iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: III Month of publication: March 2018

DOI: http://doi.org/10.22214/ijraset.2018.3260

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

A Survey on Scene Text Detection and Text Recognition

P. Balaji[1], P. Chandra Sekhar Reddy[2]
[1]*CSE, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India*
[2]*CSE, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India*

***Abstract:** Late deep learning models have shown solid capacities for arranging text and non-text segments in common images. They extract an abnormal state highlight registered all inclusive from an entire image segment (fix), where the jumbled foundation data may command genuine text highlights in the deep representation. This prompts less discriminative power and poorer vigor. Introduce another framework for scene text recognition by proposing a novel Text Attentional Convolutional Neural Network (Text CNN) that especially centers on removing text related areas and highlights from the image parts. We build up another learning component to prepare the Text CNN with multi-level and rich regulated data, including text district cover, character mark, and paired text/non text data. The rich supervision data empowers the Text CNN with a solid ability for discriminating ambiguous texts, extracting text-related regions and features from the image components. The preparation procedure is planned as a multi-undertaking learning issue, where low-level directed data significantly encourages principle errand of text/non-text order. What's more, an effective low-level locator called Contrast- Enhancement Maximally Stable Extremal Regions (CE-MSERs) is produced, which expands the generally utilized MSERs by upgrading power differentiate between text examples and foundation. This enables it to identify deeply difficult text examples, bringing about a higher review. Our approach accomplished promising outcomes on the ICDAR 2013 dataset, with a F-measure of 0.82, enhancing the best in class comes about significantly.*

***Keywords**: Maximally Stable Extremal Regions, text detector, convolutional neural networks, multi-level supervised information, multi-task learning.*

I. INTRODUCTION

As a result of human reflection and control, text in common scenes natural scenes abnormal state semantics. This property makes text present in characteristic images and videos an extraordinary, critical source of data. The rich and exact data embodied in text can be extremely helpful to an assortment of vision-based applications, for example, image look [1], target geo-area [2], human-PC connection [3], robot route [4] and modern computerization [5]. Therefore, programmed text discovery and recognition, offering a way to get to and use printed data in images and videos, have turned out to be dynamic research points in PC vision and archive examination.

In any case, confining and perusing texts in characteristic scenes are greatly troublesome undertakings. The real difficulties in scene text discovery and recognition can be generally classified into three sorts ([6], [7]):

Diversity of scene text: as opposed to characters in report images, which are generally with normal textual style, single color, predictable size and uniform course of action, texts in regular scenes may bear altogether unique textual fonts, colors, scales and orientations, even in a similar scene.

Complexity of foundation: The foundations in normal scene images and videos can be extremely complex. Components like signs, wall, blocks and grasses are for all intents and purposes undistinguishable from genuine text, and hence are effortlessly to cause disarrays and mistakes.

Interference factors: Different obstruction factors, for example, instance, noise, blur, distortion, low resolution, non-uniform illumination and partial occlusion, may offer ascent to disappointments in scene text location and recognition.

To handle these difficulties, a rich assortment of methodologies has been proposed and considerable advances have been accomplished as of late ([8]–[20]). In every one of these strategies, the investigation on portrayal is the fundamental research subject, since portrayal is the way to the adequacy and power of these algorithms. In scene text identification and recognition, portrayal includes the way and way of depicting and demonstrating text and foundation in normal scenes.

In this paper, we show a complete text re-perspective of deals with scene text discovery and recognition in the previous couple of years, primarily from the point of view of representation. This review is devoted to: (1) introduce up-to-date works and summarize

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

recent advances, (2) compare different methods and highlight state-of-the-art algorithms, and (3) analyse development tendency and predict future research directions. Also, it gives connects to valuable assets, including benchmark datasets, source codes, and online demos.

There are as of now a few amazing survey papers ([21]–[23]) in the fields of scene text identification and recognition. Be that as it may, these audit papers are to some degree outdated, since they were distributed around 10 years back and missed various imperative, powerful works that are proposed as of late. The main two close term reviews we know about are crafted by Zhang et al. [24] and Uchida et al. [25]. The study of Zhang et al. [24] has principally centered around papers identified with scene text location, however disregarded techniques on text recognition. Crafted by Uchida et al. [25] audited strategies for text identification and recognition in images and additionally videos, yet it was inclined toward works from the report investigation com-munity and ignored a few breakthrough works from the PC vision group, which have presented new bits of knowledge and thoughts ([7], [19], [20], [26]). Not quite the same as the past audit papers ([21]–[25]), this article gives a far reaching review on scene discovery and recognition in static images, with a unique emphasis on the most recent advances in these territories.

II. RECENT ADVANCES IN SCENE TEXT DETECTION AND RECOGNITION

Lately, text location and recognition in regular images have turned out to be dynamic research points in the groups of PC vision, design recognition and even report examination. Specialists from these groups have proposed a lot of original thoughts and methodologies for the extraction of literary data from in characteristic images and videos.

