

Text Localization

คือการตรวจหาและระบุตำแหน่งของ ข้อความ(text) ภายในรูปภาพโดยในปัจจุบันสามารถแบ่งแยกแบบคร่าวๆได้ 3 รูปแบบคือ

- sliding window based methods คือการทำ windows ค้นหาส่วนที่มี ความเป็นไปได้ที่จะเป็นข้อความ (possible text) บนรูปภาพ
- connected component based methods คือการค้น ลักษณะ(feature) และทำการ เชื่อมต่อส่วนที่เป็น องค์ประกอบเดียวกัน(connected component analysis) เพื่อให้ได้ possible text
- hybrid methods การทำ connected component analysis บน window เพื่อทำการค้นหาแบบเป็นบริเวณ

โดยในปัจจุบันได้มีผู้ที่นำรูปแบบพื้นนี้ไปประยุกต์ต่อทำให้เกิดกระบวนการ text localization รูปแบบใหม่ต่างๆ เช่น Maximally Stable Extremal Region (MSER) based methods ที่มีลักษณะคล้าย connected component based methods

Text Localization in Natural Images using Stroke Feature Transform and Text Covariance Descriptors

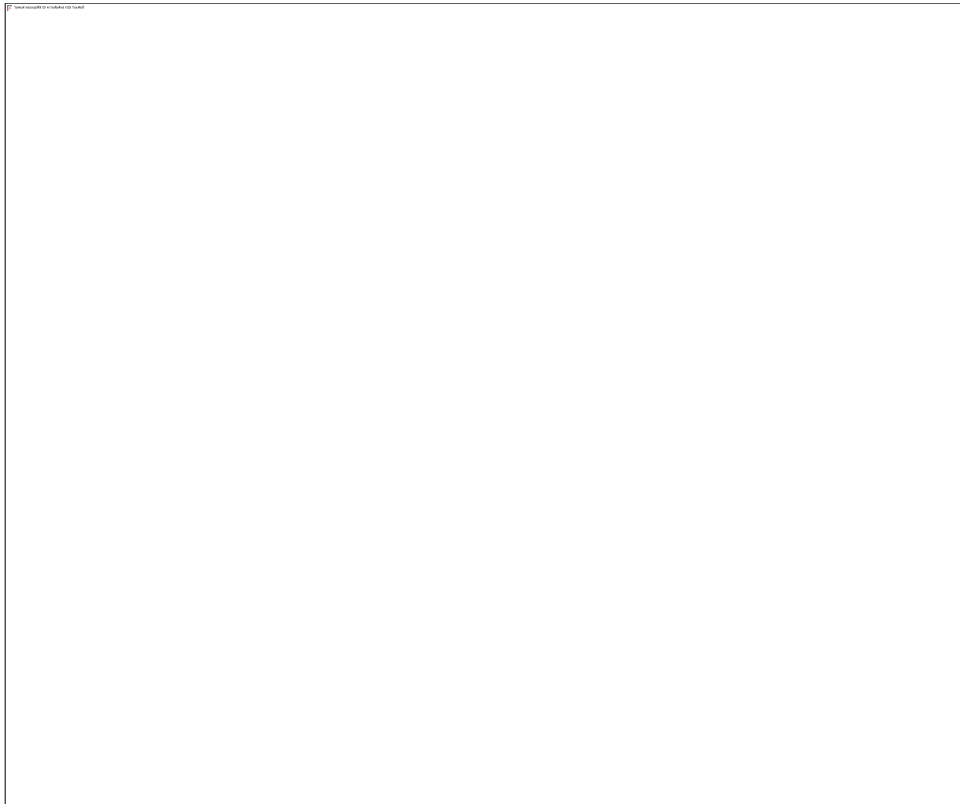
ประกอบไปด้วย 4 ขั้นตอนคือ

- 1.Component Detection using Stroke Feature Transform
- 2.Text Covariance Descriptors for Components Filtering
- 3.Text-Line Construction
- 4.Text Covariance Descriptors for Text-lines Filtering

