

# รายงานการปฏิบัติงานสาขาวิชาศึกษา

เรื่อง

ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width

Transform

**DETECTING TEXT IN MANGA USING STROKE WIDTH  
TRANSFORM**

ปฏิบัติงาน ณ มหาวิทยาลัยออกไกโด

โดย

บุญฤทธิ์ พิริย์ไชยนกุล  
รหัสประจำตัว 58070077

รายงานนี้เป็นส่วนหนึ่งของการศึกษารายวิชา สาขาวิชาศึกษา<sup>1</sup>  
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ  
ภาคเรียนที่ 1 ปีการศึกษา 2561  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

รายงานการปฏิบัติงานสาขาวิชานักศึกษา  
ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width  
Transform

**DETECTING TEXT IN MANGA USING STROKE WIDTH  
TRANSFORM**

ปฏิบัติงาน ณ มหาวิทยาลัยอุบลราชธานี  
โดย

นุญฤทธิ์ พิริย์โยธินกุล  
รหัสประจำตัว 58070077

ปฏิบัติงาน ณ มหาวิทยาลัยอุบลราชธานี  
Hokkaido University Kita 8, Nishi 5, Kita-ku,  
Sapporo, Hokkaido, 060-0808 Japan  
Web site : [www.global.hokudai.ac.jp](http://www.global.hokudai.ac.jp)

# **DETECTING TEXT IN MANGA USING STROKE WIDTH TRANSFORM**

**BOONYARITH PIRIYOTHINKUL**

**A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENT FOR COOPERATING EDUCATION PROGRAM  
THE DEGREE OF BACHELOR OF SCIENCE PROGRAM IN  
INFORMATION TECHNOLOGY  
FACULTY OF INFORMATION TECNOLOGY  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**COPYRIGHT 2018**  
**FACULTY OF INFORMATION TECHNOLOGY**  
**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

วันที่ 10 พฤษภาคม พ.ศ. 2561

เรื่อง ขอส่งรายงานการปฏิบัติงานสหกิจศึกษา  
เรียน รองศาสตราจารย์ ดร. กิตติสุชาต พสุภา<sup>1</sup>  
ที่ปรึกษาสหกิจศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ

ตามที่ข้าพเจ้า บุญฤทธิ์ พิริย์โยธินกุล นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง ได้ปฏิบัติงานสหกิจศึกษาระหว่างวันที่ 23  
กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 พฤษภาคม พ.ศ. 2561 ในตำแหน่ง ผู้ช่วยนักวิจัย ณ สถานประกอบ  
การชื่อ มหาวิทยาลัยหอโกกโภ และได้รับมอบหมายจากผู้ดูแลที่ปรึกษาให้ศึกษาและจัดทำรายงาน  
เรื่อง ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width Transform

บัดนี้ การปฏิบัติงานสหกิจศึกษาได้สิ้นสุดลงแล้ว จึงได้ขอส่งรายงานการปฏิบัติงาน สหกิจศึกษา  
ดังกล่าวมาพร้อมนี้ จำนวน 1 เล่ม เพื่อขอรับคำปรึกษาต่อไป

จึงเรียนมาเพื่อโปรดพิจารณา

ขอแสดงความนับถือ

.....  
(บุญฤทธิ์ พิริย์โยธินกุล)

## กิตติกรรมประกาศ

ตามที่ข้าพเจ้า บุญฤทธิ์ พิริย์ไชยชินกุล ได้มาปฏิบัติงานสหกิจศึกษา ณ มหาวิทยาลัยออกไกโอด ตั้งแต่ วันที่ 23 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 พฤษภาคม พ.ศ. 2561 ทำให้ข้าพเจ้าได้รับความรู้และประสบการณ์ต่าง ๆ ที่มีคุณค่ามาก many สำหรับรายงานสหกิจศึกษาฉบับนี้ สำเร็จลงได้ด้วยดี จากความช่วยเหลือและความร่วมมือสนับสนุนของหลายฝ่าย ดังนี้

1. Professor Dr. Masanori Sugimoto ตำแหน่ง ศาสตราจารย์
2. Jiang Ye ตำแหน่ง นักศึกษาปริญญาเอก ปี 2

นอกจากนี้ยังมีบุคคลท่านอื่น ๆ อีกที่ไม่ได้กล่าวไว้ ณ ที่นี่ ซึ่งให้ความกรุณาแนะนำในจัดทำรายงานสหกิจศึกษาฉบับนี้ ข้าพเจ้าจึงขอขอบพระคุณทุกท่านที่ได้มีส่วนร่วมในการให้ข้อมูลและให้ความเข้าใจเกี่ยวกับชีวิตของการปฏิบัติงาน รวมถึงเป็นที่ปรึกษาในการจัดทำรายงานฉบับนี้จนเสร็จสมบูรณ์

บุญฤทธิ์ พิริย์ไชยชินกุล

ผู้จัดทำรายงาน

วันที่ 10 พฤษภาคม พ.ศ. 2561

## ชื่อรายงานการปฏิบัติงานสาหกิจศึกษา ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke

Width Transform

ผู้รายงาน

บุญฤทธิ์ พิริย์ไบชินกุล

คณะ

เทคโนโลยีสารสนเทศ

สาขาวิชา

เทคโนโลยีสารสนเทศ

(รองศาสตราจารย์ ดร. กิตติสุชาต พสุภา)

อาจารย์ที่ปรึกษาสาหกิจศึกษา

(Professor Dr. Masanori Sugimoto)

พนักงานที่ปรึกษา

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
อนุมัติให้นบรายงานการปฏิบัติงานสาหกิจศึกษาฉบับนี้ เป็นส่วนหนึ่งของการศึกษา  
ตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

<b>ชื่อรายงาน</b>	ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width Transform
<b>ชื่อนักศึกษา</b>	บุญฤทธิ์ พิริย์โยธินกุล
<b>รหัสนักศึกษา</b>	58070077
<b>สาขาวิชา</b>	เทคโนโลยีสารสนเทศ
<b>อาจารย์ที่ปรึกษา</b>	รองศาสตราจารย์ ดร. กิตติสุชาต พสกุล
<b>ปีการศึกษา</b>	2561

## บทคัดย่อ

การ์ตูนญี่ปุ่น หรือที่รู้จักกันอย่างแพร่หลายว่า มังงะ (Manga) ถูกเป็นหนึ่งในหัวข้อที่ถูกหิบมา วิจัย ในงานวิจัยขึ้นนี้มุ่งเน้นไปที่ปัญหาการตรวจหาข้อความบนภาพมังงะ เนื่องจากปัญหาการสร้าง ชุดข้อมูลภาพมังงะ และข้อมูลประกอบ (Annotation) อย่างเช่น การระบุขอบเขตข้อความ ซึ่งต้องใช้ แรงงานคนและกินเวลาอย่างมาก ดังนั้นการพัฒนาระบบอัตโนมัติที่จะสามารถเข้ามาช่วยในงานส่วนนี้ ได้จะเป็นสิ่งที่น่าสนใจเป็นอย่างมาก โดยเราได้นำเสนอวิธีการตรวจหาข้อความบนภาพมังงะแบบใหม่ ด้วยการใช้ Stroke Width Transform (SWT) ร่วมกับการใช้ Support Vector Machine (SVM) อย่างไร ก็ดีวิธี SWT โดยดึงเดินน้ำเสียงพัฒนาขึ้นเพื่อ เพื่อการตรวจหาข้อความบนภาพถ่าย ทำให้มีความสามารถ ประยุกต์ใช้กับภาพมังงะได้ เพราะความแตกต่างระหว่างลักษณะตัวอักษรของข้อความในภาพมังงะ และภาพถ่ายนั้นคล้ายคลึงกันเกินไป ดังนั้นเพื่อให้สามารถใช้งานกับมังงะได้ เราจึงนำวิธีการตรวจ หาข้อความด้วย SWT ดึงเดินมาปรับปรุงและพัฒนาขึ้นเป็นวิธีการใหม่ของเรา โดยเราได้ปรับปรุงกฎ เกณฑ์ในการค้นหาตัวอักษรที่คล้ายคลึงอักษร (Letter Candidates) ซึ่งช่วยเพิ่มประสิทธิภาพในการตรวจ จับอักษรได้ครบถ้วนมากขึ้น และใช้ SVM เพื่อคัดแยกตัวอักษร ออกจากอักษร ช่วยในการลด False Positive ของผลลัพธ์ ในท้ายที่สุดเรานำประสิทธิภาพวิธีการของเรามาเปรียบเทียบร่วมกับวิธีการดึงเดิน และวิธีอื่น ๆ รวมถึงวิธีที่ใช้ Deep Learning เป็นส่วนประกอบ ในท้ายที่สุดประสิทธิภาพของวิธีการ ใหม่ของเรานั้นสามารถทำคะแนน F-measure ได้สูงสุดเทียบกับวิธีการอื่น ๆ ที่ 0.506

<b>Project Title</b>	Detecting Text in Manga Using Stroke Width Transform
<b>Student</b>	Boonyarith Piriyothinkul
<b>Student ID</b>	58070077
<b>Program</b>	Information Technology
<b>Advisor</b>	Associate Professor Dr. Kitsuchart Pasupa
<b>Year</b>	2018

## Abstract

The Japanese comic-book style known as manga is becoming a popular topic for researchers. This paper focuses on the problem of detecting text regions in manga pages. Because it is time-consuming and laborious to identify the text regions in images manually, an automatic approach is highly desirable. Here, we propose a new text-detection method for manga using a Stroke Width Transform (SWT) technique in conjunction with a Support Vector Machine (SVM). Conventional SWT-based text-detection techniques perform poorly with manga because both text and non-text objects have similar characteristics for strokes, lines, and shapes. To better suit manga, we propose modifying the rules for finding letter candidates, which improves the ability to capture text. An SVM is then used to classify image patches into letter and nonletter regions. We compared our proposed framework with a conventional framework and other text-detection methods including deep-learning techniques. In the results, our proposed method achieved the highest F-measure of 0.506.

# สารบัญ

	หน้า
บทคัดย่อ	I
Abstract	II
สารบัญ	III
สารบัญตาราง	IV
สารบัญรูป	V
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 วัตถุประสงค์	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 การตรวจหาข้อความในภาพถ่ายด้วยเทคนิค Stroke Width Transform	3
2.2 Histogram of Oriented Gradients	5
2.3 Support Vector Machine	6
บทที่ 3 วิธีการทดลอง	8
3.1 วิธีการใหม่ที่ลูกปั้นปรุงและพัฒนาเพิ่มเติม	8
3.2 ชุดข้อมูลสำหรับการเทรนโมเดล SVM	12
3.3 การทดลอง	12
บทที่ 4 ผลการทดลอง	15
บทที่ 5 สรุปผล	17
บรรณานุกรม	18
บทที่ ก การใช้ชีวิตในประเทศไทยปัจจุบัน	21
ก.1 ที่อยู่อาศัย	22
บทที่ ข กิจกรรมระหว่างฝึกงาน	24
ข.1 Mirai Symposium	25
ข.2 Lab Meeting	25
ข.3 กิจกรรมอื่น ๆ	26
ภาคผนวก ก ผลงานวิจัยที่ได้รับการตีพิมพ์	31

## สารบัญตาราง

ตารางที่	หน้า
4.1 ตารางแสดงการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ของเราร่วมกับวิธีการอื่น ๆ ที่เกี่ยวข้อง	15

# สารบัญ

หัวข้อ	หน้า
รูปที่	
2.1 ขั้นตอนการทำงานของ Stroke Width Transform	4
2.2 ตัวอย่างข้อมูลนำเข้าและการจำลองภาพพิศทางของ Histogram of Oriented Gradients	5
2.3 Cell และ Block ในการทำงานของ Histogram of Oriented Gradients	6
2.4 การแบ่งแยกกลุ่มข้อมูลด้วย Hyper-plane ของ SVM	7
2.5 คุณสมบัติการเปลี่ยนมิติของข้อมูลด้วย Kernel	7
3.1 ตัวอย่างผลลัพธ์จากการตรวจหาข้อความบนภาพมังงะด้วยวิธีต้นฉบับ [1] และดึงให้เห็น False Positive จำนวนมาก (ก) นักวาด: Shinoasa (ญ) นักวาด: Kousei (Public Planet)	9
3.2 แผนผังการทำงานของ (ก) วิธีการดึงเดิม [1] และ (ญ) วิธีการใหม่ของเรา	10
3.3 ตัวอย่างแสดงการเปรียบเทียบผลลัพธ์ระหว่างขอบเขตตัวอักษรที่ตรวจพบระหว่างการใช้ กฏเกณฑ์เก่าของ SWT ต้นฉบับ (ก) และกฏเกณฑ์ใหม่ในวิธีของเรา (ญ) ข้อมูลภาพถูกนำมาจากเรื่อง Arisa ©Yagami Ken	11
3.4 ตัวอย่างของ Patch: (ก) ภาพ Positive Patches and (ญ) ภาพ Negative Patches	11
3.5 ตัวอย่างแสดงการจับกลุ่มของตัวอักษร	12
4.1 ตัวอย่างขอบเขตข้อความที่วิธีการของเราตรวจพบ (ก–ญ) Love Hina ©Ken Akamatsu และ (ค–ญ) Eva Lady ©Miyone Shi	16
ก.1 ภาพหอพัก International House Kita 8 East	23
ญ.1 ภาพกิจกรรมในงาน Mirai Symposium	26
ญ.2 ภาพระหว่างงานเลี้ยงต้อนรับ	27
ญ.3 งานเลี้ยงจำลาและต้อนรับนักศึกษาปีสาม	28
ญ.4 ป้ายชื่อห้องห้องทดลองโดยมีการใช้ตัวละครจาก การ์ตูนประกอบให้น่าสนใจมากขึ้น	29

## บทที่ 1

### บทนำ

#### 1.1 ที่มาและความสำคัญ

การ์ตูนญี่ปุ่นเป็นที่รู้จักกันอย่างแพร่หลายทั่วโลกในฐานะสื่อบันเทิง หรืออีกชื่อหนึ่งคือ “มังงะ (Manga)” ในปัจจุบันมีงานวิจัยในหัวข้อมังงะอย่างหลากหลาย ในหลาย ๆ งานวิจัย [2–8] มีการใช้ชุดข้อมูลสำหรับการทดลอง เช่น Manga109 [9] ซึ่งเป็นชุดข้อมูลที่ถูกสร้างขึ้นจากภาพมังงะจำนวน 20,260 หน้า รวมรวมจากมังงะ 109 เรื่อง มังงะที่ถูกรวบรวมนี้เป็นผลงานของนักวาดมังงะมืออาชีพชาวญี่ปุ่น นอกจากภาพของมังงะแล้ว ชุดข้อมูลนี้ยังประกอบไปด้วยข้อมูลเชิงประ同胞 หรือ Annotation ต่าง ๆ เช่น ขอบเขตและตำแหน่งของใบหน้า ร่ายกาย และ ครอบภาพ เป็นต้น นอกจากนี้ยังมีข้อมูลของเขตและตำแหน่งของข้อความที่ปรากฏในภาพมังงะ โดยตำแหน่งข้อความต่าง ๆ นั้นถูกป้อนข้อมูลด้วยแรงงานคนโดยไม่เพ่งพาระบบอัตโนมัติ ในการป้อนข้อมูลดังกล่าวนั้นใช้เวลานานและต้องเพ่งพาระบบงานมนุษย์ ด้วยเหตุนี้ระบบอัตโนมัติที่จะสามารถเข้ามาร่วมในการระบุข้อมูล Annotation นั้นจึงมีประโยชน์และสามารถช่วยลดภาระงานในส่วนนี้ลงได้อย่างมาก

