IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 5, Ser. 3 (Sept. – Oct. 2024), PP 36-41
www.iosrjournals.org

Text Detection And Recognition: A Review

Manish Kushwaha CSE
M .Tech Chouksey College Of Engineering And Technology Bilaspur (CG)

Abstract –
This paper identifies and compares different stages in the process of text detection and recognition and analyses different approaches used for text extraction from color images. Two commonly used methods for this problem are stepwise methods and integrated methods, whereas this task is further divided into text detection and localization, classification, segmentation and text recognition. Important approaches used to undergo these stages and their corresponding advantages, disadvantages and applications are presented in this paper. Various text related applications for imagery are also presented over here. This review performs comparative analysis of fundamental processes in this field.

Keywords - *Text detection, Text Recognition, Localization, Classification, Segmentation*

Date of Submission: 29-09-2024 Date of Acceptance: 09-10-2024

I. Introduction

Text detection and recognition has emerged as an important problem in the past few years. Advancements in the field of computer vision and machine learning as well as increase in the applications based on text detection and recognition has resulted in this trend. Various workshops and conferences like International Conference on Document Analysis and Recognition (ICDAR) are being organized on international level giving further rise to developments in field of text processing from imagery.

Text detection and recognition from video captions as well as web pages is also getting attention. Huge work has been done in the field of text detection and extraction from natural scenes imagery. Various optical character recognition techniques are also available. Still problem of text detection and recognition is not thoroughly solved. Segmentation and extraction of text from natural scenes is still very difficult to achieve.

- This paper studies various stages in process of text detection and recognition and analyses and compares different approaches used to undergo these stages. It presents importance of every processing stage and advantages, disadvantages and applications of approaches used by various contributors to solve these problems. Various applications of text detection and recognition are also reviewed in this paper.
- The paper is organized as follows. Section 2 presents methodologies used for text detection and recognition. Important stages of text detection and recognition are presented in section 3. Various applications of text detection and recognition are discussed in section 4 and the paper concludes in section 5.

II. Methodologies

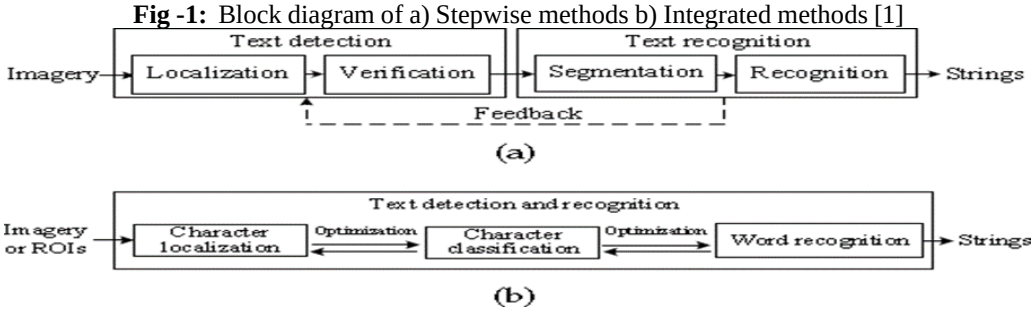
The process of detecting and recognizing text is divided into text detection stage and recognition stage [1]. Text detection deals with finding text area from input image, whereas recognition deals with converting obtained text into characters and words. Methods used for this purpose are categorized as stepwise methods and integrated methods. Stepwise methods have separate stages of detection and recognition and they proceed through detection, classification, segmentation and recognition. Integrated methods have information sharing amongst detection and recognition stages and these methods aim at recognizing words from text available. Fig. 1 shows block diagram of stepwise and integrated methods.

Stepwise methods

Stepwise methods follow stages of text detection and localization, classification, segmentation and recognition and remove background part from text in each stage. Since they are independent of lexicon size, they can be used to recognize text from images independent of volume of text. Elagouni et al. used stepwise approach in [2] for text recognition in videos using convolutional neural networks based classifier. Text detection, tracking, character segmentation, recognition and correction are important processing steps used in this approach. Neural networks based horizontal text detection is performed, followed by statistical intensity based shortest path algorithm for character segmentation. Convolutional neural classification is used for recognition along with language model. Ref. [3] also follows stepwise methods to detect texts of arbitrary orientation from natural images. Component extraction, component analysis, candidate linking and chain

analysis are four stages through which proposed system proceeds. Connected components are extracted in component extraction stage using edge detection followed by SWT. Component analysis stage eliminates non-text candidates and remaining text candidates are paired based on their adjacency in the candidate linking stage. Chains formed at this stage are analyzed by chain level classifier in chain analysis stage.

Fig -1: Block diagram of a) Stepwise methods b) Integrated methods [1]



Integrated methods

Integrated methods focus on detecting particular words from images. Integrated methods often avoid segmentation stage or replace it with word recognition or matching stage. These methods are used for applications with small size of lexicon recognizing fixed set of words. Wang and Belongie used integrated method for word spotting from natural scenes in [4]. It uses character recognition and word configuration stages in this system. It crops region around the text from an image and uses it as input image along with lexicon for word recognition. It uses Histogram of Oriented Gradients (HOG) features with nearest neighbor classifier for character detection. Word recognition represents each word in lexicon in form of chain of connected characters and matches it with output of character recognition stage to obtain nearest word for set of characters. Ref. [5] performs end-to-end text scene localization and recognition by keeping multiple segmentations of single characters until last stage of processing where character contexts in text line are known. This system detects character as external regions, i.e. regions whose outer boundary pixels have higher values than the region. System uses threshold, adjacency and color space projections as three parameters for every detected character and stores their multiple segmentations from which optimal values are selected based on contexts of characters in text line. Any single parameter does not guarantee efficient results, which causes proposal of storing multiple segmentations of three parameters. For sequence selection process, text regions are divided into text lines using exhaustive search method, followed by rejection of low confidence regions in text lines and construction of directed graph by assigning scores to each node and edge from which correct sequence of characters is selected.

III. Important Stages Of Text Detection And Recognition

In this section, four important stages of text detection and recognition are described. Text detection and localization, classification, segmentation and text recognition are described with their role and importance. Various approaches followed for this stages are also explained in this section.

Text detection and localization

Text detection deals with detecting presence of the text in the input image whereas text localization localizes position of the text and forms groups of text regions by eliminating maximum of the background. Text detection and localization process is performed using connected component analysis or region based methods.

Connected component (cc) analysis method forms graph of connected points based on color or edge features from binarized image. In [6], [7] and [8], connected component analysis is used to detect and localize text regions. Ref. [8] uses efficient cc extraction methods instead of using cc filtering approach. Ref. [7] extracts text in form of connected components by applying pixel-based constraints on components. Fig. 2 represents method of extraction of connected components.

Region based methods 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 different textual properties. Ref. [10] uses 64×32 pixels window and applies Modest AdaBoost classifier on 16 different spatial scales of image to classify text and non-text considering large variations in text size. Unsupervised learning approach is used by [11], on small 8×8 grayscale patches of image for feature learning. It further uses 32×32 pixels of image for feature extraction, text detector training and character classifier training.

Classification

After text detection and localization stage output may contain non-text regions along with text regions

as false positives. Classification stage verifies text regions and eliminates non-text regions using classification algorithms. This stage can also be called as verification. Classification algorithms are either supervised or unsupervised. Supervised algorithms knows properties of text such as color, size, texture, etc. before classification. Unsupervised algorithms do not have prior knowledge about text features.

Supervised classification algorithms need training before classification. These algorithms undergo training to be able to extract features of the text to be classified and use these features in classification phase. Ref.[12] and [13] uses supervised learning approach. Constraints on edge area as well as area, height and width constraints on block obtained in detection stage are used in [13] for classification.

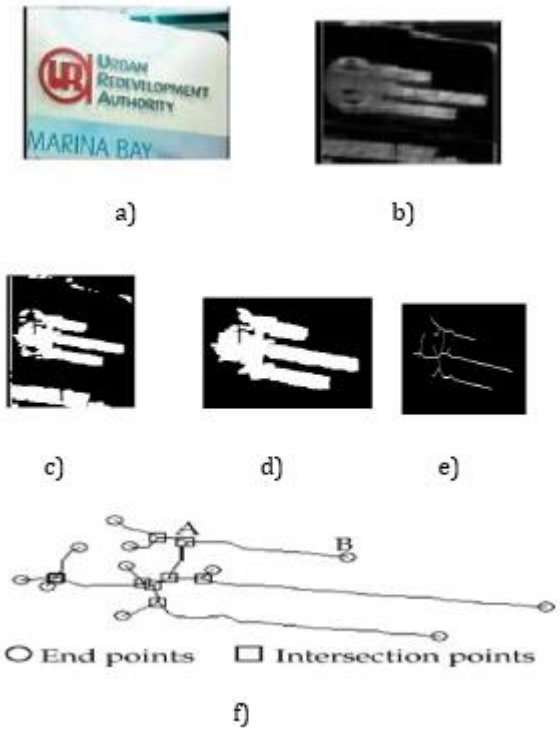


Fig -2: Extraction of connected components.
a) Original image
b) Maximum Difference map of fig.2.a
c) Text cluster
d) Connected component (CC)
e) Skeleton of CC
f) End points and Intersection points of CC [9]

Projection profile of text which is a representation of spatial pixel content distribution is used as constraint for separation of non-text regions from text regions in [12] as shown in fig. 3 .

Unsupervised classification algorithm do not undergo training. They extract features of text during the classification phase only unlike supervised classification and they use features extracted in previous classification for next one. This is similar to adaptive learning. Wavelet transform which gives successive approximation through low-pass filter and details of edges and other features from high-pass filter is used in [14]. It divides image using 16×16 sized window and obtains 36 features from each windowed image for classification of text and non-text by unsupervised method. Ref. [15] uses features like variance of stroke width, difference between contrast of text and background as well as aspect ratio of bounding box are used to form connected components which are classified using k-means based classifier. These are global features extracted from entire image unlike method used in [14] where image is divided in sub-regions.

Segmentation

Segmentation process is used to separate text from background and to extract bounded text from image. Integrated methods which focuses on word matching/recognition often combine or replace complex segmentation stage with recognition stage however stepwise methods undergo segmentation to obtain precisely extracted characters which are fed to recognition stage. Binarization, character segmentation are few of the segmentation algorithms studied in this paper.

Binarization converts color or gray-scale image into black and white image. To achieve good segmentation result irrespective of dark or bright text or background, Kim et al. uses adaptive thresholding for binarization in [16]. Ref. [17] uses k-means clustering algorithm for binarization. It uses k=3 and k=4 as cluster parameters and classify binarized image into texts using probabilistic models as shown in fig. 4.

Character segmentation is process of converting text into multiple sets of single characters. It is suitable in case of degraded text or connected characters. Gradient vector flow based method is used in [18], which is applied directly on grayscale images eliminating need of binarization. It initially identifies candidate cut pixels from the characters and then uses two pass path finding process that finds out potential cuts in forward pass and verify true cuts and remove false cuts in backward pass.

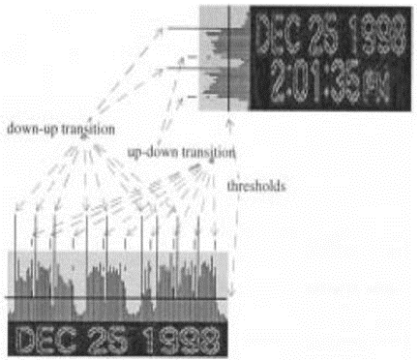


Fig -3: Use of Projection profile for identification of text lines and words [12]

Ref. [19] uses mathematical morphology (MM) operator based adaptive thresholding and extracting approach for segmenting characters from degraded text as shown in fig.5. Along with MM operator, it uses heuristics to identify candidate points for segmentation.

Text Recognition

Text recognition stage converts images of text into string of characters or words. It is important to convert images of text into words as word is an elementary entity used by human for his visual recognition. Different approaches of recognition are character recognition and word recognition.

Character recognition methods divide text image into multiple cut-outs of single characters. Separation between adjacent characters is very important for these methods.

Fig -4: Binarization of text from street view [17]

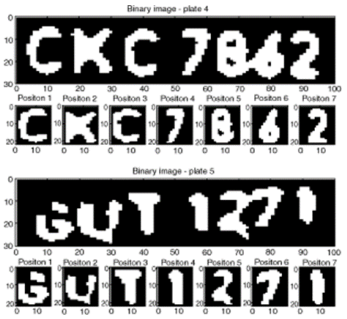


Fig -5: Results of character segmentation of degraded characters [19]

Character recognition approach using Optical Character Recognition module (OCR) is used in [20], where initially images are segmented into k classes followed by binary text image hypothesis generation which passes through connected components analysis and gray scale consistency constraint module before getting fed

to OCR. Support Vector Machine (SVM) based classifier is used for character recognition in [21]. SVM gives good support for multi-class classification, which is tested in [21] on Indic language Kannada which has total 578 characters formed by altering 34 base consonants using 16 vowels. Many of these characters falls in similar classes which has made use of SVM based layered classification approach easier.

Word recognition uses character recognition outputs along with language models or lexicons to recognize words from text image. It can be used in case of degraded characters. For applications with limited number of word possibilities in input images, word recognition is better approach than character recognition. Wachenfeld, et al. used graphical approach for word recognition in [22]. During segmentation stage, when each character is segmented, a hypothesis graph is formed to represent every segmentation and each path of graph is an ordered segmentation sequence which leads to formation of words from characters. Ref. [23] uses similarity between characters as feature along with lexicon, appearance and language properties to accurately recognize characters and words.

Table 1 and Table 2 compares methods used for text detection and recognition and approaches used for its processing stages respectively.

Table -1: Comparison of methods used for Text Detection and Recognition

Method	Features of the method
Stepwise Methods	• Separate Detection and Recognition modules • Divided into localization, classification, segmentation and recognition stages • Suitable for detection of large number of words • Less computation cost; but complexity increases due to more steps
Integrated Methods	• Detection and Recognition modules are not separate • Can avoid segmentation or replace it with word recognition • Suitable for identifying specific words from image i.e. small lexicon • Increase in lexicon size makes recognition difficult

TABLE -2: Comparison of approaches used for processing stages of Text Detection and Recognition

Processing Stage	Approach for processing	Features
Text Detection and Localization	Connected Components analysis	• Graph based method • High speed • Uses color or edge features • Not efficient for noisy images
	Region based methods	• Windowing based approach • Less speed • Use texture features or morphological operations • Efficient for noisy images also
Classification	Supervised Approach	• Supervised classifier has training phase • Classifier knows features of the text before classification starts
		• Features like color, size, projection profile are used
	Unsupervised approach	• Unsupervised classifier do not have training phase • Classifier learns from features extracted in previous classification • Wavelet, stroke width, contrast etc. are used as features
Segmentation	Binarization Binarization	• Converts color or gray-scale image into black and white image • Uses simple or adaptive thesholding • Simple algorithm • Not suitable for connected characters or degraded text
Segmentation	Character Segmentation	• Converts text into multiple sets of single characters • Uses properties of characters for segmentation • Complex algorithm • Suitable for degraded text as well as connected characters
Text Recognition	Character Recognition	• Divide text into cut-outs of single characters • Independent of lexicon • Used when number of words to be recognized are not limited
	Word Recognition	• Identifies word from text image • Recognizes small number of words provided by lexicon • Suitable only for recognizing limited number of words

IV. Applications

Various applications of text detection and recognition from images and videos have been emerged in past few years with advancements in image processing techniques. Developments of various embedded systems and increasing work in the field of computer vision and machine learning gives further rise to the increase in applications of text detection and recognition.

Text detection and recognition is used in industries for reading package labels, numbers etc. It is used to retrieve video captions as well as specific text contents from web pages. It is used for automatic number plate recognition at toll booths as well as for street boards reading purpose in case of unmanned vehicles. Text detection and recognition has very important application in form of assisting blind or visually impaired people for reading, making their daily life easy. It is also used in automatic cheque signature reading. Automatic document scanning is another application of text recognition.

V. Conclusion

In this paper, stages of text detection and recognition and various methods used for that have been presented. This process is further divided into text detection and localization, classification, segmentation and text recognition. These stages are presented in this paper along with comparison of approaches used to undergo the above mentioned stages. Analysis of advantages, disadvantages and applications of different approaches have also been performed over here.

References

[1] Qixiang Ye, David Doermann, "Text Detection And Recognition In Imagery: A Survey" Ieee Transactions On Pattern Analysis And Machine Intelligence, 2014

[2] K. Elagouni, C. Garcia And P. Sbillot," A Comprehensive Neural-Based Approach For Text Recognition In Videos Using Natural Language Processing," In Proc. Acm Conf. Multimedia Retrieval, 2011.

[3] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, And Z. Tu," Detecting Texts Of Arbitrary Orientations In Natural Images," In Proc. Ieee Int'l Conf. Computer Vision And Pattern Recognition, Pp.1083-1090, 2012.

[4] K. Wang, S. Belongie," Word Spotting In The Wild", In Proc. European Conference On Computer Vision, Pp. 591- 604, 2010.