These strategies can be comprehensively partitioned into three classes: (1) text discovery, (2) text recognition, (3) end-to-end text recognition, as illustrated. The principal classification of techniques ([9], [12], [27]–[30]) concern how to find and find the areas conceivably containing text from normal images, however don't have to perform recognition. The second class of techniques ([7], [14], [15], [31]–[33]) assumes that texts have been identified, and just spotlight on the way toward changing over the distinguished text locales into PC readable and editable symbols. The third class of techniques ([10], [11], [13], [17], [19], [34]) go for building end-to-end text recognition frameworks that achieve both the discovery and recognition tasks.

III. RELATED WORKS ON SCENE TEXT DETECTION

In the previous two decades, researchers have proposed various techniques for recognizing texts in characteristic images or videos. There are principally three kinds of strategies: urface based techniques, part based strategies and half and half techniques.

Surface based strategy ([8], [35]–[36]) regard texts as an uncommon kind of surface and makes utilization of their textural properties, for example, nearby forces, channel reactions and wavelet coefficients, to recognize text and non-text territories in the images. These strategies are typically computationally costly as all areas and scales ought to be examined. Furthermore, these techniques generally handle level texts and are sensitive to turn and scale change. In an early work, Zhong et al. [35] proposed a technique for text limitation in color images. Even spatial difference was used to generally restrict texts and afterward color division was performed inside the limited locales to discover texts. Afterward, Li et al. acquainted a text location framework with distinguish and track texts in videos. In this framework, images are decomposed by utilizing the mean of wavelet coefficients, and the main request and second-arrange minutes as neighborhood highlights.