ถึงแม้ว่าสำหรับภาพขาวรูปแบบการ์ตูนญี่ปุ่นจะมีทั้งแบบภาพวาดทั่วไปที่เป็นภาพแสดงของตัวละคร หรือทิวทัศน์ และแบบภาพมังงะ แต่ภายในงานวิจัยนี้เรามุ่งเน้นไปที่มังงะ เป็นหลักเนื่องจาก ข้อความมักปรากฏบนหนังสือการ์ตูนมากกว่าภาพวาดทั่วไปอย่างที่ทราบกันดี สำหรับวิธีการตรวจหา ข้อความในภาพมังงะนั้นมีการพัฒนามาทางลากหลายก่อนหน้านี้ [10, 11] แต่วิธีเหล่านี้ถูกพัฒนาให้เพ่งพาระบบสร้างส่วนต่าง ๆ ของภาพมังงะเป็นข้อมูลอ้างอิง เช่น ครอบช่องภาพวาด, ลักษณะของกล่องคำพูด เป็นต้น นอกจากนี้บางวิธียังคงมีความจำเป็นที่ต้องเพ่งพาระบบป้อนข้อมูลเข้าจากภายนอกทั้งจากมนุษย์ และข้อมูลอื่น ๆ ทำให้มีความสามารถทำงานได้อัตโนมัติอย่างสมบูรณ์ อย่างไรก็ได้ไม่นานมานี้มีการพัฒนาวิธี การใหม่โดยใช้วิธีการ Deep Learning อย่างเช่นเทคนิค Convolutional Neural Network เพื่อช่วยในการ สกัดลักษณะเด่น (Feature) ออกจากภาพมังงะเพื่อช่วยในการตรวจหาข้อความในภาพมังงะ [12] ซึ่งวิธี การนี้สามารถเพิ่มความแม่นยำและถูกต้องในการตรวจหาข้อความได้โดยปราศจากการเพ่งพาระบบสร้าง ต่าง ๆ ในภาพมังงะ แต่อย่างไรก็ได้ Deep Learning ยังเป็นการวิธีการที่ต้องใช้ทรัพยากรของระบบเพื่อ การคำนวนมากกว่าวิธีการอื่น ๆ ซึ่งเป็นข้อเสียสำคัญประการหนึ่ง [12]

ในงานวิจัยนี้มีจุดมุ่งหมายเพื่อพัฒนาระบบตรวจหาข้อความที่ทำงานได้กับมังงะอย่างลากหลาย และไม่ถูกจำกัดด้วยโครงสร้างหรือลักษณะทางประการของภาพมังงะ เราจึงเลือกใช้ Stroke Width Transform (SWT) ในการสกัดลักษณะเด่นของเส้นต่าง ๆ ของวัตถุที่ปรากฏในภาพออกแบบ โดยวิธี การนี้ถูกใช้เป็นขั้นตอนแรกของการตรวจหาข้อความบนภาพถ่ายมาก่อนหน้านี้ วิธีการนี้ทำงานโดยเพ่งพาระบบสมมติฐานว่าขอบของเส้นอักษรในข้อความนั้นมีขอบที่ชัดเจนและหนาแน่นปรากฏอยู่บนพื้นหลังที่รับเรียน [1] อย่างไรก็ได้การใช้วิธีการนี้กับการตรวจหาข้อความบนภาพมังงะส่งผลให้เกิดข้อผิดพลาดเชิง False Positive จำนวนมาก ปัญหานี้เกิดจากความแตกต่างของลักษณะเฉพาะตัวของภาพถ่าย และภาพมังงะ ภาพมังงะนั้นโดยส่วนใหญ่มีลักษณะเป็นภาพขาวดำ และลักษณะของวัตถุภายในมังงะ เช่น ขนาด, เส้น, และพื้นหลัง นั้นมีความคล้ายคลึงกับตัวอักษรของข้อความ ด้วยปัญหาข้างต้นเรา

จึงต้องเป้าหมายในการปรับปรุงและพัฒนา SWT ที่ถูกใช้ในภาพถ่าย [1] เพื่อให้สามารถทำงานกับภาพมังงะได้

วิธีการใหม่ของเรานี้ถูกพัฒนาขึ้นใหม่นั้นแบ่งออกเป็น 4 ส่วนดังนี้ (i) The Stroke Width Transform (ii) ค้นหาวัตถุที่เข้าข่ายลักษณะของตัวอักษร (iii) คัดแยกอักษร โดยใช้ Support Vector Machine (SVM) ร่วมกับ Histogram of Oriented Gradients Feature (iv) จัดกลุ่มอักษรที่ผ่านการคัดแยกแล้วให้เกิดเป็นบรรทัดหรือกลุ่มของข้อความ

## 1.2 วัตถุประสงค์

พัฒนาระบบค้นหาตำแหน่งข้อความสำหรับมังงะ โดยนำ Stroke Width Transform ที่ถูกใช้เป็นกระบวนการตรวจสอบในเทคนิคตรวจหาข้อความบนภาพถ่ายมาพัฒนาและปรับปรุงต่อยอดเพื่อให้สามารถใช้งานกับภาพมังงะได้มีประสิทธิภาพมากขึ้น

## 1.3 ขอบเขตของงานวิจัย

1. พัฒนาระบบตรวจหาตำแหน่งข้อความซึ่งใช้สำหรับภาพมังงะ โภนสีขาวดำ
2. ภาษาของเนื้อหาในมังงะที่นำมาใช้งาน คือ ภาษาญี่ปุ่น
3. ข้อมูลที่ใช้ในการวิจัยเพื่อการเทรนและทดสอบนำมาจากฐานข้อมูล Manga109

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้วิธีการตรวจหาข้อความใหม่ที่ถูกพัฒนาขึ้นเพื่อใช้งานร่วมกับภาพมังงะโดยเฉพาะ
2. ทำให้ทราบถึงลักษณะที่เป็นเอกลักษณ์ของมังงะซึ่งแตกต่างจากภาพถ่ายทั่วไป

## บทที่ 2

# แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การตรวจหาข้อความในภาพถ่ายด้วยเทคนิค Stroke Width Transform

วิธีการการตรวจหาข้อความในภาพถ่าย [1] นี้เป็นวิธีการที่เรานำมาใช้ศึกษาและเป็นต้นแบบในการพัฒนาเพื่อทำงานร่วมกับภาพมังงะ โดยมีขั้นตอนทั้งหมดแบ่งได้เป็น 3 ขั้นตอน ขั้นแรกคือการใช้ Stroke Width Transform ในการปรับเปลี่ยนข้อมูลให้แสดงลักษณะของความกว้างในแต่ละเส้นภายในภาพ ขั้นที่สอง ค้นหาวัตถุที่คล้ายคลึงกับตัวอักษรในภาพโดยใช้กฎเกณฑ์ที่กำหนดไว้ ขั้นสุดท้าย คือการจัดกลุ่มตัวอักษรเข้าด้วยกันเป็นบรรทัดของข้อความ

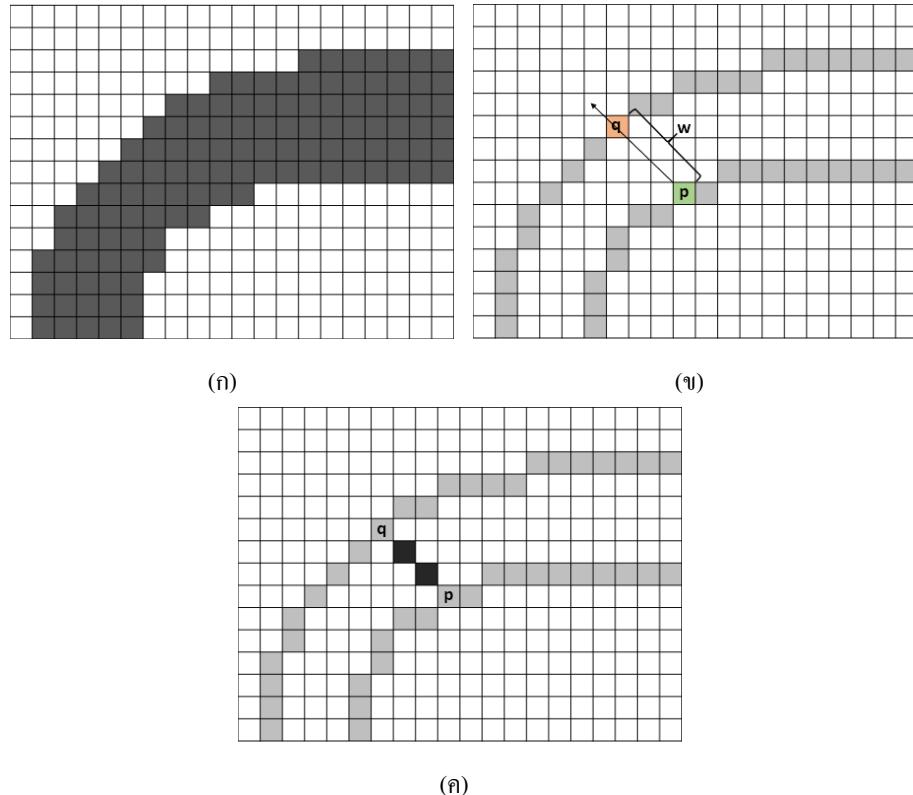
#### 2.1.1 Stroke Width Transform

Stroke Width Transform หรือ SWT เป็นเทคนิคที่ใช้ช่วยในการทำงานของระบบตรวจหาข้อความในภาพถ่าย โดยสักดิลักษณะเด่นของเส้นต่าง ๆ ในภาพ เช่น เส้นของตัวอักษร เป็นต้น [1] ด้วยลักษณะดังกล่าว ทำให้เราสามารถใช้ในการคัดแยกวัตถุที่เป็นตัวอักษรออกจากวัตถุอื่น ๆ โดยพึ่งพาลักษณะเด่นเหล่านี้

เริ่มแรกเราสร้างภาพ Output ที่มีขนาดเท่ากับภาพที่ต้องการตรวจหาข้อความ โดยแต่ละ Pixel ในภาพถูกกำหนดให้มีค่าอนันต์ ( $\infty$ ) จากนั้นใช้ Canny Edge Detection [13] ตรวจหาตำแหน่งของขอบของวัตถุในภาพ ซึ่งจากตัวอย่างในภาพ หากเรามีภาพ 2.1(ก) จากนั้นใช้การตรวจหาตำแหน่งของขอบของวัตถุ เราจะได้ผลลัพธ์ตามภาพ 2.1(ข) เมื่อตรวจหาตำแหน่งของขอบเสร็จสิ้นต่อมากำหนดความกว้างของเส้นโดยใช้ขอบที่ได้มา ความกว้างคำนวนได้จากระยะห่างระหว่างขอบของเส้นโดยพิจารณาทุก Pixel  $p$  ของขอบที่ได้จาก Canny Edge Detection เพื่อหา Pixel  $q$  ที่เข้าคู่กันอย่างที่แสดงให้เห็นในภาพ 2.1(ก) การหา  $q$  จาก  $p$  ทำได้โดยใช้ Gradient Direction ของ  $p$  ซึ่งคือ  $d_p$  โดย  $d_p$  จะชี้ไปทาง  $q$  และหาก  $d_p$  และ  $d_q$  มีทิศทางตรงกันข้ามโดยประมาณ  $d_q = -d_p \pm \pi/6$  ให้กำหนดค่าให้กับแต่ละ Pixel ที่อยู่ภายใต้แนวทางระหว่าง  $p$  และ  $q$  ให้มีค่าเท่ากับ  $\|\overrightarrow{p-q}\|$  เว้นแต่ว่า Pixel ที่จะระบุค่าให้นั้นมีค่าเดิมน้อยกว่าค่าใหม่ที่จะระบุให้ ดังนั้นหากค่าใหม่น้อยกว่าค่าเดิมใน Pixel ก็สามารถทำการระบุค่าใหม่แทนที่ค่าเดิมให้กับ Pixel นั้นได้ อย่างที่ปรากฏในภาพ 2.1(ก) เมื่อทั้งหมด Pixel ในภาพ สุดท้ายจะได้เมทริกซ์ Output ขนาดเท่ากับ Input โดยมีค่าของความกว้างเส้นถูกระบุในพื้นที่ระหว่างขอบของเส้นอย่างเช่นปรากฏในภาพ 2.1(ก)

#### 2.1.2 ค้นหาวัตถุที่มีลักษณะใกล้เคียงตัวอักษร

ในขั้นตอนนี้เราจะนำผลลัพธ์จากขั้นตอนก่อนหน้ามาจำจัดวัตถุที่ไม่มีลักษณะคล้ายคลึงอักษร เริ่มจากจับกลุ่มแต่ละ Pixel ในผลลัพธ์ของขั้นตอนก่อนหน้า การจับกลุ่มทำได้โดยเปรียบเทียบแต่ละ Pixel กับ Pixel เพื่อนบ้านรอบข้าง หาก Pixel สองตัวที่เปรียบเทียบกันมีค่าของความกว้างเส้นไม่ต่างกันเกิน 3.0 เท่า ให้ถือว่า Pixel ทั้งสองเป็นส่วนของวัตถุเดียวกันซึ่งจะถูกจัดกลุ่มเข้าด้วยกัน เมื่อการจัดกลุ่ม Pixel เสร็จสิ้น ผลลัพธ์ที่ได้คือภาพวัตถุต่าง ๆ ที่ปรากฏในภาพหรือ Connected Components อย่างไรก็ได้



รูปที่ 2.1 ขั้นตอนการทำงานของ Stroke Width Transform

หากวัตถุที่เกิดจากการรวมกลุ่มของ Pixel นั้นใหญ่หรือเล็กเกินไปก็จะถูกคัดออก การคัดวัตถุลักษณะดังกล่าวออกไปทำโดยการใช้กฎสองข้อดังนี้ (i) อัตราส่วนระหว่างเส้นผ่าศูนย์กลางต่อมัธยฐานของความกว้างของเส้นตัวอักษรนั้นต้องน้อยกว่า 10 (ii) ความสูงต้องมากกว่า 10 และน้อยกว่า 300 ตามที่แสดงในสมการ 2.1

$$f(d, h, \tilde{s}) = \begin{cases} 1, & \text{if } \frac{d}{\tilde{s}} < 10 \text{ and } 10 < h < 300 \\ 0, & \text{otherwise} \end{cases}, \quad (2.1)$$

โดย  $d$  คือ เส้นผ่าศูนย์กลางของวัตถุ,  $h$  คือ ความสูงของวัตถุ, และ  $\tilde{s}$  คือ มัธยฐานของความกว้างเส้นตัวอักษร โดยค่าความกว้างได้มามาจากค่าของ Pixel ในพื้นที่ของเส้นที่ถูกระบุไปโดยขั้นตอน SWT

### 2.1.3 จัดกลุ่มตัวอักษร

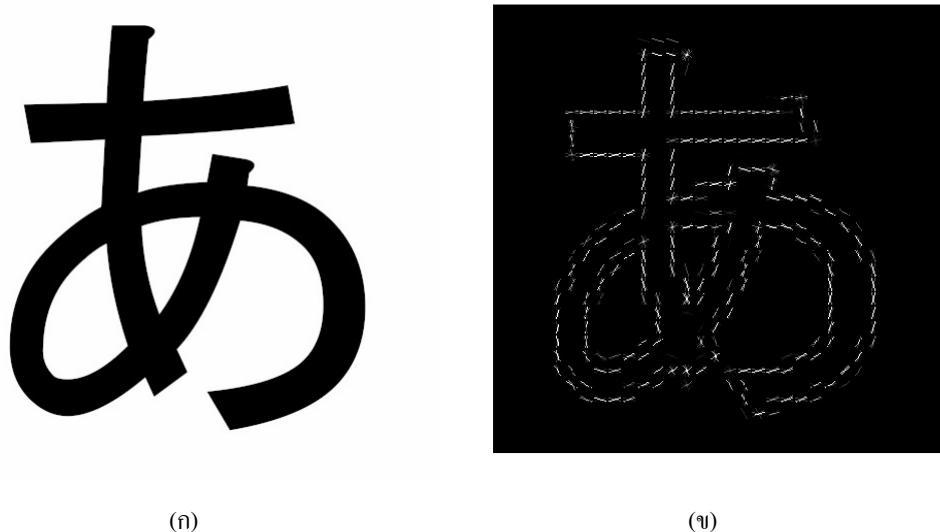
วัตถุแต่ละชิ้นที่ที่ลักษณะคล้ายคลึงกับอักษรซึ่งผ่านการคัดกรองด้วยกฎจากขั้นตอนก่อนหน้าจะถูกนำมาจับกลุ่มเป็นบรรทัดของข้อมูลในขั้นตอนนี้โดยใช้การเปรียบเทียบความคล้ายคลึงระหว่างลักษณะอักษรต่าง ดังนี้ ระยะห่างระหว่างวัตถุ, อัตราส่วนความกว้างของเส้นอักษร, และความสูงของอักษร โดยสองวัตถุจะถูกจัดกลุ่มกันต่อเมื่อ (i) อัตราส่วนระหว่างค่ามัธยฐานความกว้างเส้นของวัตถุทั้งสองมีค่าน้อยกว่า 2 เท่า (ii) ความสูงของอักษรทั้งสองต่างกันไม่เกิน 2 เท่า (iii) ระยะห่างระหว่างสอง

วัตถุนั้นมีค่าไม่เกิน 3 เท่าของวัตถุที่กว้างที่สุดในคู่อักษรที่ใช้เปรียบเทียบ หลังจากการจัดกลุ่มนี้เราจะได้ใช้ของอักษรที่ถูกจัดกลุ่มเข้าด้วยกัน แต่ละโซ่อักษรจะประกอบไปด้วยอักษรสองตัวที่ถูกจัดกลุ่ม ต่อมานั้นแต่ละโซ่อักษรรวมเข้าด้วยกันหากโซ่อักษรนี้มีอักษรในโซ่อักษรร่วมกันโซ่อื่น ๆ และทิศทางของโซ่อักษรจะมีความใกล้เคียงกัน