[5] L. Neumann, J. Matas," On Combining Multiple Segmentations In Scene Text Recognition," In Proc. Ieee Int'l Conf. Document Analysis And Recognition, Pp. 523- 527, 2013.

[6] A.K. Jain And B. Yu," Automatic Text Location In Images And Video Frames," Pattern Recognition, Vol. 31, No. 12, Pp. 2055-2076, 1998.

[7] S.M. Hanif, L. Prevost, P.A. Negri," A Cascade Detector For Text Detection In Natural Scene Images," In Proc. Ieee Int'l Conf. Pattern Recognition, Pp. 1-4, 2008.

[8] H. Koo, D.H. Kim," Scene Text Detection Via Connected Component Clustering And Non-Text Filtering," Ieee Trans. Image Processing, Vol. 22, No. 6, Pp. 2296-2305, 2013.

[9] P. Shivakumara, T.Q. Phan And C.L. Tan," A Laplacian Approach To Multi-Oriented Text Detection In Video," Ieee Trans. Pattern Analysis And Machine Intelligence, Vol. 33, No. 2, Pp. 412-419, 2011.

[10] J. Lee, P. Lee, S. Lee, A. Yuille And C. Koch," Adaboost For Text Detection In Natural Scene," In Proc. Ieee Int'l Conf. Document Analysis And Recognition, Pp. 429-434, 2011.

[11] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu, Andrew Y. Ng," Text Detection And Character Recognition In Scene Images With Unsupervised Feature Learning," In Proc. Ieee Int'l Conf. Document Analysis And Recognition, Pp. 440-445, 2011.

[12] R. Lienhart And A. Wernicke, "Localizing And Segmenting Text In Images And Videos," Ieee Trans. Circuits System On Video Technology, Vol. 12, No. 4, Pp. 256-268, 2002.

[13] M. Li, C. Wang," An Adaptive Text Detection Approach In Images And Video Frames," In Proc. Int'l Joint Conf. Neural Network, Pp. 72-77, 2008.

[14] H. Li, D. Doermann, And O. Kia," Automatic Text Detection And Tracking In Digital Video," Ieee Trans. Image Processing, Vol. 9, Pp. 147-156, 2000.

[15] A. Mosleh, N. Bouguila, A. Ben Hamza," Image Text Detection Using A Bandlet-Based Edge Detector And Stroke Width Transform," In Proc. British Machine Vision Conference, Pp. 1-2, 2012.

[16] W. Kim And C. Kim," A New Approach For Overlay Text Detection And Extraction From Complex Video Scene," Ieee Trans. Image Processing, Vol. 18, No. 2, Pp. 401-411, 2009.

[17] J.J. Weinman, Z. Butler, D. Knoll, J. Feild," Toward Integrated Scene Text Reading," Ieee Trans. Pattern Analysis And Machine Intelligence, Vol. 3, No. 2, Pp. 375- 387, 2014.

[18] T. Phan, P. Shivakumara, B. Su And C.L. Tan," A Gradient Vector Flow-Based Method For Video Character Segmentation," In Proc. Ieee Int'l Conf. Document Analysis And Recognition, Pp. 1024-1028, 2011.

[19] S. Nomura, K. Yamanak, O. Katai, H. Kawakami, T. Shiose, "A Novel Adaptive Morphological Approach For Degraded Character Image Segmentation," Pattern Recognition, Vol. 38, No. 11, Pp. 1961-1975, 2005.

[20] D. Chen, J.M. Odobez, H. Bourlard,"Text Detection And Recognition In Images And Video Frames," Pattern Recognition, Vol. 37, No. 3, Pp. 596-608, 2004.

[21] K. Sheshadri, S.K. Divvala," Exemplar Driven Character Recognition In The Wild," In Proc. British Machine Vision Conference, Pp. 1-10, 2012.

[22] S. Wachenfeld, H. Klein And X. Jiang," Recognition Of Screen-Rendered Text," In Proc. Ieee Int'l Conf. Pattern Recognition, Pp. 1086-1089, 2006.1733-1746, 2009.

[23] J.J. Weinman, E. Learned-Miller And A. Hanson," Scene Text Recognition Using Similarity And A Lexicon With Sparse Belief Propagation," Ieee Trans. Pattern Analysis And Machine Intelligence, Vol. 31, No. 10, Pp.