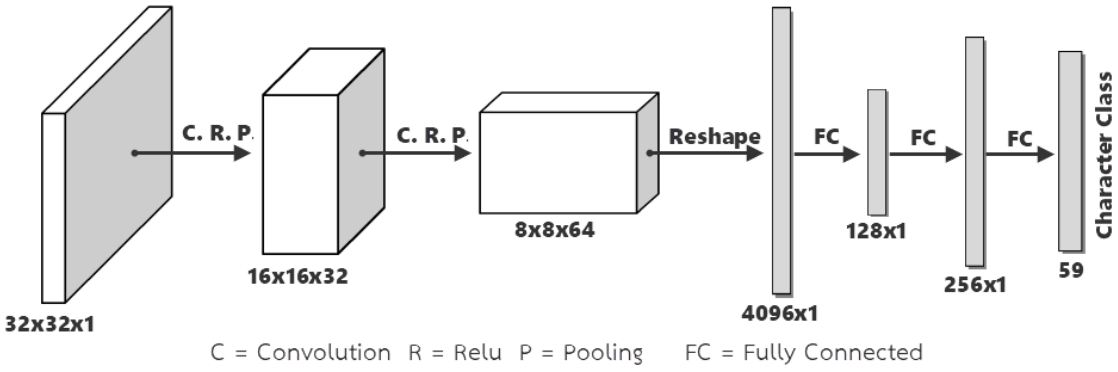
1. การใช้ Feature based ร่วมกับซัพพอร์ตเวคเตอร์แมชชีน (SVM)

วิธีการนี้จะนำข้อมูลที่ได้จากกระบวนการสกัดคุณลักษณะ มาผ่านการจำแนกตัวอักขระด้วยซัพพอร์ตเวคเตอร์แมชชีน โดยคุณลักษณะที่ถูกนำมาจำแนกตัวอักขระได้แก่ Chain code, Zoning, Histogram Projection และ HOG

2. Convolution Neural Network (CNN)

ในการดำเนินงานครั้งนี้ได้มีการประยุกต์ใช้ Convolution Neural Network ในการจำแนกตัวอักขระและรูปแบบของตัวอักขระที่มีความแตกต่างกัน โดยการสร้างเครือข่าย Convolution Neural Network จะแบ่งออกเป็น 7 ชั้น ดัง

รูปที่ 164



รูปที่ 164 โครงสร้างของ Convolution neural network (CNN)

ชั้นที่ 1) Input Layer เป็นชั้นที่นำเข้าภาพตัวอักขระ ที่ผ่านการแปลงภาพให้เป็นภาพระดับเทา (Grayscale) ที่มีขนาด 32x32 พิกเซล

ชั้นที่ 2-3) Convolution Layer ในการดำเนินงานครั้งนี้จะใช้ Convolution Layer ทั้งหมด 2 ชั้น โดยแต่ละชั้นจะมีกระบวนการดังต่อไปนี้

(1) Convolution เป็นกระบวนการนำตัวกรอง (Filter) ขนาด 3x3 มาผ่านการคำนวณที่ได้อธิบายไว้แล้วในบทที่ 2 เรื่องการ Convolution โดยผลลัพธ์ในกระบวนการนี้จะได้ภาพเอาต์พุต (Output) หรือ Feature maps ของชั้นแรกจำนวน 32 เอาต์พุต และชั้นที่สองจำนวน 64 เอาต์พุต

(2) Relu เป็น Activating function ที่มีการกำหนดค่าเทรชโฮลด์ (Threshold) โดยการเปลี่ยนแปลงค่าที่เป็นค่าเชิงลบ (Negative) ให้เป็นค่าเชิงบวก (Positive) คือจากค่าที่ถูกส่ง

มาจากกระบวนการ Convolution หากมีพิกเซลใดที่มีค่าติดลบก็จะถูกเปลี่ยนให้มีค่าเป็น 0 ซึ่งกระบวนการ Relu จะดำเนินการกับทุก ๆ ภาพที่ถูกส่งเข้ามาจากกระบวนการ Convolution

(3) Pooling หรือ Max Pooling เป็นกระบวนการในการลดสัดส่วนของภาพ ที่จะดำเนินการโดยการสุ่มตัวอย่าง (down-sampling) ด้วยการกำหนดขนาดของหน้าต่างที่มีขนาด 2x2 และจะพิจารณาค่าที่มีอยู่ภายในหน้าต่างเท่านั้น ซึ่งการพิจารณาจะเลือกค่าที่มีค่ามากที่สุดเพียงหนึ่งค่าเท่านั้น หลังจากนั้นจะดำเนินการเลื่อนหน้าต่างไปยังตำแหน่งต่อไป เพื่อทำการพิจารณาค่าภายในพื้นที่ถัดไป ซึ่งการเลื่อนหน้าต่างนี้จะไม่มีการทับซ้อนกันของหน้าต่างที่เคยเลื่อนผ่านมาแล้ว ทำให้ผลลัพธ์ที่ได้มีขนาดน้อยลง ทั้งนี้กระบวนการ Pooling จะกระทำกับทุก ๆ ภาพที่ถูกส่งมาจากกระบวนการ Relu โดยการทำ Pooling ในชั้นแรกขนาดของ Feature maps จะถูกลดลงเหลือ 16x16 และเมื่อผ่านชั้นที่สองขนาดจะลดลงเหลือ 8x8

ชั้นที่ 4-6) Fully Connected เป็นชั้นที่มีการดำเนินการในรูปแบบของโครงข่ายประสาทเทียม (Neural network) โดยชั้นที่ 4 เป็นชั้นที่จะนำข้อมูลที่ผ่านการดำเนินการในชั้นที่ 3 ที่มีขนาด 8x8x64 มาจัดเรียงหรือเปลี่ยนรูปแบบใหม่ ที่จะมีการเปลี่ยนขนาดเป็น 4096x1 ซึ่งข้อมูลที่ได้คือคุณลักษณะเด่นของภาพตัวอักขระ และถูกส่งผ่านต่อไปยังชั้นที่ 5 และชั้นที่6 ที่เป็นชั้น Hidden Layer ที่มีจำนวนโหนด (Node) อยู่ในชั้นที่ 5 จำนวน 128 โหนด และชั้นที่ 6 มีจำนวนโหนดอยู่ที่ 255 โหนด