สุดท้ายขั้นตอนนี้จะจบลงเมื่อไม่มีโซ่อักษรใด ๆ ถูกเชื่อมต่อเพิ่มเติม ในที่สุดเราจะได้กลุ่มหรือโซ่อักษรที่เกิดจากการจัดกลุ่มด้วยความคล้ายคลึงของอักษรและทิศทางของข้อความ อีกนัยนึงคือเราได้กลุ่มบรรทัดของแต่ละประโยคออกจากภาพถ่ายเรียบร้อยในขั้นตอนนี้

## 2.2 Histogram of Oriented Gradients

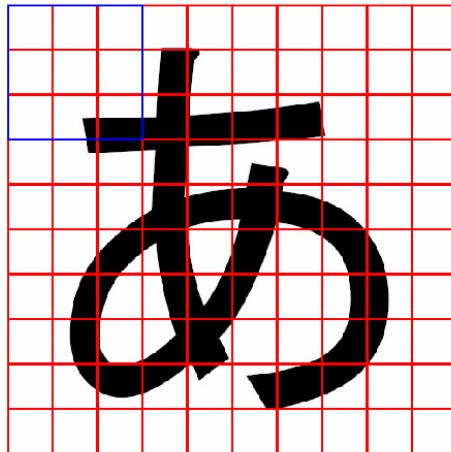
Histogram of Oriented Gradients (HOG) เป็นการสกัด Feature ของภาพโดยอาศัยรูปแบบ Histogram ของทิศทางเนคต์ในภาพ หรือ Gradient direction เพื่อพิจารณาลักษณะของวัตถุต่าง ๆ อย่างที่แสดงในภาพ 2.2 โดยภาพ 2.2(ก) คือ ตัวอย่างอักษรภาษาญี่ปุ่น และภาพ 2.2(ข) คือภาพแสดงทิศทางของเนคต์ของภาพอักษรโดยใช้เส้นขนาดเล็กแสดงทิศทางของเนคต์ ด้วยความสามารถนี้ จึงมีการนำ HOG มาใช้สำหรับสกัดลักษณะเด่นเพื่อใช้ในการจำพวกการตรวจจับวัตถุอย่างหลากหลาย ทั้ง การตรวจจับท่าทางของมือ [14], การตรวจจับรถบนท้องถนน [15], การตรวจจับมนุษย์ในภาพ [16] และไม่เพียงแค่สามารถใช้กับงานตรวจจับวัตถุ แต่ยังสามารถใช้กับงานด้านตรวจหาข้อความในภาพได้เช่นกัน [17–19]



รูปที่ 2.2 ตัวอย่างข้อมูลนำเข้าและการจำลองภาพทิศทางของ Histogram of Oriented Gradients

การสกัดลักษณะเด่นของ HOG นั้นทำได้โดยเริ่มจากแบ่งภาพเป็นส่วนเล็ก เรียกว่า Cell หรือช่องสี่เหลี่ยมตามที่แสดงในภาพ 2.3 จากนั้นสร้าง Histogram สำหรับ Cell นั้น ๆ ด้วยค่า Gradient Direction และ Magnitude โดย Histogram นี้จะเป็นตัวแทนของลักษณะของรูปร่างที่อยู่ภายใน Cell นั้น ๆ จากนั้นจะทำการ Normalization กับ Histogram ของแต่ละ Cell ด้วยกลุ่มของ Cell หรือที่เรียกว่า Block อย่างที่เห็นเป็นช่องสี่เหลี่ยมในภาพ 2.3 สุดท้ายเราจะได้ Histogram ของ Gradient Direction จากทุก ๆ Cell

ของภาพซึ่งเป็นตัวแทนของรูปปั้งวัตถุแต่ละส่วน ด้วย Histogram ที่ได้มาจะถูกนำไปใช้กระบวนการ Vectorization เพื่อให้สามารถใช้ในงานอื่น ๆ ได้ต่อไป



รูปที่ 2.3 Cell และ Block ในการทำงานของ Histogram of Oriented Gradients

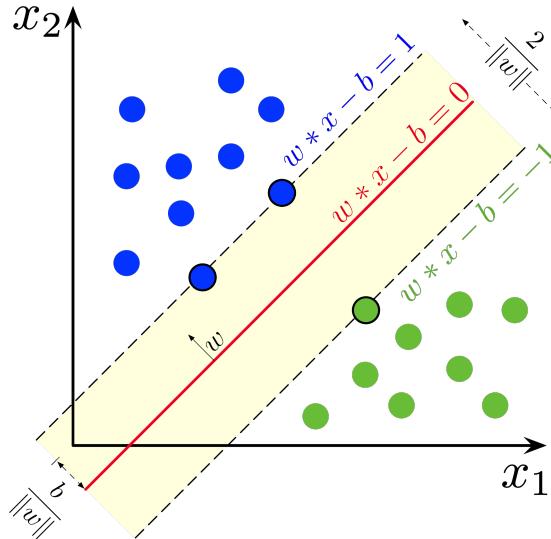
อย่างไรก็ได้การที่จะได้ HOG ของวัตถุที่เราต้องการตรวจสอบจำเป็นต้องใช้ภาพของวัตถุนั้น ๆ ใน การคำนวณ แต่ในสถานการณ์จริงภาพของวัตถุอาจอยู่ในภาพถ่ายขนาดใหญ่ที่ประกอบไปด้วยหลาย วัตถุ เราจึงต้องตัดภาพของวัตถุเป็นส่วนย่อยเพื่อใช้คำนวณกับ HOG เราเรียกภาพส่วนย่อยที่ถูกตัดออก นานัปการ ว่า Patch ดังนั้นเราจำเป็นต้องใช้ Patch ที่มีขนาดใกล้เคียงกับวัตถุนั้น ๆ เป็นเหตุให้หากภาพมี ขนาดใหญ่แต่วัตถุที่ต้องการตรวจพบมีขนาดเล็ก จำนวน Patch ก็จะมากขึ้น

ในการนิยามมังงะ ตัวอักษรนักมีขนาดเล็ก ( $20\text{px} - 40\text{px}$  โดยส่วนใหญ่อ้างอิงจาก Dataset ของเรา) เมื่อเทียบกับขนาดภาพมังงะ ( $1170\text{px}$  อ้างอิงจาก dataset ของเรา) ซึ่งมีขนาดใหญ่กว่าหลายเท่า ดังนั้น จำนวนของ Patch ที่ต้องสร้างและคำนวนด้วย HOG จึงมีมหาศาลและสร้างภาระแก่การคำนวน ลักษณะเด่นด้วย SVM อย่างมาก ซึ่งปัญหาส่วนนี้ส่งผลกระทบต่อความเร็วในการทำงานของเรา เราจึงใช้ SWT สำหรับกำหนดพื้นที่ที่คาดว่าเป็นอักษรเพื่อใช้สร้าง Patch แทนการสร้าง Patch จากทุกส่วนของภาพ ด้วยการใช้ Sliced Window

### 2.3 Support Vector Machine

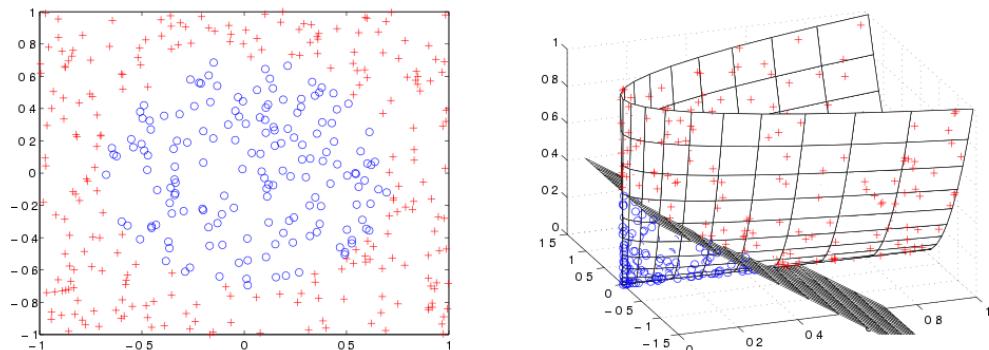
Support Vector Machine (SVM) [20] เป็นเทคนิค Pattern Recognition แบบ Supervised Learning ซึ่งถูกใช้ทั้งในงานเพื่อ Classification และ Regression ซึ่งภายในงานนี้ได้ใช้งานเพื่อ Classification โดย ทำงานด้วยการสร้าง Hyper-plane ที่เหมาะสมที่สุด (Optimal) เพื่อจำแนกแยกข้อมูลสองกลุ่มอย่างที่ แสดงในภาพ 2.4

เพื่อที่จะแยกข้อมูลทั้งสองกลุ่มด้วย Optimal Hyper-plane นั้น  $w \times x - b = 0$  จะทำหน้าที่ แบ่งข้อมูลสองกลุ่มออกจากกันโดยมี Support Vector ทำหน้าที่เป็นกันชนระหว่างข้อมูลที่ใกล้กันที่สุด ระหว่างกลุ่มข้อมูลทั้งสอง ซึ่ง SVM นั้นจะสร้างพื้นที่การตัดสินใจขึ้นมา หรือก็คือพื้นระหว่าง  $w \times x - b = 1$  และ  $w \times x - b = -1$  โดยจะปรับให้ระยะห่างหรือความกว้างระหว่างทั้งสองนี้มีค่า



รูปที่ 2.4 การแบ่งแยกกลุ่มข้อมูลด้วย Hyper-plane ของ SVM

สูงสุด โดยระยะห่างนั้นมีค่าเท่ากับ  $\frac{2}{\|w\|}$  อย่างไรก็ได้หลาย ๆ ครั้งข้อมูลไม่สามารถแบ่งแยกได้ด้วยเส้นตรง และจำเป็นต้องใช้การแบ่งข้อมูลแบบ Non-linear ซึ่งสำหรับ SVM แล้วนั้นสามารถใช้ Kernel เป้ามาช่วยในการเปลี่ยนมิติของข้อมูลเพื่อให้สามารถแบ่งแยกข้อมูลทั้งสองกลุ่มได้ด้วย Linear Hyper-plane ตามที่แสดงในภาพ 2.5



รูปที่ 2.5 คุณสมบัติการเปลี่ยนมิติของข้อมูลด้วย Kernel

## บทที่ 3

### วิธีการทดลอง

ในบทนี้เรากล่าวถึงวิธีการใหม่ของเรารather ได้ปรับปรุงและพัฒนาขึ้นมาโดยใช้ SWT เป็นส่วนหลักในระบบใหม่ โดยมีจุดประสงค์เพื่อทำให้สามารถใช้งานร่วมกับภาพมังงะได้อย่างมีประสิทธิภาพ และดำเนินการทดลองเพื่อวัดประสิทธิภาพของวิธีการใหม่ของเราว่าสามารถทำงานได้ดีขึ้นหรือไม่อย่างไร เมื่อเปรียบเทียบกับวิธีการต้นฉบับ [1] และวิธีการอื่น ๆ ที่ถูกพัฒนามาก่อนหน้า

#### 3.1 วิธีการใหม่ที่ถูกปรับปรุงและพัฒนาเพิ่มเติม

สำหรับวิธีการอย่างที่ได้กล่าวไว้ในบทที่ 1 วิธีการใหม่ของเราราได้ใช้ประโยชน์จาก SWT ร่วมกับความสามารถของ SVM โดยใช้ HOG เป็นลักษณะเด่น หรือ Feature อย่างไรก็ตามที่ได้กล่าวไว้ในบทที่ 1 จุดประสงค์หลักของ SWT ที่ถูกนำเสนอไปในงานวิจัยก่อนหน้านี้ถูกออกแบบเพื่อการตรวจหาข้อความบนภาพถ่ายเป็นภาษาไทยหลัก ด้วยเหตุผลนี้ทำให้การทำงานร่วมกับภาพมังงะไม่สามารถทำงานได้ดีอย่างที่ควร ก่อให้เกิด False Positive จำนวนมาก ตามที่แสดงให้เห็นในภาพ 3.1 ซึ่งจำนวน False Positive ที่มากนั้นแสดงถึงประสิทธิภาพที่ต่ำของระบบ สาเหตุหลักคือความแตกต่างเชิงเอกลักษณ์ของวัสดุในภาพจริงและภาพคอมพิวเตอร์ นอกเหนือไปจากการที่ตัวอักษรในภาพมากกินไป เช่น เส้นของตัวหนังสือ, เส้นผ่านศูนย์กลางของตัวอักษร, และรายละเอียดบนพื้นหลัง อย่างที่แสดงในภาพ 3.1(g) และภาพ 3.1(h) ด้วยเหตุนี้วิธีการของเราจึงปรับปรุงขั้นตอนของระบบดังเดิม โดยเราได้ปรับปรุงขั้นตอนการค้นหาวัตถุที่คล้ายคลึงอักษร การจับกลุ่มอักษร, และเพิ่มขั้นตอนใหม่สำหรับการคัดแยกอักษรเพิ่มอีกหนึ่งขั้นตอน เพื่อให้สามารถใช้งานกับภาพมังงะได้อย่างมีประสิทธิภาพและแม่นยำมากขึ้น

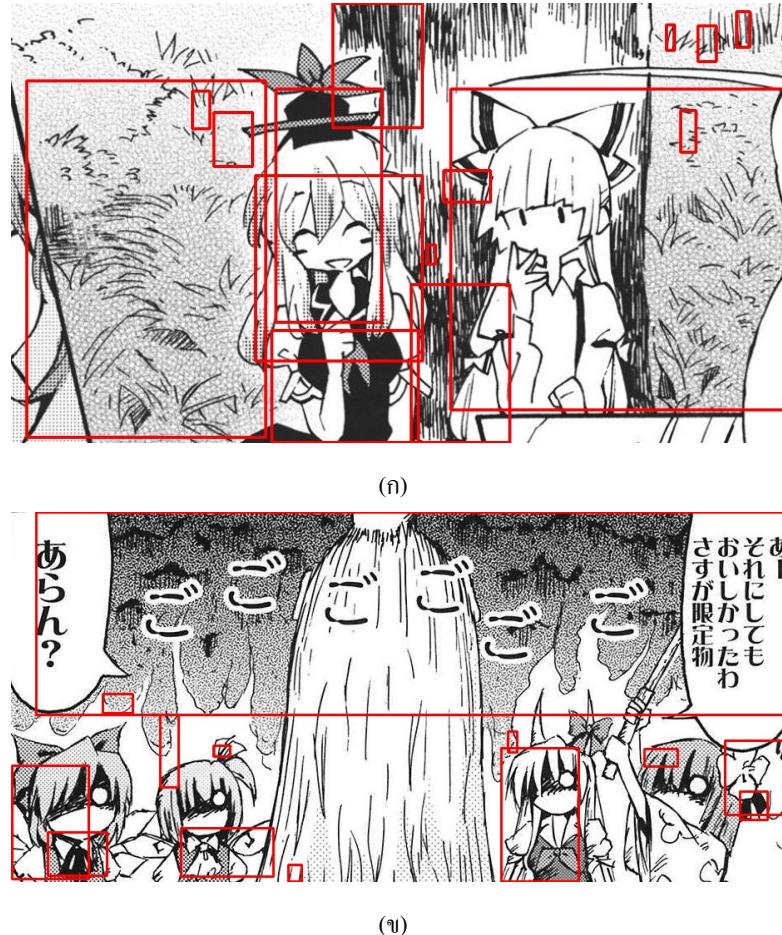
ความแตกต่างของวิธีต้นฉบับและวิธีการใหม่ของเรานั้นถูกแสดงให้เห็นในภาพ 3.2 อย่างที่เห็นในภาพ 3.2(h) เราได้เพิ่มขั้นตอนการคัดแยกตัวอักษรเข้ามา โดยวิธีการใหม่จะคัดแยกอักษรออกจากวัสดุอื่น ๆ ที่ไม่มีความเกี่ยวข้องกันที่จะจับกลุ่มตัวอักษรเข้าเป็นประยุกต์ ขั้นตอนการคัดแยกนี้ใช้ความสามารถของ SVM Classification เพื่อช่วยลด False Positive ของผลลัพธ์การค้นหาวัตถุที่คล้ายตัวอักษรจากขั้นตอนก่อนหน้า

