International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 05,
2019 ISSN: 1475-7192

Text Detection Using Image Processing: A Survey

[1]Laxmi Prasad mishra, [2]Mihir Narayan Mohanty

Abstract--- *The aim of text recognition is to recognize the text from written hard copy documents to the required format. The process of text recognition including many steps as well as preprocessing, image segmentation, feature extraction, classification, post-processing. Preprocessing is for doing the essential operation on input image like binarization that converts gray Scale image into Binary Image, noise reduction that removes the noisy signal from an image. Segmentation process for the segment the given image into line by line and segment every character from the segmented line. Future extraction calculates the characteristics of a character. A text classification contains the information and will the comparison. Today, it plays a crucial role within the workplace, university, etc. necessary approaches wont to undergo these stages and their corresponding advantages, disadvantages, and application are presented during this article, numerous text-related applications for imagery also is presented over here. This review performs a comparative analysis of elementary processes during this field.*

Index Terms--- *Text Detection, Classification, Preprocessing, Segmentation, Object Detection.*

I. INTRODUCTION

Text detection and recognition have emerged as a very important drawback within the past few years. Advancements within the field of Image processing and machine learning, similarly as an increase within the applications based on text detection using image processing techniques, have resulted during this trend[1]. Text recognition from video captions further as web content also are getting attention, large work has been done in the field of text detection and recognition from natural scenes image. The number of optical character recognition techniques also are on the market. Still, the problem of text detection and recognition isn't completely solved. Segmentation and extraction of text from natural scenes are still very difficult to achieve[2]. Text detection employed in an official task during which the massive information has to be compelled to type like post offices, banks, colleges, etc., in real-life applications wherever this paper would like to collect some data from text written images. Individuals want to scan during a document and have the text of that document accessible in a .txt or doc format[3].

Different ways of text detection from the image and video are proposed for the last decade. Most of them are divided into the subsequent stages, spatial text extraction, temporal text detection, image binarization image segmentation, and character recognition[4]. From these stages, the foremost crucial is the stage of spatial image detection that really concentrates the focus of most researchers during this field[5]. The performance of the complete text detection system depends on the accurate localization of text in multimedia, a lot of lesser work has been finished the timely detection of

Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, laxmimishra@soa.ac.in, mihirmohanty@soa.ac.in

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

text in videos, whereas, typically the only static text is considered. Some strategies have also been proposed for the binarization of the text image, though most researchers use progressive algorithms from the classic document analysis research area[6].

Finally, a really important aspect that has not been sufficiently studied is the development of the corresponding evaluation protocols that are necessary for optimizing the algorithms similarly to comparing many ways within the literature[7]. During this section, this paper will define the techniques found in the literature for various stages of text detection from images and video frames. The text regions are extract under the assumption that the region of the image background and the object regions are sparser than those of the text regions. However, this kind of approaches is not very effective to extract texts with large font size[8].

II. RELATED WORK

The analysis of a text extraction method is an aspect not as trivial because it might seem. It resembles the generic problem of object detection analysis having, in addition, its own problems. Most researchers use for his or her experimentation straightforward boxed-based or area-based strategies, whereas, [9]only a few works have targeted the particular problem of analysis. An overall measure of text extraction during a frame, a box-based referred to as Frame Detection Accuracy (FDA). The evaluation ways of this type have supported the mapping between ground truth and objects detected. Particularly for the text extraction problem,[10] text lines are thought-about to be the objects wherever a text line is typically defined as aligned series of characters with a small intermediate distance relative to their height. Text detection deals with detecting the presence of the text within the input image,[11] whereas, text localization localizes the position of the text and forms groups of text regions by eliminating most of the background. Text detection and localization method are performed using connected component analysis or region-based strategies.

Text Detection:

In this section, different stages of text extraction and recognition are represented. Text detection and localization, image classification, image segmentation, and text recognition are represented with their role and importance[12]. Numerous methods followed for these stages are explained during this section. Text extraction deals with detecting the presence of the image text within the input image whereas text localization localizes the position of the text and forms teams of text regions by eliminating most of the background. Text extraction and localization methods are performed using connected component analysis or region-based techniques[13].

The connected component analysis methodology forms a graph of connected points supported color or edge features from the binarized image. Region-based strategies divide images into small regions using windows and search these regions for the presence of text using texture or morphological operations since text and non-text regions have totally different textual properties[14], [15]. Multimedia classification to classify text and non-text considering giant variations in text size. The unsupervised machine learning techniques is employed by, on little 8×8 grayscale patches of image for feature extraction learning. It uses 32×32 pixels of image for feature detection, text detection training, and character

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

classifier training[16].

Classification:

After text extraction stage output could contain non-text regions together with text regions as false positives. The text classification stage verifies text regions and eliminates non-text regions using image classification algorithms[17]. This stage may be referred to as verification, classification algorithms are either supervised or unsupervised. Supervised machine learning algorithms are aware of the properties of text detection like color, size, texture, etc. before text classification[18]. Unsupervised machine learning algorithms don't have prior data regarding text features. Supervised classification algorithms would like training before classification. These algorithms undergo training to be able to detection features of the text to be classified and use these features in the classification section[19]. Supervised machine learning algorithm constraints on edge area additionally as area, image height and dimension constraints on block obtained within the text detection stage are employed in for text classification.

Text classification is done using the features detection in the previous step, which corresponds to each character recognize[20]. These text features are analyzed using the set of rules and labeled as belonging to different classes. This text classification is generalized such that it works for single font type.

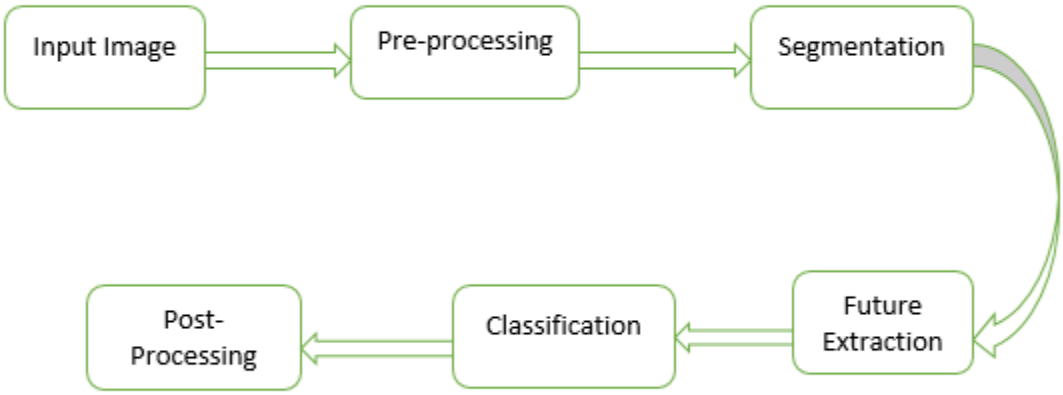

Fig. 1: Text Detection and Recognition

After text classification, algorithm will check the classified data with the database where this paper has already saves the classification of text, text numbers [21]. Then algorithm will display the result according to the comparison. Here this algorithm will check the pattern of inputted character with database. Unsupervised text classification algorithm doesn't undergo the training phase. The text extract features throughout the image classification phase-only not like supervised classification and that they use features detection within the previous text classification for the next one [22]. This can be similar to an adaptive learning algorithm. Wavelet transform which provides sequential approximation through low-pass filter and details of edges and different features from the high-pass filter is used in.

Segmentation:

The image segmentation method is employed to separate text detection from the background and to detection bounded text from an image. Integrated ways that focus on word matching/recognition usually mix or replace advanced segmentation stage with recognition stage but stepwise strategies undergo segmentation to get exactly extracted characters that are fed to the recognition stage[23]. Binarization, text segmentation are few of the

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

segmentation algorithms studied during this article. Binarization converts color or gray-scale images into black and white images. To realize good segmentation result no matter dark or bright text or background, adaptive thresholding for binarization uses the k-means clustering algorithm for binarization. It uses k=4 and k=5 as cluster parameters and text classifies the binarized image into texts using probabilistic models[24].

For text segmentation, the image must be divided row-wise (line segmentation), then every row has to be divided column-wise (word segmentation). Finally, the text will be detected using suitable algorithms like edge detection technique;[25] histogram-based techniques or connected component analysis. Connected components are the algorithm of graph theory, wherever subsets of connected components are unambiguously labeled based on a given heuristic[26].

The connected component algorithm is employed in computer vision to detect connected regions in binary digital images, though color images and data with higher-dimensionality may also be processed[27]. When integrated into an image recognition system or human-computer interface, connected component labeling will operate on a spread of data technique.

III. TEXT RECOGNITION

The text recognition process converts images of text into a string of characters or words. It's important to convert images of text into words because the word is an elementary entity employed by humans for his visual recognition[28]. The completely different method of recognition is character recognition and word recognition. Character recognition techniques divide text images into multiple cut-outs of single characters. The separation between adjacent characters is extremely important for these ways[29]. These techniques are principally based on the concepts of binarization, edge detection and spatial frequency analysis of images. For instance, a system was proposed that performs text detection, image segmentation, and text recognition, during a single framework that helps in accommodating contextual relationships[30]. They performed the blue and white color segmentation on the traffic panel images to search out the interesting key points within the image then perform the character recognition within the interested area. Text recognition from traffic signs framework,[31] for localizing the texts they proposed a system based on region proposal mechanism and used a Convolutional Neural Network (CNN) for recognition purposes.

The text recognition in video frames is more difficult than that of natural scene images. Only a number of items of work exist for text recognition from video frames. The difficulties in video text analysis are low-resolution frames, advanced background,[32] variation in color, non-uniform illumination, font, font size, camera motion, and different blurring artifacts, etc. Hence, presently available OCR techniques provide poor performance in video text images. Most of the recognition techniques perform the pre-processing of images before applying the standard segmentation algorithm on video frames[33].

Many pieces of work are proposed previously on the binarization of images. Binarization technique based on a Convolutional Neural Network (CNN) for the color text area of video frames and its performance depends on the number of training samples used. In these approaches, binarization in several color channels was also not explored. In pixel-level strategies,[23] an end-to-end fully convolutional network learns to come up with a dense prediction map indicating

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

whether or not every pixel within the original image belongs to any text instances or not. Post-processing techniques then group pixels together depending on that pixel of image belong to an equivalent text instance. Since text can appear in clusters that make predicted pixels connected to every different, the core of pixel-level strategies is to separate text instances from one another[34].

Proposed a completely unique feature extraction technique for color scene/video images during this article. Most of the present work on retrieval of words from scene/video images in the main concentrate on the binarizing image within the first step and then features are extracted from those binarized images[35].

However, binarization is troublesome in scene/video images because of a very uncertain variation of light and illumination of the pictures because of improper binarization, important data could get lost unintentionally, [36], [37] which could be a huge hindrance to word retrieval in scene/video images. Here paper is got handled this constraint of binarization for video/scene pictures using a novel feature extraction technique.

IV. CONCLUSION

In this paper, introduce a text detection using image processing that supported a scalable feature learning algorithm and applied it to the pictures of text in natural scenes. This paper tend to demonstrate that with larger banks of features. This paper tend to be able to achieve increased accuracy with high performance corresponding to different systems, similar to results determined in different areas of computer vision and machine learning. Recognition results obtained in several data sets are encouraging. The experimental analysis shows that the binarization step can be avoided by selecting a correct color channel of a picture patch for feature detection that contains a lot of details concerning the text info. The potency of color channel choice is taken forward for text localization in video or scene images. Besides, color channel selection primarily based feature extraction can be employed in completely different texture classification tasks. In this work, this paper tend to consider a single color channel for feature extraction whereas this may be extended to multiple color channel selection for feature extraction and completely different weights may be assigned to the features of the color channel depending on the importance of the color channel.

REFERENCES

[1] S. Paudel, P. W. C. Prasad, and A. Alsadoon, Feature Selection Approach for Twitter Sentiment Analysis and Text Classification Based on Chi-Square and Naïve Bayes. 2018.
[2] J. Kumar, S. Taterh, and D. Kamnthania, “Study and Comparative Analysis of Various Image Spamming Techniques,” in Advances in Intelligent Systems and Computing, 2018.
[3] S. Yadav and A. K. Thripati, “Comparative analysis of Canny Edge based image steganography with RSA encryption,” Int. J. Adv. Sci. Technol., 2017.
[4] A. K. Jain and B. B. Gupta, “Phishing detection: Analysis of visual similarity based approaches,” Security and Communication Networks. 2017.
[5] A. Roy, A. Paul, H. Pirsiavash, and S. Pan, “Automated detection of substance use-related social media posts based on image and text analysis,” in Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2018.
[6] S. P. Faustina Joan and S. Valli, “A Survey on Text Information Extraction from Born-Digital and Scene Text Images,” Proceedings of the National Academy of Sciences India Section A - Physical Sciences. 2019.
[7] R. S. Prasad, S. Prasad, and V. Prasad, “Pattern recognition in thought-form images using radon transform and histograms,” in ACM International Conference Proceeding Series, 2017.

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

[8] D. Holland et al., "AUTOMATING THE PHOTOGRAMMETRIC WORKFLOW IN A NATIONAL MAPPING AGENCY Commission IV , Working Group IV / 3 : Mapping from High Resolution Data," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012.
[9] 9. Retrieved from aarssenl@queensu.ca Aarssen, L. W., & Crimi, L. (2016). Legacy, leisure and the 'work hard—Play hard' hypothesis. The Open Psychology Journal et al., "Playing the objectification game: How women's self-esteem impacts the existential consequences of objectification.," 2018.
[10] M. Shah and R. Kapdi, "Object detection using deep neural networks," in Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 2017.
[11] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," IEEE Trans. Pattern Anal. Mach. Intell., 2015.
[12] X. C. Yin, Z. Y. Zuo, S. Tian, and C. L. Liu, "Text Detection, Tracking and Recognition in Video: A Comprehensive Survey," IEEE Transactions on Image Processing. 2016.
[13] V. K. Pham and G. S. Lee, "Robust text detection in natural scene images," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016.
[14] H. Li, P. Wang, and C. Shen, "Towards End-to-End Text Spotting with Convolutional Recurrent Neural Networks," in Proceedings of the IEEE International Conference on Computer Vision, 2017.
[15] L. Neumann and J. Matas, "Scene text localization and recognition with oriented stroke detection," in Proceedings of the IEEE International Conference on Computer Vision, 2013.
[16] C. Yao, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," IEEE Trans. Image Process., 2014.
[17] C. Bartz, H. Yang, and C. Meinel, "See: Towards semi-supervised end-to-end scene text recognition," in 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 2018.
[18] L. Sun, Q. Huo, W. Jia, and K. Chen, "A robust approach for text detection from natural scene images," Pattern Recognit., 2015.
[19] K. Prasad, D. C. Nigam, A. Lakhotiya, D. Umre, and B. I. T. Durg, "Character Recognition Using Matlab ' s Neural Network Toolbox," Int. J. u- e- Serv. Sci. Technol., 2013.
[20] V. Abburu, S. Gupta, S. R. Rimitha, M. Mulimani, and S. G. Koolagudi, "Currency recognition system using image processing," in 2017 10th International Conference on Contemporary Computing, IC3 2017, 2018.
[21] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to Iris, fingerprint, and face recognition," IEEE Trans. Image Process., 2014.
[22] M. Opitz, M. Diem, S. Fiel, F. Kleber, and R. Sablatnig, "End-to-end text recognition using local ternary patterns, MSER and deep convolutional nets," in Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014, 2014.
[23] C. Shi, C. Wang, B. Xiao, Y. Zhang, and S. Gao, "Scene text detection using graph model built upon maximally stable extremal regions," Pattern Recognit. Lett., 2013.
[24] A. Ramisa, F. Yan, F. Moreno-Noguer, and K. Mikolajczyk, "BreakingNews: Article Annotation by Image and Text Processing," IEEE Trans. Pattern Anal. Mach. Intell., 2018.
[25] A. Lavin and S. Gray, "Fast Algorithms for Convolutional Neural Networks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.
[26] S. Emami and V. P. Suciu, "Facial Recognition using OpenCV," J. Mobile, Embed. Distrib. Syst., 2012.
[27] Y. Lecun, Y. Bengio, and G. Hinton, Artificial Neural Networks as Models of Neural Information Processing. 2015.
[28] G. Li and Y. Yu, "Visual saliency detection based on multiscale deep CNN features," IEEE Trans. Image Process., 2016.
[29] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards accurate region proposal generation and joint object detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.
[30] A. Polzounov, A. Ablavatski, S. Escalera, S. Lu, and J. Cai, "Wordfence: Text detection in natural images with border awareness," in Proceedings - International Conference on Image Processing, ICIP, 2018.
[31] S. Qin and R. Manduchi, "Cascaded Segmentation-Detection Networks for Word-Level Text Spotting," in Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2018.
[32] R. Vedantam, C. L. Zitnick, and D. Parikh, "CIDEr: Consensus-based image description evaluation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015.
[33] M. M. Kasar, D. Bhattacharyya, and T. H. Kim, "Face recognition using neural network: A review," Int. J. Secur. its Appl., 2016.

Received: September 2019 | Revised: October 2019 | Accepted: November 2019

[34] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in Proceedings - International Conference on Image Processing, ICIP, 2011.
[35] N. S. Roubtsova, R. G. J. Wijnhoven, and P. H. N. de With, "Integrated text detection and recognition in natural images," in Image Processing: Algorithms and Systems X; and Parallel Processing for Imaging Applications II, 2012.
[36] S. Theodoridis, Machine Learning: A Bayesian and Optimization Perspective. 2015.
[37] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," IEEE Trans. Image Process., 2015.

Received: September 2019 | Revised: October 2019 | Accepted: November 2019