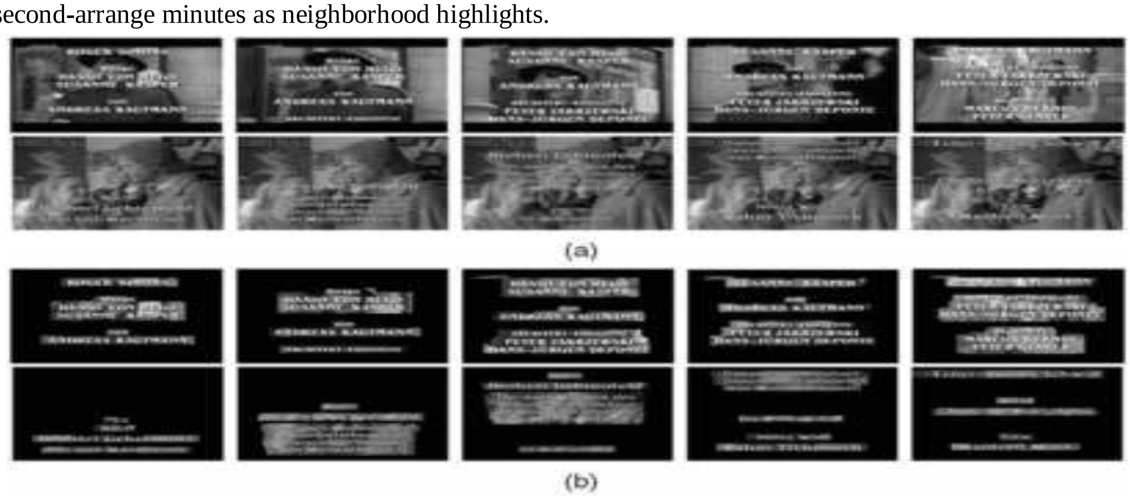


Fig.1: Text detection examples of the algorithm of Kim *et al* [36]. This algorithm is a representative work of early stage methods for text detection. It is only applicable to relatively simple scenes.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Kim et al. [36] prepared a SVM classifier to arrange every pixel by specifically utilizing the crude pixel power as nearby component. Text zones were looked for through Mean Shift in likelihood maps. The technique produces probability location brings about images or videos (Fig. 1) with basic foundations, however it is hard to sum up this technique to complex characteristic scene images or videos. To deal with multilingual texts (principally Chinese and English) in videos, Lyu et al. proposed a coarse-to-fine multi-scale look conspire. The plan utilized properties, for example, solid edge and high difference of texts to recognize text and non-text areas. Also, this algorithm gives a neighborhood versatile twofold methodology to portion identified text zones. Like numerous different methodologies, this technique includes various principles and parameters, so it is hard for it to manage videos of various characteristics and texts of various kinds. Not quite the same as ordinary strategies, Zhong et al. proposed an interesting algorithm that can specifically identify text in the Discrete Cosine Transform (DCT) space. The advantage of this algorithm lies in its high proficiency, as it isn't important to unravel the image before discovery. Nonetheless, the discovery exactness of this technique is constrained. With a specific end goal to accelerate the text identification technique, Chen et al. [8] proposed a quick text indicator. The identifier is a course Adaboost classifier, in which each frail classifier is prepared from an arrangement of highlights. The component pool incorporates mean quality, power change, level contrast, vertical distinction, and angle histogram. The location proficiency of this strategy is fundamentally higher than different algorithms, however the identification exactness on true images is constrained. As of late, Wang et al. proposed a strategy for finding particular words from regular scenes. , single characters are identified by sliding window. At that point, conceivable blends are scored by the auxiliary connections between characters. At long last, the most comparable mixes are chosen from the given rundown as the yield comes about. Not at all like traditional text detection methods, this algorithm can just identify words in the given rundown, incapable for taking care of words out of the given rundown. Truly, in any case, a word list that contains every single conceivable word isn't generally accessible for each image. This influences the relevance to scope of the technique limit, contrasted with other text recognition strategies. Part based strategies ([9], [12], [28], [29]) first concentrate applicant segments through an assortment of ways (e.g., color clustering or extreme region extraction), and after that sift through non-text segments utilizing physically outlined standards or naturally prepared classifiers. As a rule, these techniques are substantially more effective, in light of the fact that the quantity of segments to be prepared is moderately little. What's more, these techniques are uncaring to turn, scale change and text style variety. Lately, segment based strategies have turned into the standard in the field of scene text location. The technique proposed by Jain et al. decomposed images into a few non-covering parts by color clustering, assembled segments into text lines through segment investigation, and after that evacuated non-text segments as indicated by geometric guidelines. On account of the misleadingly characterized Principles and parameters, this technique performs poorly on complex characteristic images.

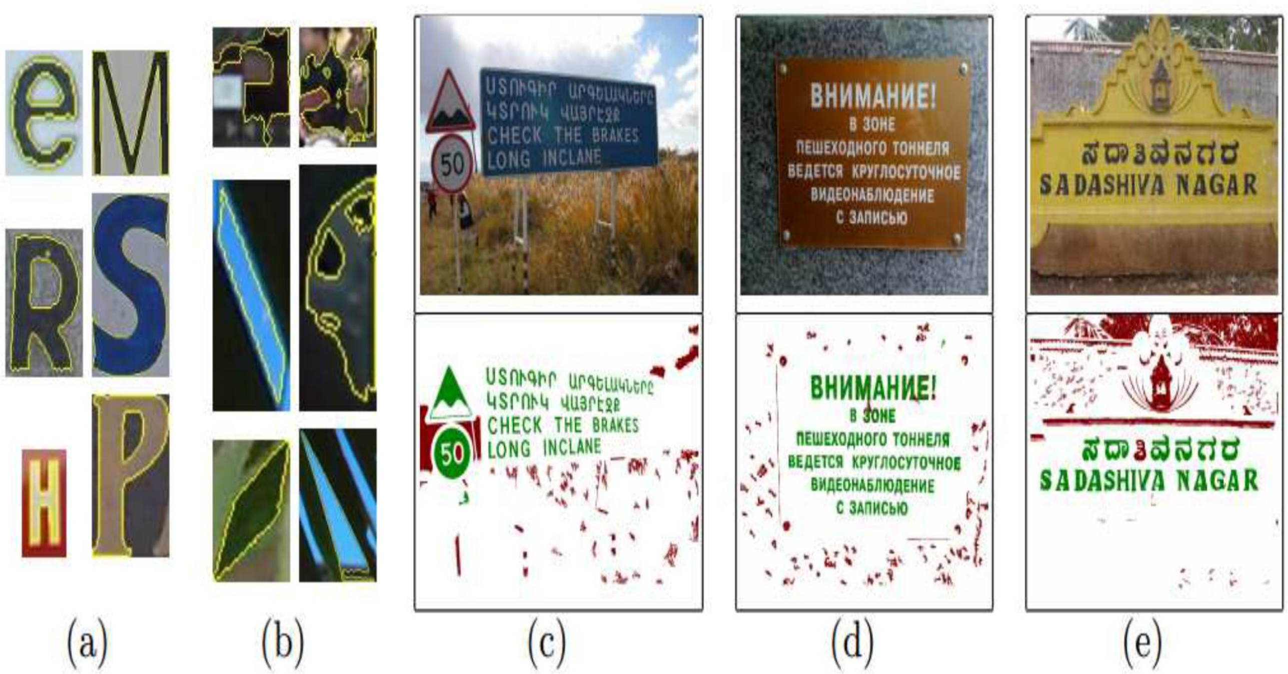


Fig. 2: Text detection examples of the algorithm of Neumann *et al.* [10]. This work is the first that introduces MSER into the field of scene text detection.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Fig. 3: Text detection examples of the algorithm of Yao et al. [12]. Different from previous methods, which have focused on horizontal or near-horizontal texts, this algorithm is able to detect texts of varying orientations in natural images.