1. Component Detection using Stroke Feature Transform

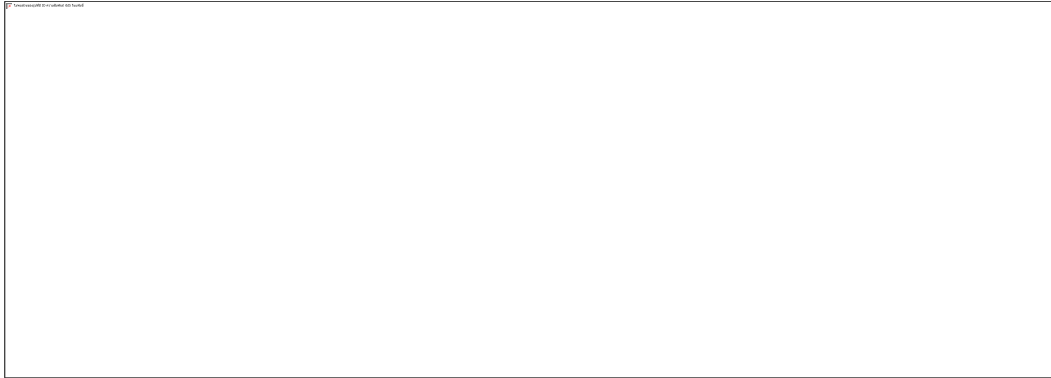
a. The Stroke Width Transform (SWT)

การสร้างรูปภาพที่มีขนาดเท่ากับภาพ input โดยที่ค่าในแต่ละ pixel คือระยะห่างระหว่าง เส้นขอบ ณ จุด p กับเส้นขอบ ณ จุด q ที่อยู่ใกล้ที่สุดโดยดูจากทิศทางของ gradient ต้องมีทิศทางสวนกับ ขอบแรก ซึ่งหาได้จากวิธีการ Canny Edge Detection



b. The Stroke Feature Transform (SFT)

การเพิ่มเติมลักษณะของ SWT โดยแบ่งเป็น 2 ลักษณะ ความสม่ำเสมอของสี(color uniformity) และความสัมพันธ์ของเส้น(local relationships of edge pixels) ได้ผลลัพธ์มาเป็น 2 ขั้นตอนคือ Stroke color map และ Stroke width map



โดยกระบวนการเริ่มจากหา เส้น(ray) มีผ่านเงื่อนไขอย่างใดอย่างหนึ่งใน 2 ข้อ

- Stroke width constraint : จะเหมือนกับ SWT ทั่วไป
- Stroke color constraint : สมมุติจุด q เป็นจุดสิ้นสุดของ ray ค่าสีของจุด q ต้องมีความต่างของใน channel rgm ไม่เกินค่าๆหนึ่ง

c. Component Generation


คือการรวมกลุ่มของ output ที่ได้จาก SFT ให้กลายเป็น component เดียวกัน ซึ่งกระบวนการที่ใช้สามารถเลือกใช้ได้หลายวิธีเช่น region growing เป็นต้น

2. Text Covariance Descriptors for Filtering

TCD for Components (TCD-C)

การสร้าง Descriptors เพื่ออธิบาย feature ของ component ในข้อ 1. เพื่อจะสามารถกรอง component ที่มีความเป็นไปได้น้อยที่จะเป็น character ของ text โดยที่ feature ใช้ประกอบด้วย

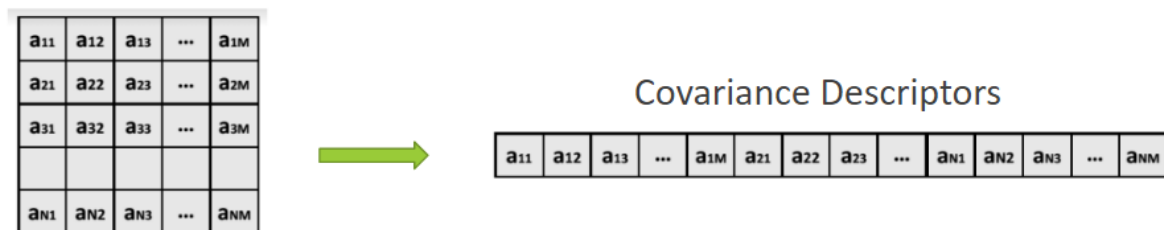
- Normalized pixel coordinates : รูป output ที่มีค่า pixels เป็น coordinate ของ x', y' โดยมีสมการ