##### 3.1.1 The Stroke Width Transform

ในขั้นตอนนี้ใช้วิธีการเดียวกับงานวิจัย [1] ที่กล่าวไว้ในบทที่ 2 โดยเรานำ SWT มาดำเนินการบนภาพมังงะเพื่อให้อยู่รูปแบบตัวดำเนินการ SWT โดยข้อมูล Output ของขั้นตอนนี้คือเมทริกซ์ขนาดเท่ากับภาพ Input ซึ่ง Output นี้จะถูกใช้ในขั้นตอนต่อไป

##### 3.1.2 ค้นหาวัตถุที่ใกล้เคียงอักษร

ในมังงะนั้นข้อความหรืออักษรที่หลายเมื่นาดที่หลากรายละเอียดต่างไปจากภาพถ่าย เราจึงต้องนำกฎเกณฑ์ที่ใช้ในการคัดกรองวัตถุกับตัวอักษรบนภาพถ่ายมาใช้ทำการคัดเปลี่ยนให้เหมาะสมกับ

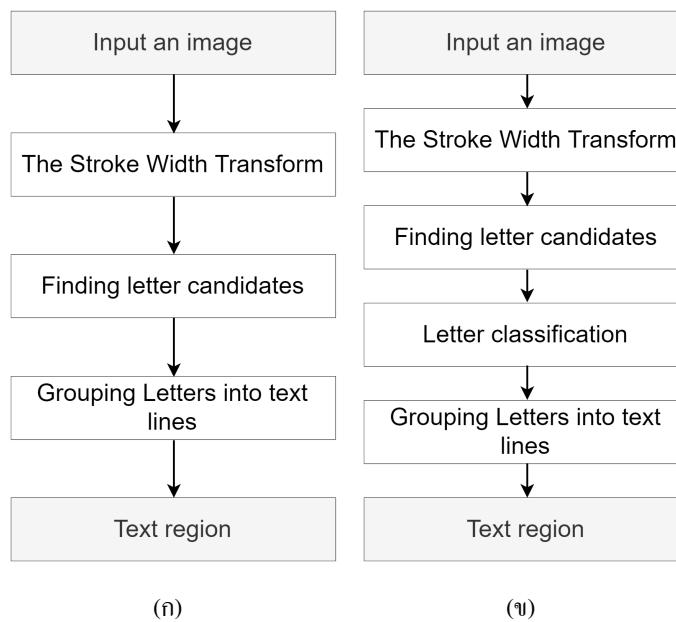


รูปที่ 3.1 ตัวอย่างผลลัพธ์จากการตรวจหาข้อความบนภาพมังงะด้วยวิธีต้นฉบับ [1] แสดงให้เห็น False Positive จำนวนมาก (ก) นักวาด: Shinoasa (ว) นักวาด: Kousei (Public Planet)

สภาพลักษณะเฉพาะของอักษรในมังงะ โดยกลุ่มกล่าวถูกดังแบ่งให้อยู่ในรูปแบบสมการ 3.1

$$f(d, h, w, \tilde{s}) = \begin{cases} 1, & \text{if } 1 < \frac{d}{\tilde{s}} < 15 \text{ and } \tilde{s} \leq 80 \text{ and} \\ & 5 < h, w < 50 \\ 0, & \text{otherwise,} \end{cases} \quad (3.1)$$

โดยตัวแปรใหม่ที่ถูกเพิ่มเข้ามาคือ  $w$  ซึ่งคือ ค่าความกว้างของวัตถุนั้น ๆ เมื่อเราคำนวณวัตถุที่ไม่มีลักษณะคล้ายอักษรออกไปแล้ว เราจะได้กลุ่มของวัตถุที่มีลักษณะใกล้เคียงอักษร อย่างที่แสดงในภาพ 3.3 โดยในขั้นตอนนี้ของวิธีการ SWT ต้นฉบับไม่สามารถตรวจจับวัตถุที่คล้ายคลึงอักษรได้ครบถ้วนเพียงพอตามที่แสดงให้เห็นในภาพ 3.3(ก) แต่ก็ใหม่ของเราระบบที่ถูกปรับปรุงแล้วนั้นสามารถตรวจจับวัตถุที่คล้ายอักษรได้ครอบคลุมมากขึ้น อย่างไรก็ได้ก็เกณฑ์ใหม่นั้นสร้าง False Positive ที่มากขึ้นตาม ซึ่งมากกว่าผลลัพธ์จากกลุ่มเกณฑ์ดั้งเดิมของ SWT ต้นฉบับ อย่างไรก็ได้ปัญหาดังกล่าวจะถูกแก้ไขด้วยขั้นตอนคัดแยกอักษรด้วย SVM ต่อไป ซึ่ง SVM จะทำหน้าที่กำจัด False Positive ออกไป



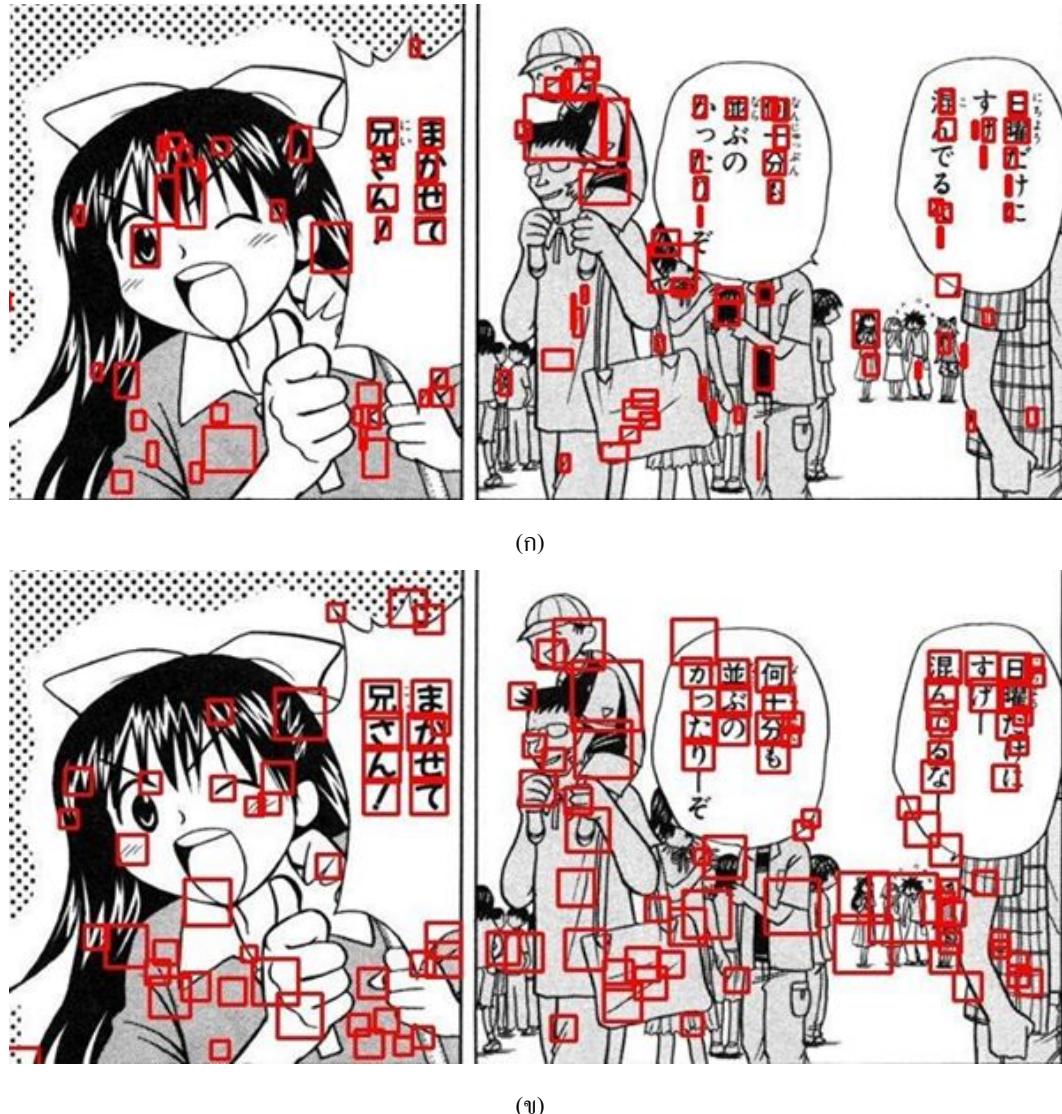
รูปที่ 3.2 แผนผังการทำงานของ (ก) วิธีการดึงเดิม [1] และ (ข) วิธีการใหม่ของเรา

### 3.1.3 คัดแยกอักษรด้วย SVM

ในขั้นตอนนี้เราสร้างภาพนาดเล็ก (Patch) จากภาพมังงะ Input โดยพิ่งพาตำแหน่งและขนาดของวัตถุที่คล้ายคลึงอักษรจากขั้นตอนก่อนหน้าในการสร้างขอบเขตของภาพนาดเล็กนั้น ๆ โดยภาพนาดเล็กเหล่านี้จะถูกคัดแยกเป็นกลุ่มที่เป็นอักษรและกลุ่มที่ไม่ใช่ตัวอักษรด้วย SVM ซึ่งจะช่วยลด False Positive ให้ต่ำลงทำให้ได้ผลลัพธ์การตรวจหาข้อความที่แม่นยำมากขึ้น

เราได้นำ SVM มาดำเนินการในขั้นตอนนี้โดย SVM คือ เทคนิค Supervised Learning แบบหนึ่ง ซึ่งมักถูกใช้ในงานด้านคัดแยก (Classification) และ สมการถดถอยต่อเนื่อง (Regression) [20] สำหรับ ชุดข้อมูลสำหรับเทรนโมเดลของ SVM ในขั้นตอนนี้จะถูกสร้างจากภาพหน้าเด็ก หรือ Patch โดยแบ่ง ออกเป็น ภาพที่เป็นอักษร (Positive) และ ภาพที่ไม่เป็นอักษร (Negative) อย่างที่แสดงในภาพ 3.4 ภาพ หน้าเด็กสำหรับเทรนนิ่งและทดสอบเหล่านี้สร้างจาก Manga109 สำหรับภาพ Positive และ Negative ที่สร้างขึ้นมาจะถูกนำไปสกัดลักษณะเด่นหรือ Feature ด้วย Histogram of Oriented Gradients หรือ HOG [14] ซึ่งเป็นเทคนิคการสกัดข้อมูลเชิงรูป่างของวัตถุในภาพด้วยการพิ่งพารากระกระจายตัวของ ทิศทางโgnus โดยลักษณะเด่นที่ถูกสกัดมาสำหรับใช้งานในงานวิจัยนี้นั้นเป็นข้อมูลรูปแบบเวกเตอร์ 2,916-dimension

เช่นเดียวกับข้อมูลสำหรับ tren ภาพขนาดเล็กของวัตถุที่คัดลักษณะอักษรจากขั้นตอนก่อนหน้าจะถูก HOG สกัดลักษณะเด่นออกมานำสู่ SVM ดำเนินการคัดแยกภาพที่เป็นอักษรและไม่ใช่อักษรออกจากกัน หลังจากเสร็จสิ้นกระบวนการ ส่วนภาพที่เป็นอักษรจะถูกนำสู่การจัดกลุ่มอักษรให้เกิดเป็นข้อความในขั้นตอนต่อไป



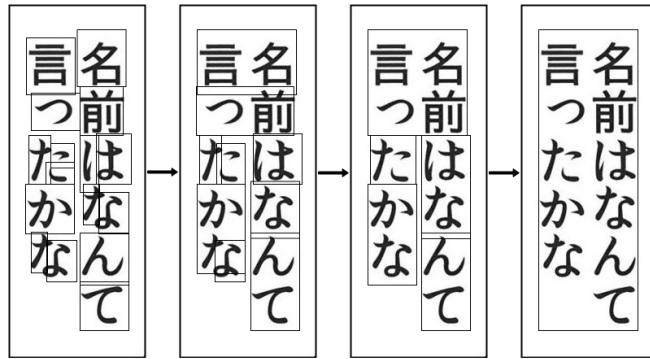
รูปที่ 3.3 ตัวอย่างแสดงการเปรียบเทียบผลลัพธ์ระหว่างข้อมูลตัวอักษรที่ตรวจพบระหว่างการใช้กลุ่มที่เก่าของ SWT ด้านบน (n) และกลุ่มที่ใหม่ในวิธีของเรา (u) ข้อมูลภาพถูกนำมารีอิง Arisa ©Yagami Ken



รูปที่ 3.4 ตัวอย่างของ Patch: (g) ภาพ Positive Patches and (u) ภาพ Negative Patches

#### 3.1.4 จัดกลุ่มอักษรเป็นข้อความ

ในวิธีการด้านบน [1] ขั้นตอนจัดกลุ่มวัตถุที่คล้ายคลึงอักษรเข้าด้วยกันเป็นประโยชน์หรือบรรทัดของข้อความ ได้ใช้หลักการเปรียบเทียบความคล้ายคลึงลักษณะต่าง ๆ ของตัวอักษร ประกอบไปด้วย



รูปที่ 3.5 ตัวอย่างแสดงการจับกลุ่มของตัวอักษร

ของความสูง ขนาดเส้น ทิศทาง และ ระยะห่าง โดยวัตถุที่ไม่ถูกจับคู่จะถูกกำจัดทิ้งไป กระบวนการนี้ใช้สมมติฐานว่าประโภคหรือข้อความมักเกิดจาก การรวมตัวกันของอักษรมากกว่าหนึ่งตัวและจัดเรียงอยู่ในทิศทางเดียวกันกับตัวอักษรอื่น ๆ ที่ขนาดใกล้เคียงกันตามที่ได้กล่าวไว้ใน 2 ขั้นตอนนี้ช่วยกำจัดข้อมูลรบกวนอื่น ๆ เช่น วัตถุที่ไม่ใช้อักษรที่กระจายอยู่ในภาพ แต่อย่างไรก็ดี วิธีการของเรานี้ได้กำจัดข้อมูลรบกวนเหล่านี้ออกไปแล้วในขั้นตอนการคัดแยกอักษรด้วย SVM ดังนั้นเราจึงใช้เพียงระยะห่างระหว่างอักษรเป็นปัจจัยในการจับกลุ่มอักษร

วิธีการจัดกลุ่มอักษรของเรานี้จะใช้อักษรที่ถูกคัดแยกด้วย SVM และมาจัดกลุ่มเป็นประโภคโดยการจัดกลุ่มแต่ละอักษรที่อยู่ห่างกันไม่เกิน 1.5 เท่าของตัวอักษรที่แคนที่สุดของคู่อักษรที่ใช้เปรียบเทียบ หากตัวอักษรใดที่ห่างจากกันเกินกว่าค่าที่กำหนดจะถือว่าเป็นอักษรของคนละประโภคซึ่งจะไม่ถูกจับกลุ่มเข้ามา โดยตัวอย่างในภาพ 3.5 แสดงถึงตัวอย่างขั้นตอนการจับกลุ่มด้วยระยะห่าง นอกจากนี้หลังจากการจัดกลุ่มแต่ละประโภคด้วยระยะห่างเสร็จสิ้นแล้ว กลุ่มอักษรเหล่านี้ถูกนำมาพิจารณาขนาดด้วยเช่นกัน โดยแต่ละกลุ่มของอักษรหรือแต่ละข้อความต้องมีพื้นที่ (ความกว้าง  $\times$  ความสูง) เกินกว่า 2,550px อ้างอิงจากการทดลองกับชุดข้อมูลของเรา สุดท้ายเราจะได้กลุ่มของอักษรหรือประโภคข้อความจากภาพในมังงะอ กามา

### 3.2 ชุดข้อมูลสำหรับการเทรนโมเดล SVM

เราได้นำภาพจากชุดข้อมูลภาพมังงะขนาดใหญ่ Manga109 [9] ประกอบไปด้วยภาพมังงะพร้อมข้อมูลประกอบ หรือ Annotation ของมังงะ 109 เรื่อง จัดทำโดยห้องทดลอง Aizawa Yamasaki แห่งมหาวิทยาลัยโตเกียว มังงะทั้งหมดในชุดข้อมูลนี้ถูกวัดโดยนักความมั่งจะมีอาชีพชาวญี่ปุ่นและถูกจัดจำแน่นายในช่วงปี 1970 ถึงปี 2010 แต่ละหน้าของมังงะถูกระบุตำแหน่งของข้อความในภาพซึ่งข้อมูลดังกล่าวเป็นเครื่องมือแก่การใช้トレนโมเดลและทดสอบวิธีการของเรา

### 3.3 การทดลอง

เราดำเนินการทดลองในรูปแบบเดียวกับงานวิจัยของ Aramaki et al. [12] ซึ่งจะทำให้เราสามารถเปรียบเทียบผลการทดลองประสิทธิภาพวิธีการของเรากับงานวิจัยอื่น ๆ ที่เกี่ยวข้องกับการตรวจหา

