

Objective

To develop a novel language agnostic text detection method utilizing edge enhanced Maximally Stable Extremal Regions in natural scenes by defining strong charactereness measures.

Introduction

- Text co-occurring in images and videos serve as a warehouse for valuable information for describing images.
- A few interesting applications are
 - Extract street names, numbers, textual indications such as “diversion ahead”
 - Autonomous vehicles- follow traffic rules based on road sign interpretation
 - Indexing and tagging of images

Performing the above tasks is trivial for humans but segregating it against a challenging background still remains as a complicated task for machines.

Related Works

- Maximally Stable Extremal Regions (MSERs)
 - With Canny Edge Detector
 - MSER is applied to the image to determine regions with characters
 - Pixels outside of Canny Edges are removed
 - With Graph Model
 - Apply MSER for generating blobs
 - Generate a graph model using the positioning, color etc of graphs
 - Then define cost functions to separate foreground and background regions
- Stroke Width Transform
 - Finds stroke width for each image pixel
 - A stroke is a contiguous part of an image that forms a band of nearly constant width

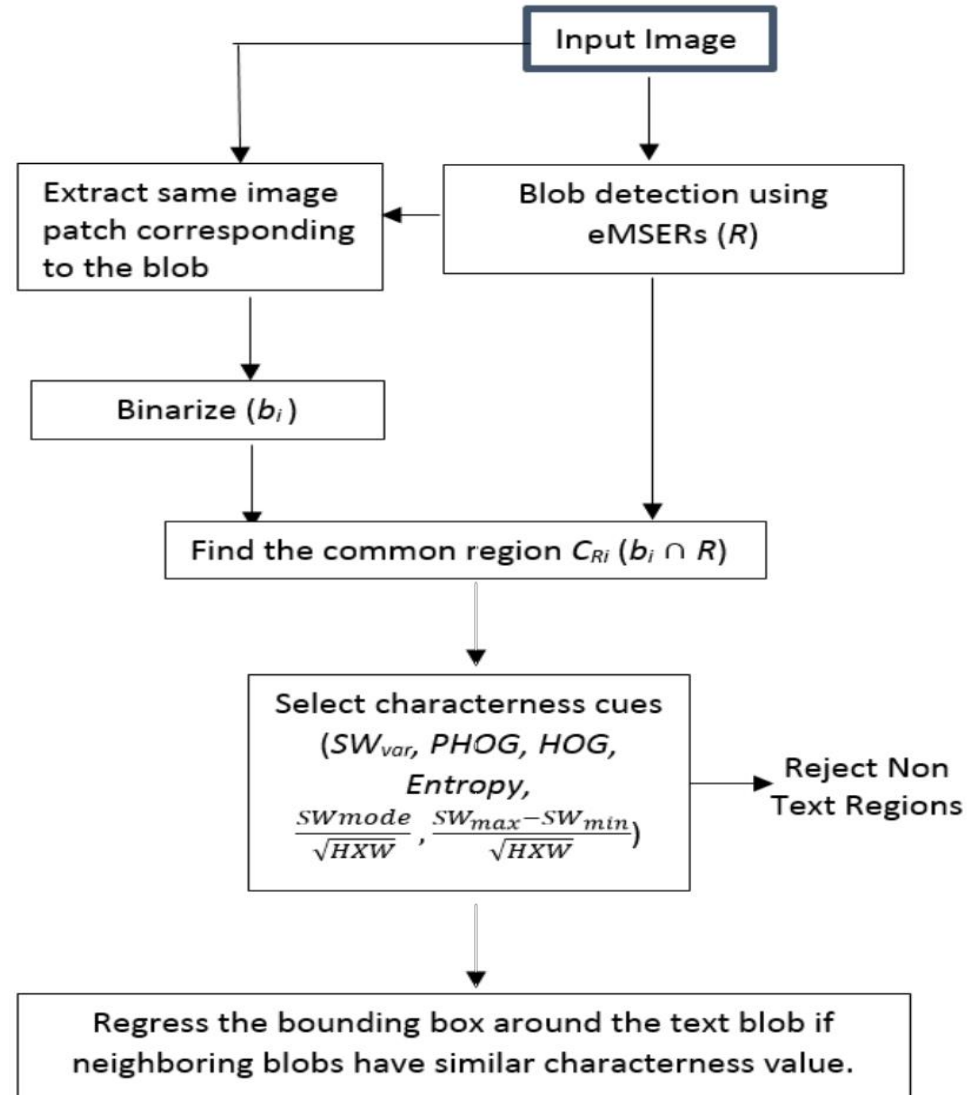
Related Works

- Feature based techniques
 - Histogram of Oriented Gradients
 - Gabor based features
 - Shape descriptors
 - Fourier Transform
 - Zernike moments
- Characterness
 - Text specific saliency detection method
 - Uses saliency cues to accentuate boundary information