คือ  โดย y ก็เหมือนกันและ ค่า max ,min คือ maximum และ minimum pixel ในรูป input

- Pixel intensities : ค่า I และ ค่า RGB IR, IG, and IB ใน the stroke color map
- Stroke width values : ค่าที่ได้จาก the stroke width map
- Stroke distance values : หาได้จากการนำค่าใน the stroke width map ไปผ่านกระบวนการ the Euclidean distance transform
- Per-pixel edge labeling : คือค่าที่ได้จาก Canny edge detection โดย edge มีค่าเป็น 1 และ non-edge มีค่าเป็น 0

นำ feature ที่ได้ทั้งหมดไปสร้าง covariance matrix จะได้เป็น matrix 9×9 หรือก็คือจะได้ descriptor ขนาด 45 และเพิ่ม feature อีก 3 อย่างคือ (1). the aspect ratio อัตราส่วนระหว่างความสูงและความกว้าง (2). the occupation percentage อัตราส่วนระหว่าง pixel ใน component และจำนวน stroke pixels ใน component (3). the ratio of the component scale อัตราส่วนระหว่าง ค่ามากที่สุดระหว่างความสูงกับความกว้างกับค่าเฉลี่ยของ stroke width map value

นำ descriptor ขนาด 48 ไปฝึกกับ random forests classifier เพื่อสร้างออกมาเป็นค่า confident score ของ component



3. Text-Line Construction

การสร้างกลุ่มของ character ซึ่งก็คือ text จาก component หลายๆอันโดยอาศัยลักษณะพื้นฐานของ text คือ

1. ค่า stroke width ของแต่ละ component ควรมีค่าใกล้เคียงกัน
2. ค่าอัตราส่วนความสูงของ component ควรไม่เกิน 2.0 (คืออัตราส่วนความแตกต่างของอักษรตัวเล็กและตัวใหญ่)
3. ความห่างระหว่าง component ควรไม่เกิน 1 เท่าของ component

4. Text Covariance Descriptors for Filtering

TCD for Text-lines (TCD-T)

หลังจากที่ได้ component ที่น่าจะเป็น text แล้วก็สร้าง descriptor เพื่ออธิบาย feature ของ component ตัวใหม่นี้โดยที่ feature ใช้ประกอบด้วย

- 7 feature ที่ใช้ใน TCD-C รวมกับค่าเฉลี่ยของ I , IR , IG , IB , $Sswm$, $Sdist$
- The coordinate
- ความสูงของ component โดย normalize ด้วยความสูงของ text line ทั้งหมด
- ค่า Cosine ของ component ปัจจุบันกับ component ด้านข้างโดยดูจาก ทิศทางของ component ปัจจุบัน
- Horizontal distance ของ component ปัจจุบันกับ component ด้านข้าง
- จำนวนของ component ใน text-line โดยผ่านการ normalize ค่าสูงสุดของ component ใน text-line เช่น 10

- ค่าเฉลี่ยของ confident score ที่ได้จาก TCD-C
ผลลัพธ์สุดท้ายนี้จะได้ descriptor ขนาด 216 และนำไป train เหมือนกับใน TCD-C