ชั้นที่ 7) Output เป็นชั้นผลลัพธ์หรือคำตอบ หลักจากโครงข่ายประสาทเทียมได้มีการคำนวณเสร็จแล้ว ผลที่ได้จะอยู่ในรูปแบบของคลาส (Class) ของตัวอักขระที่มีจำนวน 59 คลาส

4.4 การทดลองและผลลัพธ์ (Experiment and Results)

1. ข้อมูลรูปภาพ (Image Data) ข้อมูลรูปภาพตัวอักขระที่ใช้สำหรับการจำแนกประเภทตัวอักขระภาษาไทย ประกอบไปด้วยรูปภาพตัวอักขระภาษาไทย (หนึ่งภาพต่อหนึ่งตัวอักขระ) ทั้งหมดจำนวน 2,215 ภาพ และรูปภาพจำนวน 355 ภาพสำหรับใช้เป็นชุดทดสอบ โดยแบ่งกลุ่มตัวอักขระออกเป็น 59 กลุ่ม ดังตาราง 5

ตาราง 5 แสดงกลุ่มตัวอักขระและจำนวนในแต่ละกลุ่ม

ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน
ก	100	ณ	22	ร	129	ไม้หันอากาศ	66

ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน	ตัวอักขระ	จำนวน
ข	35	ด	40	ล	50	สระเอ	55
ค	49	ต	36	ว	59	สระแอ	21
ฅ	3	ถ	13	ศ	39	สระโอ	19
ง	83	ท	41	ษ	13	สระอะ	42
จ	30	ธ	25	ส	68	สระอา	156
ฉ	3	น	118	ห	59	สระอำ	17
ช	44	บ	56	อ	88	สระอิ	72
ซ	6	ป	43	ไม้เอก	32	สระอี	46
ญ	7	ผ	3	ไม้โท	49	สระอึ	8
ฎ	2	ฝ	2	ไม้ตรี	2	สระอื	14
ฏ	8	พ	38	ไม้จัตวา	2	สระอุ	46
ฐ	4	ภ	14	ไม้ไต่คู้	7	สระอู	21
ฑ	3	ม	68	ไม้ม้วน	6	การันต์	37
ฒ	5	ย	79	ไม้มลาย	12		

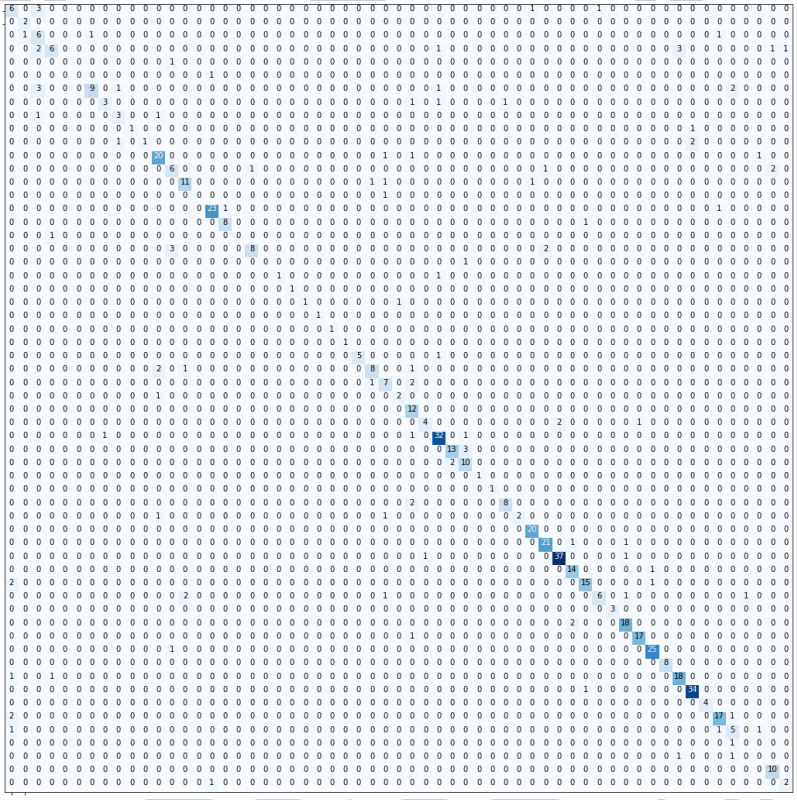
2. การทดลอง (Experiments) การทดลองมีการเปรียบเทียบวิธีการ 2 วิธีคือ วิธีการแรกเป็นการสกัดคุณลักษณะ 2 วิธีได้แก่ 1) Chain code, Zoning, Histogram Projection และ 2) Histograms of Oriented Gradients (HOG) ร่วมกับเทคนิคการจำแนกประเภทด้วย ซัพพอร์ตเวคเตอร์แมชชีน (SVM) และวิธีการที่สองเป็นการเรียนรู้เชิงลึก (Deep Learning) ด้วยวิธีการ Convolution Neural Network (CNN) โดยผลทดลองวิธีการที่ได้ดังตาราง 6

ตาราง 6 ผลของการจำแนกอักขระ

	Accuracy
Chain code, Zoning, Histogram Projection + SVM	56%
HOG + SVM	76%
CNN	**82%**

จากผลการทดลองพบว่า วิธีการสกัดคุณลักษณะด้วย Chain code, Zoning, Histogram Projection ร่วมกับการจำแนกประเภทด้วยซัพพอร์ตเวคเตอร์แมชชีน (SVM) มีประสิทธิภาพคิดเป็นร้อยละ 56% วิธีการสกัดคุณลักษณะเด่นด้วย Histograms of Oriented Gradients (HOG) ร่วมกับการจำแนกประเภทด้วยซัพพอร์ตเวคเตอร์แมชชีน มีประสิทธิภาพคิดเป็นร้อยละ 76% เมื่อพิจารณาวิธีการทั้งสองยังมีประสิทธิภาพในการจำแนกน้อยกว่าวิธีการ Convolution Neural Network (CNN) ที่มีประสิทธิภาพคิดเป็นร้อยละ 82%