Contributions

- We develop a language agnostic text identification framework using text candidates obtained from edge based MSERs and combination of various characteriness cues. This is followed by an entropy assisted non-text region rejection strategy. Finally, the blobs are refined by combining regions with similar stroke width variance and distribution of characteriness cues in respective regions
- We provide comprehensive evaluation on popular text datasets against recent text detection techniques and show that the proposed technique provides equivalent or better results.

Methodology



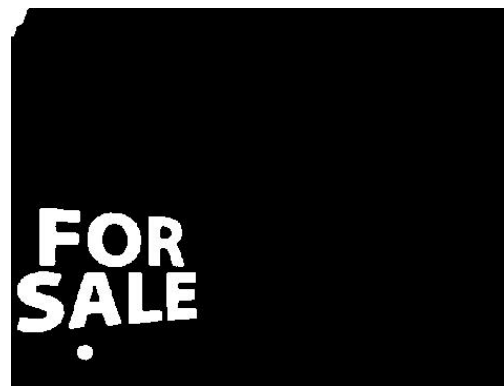
Methodology

Text candidate generation using eMSERs:

- Generate initial set of text candidates using edge enhanced Maximally Stable Extremal Regions (eMSERs) approach.
 - MSER is a method for blob detection which extracts the covariant regions.
 - It aggregates region with similar intensity at various thresholds.
 - In order to handle presence of blur, eMSERs are computed over the gradient amplitude based image.
- Two sets of regions are generated: dark and bright; dark regions are those with lower intensity than their surroundings and vice-versa
- Non text regions are rejected based on geometric properties such as aspect ratio, number of pixels(to reject noise) and skeleton length.



Original Image



Lighter side

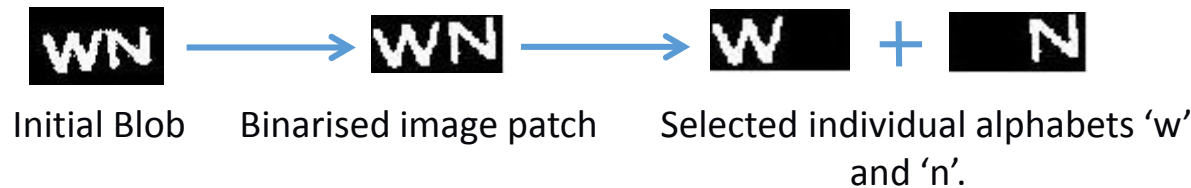


Darker Side

Methodology

Elimination of non-text regions:

- Text usually appears on a surrounding having a distinctive intensity.
 - Find corresponding image patches, R , for eMSER blobs. As the patch may contain spurious data, we obtain binarized image patch b_i using Otsu's threshold for that region and common region, C_{R_i} between b_i and R . Retain blob if $(b_i \cap R > 90\%)$.



- Define various characteriness cues:
 - *Stroke width variance*: For every pixel p in the skeletal image of region (r) to the boundary of the region, $SW(p)$ distribution is obtained and following are evaluated:

$$\frac{\text{var}(SW)}{\text{mean}(SW)^2}$$

$$\frac{\text{max}(SW) - \text{min}(SW)}{\sqrt{HXW}}$$

$$\frac{\text{mode}(SW)}{\sqrt{HXW}}$$

- *HOG and PHOG*: HOG is invariant to geometric and photometric transformations. PHOG helps in providing a spatial layout for the local shape of the image.

- *Entropy*: Calculated as Shannon's entropy for the common regions ($b_i \cap R$) given as,

$$H = -\sum_{i=0}^{N-1} p_i \log p_i$$

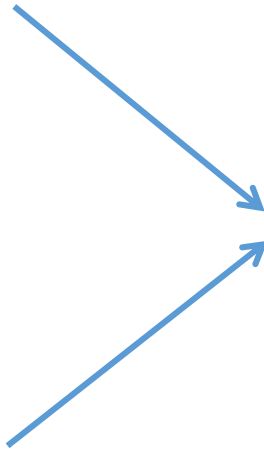
where $N = \#$ gray levels ;

$p_i =$ probability associated to the gray level i

Methodology

Bounding Box Refinement:

- Characterness cue distribution is defined by computing values for ICDAR 2013 dataset.
- Using above distribution, stroke width distribution and stroke width difference combine the neighboring candidate regions and aggregate them into one larger text region.
- Combine all the neighboring regions into a single text candidate.



Smaller regions selected as individual blobs










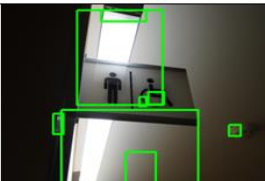




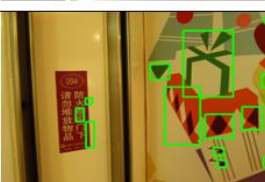







Final result after combining them

Results

Training and Testing:

Training is performed on ICDAR 2013 dataset while the test set consists of MSRATD and KAIST datasets. This setting makes the evaluation potentially challenging as well as allows to evaluate the generalization ability of various techniques.

Qualitative Results

Original Image & Ground Truth	<u>Characterness</u>	Blob Detection	Proposed Approach (individual)	Proposed Approach (combined)
				
				
				
				

Results

Quantitative Results

MSRATD

	Precision	Recall	F- Measure
Proposed	0.85	0.33	0.46
Characterness [1]	0.53	0.25	0.31
Blob Detection [2]	0.8	0.47	0.55
Epshtein et al. [3]	0.25	0.25	0.25
Chen et al. [4]	0.05	0.05	0.05
TD-ICDAR [5]	0.53	0.52	0.5
Gomez et al. [6]	0.58	0.54	0.56

KAIST - English

	Precision	Recall	F- Measure
Proposed	0.8485	0.3299	0.4562
Characterness	0.5299	0.2467	0.3136
Blob Detection	0.8047	0.4716	0.5547

KAIST - Korean

	Precision	Recall	F- Measure
Proposed	0.9545	0.3556	0.4994
Characterness	0.7263	0.3209	0.4083
Blob Detection	0.9091	0.5141	0.6269

KAIST - Mixed

	Precision	Recall	F- Measure
Proposed	0.9702	0.3362	0.4838
Characterness	0.8345	0.3043	0.4053
Blob Detection	0.9218	0.4826	0.5985

KAIST - All

	Precision	Recall	F- Measure
Proposed	0.9244	0.3407	0.4798
Characterness [1]	0.6969	0.2910	0.3757
Blob Detection [2]	0.8785	0.4898	0.5933
Gomez et al. [6]	0.66	0.78	0.71
Lee et al. [7]	0.69	0.60	0.64

Conclusion

- Proposed a language agnostic text identification scheme using text candidates obtained from edge based eMSERs.
- Processing steps are used to reject the non-textual blobs and combine smaller blobs into one larger region by utilizing stronger characteriness measures.
- The effectiveness has been analyzed with precision, recall and F-measure evaluation measures showing that the proposed scheme performs better than the traditional text detection schemes.

References

- [1] Li, Yao, Wenjing Jia, Chunhua Shen, and Anton van den Hengel. "Characterness: An indicator of text in the wild." *IEEE transactions on image processing* 23, no. 4 (2014): 1666-1677.
- [2] Jahangiri, Mohammad, and Maria Petrou. "An attention model for extracting components that merit identification." In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 965-968. IEEE, 2009.
- [3] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2963-2970. IEEE, 2010.
- [4] Chen, Xiangrong, and Alan L. Yuille. "Detecting and reading text in natural scenes." In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II-II. IEEE, 2004.
- [5] Yao, Cong, Xiang Bai, Wenyu Liu, Yi Ma, and Zhuowen Tu. "Detecting texts of arbitrary orientations in natural images." In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1083-1090. IEEE, 2012.
- [6] Gomez, Lluís, and Dimosthenis Karatzas. "Multi-script text extraction from natural scenes." In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pp. 467-471. IEEE, 2013.
- [7] Lee, SeongHun, Min Su Cho, Kyomin Jung, and Jin Hyung Kim. "Scene text extraction with edge constraint and text collinearity." In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 3983-3986. IEEE, 2010.



Thank You