Volume 6, Issue 8, August 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Recent Survey on Various Scene Text Recognition and Detection in Images

Imran Siddiqui*
Research Scholar, Computer Science and Engineering
RKDF IST, Bhopal, India

Dr. Varsha Namdeo
Associate Professor, Computer Science and Engineering
RKDF IST, Bhopal, India

Abstract—Scene text recognition brings many new challenges. research area in the image processing area because of complex background, unidentified outline, scene text, numerous character fonts and sizes and variability in imaging conditions with uneven lighting, shadowing and aliasing, resolution and thus can provide useful information for a wide range of vision tasks of text or image and illumination changes.

Index Terms—Text Detection, Optical Character Recognition (OCR), scene text, Text regions

I. INTRODUCTION

Text detection in natural images has received much attention from the communities of computer vision and document analysis. The problem of finding text in an arbitrary image of a scene can be radically more complex. First, the contents of the input image are generally more varied. From urban structures to more natural subjects like trees, the variety of potential image contents is vast, and it occupies the entire input image. Text regions in scene images need not be well-bounded the way they usually are in documents. Furthermore, text in scenes is often only a few words in one place. There are no large paragraphs or long lines to analyze. Although text is generally designed to be readable, there are often adverse effects of the imaging conditions that can make it difficult to identify. Text detection in document processing is frequently taken care of as a very simple procedure. In general, this involves a search for lines of text in a binarized image. Other methods include dealing out and categorizing connected parts [1]. Distance from the camera can make text small and low resolution, without much detail. Specularities can mix text regions with a reflected image. Perspective distortion can produce text with a varying font size or orientation. Moreover, unlike document processing, there is no global page model whose transform parameters can be estimated. Because text may be printed on bricks, wood or complex backgrounds, the simple binarization and text zoning algorithms of document processing will be insufficient. Small amounts of text can appear anywhere, at any size, with any world orientation, and on any surface. All of these problems tend to make text location in scene images generally more challenging than document text detection and pre-processing.

Most of the existing methods for text detection [2], [3] and recognition [4] can only handle horizontal or near-horizontal texts. This is a rigorous disadvantage as a big part of the texts in real-world circumstances are non-horizontal. Such drawback would make it fail to capture the information embodied in non-horizontal texts and thus critically confines the achievability of these methods.

Figure-1: Images for scene text reading

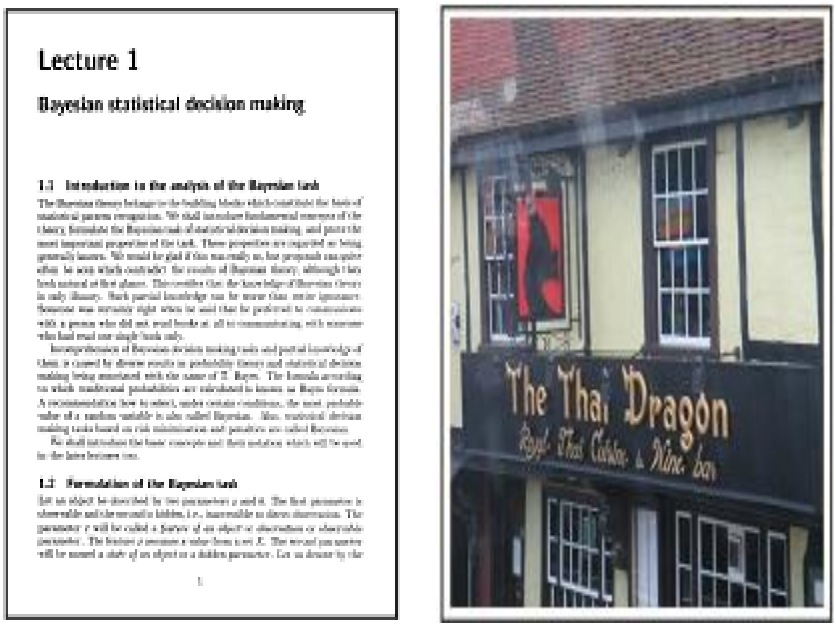
The approaches of Epshtein et al. [2] are modern efforts in the direction of end-to-end text recognition systems. These approaches chiefly address the problem of text detection and achieve subsequent recognition using off-the-shelf OCR engines. Epshtein et al. [2] proposed a new image operator, called Stroke Width Transform (SWT) to confine texts in natural scene images. For recognition, off-the-shelf OCR engine was accepted to identify the masks created by SWT. These techniques have accomplished brilliant concert on natural images and encouraged various examiners in this area [5-7]. On the other hand, these techniques pleasure text detection and recognition as separated phases and only focus on horizontal texts. The recognition correctness of these techniques is fairly limited as off-the-shelf OCR engines are particularly considered for texts in document images not natural images.