ขั้นตอน CNN ทำการสกัดเอกลักษณ์แบบนามธรรม (Abstract Feature) ซึ่งจะขึ้นอยู่กับข้อมูลที่ทำการอินพุตเข้าไปในเครือข่าย จะเห็นได้ว่าประสิทธิภาพของเทคนิคมักจะขึ้นกับรูปแบบของข้อมูลที่ทำการอินพุตเข้าไปในเครือข่าย



รูปที่ 165 แสดง confusion matrix ของการรู้จำอักขระในภาพ

จากภาพที่ 165 แสดง confusion matrix ของการรู้จำอักขระในภาพ ซึ่งพบว่าอักขระที่มีการตรวจจับผิดพลาดนั้นได้แก ไม้ตรี ไม้จัตรา ต ฉ ช สระอึ และสระอี สาเหตุของความผิดพลาดเกิดจาก

1. จำนวนของอักขระในกลุ่มข้อมูลฝึกฝนนั้นน้อย ตัวอย่างเช่น ไม้จัตวา และด้วยลักษณะของอักขระนั้นมีรูปร่างที่คล้ายกับ ไม้เอก (ซึ่งมีจำนวนอักขระใช้ชุดฝึกฝนนั้นจำนวนมาก) ทำให้มันจัตวาถูกจำแนกเป็นไม้เอก

2. รูปร่างของอักขระที่มีความคล้ายคลึงกันเมื่อขนาดของอักขระนั้นมีขนาดเล็ก เช่น สระอึ และสระอี ทำให้ผลการจำแนกอักขระประเภทดังกล่าวไม่ดีนัก

กาวัดประสิทธิภาพการรู้จำข้อความจะประเมินด้วยเทคนิค Larvenstine Distance ซึ่งเทคนิคนี้จะประวัดการจากการรับ Operation ที่จะใช้ในการเปลี่ยนจากข้อความ s ไปเป็นข้อความ t

อัลกรึทึม Lavenstine Distance
1. กำหนดให้ n เป็นความยาวของข้อความ s 2. กำหนดให้ m .เป็นความยาวของข้อความ 3. ทำการสร้าง matrix ขนาด mxn 4. กำหนดค่าให้แถวแรกของ matrix มีค่า 0 ถึง n และกำหนดค่าคอลัมน์แรกมีค่า 0 ถึง m 5. ตรวจสอบอักขระในข้อความ s และ t 5.1 ถ้า s[i] เท่ากับ t[j] จะกำหนดค่าให้เป็น 0 5.2 ถ้า s[i] ไม่เท่ากับ t[j] จะกำหนดค่าให้เป็น 1 6. กำหนดค่าให้ cell d[i,j] ใน matrix มีค่าที่น้อยที่สุดจาก 6.1 เซลที่อยู่ด้านบน + 1 : d[i-1,j]+1 6.2 เซลที่อยู่ด้านซ้าย +1 : d[i, j+1]+1 6.3 เซลที่อยู่แนวทแยง + d[i,j] : d[i+1, j+1]+d[i,j] 7. เมื่อทำกระบวนการที่ 5 และ 6 เสร็จสิ้นแล้ว distance ที่ได้จะอยู่ที่ช่อง d[n,m]

จากอัลกอรึทึมข้างต้น จะกำหนดให้ s เป็นข้อความที่ได้จากการทำนาย และ t เป็นข้อความของผลเฉลย (Ground Truth) พบว่าอัตราข้อผิดพลาด (Character Error Rate: CRE) อยู่ที่ 0.27 ตัวอย่างของผลลัพธ์แสดงดังภาพที่ 166

ภาพ	ข้อความ	ผลการทำนาย
		เลพาะเจ้าหทาที
		ก๋วยจ๊บพีพีง
		อิ่มอร่อย ห บาท
		บายยา
		โดย อ เกสัชกร

รูปที่ 166 แสดงผลลัพธ์การตรวจจับแล้วทำนายผลข้อความ

บทที่ 5
สรุปผลการดำเนินงาน

ปริญญานิพนธ์ฉบับนี้นำเสนอวิธีการในการตรวจจับและรู้จำข้อความในภาพป้ายโฆษณา โดยปริญญานิพนธ์นี้ได้ทำการแบ่งการประมวลประมวลผลออกเป็น 2 ส่วนด้วยกัน ได้แก่ การตรวจจับข้อความในภาพ และการรู้จำข้อความ ในการตรวจจับข้อความในภาพนั้นจะอาศัยข้อมูลก่อนหน้าในการแผนที่ความน่าจะเป็นของข้อความในหลาย ๆ ระดับ เพื่อทำการระบุตำแหน่งข้อความในรูปแบบของ Window-based Technique สำหรับการรู้จำข้อความในภาพจะอาศัยวิธีการของ Character-based Technique ในการสกัดเอกลักษณ์จากตัวอักขระโดยวิธีการอาศัยรูปร่าง และการใช้การสลัดข้อ Latent ที่อยู่ในตัวอักขระ ก่อนที่จะมีการจำแนกอักขระออกเป็นประเภทอักขระ และนำมารวมกันตามรูปแบบการเรียงของอักขระในภาพ เพื่อสร้างเป็นข้อความ

ในบทนี้จะทำการอภิปรายเทคนิคที่ได้นำเสนอในปริญญานิพนธ์นี้รวมถึงรายงานผลการดำเนินการวิจัย โดยจะแบ่งหัวข้อออกเป็นดังต่อไปนี้ 5.1 การตรวจจับข้อความในภาพ 5.2 การรู้จำข้อความในภาพ 5.3 สรุปผลการดำเนินงาน 5.4 ประโยชน์ที่ได้รับจากการวิจัย (Thesis Contributions) และ ข้อเสนอแนะและการดำเนินงานในอนาคต

5.1 การตรวจจับข้อความในภาพ

การรู้จำข้อความในภาพถือได้ว่าเป็นส่วนสำคัญในในการพัฒนาแอฟพลิเคชันทางด้านคอมพิวเตอร์วิทัศน์ ที่สามารถนำไปประยุกต์ใช้งานในแอฟพลิเคชันต่าง ๆ ตัวอย่างเช่น แอฟพลิเคชันบนมือถือ เป็นต้น อย่างที่ได้กล่าวข้างต้นปัญหาและความท้าทายของการตรวจจับข้อความในภาพคือ ภาพจะมีความหลากหลายทางการกายภาพ ไม่ว่าจะเป็น ขนาด สี ตำแหน่ง แสดง ของข้อความในภาพ ด้วยเหตุนี้จะมีการพัฒนาเทคนิควิธีการในการตรวจจับข้อความในภาพด้วยเทคนิคต่าง ๆ ได้แก่