ข้อความในภาพมังงะซึ่งเคยถูกทดสอบมาก่อนหน้าแล้วได้ เราได้เลือกภาพมังงะด้วยวิธีการสุ่มเลือก 100 หน้าสำหรับการเทรน และอีก 100 หน้าสำหรับทดสอบประสิทธิภาพ โดยภาพมังงะทั้งหมดนี้ถูกสุ่มเลือกจากมังงะ 6 เรื่อง ได้แก่ *Aosugiru Haru, Arisa 2, Bakuretsu Kung Fu Girl, Dollgun, Love Hina, และ Uchiha Akatsuki EvaLady*

เนื่องจากวิธีการของเราราใช้ SVM ซึ่งต้องสร้างโมเดลสำหรับใช้งานคัดแยกภาพระหว่างภาพขนาดเล็ก (Patch) ระหว่างกลุ่มที่เป็นอักษรและไม่ใช้อักษรตามที่แสดงไปในภาพ 3.4 เราจึงได้สร้างชุดข้อมูลประกอบด้วยภาพอักษร 5,201 ภาพ และ ภาพขนาดเล็กอีก 6,75 ของวัตถุที่ไม่ใช้อักษรอีก 5,201 ภาพ กล่าวคือแบ่งเป็นข้อมูล Positive และ Negative ส่วนละ 50% เท่า ๆ กัน โดยภาพเหล่านี้สร้างจากภาพสำหรับการเทรนที่ได้ถูกกล่าวไว้ในย่อหน้าก่อนหน้า โดยใช้ขั้นตอน ค้นหาวัตถุที่ใกล้เคียงอักษร ของเรานในการค้นหาตำแหน่งและขอบเขตของวัตถุที่คล้ายคลึงอักษรและนำตำแหน่งเหล่านั้นสร้างภาพขนาดเล็กเหล่านี้ออกมาน

สำหรับ SVM เราใช้ Radial Basis Function Kernel โดย Hyperparameter ที่ร่วมใช้งานประกอบไปด้วย  $C$  และ  $\gamma$  เราใช้ Grid Search บนเครื่องคอมพิวเตอร์ *Google Cloud Compute Engine n1-highcpu-8* ในการค้นหาค่า Hyperparameter ที่ดีที่สุด ในช่วง  $2^{-10}$  ถึง  $2^{10}$  โดยได้ค่า  $C$  และ  $\gamma$  ที่ดีที่สุดที่  $2^5$  และ  $2^{-6.75}$  ตามลำดับ โมเดลที่ถูกปรับปรุงให้เหมาะสม (Optimized) และนี้จะถูกนำไปใช้ในขั้นตอนคัดแยกอักษรด้วย SVM

การประเมินวิธีการของเราราที่ถูกพัฒนาขึ้นใหม่นี้ได้ใช้รูปแบบการประเมินเดียวกับที่ใช้ใน ICDAR 2013 Robust Reading Competition [21] โดยถ้าอัตราส่วนระหว่างพื้นที่ Overlapped ต่อ พื้นที่ Ground-truth นั้นมากกว่าค่า  $t_p$  และอัตราส่วนระหว่างพื้นที่ Overlapped ต่อ พื้นที่ Detected Region มากกว่า  $t_r$  ให้ถือว่า พื้นที่ขอบเขตอักษรที่ถูกตรวจพบจากวิธีการของเรานั้นถูกต้อง โดย  $t_p$  และ  $t_r$  นั้นมีค่าเท่ากับ 0.5 ถ้างบค่าดังกล่าวตามงานวิจัยของ Aramaki et al. [12] สำหรับ Precision และ Recall เราได้คำนวณตามสมการดังต่อไปนี้ 3.2, 3.3 ตามลำดับ

$$P = \frac{\text{#Correctly Detected Rectangles}}{\text{#Detected Rectangles}} \quad (3.2)$$

$$R = \frac{\text{#Correctly Detected Rectangles}}{\text{#Rectangles of the Ground-truth}} \quad (3.3)$$

สำหรับ F-Measure เราคำนวนด้วยสมการดังนี้ 3.4

$$F = 2 \cdot \frac{P \cdot R}{P + R} \quad (3.4)$$

เราได้เปรียบเทียบผลการทดลองของวิธีการใหม่ของเราร่วมกับวิธีการต้นฉบับ [20] ซึ่งถูกใช้กับภาพถ่ายนอกจากราชีพเปรียบเทียบกับงานวิจัยการตรวจหาข้อความอื่น ๆ ด้วยดังนี้ Basic Grouping+ImageNet Classification Model (BG+ImN) [12], Basic Grouping+Illustration2Vec Model (BG+I2V) [12], Scene Text Detection (STD) [22], Speech Balloon Detection (SBD) [10], และ Text Line Detection (TLD) [23] วิธีข้างต้นที่กล่าวถึงมีเทคนิคในการตรวจหาข้อความที่หลากหลายแตกต่างกัน เช่น การยึดหลักสมมติฐาน

พื้นฐาน (ทิศทางของข้อความ, รูปแบบการจัดวาง, ลักษณะของกล่องคำพูด) และ Convolutional neural network

สำหรับ BG+ImN, BG+I2V, STD, SBD, และ TLD นั้นเราได้นำผลลัพธ์การทดลองจากงานวิจัยของ Aramaki et al. [12] มาใช้ในการเปรียบของเราโดยตรง ซึ่งการเปรียบเทียบโดยตรงนี้สามารถทำได้เนื่องจากเราได้ดำเนินการทดลองในสภาพแวดล้อมเดียวกับงานวิจัยดังกล่าว โดยผลลัพธ์การเปรียบเทียบและตัวอย่างขอบเขตข้อความที่วิธีการของเราที่นี่สามารถตรวจสอบลูกแลดงให้บทที่ 4

## บทที่ 4

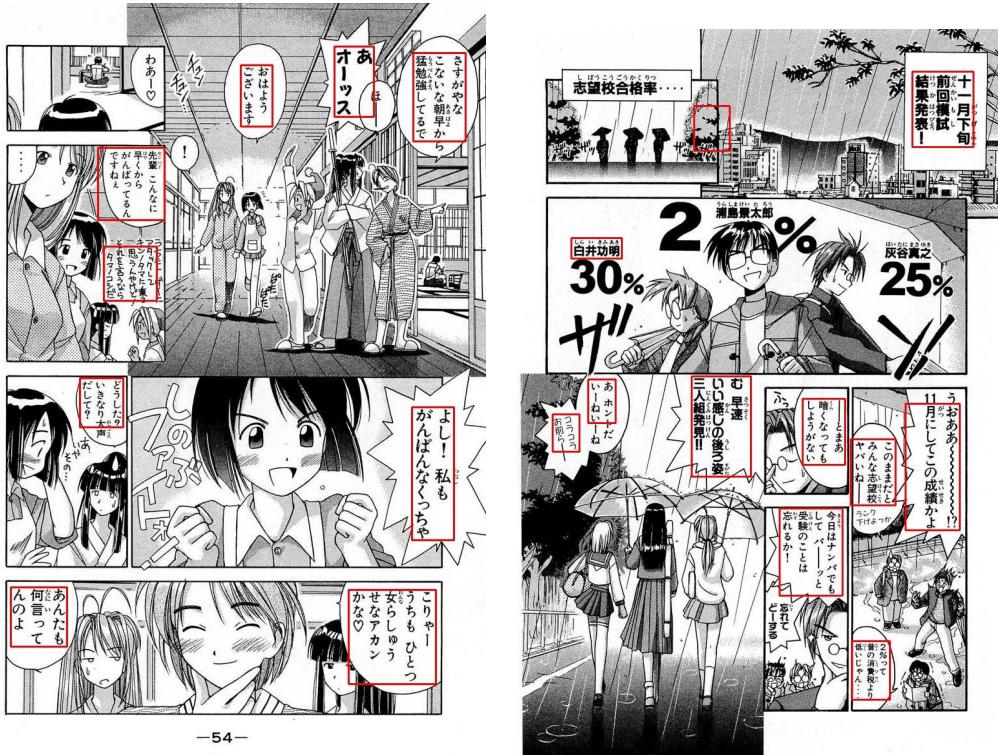
### ผลการทดลอง

ผลลัพธ์การทดสอบประสิทธิภาพในวิธีการใหม่ของเราและเปรียบเทียบกับงานวิจัยอื่น ๆ ที่เกี่ยวข้องแสดงในตารางที่ 4.1 จากรายงานดังกล่าว วิธีการของเราได้รับ F-measure สูงที่สุด ที่ 0.506 ซึ่งแสดงให้เห็นชัดเจนว่าวิธีการของเรา มีประสิทธิภาพดีกว่าวิธีการ SWT ต้นฉบับ [1] ยิ่งไปกว่านั้นวิธีการของเรา ยังได้รับ F-measure ที่มากกว่าวิธีการ BG+ImN [12] และ BG+I2V [12] ซึ่งทั้งสองวิธีนี้ใช้เทคนิค Deep Learning เป็นส่วนหนึ่งในการตรวจหาข้อความในภาพอย่างไรก็ได้ค่า Precision และ Recall สูงสุดของ การทดลองนี้อยู่ที่ 0.715 และ 0.481 เป็นของ BG+I2V และ BG+ImN ตามลำดับ สำหรับตัวอย่างพื้นที่ของข้อความที่วิธีการใหม่ของเราระบุถูกแสดงในภาพ 4.1

Method	Precision	Recall	F-measure
STD [22]	0.165	0.051	0.078
SBD [10]	0.180	0.102	0.130
TLD [23]	0.095	0.095	0.095
BG + ImN [12]	0.451	<b>0.481</b>	0.466
BG + I2V [12]	<b>0.715</b>	0.191	0.301
Baseline [1]	0.068	0.336	0.113
วิธีการของเรา	0.564	0.458	<b>0.506</b>

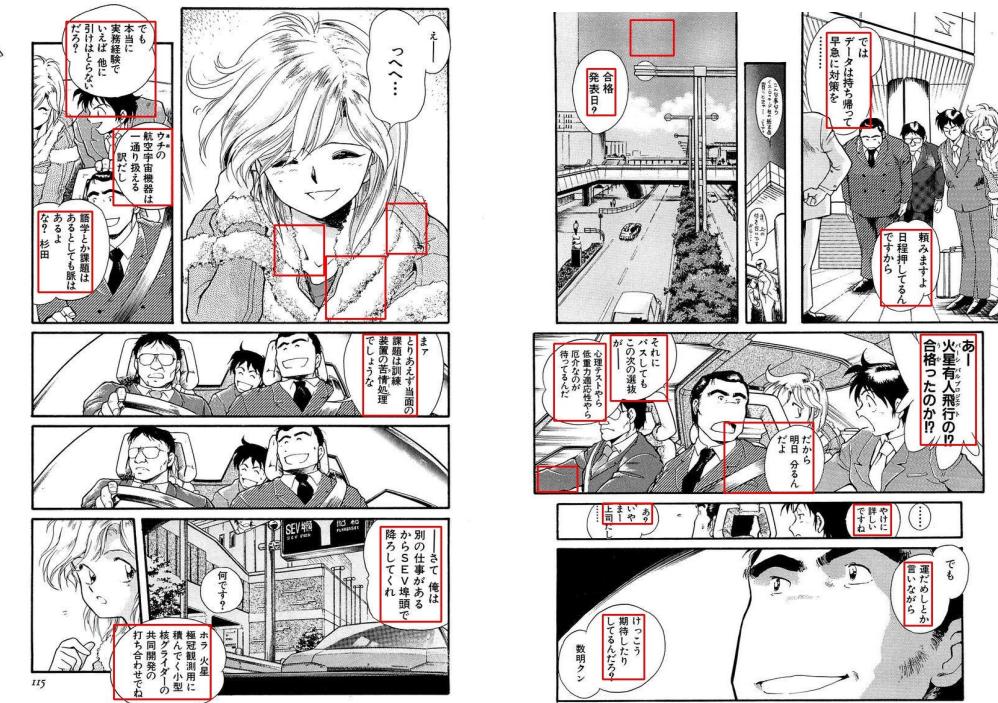
ตารางที่ 4.1 ตารางแสดงการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ของเราร่วมกับวิธีการอื่น ๆ ที่เกี่ยวข้อง

เป็นที่น่าสนใจอย่างมากที่วิธีการของเราสามารถทำงานได้ดีกว่าเทคนิค Deep Learning ทั้งสองวิธี สมมติฐานแรกคือ BG+ImN นั้นใช้ ImageNet Classification Model [24] ซึ่งถูกтренนนภาพถ่ายของวัตถุจริง อย่างไรก็ได้ภาพมักจะของวัตถุต่าง ๆ นั้นมีความแตกต่างจากภาพวัตถุจริงอย่างชัดเจนซึ่ง ณ จุดนี้ทำให้วิธีการนี้ไม่สามารถทำงานได้เต็มประสิทธิภาพ อีกวิธีการที่ใช้ Deep Learning คือ BG+I2V ถึงแม้ว่าวิธีการนี้จะได้รับ Precision สูงที่สุดในการทดลองของเราแต่คะแนน Recall นั้นต่ำกว่าทั้ง BG+ImN และวิธีการของเรา วิธีการนี้ใช้โมเดล Illustration2Vec [25] เป็นโมเดลสำหรับคัดแยกข้อความจากวัตถุ อื่น ๆ ที่ปรากฏในภาพมังงะ ส่วนของโมเดลนั้นถูก trenn ภาพวาด Anime (ภาพการ์ตูนแบบญี่ปุ่น) และภาพมังงะจากหลากหลายแหล่ง อันประกอบไปด้วย Danbooru และ Safebooru ซึ่งมีลักษณะงานคล้ายกับข้อมูลที่เราใช้ทดสอบในวิธีการของเรา แต่โมเดลนี้ถูกออกแบบมาเพื่อการทำงานป้ายกำกับ (Tag Prediction) และค้นหาภาพที่คล้ายคลึงกัน ดังนั้นการเลือกใช้วิธีการที่ไม่ต้องลักษณะโดยตรง ในลักษณะนี้จึงอาจเป็นเหตุผลว่าทำไมโมเดลนี้จึงไม่มีประสิทธิภาพเท่าที่ควรในการทดลองนี้



—54—

(v)



(ก)

(۹)

รูปที่ 4.1 ตัวอย่างขอบเขตข้อความที่วิธีการของเราราตรวจ (ก-ป) Love Hina ©Ken Akamatsu และ (ค-ง) Eva Lady ©Miyone Shi

## บทที่ 5

# สรุปผล

ในการทดลองนี้ เราได้เสนอวิธีการตรวจหาข้อความบนภาพมังงะด้วยเทคนิค SWT ร่วมกับการใช้ SVM และลักษณะเด่น (Feature) HOG ในการลด False Positive ที่เกิดขึ้น การทดลองของเรามาดำเนินการบนชุดข้อมูล Manga109 ซึ่งเป็นชุดข้อมูลภาพมังงะที่มีกรอบ Annotation ที่เกี่ยวข้องมาเรียบร้อยแล้ว วิธีการของเรานั้นสามารถทำงานได้ผลลัพธ์ F-measure ที่ดีที่สุดในการเปรียบเทียบกับวิธี Baseline และ วิธีการอื่น ๆ ที่เกี่ยวข้องรวมถึงวิธีการที่ใช้ Deep Learning เป็นส่วนประกอบในการคัดแยกอักษรหรือ ข้อความจากวัตถุอื่น ๆ ในภาพ ถึงแม้ว่างานของเรานั้นได้ทดสอบบนการ์ดูนญี่ปุ่นสามารถทำงานได้เป็น ผลดีเยี่ยม แต่วิธีการของเรานั้นยังต้องมีการพัฒนาและปรับปรุงเพิ่มเติมเพื่อปรับปรุงประสิทธิภาพและ ทำให้สามารถใช้งานร่วมกับภาษาอื่น ๆ ได้

