IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 4 ISSUE : 03 Print / Issue Publication Date: 02-Sep-2019

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2019.v04i03.062

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 3, ISSN No. 2455-2143, Pages 382-390
Published Online July 2019 in IJEAST (http://www.ijeast.com)

A SURVEY ON RECENT METHODOLOGIES IN MULTILINGUAL CHARACTER DETECTION AND RECOGNITION

Snehal S. Gaikwad
Department of Electronics and Telecommunication
Dr. B. A. T. University, Raigad, Maharashtra, India

S. L. Nalbalwar
Department of Electronics and Telecommunication
Dr. B. A. T. University, Raigad, Maharashtra, India

Abstract— **Content discovery and recognition has risen as a significant issue in the previous couple of years. Headways in the field of computer vision, Artificial Intelligence and some constant applications dependent on content identification and recognition has taken more consideration of researchers. Different workshops and conferences are being sorted out on global level giving further ascent to advancements in field of character detection and recognition. Multilingual character detection and recognition from video subtitles, scenes and documents is additionally getting high consideration on this subject. Lot of work has been done for English language but still the issue of handwritten Hindi character recognition isn't yet comprehended sufficient. Different optical character recognition systems perform good for English characters but the accuracy for Hindi character recognition is not up to the mark. This paper surveys different recent ongoing strategies for English and Hindi character recognition. The methodologies include recent features extraction and classification techniques. Moreover, this study includes accepted and published work by research community in recent years. This study also benefits its readers by talking about limitations of existing systems in this field.**

Keywords— **Character detection, Character recognition, Neural network, Line segmentation, SVM classifier.**

I. INTRODUCTION

The area of character detection and recognition focuses on the problem of recognizing text/characters in images which includes natural scene text, street signs, business signs, grocery item labels, and license plates, historical documents, logos etc. multilingual character detection and recognition is also a challenging problem. Character recognition includes various applications like improving navigation for people with low vision, recognizing and translating text into other languages, improving image retrieval and aiding autonomous navigation for cars and robots etc. Sometimes images of documents that have standard fonts, structured text on a plain background and are usually captured in a controlled setting also needs to be recognized due to the text degradation. Natural scenes often contain more extreme lighting variation, may include unusual or highly stylized fonts, often vary widely in color and texture and may be captured from a variety of different viewing angles due to which the text is unrecognizable. Document analysis goes for comprehension of physical structure of images and recognition assigns the names or labels to those distinguished structures.

Optical character recognition (OCR) system works well for English characters but the accuracy decreases for Hindi characters. OCR gives satisfiable recognition rate on printed records; however, their exactness seriously debases on handwritten records; the fundamental reason of this debasement is diverse composition styles of different people. Scientists are still attempting to improve exactness of character recognizers which endures because of handwriting varieties, low-quality pictures, character recognition systems, grouping issues and different other factors. Comparing with the text in the documents, the text in media is in small quantity which often carries information of media contents. They usually present important names, locations, brands of the products, date and time. The features which are widely used for character recognition are structural features like loops, intersections, endpoints, extreme points and statistical features like moments, histograms, profile projections, zoning [1]. The aim of this paper is to find proper way of character detection and recognition from arbitrary complex images of natural scenes and documents. Also, it introduces some widely used databases for character recognition. The various stages for character recognition are shown in Figure 1. Image acquisition is the first stage which captures the image and converts into the digital format. This format is easy to process as the pixels values has form digital. Preprocessing stage removes some artifacts presents in the image of text using some filtering techniques etc. Next, this output is given to feature extraction stage which will extract the features useful for recognition. According to these features, the text is segmented from the background and then classified as text or non-text. The purpose of carrying this

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 3, ISSN No. 2455-2143, Pages 382-390
Published Online July 2019 in IJEAST (http://www.ijeast.com)

study is to introduce the recently used methods for characters recognition and their limitations which would provide the researcher beneficial directions for the research in this field.

The rest of the paper is organized as follows. Section II contains the detailed survey of each character recognition methods for English and Hindi with the limitations of their work, Section III introduces some widely used databases, Section IV shows the applications of character recognition, Section V concludes the survey with future directions for the researchers.