เทคนิคการทำ Intensity Thresholding

เทคนิคการใช้ Distance Transform

เทคนิค Stroke Width Transform (SWT)

เทคนิค Maximally Stable Extremal Regions (MSER) เป็นต้น

เทคนิคดังกล่าวนั้นจะอาศัยมูลจากพิกเซลในการต่อเชื่อมกันเพื่อสร้างเป็นกลุ่มพิกเซลที่เป็นตัวอักขระ ข้อความในภาพนั้นมีความหลากหลายและหากพิจารณาในระดับพิกเซลแล้วจะเห็นว่า พิกเซลที่เป็นอักขระและพิกเซลที่ไม่ใช่อักขระนั้นมีคุณสมบัติที่ใกล้เคียงกัน จึงเป็นสาเหตุหลักที่ทำให้การตรวจจับ

ข้อความในภาพด้วยเทคนิคดังกล่าวนี้ให้ผลลัพธ์ที่ไม่ดีเท่าที่ควร ดังนั้นปริญญานิพนธ์นี้จึงได้นำเสนอเทคนิควิธีในการตรวจจับตำแหน่งของข้อความในภาพ โดยนำเสนอวิธีการ 2 วิธีการด้วยกัน คือ 1) Adaptive Maximally Stable Extremal Regions (AMSER) และ 2) การใช้ข้อมูลก่อนหน้าในการระบุตำแหน่งข้อความ (Prior Information-based Technique) โดยเทคนิค AMSER จะทำการหาค่าที่เหมาะสมที่สุดของ (Optimize) ค่าเทรชโฮลด์ต่ำสุดและมาสุด ด้วยเทคนิค Graph-Cut สำหรับเทคนิค Prior Information-based จะทำการสร้างแผนที่ความน่าจะเป็น (Probability Maps) จากภาพในหลาย ๆ ระดับ เพื่อเป็นข้อมูลในการใช้เพื่อพิจารณาตำแหน่งของความในภาพ โดยแผนที่ความน่าจะเป็นนี้จะทำสร้างจากชุดข้อมูลก่อนหน้า (Prior information) ในการประมาณตำแหน่งของข้อความในภาพ ภาพจะถูกแบ่งออกเป็น Window ที่ไม่ซ้อนทับกัน จากนั้นในแต่ละ Window จะถูกนำมาประมวลผลเพื่อคำนวณหาความน่าจะเป็นที่มีส่วนของข้อความอยู่ใน Window นั้น นอกจากนั้น Window ที่ใกล้เคียงกันและมีคุณสมบัติในแบบเดียวกันก็จะถูกนำมารวมกัน (Window Merging) โดยใช้เทคนิค Markov Random Walk (MRW) และ Graph-Cut เพื่อสร้างเป็นพื้นที่ข้อความในภาพ

5.2 การรู้จำข้อความในภาพ

เมื่อทำการตรวจจับข้อความในภาพได้แล้ว ขั้นตอนต่อไปคือการนำข้อความที่ตรวจจับได้นั้นมาทำการรู้จำ โดยปริญญานิพนธ์นี้ได้ทำการนำเสนอวิธีการในรู้จำข้อมูลอักขระเพื่อสร้างเป็นข้อความ 2 วิธีด้วยกัน (เพื่อใช้ในการเปรียบเทียบประสิทธิภาพ) คือ 1) การใช้เอกลักษณ์เชิงรูปร่างของอักขระ และ 2) การใช้เอกลักษณ์แบบ Latent Feature เมื่อทำการตรวจจับข้อความในภาพแล้ว ข้อความจะถูกนำมาพิจารณาการเรียบแนว (Layout) เพื่อทำการหาลำดับของอักขระที่ทำการรู้จำ จากนั้นจะทำการแยกอักขระในภาพออกเป็นทีละตัว และนำไปสกัดเอกลักษณ์ สำหรับการสกัดเอกลักษณ์เชิงรูปร่างนั้นจะใช้เทคนิค

Histogram of Oriented Gradients

Chain Code

Moments

เพื่อทำการสร้างเวกเตอร์เอกลักษณ์ และทำไปทำการทำนายกลุ่มต่อไป สำหรับการสกัดเอกลักษณ์ Latent จากอักขระนั้นจะใช้เทคนิคของ Convolution Neural Network ในการสร้าง Network 2 Network แบบขนานเพื่อใช้ในการเปรียบเทียบระหว่างอักขระปกติ และอักขระที่มีการเพิ่มส่วนข้อมูล

ด้วยการปรับรูปร่าง และ ขนาด (Augmentation) จากนั้นจะทำการหาค่าที่เหมาะสมที่สุดใน Network เพื่อนำน้ำหนักใน Network นั้นมาสร้างเป็นเอกลักษณ์ของภาพ

5.3 สรุปผลการดำเนินงาน

การทดลองได้ทำการทดลองโดยแบ่งออกเป็น 2 ส่วนการทดลอง จาดภาพป้ายข้อความที่มีการเก็บในหลาย ๆ รูปแบบทั้งหมด 1,200 ภาพ

ในการตรวจจับภาพข้อความในภาพนั้น ภาพผลเฉลยจะทำการเตรียมด้วยอุปกรณ์ช่วยโดยจะทำการระบุพื้นที่เป็นข้อความด้วยตำแหน่งข้องจุดเริ่มต้นของข้อความในภาพด้วยค่า x และ y ในระบบ Cartesian และความกว้าง ความสูงของข้อความในภาพ การวัดประสิทธิภาพนั้นจะพิจารณาจากอัตราของพื้นที่ซ้อนทับ จากการวัดประสิทธิภาพพบว่าวิธีการใช้ข้อมูลก่อนหน้า (Prior Information) ให้ผลลัพธ์ที่ดีที่สุด (รายละเอียดในบทที่ 3)

