Text Detection for Multi-Orientation Scene Images using Adaptive Clustering

Baviskar Vaibhav G.
M. E. Computer (Student)
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik
S. P. P. U.

Mankar J. R.
Assistant Professor
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik
S. P. P. U.

ABSTRACT
Detection of text in camera-based images is a vital requirement for several computer vision applications. Text detection task is frequently challenging due to difficulties like composite backgrounds, dissimilarities of text orientations, font, size, color. The aim is to recognize text in a combine manner by searching for words from the image into text areas or single character candidates. Text captured in natural scenes is most of the times with multiple orientations and point of distortions. Currently most research efforts focuses on horizontal orientation from images. To address same issues a novel approach unified distance metric learning framework is proposed an adaptive hierarchical clustering, which learns weights of the character candidates once at a time and adaptively integrate different feature similarities. An effective multi-orientation text detection system, which constructs the text character candidates by grouping characters based on an adaptive clustering.

Keywords
Distance metric learning, Multi-orientation, Scene text detection

1. INTRODUCTION
Detection and extraction method of text plays a vital role in many computer vision applications and in great demand for applications like reading, image retrieval, Image analysis, etc. Information gathering in form of extracting of text from natural scene images is a bid task because of dissimilarities of text format like font, color, scale, and orientation alignment, shape, composite backgrounds and texture, geometry of text, image low and high resolution, image illumination, layout, image distortion, blurring and lighting condition. A text extraction in natural scene contains useful and valuable information and makes it easy to understand for human beings and computer machines. This research topic is very active and challenging task in many computer vision applications.

In real natural scenes most of the times text captured in multiple orientations and perspective distortions, however text detection is attentive on horizontal scene text detection. There are not many techniques suggested for multi-orientation text detection, and extraction for natural scene can be roughly analyze into three techniques: Region based technique, texture based techniques and hybrid technique [3].

Fig 1 Examples of Multi-orientation text natural scene Images

Region based technique is technique that uses a bounding box or sliding window to detect a text from a natural scene and uses different technique to recognize text. In this approach a text region is identified from a complex background and eliminates the false or non-text region. This approach is based upon color, edge, shape, and contour and geometry features [2]. On the basis of these features separates text or false text region. The performance of region based technique is slow as compared to other techniques [12].

Edge based and Connected Component is a further classification of the region based approach. The texture based technique [8] uses different texture properties to extract a text from a complex image. Various techniques are used in textual information like Wavelets, Fourier Transform and Gabor filters, Transform Wavelet etc. A train classifier is preferred to extract the features of the target image region [2]. The main aim of train classifier is to distinguish the text or non-text region for a scene [8].

The hybrid technique [4] uses a combination of both techniques, i.e. region based and texture based technique. In this, first step region based technique is preferred to detect a text or character candidate using the Character Component [CC]method. The features are extracted from text region and use a classifier to decide which region contains a text or non-text on the basis of texture based technique. The main disadvantage of these techniques that the single technique is compatible for all the natural scene images due to size, color, font variation varies from one image to another image

For better understanding of the concepts it is divided into sections: Section 1 describes the introduction of the system and motivation of the proposed system. Section 2 describes the related work in which motivational survey, efficiency and drawbacks of previous system are discussed. Section 3 describes the detailed design of the proposed system. Section 4 presents the experimental results for demonstrating the validity of the proposed system for the large scale datasets and Section 5 describes the conclusion.

2. LITERATURE SURVEY

This Section explains the methods studies of all existing systems are included which were used pre-viously. By studying this, it gives all information, advantages, disadvantages and limitations of the existing systems.

X.-C. Yin, X. Yin [9] designed hybrid and robust technique to detection and localization of text in natural scene using scale-adaptive binarization to extract a candidate character. Then apply conditional random field (CRF) model is used to filter the non-text region. At last step, the energy minimization method is used to club the text line or text character region. The images come from ICDAR 2005 database and achieve a good result.

Anhar Risnumawan [10] develop a robust method for text detection in natural scene use a properties or features, i.e. Mutual Magnitude Symmetry (MMS), Mutual Detection Symmetry (MMD) and Gradient Vector Symmetry (GVS) to detect the text candidate from natural scenes. Local descriptor SIFT exploring the pixel text , identify the text candidate or remove the non-text pixels(candidate).Then apply the ellipse growing method which is based upon the text orientation , extract the text , restore and eliminate the non-text character in it. In this, the proposed method work on three datasets and not depend upon contrast, orientation, fonts, resolution and text size.