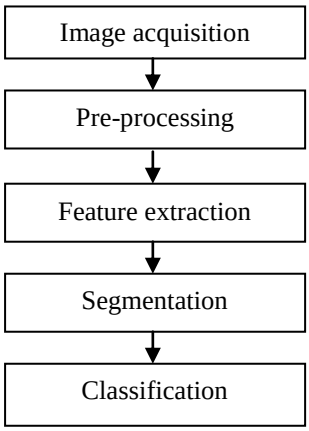

Fig. 1. Stages of character recognition

II. RELATED WORK

A. English character recognition

The character is only organized edges, arrangement of uniform shading regions, a gathering of strokes or a sort of structure. There are numerous objects in nature, for example, leaves, windows, trees which have comparative edges, strokes with character which makes hard to recognize the real content and nontext patterns. In any case, in the pattern that one can profoundly focus on extraordinary variety of text style, shading, perspective proportion then he may effortlessly recognize those examples. A Bayesian based system for content location and recognition has been contemplated. This system has three noteworthy segments viz. Text tracking, Text detection, and text recognition. In this methodology first approach is performed in each edge separately. Each frame is acquired from following directions from the earlier edges and recognition. Bayesian theorem provides a way to calculate posterior probability. This gives Maximum a posteriori (MAP) calculation problem which is to be solved. The video text dataset (USTB-VidTEXT) is used for the experimentation work and results are evaluated using three evaluation metrics like recall, precision and f-score which are formulated below:

$$\text{Recall} = TP/(TP+FN) \quad (1)$$

$$\text{Precision} = TP/(TP+FP) \quad (2)$$

$$\text{F score} = 2.\text{recall}.\ \text{precision} / (\text{recall} + \text{precision}) \quad (3)$$

Where,

TP = similarity between annoted text and recognized text,

TP + FN = total number of ground truths,

TP + FP = total predicted positives.

Their experimental results for text recognition on biscuit joiner sequence from USTB-VidTEXT dataset gives recall of 0.7007, precision of 0.7000 and f score of 0.7003. The performance of their methods degrades when it comes for complex videos like web videos which is still a challenging problem [1].

CNN based approach is introduced by Wenhao He, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu which is having two tasks of localization of text region. First is by a down sampled segmentation based module and to regress the boundaries of text region. Secondly, they proposed a scene text detection framework based on fully convolutional network with bi-task prediction module in which one is pixel wise classification between text and non-text and another one is pixel wise regression to determine the vertex co-ordinates of quadrilateral text boundaries. The feature extraction structures are VGG-16, ResNet-50 and S-VGG. For VGG-16 they used sub network by removing fully connected layers and including additional 3X3 convolutional layers to enlarge respective field. For ResNet-50 they removed the average pooling and fully connected layers. S-VGG follows the design of VGG-16 but adopts less kernels and inserts batch normalization. Second module is multilevel feature fusion which is realized by an iterative process. Third module is Multitask output which is the classification task and Last module is post processing to eliminate densely overlapped quadrilaterals for a word or text line. They evaluated their method on various multilingual datasets like ICDAR2015 Incidental Text Dataset, MSRA-TD500, ICDAR2013 Focused Text Dataset, RCTW-17, CASIA-10K and MLT-17 [2]. The results can be improved to some extent by combining line level and character level segmentation methods on scene text detection.

Gosavi, A., Gurav, A., Bisht, G., represents the Optical Character Recognition (OCR) based text detection and translation on Android platform where the image was captured using mobile phone camera and histogram equalization was applied for contrast adjustment. Then they applied the character recognition task on that contrast adjusted image and removed non-text region based on basic geometric properties such as aspect ratio, Eccentricity, Euler number. Stroke width variation and merged text region is used to remove non-text region for final detection [3].

The unpredictability of conditions, adaptable image acquisition styles and variety of content substance present different difficulties like scene complexity, uneven lightening, blurring and degradation, aspect ratios, distortion, fonts and

multilingual environments. The hole between the technical status and the required performance shows that text detection and recognition stay unsolved issues which allows numerous research opportunities for end-to-end recognition, open vocabulary recognition, processing incidental text, processing multilingual text and real time detection and recognition [4].

Cascaded segmentation detection network based on fully convolutional network is proposed for text-based traffic sign detection where fully convolutional Network is used to detect traffic sign areas providing region of interest (ROI), followed by fast neural Network to detect extracted RoI. This framework for text-based traffic sign detection composed of two parts namely FCN based traffic sign detector and fast text detector. They did the traffic sign detection training where pixels within bounding box of each traffic sign were treated as positive region that is it only focuses on sign area. In testing of traffic sign detection, a salience map of a traffic panel is generated in which two simple morphological operators are used such erosion and dilation. Next, they proposed text detector which is fully convolutional and uses multiple stages for prediction to cover the different respective field. The text box layers are properly designed to predict the output results which are used to form the bounding boxes of text in the image [5].

Yang et al. [6] shows tracking based multi orientation scene text detection. They locate possible character candidates and extracts text region with multiple channels and scales, secondly an optimal tracking trajectory is learned and linked globally over consecutive frames by dynamic programming to refine all detection, recognition and prediction information. They used adaboost and bayesian classifiers to verify extracted text regions. Experiments on multioriented scene text detection gives good results for precision and f-score but it fails for recall performance.