ในการรู้จำข้อความนั้นจะทำการแบ่งการทดลองและประเมินผลออกเป็น 2 ส่วน ได้แก่ 1) การรู้จำประเภทของอักขระ และ 2) การรู้จำข้อความในภาพ ในการรู้จำอักขระในภาพพบว่าการใช้ Latent Features ให้ประสิทธิภาพที่ดีกว่าการใช้ เอกลักษณ์เชิงรูปร่าง และการรู้จำข้อความนั้นให้ประสิทธิภาพโดยพิจารณาจากเทคนิค Lavenstine ซึ่งให้ค่า Recognition Loss Rate อยู่ที่ 32.12%

5.4 ประโยชน์ที่ได้รับจากการวิจัย (Thesis Contributions)

1. การพัฒนาเทคนิคในการตรวจจับข้อความในภาพด้วยวิธีการ AMSER ที่มีการประยุกต์ใช้เทคนิค DE มาช่วยในการหาค่าที่เหมาะสมที่สุดของพารามิเตอร์

2. การพัฒนาเทคนิคในการตรวจข้อความในภาพด้วยการใช้ข้อมูลก่อนหน้า (Prior Information) มาช่วยในการหาตำแหน่งของข้อความในภาพ

3. การพัฒนาเทคนิคในการรู้จำข้อความจากเอกลักษณ์ Latent Features จากข้อมูลตัวอักขระ

5.5 ข้อเสนอแนะและการดำเนินงานในอนาคต

สำหรับข้อเสนอแนะและการพัฒนาในอนาคตมีดังต่อไปนี้

1. ข้อมูลอาจจะต้องมีการทำให้มีมาตรฐานมากกว่านี้ เนื่องจากข้อมูลบางส่วนมีการเก็บจากระบบอินเทอร์เน็ตดังนั้นคุณภาพของภาพนั้นอาจจะไม่เท่ากัน เมื่อเปรียบเทียบจากการเก็บด้วยกล้องดิจิทัล ซึ่งจะมีผลต่อการประมวลอย่างมาก

2. การตรวจจับตำแหน่งของข้อความในภาพอาจจะมีการพัฒนาต่อยอดโดยการนำเอาเทคนิคทางด้าน Deep Learning มาใช้ในการประมาณตำแหน่งของข้อความในภาพ

3. ในการรู้จำข้อความภาพ การใช้ Character-based เทคนิคนั้นจำเป็นอย่างยิ่งที่จะได้ข้อมูลจากการแยกอักขระจากภาพที่ดี ดังนั้นผลลัพธ์ของวิธีการดังกล่าวนอกจากจะขึ้นกับเอกลักษณ์ที่จะใช้ในการรู้จำอักขระแล้ว ยังต้องอาศัยผลของการแยกอักขระในภาพที่ดีด้วย แนวทางในการพัฒนานั้นอาจจะใช้เทคนิคของ Deep Learning ในการเรียนรู้แบบ Image-based ที่จะพิจารณาถึงการเกิดขึ้นของส่วนภาพที่มีผลต่อเนื่องกัน เพื่อนำไปทำนายข้อความในภาพต่อไป

บรรณานุกรม

บรรณานุกรม

1. Judd, T., et al., *Learning to predict where humans look*, in *IEEE 12th International Conference on Computer Vision (ICCV)*. 2009. p. 2106–2113.
2. Shi, C., et al., *Scene text recognition using part-based tree-structured character detection.* Proc. IEEE CVPR, 2013: p. 2961–2968.
3. นันทิยา พูลสวัสดิ์, การพัฒนาระบบสารสนเทศ เพื่อแนะนำสินค้าอิเล็กทรอนิกส์ โดยใช้เทคนิคเคมีนร่วมกับขั้นตอนวิธีการหาเพื่อนบ้านที่ใกล้ที่สุด. 2554, ปัญหาพิเศษวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิชาระบบสารสนเทศเพื่อการจัดการ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ.
4. นเรศ ผ่องสวัสดิ์กุล and จีรพร วีระพันธุ์, อัลกอริทึมเคมีนคลัสเตอร์ริงแบบขนานที่มีประสิทธิภาพบนระบบคลัสเตอร์. 2552, ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
5. อุไร ทองหัวไผ่, ระบบค้นคืนสารสนเทศ. *Information Retrieval*. ภาควิชาวิทยาการคอมพิวเตอร์. คณะวิทยาศาสตร์. มหาวิทยาลัยรามคำแหง. .
6. เชาวน์ ปอแก้ว, การพัฒนาระบบรู้จำอักษรล้านนาโดยอาศัยเคเนียร์เรสเนเบอร์. 2555, วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่.
7. Vladimir N and Vapnik, *An Overview of Statistical Learning Theory*, in *IEEE TRANSACTIONS ON NEURAL NETWORKS*. 1999.
8. Cristianini, N. and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. 2000, Cambridge university press.
9. สุพจน์ จันทร์วิพัฒน์, การแก้ปัญหาการจัดกลุ่มข้อมูลด้วยวิธีซัพพอร์ตเวคเตอร์แมชชีน และตัวอย่างการประยุกต์ใช้งาน. วารสารวิชาการเทคโนโลยีอุตสาหกรรม ปีที่ 7, 2554.
10. Burges and C. JC, *A Tutorial on Support Vector Machines for Pattern Recognition*, in *Data Mining and Knowledge Discovery*. 1998. p. 121-167.