## បរទេសក្រុម

- [1] B. Epshtain, E. Ofek, and Y. Wexler, “**Detecting text in natural scenes with stroke width transform**,” in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, CA, USA, Jun 2010, pp. 2963–2970.
- [2] H. Yanagisawa, T. Yamashita, and H. Watanabe, “**A study on object detection method from manga images using CNN**,” in *Proceedings of the International Workshop on Advanced Image Technology (IWAIT 2018)*, Chiang Mai, Thailand., Jan 2018, pp. 1–4.
- [3] X. Liu, C. Li, H. Zhu, T.-T. Wong, and X. Xu, “**Text-aware balloon extraction from manga**,” *The Visual Computer*, vol. 32, no. 4, pp. 501–511, Apr 2016. [Online]. Available: <https://doi.org/10.1007/s00371-015-1084-0>
- [4] X. Pang, Y. Cao, R. W. Lau, and A. B. Chan, “**A Robust Panel Extraction Method for Manga**,” in *Proceedings of the 22nd ACM International Conference on Multimedia (MM 2014)*. New York, NY, USA: ACM, 2014, pp. 1125–1128. [Online]. Available: <http://doi.acm.org/10.1145/2647868.2654990>
- [5] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “**Interactive segmentation for manga using lossless thinning and coarse labeling**,” in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA 2015)*, Hung Hom, Kowloon, Hong Kong, Dec 2015, pp. 293–296.
- [6] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, and K. Aizawa, “**Object Detection for Comics using Manga109 Annotations**,” *CoRR*, vol. abs/1803.08670, 2018. [Online]. Available: <http://arxiv.org/abs/1803.08670>
- [7] S. Kovánen and K. Aizawa, “**A layered method for determining manga text bubble reading order**,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2015)*, Quebec City, QC, Canada, Sept 2015, pp. 4283–4287.
- [8] Y. Matsui, T. Shiratori, and K. Aizawa, “**DrawFromDrawings: 2D Drawing Assistance via Stroke Interpolation with a Sketch Database**,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 7, pp. 1852–1862, 2017.
- [9] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, “**Sketch-based manga retrieval using manga109 dataset**,” *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 811–21 838, Oct 2017. [Online]. Available: <https://doi.org/10.1007/s11042-016-4020-z>

- [10] H. Tolle and K. Arai, “**Manga content extraction method for automatic mobile comic content creation**,” in *Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS 2013)*, Bali, Indonesia, Sept 2013, pp. 321–328.
- [11] C. Rigaud, T. Le, J. . Burie, J. Ogier, S. Ishimaru, M. Iwata, and K. Kise, “**Semi-automatic Text and Graphics Extraction of Manga Using Eye Tracking Information**,” in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, Santorini, Greece, April 2016, pp. 120–125.
- [12] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “**Text detection in manga by combining connected-component-based and region-based classifications**,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*, Phoenix, AZ, USA, Sept 2016, pp. 2901–2905.
- [13] J. Canny, “**A Computational Approach to Edge Detection**,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [14] W. T. Freeman, W. T. Freeman, M. Roth, and M. Roth, “**Orientation Histograms for Hand Gesture Recognition**,” in *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 1994, pp. 296–301. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.618>
- [15] S. Bougarriou, F. Hamdaoui, and A. Mtibaa, “**Linear SVM classifier based HOG car detection**,” in *Proceedings of the 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA 2017)*, Monastir, Tunisia, Dec 2017, pp. 241–245.
- [16] N. Dalal and B. Triggs, “**Histograms of oriented gradients for human detection**,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA, Jun 2005, pp. 886–893.
- [17] D. Wang, H. Wang, D. Zhang, J. Li, and D. Zhang, “**Robust Scene Text Recognition Using Sparse Coding based Features**,” *CoRR*, vol. abs/1512.08669, 2015. [Online]. Available: <http://arxiv.org/abs/1512.08669>
- [18] S. Tian, S. Lu, B. Su, and C. L. Tan, “**Scene Text Recognition Using Co-occurrence of Histogram of Oriented Gradients**,” in *2013 12th International Conference on Document Analysis and Recognition*, Washington, DC, USA, Aug 2013, pp. 912–916.

- [19] A. K. Sah, S. Bhowmik, S. Malakar, R. Sarkar, E. Kavallieratou, and N. Vasilopoulos, “**Text and non-text recognition using modified HOG descriptor**,” in *2017 IEEE Calcutta Conference (CALCON)*, Dec 2017, pp. 64–68.
- [20] J. Suykens and J. Vandewalle, “**Least Squares Support Vector Machine Classifiers**,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, Jun 1999. [Online]. Available: <https://doi.org/10.1023/A:1018628609742>
- [21] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. de las Heras, “**ICDAR 2013 Robust Reading Competition**,” in *2013 12th International Conference on Document Analysis and Recognition*, Kolkata, India, Aug 2013, pp. 1484–1493.
- [22] L. Gómez and D. Karatzas, “**Multi-script Text Extraction from Natural Scenes**,” in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, Washington, DC, USA, Aug 2013, pp. 467–471.
- [23] C. Rigaud, D. Karatzas, J. Van De Weijer, J.-C. Burie, and J.-M. Ogier, “**Automatic text localisation in scanned comic books**,” in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, Barcelona, Spain, Feb 2013. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00841492>
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “**ImageNet Classification with Deep Convolutional Neural Networks**,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012)*, vol. 1. Lake Tahoe, Nevada, USA: Curran Associates Inc., Dec 2012, pp. 1097–1105. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999257>
- [25] M. Saito and Y. Matsui, “**Illustration2Vec: A Semantic Vector Representation of Illustrations**,” in *SIGGRAPH Asia 2015 Technical Briefs*. Kobe, Japan: ACM, Nov 2015, pp. 5:1–5:4. [Online]. Available: <http://doi.acm.org/10.1145/2820903.2820907>

## ภาคผนวก ก

## การใช้ชีวิตในประเทศไทยญี่ปุ่น

สำหรับการฝึกงานร่วมกับมหาวิทยาลัยออกไกโดในประเทศไทยญี่ปุ่น ได้ทำการฝึกงานรวมกันเป็นระยะสี่เดือนกับอีกสิบวัน โดยได้เข้ามาทำงานในห้องทดลอง Intelligent Information System ภายใต้การดูแลของอาจารย์ประจำห้องทดลอง Professor Dr. Masanori Sugimoto และในช่วงเวลาฝึกงานนี้ ยังได้รับความช่วยเหลือจากสมาชิกภายในห้องทดลองอีกท่าน Jiang Ye นักศึกษาปริญญาเอก ปี 2 จากประเทศจีน ได้ให้ความช่วยเหลือและการดูแลในเรื่องการดำเนินการเอกสารที่จำเป็นต่าง ๆ เช่น เอกสารการขยายน้ำมันสำเร็จรูปและเอกสารดูแลในเรื่องการดำเนินการเอกสารที่จำเป็นต่าง ๆ เช่น เอกสารการขยายน้ำมันสำเร็จรูปและการขยายน้ำมันสำเร็จรูป นอกจากนี้ยังได้รับความช่วยเหลือในเรื่องความเป็นอยู่และสิ่งของทั่วไปในการใช้ชีวิตจากทั้งกลุ่มนักศึกษาชาวไทยและชาวต่างชาติ ซึ่งล้วนศึกษาอยู่ที่มหาวิทยาลัยออกไกโด

การใช้ชีวิตในต่างแดน เช่น ประเทศไทยญี่ปุ่นนั้นมีอุปสรรคหลายอย่าง ปัจจุบันประการหนึ่งที่ประสบบ่อยครั้งที่สุดคือ เรื่องการสื่อสารภาษาญี่ปุ่น อย่างที่ทราบกันดีว่าชาวญี่ปุ่นส่วนใหญ่ไม่สามารถสนทนาภาษาอังกฤษได้ ดังนั้นจึงมีความจำเป็นต้องเรียนรู้ภาษาญี่ปุ่นพื้นฐานเพื่อใช้ในการดำเนินชีวิตประจำวัน รวมถึงสมาชิกภายในห้องทดลองที่ส่วนใหญ่ไม่สามารถสื่อสารด้วยภาษาอังกฤษ เช่น กัน สมาชิกภายในห้องทดลองส่วนมากจะสามารถสื่อสารได้เพียงประโภคพื้นฐาน ไม่สามารถสนทนาเป็นบทสนทนาที่ยาวหรือเป็นการเล่าเรื่องได้ แต่อย่างไรก็ได้ พากษาสามารถอ่านและเขียนภาษาอังกฤษได้ค่อนข้างดี ดังนั้นเมื่อมีปัญหาในการพูดคุยจึงใช้การเขียนหรือพิมพ์ทคแทน อย่างไรก็ได้ภาษาในห้องทดลองก็มีสมาชิกชาวญี่ปุ่นบางคนที่มีความสามารถภาษาอังกฤษในขั้นดีมาก อย่างเช่นนักศึกษาปริญญาโทชาวญี่ปุ่นท่านหนึ่งที่เคยได้ฝึกงานในนิวซีแลนด์และฟิลิปปินส์มาก่อนหน้านี้ ทำให้สามารถสื่อสารภาษาอังกฤษได้เป็นอย่างดี

นอกจากตัวผู้ที่เป็นนักศึกษาฝึกงานจากไทย ภายในห้องทดลองที่มีฝึกงานก็ยังมีนักศึกษาต่างชาติอีกสองคน คนแรกอย่างที่กล่าวไปในตอนต้น คือ นักศึกษาปริญญาเอกจากประเทศไทย และคนที่สองคือนักศึกษาปริญญาโทจากบรัสเซล สำหรับชาวจีนนี้ใช้ภาษาญี่ปุ่นเป็นภาษาหลักและสามารถสื่อสารกับคนญี่ปุ่นในห้องทดลองได้อย่างราบรื่น และยังพูดภาษาอังกฤษได้อีกด้วย

### ก.1 ที่อยู่อาศัย

หอพักที่ใช้อาศัยตลอดโครงการฝึกงานนี้ถูกจัดทำให้โดยทางมหาวิทยาลัยออกไกโด หอพักมีชื่อว่า International House Kita 8 East เป็นหอพักชั้นต้น สำหรับนักศึกษาชาวต่างชาติเท่านั้น โดยภายในห้องจะมีเพียงตู้, เตียง, โต๊ะ, คอมไฟ, ไฟเพดาน, ตู้เย็น, ฮีทเตอร์, และลังขยะ อย่างที่แสดงในภาพ ก.1(ก) และ ก.1(ข) สำหรับห้องน้ำ และห้องซักรีดต้องใช้ของส่วนกลางเท่านั้น โดยจะจัดแยกในแต่ละชั้น สำหรับให้ผู้อยู่อาศัยใช้ร่วมกัน

สำหรับห้องครัวและห้องรับประทานอาหารมีจัดเตรียมให้ที่ชั้นแรกของหอพักโดยต้องใช้ร่วมกัน ทั้งหอพัก เตาแก๊สและอ่างล้างจานถูกติดตั้งไว้เรียบร้อยสามารถใช้งานได้ในทันที สำหรับของใช้ส่วนตัว รวมถึงเครื่องครัวผู้อยู่อาศัยต้องหาซื้อด้วยตนเอง ในกรณีของผู้มีความจำเป็นซึ่งมีค่าใช้จ่ายมากขอน

ชื่อและแก้ว สำหรับอุปกรณ์ทำอาหารได้รับมาจากนักศึกษาชาวจีนที่กำลังจะกลับประเทศจีน ภาพห้องครัวแสดงในภาพ ก.1(ก) และ ก.1(ง)



(ก) ห้องนอน



(ข) ห้องนอน



(ค) ห้องอาหาร



(ง) ห้องครัว

รูปที่ ก.1 ภาพหอพัก International House Kita 8 East

ภาคผนวก ๖

## กิจกรรมระหว่างฝึกงาน

นอกจากการทำโปรเจกแล้วนั้น ระหว่างสี่เดือนนี้ผู้ฝึกงานได้ร่วมกิจกรรมด้านวิชาการต่าง ๆ ดังเช่น การประชุมห้องทดลองรายสัปดาห์ และ การร่วมนำเสนอผลงานระหว่างห้องทดลอง เป็นต้น โดยรายละเอียดจะกล่าวต่อจากนี้

### **๔.๑ Mirai Symposium**

กิจกรรมนี้จัดขึ้นที่เริ่วักงหรือสอร์ทแบบญี่ปุ่น ซึ่งมีอ่อนเช็นหรือบอน้ำพุร้อนให้ได้แช่อก็ด้วยค่าใช้จ่ายห้องน้ำเดียว ห้องทดลองออกให้ห้องน้ำเดียวในงานจะมีกิจกรรมหลักคือการให้นักศึกษาทุกคนของห้องทดลองต่าง ๆ ที่มาร่วมงานนำผลงานตนเองมานำเสนอด้วยโปสเตรอร์ในห้องประชุม หมุนเวียนไปเป็นช่วงเวลาคนละ 30 นาที หากสนใจงานของครุภัณฑ์สามารถเดินไปสอบถามได้ อย่างที่เห็นในภาพ ๔.๑(ก) การนำเสนอไม่ได้จำกัดว่างานต้องเป็นผลงานที่เสร็จสิ้นแล้ว อาจเป็นความก้าวหน้าของงานที่พัฒนาอยู่ได้ เช่นเดียวกัน สำหรับครั้งนี้มีห้องทดลองร่วมงานสามห้องทดลอง สำหรับนักศึกษาต่างชาติจะมีนักศึกษาท่านอื่นเข้ามาสอบถามเพียงเล็กน้อยเนื่องจากต้องสื่อสารด้วยภาษาอังกฤษและนักศึกษาญี่ปุ่นส่วนใหญ่ไม่สามารถพูดอังกฤษได้ สำหรับงานวิจัยที่ผู้สอนนำก็ได้นำไปนำเสนอในงานนี้ด้วยเช่นเดียวกัน

เมื่อช่วงนำเสนอเสร็จสิ้น ทุกคนจะได้พักผ่อนตามอัธยาศัยโดยส่วนใหญ่จะเลือกไปแช่น้ำพุร้อนกัน สำหรับบ่อน้ำพุร้อนก็มีทั้งภายในออกและภายในอาคาร จากนั้นเป็นมื้อเย็นซึ่งเป็นอาหารชุดญี่ปุ่น ๔.๑(ข)

สำหรับวันที่สองก็มีกิจกรรมสร้างสรรค์ เป็นกิจกรรมที่ให้แยกกลุ่มกันและแก้โจทย์ปัญหาโดยให้แก่ภายในเวลาที่กำหนด ตัวอย่างคำถาม เช่น มีเครื่องยนต์ร้อนหรือเย็นมีทั้งที่หมายหน้าก้ออยและหัวอย่าง ลักษณะ ต้องการแบ่งกลุ่มหรือยุสลงกลุ่มโดยที่แต่ละกลุ่มนี้จำนวนหน้าหรือยกและก้ออยท่า ๆ กัน โดยที่ผู้แบ่งมองไม่เห็นหรือยุสจะทำได้อย่างไร เป็นต้น เมื่อเสร็จกิจกรรมสร้างสรรค์ก็ถือเป็นการจบกิจกรรม Mirai Symposium แต่เพียงเท่านี้และเดินทางกลับมหาวิทยาลัย

### **๔.๒ Lab Meeting**

สำหรับห้องทดลองที่ได้เข้ามาฝึกงานนั้นมีกิจกรรม Lab Meeting ทุกวันพุธ คือ กิจกรรมที่ให้สามารถภายในห้องทดลองร่วมกันนำเสนอผลงานวิจัยต่าง ๆ ที่ได้อ่านมานำเสนอต่อสมาชิกท่านอื่น ๆ ภายในห้องทดลอง โดยแต่ละสมาชิกจะถูกกำหนดวันเพื่อให้แต่ละคนที่ถูกกำหนดในวันนั้นนำ Conference Paper หรือ Journal ที่ถูกตีพิมพ์ในงานประชุมวิชาการ ขึ้นดังของโลกนาร่วมกันนำเสนอแบบสรุปพร้อมสไลด์ กิจกรรมนี้ไม่มีคะแนนหรือรางวัลใด ๆ แต่เป็นการแลกเปลี่ยนความรู้ที่เกี่ยวข้องกับงานของตนเองที่ทำอยู่

สำหรับผู้ได้มีโอกาสนำเสนอผลงานวิจัยที่ถูกตีพิมพ์ใน Journal เรื่อง *Text-aware balloon extraction*



(ก) บรรยายค่าคระหว่างการนำเสนอผลงานด้วยโปสเตอร์



(ข) ชุดอาหารจัดเลี้ยงมื้อเย็น



(ค) ห้องอาหารจัดเลี้ยง

### รูปที่ บ.1 ภาพกิจกรรมในงาน Mirai Symposium

from manga โดยหลังจากนำเสนอเรื่องลึกลับที่ได้รับคำชมและความคิดเห็นจากสมาชิกและอาจารย์ของห้องทดลองอย่างหลากหลาย ซึ่งถือเป็นโอกาสที่ดีในการเรียนรู้มุมมองและความคิดเห็นที่แปลกใหม่ต่องานของเรา