II. THEORETICAL BACKGROUND

Text detection and recognition in natural images have obtained rising attention in computer vision and image understanding, due to its various possible applications in image retrieval, scene accepting, visual assistance, etc. Though tremendous attempts have in recent times been dedicated to improving its presentation reading texts in unconstrained environments is still enormously demanding and continues an open problem. Since scene text images usually contain a large amount of irrelevant elements as well to valuable text contents, text detection is a serious process to localize texts and remove inappropriate parts. Then text recognition is need to understand the signs in the localized text areas and to convert them into computer readable and editable characters, thus making accessible the high level semantics embodied in texts and thus can provide useful information for a wide range of vision tasks, such as image understanding, image search [8]. This property has made text detection and recognition in natural images active research topics in computer vision [9-11]. It is also important to contrast the nature of recognizing text in documents and scene images. As mentioned above, a region of text in a scene often involves only one to five words. This makes several techniques that might be used by document recognition systems less applicable. For instance, powerful language models that rely on complete sentences or longer phrases may not be useful. Also, techniques for identifying the font face of a text sample to aid recognition may not be robust with only a handful of characters. In addition to the small sample problem, the sheer variety of fonts that may be present in signs and scenes is often much greater than that found in most documents. Text recognition algorithms for scene images must handle a great number of fonts, many of which will be entirely novel for the system. Perspective distortions and complex backgrounds, as described above, will also require recognition algorithms to be robust to character "deformations" and non-binary input. In low resolution situations, it will not be reasonable to binarize text regions before recognition. This is in contrast to most document recognition systems, which often rely on binary images to at least perform word segmentation prior to recognition, if not a heuristic over-segmentation of binarized characters.

III. TEXT FEATURES RECOGNITION TASKS

At the present time, there is ever-increasing require for the software systems to identify characters in computer system when data is scanned during paper documents as we know that they have amount of newspapers and books which are in printed format associated to different subjects. These days there is an enormous order in "storing the data available in these paper documents in to a computer storage disk and then afterward recycling this data by searching process". One uncomplicated approach is to store information in these paper documents in to computer system is to initial scan the documents. Whenever they scan the documents all the way through the scanner the documents are accumulated as images format in the computer system. Various techniques for scene text localization and recognition aspire to discover all regions in an image or a video that would be measured as text by human mark boundaries of the regions frequently by rectangular bounding boxes and output a sequence of Unicode characters connected with its substance. They permit for real-world images and video processing i.e. dealing out of images/videos taken by a typical camera or a mobile phone and "reading" content of each distinguished area into a digital text layout that can be additional procedure by a computer. Scene text localization and recognition that is also known as text localization and recognition in real-world images, nature scene OCR is an open problem distinct printed document recognition where modern schemes are capable to recognize properly more than 99% of characters [12]. Factors causal to the complexity of the problem include: non-uniform background, the need for reimbursement of perspective consequences i.e. for documents, rotation or/and scaling is adequate; real-world texts are often short snippets written in different texts and languages; text alignment does not follow strict rules of printed texts; many statements are appropriate names which avoids an effective use of a dictionary.



(a) A scanned book page. (b) Robust Reading dataset

Figure 1: The difference between printed document recognition (OCR) and scene text localization and recognition.

The recognition of materials, objects, and scene categories from photographs are among the most central problems in the area of computer vision. on the other hand, even though decades of exhaustive research even the most sophisticated recognition systems today [13] for a summary of the state of the art remain incapable of handling more than just a few simple classes, or of functioning under unconstrained real-world conditions. What makes recognition so difficult is the seemingly limitless variability of natural imagery that arises from viewpoint and lighting changes, progress and twist of non-rigid or articulated objects, intra-class appearance variations, and the presence of occlusion and background clutter. Indeed, using local features as image primitives has several important advantages:

- ***Robustness:*** Because local features are relatively small and compact, they can be preserved even when large portions of an image are affected by clutter or occlusion.
- ***Repeatability:*** Many existing local feature detectors can reliably identify corresponding features in different images despite geometric transformations, changes in lighting, or minor appearance variations. These may be features corresponding to the same surface patch in two views of the same object, or features corresponding to analogous structures, such as eyes, on different instances of the same class.
- ***Expressiveness:*** Unlike the geometric features historically used in computer vision (points or line segments), today's local features contain information not only about their shape (circular, elliptical, or rectangular), but also about their appearance. Rich high-dimensional descriptors of appearance provide strong consistency constraints for matching tasks.
- ***Local geometric invariance:*** Depending on the requirements of a particular application, one may choose to use scale-, rotation- or affine-invariant local characteristics. Particularly, affine invariance presents strength to a wide range of geometric transformations that can be locally approximated by a linear model, including perspective distortions and non-rigid deformations.
- ***Compactness or Sparsity:*** The number of features returned by most detectors is typically orders of magnitude smaller than the total number of pixels in the original image. The resulting patch-based description is extremely compact, thus reducing processing time and storage requirements.

Texture Recognition: We want to recognize images of textured surfaces subject to viewpoint changes and non- rigid deformations. For this task, we use a local model in which spatial constraints are absent, i.e., each patch is considered separately, without any information about its neighborhood or its position in the image. The distribution is learned by quantizing the descriptors in the image and forming a signature, or a set of all cluster centers together with weights indicating the relative sizes of the clusters. Signatures of different images are compared using Earth Mover's Distance (EMD), which solves a partial matching problem between sets of possibly unequal cardinality, and is robust to noise, clutter, and outliers. We also investigate an alternative bag-of-features approach, in which local features are quantized into "textons" or "visual words" drawn from some universal vocabulary, and their distributions in images are signified as histograms of texton labels. This approach is analogous to the bag-of-words paradigm for text document analysis [14]. A major shortcoming of bag-of-features methods is their disregard of spatial relations.

For many natural textures, the spatial layout of local features captures perceptually important information about the class. Augmenting our texture representation with geometric information can be expected to increase its ability to discriminate between textures that have similar local elements but different geometric patterns.

Object Recognition: Our second target problem is recognizing object categories despite 3D viewpoint transforms, intra-class emergence discrepancies, non-rigid movements as well as clutter and occlusion in the test images. At the most basic level, object recognition may be considered as a whole-image classification problem, i.e., identifying which object class is present in an image, without attempting to segment or localize that object. For this admittedly simplified task, it is possible to represent the "visual texture" of images enclosing objects using the order less bag-of-features model described. Such models have been used in several recent approaches to visual categorization [15], unsupervised discovery of visual topics and video retrieval.

Even though its realistic benefits a bag of characteristics is an enormously weakened representation for object classes, since it ignores all geometric information about the object class, fails to distinguish between foreground and background features, and cannot segment an object from its surroundings. An object model consists of a collection of multiple semi-local parts, and these parts, in turn, can be connected to each other by looser geometric relations. Semi-local parts are detected using an alignment-like geometric search, which can in principle work with any rigid 2D transformation model, from translation to projective.

Scene Recognition: Our third problem is recognizing semantic scene group for example beach, mountain, school, office, etc. At the same time as texture and object recognition, this task can be approached using purely local models [16]. However, in this work that improved performance can be achieved by a global model that takes into account the absolute positions of the features. Intuitively, a global model has improved discriminative power because it captures spatial regularities that are important to our perception of natural scenes (for example, sky is usually above ground or water, building walls are usually vertical, the horizon is usually a horizontal line, etc.

IV. LITERATURE SURVEY

Wang et al. [17] introduced an end-to-end scene text recognition system, which aimed to tackle a special case of the scene text understanding problem where in addition to the natural image it is also given a list of words (i.e., a lexicon) to be detected and read. However, the usability of this algorithm in general text understanding scenarios is limited since a lexicon with probable words for each individual image is not always available.

Jing Zhang and Rangachar Kasturi in [18], we propose a new unsupervised text detection approach which is based on Histogram of Oriented Gradient and Graph Spectrum. By exploring the properties of text edges the suggested method initially removes text edges from an image and confines applicant character blocks using Histogram of Oriented Gradients then Graph Spectrum is exploited to confine comprehensive association among applicant blocks and cluster applicant blocks into groups to produce bounding boxes of text objects in the image. The suggested technique is robust to the color and size of text.