11. Green, B. *Canny Edge Detection Tutorial*. 2014 [cited 1 February 2014]; Available from: http://dasl.mem.drexel.edu/alumni/bGreen/www.pages.drexel.edu/_weg22/can_tut.html.
12. *Neural Network*. 2014 [cited 1 February 2014]; Available from: http://4.bp.blogspot.com/.
13. สมหญิง พรหมเจริญ and ยุทธพงษ์ รังสรรค์เสรี, การจำแนกประเภทข้อมูลภาพถ่ายดาวเทียม *JERS*-1 ระบบ *OPS* โดยใช้โครงข่ายประสาทเทียม, in การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 37.
14. Rumelhart, D.E. and J.L. McClelland, *Parallel distributed processing*, in *Psychological and biological models* 2 1987.
15. Benediktsson, J.A. and P.H. Swain, *Neural network approaches versus statistical methods in classification of multisource remote sensing data*, in *IEEE Transactions on geoscience and remote sensing* 28.4 1990. p. 540-552.
16. เทคโนโลยีการประมวลผลภาพ (*Image processing*). 2559 [cited 6 มีนาคม 2559]; Available from: https://silllovely.wordpress.com/2013/06/11/เทคโนโลยีการประมวลผลภาพ.
17. สมเกียรติ อุดมหรรษากุล, *Fundamentals of Digital Image Processing*. 2554, กรุงเทพ: ท๊อป 208.
18. บุญธรรม ภัทราจารุกุล, การประมวลผลภาพดิจิตอลเบื้องต้น (*Fundamentals of digital Image Processing*). 2556, กรุงเทพ :ซีเอ็ดยูเคชั่น. 384.
19. ยุทธพงษ์ รังสรรค์เสรี and พรพรรณ์ ดุลยกาญจน์, การจำแนกข้อมูลภาพโดยการทำเธรโชลด์หลายระดับของค่าทางสถิติ อันดับที่สองของระดับสีเทา, in การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 38.
20. *Otsu's method*. 2014 [cited 1 February 2014]; Available from: http://en.wikipedia.org/wiki/Otsu's_method.
21. ชาตรี กอบัวแก้ว, การจำแนกพระผงโดยการเปรียบเทียบลักษณะพิเศษ. 2550, วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปกร
22. วิทยากร อัศดรวิเศษ และคณะ, การประยุกต์ใช้ *Matlab*. 2555, กรุงเทพ :บริษัทแอคทีฟพริ้นท์ จำกัด.

23. Bhardwaj, S. and A. Mittal, *A Survey on Various Edge Detector Techniques*, in *Procedia Technology*. 2012. p. 220-226.

24. *Canny edge detector*. 2014 1 February 2014]; Available from: http://en.wikipedia.org/wiki/Canny_edge_detector.

25. ธนารมย์ สายเชื้อ, การตรวจจับและวิเคราะห์ต้นปาล์มจากภาพถ่ายทางอากาศ. 2555, สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

26. *Connected-component_labeling*. 2014 1 February 2014]; Available from: http://en.wikipedia.org/wiki/Connected-component_labeling.

27. คอมพิวเตอร์วิทัศน์. 2559 5 มีนาคม 2559]; Available from: http://th.wikipedia.org/wiki/คอมพิวเตอร์วิทัศน์.

28. การกรองข้อมูลภาพ. 2559 5 มีนาคม 2559]; Available from: http://www.getdd.net/basiccom/51-imageprocessing.html.

29. ชาติชาย เกตุทับทิม, วิธีการตรวจสอบและจำแนกความบกพร่องของผ้าโดยใช้กาบอร์ฟิลเตอร์และระบบการปรับตัวการเรียนรู้นิวรอลฟัซซี. 2554, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

30. วิไลลักษณ์ คิดสร้าง, ตัวกรองกาบอร์ประยุกต์เพื่อการตรวจจับจุดบกพร่องในสิ่งทอ. 2553, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีสุรนารี.

31. Ezaki, N., M. Bulacu, and L. Schomaker, *Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons*, in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on IEEE 2004.*

32. Liu, C., C. Wang, and R. Dai, *Text Detection in Images Based on Unsupervised Classification of Edge-based Features*, in *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on. IEEE 2005.*

33. Phan, T.Q., P. Shivakumara, and C.L. Tan, *A Laplacian Method for Video Text Detection*, in *10th International Conference on Document Analysis and Recognition*. 2009.

34. Epshtein, B., E. Ofek, and Y. Wexler, *Detecting Text in Natural Scenes with Stroke Width Transform*, in *Computer Vision and Pattern Recognition (CVPR)*. 2010: IEEE Conference on. IEEE. p. 2963-2970.

35. Subramanian, K., et al., *Character-stroke detection for text-localization and extraction*, in *Document Analysis and Recognition*. 2007: Ninth International Conference on IEEE.

36. Jung, C., Q. Liu, and J. Kim, *A stroke filter and its application to text localization*, in *Pattern Recognition Letters* 30.2 2009. p. 114-122.

37. โสภณ ผู้มีจรรยา, *MLRES Active Contour for Image Segmentation with Multiple Objects*, in *The 5th NPRU National Academic Conference* 2013 2013.

38. Sum, K.W. and P.Y.S. Cheung, *Boundary vector field for parametric active contours*, in *Pattern Recognition*. 2007. p. pp.1635-1645.

39. Xie, X. and M. Mirmehdi, *MAC: Magnetostatic Active Contour Model*, in *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*. 2008. p. pp.632-647.

40. Wang, T., I. Cheng, and A. Basu, *Fluid Vector Flow and Applications in Brain Tumor Segmentation*, in *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*. 2009. p. pp.781-789.

41. Li, B. and S.T. Acton, *Active Contour External Force Using Vector Field Convolution for Image Segmentation*, in *IEEE Transactions on Image Processing*. 2007. p. pp. 2096-2106.

42. Chan, T.F. and L.A. Vese, *Active Contours Without Edges*, in *IEEE Transactions on Image Processing*. 2001. p. pp. 266-277.

43. Anthony Yezzi, J., A. Tsai, and A. Willsky, *A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations.* Journal of Visual Communication and Image Representation 13, 2002: p. pp.195-216.

44. Michailovich, O., Y. Rathi, and A. Tannenbaum, *Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow*, in *IEEE Transactions on Image Processing*. 2007. p. pp.2787-2801.

45. MILLE, J. and L.D. COHEN, *A local normal-based region term for active contours.* Energy Minimization Methods in Computer Vision and Pattern Recognition, 2009.

46. Ronfard, R., *Region-based strategies for active contour models.* International Journal of Computer Vision 1994. 13(2): p. pp.229-251.

47. Lankton, S. and A. Tannenbaum, *Localizing region-based active contours*, in *IEEE Transactions on Image Processing*. 2008. p. pp.2029-2039.

48. โสภณ ผู้มีจรรยา, แอ๊กทิฟคอนเท้วร์แบบใช้ข้อสนเทศบริเวณท้องถิ่นบนเส้นค้นหาที่ยืดได้สำหรับการแบ่งส่วนภาพ. 2552, วิทยานิพนธ์วิศวกรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

49. Kass, M., A. Witkin, and D. Terzopoulos, *Snakes: Active Contour Models*, in *International Journal of Computer Vision*. 1988. p. pp.321-331.