Unsupervised stroke feature learning is used for multilingual text detection with nonlinear neural Network. In this study they focused on the stroke of the text. They used improved discriminative clustering algorithm to obtain language independent stroke feature. Sparse coding is to construct a dictionary and minimize the error in reconstruction. Their initialization framework for stroke detection includes estimating local dispersion metric for each set of data, then selecting the data which have higher metric than a threshold as candidates for initial features and third is to determine initial stroke features from candidates. K means algorithm is used for stroke feature learning which has been identified as fast and effective method to learn images. For detection of the text they proposed two components. Firstly, they used the unsupervised clustering algorithm to generate a set of stroke features and secondly, they built a hierarchy network and combine it with stroke features [7].

Bigrams and trigrams are formed between the perceived characters by considering the spatial pairwise requirements for fine grained item characterization and logo recovery without any training required. In this work they have used ABBYY commercial engine. They proposed a state-of-the-art fine-grained classification approach which combines textual and visual cues to distinguish objects. The experimental performs well for character recognition with recall of 79%. The results can be improved by combining the textual and visual cues which gives better results for logo retrieval [8].

Text-Attentional Convolutional Neural Network which is highly trained with text information for scene text detection has implemented which filters the text component and does classification. Secondly, they developed a powerful low-level detector called contrast enhancement maximally stable extremal regions (CE-MSER) which enhances the intensity contrast between text patterns and background. The original MSER has the disadvantages of easy distortion of text components with various complicated backgrounds which leads incorrect detections of characters. So, authors modified the traditional MSER by adding the cluster-based approach for computing the cluster cues for image globally and then improving the contrast of small regions in an image by using color space smoothing method [9].

Connected component analysis is used to form the regions. All these regions with minimum intensities variation through different thresholds are considered as maximally stable regions. They also used Stroke Width Transform (SWT) to compute stroke width for each pixel to remove nontext regions by using total number stroke widths in a region, their standard deviation and variance [10].

A novel approach towards text detection using superpixel based stroke feature transform (SSFT) and deep learning based region classification (DLRC) for candidate character (CCR) extraction performs well. SSFT mainly composed of three steps:

- *Superpixel segmentation and clustering:* In this input image is first resized into a particular size and then smoothed with an edge preserving filter. Next, it is over segmented using the linear iterative clustering (SLIC) algorithm.
- *Background region removal:* Fast edge detection algorithm is used to extract an edge probability map and gradient orientation map in which each pixel represents its probability of being an edge point and then edge of the image region is extracted by setting the threshold value.
- *Region refinement:* This stage is proposed to remove some of the remaining background regions which are incorrectly segmented.

Deep learning based region classification (DLRC) focuses on extracting hand crafted features and deep CNN based features for region classification. For feature extraction they concentrated on colour features, texture features, geometric features and deep features and these features are fused by

using two fully connected networks (FCNs) for region classification, then these networks are trained by using random forest and SVM. After extraction of text, text region are detected. Experimental results were evaluated on ICDAR2011, ICDAR2013 and SVT dataset [11].

Study towards the fine grained classification where it depends upon the overall appearance of the objects like flower types, species of dogs and birds etc. attracts the researchers to work on it. Karaoglu et al. [12] summarizes about textual contents in images for fine-grained business place classification and logo retrieval. Textual cues are extracted from the image by using two-step procedure. In the initial step, word box proposition are produced to find the words in the image. For word box proposition they detected the character candidates by using text saliency and MSERs. To compensate with the extreme illumination conditions, character candidates are computed using variety of colour spaces like RGB, HSV etc. Then the character candidates are filtered from non-characters regions by considering the features like aspect ratio, size, solidity, contrast. In the second step, the word proposition are utilized as input to a word recognizer to frame the word-level portrayal using 4 layered convolutional neural network. This strategy achieves a word recognition recall of 74.65%

Spatial filters and temporal filters and non-local means (NLM) algorithm are used for noise suppression problem which is based on movement detector (BMD) [13]. The results are evaluated using the metrics like PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) for determining the quality of images and videos. Ahn, Ryu, Koo and Cho [14] proposed Binarization algorithm for text line detection in degraded historical documents. In binarization they first identified the non-text region by doing binarization on gray image using Otsu's method and morphological dilation is done to remove some connected components from the ROI. In text line detection they applied the non-text connected component filtering on binarized image using scale invariance with median height of characters and then main textbox is detected using projection profile analysis. CC grouping is done by using stroke width and by estimating the skew angle for final text line detection. The experiments were carried out on ICDAR 2015 text line detection in historical documents with the evaluation measure like an origin point (OP) based scoring method which gives the coordinates as an intersection of the baseline of each text line and the left-most edge of the first character, recognition accuracy, detection rate and F measure.

A new comprehensive dataset for scene text detection and recognition task consists of horizontal, multi oriented and curved text with 1555 scene images, 9330 annoted words. They used DecovNet for semantic segmentation for text detection [15].

Shi et al. [16] introduced the ASTER (An attentional scene text recognizer) which is end-to-end system with filtering network and recognition network. Filtering network first predicts the set of control points through localization network, then Thin Plane Spline Transformation is calculated from control points and passed towards the grid generator and sampler to generate rectified image. Recognition task is divided into two steps which are Encoder and Decoder. Encoder extracts the feature map from the input image and given to bidirectional LSTM to analyze the feature sequence bidirectionally. Decoder converts this feature sequence into character sequence. Shi, B., Bai, X. and Yao, C., [17], the solution for the problem of scene text recognition using neural network architecture has been explained. This neural network model is combination of deep CNN and recurrent NN. Recurrent NN includes three layers namely Convolutional layers (extracts the feature sequence from each input image), Recurrent layers (predicts each frame of feature sequence) and Transcription layer (translates the per frame predictions by the recurrent layers into a label sequence). Convolutional layer takes the input image and extracts the features which are arranged in a vector form for further processing. In recurrent layer these features are stored using Long-Short term memory (LSTM) which acts as a memory device and then those stored features are distributed and predicted in transcription layer to get the exact sequence of characters. L0- regularized intensity and gradient based technique for image deblurring is shown in [18]. An algorithm based on half quadratic splitting technique has been proposed for latent image estimation and then estimated the blur kernel, to remove the artifacts from the image. The limitation of this method is it fails when blurred image contains a large amount of noise. Ramisa, Yan, Noguer and Mikolajczyk proposed an adaptive CNN architecture for automatic image retrieval for complex textual cases [19]. They focused on the textual contents in news articles where they used new adaptive CNN architecture for source detection, article illustration and geolocation of articles. For article illustration they used deep canonical correlation analysis and for geolocation they proposed new loss function based on great circle distance. Limitation of their work is, the quality of generated captions is much lower for news illustration task compared to another dataset.

The markov random field (MRF) structure for looking through content in the documents has been clarified in [20]. It has two primary segments. First is proficient handling and coordinating of visual highlights where the SIFT highlights are separated utilizing Fast corner identifier from each local patch placed over the corner points. Hierarchical k-means is utilized for quantizing SIFT descriptors to visual terms.

Reference [21] explains the reviews on the methods for text feature extraction based on deep learning which learns feature representation automatically from big data. Filtering is the quick methods to extract the features from large scale text which includes finding the word frequency in the text, mutual information between two objects and information gain. Different classifiers can be integrated with the weighting methods which assigns weight between (0, 1) to train each feature. Commonly used Mapping methods are LSI (latent semantic index) and PCA gives best results to text classification

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 3, ISSN No. 2455-2143, Pages 382-390
Published Online July 2019 in IJEAST (http://www.ijeast.com)

Reference [22] goes for restoring total character forms in video/scene pictures from dim qualities, as opposed to the customary methods that thinks about binary data as contributions for character detection and recognition. Zero crossing points calculated by using Laplacian operations on the image to recognize stroke candidate pixels (SPC). These Laplacian images are calculated by using following formula:

$$I_x = I * Mask_x \qquad x \in \{1, 2, 3, 4\} \qquad (4)$$

Where,

$Mask_x$ = one- dimensional Laplacian mask which is convolved with image I

x =1,2,3,4 denote the horizontal, the vertical, the diagonal and the secondary diagonal Laplacian images respectively.

For each SPC pair, filtering method is applied to check gradient symmetry, distance symmetry and colour symmetry. Histogram operation is performed on each PSCPs to pick seed stroke candidate pairs (SSCP). The PSCP that gives a dominant peak is considered as the actual PSCP pair. At long last, an iterative calculation is proposed for SSCP to restore total character forms. Quality measures like MSE (mean square error), PSNR (peak signal to noise ratio) and SSIM (structural similarity index measure) are used to evaluate the contours. Proposed method has good SSIM when evaluated on ICDAR2013 and ICDAR2015 video data but it fails for PSNR and MSE.

B. Hindi character recognition

Work on OCR of Indian scripts began in the decade of 1970. The Devanagari content gives a rich composition framework to various Indian counting Sanskrit, Hindi, Marathi and Nepali among others. It contains 11 vowels and 33 consonants shown in Figure. 2. Research on Devanagari OCR started a couple of decades prior and a few OCR systems for the Devanagari character recognition have surfaced from that point [23].

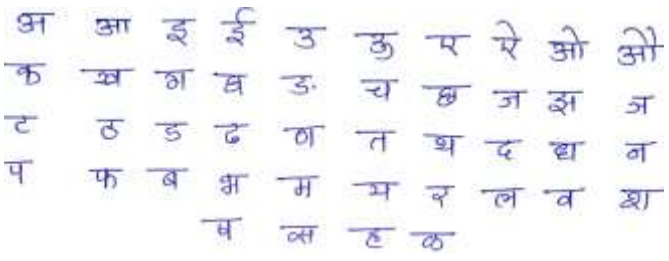

Fig. 2. Devanagari vowels and consonants

The progressing exploration to improve reading abilities of computers offered ascend to the field of document analysis and recognition (DAR). DAR analyses the handwritten or printed documents and does their recognition. In recognition process, the labels are assigned to those recognized structures. Some commonly used applications of optical character recognition (OCR) systems of Hindi language are automated postal addressing, zip code reading, data acquisition in bank checks and institutional records [24]. Each stage of character recognition system shown in Figure. 1 should be able to perform well towards the recognition process, otherwise performance degrades. Recognition of handwritten Hindi characters is still a very challenging task which suffers from handwriting variations, low-quality images, character recognition techniques, classification problems and various other factors. A very limited work has been done on word level skew detection and correction for handwritten Hindi characters. Researchers are also attracted towards the ancient documents which are preserved in the museums. The characters in ancient documents are often overlapped, so thinning algorithm can badly distort them. Touching character and half character is considered one of the major problems in handwritten word recognition. The recent recognition methods are good at recognizing the good quality characters but fails in recognizing overlapped, similar shaped characters which reduces the recognition rate.

Any alphabet can be divided into the features like vertical bar, intersections, shirorekha touch frequency and number of endpoints and slopes of ending curves [25]. Preprocessing is a very important stage in character recognition which contains smoothening, binarization, thinning and spur removal from the character. Hindi character recognition has also a very important application of learning akshara orthography. The challenge in this case is Phonological awareness at phoneme level is influenced by orthographic representation [26]. Reference [27] focused on DCT features of ancient characters. They used 5484 samples of presegmented basic Devanagari characters dataset. These features are then classified using support vector machine (SVM), Naïve Bayes and Decision tree classifiers. They added adaboost (Adaptive boosting) and bagging with classifiers to improve the recognition results. They decided the feature vector length for classification among which length of 100 feature vector performs good with RBF SVM classifier. But this work fails for modifiers and conjuncts.

Conjuncts in the Devanagari scripts are very difficult to segment. [28] Presents a novel approach for segmentation of Devanagari characters from ancient documents. Document image is segmented into lines, words and characters. For segmentation they used piecewise projection profile concepts where they concentrated on vertical and horizontal projection of character, average line height and connected components. They achieved line segmentation accuracy of 97%, character accuracy of 98.5% and overlapping and touching components accuracy of 100% and 96% respectively. If the average line height is not computed correctly then accuracy degrades.

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 3, ISSN No. 2455-2143, Pages 382-390
Published Online July 2019 in IJEAST (http://www.ijeast.com)

A study of recognition for handwritten English, Hindi and Punjabi characters is presented in [29]. They reported some challenges which affects the text recognition rate which are mentioned below:

- Similar shaped characters with same meaning
- Extraction of appropriate features and selection feature extraction techniques
- Low quality images
- Cursive case of handwriting
- Inappropriate scanning and binarization

For classification they used KNN (K nearest neighbor), SVM (support vector machine) and MLP (multilayer perceptron) classifiers. They achieved accuracy rate of 84.67% for Hindi characters.

Characters has two features namely structural and statistical. Structural features include the shape and overall structure of character where statistical features includes horizontal and vertical peak content, centroid, intersection and open end points [29, 30].

Line segmentation in Devanagari script has gain a lot of attention of researchers as it contains the conjuncts, overlapped characters and slanting lines. So, it becomes a challenging task to recognize those. Reference [30] represents the Devanagari characters recognition from ancient documents. Image is divided into fixed size vertical stripes. For each vertical stripe horizontal projection profile (HPP) is calculated. Depending upon the threshold for each HPP, piecewise separating lines (PSLs) are considered for each row. Then average line height is calculated to handle over and under segmentation.

III. DATASETS

A. Dataset for English characters

This section shows some commonly used datasets for English language along with their features including the category, orientation and language.

1. ICDAR'2003

This dataset contains natural scene images with horizontal orientation available in English language. This dataset is intended to be used for evaluations of block-based text detection algorithms [31].

2. MSRA-I (2004)

This dataset contains Graphic & scene text with Horizontal orientation available in English, Spanish, Chinese language [32].

3. KIST (2010)

This dataset contains distorted natural scene images available in English and Korean language [33].

4. SVT (2010)

This dataset contains natural scene images with horizontal orientation available in English language [34].

5. Tan (2011)

This dataset contains multioriented Graphic & scene text available in English and Chinese language [35].

6. ICDAR'11 (2011)

This dataset contains natural scene images with horizontal orientation which is available in English language and distorted Graphic text available in English language [36].

7. IIIT5K Word (2012)

This dataset contains distorted Graphic & scene text available in English language [37].

8. MSRA-II (2012)

This dataset contains multioriented scene text available in English and Chinese language [38].

9. ICDAR'13 (2013)

This dataset contains natural scene images with horizontal orientation which is available in English language and distorted Graphic text available in English language. It contains 229 training images and 233 test images which are well captured and clear [39].

10. ICDAR'2015

Incidental Text Dataset. This dataset contains 1000 training images and 500 test images. This dataset contains texts with various scales, blurring, orientations and view-points. The annotation of bounding box (actually quadrilateral) has 8 coordinates of four corners in a clock-wise manner. In evaluation stage, word-level predictions are required [40].

11. MSRA-TD500

This dataset contains 300 training images and 200 test images, where there are many multi-oriented text lines. Texts in this dataset are stably captured with high resolution and are bi-lingual of both English and Chinese. In evaluation stage, line-level predictions are required [41].

12. ICDAR'2017/MLT-17

Multi-lingual scene text detection (MLT-17) deals with various scripts and languages in natural scenes. The dataset is composed of complete scene images which come from 9 languages representing 6 different scripts. There are 7200 training images, 1800 validation images and 9000 test images [42].

B. Dataset for Hindi characters

To the best knowledge of authors, there are three standard databases are available for Hindi characters. The details about these databases are given below.

1. ISI database

This is Hindi characters database having 30,000 samples of basic Hindi characters. These characters are written by 1049 different writers [43].

2. Vijay Dongre database

This database is created by researcher named Vijay Dongre which contains 5137 numerals and 20,305 isolated characters. These numerals and characters are written by 750 different writers [44].

3. HPL offline dataset

This database is generated using from digital equipment and not collected by scanning character images [45].

IV. APPLICATIONS

Various applications of character detection and recognition from images and videos have gained many advantages in the domestic life in past few years with advancements in image processing techniques.

Character 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, newspaper reading out system, signage recognition and conversion to text. It is used for automatic number plate recognition at toll booths, for street boards reading purpose in case of unmanned vehicles as well as travelling assistant system. Text detection and recognition has very important application in 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. Application of text detection and recognition also gained high concentration in security based application for logo retrieval also.

V. CONCLUSION AND FUTURE SCOPE

Character or text is one of the most expressive means of communications, and can be embedded into documents or into scenes as a means of communicating information. This is done in the way that makes it noticeable and/or readable by others. The process is divided into text detection and localization, classification, segmentation and text recognition. This paper also introduces the recent techniques for text detection and recognition and attempts the comprehensive literature survey of text detection and recognition research for English and Hindi language. It summarizes the various approaches and also highlights the state-of-the-art methods with publicly available datasets.

The brief review of text/character detection and recognition methods come up with some unsolved problems which provides possible research direction.

- In scene images, background region could be hardly distinguished from text without context information. Illumination also effects on this problem. So, Text detection algorithm should be more optimized to increase the processing speed to real time.
- Incidental text suffers from image degradation, distortions, font variations and cluttered background. Many approaches have been applied to capture images with offline processing modes. Combining text detection and recognition with text tracking algorithm will improve text detection and recognition accuracy and also enhance time performance.
- Fully convolutional based traffic sign detection method fails for oriented angle of text where more improvement is required for efficiency and accuracy.
- Line level methods fails to detect some characters at the end of text lines and character level methods ignores global information of text line. So, line level and word level methods can be combined for better results.
- Due to motion blur and perspective distortion there is degradation in text tracking. Also, scene text detection in video is time consuming. So, there is a need to develop an algorithm at segmentation level which will increase the accuracy.
- Stroke feature to detect text is not sufficient and its implementation is essential for English and Hindi language.
- MSER is negatively affected by blurriness and over resolution. While calculating stroke widths, edge map of the image is acquired poorly. Also, OCR assistance improves precision but decreases recall performance, so OCR assistance step can be avoided to increase recall performance.
- There is need to develop the state-of-the-art methods for Hindi conjuncts and half characters.

Funding: This study was not funded.

VI. REFERENCE

[1] Tian, S., Yin, X.C., Su, Y. and Hao, H.W., 2017. A unified framework for tracking based text detection and recognition from web videos. IEEE transactions on pattern analysis and machine intelligence, 40(3), pp.542-554.

[2] He, W., Zhang, X.Y., Yin, F. and Liu, C.L., 2018. Multi-oriented and multi-lingual scene text detection with direct regression. *IEEE Transactions on Image Processing*, *27*(11), pp.5406-5419.

[3] Gosavi, A., Gurav, A., Bisht, G., 2016. Text Detection and Translation. International Journal of Engineering Science and Computing, 6(4), pp.3306-3309.

[4] Yin, X.C., Zuo, Z.Y., Tian, S. and Liu, C.L., 2016. Text detection, tracking and recognition in video: a comprehensive survey. *IEEE Transactions on Image Processing*, *25*(6), pp.2752-2773.

[5] Zhu, Y., Liao, M., Yang, M. and Liu, W., 2017. Cascaded segmentation-detection networks for text-based traffic sign detection. IEEE transactions on intelligent transportation systems, 19(1), pp.209-219.

[6] Yang, C., Yin, X.C., Pei, W.Y., Tian, S., Zuo, Z.Y., Zhu, C. and Yan, J., 2017. Tracking based multi-orientation scene text detection: A unified framework with dynamic programming. IEEE transactions on image processing, 26(7), pp.3235-3248.

[7] Li, L., Yu, S., Zhong, L. and Li, X., 2015. Multilingual text detection with nonlinear neural network. Mathematical Problems in Engineering, 2015.

[8] Karaoglu, S., Tao, R., van Gemert, J.C. and Gevers, T., 2017. Con-text: Text detection for fine-grained object classification. *IEEE transactions on image processing*, *26*(8), pp.3965-3980.

[9] He, T., Huang, W., Qiao, Y. and Yao, J., 2016. Text-attentional convolutional neural network for scene text detection. IEEE transactions on image processing, 25(6), pp.2529-2541.

[10] Özgen, A.C., Fasounaki, M. and Ekenel, H.K., 2018, May. Text detection in natural and computer-generated images. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE..

[11] Tang, Y. and Wu, X., 2018. Scene text detection using superpixel-based stroke feature transform and deep learning based region classification. IEEE Transactions on Multimedia, 20(9), pp.2276-2288.

[12] Karaoglu, S., Tao, R., Gevers, T. and Smeulders, A.W., 2016. Words matter: Scene text for image classification and retrieval. IEEE transactions on multimedia, 19(5), pp.1063-1076.

[13] Kaplun, D.I., Voznesenskiy, A.S., Klionskiy, D.M., Geppener, V.V. and Serzhenko, F.L., 2017. Study of spatiotemporal processing algorithms for video and image processing. Pattern Recognition and Image Analysis, 27(4), pp.686-694.

[14] Ahn, B., Ryu, J., Koo, H.I. and Cho, N.I., 2017. Textline detection in degraded historical document images. EURASIP Journal on Image and Video Processing, 2017(1), p.82.

[15] Ch'ng, C.K. and Chan, C.S., 2017, November. Total-text: A comprehensive dataset for scene text detection and recognition. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 935-942). IEEE.

[16] Shi, B., Yang, M., Wang, X., Lyu, P., Yao, C. and Bai, X., 2018. Aster: An attentional scene text recognizer with flexible rectification. IEEE transactions on pattern analysis and machine intelligence.

[17] Shi, B., Bai, X. and Yao, C., 2016. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 39(11), pp.2298-2304.

[18] Pan, J., Hu, Z., Su, Z. and Yang, M.H., 2016. L_0 -regularized intensity and gradient prior for deblurring text images and beyond. IEEE transactions on pattern analysis and machine intelligence, 39(2), pp.342-355.

[19] Ramisa, A., Yan, F., Moreno-Noguer, F. and Mikolajczyk, K., 2017. Breakingnews: Article annotation by image and text processing. IEEE transactions on pattern analysis and machine intelligence, 40(5), pp.1072-1085.

[20] Yalniz, I.Z. and Manmatha, R., 2017. Dependence models for searching text in document images. IEEE transactions on pattern analysis and machine intelligence, 41(1), pp.49-63.

[21] Liang, H., Sun, X., Sun, Y. and Gao, Y., 2017. Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017(1), p.211.

[22] Wu, Y., Shivakumara, P., Lu, T., Tan, C.L., Blumenstein, M. and Kumar, G.H., 2016. Contour restoration of text components for recognition in video/scene images. IEEE Transactions on Image Processing, 25(12), pp.5622-5634.

[23] Yadav, M., Purwar, R.K. and Mittal, M., 2018. Handwritten Hindi character recognition: a review. IET Image Processing, 12(11), pp.1919-1933.

[24] Chaudhuri, A., Mandaviya, K., Badelia, P., K Ghosh, S., 2017. Optical Character Recognition Systems for Different Languages with Soft Computing. pp.352.

[25] Sharma, R. and Mudgal, T., 2019. Primitive Feature-Based Optical Character Recognition of the Devanagari Script. In Progress in Advanced Computing and Intelligent Engineering (pp. 249-259). Springer, Singapore.

[26] Bhide A., Perfetti C. (2019) Challenges in Learning Akshara orthographies for Second language Learners. In: Joshi R., McBride C. (eds) Handbook of Literacy in Akshara Orthography. Literacy Studies (Perspectives

from Cognitive Neurosciences, Linguistics, Psychology and Education), vol 17. Springer, Cham

[27] Narang, S.R., Jindal, M.K. and Kumar, M., 2019. Devanagari ancient character recognition using DCT features with adaptive boosting and bootstrap aggregating. Soft Computing, pp.1-12.

[28] Narang, S.R., Jindal, M.K. and Kumar, M., 2019. Drop flow method: an iterative algorithm for complete segmentation of Devanagari ancient manuscripts. Multimedia Tools and Applications, pp.1-26.

[29] Kumar, M. and Jindal, S.R., 2019. A Study on Recognition of Pre-segmented Handwritten Multi-lingual Characters. Archives of Computational Methods in Engineering, pp.1-13.

[30] Narang, S., Jindal, M.K. and Kumar, M., 2019. Devanagari ancient documents recognition using statistical feature extraction techniques. Sādhanā, 44(6), p.141.

[31] Lucas, S.M., Panaretos, A., Sosa, L., Tang, A., Wong, S. and Young, R., 2003, August. ICDAR 2003 robust reading competitions. In Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. (pp. 682-687). IEEE.

[32] Hua, X.S., Wenyin, L. and Zhang, H.J., 2004. An automatic performance evaluation protocol for video text detection algorithms. IEEE transactions on circuits and systems for video technology, 14(4), pp.498-507.

[33] Lee, S., Cho, M.S., Jung, K. and Kim, J.H., 2010, August. Scene text extraction with edge constraint and text collinearity. In 2010 20th International Conference on Pattern Recognition (pp. 3983-3986). IEEE.

[34] Wang, K. and Belongie, S., 2010, September. Word spotting in the wild. In European Conference on Computer Vision (pp. 591-604). Springer, Berlin, Heidelberg.

[35] Shivakumara, P., Sreedhar, R.P., Phan, T.Q., Lu, S. and Tan, C.L., 2012. Multioriented video scene text detection through Bayesian classification and boundary growing. IEEE Transactions on Circuits and Systems for Video Technology, 22(8), pp.1227-1235.

[36] Shahab, A., Shafait, F. and Dengel, A., 2011, September. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. In 2011 international conference on document analysis and recognition (pp. 1491-1496). IEEE.

[37] Mishra, A., Alahari, K. and Jawahar, C.V., 2012, June. Top-down and bottom-up cues for scene text recognition. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2687-2694). IEEE.

[38] Yao, C., Bai, X., Liu, W., Ma, Y. and Tu, Z., 2012, June. Detecting texts of arbitrary orientations in natural images. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1083-1090). IEEE.

[39] Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., i Bigorda, L.G., Mestre, S.R., Mas, J., Mota, D.F., Almazan, J.A. and De Las Heras, L.P., 2013, August. ICDAR 2013 robust reading competition. In 2013 12th International Conference on Document Analysis and Recognition (pp. 1484-1493). IEEE.

[40] Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L., Chandrasekhar, V.R., Lu, S. and Shafait, F., 2015, August. ICDAR 2015 competition on robust reading. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1156-1160). IEEE.

[41] Yao, C., Bai, X., Liu, W., Ma, Y. and Tu, Z., 2012, June. Detecting texts of arbitrary orientations in natural images. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1083-1090). IEEE.

[42] Nayef, N., Yin, F., Bizid, I., Choi, H., Feng, Y., Karatzas, D., Luo, Z., Pal, U., Rigaud, C., Chazalon, J. and Khlif, W., 2017, November. Icdar2017 robust reading challenge on multi-lingual scene text detection and script identification-rrc-mlt. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 1454-1459). IEEE.

[43] Bhattacharya, U. and Chaudhuri, B.B., 2005, August. Databases for research on recognition of handwritten characters of Indian scripts. In Eighth International Conference on Document Analysis and Recognition (ICDAR'05) (pp. 789-793). IEEE.

[44] Dongre, V.J. and Mankar, V.H., 2012. Development of comprehensive devnagari numeral and character database for offline handwritten character recognition. Applied Computational Intelligence and Soft Computing, 2012, p.29.

[45] Bharath, A. and Madhvanath, S., 2009. Online handwriting recognition for Indic scripts. In Guide to OCR for Indic scripts (pp. 209-234). Springer, London.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143