S. No.	Paper	Author	Advantages	Issues
1	End-to-end scene text recognition	K. Wang, B. Babenko, and S. Belongie. [20]	This approach enables us to train highly accurate text detection and character recognition modules.	Here they aimed to tackle a special case of the scene text understanding problem where in addition to the natural image
2	Multi-orientation scene text detection with adaptive clustering	X. C. Yin, W. Y. Pei, J. Zhang, and H. W. Hao [21]	This can be evaluated on several public scene text databases to construct and release a practical challenging multi-orientation scene text data set	Content-based image analysis tasks, mainly efforts only focus on horizontal or near horizontal scene text.
3.	Robust scene text detection with convolution neural network induced MSER trees.	W. Huang, Y. Qiao, and X. Tang [22]	Using this method by enhancing intensity contrast between text patterns and background to detecting text patterns and resulting in a higher recall	High-level feature globally computed true text features in the deep representation. This show the ways to fewer discriminative power and not as good as strength
4.	A Performance Evaluation Methodology for Historical Document Image Binarization.	Ntirogiannis, Konstantinos, Basilis Gatos, and Ioannis Pratikakis. [23]	This method has giving the high recall and precision evaluation measures are properly modified using a weighting scheme	Pixel-based binarization evaluation methodology for historical handwritten/machine-printed document images
5.	Text line extraction from handwritten document pages using spiral run length smearing Algorithm.	Malakar, Samir [24]	Extraction of text lines from document images using this technique extracts 87.09% and 89.35% text lines effectively from the said databases respectively	Handwritten document images, presence of twisted touching or not be separating text line(s) makes this procedure a real challenge.
6.	Retrieval of Rashi Semi-Cursive Handwriting via Fuzzy Logic	Gur, Eran, and ZeevZelavsky [25]	This approach combines letter statistics and correlation coefficients in a set of fuzzy based regulations to enabling the recognition of distorted letters that may not be retrieved otherwise	Text recognition and retrieval of using OCR tools do not supply a complete solution and in most cases human inspection is required

Nidhi Sharma and Mohit Khandelwal in [19] also presented a technique for improving the recognition accuracy of Hindi OCR method by increasing the idea for detection of bold, italic word of unusual fonts. Detection of method words is not only helpful for development of OCR presentation but also helpful in automatic Indexing, because it is noted that important terms are often printed in style. Most of the Indian scripts are composed in two dimensions that make them different from Roman script. Another approach used in which stroke pattern analysis operating on wavelet decomposed word images to distinguish the existence of italic style. It initially removes the normalized total height H of the vertical straight line segments (VSLS) and the normalized total length L of the long continuous diagonal strokes (CDS). Here author has defined some thresholds by experimental result to make a decision the range of H and L of the italic and normal words. This technique takes benefit of 2-D wavelet decomposition on each word image and executes statistical analysis on stroke patterns.

In this paper author has proposed a new method in [26], text in a digital image contains significant data to the scene considerate and can be valuable for many applications. Detecting and extracting such text is a complicated job. The most important difficulty in removing text from natural images is reasoned by numerous causes including font size dissimilarity, arrangement of text and variation of font colors. In this paper, here author propose a connected component based technique to repeatedly detect the text region from natural images. Since text regions in images contain frequently replication of vertical strokes. Once the group of edges is originating, neighboring vertical edges are associated to each other. Connected regions with geometric aspects out of suitable conditions are measured as outliers and removed.

Here author has described in [27] a simple and fast algorithm for detection of italic and bold characters in dissimilar fonts in Devnagari script without recognition of the genuine character. Only a small amount of works has been done for printed devanagari text in the region of optical character recognition. Here they present automatic information which tells

us about the font category phase in the method of weight and slope. The procedure of recognition and classification of italic and bold character can be utilized for construction correctness of the text recognition system in the OCR. This simple and fast algorithm gives high correctness and very simple to put into operation.

X. F. Wang et. al [28] proposed a new system for embedded text segmentation. This system is based on two statements of embedded texts: i) the color of text pixels go behinds Gaussian distribution; ii) the local part of the embedded text has the similar color distribution with the comprehensive part. By these two statements, here they expanded a two-step text segmentation approach: in the initial step i.e. coarse segmentation step and here they used a 1-D Gaussian function to generate a model for the color distribution of text pixels. Then the self-assured text region was removed using a stroke operator to get hold of the model constraints and the considerations are approximated from a expanded heuristic process

In this paper author [29] has presented one method that was used for detection and removal of text from images. The system become aware of text using morphological process, related component labeling and a set of selection criteria which assists to filter out non text regions. So, the consequential image is the image with only texts. Text Inpainting is done in two steps. The initial step identifies the text region repeatedly, without user interaction and in the subsequent step; the text is eliminated from the image using standard based Inpainting algorithm.

V. CONCLUSION

In this paper, we have reviewed and analyzed different methods to find text recognition and detecting characters from images of text in natural scenes. Thus, while much research has focused on developing the models and features used in scene-text applications using a more automated and scalable solution.

REFERENCES

[1] Bargeron, David, Viola, Paul, and Simard, Patrice. Boosting-based transductive learning for text detection. In Proc. Intl. Conf. on Document Analysis and Recognition (2005), pp. 1166–1171.

[2] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. of CVPR, 2010.

[3] A. Ikica and P. Peer, "An improved edge profile based method for text detection in images of natural scenes," in Proc. of EUROCON, 2011.

[4] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky, "Large-lexicon attribute-consistent text recognition in natural images," in Proc. Of ECCV, 2012.

[5] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. Image Processing, vol. 20, no. 3, pp. 800–813, 2011.

[6] 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 Proc. of ICIP, 2011.

[7] A. Mosleh, N. Bouguila, and A. B. Hamza, "Image text detection using a bandlet-based edge detector and stroke width transform," in Proc. Of BMVC, 2012.

[8] S. Tsai, H. Chen, D. Chen, G. Schroth, R. Grzeszczuk, and B. Girod, "Mobile visual search on printed documents using text and low bit-rate features," in Proc. of ICIP, 2011.

[9] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in Proc. of CVPR, 2012.

[10] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky, "Large-lexicon attribute-consistent text recognition in natural images," in Proc. Of ECCV, 2012.

[11] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multi-scale representation for scene text recognition," in Proc. of CVPR, 2014.

[12] X. Lin. Reliable OCR solution for digital content re-mastering. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Dec. 2001

[13] C. Schmid, J. Ponce, M. Hebert, and A. Zisserman, editors. Towards Category-Level Object Recognition. Springer Lecture Notes in Computer Science, 2006.

[14] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, 2003.

[15] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. "Visual categorization with bags of keypoints" In ECCV Workshop on Statistical Learning in Computer Vision, 2004

[16] Lu, K. Toyama, and G. Hager "A two-level approach for scene recognition". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 688–695, 2005.

[17] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in Proc. of ICCV, 2011.

[18] Jing Zhang and Rangachar Kasturi, "Text Detection Using Edge Gradient and Graph Spectrum", International Conference on Pattern Recognition IEEE,pp 3979-3982, 2010.

[19] N. Sharma, M. Khandelwal, "Detection of Bold Italic and Underline Texts for Hindi TR", In: International Journal of Computer Trends and Technology (IJCTT), vol. 4, Issue 8, pp. 2425-2428, 2013.

[20] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in Proc. of ICCV, 2011.

[21] X. C. Yin, W. Y. Pei, J. Zhang, and H. W. Hao, "Multi-orientation scene text detection with adaptive clustering," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1930–1937, Sep. 2015.

[22] W. Huang, Y. Qiao, and X. Tang, "Robust scene text detection with convolution neural network induced MSER trees," in Proc. 13th Eur. Conf. Comput. Vis. (ECCV), 2014,

[23] Ntirogiannis, Konstantinos, Basilis Gatos, and Ioannis Pratikakis. "A Performance Evaluation Methodology for Historical Document Image Binarization." IEEE International Conference on Document Analysis and Recognition, 2013.

[24] Malakar, Samir, et al. "Text line extraction from handwritten document pages using spiral run length smearing algorithm." IEEE International Conference on Communications, Devices and Intelligent Systems (CODIS), 2012.

[25] Gur, Eran, and ZeevZelavsky, "Retrieval of Rashi Semi-Cursive Handwriting via Fuzzy Logic", Frontiers in Handwriting Recognition (ICFHR), International Conference on. IEEE, 2012

[26] Manoj Kumar , Gueesang Lee, "Automatic Text Location from Complex Natural Scene Color images" The 2nd International Conference on Computer and Automation Engineering (ICCAE) IEEE vol. 3, pp. 594-597, 2010.

[27] R. K. Yadav, B. D. Mazumdar, "Detection of Bold and Italic Character in Devanagari Script", In: International Journal of Computer Applications, vol. 39, no. 2, pp. 19-22, 2012.

[28] Jian Zhang, Renhong Cheng, Kai Wang, Hong Zhao, "Research on the text detection and extration from complex images", Fourth International Conference on Emerging Intelligent Data and Web Technologies. Vol. 10, 2013, Page no. 708-713.

[29] Khyati Vaghela, Narendra Patel," Automatic Text Detection Using Morphological Operations and Inpainting", International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 5, May 2013.