50. ศราวุธ แต้โอสถ, การแบ่งส่วนภาพโดยวิธีเลเวลเซตร่วมกับความรู้เชิงรูปร่าง. 2549, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

51. Cohen, L.D., *On active contour models and balloons*, in *CVGIP: Image understanding*. 1991. p. pp.211-218.

52. Malladi, R., J. Sethian, and B.C. Vemuri, *Shape modeling with front propagation: A level set approach*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1995. p. pp.158-175.

53. Caselles, V., R. Kimmel, and G. Sapiro, *Geodesic Active Contours*, in *International Journal of Computer Vision*. 1997. p. pp.61-79.

54. Zhang, Q. and R. Pless, *Segmenting multiple familiar objects under mutual occlusion*, in *IEEE International Conference on Image Processing*. 2006. p. pp. 197-200.

55. Osher, S. and J.A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations.* Journal of computational physics, 1988. 79: p. 12-49.

56. Cui, X., et al., *An improved image segmentation algorithm based on the watershed transform*, in *Information Technology and Artificial Intelligence Conference (ITAIC)*. 2014: 2014 IEEE 7th Joint International. p. pp. 428-431.

57. Pinto, T.W., et al., *Image segmentation through combined methods: Watershed transform, unsupervised distance learning and Normalized Cut*, in *Image Analysis and Interpretation (SSIAI)*. 2014: 2014 IEEE Southwest Symposium on. p. pp. 153-156.

58. Amankwah, A. and C. Aldrich, *Automatic estimation of bubble size distributions in flotation froths by use of a mean shift algorithm and watershed transforms*, in *Geoscience and Remote Sensing Symposium (IGARSS)*. 2014: 2014 IEEE International. p. pp. 1608-1611.

59. Vincent, L. and P. Soille, *Watersheds in digital spaces: an efficient algorithm based on immersion simulations*, in *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6. 1991. p. pp.583-598.

60. ชัยพร ปานยินดี, การตรวจหาก้อนเนื้องอกในปอดจากภาพ *PET/CT* โดยใช้การแปลงวอเตอร์เชด และการจับคู่ตำแหน่งภายในภาพ. 2551, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

61. พรพจน์ โพธิ์พงศ์วิวัฒน์, การแบ่งส่วนภาพโดยทฤษฎีกราฟของภาพจากการต่อแบบหลายความละเอียด. 2545, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

62. Morris, O.J., M.d.J. Lee, and A.G. Constantinides, *Graph theory for image analysis: an approach based on the shortest spanning tree*, in *Communications, Radar and Signal Processing, IEE Proceedings* 1986. p. 146-152.

63. Morris, O.J., M.d.J. Lee, and A.G. Constantinides, *A unified method for segmentation and edge detection using graph theory*, in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86*. 1986.

64. ศิรวิช ผสมกุศลศีล, การแบ่งส่วนภาพโดยใช้ทฤษฎีกราฟบนภาพ ที่ผ่านการทำให้ราบเรียบและรักษาของภาพ. 2545, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

65. Srisuk, S., et al., *A new shape matching measure for nonlinear distorted object recognition*, in *Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T.(Eds.).* 2003.

66. Belongie, S., J. Malik, and J. Puzicha, *Shape Matching and Object Recognition Using Shape Contexts*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002. p. pp.509-522.

67. Wiengsamotha, S., *Car Detection*, in *J Sci Technol MSU*. 2013. p. pp.725-732.

68. Siriboon, K., A. Jirayusakul, and B. Kruatrachue. *HMM Topology Selection for On-line Thai Handwritten Recognition*. in *Proceedings of the First International Symposium on Cyber Worlds (CWi02)* 2002.

69. Panggabean, M. and L.A. Rønningen, *Character recognition of the Batak Toba alphabet using signatures and simplified chain code*, in *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on*. 2009. p. pp.215 - 220.

70. Siddiqi, I. and N. Vincent, *A Set of Chain Code Based Features for Writer Recognition*, in *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*. 2009: Barcelona. p. pp.981 - 985.

71. Phuangsuwan, P. and T. Boriboon, *Hand Written Thai Characters Recognition for Mobile Phone*, in *The Tenth National Conference on Computing and Information Technology* 2014. p. pp.491-496.

72. Untao, A. and S. Valuvanathon, *Thai Hand Shape Recognition Using HOG - PCA and SVM*, in *The Tenth National Conference on Computing and Information Technology*. 2014.

73. Chatklaw Jareanpon, *INTRODUCTION TO DATA MINING*. 2013, Computer Science, Informatics, Mahasarakham University.

74. จักรภัทร แก้วทอง and ไตรปิฎก อินทสุวรรณ, โปรแกรมตรวจจับวัตถุและข้อความบนป้ายโฆษณา. 2560, ปริญญานิพนธ์วิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

75. wbi.msu.ac.th. การวัดประสิทธิภาพ. 2559 7 มีนาคม 2559]; Available from: www.wbi.msu.ac.th/file/648/doc_31.ppt.

76. เอกสิทธิ์ พัชรวงศ์ศักดา, การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น. 2557, กรุงเทพ:บริษัท เอเซีย ดิจิตอลการพิมพ์ จำกัด. 123.

77. Chen, T.B., D. Ghosh, and S. Ranganath, *Video-text extraction and recognition*, in *TENCON* 2004. 2004 *IEEE Region* 10 *Conference*. 2004 IEEE. p. 319-322.

78. Yi, J., Y. Peng, and J. Xiao, *Color-based clustering for text detection and extraction in image*, in *Proceedings of the* 15*th international conference on Multimedia*. 2007: ACM.

79. Lyu, M.R., J. Song, and M. Cai, *A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction*, in *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*. 2005.

80. ANOUAL, H., S. ELFKIHI, and A. JILBAB, *Features Extraction for Text Detection and Localization*, in *I/V Communications and Mobile Network (ISVC),* 2010 5*th International Symposium on. IEEE*. 2010.

81. Ma, L., C. Wang, and B. Xiao, *Text Detection in Natural Images Based on Multi-Scale Edge Detetion and Classification*, in 3*rd International Congress on Image and Signal Processing (CISP*2010) 2010.

82. Yi, C. and Y. Tian, *Text String Detection From Natural Scenes by Structure-Based Partition and Grouping*, in *IEEE TRANSACTIONS ON IMAGE PROCESSING*. 2011.