Canny Text Detector: Fast and Robust Scene Text Localization Algorithm

1. Character Candidate Extraction

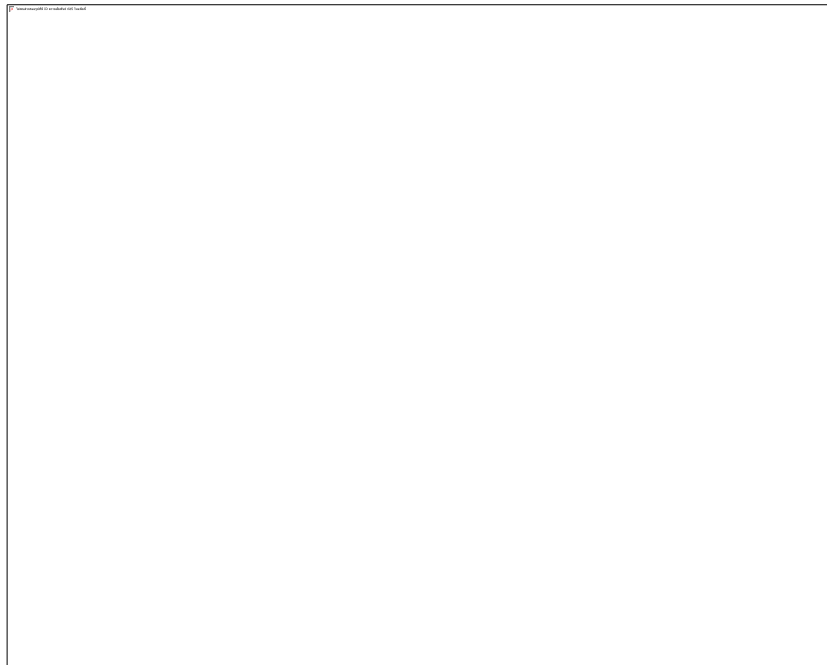
Extremal Regions (ERs)

พื้นที่ที่มีลักษณะตามสมการคือ



โดย

- x, y ค่า index ของรูป
- t คือ threshold ที่ใช้สร้าง R
- $B(R_t)$ คือ set ของ pixels ที่เป็นส่วนนอกของ R_t



2. Non-maximum Suppression

เนื่องจาก character component ที่ได้จาก ข้อ 1. จะมีส่วนที่ซ้อนทับกันมากเกินไปซึ่ง non-maximum suppression มีไว้เพื่อกรองส่วนที่ซ้อนทับกันนั้นโดยเริ่มจาก

- กรอง R_t ที่ค่า $O(R_{t-k}, R_t) > 0.7$ โดยสมการคือ



โดยที่ R_{t-k} คือ parent region

- คำนวณจำนวน pixel ที่ซ้อนทับกันระหว่าง R_{t-k} และ R_t โดยที่ค่าที่ได้ต้องมากกว่า 3

- เลือก region ที่มีค่า $S(R_t)$ มากที่สุดโดยสมการคือ



โดยที่ t' ในกรณีนี้คือ 2

3. Double Threshold Classification

กรอง region ที่มีความเป็นไปได้ว่าจะเป็น text โดยแบ่งผลลัพธ์ไว้ 3 แบบคือ strong text ที่หมายถึง text แน่ๆ และ weak text ที่เป็นไปได้ทั้ง text และ non-text และ non-text ซึ่งใช้ AdaBoost และ multiple cascades โดย feature ที่ใช้คือ mean local binary pattern (MLBP)

4. Text Tracking by Hysteresis

การเลือก weak text เพื่อเปลี่ยนเป็น strong text โดยมีหลักการคือ สำหรับแต่ละ strong text ให้ดูหา weak text ที่อยู่ใกล้แล้วเปรียบเทียบลักษณะโดยถ้ามีลักษณะใกล้เคียงกันก็เปลี่ยน weak text เป็น strong text ซึ่งลักษณะที่ใช้พิจารณาประกอบด้วย

- The spatial location : ระยะห่างระหว่าง text ต้องน้อยกว่า 2 เท่าของค่าสูงสุดของความสูงและความกว้างของ strong text
- The size : ความต่างของความสูงและความกว้างต้องไม่มากกว่า 2 เท่าของความกว้างและความสูง
- The color YCrCb : ค่าสีใน channel YCrCb ต้องไม่ห่างกันเกิน 25
- The ratio ของ large และ small stroke widths : ต้องน้อยกว่า 15

5. Text Grouping

สร้าง bounding boxes สำหรับ text โดยอาศัย minimum-area encasing rectangle