Making utilization of the property that characters have almost consistent stroke width, Epshtein et al. [9] proposed another image administrator: Stroke Width Transform (SWT). This administrator gives a simple method to recuperate character strokes from edge maps and can effectively separate text parts of various scales and headings from complex scenes. This technique additionally accompanies a progression of human-characterized standards and parameters, and just thinks about level texts. Neumann et al. [10] proposed a text discovery algorithm in view of Maximally Stable Extremal Regions (MSER). This algorithm removes from the first images MSER locales as competitors, and takes out invalid hopefuls utilizing a prepared classifier (Fig. 2). At a later stage, the remained hopefuls are assembled into text lines through a progression of association rules. In any case, such association tenets can just adjust to even or almost flat texts, hence this algorithm can't deal with texts with bigger slant edge. SWT [9] and MSER [10] are two delegate techniques in the field of scene text location, which constitute the premise of a considerable measure of ensuing works ([12]–[14], [29], [30], [34]). The immense achievement of sparse representation in confronts recognition and image denoising has inspired various researchers. For instance, Zhao et al. built a scanty lexicon from preparing tests and utilized it to judge whether a specific territory in the image contains text. Nonetheless, the speculation capacity of the educated meager word reference is limited, with the goal that this technique can't deal with issues like rotation and scale change. Not the same as the previously mentioned algorithms, the approach proposed by Yi et al. [28] can identify tilted texts in characteristic images. Firstly, the image is partitioned into various locales as indicated by the circulation of pixels in color space, and afterward areas are joined into associated segments as per the properties, for example, color likeness, spatial separation and relative size of districts. At last, non-text segments are disposed of by an arrangement of guidelines. Be that as it may, the pre-essential of this technique is that it expect the info images comprise of a few principle hues, which isn't really valid for complex regular images. What's more, this technique depends on a great deal of artificially designed sifting and parameters, with the goal that it is hard to sum up to substantial scale complex image informational collections. Shivakumara et al. likewise proposed a technique for multi-situated text recognition. The strategy removed applicant areas by bunching in the Fourier-Laplace space and separated the locales into unmistakable parts utilizing skeletonization. Hence, these parts for the most part don't relate to strokes or characters, however just text pieces. This technique can't contrast and different strategies quantitatively, since it can't recognize characters or words specifically. In view of SWT [9], Yao et al. [12] proposed a algorithm that can identify texts of subjective introductions in characteristic images (Fig. 3). This algorithm is outfitted with a two-level arrangement plan and two arrangements of rotation and revolution invariant highlights uniquely intended for catching the inherent qualities of characters in regular scenes. Huang et al. [29] introduced another administrator in light of Stroke Width Transform, called Stroke Feature Transform (SFT). In order to solve the mismatch problem of edge points in the original Stroke Width Transform, SFT presents color consistency and compels relations of neighborhood edge focuses, delivering better part extraction comes about. The location execution of SFT on standard datasets is fundamentally higher than different strategies, however just for level texts. In [30], Huang et al. proposed a novel system for scene text identification, which incorporated Maximally Stable Extremal Regions (MSER) and Convolutional Neural Networks (CNN). The MSER administrator works in the front-end to extricate text competitors, while a CNN based classifier is connected to accurately distinguish genuine text applicants and separate the associations of various characters in parts. This algorithm achieves altogether improved execution over ordinary strategies.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Half and half techniques: [27] is a mix of surface based strategies and part based strategies, which make utilization of the upsides of these two sorts strategies. In the strategy proposed by Liu et al. edge pixels of all conceivable text areas were removed utilizing an intricate edge location system, and the inclination and geometrical properties of re-gion shapes are confirmed to create competitor text districts, trailed by a surface examination technique to recognize genuine text locales from non-text locales. Not at all like the cross breed technique proposed by Skillet et al. extricated competitor parts from multi-scale likelihood maps. The likelihood maps are evaluated by a classifier, which is prepared on an arrangement of surface highlights (Hoard highlights) processed with a gathering of predefined designs. A Conditional Random Filed (CRF) demonstrates consolidating unary part properties and twofold relevant connections, is used to segregate text segments from non-text segments. Like most different algorithms, these two techniques can just detect horizontal texts.

IV. RELATED WORKS ON SCENE TEXT RECOGNITION

Since the properties of characteristic images are extraordinarily not quite the same as report images, there would be numerous obstructions if applying conventional character recognition techniques to normal images. For instance, these techniques may create a lot of false alarms and gibberish, hen running on characteristic images. as to handle these issues, Sawaki et al. genius represented a technique which can consequently make character layouts as indicated by the qualities of common images. Zhou et al. utilized surface fitting classifier and particularly outlined character recognition algorithm to recognize characters in Web images (counting straightforward engineered images and common images). Hence, these algorithms were not assessed on complex common images, so the flexibility of these techniques have not been sufficiently approved. In de Campos et al. tried, looked at and dissected the current normally utilized element descriptors and order algorithms in PC vision and example recognition. Likewise they discharged a image dataset, called Chars74K, for assessing character recognition algorithm. Chars74K has been generally acknowledged and utilized as a part of the field of character recognition in common images. Be that as it may, not at all like the standard character recognition strategies, which regard word as the essential unit, the strategy for de Campos et al. just, think about the issue of individual character recognition.