83. Huang, X., K. Liu, and L. Zhu, *Auto Scene Text Detection Based on Edge and Color Features* in *International Conference on Systems and Informatics*. 2012: ICSAI 2012. p. 1882-1886.

84. Kim, K.C., et al. *Scene text extraction in natural scene images using hierarchical feature combining and verification*. in *Pattern Recognition,* 2004. *ICPR* 2004. *Proceedings of the* 17*th International Conference on*. 2004.

85. Wei, Y.C. and C.H. Lin, *A robust video text detection approach using SVM*, in *Expert Systems with Applications*. 2012. p. 10832-10840.

86. Mariano, V.Y. and R. Kasturi. *Locating Uniform-Colored Text in Video Frames*. in *Pattern Recognition,* 2000. *Proceedings.* 15*th International Conference on IEEE* 2000.

87. Sharma, N., et al., *A New Method for Arbitrarily-Oriented Text Detection in Video*, in 10*th IAPR International Workshop on Document Analysis Systems*. 2012.

88. Zhou, J., et al., *A Robust System for Text Extraction in Video*, in *ICMV*. 2007. p. 119-124.

89. Pise, A. and S.D.Ruikar, *Text Detection and Recognition in Natural Scene Images*, in *International Conference on Communication and Signal Processing*. 2014: India.

90. Khurshid, K., et al., *Comparison of Niblack inspired Binarization methods for ancient documents*, in *S&T/SPIE Electronic Imaging. International Society for Optics and Photonics*. 2009.

91. Lee, S.-W., D.-J. Lee, and H.-S. Park, *A new methodology for gray-scale character segmentation and recognition*, in *Pattern Analysis and Machine Intelligence*. 1996: IEEE Transactions on 18. p. 1045-1050.

92. Dongre, V.J. and V.H. Mankar, *Devnagari document segmentation using histogram approach.* International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2011. 1: p. 1109-1247.

93. Xu, D., Z. Peng, and Y. Yong, *An Improved Image Segmentation Method Based on Fast Level Set Combining with CV Model*, in *Engineering and Technology (S-CET)*. 2012: 2012 Spring Congress. p. pp. 1-4.

94. Shi, Y. and W.C. Karl, *A real-time algorithm for the approximation of level-set-based curve evolution*, in *IEEE TRANSACTIONS ON IMAGE PROCESSING*. 2008. p. 645-656.

95. Mumford, D. and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems.* Communications on pure and applied mathematics, 1989. 42: p. 577-685.

96. Tanprasert, C. and S. Sae-Tang. *Thai type style recognition*. in *In Circuits and Systems,* 1999. *ISCAS'*99. *Proceedings of the* 1999 *IEEE International Symposium*. 1999.

97. Sawaki, M., H. Murase, and N. Hagita, *Character Recognition in Bookshelf Images using Context-based Image Templates*, in *Document Analysis and*

Recognition ICDAR'99. 1999: Proceedings of the Fifth International Conference. p. pp. 79-82.

98. Saidane, Z. and C. Garcia, *Automatic scene text recognition using a convolutional neural network*, in *Proceedings of the Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR)*. 2007.

99. Zhou, X.-D., et al., *Online handwritten Japanese character string recognition incorporating geometric context*, in *Document Analysis and Recognition*. 2007: CDAR 2007. Ninth International Conference. p. pp. 48-52.

100. Delakis, M. and C. Garcia, *Text detection with convolutional neural networks*, in *International Conference on Computer Vision Theory and Applications*. 2008 VISAPP (2). p. 290-294.

101. Mammeri, A., E.-H. Khiari, and A. Boukerche, *Road-Sign Text Recognition Architecture for Intelligent Transportation Systems*, in *Vehicular Technology Conference (VTC Fall)*. 2014: 2014 IEEE 80th. p. pp. 1-5.

102. Yuzhe, S., L. Shanzhen, and L. Shaobin, *License Plate Character Recognition Research Based on Shape Context*, in *Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*. 2014: 2014 Fourth International Conference. p. pp. 489-492.

103. วิทยา จิรัฐิติเจริญ, การตรวจจับและตัดแยกตัวอักษรภาษาไทยในฉากธรรมชาติ. 2550, วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

104. รุจิพันธุ์ โกษารัตน์, การรู้จำตัวอักขระภาษาไทยโดยใช้โครงข่ายประสาทเทียม. 2551, วิทยาพิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่.

105. Dalal, N. and B. Triggs, *Histograms of oriented gradients for human detection*, in *Computer Vision and Pattern Recognition. 2005: CVPR 2005. IEEE Computer Society Conference*. 2005. p. 886-893.

106. นัฐธริยา เหล่าประชา, อุมาภรณ์ สายแสงจันทร์, and ร. ซ่ำชอง, การพัฒนาวิธีการตรวจสอบความสมบูรณ์ของรูปปั้นโบราณด้วยหลักการประมวลผลภาพ. KKU Engineering Journal, 2015. 42(2): p. 203-210.

ประวัติผู้เขียน

ชื่อ	เกรียงศักดิ์ รักภักดี
วันเกิด	วันที่ 4 กรกฎาคม พ.ศ. 2520
สถานที่เกิด	อำเภอ เมือง จังหวัด อุบลราชธานี
สถานที่อยู่ปัจจุบัน	11/1 ถนน ราชวงศ์ ตำบล ในเมือง อำเภอ เมือง จังหวัด อุบลราชธานี รหัสไปรษณีย์ 34000
ตำแหน่งหน้าที่การงาน	พนักงานมหาวิทยาลัย สายวิชาการ
สถานที่ทำงานปัจจุบัน	คณะบริหารธุรกิจและการจัดการ มหาวิทยาลัยราชภัฏอุบลราชธานี เลขที่ 2 ถนน ราชธานี ตำบล ในเมือง อำเภอ เมือง จังหวัด อุบลราชธานี รหัสไปรษณีย์ 34000 โทรศัพท์ 0-4535-2000
ประวัติการศึกษา	พ.ศ. 2542 ปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี
	พ.ศ. 2549 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) เทคโนโลยีสารสนเทศเพื่อการเกษตรและพัฒนาชนบท มหาวิทยาลัยอุบลราชธานี
	พ.ศ. 2562 ปริญญาปรัชญาดุษฎีบัณฑิต (ปร.ด.) สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม