

An Overview of Text Detection in Natural Scene Images

A.Vishnuvardhan, M.N.Sriharsha

Abstract—Finding and ousting the words or sentences from the essential pictures is one of the standard testing district in PC vision which generally requires post-organizing steps. It is a framework which is used to see and confine the perfect substance from the photos. In spite of how examination is done on this, yet in the interim there is need to improve the precision of the ousted bit of the substance. Thusly a solid structure is required.

Keywords — Text detection, Text Extraction, Natural Image, Computer vision.

1. INTRODUCTION

Earlier than starting off to the substance zone there is a want to study the one-of-a-kind solicitations of pictures concerning the shading and structures; those are customary pictures, organized pictures and uncertain snap shots [4]. center is commonly on to look regular and made. ordinary snap shots which might be typical referred to as optical snap shots which are crossed the digicam or PDAs and those are accrued into 3 kinds which can be outdoor, indoor and films. alternatively made photographs are enlisted reliant on some different type of flag, activities of made photographs are CAT Scans in which 'C' in CAT tends to organized and radar photographs which sets SAR in which the 'S' addresses organized.

All around message affirmation become dominantly installed on killing surveyed records, regardless without scarcely lifting a finger purchaser matters, for example, mobile telephones, capsules and desktops, so the patron applications associated with automated image managing have ended up being noteworthy [3]. anyhow the changed substance request in conventional pics is a to this point trying out one. because the revealed records is to be had anywhere i.e., each inside and outside. There are extensive assembling of groups that joins, support to surely hindered human beings, modernized direction and scene understanding, and so forth [1].,

2. CHALLENGES

If you have to see and withdraw the printed substance from the trademark picture, the standard preventions that are looked by the authorities are wide style of inventive substance appearance in setting on breathtaking substance styles, thicknesses, tints, sizes, surfaces, notwithstanding the closeness of geometrical twistings and midway obstacles in the pics, uncommon lights conditions and photograph objectives [2]. exact and weighty substance domain in home made scene photos is starting in the no so distant past an open issue in this field due to exceptional characteristics of scene tasteful substance. driving a wonderful reputation rate for scene dynamic substance remains a lessen need in view of nonattendance of strong and unbelievable binarization procedure that segregates nearer watch (printed substance) and establishment (non-canny substance) enjoyably. achieving confirmation charge for bowed printed content looks for after is other than an unexplored region in this subject.

3. EXISTING METHODOLOGIES

Content insistence in a trademark picture is an extraction of substance zones which all things considered incorporates five phases:

- i) Acquiring the image through the camera
- ii) Perform the pre-getting ready handle the image for the subsequent advances
- iii) Extract the contender character zone reliant on the some substance features
- iv) Ensuring the character spaces using the classifiers.
- v) These character spaces are gathered into substance regions

There are ordinarily three existing rationalities for substance disclosure in scene pictures: a) Texture based procedure or Sliding Window based systems b) Connected piece or Region based approachs. what's more, c) Hybrid structures

a) Texture based structures: These theories treat messages as a wonderful sort of surface and use their textural properties, for instance, neighborhood powers, channel responses and wavelet coefficients, to see content and non content zones in the photos. These systems are ordinarily computationally exorbitant as all areas and scales should be inspected. Basically, these frameworks overall handle even messages and are sensitive to turn and scale change.

Revised Manuscript Received on May 29, 2019.

A.Vishnuvardhan Asst. Professor, VVIT, Nambur, Guntur(dt.), AP, India

M.N.Sriharsha Asst. Professor, VVIT, Nambur, Guntur(dt.), AP, India.



Texture based Methods				
Author (s)	Feature (s)	Classifier	Methodology	Performance
Kim, K.I., Jung, K., Jin, H.K [5]	By using with the feature of color clustering, candidate character regions were extracted	SVM is used to ensure that the extracted region was a character.	By using with the help of Continuous Adaptive shift algorithm	Detection Rate is 87% using 50 images of different size, fonts and formats.
Pan, Y.F., Hou, X., Liu, C.L [6]	HOG & LBP	Cascade Adaboost	Window grouping method is used to generate text lines and then Markov Random Fields is used to filter the non-text regions.	Precision - 68 % Recall - 67 % on ICDAR 2003 Dataset.
Ye, J., Huang, L.L., Hao, X [7]	Gray Scale Invariance of LBP and Modified LBP	Polynomial Neural Networks	Verification and fusion is used to provide the text regions.	Precision – 68% Recall – 87.7% on ICDAR 2003 Dataset.
Lee, J., Lee, P.H., Lee, S.W., et al [8]	Six different classes of features	Modest Adaboost	Multi scale sequential search	Precision – 66% Recall – 75% on ICDAR 2003 Dataset.
Song, Y., He, Y., Li, Q., et al [9]	Haar like features i.e, edge, line and center surround features.	Adaboost	Adaboost is combined with Haar like features to obtain cascade classifiers for text regions extraction.	Precision – 72.6% Recall – 79.9% using 128 street view images.
Gllavata, J., Ewerth, R., Freisleben, B [10]	Wavelet transform	Distribution of high frequency wavelet coefficients.	K-means and projection analysis.	Precision – 87% Recall – 90% on video frames taken from the MPEG-7 video test set.
Shivakumara, P., Phan, T.Q., Tan, C.L [11]	Wavelet transform	Distribution of high frequency wavelet coefficients.	New Fourier-Statistical Features (FSF) in RGB space to detect texts of different fonts, size and scripts	Precision – 92.7% Recall – 93.2% on video frames taken from the MPEG-7 video test set.

a) Connected Component Based Methods: These techniques first concentrate contender partitions through a social affair of ways (e.g., shading bundling or silly area extraction), and after that channel through non-content parts using physically formed benchmarks or thusly sorted out classifiers. As a last resort, these structures are extensively

coherently stunning, in light of the way in which that the proportion of parts to be readied is tied in with nothing. In like manner, these approaches are unfeeling toward turn, scale change and academic style mix. These procedures are reasonably new when risen up out of the surface based structures

Connected Component Based Methods			
Author(s)	Approach to extract the character regions.	Methodology to get the text regions.	Results
Lyu, M.R., Song, J., Cai, M [12]	Local thresholding and hysteresis edge recovery are applied to get the character regions.	A local adaptive binary strategy to segment detected text areas.	Detection Rate – 91.1% Detection Accuracy – 90.8%
Zhao, M., Li, S., Kwok, J [13]	Simple classification procedure with two learned discriminative dictionaries are applied to get character regions.	Projection analysis is used to group the character regions into text and then adaptive run-length smoothing algorithm was used to refine the text areas.	Precision – 98.8% Recall – 94.2% on Microsoft common test set

Epshtein, B., Ofek, E., Wexler, Y [14]	Detect the connected components with the similar stroke width as character regions.	Text lines are built with the features of shape and distance.	Precision – 73% Recall – 60% on ICDAR 2003 and ICDAR 2005 Data sets
Yao, C., Bai, X., Liu, W., et al [15]	Detect the connected components with the similar stroke width as character regions.	Color and shape of the regions are used to build the multi direction text lines.	Precision – 77% Recall – 73% on OSTD
Yin, X.C., Yin, X., Huang, K., et al [16]	An efficient pruned exhaustive search algorithm is used to filter out the nesting or duplicate regions.	Morphological features and Single link algorithm are used to group the character regions into text regions	Precision – 86.29% Recall – 68.26% on ICDAR 2011 Data set
Chucal Yi, YingLiTian [17]	SVM learning classifier is used to detect the character regions.	Text string detection method is used. It consist two step: -Image partition to find text character. -Character candidate grouping to detect text strings.	Precision – 71% Recall – 62% on Robust reading Data set
Zheng Zhang, Wei Shen, Cong Yao, Xiang Bai [18]	CNN classifier is used to detect the character regions	Sliding window and connected component Extraction are method for scene text detection	Precision – 88% Recall – 74% on ICDAR 2013 data set
Tong He, Weilin Huang, Yu Qiao, and Jian Yao [19]	Text CNN Classifier	Recent deep learning Learning models are used in this paper for text detection. It has capable of high level feature from whole image.	Precision – 91% Recall – 74% on ICDAR 2011 data set

b) Hybrid Methods: These methods are a combination of texture based methods and component based methods, which make use of the advantages of these two types methods.

Hybrid Methods		
Author(s)	Methodology	Results
Y. Liu, S. Gotoand, and T. Ikenaga.[20]	Edge pixels of all possible text regions were extracted using an elaborate edge detection strategy, and the gradient and geometrical properties of region contours are verified to generate candidate text regions	Precision – 73.4% Recall – 79.3% on 529 ICDAR Contest images
Y. Pan, X. Hou, and C. Liu.[21]	Candidate components are extracted from multi-scale probability maps. The probability maps are estimated by classifier, which is trained on a set of texture features	Precision – 67.4% Recall – 69.7% on Multilingual data set
J. D. Lafferty, A. McCallum, and F. C. N. Pereira [22]	A Conditional Random Filed (CRF) model combining unary component properties and binary contextual relationships, is utilized to discriminate text components from non-text components.	On 2, 000 training and 500 test samples, trained to convergence of the iterative scaling algorithm, the CRF error is 4.6%

4. FUTURE TRENDS:

In real conditions, sytheses can be in different presentations But winning piece of the overseers in the fields just gave their plan to level messages. To manhandle hypothetical information in regular scenes, it is essential to examine works of different presentations. By a wide edge most of the present systems are concerned over substance in

English Considering practicability, it is basic to make ID and confirmation structures that can administer works of different vernaculars. The mix of gigantic learning structures and huge degree of organizing data seems standard the fields of scene content region and verification. Past important

learning based strategies just understood make systems from various locales and achieved execution bolsters over standard estimations. Further improvement in obvious affirmation and statement accuracy can be created, if the significant learning structure is used to discover and exhibit the properties of scene content from colossal volume of data.

REFERENCES

1. Andrei Polzounov,ArtsiomAblavatski , Sergio Escalera, Shijian Lu, JianfeiCai “Wordfence: Text Detection In Natural Images With Border Awareness” arXiv:1705.05483v1 [cs.CV] 15 May 2017
2. ShangxuanTian, Yifeng Pan, Chang Huang, Shijian Lu, KaiYu, and Chew Lim Tan, “Text flow: A unified text detection system in natural scene images,” in Proceedings of the IEEEInternational Conference on Computer Vision, 2015, pp. 4651–4659.
3. A. Gonzalez, L.M. Bergasa“A Text Reading Algorithm for Natural Images” Department of Electronics, University of Alcalá, University Campus, 28871 AlcaladeHenares (Madrid), Spain.
4. Maurizio Garbarino “Automatic Classification of Natural and Synthetic Images” Master of Science Thesis Stockholm, Sweden 2008
5. Kim, K.I., Jung, K., Jin, H.K.: Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. *IEEE Trans.Pattern Anal. Mach. Intell.* 25(12), 1631–1639 (2003)
6. Pan, Y.F., Hou, X., Liu, C.L.: A robust system to detect and localize texts in natural scene images. In: The Eighth IAPR International Workshop on Document Analysis Systems, pp. 35–42. IEEE (2008)
7. Ye, J., Huang, L.L., Hao, X.: Neural network based text detection in videos using local binary patterns. In: Chinese Conference on Pattern Recognition, CCPR 2009, pp. 1–5 (2009)
8. Lee, J., Lee, P.H., Lee, S.W., et al.: AdaBoost for text detection in natural scene. In: International Conference on Document Analysis and Recognition, pp. 429–434. IEEE Computer Society (2011)
9. Song, Y., He, Y., Li, Q., et al.: Reading text in street views using Adaboost: towards a system for searching target places. In: 2009 IEEE Intelligent Vehicles Symposium, pp. 227– 232. IEEE (2009)
10. Gllavata, J., Ewerth, R., Freisleben, B.: Text detection in images based on unsupervised classification of high-frequency wavelet coefficients. *Proc. Int. Conf. Pattern Recogn.* 1(3), 425–428 (2004)
11. Shivakumara, P., Phan, T.Q., Tan, C.L.: New fourier-statistical features in RGB space for video text detection. *IEEE Trans. Circ. Syst. Video Technol.* 20(11), 1520–1532 (2010)
12. Lyu, M.R., Song, J., Cai, M.: A comprehensive method for multilingual video text detection, localization, and extraction. *IEEE Trans. Circ. Syst. Video Technol.* 15(2), 243–255 (2005)
13. Zhao, M., Li, S., Kwok, J.: Text detection in images using sparse representation with discriminative dictionaries. *Image Vis. Comput.* 28(12), 1590–1599 (2010)
14. Epshtein, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2963–2970. IEEE (2010)
15. Yao, C., Bai, X., Liu, W., et al.: Detecting texts of arbitrary orientations in natural images. In: IEEE Conference on Computer Vision & Pattern Recognition, pp. 1083–1090 (2012)
16. Yin, X.C., Yin, X., Huang, K., et al.: Robust text detection in natural scene images. *IEEE Trans. Pattern Anal. Mach. Intell.* 36(5), 970–983 (2013)
17. C. Yi and Y. Tian, “Text string detection from natural scenes by structure-based partition and grouping,” *IEEE Trans. Image Processing (TIP)*, vol. 20, pp. 2594–2605, 2011
18. Z. Zhang, W. Shen, C. Yao, and X. Bai, “Symmetry-based text line detection in natural scenes,” 2015, in *IEEE Computer Vision and Pattern Recognition (CVPR)*.
19. Tong He, Weilin Huang, Yu Qiao, and Jian Yao, “Text AttentionalCovnvolutional Neural Network for Scene Text Detection” arXiv:1510.03283v2 [cs.CV] 24 Mar 2016
20. Y. Liu, S. Gotoand, and T. Ikenaga. A contour-based robust algorithm for text detection in color images. *IEICE Transactions on Informationand Systems*, 89(3):1221–1230, 2006.
21. Y. Pan, X. Hou, and C. Liu. A hybrid approach to detect and localize texts in natural scene images. *IEEE Trans. Image Processing*, 20(3):800–813, 2011.
22. J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data In *Proc. of ICML*, 2005.