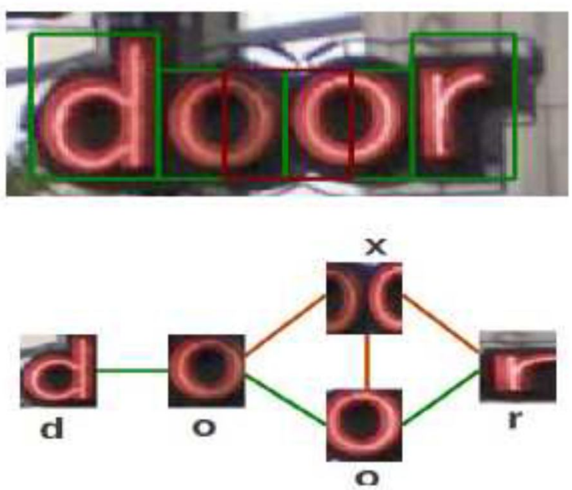


Fig. 4: Bottom-up and top-down cues for scene text recognition [31]. In this method, cues from low level (character detection) and high level (language prior) are integrated into a unified framework.

Mishra et al. [31] utilized base up and top-down signs for scene text recognition, which works in a blunder rectification way. Because of the nearness of complex foundations in characteristic scenes, it is exceptionally hard to directly section characters from neighborhood foundation. So this technique utilizes sliding window to distinguish conceivable characters, and regard the discovery comes about as the base up data. The best down data originates from the insights of an expansive word reference. The base up and top-down data are coordinated in a bound together model through Conditional Random Field (CRF). One of the upsides of this strategy is that is can find mistakes in character recognition. As appeared in Fig. 4, the district between two 'o's is viewed as the character 'x', however as indicated by the earlier data, the likelihood of 'oor' is higher than 'oxr', so the word is recognized as 'entryway' at long last.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

As of late, Mishra et al. proposed another text recognition strategy [15] in view of the algorithm in [31]. This strategy presents a mistake redress show, which take full favorable position of higher request earlier data, additionally boosting the recognition exactness. Novikova et al. [14] proposed to characterize character appearance and the connection between characters by means of a brought together probabilistic model. Not at all like the algorithm in [31], are character competitors separated utilizing MSER. This technique receives Weighted Limited State Transducers as the probabilistic model and hunts the in all probability word by a proficient thinking algorithm. In any case, the technique of this strategy is convoluted and its pledge recognition execution has no undeniable preferred advantage over other mistake correction strategies that likewise uses statistic language models ([15], [31]).

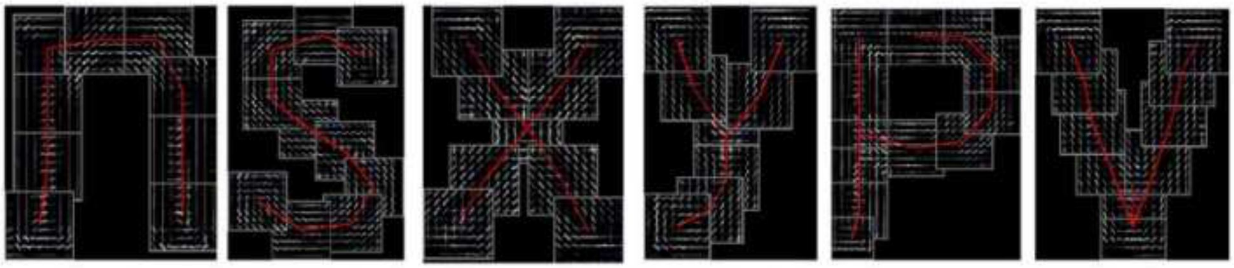


Fig. 5: Part-based tree-structured model for scene text recognition [32]. The structures of characters are manually designed and the parts are annotated by hand.

Fig. 6: Strokelets for scene text recognition [7]. In contrast to [32], the parts of characters (i.e. strokelets) are automatically learned from training data.

Rodriguez-Serrano et al. investigated another approach for text recognition, in which name installing was used to specifically perform coordinating amongst strings and images, by-passing pre-or post-preparing tasks. Last two years, part based text recognition algorithms ([7], [32]) have turned out to be exceptionally well known. Shi et al. [32] ace represented a section based tree-organized model to perceive characters in trimmed images. This algorithm is robust to nose, obscure, incomplete impediment and textual style variety. Be that as it may, this al-gorithm relies upon the nitty gritty explanation data, including character models and part comments (Fig. 5). In [7], Yao et al. proposed a novel portrayal; call Strokelets, which comprises of an arrangement of multi-scale mid-level components (Fig. 6). Strokelets can be naturally gained from character level marks and can catch the auxiliary properties of characters at various granularities. In addition, strokelets give an elective method to precisely distinguish singular characters and form a histogram highlight to successfully depict characters. The scene text recognition algorithm in light of strokelets has turned out to be both powerful and robust.

Fig. 7: End-to-end recognition examples of [11]. This algorithm accomplishes both text detection and recognition, but requires that a lexicon is given for each test image in advance.

V. RELATED WORKS ON END-TO-END TEXT RECOGNITION

The previously mentioned techniques just concern one part of the issue in text data extraction [22] (either text identification or text recognition). There are an assortment of techniques that endeavor to build a bound together system for both text identification and recognition.

Base on the work in Wang et al. [11] proposed a conclusion to-end text recognition system1) (see Fig. 7). Enlivened by general protest recognition algorithms in PC vision, this technique regards words as an uncommon sort of question, and characters as parts of the protest. It looks through the most conceivable discovery and recognition comes about by displaying each single character and the spatial connection between characters. Examinations demonstrate that this technique acquires magnificent execution on numerous standard datasets, this algorithm can just deal with words that are inside the given word list, along these lines it isn't material to images without a word list.

The primary genuine end-to-end text recognition framework for normal images is proposed by Neumann et al. [10], which does not require a word list. This framework removes the character hopefuls by means of MSER and takes out non-text applicants through a prepared classifier. The rest of the applicants are bolstered into a character recognition module, which is prepared utilizing a lot of engineered characters. In light of [10], Neumann et al. presented new component extraction techniques and mix methodologies, which fundamentally enhances the precision and productivity of this framework. Afterward, Neumann et al. [13] additionally broaden the techniques in [10] to achieve continuous text discovery and recognition2) (Fig.8). As of late, Neumann et al. displayed another framework for scene text restriction and recognition, which joins the upsides of sliding-window based and part based techniques. In this framework, character parts (strokes) are demonstrated by arranged bar channels. These situated bar channels are used to perform both character discovery and recognition.

In light of [12], Yao et al. [34] built a conclusion to-end framework that achieves scene text discovery and recognition simultaneously. This is the main work that can limit and read texts of self-assertive introductions in characteristic images.

The considerable great success of deep learning strategies in different PC vision assignments has illuminated specialists in the region of scene text identification and recognition. Coates and Wang et al. utilized CNN with unsupervised relating for text identification and character recognition. Bissacco et al. [17] manufactured a framework, called PhotoOCR, which can read characters in uncontrolled conditions. The center of PhotoOCR is a DNN display running on Hoard features (fig. 9), in-stead of image pixels. Jaderberg et al. [19] proposed another CNN design, which permits highlight sharing for character location, character order and bigram grouping, deep learning based systems, once trained with tremendous data, generally outperform conventional methods by a considerable margin, once prepared with gigantic information, by and large outflank customary strategies by an extensive edge. There are for the most part two downsides of these techniques: (1) They all only handle horizontal or near-horizontal texts; (2) The computational burdens of these algorithms are extremely high. These two disadvantages may compel the advancement and utilization of such algorithms.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Fig. 8: End-to-end recognition examples of [13]. This is the first real end-to-end system for scene text recognition.

VI. RELATED APPLICATIONS AND PRODUCTS OF SCENE TEXT DETECTION AND RECOGNITION

As of late, text identification and recognition in normal scenes have turned out to be dynamic research points. Thus, a great deal of important hypotheses, models, algorithms and frameworks have been proposed created. In the interim, specialists in related fields, for example, robot and media embrace these advancements in applications in robot route, image pursuit and question recognition, accomplishing satisfactory outcomes.

Analysts from the Grip research center in the College of Pennsylvania have effectively invested a robot called 'Graspy'3) with the capacity of finding and perusing characters in normal scenes. While moving in the room, this robot can see the encompassing condition, perceive the characters, doorplates and signs on the divider, and gather its situation as per such data.

Fig.9 Filters learned from characters in natural images [19]. As can be seen, the learned filters can capture the shapes of the corresponding characters.

Tsai et al. [1] has built up a record scan framework for advanced cells. This framework enables the client to take a photo of the interested record, and after that it will naturally read the title of the report and profit the archive put away for the server. Karaoglu et al. joined literary data in characteristic images into the conventional question recognition structure, which additionally enhances the recognition exactness.

What's more, some business items additionally have functionalities identified with scene text discovery and recognition. For instance, the Google Goggles application can read characters on books, Cds and items. The Amazon Firefly application can perceive web locations and telephone numbers in regular scenes.

