

# Text Detection in Natural Images

## **Guided By:**

Dr. Vinod Pankajakshan

## **Submitted By:**

Arjun Krishna 15115030

A S Akil Arif Ibrahim 15116001

# Contents

- Introduction
- MSER
- Stroke Width Transform
- Finding Letter Candidates
- Grouping Letter Candidates into Text Regions
- Experiments
- Results
- Conclusion
- References

# Introduction

- Helping a foreigner to understand the contents on an information board (by translating the recognized text into his/her native language) .
- Computerized aid for visually impaired.
- Robotic navigation and intelligent transport system.

# Related Works

The available methods for text detection can be classified into two groups:

- Texture based
- Region based

# MSER

Source



Mser output



# Stroke Width Transform

Edge detection using Canny Edge Detector

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Create gradient map using sobel filter

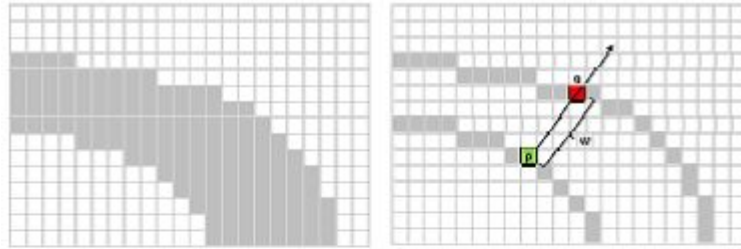
-1	-2	-1
0	0	0
1	2	1

Horizontal

-1	0	1
-2	0	2
-1	0	1

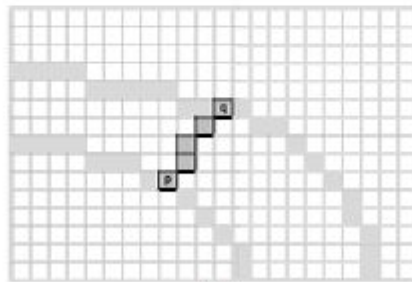
Vertical

# Stroke Width Transform



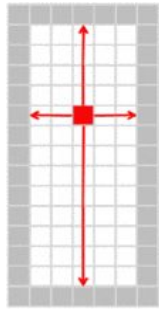
(a)

(b)

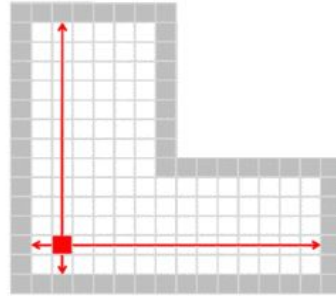


(c)

# Stroke Width Transform



(a)



(b)



# Finding Letter Candidates

(SWT ratio is less than 3.0 is a connected component)

- **Variance of SWT:** The selected threshold is half the average stroke width of a particular connected component.
- **Aspect Ratio:** We limit the aspect ratio to be between 0.1 and 10.
- **Ratio between the diameter of the component and its median stroke width:** We limit this ratio to be a value less than 10.
- **CC size:** We limit the width and font height to be between 10 and 300 pixels.

# Grouping Letter Candidates into Text Regions

1. The ratio between the median stroke-widths must not exceed 2.0.
2. The ratio between the heights of the letters must not exceed 2.0. This is due to capital letters next to lower case letters.
3. The distance between letters must not exceed three times the width of the wider one.

# Grouping Letter Candidates into Text Regions

- The candidate pairs determined are clustered together into chains.
- Initially, each chain consists of a single pair of letter candidates.
- Two chains can be merged together if they share one end and have similar direction. The process ends when no chains can be merged.
- Each produced chain of sufficient length (at least 3 letters) is considered to be a text line.

# Experiments

The implementation can be found here:

<https://github.com/akilarif/Text-Detection-in-Natural-Images-Using-SWT-and-MSER-followed-by-SWT>

# Results- SWT Only



Source Image



Labels



Final Result

# Results- SWT Only



Source Image



Labels



Final Result

# Results- SWT Only



Source Image



Labels



Final Result

## Results- SWT Only



Source Image



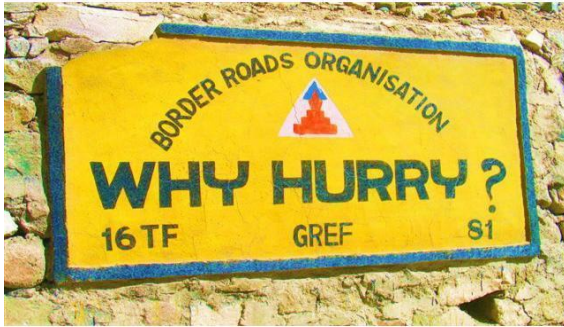
Labels



Final Result



# Results- SWT Only- Fail Cases



Source Image



Labels



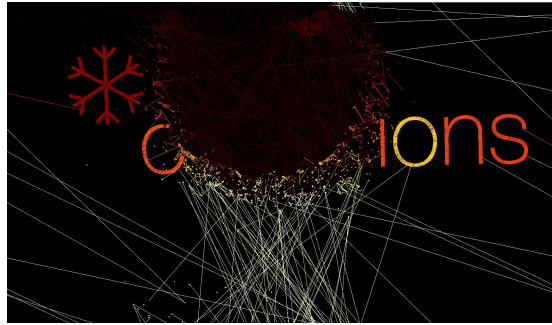
Final Result

Round and curved letters may not be detected. If the thresholds in the 'grouping letter candidates into text regions' step are relaxed, then the curved letters might be recognized, but there will be more noise.

# Results- SWT Only- Fail Cases



Source Image



Labels



Final Result

This algorithm fails when there are strong highlights in the image.

# Results- MSER Detection Followed By SWT



MSER Detection



Final Result

The final result in this case is worse than that obtained by SWT only algorithm. This is because the MSER detection didn't detect the text regions correctly.

# Results- MSER Detection Followed By SWT



MSER Detection



Final Result

There is slightly less noise in the final result of this algorithm compared to SWR only algorithm.

# Results- MSER Detection Followed By SWT



MSER Detection



Final Result

The final results of this algorithm and SWT only algorithm are pretty much the same.

# Results- MSER Detection Followed By SWT



MSER Detection



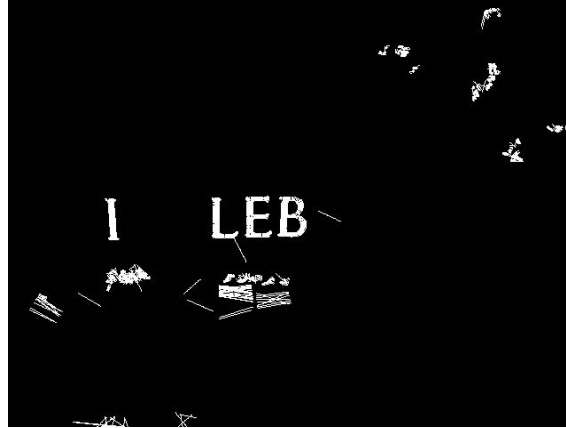
Final Result

The noise in the top right corner in the final result of SWT only algorithm is not present in the final result of this algorithm. But, there is a small addition of noise near the end of second line. Both the final results are very similar.

# Results- SWT Only vs MSER+SWT



MSER Detection



Final Result- SWT Only



Final Result- MSER + SWT

In this case, SWT gave a worse result due to the presence of foliage close to the text region. MSER detection eliminated some of the foliage.

# Results

**Strengths of the algorithm:** Can detect letters of different languages. The texts can be of varying sizes.

**Weakness of the algorithm:** Noise can be detected in certain cases. This becomes more prominent when there is foliage in the image. Round and curved letters may not be detected. Also, this algorithm fails when there are strong highlights in the image.



## Future work

1. The grouping of letters can be improved by considering the directions of the recovered strokes. This may allow the detection of curved text lines as well.
2. Better connected component algorithm to improve grouping.
3. Better edge detection stage by using a better edge detector that can compensate for noise to a certain extent (for example, **bandlet based edge detector**).

# References

1. Boris Epshtein, Eyal Ofek, and Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform", Proc. 23rd IEEE Conference on Computer Vision and Pattern Recognition, pp. 2963-2970, 2010.
2. X. Chen, A. Yuille, "Detecting and Reading Text in Natural Scenes", Computer Vision and Pattern Recognition (CVPR), pp. 366-373, 2004.
3. R. Lienhart, A. Wernicke, "Localizing and Segmenting Text in Images and Videos", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 4, April 2002, pp. 256- 268.
4. J. Gllavata, R. Ewerth, B. Freisleben, "Text Detection in Images Based on Unsupervised Classification of High-Frequency Wavelet Coefficients", 17th International Conference on Pattern Recognition (ICPR 2004) - Vol. 1, pp. 425-428 .
5. H. Li, D. Doermann, O. Kia, "Automatic Text Detection and Tracking in Digital Video", IEEE Transactions on Image Processing, Vol. 9, No. 1, January 2000 .
6. Text Detection On Scene Images Using MSER,international Journal of Research in Computer and Communication Technology, Vol 4, Issue 7 , July -2015
7. A. Jain, B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition 31(12): 2055-2076 (1998)

# References continued

1. H-K Kim, "Efficient automatic text location method and content-based indexing and structuring of video database". J Vis Commun Image Represent 7(4):336–344 (1996)
2. Y. Liu, S. Goto, T. Ikenaga, "A Contour-Based Robust Algorithm for Text Detection in Color Images", IEICE TRANS. INF. & SYST., VOL.E89–D, NO.3 MARCH 2006
3. Image Text Detection Using a Bandlet-Based Edge Detector and Stroke Width Transform.
4. K. Jung, K. Kim, A. K. Jain, "Text information extraction in images and video: a survey", Pattern Recognition, p. 977 – 997, Vol 5. 2004.
5. X. Chen, A. Yuille, "Detecting and Reading Text in Natural Scenes", Computer Vision and Pattern Recognition (CVPR), pp. 366-373, 2004
6. R. Lienhart, A. Wernicke, "Localizing and Segmenting Text in Images and Videos" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 12, NO. 4, APRIL 2002, pp. 256- 268
7. A. Jain, B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition 31(12): 2055-2076 (1998)
8. H-K Kim, "Efficient automatic text location method and content-based indexing and structuring of video database". J Vis Commun Image Represent 7(4):336–344 (1996)

THANK YOU