### บ.3 กิจกรรมอื่น ๆ

นอกจากกิจกรรมเชิงวิชาการแล้ว ผู้เข้าร่วมในกิจกรรมสังสรรค์อื่น ๆ เช่น งานเลี้ยงตามโอกาสต่าง ๆ โดยผู้มีส่วนได้ร่วมงานเลี้ยงต้อนรับตัวผู้ชั่งจัดในช่วงเดือนตุลาคมที่ผ่านมา สาเหตุที่จัดขึ้นจากเดือนที่เข้าฝึกงานวันแรกไปหลายเดือนเนื่องจากกำหนดการในตอนแรกนั้นคือหนึ่งเดือนหลังจากผู้มาอยู่ที่ประเทศญี่ปุ่น แต่ในช่วงกำหนดการเกิดเหตุการณ์แผ่นดินไหวใหญ่บันดาลใจให้ทำให้ต้องเลื่อนกำหนดการออกไป โดยงานเลี้ยงนี้จัดที่ร้านเนื้อย่างเจงกิสข่าน เป็นเนื้อแกะย่างบนเตาเหล็ก

ลักษณะคล้ายหนูกระทะในประเทศไทยแตกต่างกันเพียงหน้าไปที่เนื้อแกะเป็นหลัก ภาพกิจกรรมแสดงในภาพ ข.2



(ก) สมาชิกในห้องทดลองระหว่างกินเลี้ยง



(ข) โต๊ะอาหารในร้านอาหารเจงกิสخ่าน



(ค) กระบวนการรับประทานเจงกิสخ่าน

## รูปที่ ข.2 ภาพระหว่างงานเลี้ยงต้อนรับ

นอกจากงานเลี้ยงต้อนรับแล้วยังมีงานเลี้ยงต้อนรับนักศึกษาปีสามที่เข้าเป็นสมาชิกใหม่ในห้องทดลองนี้อีกด้วย ซึ่งงานนี้จัดรวมเป็นงานเลี้ยงสำหรับพ่อแม่ ฯ กันเนื่องจากจัดให้ล้วนสิ้นสุดการฝึกงาน งานเลี้ยงจัดในห้องทดลองพร้อมด้วยอาหารหลากหลายชนิด งานไม่มีกิจกรรมอะไรมากมายเป็นเพียงการกินเลี้ยงเพื่อทำความรู้จักกับสมาชิกใหม่ของห้องทดลองและนั่งเล่น侃คุยกันตามที่แสดงในภาพ ข.3

อธิบายเพิ่มเติมสำหรับงานเลี้ยงต้อนรับนักศึกษาปีสาม สำหรับห้องทดลองของคณะวิศวกรรมศาสตร์ที่ได้มารอยู่นี้จะมีการเปิดห้องทดลองให้นักศึกษาในคณะชั้นปีสามได้เข้าเยี่ยมชมห้องทดลองที่ตนเองสนใจก่อนจะเลือกเข้ามาเป็นสมาชิกในห้องทดลองที่เกี่ยวข้องกับงานที่สนใจจะทำ โดยงานที่สนใจจะทำจะจำกัดเป็นชื่นงานจากการศึกษาคล้ายกับในมหาวิทยาลัยของไทย โดยในช่วงเวลาหนึ่งปีแต่ละห้อง



(ก) อาหารต่าง ๆ ในงานเลี้ยง



(ข) สมาชิกในห้องทดลองระหว่างเล่นเกมในงานเลี้ยง

### รูปที่ ข.3 งานเลี้ยงอาหารและต้อนรับนักศึกษาปีสาม

ทดลองกิจกรรมการโฆษณาห้องทดลองของตัวเองและเปิดโอกาสให้เข้าเยี่ยมชมงานภายในห้องทดลองว่า ทำวิจัยเรื่องอะไร อย่างเช่นภาพ ข.4 ที่เป็นป้ายเชิญเข้าห้องทดลองด้านเสียงและการใช้ตัวละครจาก การ์ตูนเรื่อง Kemono Friend ประกอบให้น่าสนใจและดึงดูดมากขึ้น



รูปที่ ข.4 ป้ายเชิญชวนชมห้องทดลองโดยมีการใช้ตัวละครจากภาร์ตูนประกอบให้น่าสนใจมากขึ้น

**ភាគធនវក ៩**

**ผลงานวิจัยที่ได้รับการตีพิมพ์**

# Detecting Text in Manga using Stroke Width Transform

Boonyarith Piriyothinkul\*, Kitsuchart Pasupa<sup>†</sup>  
 Faculty of Information Technology  
 King Mongkut's Institute of Technology Ladkrabang  
 Bangkok 10520  
 Thailand  
 Email: \*58070077@kmitl.ac.th, <sup>†</sup>kitsuchart@it.kmitl.ac.th

**Abstract**—The Japanese comic-book style known as manga is becoming a popular topic for researchers. This paper focuses on the problem of detecting text regions in manga pages. Because it is time-consuming and laborious to identify the text regions in images manually, an automatic approach is highly desirable. Here, we propose a new text-detection method for manga using a Stroke Width Transform (SWT) technique in conjunction with a Support Vector Machine (SVM). Conventional SWT-based text-detection techniques perform poorly with manga because both text and non-text objects have similar characteristics for strokes, lines, and shapes. To better suit manga, we propose modifying the rules for finding letter candidates, which improves the ability to capture text. An SVM is then used to classify image patches into letter and nonletter regions. We compared our proposed framework with a conventional framework and other text-detection methods including deep-learning techniques. In the results, our proposed method achieved the highest F-measure of 0.506.

## I. INTRODUCTION

Japanese comics are well known across the world as “manga.” There has been extensive research in manga [1]–[7] and benchmark datasets for this research, including Manga109 [8]. Manga109 contains more than 20,260 manga pages collected from 109 volumes and drawn by professional manga artists in Japan. The annotations in the dataset refer to face, body, frame, and text regions, with the text regions being identified and annotated manually. The annotations were undertaken by human without using any support from automation. This task was time-consuming and laborious. Therefore, an automatic annotation system is highly desirable.

In comparison to regular artworks, manga contains much more text and this research focuses on manga. Several text-detection methods for manga have been proposed [9], [10] but these methods are limited to specific layouts, dialogue-balloon positions, and art frames. Moreover, the proposed frameworks are not fully automated and robust to the various text styles present in manga. Recently, a deep-learning technique, the convolutional neural network, was applied to the feature-extraction aspect of text detection in manga [11]. It performed well and was able to reduce the false-positive rate without limitations on layout structure. However, its use of a deep-learning technique results in a high computational cost [11].

In this research, we focus on developing a framework for text detection that is robust to manga components. The

Masanori Sugimoto  
 Graduate School of Information Science & Technology  
 Hokkaido University  
 Hokkaido 060-0808  
 Japan  
 Email: sugi@ist.hokudai.ac.jp

Stroke Width Transform (SWT) technique is used to describe the characteristics of a component “stroke.” It was proposed initially as part of a framework for detecting text in natural scenes, based on the assumptions that there would be high-density edges to text areas and smooth backgrounds [12]. However, employing the framework proposed by [12] to detect text in manga can generate many false-positive text predictions that reduce performance. This is because manga involves greyscale image and object details such as size, stroke, and background that are similar to text. We therefore aim to adopt an SWT similar to that proposed by [12] and improve its framework to suit manga.

The proposed technique can be divided into four main steps: (i) the Stroke Width Transform, (ii) finding letter candidates, (iii) letter classification, which aims to reduce false-positive results using the Support Vector Machine (SVM) with the histogram of oriented gradients (HOG) as a feature, and (iv) grouping letters into lines.

The paper is structured as follows. Detecting text in natural scenes with the SWT [12] is briefly explained in Section II. Our proposed framework is described in Section III. Section IV discusses our experiments and their results. Our conclusions are contained in Section V.

## II. DETECTING TEXT IN NATURAL SCENES WITH THE SWT

### A. The Stroke Width Transform

The SWT is a text-detection technique for photographs that uses the “stroke” features of an object [12], which can differentiate between text and nontext objects. The extraction process begins with the output image being initialized (by assigning  $\infty$  to all its pixels). The output image is the same size as an input image. Next, the Canny edge detection [13] is employed to extract the edges objects in the image. The distance between the borders of a stroke can be calculated by considering each edge pixel  $p$  and finding a corresponding pixel  $q$  on the other edge. As shown in Fig. 1(b), the gradient direction of edge pixel  $p$  ( $d_p$ ) points to edge pixel  $q$ . If  $d_p$  is approximately opposite  $d_q$ ,  $d_q = -d_p \pm \pi/6$  and any pixels that are located in between  $p$  and  $q$  will be assigned to  $\|p - q\|$ , as shown in Fig. 1(c). If the pixel already contains a width  $\|p - q\|$  value and the new value of  $\|p - q\|$  is smaller than the existing one, it will be replaced by the smaller value.

Otherwise, it remains the same. Eventually, a matrix with stroke-width values assigned to each element will be returned.

### B. Finding letter candidates

In this step, the SWT output is used to find letter candidates by eliminating unrelated components. Each pixel is checked and compared with its neighbors. If their stroke-width ratio does not exceed 3.0, they will be grouped together. However, if the connected components are too big or too small, they will be eliminated according to these two rules: (i) the ratio of the diameter of the connected component to its median stroke width must be less than 10 and (ii) its height must be between 10 and 300 pixels, as shown in (1).

$$f(d, h, \tilde{s}) = \begin{cases} 1, & \text{if } \frac{d}{\tilde{s}} < 10 \text{ and } 10 < h < 300 \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where  $d$  is the diameter of the connected component,  $h$  is its height in pixels, and  $\tilde{s}$  is its median stroke width.

### C. Grouping letters into text lines

The letter candidates will be grouped into lines based on the distance between letters, the stroke-width value, and their height. Two letter candidates can be grouped together if (i) their median stroke-width ratio is less than 2.0, (ii) their height ratio does not exceed 2.0, and (iii) the distance between them does not exceed three times that of the wider candidate. After grouping, all letter chains can be merged to form a word. Chains can be merged if two chains have shared letters and the same direction. This step terminates when there are no more chains to be merged. Finally, the group or chain for a word is obtained from the image.

## III. PROPOSED METHOD

In this section, we describe our proposed method, which utilizes an SWT in conjunction with an SVM and a HOG. As mentioned in Section I, the purpose of the SWT is to detect text in natural scenes. Unfortunately, when an SWT is simply applied to manga, it can lead to many false-positive results, as shown in Fig. 2. This is because of the characteristic differences between objects in manga and in the real world. Moreover, various graphic components in manga are very similar to text characters, such as grass, a character's hair, and background details, as shown in Figs. 2(a) and 2(b). Our proposal therefore modifies the steps involving finding letter candidates and grouping letters and adds a new step to make the approach more suited to manga, as shown in Fig. 3. In the last step, it utilizes an SVM to reduce the false positives in the previous finding-letter-candidates step.

### A. The Stroke Width Transform

This step is similar to that proposed in [12]. The input is a manga image that has been transformed by the SWT operator. The size of the output image is the same as the input.

### B. Finding letter candidates

In manga, there may be various text sizes to compare with the natural scenes. Therefore, we modify the conditions for finding letter candidates to enhance the performance, as shown in (2).

$$f(d, h, w, \tilde{s}) = \begin{cases} 1, & \text{if } 1 < \frac{d}{\tilde{s}} < 15 \text{ and } \tilde{s} \leq 80 \text{ and} \\ & 5 < h, w < 50 \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where  $w$  is the width of the connected component in pixels.

Finally, we obtain the letter candidates shown in Fig. 4. Note that the baseline method fails to capture all the letters while our proposed method captures them all. However, there are many nonletter image patches. Their elimination is the purpose of the subsequent step.

### C. Letter classification

In this step, we aim to classify image patches obtained from the previous step into letter and nonletter patches. This will reduce the number of false positives in the letter candidates. We adopt an SVM to perform this task. An SVM is a supervised learning technique that is well known and widely used for classification and regression tasks [14]. The dataset comprises two classes of letter (positive) and nonletter (negative) image patches created from Manga109 [8], as shown in Fig. 5. More details will be described in Section IV. Image-patch features are extracted by using the HOG [15], which is a descriptor for the gradient-direction patterns of objects in terms of their distribution. In this work, a HOG feature is represented by a 2,916-dimensional vector. After this process, the letter patches are grouped into lines in the next step.

### D. Grouping letters into lines

The conventional (baseline) approach [12] groups the letter candidates into lines by using rules about height and median stroke-width values, as explained in Section II. It aims to remove scattered noise, i.e., the nonletter image patches. However, our approach has already dealt with nonletter image patches by using an SVM in conjunction with a HOG. Therefore, we use only the distance between letters to form groups.

Our grouping method uses the classified letters to create lines of letters. A line is created by including in the group any character that is closer to another than 1.5 times the narrower character's width. Characters spaced out less than this are considered as part of the same line, as shown in Fig. 6. Moreover, each line must contain at least three characters and have an area (width  $\times$  height) exceeding 2,550 pixels, according to the experiments using our dataset. Finally, the letters are grouped into lines.

## IV. EXPERIMENTS

### A. Dataset

We used images from Manga109 [8], which comprises 109 annotated volumes of manga created by the Aizawa Yamasaki Laboratory at the University of Tokyo. All manga volumes in the dataset were drawn by professional manga artists in

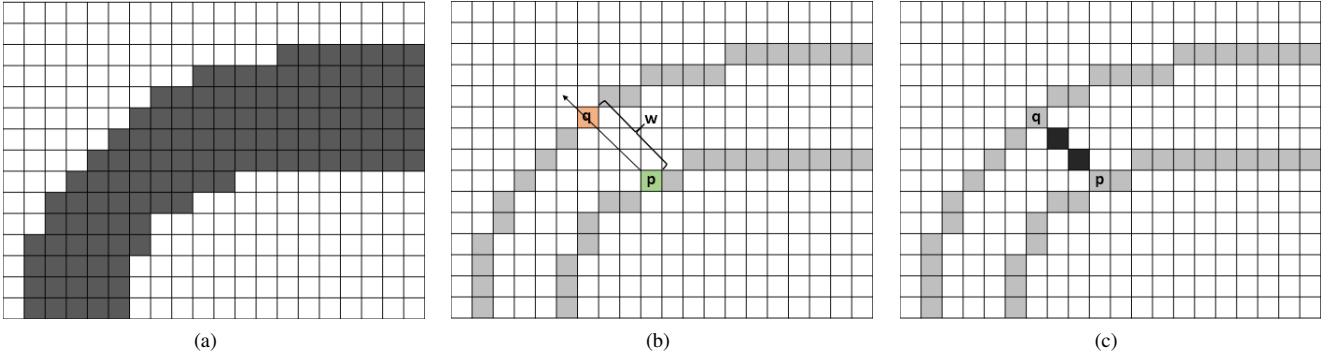


Fig. 1: (a) Grey pixels representing a stroke, (b) Pixel  $p$  is considered to be at the edge of the stroke, with the gradient direction  $d_p$  of  $p$  pointing to  $q$ , the corresponding pixel on the other edge.  $w$  is the distance between the two edge pixels. (c) The pixels along the line connecting  $p$  and  $q$  are assigned the distance  $w$  between  $p$  and  $q$ .

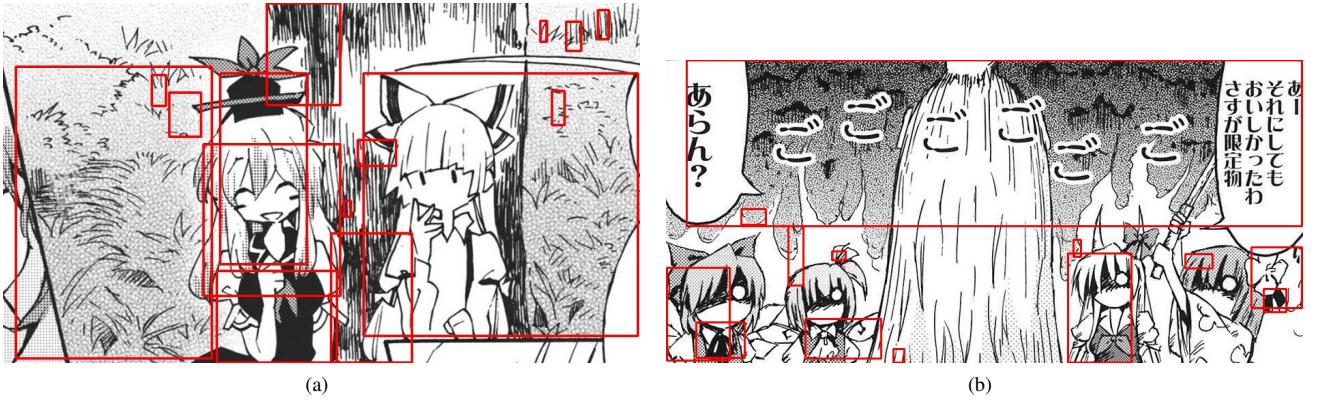


Fig. 2: Example of results generated by a conventional SWT method (the baseline method), where there are many false positives because the characteristics of text and graphic elements are very similar [12]. (a) Artist: Shinoasa (b) Artist: Kousei (Public Planet)

Japan and were available commercially to the public between the 1970s and 2010s. Each page is annotated with text-region areas suitable for training and evaluating our proposed method.

### B. Experimental settings

We used experimental settings similar to those of Aramaki et al. [11], enabling us to make direct comparisons between their results and our results. We selected 100 pages of training data and 100 pages of test data at random from six manga titles, i.e., *Aosugiru Haru*, *Arisa 2*, *Bakuretsu Kung Fu Girl*, *Dollgun*, *Love Hina*, and *Uchiha Akatsuki EvaLady*.

Because our approach requires us to build a model for classifying image patches as letters or nonletters, as shown in Fig. 5, we created a dataset comprising 5,201 letter patches and 5,201 nonletter patches. These image patches were obtained from 100 pages selected at random from the Manga109 data set. The data were partitioned into 50 % for training and 50 % for validation. We used an SVM with a radial basis function kernel. This has two hyperparameters that require tuning, i.e., the regularization ( $C$ ) and kernel ( $\gamma$ ) parameters. A grid search was performed, with  $C$  and  $\gamma$  ranging from  $2^{-10}$  to  $2^{10}$ , to train the model and evaluate its F-measure on

the validation set. We performed the search using the Google Compute Engine n1-highcpu-8. The optimal  $C$  and  $\gamma$  values, returning the best F-measure, were  $2^5$  and  $2^{-6.75}$ , respectively. This optimized model was used in the letter-classification step for our proposed method.

Our evaluation method for text detection was the same as for the ICDAR 2013 Robust Reading Competition [16]. If the ratio of the overlapped area to the ground-truth area is greater than  $t_p$  and the ratio of the overlapped area to the detected region is greater than  $t_r$ , correct detection has been achieved. Here,  $t_p$  and  $t_r$  were set to 0.5, as for Aramaki et al. [11]. Precision and recall can be calculated by (3) and (4), respectively.

$$P = \frac{\text{\#Correctly Detected Rectangles}}{\text{\#Detected Rectangles}} \quad (3)$$

$$R = \frac{\text{\#Correctly Detected Rectangles}}{\text{\#Rectangles of the Ground-truth}} \quad (4)$$

The F-measure can then be calculated by

$$F = 2 \cdot \frac{P \cdot R}{P + R}. \quad (5)$$

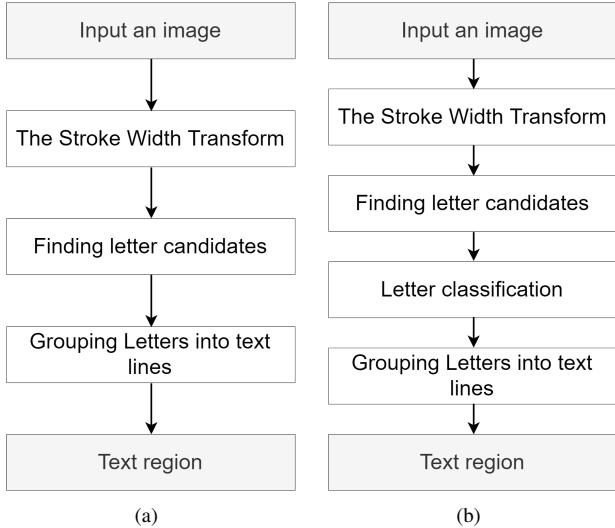


Fig. 3: The flowchart for the baseline method [12] comprises three steps whereas our proposed method has four steps: (a) baseline method and (b) proposed method.

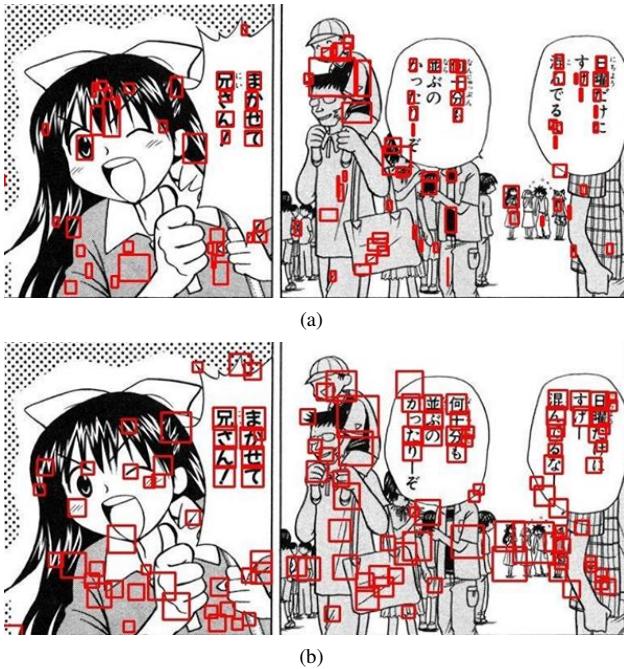


Fig. 4: The results of generating letter-candidate regions in Arisa ©Yagami Ken. The image patches with bounding boxes are the letter-candidate regions: (a) baseline method and (b) proposed method.

We compared our proposed method with a conventional (baseline) SWT method [14] and with previous published results for this dataset using various text-detection methods reported in literature. These other approaches were the Basic Grouping+ImageNet Classification model (BG+ImN) [11], the Basic Grouping+Illustration2Vec model (BG+I2V) [11],

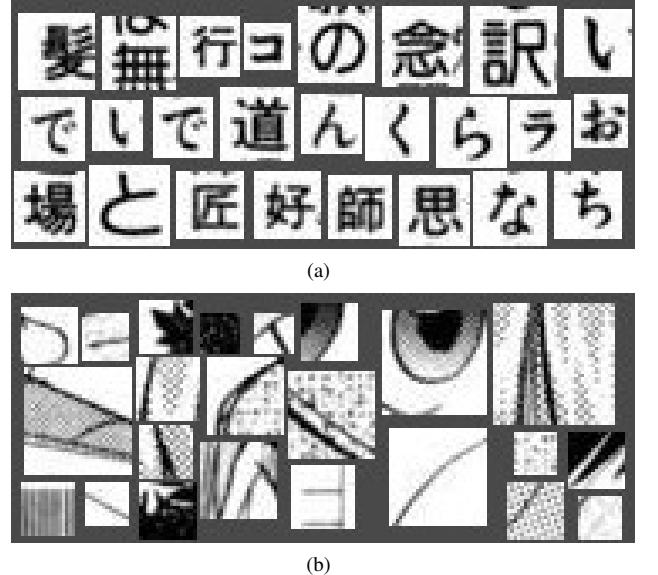


Fig. 5: Examples of letter image patches: (a) positive image patches and (b) negative image patches.

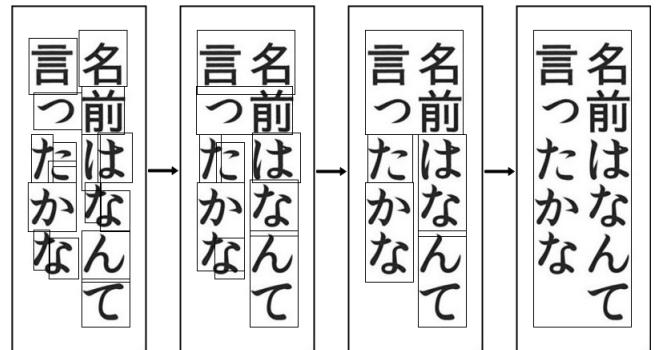


Fig. 6: An example of the process of grouping letters into lines.

Scene Text Detection (STD) [17], Speech Balloon Detection (SBD) [9], and Text Line Detection (TLD) [18]. These approaches use a variety of text-detection methods, including heuristic assumptions (direction of text, layout, dialogue balloon) and convolutional neural networks. The results are shown in Table I. Note that our proposed method yields the highest F-measure of 0.506. The proposed method clearly outperforms the baseline method for all performance measures. Moreover, its F-measure is higher than that for both BG+ImN and BG+I2V. Note also that both BG+ImN and BG+I2V utilize deep-learning methods. However, the best precision and recall is achieved by BG+I2V and BG+ImN at 0.715 and 0.481, respectively. Examples of text regions detected by our proposed technique are shown in Fig. 7.

It is interesting that our method can perform better than deep-learning techniques. This might be because BG+ImN uses the ImageNet Classification model [19], which was trained using a large number of real-world object images. However, manga objects are different from objects in natural

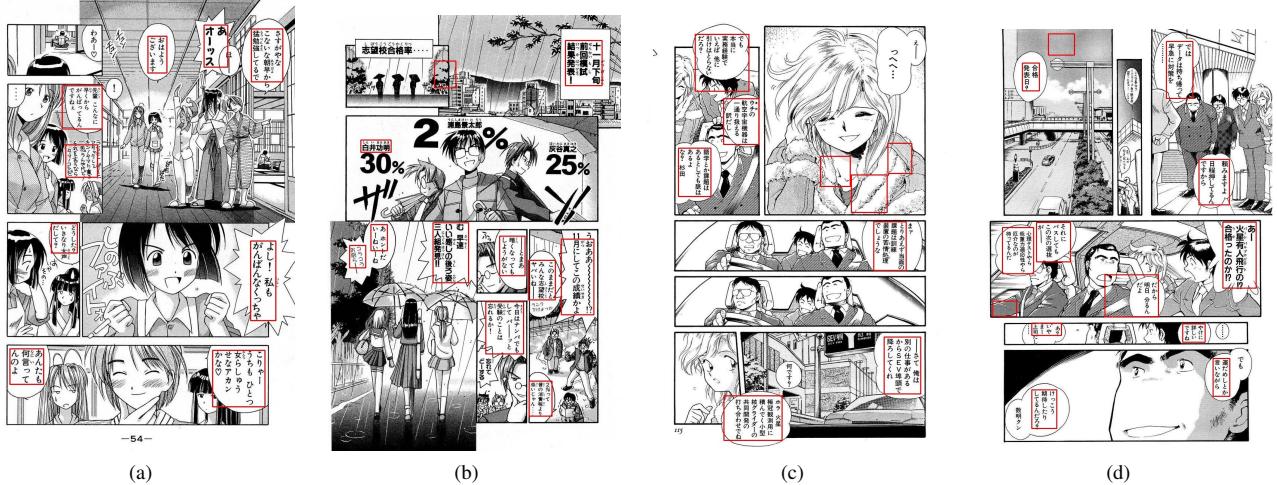


Fig. 7: Examples of text regions detected by our proposed method on (a–b) Love Hina © Ken Akamatsu and (c–d) Eva Lady © Miyone Shi.

TABLE I: The experimental results with Manga109 [8]. The proposed method is compared with the baseline method [12], the Basic Grouping+ImageNet Classification model (BG+ImN) [11], the Basic Grouping+Illustration2vec model (BG+I2V) [11], Scene Text Detection (STD) [17], Speech Balloon Detection (SBD) [9], and Text Line Detection (TLD) [18].

Method	Precision	Recall	F-measure
STD [17]	0.165	0.051	0.078
SBD [9]	0.180	0.102	0.130
TLD [18]	0.095	0.095	0.095
BG + ImN [11]	0.451	<b>0.481</b>	0.466
BG + I2V [11]	<b>0.715</b>	0.191	0.301
Baseline [12]	0.068	0.336	0.113
Our method	0.564	0.458	<b>0.506</b>

scenes. Another deep-learning model is BG+I2V. Although it achieved the best precision in our experiments, its recall is lower than both BG+ImN and our proposed technique. This method uses Illustration2Vec [20] as a classification model. It was trained using a dataset of manga and anime illustrations (the Japanese animation art) from several sources (e.g., Danbooru and Safebooru) that is similar to our approach. The model was not built specifically for text classification but for other purposes such as tag prediction and finding similar images. This might explain why the model appeared relatively unsuccessful in our experiments.

## V. CONCLUSION

In this research, we have proposed a text-detection framework for manga using an SWT in conjunction with an SVM and a HOG. Experiments were conducted using a well-known Manga109 dataset. The proposed method yielded the best F-measure in comparison with a baseline method and those involving deep learning. To date, we have only tested our method on Japanese comics, where it works well. However, more investigations are required to find additional improvements and we aim to extend the method to other languages.

## ACKNOWLEDGMENT

We would also like to show our gratitude to Xiaoyong Zhang, Sendai National College of Technology for sharing their wisdom. We also thanks to Napassorn Thammaviwatnukoon and ZUN for inspiration on this research. This research is supported by King Mongkuts Institute of Technology Ladkrabang.

## REFERENCES

- [1] H. Yanagisawa, T. Yamashita, and H. Watanabe, “A study on object detection method from manga images using CNN,” in *Proceedings of the International Workshop on Advanced Image Technology (IWAIT 2018)*, Chiang Mai, Thailand., Jan 2018, pp. 1–4.
- [2] X. Liu, C. Li, H. Zhu, T.-T. Wong, and X. Xu, “Text-aware balloon extraction from manga,” *The Visual Computer*, vol. 32, no. 4, pp. 501–511, Apr 2016. [Online]. Available: <https://doi.org/10.1007/s00371-015-1084-0>
- [3] X. Pang, Y. Cao, R. W. Lau, and A. B. Chan, “A robust panel extraction method for manga,” in *Proceedings of the 22nd ACM International Conference on Multimedia (MM 2014)*. New York, NY, USA: ACM, 2014, pp. 1125–1128. [Online]. Available: <http://doi.acm.org/10.1145/2647868.2654990>
- [4] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “Interactive segmentation for manga using lossless thinning and coarse labeling,” in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA 2015)*, Hung Hom, Kowloon, Hong Kong, Dec 2015, pp. 293–296.
- [5] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, and K. Aizawa, “Object detection for comics using manga109 annotations,” *CoRR*, vol. abs/1803.08670, 2018. [Online]. Available: <http://arxiv.org/abs/1803.08670>
- [6] S. Kovánen and K. Aizawa, “A layered method for determining manga text bubble reading order,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2015)*, Quebec City, QC, Canada, Sept.
- [7] Y. Matsui, T. Shiratori, and K. Aizawa, “Drawfromdrawings: 2D drawing assistance via stroke interpolation with a sketch database,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 7, pp. 1852–1862, 2017.
- [8] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, “Sketch-based manga retrieval using manga109 dataset,” *Multimedia Tools and Applications*, vol. 76, no. 20, pp.

- 21 811–21 838, Oct 2017. [Online]. Available: <https://doi.org/10.1007/s11042-016-4020-z>
- [9] H. Tolle and K. Arai, “Manga content extraction method for automatic mobile comic content creation,” in *Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS 2013)*, Bali, Indonesia, Sept 2013, pp. 321–328.
- [10] C. Rigaud, T. Le, J. . Burie, J. Ogier, S. Ishimaru, M. Iwata, and K. Kise, “Semi-automatic text and graphics extraction of manga using eye tracking information,” in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, Santorini, Greece, April 2016, pp. 120–125.
- [11] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “Text detection in manga by combining connected-component-based and region-based classifications,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*, Phoenix, AZ, USA, Sept 2016.
- [12] B. Epshtain, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, CA, USA, Jun 2010, pp. 2963–2970.
- [13] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [14] J. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, Jun 1999. [Online]. Available: <https://doi.org/10.1023/A:1018628609742>
- [15] W. T. Freeman, W. T. Freeman, M. Roth, and M. Roth, “Orientation Histograms for Hand Gesture Recognition,” in *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 1994, pp. 296–301. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.618>
- [16] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazn, and L. P. de las Heras, “Icdar 2013 robust reading competition,” in *2013 12th International Conference on Document Analysis and Recognition*, Kolkata, India, Aug 2013, pp. 1484–1493.
- [17] L. Gómez and D. Karatzas, “Multi-script text extraction from natural scenes,” in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, Washington, DC, USA, Aug 2013, pp. 467–471.
- [18] C. Rigaud, D. Karatzas, J. Van De Weijer, J.-C. Burie, and J.-M. Ogier, “Automatic text localisation in scanned comic books,” in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, Barcelona, Spain, Feb 2013. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00841492>
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012)*, vol. 1. Lake Tahoe, Nevada, USA: Curran Associates Inc., Dec 2012, pp. 1097–1105. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999257>
- [20] M. Saito and Y. Matsui, “Illustration2Vec: A semantic vector representation of illustrations,” in *SIGGRAPH Asia 2015 Technical Briefs*. Kobe, Japan: ACM, Nov 2015, pp. 5:1–5:4. [Online]. Available: <http://doi.acm.org/10.1145/2820903.2820907>