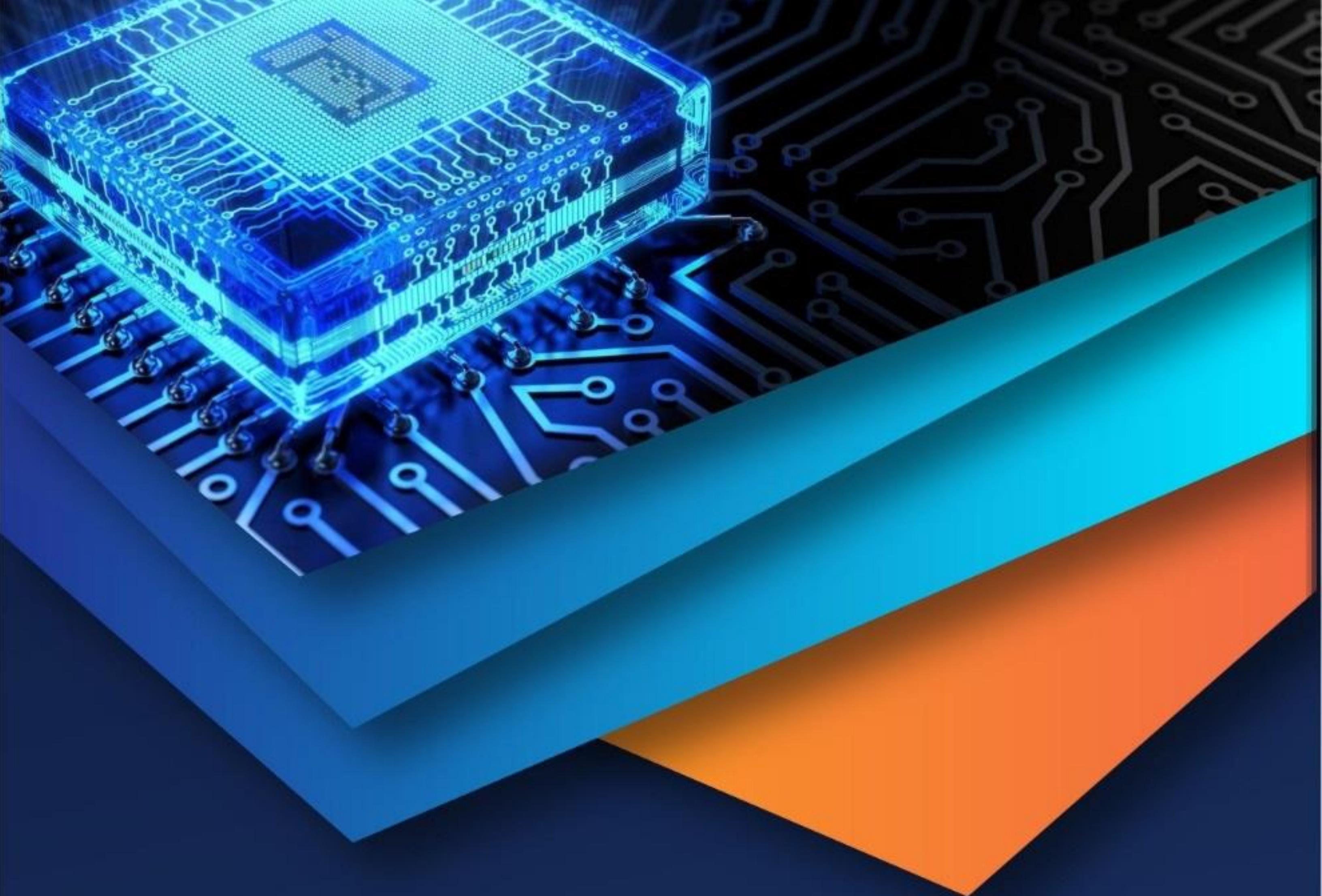
VII. CONCLUSIONS

Text is conceived as an explicit carrier of abnormal state semantics. This one of a kind property makes text not quite the same as other nonspecific visual prompts, for example, shape, color and surface. The rich and exact data embodied in text can help an extensive variety of certifiable applications. Hence, distinguishing and perceiving texts in common scenes have turned out to be critical and dynamic research regions in PC vision. This text audit is gone for following the current advances in scene text identification and recognition, and furnishing different specialists with a full reference to helpful assets in these fields.

Through the endeavors of incalculable researchers, significant advances have been made in scene text location and recognition as of late. In any case, there are a lot of issues that should be tended to later on. To assemble practical frameworks that can precisely and robustly extract textual data from normal scenes, there is as yet far to go. We trust the following viewpoints are qualified to investigate in the following decade:

REFERENCES

[1] 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.
[2] D. B. Barber, J. D. Redding, T. W. McLain, R. W. Beard, and C. N. Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. Journal of Intelligent and Robotic Systems, 47(4):361–382, 2006.
[3] B. Kisacanin, V. Pavlovic, and T. S. Huang. Real-Time Vision for Human-Computer Interaction. Springer, Heidelberg, 2005.
[4] G. N. DeSouza and A. C. Kak. Vision for mobile robot navigation: A survey. IEEE Trans. PAMI, 24(2):237–267, 2002.
[5] Y. K. Ham, M. S. Kang, H. K. Chung, R. H. Park, and G. T. Park. Recognition of raised characters for automatic classification of rubber tires. Optical Engineering, 34(1):102–109, 2005.
[6] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu. Rotation-invariant features for multi-oriented text detection in natural images. PLoS One, 8(8), 2013.
[7] C. Yao, X. Bai, B. Shi, and W. Liu. Strokelets: A learned multi-scale representation for scene text recognition. In Proc. of CVPR, 2014.
[8] X. Chen and A. Yuille. Detecting and reading text in natural scenes. In Proc. of CVPR, 2004.
[9] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In Proc. of CVPR, 2010.
[10] L. Neumann and J. Matas. A method for text localization and recogni-tion in real-world images. In Proc. of ACCV, 2010.
[11] K. Wang, B. Babenko, and S. Belongie. End-to-end scene text recog-nition. In Proc. of ICCV, 2011.
[12] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu. Detecting texts of arbitrary orientations in natural images. In Proc. of CVPR, 2012.
[13] L. Neumann and J. Matas. Real-time scene text localization and recog-nition. In Proc. of CVPR, 2012.
[14] T. Novikova, , O. Barinova, P. Kohli, and V. Lempitsky. Large-lexicon attribute-consistent text recognition in natural images. In Proc. of ECCV, 2012.
[15] A. Mishra, K. Alahari, and C. V. Jawahar. Scene text recognition using higher order language priors. In Proc. of BMVC, 2012.
[16] J. J. Weinman, Z. Butler, D. Knoll, and J. Feild. Toward integrated scene text reading. IEEE Trans. on PAMI, 36(2):375–387, 2013.
[17] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven. PhotoOCR: Read-ing text in uncontrolled conditions. In Proc. of ICCV, 2013.
[18] T. Q. Phan, P. Shivakumara, S. Tian, and C. L. Tan. Recognizing text with perspective distortion in natural scenes. In Proc. of ICCV, 2013.
[19] M. Jaderberg, A. Vedaldi, and A. Zisserman. Deep features for text spotting. In Proc. of ECCV, 2014.
[20] J. Almazan, A. Gordo, A. Fornes, and E. Valveny. Word spot-ting and recognition with embedded attributes. IEEE Trans. PAMI, 36(12):2552–2566, 2014.
[21] D. Chen, J. Luettin, and K. Shearer. A survey of text detection and recognition in images and videos. Technical report, IDIAP, August 2000.
[22] K. Jung, K. Kim, and A.K. Jain. Text information extraction in images and video: a surveyâAI. PR, 37(5):977–997, 2004.
[23] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. IJDAR, 7(2):84–104, 2005.
[24] H. Zhang, K. Zhao, Y. Z. Song, and J. Guo. Text extraction from natu-ral scene image: A survey. Neurocomputing, 122(1):310–323, 2013.
[25] S. Uchida. Text localization and recognition in images and video. Handbook of Document Image Processing and Recognition, pages 843–883, 2014.
[26] L. Kang, Y. Li, and D. Doermann. Orientation robust text line detection in natural images. In Proc. of CVPR, 2014.
[27] Y. Pan, X. Hou, and C. Liu. A hybrid approach to detect and lo-calize texts in natural scene images. IEEE Trans. Image Processing, 20(3):800–813, 2011.
[28] C. Yi and Y. Tian. Text string detection from natural scenes by structure-based partition and grouping. IEEE Trans. Image Process-ing, 20(9):2594–2605, 2011.
[29] W. Huang, Z. Lin, J. Yang, and J. Wang. Text localization in natural images using stroke feature transform and text covariance descriptors. In Proc. of ICCV, 2013.
[30] W. Huang, Y. Qiao, and X. Tang. Robust scene text detection with convolution neural network induced mser trees. In Proc. of ECCV, 2014.
[31] A. Mishra, K. Alahari, and C. V. Jawahar. Top-down and bottom-up cues for scene text recognition. In Proc. of CVPR, 2012.
[32] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang. Scene text recognition using part-based tree-structured character detection. In Proc. of CVPR, 2013.
[33] C. Y. Lee, A. Bhardwaj, W. Di, V. Jagadeesh, and R. Piramuthu. Region-based discriminative feature pooling for scene text recognition. In Proc. of CVPR, 2014.
[34] C. Yao, X. Bai, and W. Liu. A unified framework for multi-oriented text detection and recognition. IEEE Trans. Image Processing, 23(11):4737–4749, 2014.
[35] Y. Zhong, K. Karu, and A. K. Jain. Locating text in complex color images. Pattern Recognition, 28(10):1523âA¸S–1535, 1995.
[36] [36]. K. I. Kim, K. Jung, and J. H. Kim. Texture-based approach for text detection in images using support vector machines and continuously.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY