



รายงาน

Image processing ในชีวิตประจำวัน
หัวข้อเรื่อง การตัดตัวอักษรออกจากภาพเอกสารเชิงดิจิทัล

จัดทำโดย

นาย ธนภัทร นันทสิริโยธิน รหัสนิต 6230300435

เสนอ

อาจารย์ ไพรัช สร้อยทอง

รายงานนี้เป็นส่วนหนึ่งของรายวิชา

การประมวลผลภาพดิจิทัล รหัสวิชา 03603372-60

มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา ภาคการเรียนที่ 2 ปีการศึกษา 2564

คำนำ

รายงานฉบับนี้เป็นส่วนหนึ่งของวิชา Image processing รหัสวิชา 03603372-60 ซึ่งได้ถูกมอบหมายในหัวข้อที่ว่าด้วย “Image processing” ในชีวิตประจำวัน โดยข้าพเจ้าเลือกที่ศึกษาหาความรู้จากการตัดตัวอักษรออกจากภาพเอกสารเชิงดิจิทัล โดยในเนื้อหาข้าพเจ้า จะยิบยกตัวอย่าง และความหมายคำศัพท์ที่ควรรู้ไว้ก่อนด้วย

ข้าพเจ้า ขอขอบคุณ อาจารย์ ไพรัช สร้อยทอง ที่ให้ความรู้ ให้คำปรึกษา และหวังใจเป็นอย่างยิ่งว่าผู้ที่มาศึกษาเอกสารฉบับนี้ จะได้รับความรู้เพิ่มมากขึ้น และสามารถนำไปใช้ในชีวิตประจำวันได้ หากเอกสารชุดนี้มีความผิดพลาดประการใด ข้าพเจ้าขอน้อมรับและจะนำไปปรับแก้ไข

คณะผู้จัดทำ

นาย ธนภัทร นันทสิริโยธิน

สารบัญ

เรื่อง	หน้า
บทนำ	1
กระบวนการและขั้นตอนเบื้องต้น	2-3
รูปภาพสี RGB	4
ภาพสีขาวเทาหรือ Grey Scale	4-5
ภาพขาวดำ	5
แยกวัตถุและภาพพื้นหลัง	6
การกำจัดสัญญาณรบกวน (Noise Reduction)	7
เอกสารอ้างอิง	8

การตัดตัวอักษรออกจากภาพเอกสารเชิงดิจิทัล

บทนำ

ในยุคของดิจิทัล เกิดจากความก้าวหน้าทางด้านเทคโนโลยี มากมายไม่ว่าจะเป็นระบบสื่อสาร ระบบเทคโนโลยีของคอมพิวเตอร์ เทคโนโลยีของภาพถ่าย เทคโนโลยีการบันทึกภาพถ่ายที่ถูกพัฒนาแบบก้าวกระโดด เนื่องจาก เราไม่สามารถปฏิเสธได้เลยว่า ในยุคปัจจุบันกล้องมือถือมีความละเอียดที่มากเพียงพอต่อการใช้งาน และผู้คนในปัจจุบัน นิยมเก็บเอกสารด้วยรูปถ่ายที่ถูกบันทึกด้วยโทรศัพท์เคลื่อนที่ (มือถือ) เป็นส่วนใหญ่เพราะไม่จำเป็นต้องถือเอกสาร ฉบับจริงติดตัวไปด้วยทุกที่ อาจจะแค่เพียงต้องการดูแค่ชั่วคราวเท่านั้น

ดังนั้นในยุคของดิจิทัลนั้น จึงเกิดเทคโนโลยีชนิดหนึ่งขึ้น ที่ช่วยแปลงภาพเอกสาร ที่เป็นสี และมีแสงจากธรรมชาติ เป็น รูปภาพที่คล้ายกับการถูกสแกนมาจากเครื่องสแกนเนอร์ ดังนั้นจึงมีการประยุกต์ใช้ความรู้ทางด้าน (image Processing) เพื่อปรับเปลี่ยนเอกสารให้เป็นรูปภาพเชิงดิจิทัล และนำไปผ่านกระบวนการรู้จำตัวอักษร (Optical Character Recognition OCR) เพื่อแปลงรูปภาพเชิงดิจิทัล ให้อยู่ในรูปแบบของเอกสารอิเล็กทรอนิกส์

เอกสารเชิงดิจิทัล เอกสารเชิงดิจิทัล คือการรวมกันของสองคำคือ “เอกสาร” รวมกับ “ดิจิทัล” ซึ่งเอกสารในหัวข้ออาจประกอบด้วยตัวอักษรเป็นส่วนมาก และ ดิจิทัล เป็นชื่อเฉพาะ อาจสะกดเป็นดิจิตอล หรือ ดิจิตอล คือการนำเอาข้อมูลหรือระบบข้อมูล มาทำให้อยู่ในรูปแบบสื่อสารสนเทศบนคอมพิวเตอร์ โดยเฉพาะนิยมถูกเปลี่ยนเป็นในเลขฐานสอง เพื่อให้ข้อมูลนั้นถูกบันทึกลงคอมพิวเตอร์ ดังนั้น เอกสารดิจิทัลคือการนำเอกสารในชีวิตประจำวันแปลงไปเป็นรูปแบบของคอมพิวเตอร์ ซึ่งในรายงานฉบับนี้ จะกล่าวว่าเอกสารเชิงดิจิทัลเป็นเอกสารที่เป็นรูปภาพเท่านั้น

กระบวนการและขั้นตอน

การประมวลผลภาพนำไปสู่การหาคุณลักษณะพิเศษ ของรูปภาพตัวอักษร ซึ่งเรียกขั้นตอนขั้นดังกล่าวว่า ขั้นตอนประมวลผลเบื้องต้น ซึ่งประกอบไปด้วย 4 ขั้นตอนหลัก คือ การกำจัดพื้นหลัง การจำแนกบรรทัดข้อความ การจำแนกตัวอักษรออกจากบรรทัดข้อความ และการหาคุณลักษณะพิเศษ ซึ่งการประมวลผลเบื้องต้น เริ่มต้นจากการนำรูปภาพตัวอักษรที่เป็นภาพสี (RGB image) มาแปลงเป็นภาพสีระดับเทาขาว (Grey level) จากนั้นเปลี่ยนจากภาพ ขาวเทาให้เป็นภาพ ขาวดำ (Black and white image) โดยจะผ่านการกรองค่า เธรดโฮลด์ (threshold value) เป็นตัวแยกวัตถุ ในกรณีนี้คือ ตัวอักษร ออกจากพื้นหลัง

เมื่อผ่านกระบวนการแปลงภาพให้เป็นภาพขาวดำจากนั้น ทำการกำจัด สัญญาณรบกวนออกจากภาพ เพื่อให้ได้ภาพที่มีความชัดเจนขึ้น และทำการจำแนกบรรทัดข้อความ และจำแนกตัวอักษรออกจากบรรทัดข้อความ และจำแนกตัวอักษรออกจากบรรทัดข้อความ จากนั้นนำรูปภาพตัวอักษรที่ได้จากขั้นตอนข้างต้นมา เพื่อหาคุณลักษณะพิเศษ ของรูปภาพตัวอักษรแต่ละตัว เพื่อนำคุณลักษณะพิเศษที่ได้ไปเข้าสู่กระบวนการ รู้จำ เพื่อจดจำตัวอักษร

ผลลัพธ์ที่ได้จากกระบวนการเรียนรู้จำ คือเอกสารอิเล็กทรอนิกส์ ซึ่งอาจอยู่ในรูปแบบ ของ Text File หรือโปรแกรมประมวลผลคำ Word Processing ซึ่งทำให้ง่ายต่อการนำไปประยุกต์

แต่สำหรับในรายงานฉบับนี้ ขั้นตอนการแสดงผลลัพธ์ที่ออกมาในรูปแบบของ Text File และขั้นตอนการเรียนรู้จำอาจไม่ได้ แสดงให้เห็น เพราะอยู่นอกเหนือของความรู้วิชาที่กำลังศึกษาอยู่

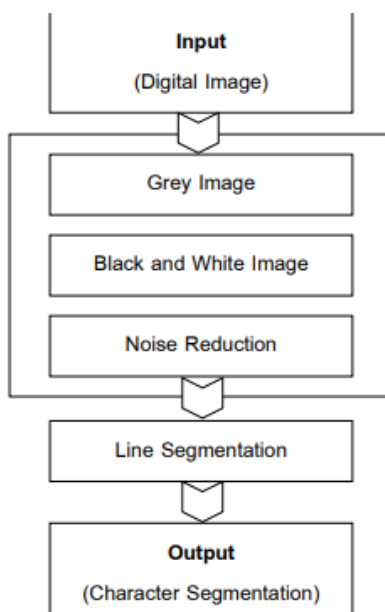


Figure 1 Framework of Thai handwritten character segmentation

ที่มา [Microsoft Word - paper - image.doc \(msu.ac.th\)](http://msu.ac.th/paper-image.doc)

ภาพเชิงดิจิทัล คือฟังก์ชัน 2 มิติ $f(x,y)$ ของความเข้มแสง intensity โดยที่ x และ y คือค่าแสดงตำแหน่งในระบบพิกัดฉาก และค่าของ ฟังก์ชัน f ที่ตำแหน่งใดๆ จะเป็นสัดส่วนของความสว่างของแสง ณ ตำแหน่งนั้นๆ ดังรูปข้างล่าง

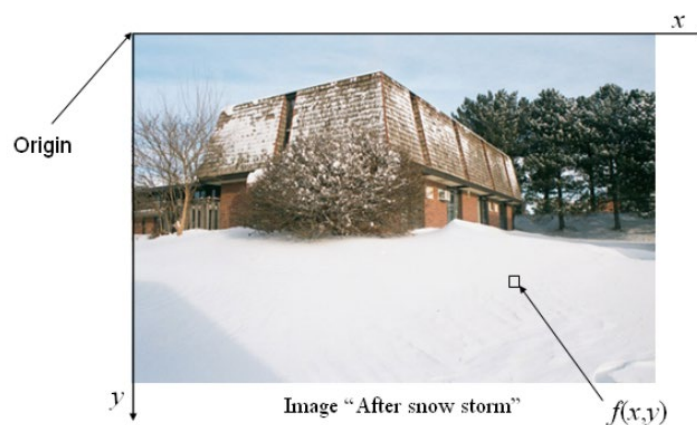


Figure 2 An example of digital image

ที่มา [Microsoft Word - paper - image.doc \(msu.ac.th\)](http://msu.ac.th/paper-image.doc)

รูปภาพสี RGB

ค่าในแต่ละพิกเซลของ RGB ต้องประกอบไปด้วยเวกเตอร์ของสีแดง สีเขียว สีน้ำเงิน อย่างละ 8 บิต ดังนั้นรูปภาพจะประกอบไปด้วยจำนวนบิต ทั้งหมด 24 บิต ดังนั้น สีที่เป็นไปได้ทั้งหมด มากกว่า 16 ล้านสี

					155	187	209	58	7	
					14	125	233	201	98	159
253	144	120	251	41	147	204				
67	100	32	241	23	165	30				
209	118	124	27	59	201	79				
210	236	105	169	19	218	156				
35	178	199	197	4	14	218				
115	104	34	111	19	196					
32	69	231	203	74						

ที่มาของรูปภาพ <https://www.researchgate.net/>

ภาพสีขาวเทาดำ หรือ Grey Scale

เป็นค่าในแต่ละพิกเซลของ ค่าความเข้มแสง ณ ตำแหน่งใดๆของพิกเซล ซึ่งขั้นตอนการแปลงภาพสีให้เป็นภาพระดับเทา ทำได้โดยการแยกระดับสีแต่ละพิกเซลออกจากกันในรูปแบบสี RGB จากนั้นนำค่าสี RGB มาเข้าสู่สมการเพื่อคำนวณ หาค่าสีเทาและนำค่าที่ได้ไปแทนที่จุดพิกเซลเดิม โดยสมการคือ

$$G' = 0.3R + 0.59G + 0.11B \quad \text{หรือ}$$

$$G' = \frac{R + G + B}{3}$$

โดยกำหนดให้

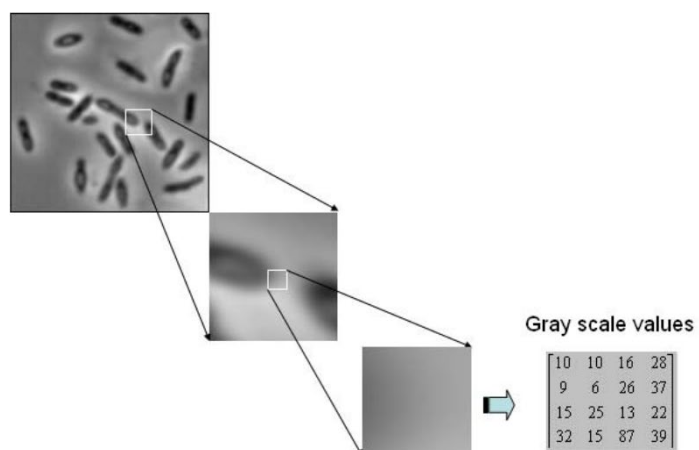
G' คือค่าระดับสีเทา

R คือค่าระดับสีแดง

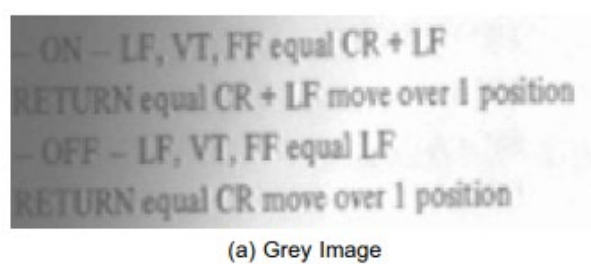
G คือค่าระดับสีเขียว

B คือค่าระดับสีน้ำเงิน

ตัวอย่างรูปภาพ และค่าที่ได้จากการคำนวณ

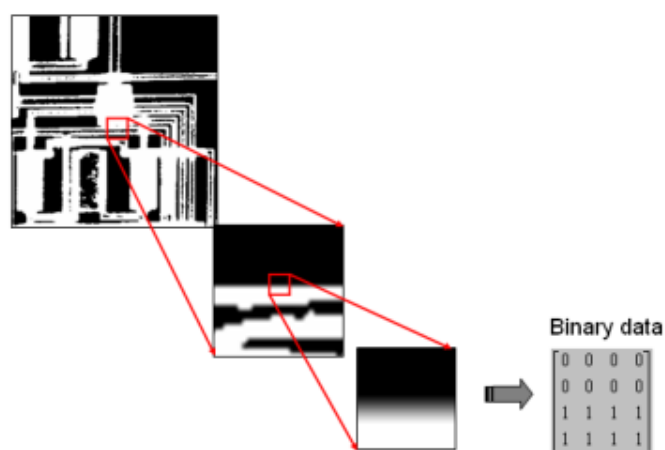


รูปภาพตัวอย่างของไฟล์เอกสารที่ถูกเปลี่ยนเป็น ภาพสีเทาเทาดำ หรือ Grey Scale4



ภาพขาวดำ

ค่าในแต่ละพิกเซลของ black and white image จะใช้แค่ 1 บิตที่มีความเป็นไปได้ เช่นสี ขาว คือ 1 และสีดำ คือ 0 เท่านั้น



แยกวัตถุและภาพพื้นหลัง

ขั้นตอนการแปลงภาพสีเทา ให้ กลายเป็นภาพขาวดำ จะทำให้สามารถแยกวัตถุ (อักษร) ออกจากพื้นหลังได้ โดยอาศัยวิธีการทำเรตโซลด์ (Threshold) เรตโซลด์เป็นวิธีที่ใช้ในการแปลงภาพสี ต่าง ๆ ให้อยู่ในรูปแบบของภาพระดับขาวดำ โดย ใช้ค่าเรตโซลด์ในการจำแนกวัตถุ และพื้นหลัง ออกจากกัน หากจุดใดในภาพนั้นมีความเข้ม น้อยกว่าค่าเรตโซลด์จุดภาพนั้นก็จะถูกปรับให้ เป็นสีขาว แต่ถ้าจุดใดในภาพมีความเข้ม มากกว่าค่าเรตโซลด์ ก็จะถูกปรับให้เป็นสีดำ วิธีการหาค่าเรตโซลด์สามารถหาได้ด้วยวิธี ต่อไปนี้

1. Global Threshold เป็นการหาค่าเรตโซลด์ค่าเดียวรวมกันทั้งภาพโดยใช้ Histogram ในการหาเรตโซลด์ได้จากค่าที่น้อยที่สุดที่อยู่ระหว่างยอดทั้งสอง มีหลักการในการแบ่ง คือค่าที่อยู่ทางซ้ายของค่าเรตโซลด์จะเป็นสีดำ ส่วนค่าที่อยู่ทางขวาของเรตโซลด์เป็นสีขาว
2. Adaptive Threshold การหาค่าเรตโซลด์ด้วยวิธีนี้เหมาะสำหรับภาพที่มีความสว่างไม่สม่ำเสมอ ดังนั้นจึงหาค่าเรตโซลด์หลายค่า เพราะแต่ละบริเวณจะใช้ค่าเรตโซลด์ไม่เท่ากัน ดังนั้นจึงจำเป็นต้องใช้ mask หรือ windows ขนาด $N \times N$ โดยที่ N นั้นควรเป็นเลขคี่ ไปวางไว้บริเวณภาพที่มีค่า Grey Scale ของทุกบริเวณ แล้วนำค่าในขอบเขตใน windows มาบวกกันแล้วหารด้วยจำนวนช่องทั้งหมดของ windows จะได้ ค่า เรตโซลด์ที่อยู่ภายใน windows นั้นๆ จากนั้นทำการหาค่าเรตโซลด์ เช่นนี้ไปเรื่อย ๆ กับบริเวณที่ไม่ซ้ำกัน จนได้ค่าเรตโซลด์ครบในทุกๆ พิกเซล ถ้าค่า Grey Level ของพิกเซลนั้นมีค่ามากกว่าค่าเรตโซลด์ของพิกเซลนั้นๆ จะต้องกำหนดให้เป็นสีขาว แต่ถ้าค่า Grey Level ของพิกเซลนั้นๆ น้อยกว่า เรตโซลด์ของพิกเซลนั้นๆ แล้วจะกำหนดให้เป็นสีดำ ทำเช่นนี้ไปจนครบทุกพิกเซลจะได้ผลลัพธ์สีดำ ทำไปเรื่อยๆ จะได้ภาพเป็นขาวดำ

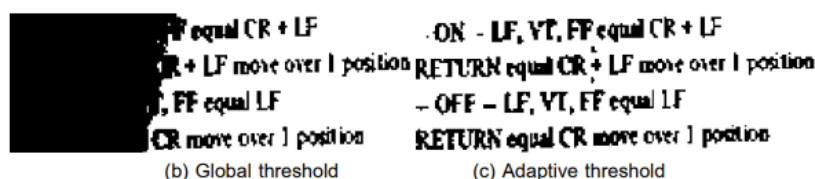
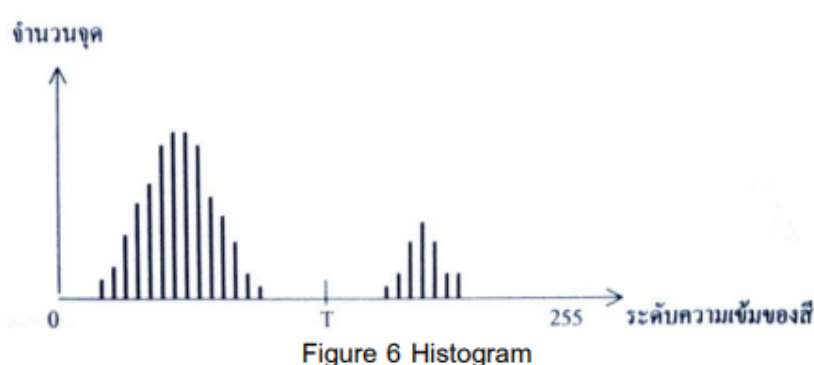


Figure 7 Convert image from grey image to black and white image using threshold

การกำจัดสัญญาณรบกวน (Noise Reduction)

เมื่อได้ภาพขาวดำแล้ว จากรูปด้านบน จะเห็นได้ว่า ทางซ้ายของภาพมีส่วนที่ยังไม่ชัดเจนอยู่ จึงต้องอาศัยวิธีการกำจัดสัญญาณรบกวนออกจากรูปภาพ ผลลัพธ์ที่ได้นั้นจะทำให้ภาพชัดเจนมากขึ้น โดยใช้วิธีต่อไปนี้

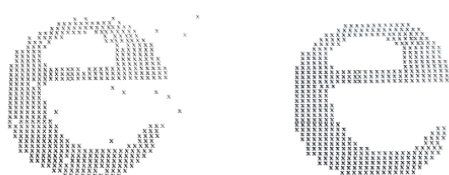
1. Morphological Image Processing เป็นกระบวนการนำโครงสร้าง 0 หรือ 1 ขนาดหนึ่งไปวางบนภาพที่แต่ละตำแหน่งบนภาพจะใช้การอนุมานด้วยเหตุผลระหว่างโครงสร้างกับภาพที่อยู่ใต้โครงสร้าง ได้ผลลัพธ์มาปรับค่าที่ตำแหน่งนั้น การทำงาน มี 2 แบบคือ

1. Erosion คือการกำจัดจุดขอบทุกจุดจะทำให้วัตถุเล็กลง 1 Pixel
2. Dilation คือการเพิ่มจุดรอบขอบวัตถุอีก 1 Pixel

หากทำ Erosion และทำต่อด้วย Dilation จะเรียกว่า Opening คือทำการให้วัตถุที่เล็กนั้นถูกกำจัดออกไปจะแยกวัตถุที่เชื่อมต่อกันด้วยเส้น บางๆออกจากกัน จากนั้นขึ้นทำให้วัตถุมีขนาด ใหญ่ขึ้น ขอบเรียบขึ้น ผลลัพธ์จะได้วัตถุที่มีขนาด เท่าเดิม

แต่ถ้าหาก หากทำ Erosion และทำต่อด้วย Dilation จะเรียกว่า Closing คือทำการให้วัตถุที่เล็กนั้นถูกกำจัดออกไปจะแยกวัตถุที่เชื่อมต่อกันด้วยเส้น บางๆออกจากกัน จากนั้นขึ้นทำให้วัตถุมีขนาด ใหญ่ขึ้น ขอบเรียบขึ้น ผลลัพธ์จะได้วัตถุที่มีขนาด เท่าเดิม

2. Text Noise Filters เป็นการกรองสัญญาณ รบกวนของภาพโดยที่ตัวอักษรบนภาพมี ความชัดไม่เปลี่ยนไป พื้นที่ขนาด 1 Pixel ที่ปรากฏโดด ๆ เป็นรู หรือ เป็นส่วนที่นูนออกมา จะตรวจสอบได้ โดยใช้โครงสร้างขนาด 3×3 Pixel ส่วนพื้นที่ที่มีขนาดใหญ่กว่า 1 พิกเซลใช้ การกรองแบบ kFill ตรวจสอบได้ ซึ่ง kFill เป็นการ ใช้โครงสร้างขนาด $k \times k$ พิกเซลซึ่งประกอบด้วย ส่วนที่อยู่ตรงกลางมีขนาด $(k - 2) \times (k - 2)$ พิกเซล และล้อมรอบตรงกลางอีก $4(k - 1)$ เช่น 3×3 จะมีตรงกลาง $(3-2) \times (3-2) = 1$ พิกเซลและล้อมรอบ ด้วย $4(3-1) = 8$ พิกเซลในส่วนตรงกลางจะถูก กำหนดค่าให้เหมือนกันหมด (fill) เป็น 1 (ON) หรือ 0 (OFF) การพิจารณาค่าเป็น ON (หรือ OFF) นั้นจะต้องดูว่าภาพที่ส่วนตรงกลางของ โครงสร้างที่บออยู่นั้นต้องเป็น 0 (หรือ 1) ทั้งหมด และเงื่อนไขซึ่งขึ้นอยู่กัค่าของตัวแปร 3 ตัวที่ได้ จากค่าของพิกเซลในส่วนที่ล้อมรอบอยู่ต่อไป นี้ ต้องเป็นจริงดังสมการ $(c = 1)$ and $\{ (n > 3k - 4) \text{ or } [(n = 3k - 4) \text{ and } r = 2] \}$ กำหนดให้ n เป็นจำนวนพิกเซลในส่วนที่ ล้อมรอบที่มีค่าเป็น 1 (หรือ 0) c จำนวนกลุ่มของพิกเซลที่มีค่า เป็น 1 ที่อยู่ติดต่อกันในส่วนที่ล้อมรอบ r จำนวนพิกเซลที่อยู่มุมที่มีค่าเป็น 1 (หรือ 0)



(a) Noise black and white image (b) Black and white image
Figure 8 An example showing a black and white image before and after noise reduction

เอกสารอ้างอิง

[Microsoft Word - paper - image.doc \(msu.ac.th\)](#)