Xu-Cheng Yin [11] propose an accurate and robust text detection technique based upon the Maximally Stable Extremal Region (MSERs) the text character is extracted weather the condition of the image is bad. A self-trained distance metric method is used that learns weights and single link methods use the learned parameters. The classifier is used to identify the posterior probability of text character and remove the non-text candidate character. The proposed technique is very effective used the ICDAR 2013 dataset in "Text Localization in Born-Digital image" and "Text Localization in Real Scène". In this paper, the ICDAR 2011 database is used and gives a 76% f-measure which is good. the candidate text region is a real text or not in image layer. In the last step, OCR package is used to recognize the real text which is localized by the trained classifier. The accuracy or recognition rate of the text region improves on this layer method.

Chucai Yi and Yingli Tian [12] present a framework to extract and localization a text from the complex natural scene by using three steps, i.e. boundary clustering, stoke segmentation and string fragmentation classification method. The text is automatically extracted and localization from natural scene by using three phases: pixel, character and string on the basis of the features. They proposed a two method to combine a stoke text and filter the non-text region. After the Gabor-based method is used to string fragment classification on the basis of text features. This method work upon a natural scene text, born-digital images, pictures captured by a blind person, broadcast videos, ICDAR 2003 and ICDAR 2011 dataset.

Cong Yao [13] proposed a unified method of text detection and recognition of the multi-oriented natural scene using same features and classification method. They proposed a new dictionary search based method is used to detect and recognition errors and correct them. This technique work on different font, scales, orientation and color text in natural scenes .The proposed method mainly focus on multi-oriented text and achieves a performance parameter of f- measure 73 % work on four databases.

For improving the performance of the large scale image dataset, it is key to find out the real and well organized method for image retrieval. To improve the accuracy the MPEG-7 descriptor is used.

3. IMPLEMENTATION DETAILS

The proposed system has several advantages over existing one:

1. Proposed system is better enough to detect texts with Multi-orientations as well as scales with complex background in Natural scene images.
2. Proposed method can well preserve the text characters integrality and under maximally extremal regions (MESR).
3. Morphology clustering, Orientation clustering and Projection clustering for effectively use as the key feature.

3.1 Adaptive Hierarchical Clustering

A hierarchical structure based 2-dimensional proximity matrix is design with the help of an hierarchical clustering and also arrange data into a hierarchical structure manner. The outcomes are typically presented by a binary tree or dendrogram. From these outcomes different clusters of the data formation is done. Commonly, two major difficulties are faced in hierarchical clustering technique

(1)The measure of the proximity of data

(2)Determination of cluster numbers for obtaining the accurate clustering outcomes.

Usually segmentation and clustering-based textual information extraction and detection techniques are tedious]. Additionally existing considerable clustering techniques with, the following procedure is metric learning has unique attention on partitional clustering[23]. Adaptive hierarchical clustering technique involves metric learning have couple mentioned difficulties [24]. This system is simply constructed with help of three fundamental phases

1) Sample selection,

2) Weight conversion,

3) Model determination.

Single-link clustering technique [11] with distance metric learning is strongly preferred for horizontal scene text detection and can be seen as a precise case of this system. The system can easily involve various hierarchical clustering techniques e.g., Single link clustering and divisive hierarchical clustering, Therefore an analysis of the different classes of available clustering techniques with big datasets may provide outstanding and useful conclusions. A set of data samples are represented by $X=\{x_1,\ldots..,x_j,\ldots..x_N\}$ where $x_j=(x_{j1}\ldots x_{j2},\ldots.,x_{jn})T \in R^{\eta}$ known as feature vector. Tree like nested structure constructed with the help of this proposed technique for data pattern X. Some initial knowledge for data variables, from the similarity view, Link points are,

$$d(x_i, x_j) = \epsilon \quad (x_i, x_j) \in S$$

Where ϵ is the threshold value. Non link points are,

$$d(x_i, x_j) > \epsilon \quad (x_i, x_j) \in D$$

The threshold value ϵ is used to get the cluster numbers.

3.2 Distance Metric Learning Framework

Distance metric learning form is expressed as

$$d(x_i, x_j: \omega) = \omega^T \text{vec}(x_i, x_j), \quad (3.2.1)$$

where weight vector ?, vec (xi, xj) is the similarity vector of two variables xi and xj. In this, the aim of this framework is to forming the two sets of clusters i.e. set S for same pair of points and set D for different pair of points. The distance of pair of points in set D is maximized and the same is minimized in set S. this framework is also capable of providing tough and indicative problems which are responsible for the formation of number of representative part of the problem, i.e., given the labeled cluster set {Ck}mk=1(with m clusters), the following strategy is used to compute D and S

$$D = \{(\hat{x}_k, \hat{y}_k) = \underset{x \in Ck, y \in C-k}{\arg\min} d(x, y; \omega)\}^m_{k=1} \quad (3.2.2)$$

$$S = \{(\ddot{x}_k, \ddot{y}_k) = \underset{x \in Ck}{\arg} \max_{y \in Ck} \min d(x, y; \omega)\}^m_{k=1}$$

For weight conversion the same procedure

$$\theta = [-\epsilon \omega]^T, \text{ Hence, } d(x_i, y_j; \theta) \quad (3.2.3)$$

For 2nd phase expression is denoted by,

$$\theta^* = \arg\min J(\theta: D, S) \quad (3.2.4)$$

Where J (θ: D, S) is the objective function.

3.3 Process Block Diagram

First an adaptive hierarchical clustering algorithm with a unified distance metric learning framework is proposed. This framework has the advantages that both similarity weights and the clustering threshold can be simultaneously optimized. Consequently, this approach select and learn many parameters used in this text detection system.

Second, text candidate's construction includes several sequential character grouping steps: Morphology clustering first coarsely groups character candidates with similar appearance then orientation clustering group's character pairs with consistent orientation. Finally, projection clustering finely separates text lines in the same orientation and grouping, constructing of text candidates with multiple orientations is based on maximally stable extremal region (MSER) technique.

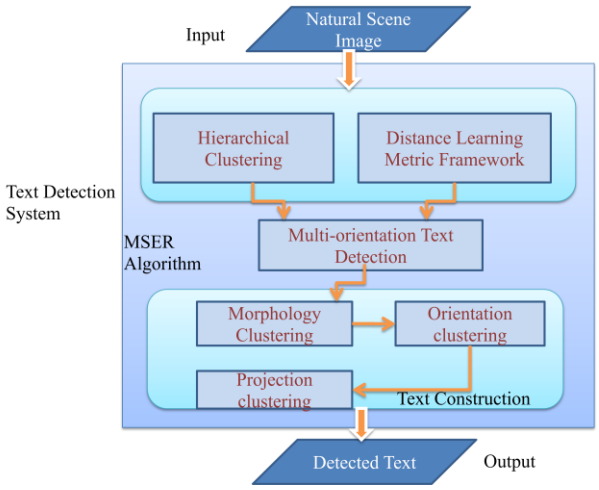

Fig. 2: Pipeline of the proposed Text Detection System

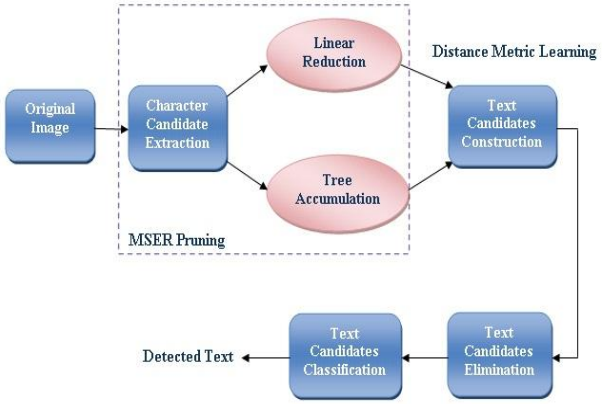

Fig. 3: Steps in Multi-orientation Text Detection.

4. RESULTS AND DISCUSSION

In order to assess the performance of the proposed system A challenging multi-orientation natural scene text data set (USTB-SV1K)1, images of which are directly crawled from Google Street View. Google Street View is actually composed of 91 patch images with 512 * 512 size and their ground truth will be the categorization of the text detection system. The accuracy is measured using Mean Average Precision.

Data set includes 2955 text regions, and the mean and standard deviation of the number of text regions per image are 2.96 and 2.08 respectively.

Fig 4 Input Image from USTV-1k Dataset

Fig 5. Segmentation of Input Image

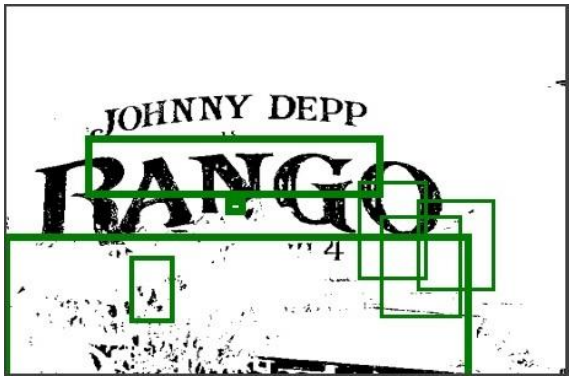

Fig 6. Character candidate Extraction of input image

Fig 7. Candidate Construction of input Image.

Fig 8 Text detection and recognition from input Image.

The infrastructure of the text detection system provides the flexibility for incorporating techniques addressing multi-orientation cases and extend this framework by modifying the text candidates construction stage, and propose a unified multi-orientation text detection system, thereby enabling the effective detection of both multi-orientation and perspective-distortion scene text. Following are stages of multi-orientation scene text detection system:

A) Input Image: The proposed methodology was tried for both horizontal and Non-horizontal picture datasets. The input image is taken from USTSV-1k dataset shown in figure 4.

B) Segmentation: The function of text line segmentation is to convert a region of multiple text lines into multiple sub-regions of single text lines. Text line segmentation and character segmentation algorithms to obtain the precisely bounded characters as shown in figure 5.

C) Character Candidate Extraction: In this step text components usually have significant color contrast with backgrounds and tend to form homogenous color regions. The MSER algorithm that adaptively detects stable color regions provides a viable solution for localizing text . The approach [208] that uses a pruning algorithm to select appropriate MSERs as character candidates and hybrid features to validate the candidates achieved as shown in figure 6.

D) Character Candidate Construction: Text candidates are constructed by three sequential coarse-to-fine grouping steps with adaptive clustering and several parameters are learned by distance metric learning (see Fig. 7): morphology clustering, orientation clustering and projection clustering. Morphology clustering first coarsely groups character candidates with similar appearance together then orientation clustering refines to group character pairs with consistent orientation together; finally projection clustering finely separates text lines in the same orientation as detected text shown in figure 8.

4.1 Performance Metrics

Accuracy is measured in terms of average precision at different recall values.

It is given by formula

$$\text{precision} = \frac{\sum_{j=1}^{|D|} Matc\,h_D(D_j)}{|D|},$$

$$\text{recall} = \frac{\sum_{i=1}^{|G|} Matc\,h_G(G)}{|G|}$$

The harmonic mean of p and r is denoted as standard F-measure *f* .It is defined as the following

$$f = 2\,\frac{Recall\ .\ precision}{Recall\ *\ Precision}$$

where G is the set of groundtruth rectangles.

D is the set of detected rectangles.

Table 1.Comparision Table

Performance Parameter	Precision	Recall	Harmonic Mean
Current System	0.81	0.63	0.71
Proposed System	0.89	0.68	0.79

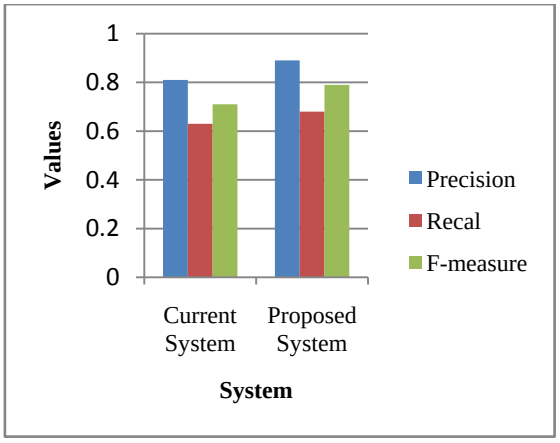

Fig.2.Comparative results between proposed method and Existing Method

4.2 Mathematical Model

This system takes an image from a set of N images as input, where N is number of images in database. Set of operations are then performed on those images. The output of previous function will act as an input to the next function.

The proposed system S is defined as follows

S = {I, F, O}

I denote input given to system. It is defined by,

I = {I1, I2, I3, I4, I5}

I1 = Training Image pairs dataset. It consists of multi-orientation natural scene text data.

I2 = 360 angle-view street view Patches Images.

I3 = Dataset with Horizontal & Nearer Horizontal Orientation Images.

I4 = Text Candidates.

I5 = Partial Candidate Image Patches.

O denotes output of the system. It is defined by,

O = {O1, O2, O3, O4, O5}

O1 = Gray scale image from original Input Image.

O2 = Segmentation image of original Input Image.

O3 = Text character pair by Adaptive Clustering.

O4 = Text character pair extraction and construction by MSER of original Input Image.

O5 = Final Detected Text.

F denotes set of function. It is defined by,

F = {F1, F2, F3, F4, F5}

F1 = It is function of preprocessing an image converting into Gray Scale. Training Dataset having Natural scene images Multi-orientation text is provided as input to this function..

F2 = It is function of segmentation i.e. partition in an image with set of pixels.

F3 = It is function of Adaptive clustering and distance metric learning for calculating proximity data of dataset.

F4 = It is function of text extraction and construction using MSER technique from input image.

F5 = It is a function of detecting final text candidate which is calculated from adaptive clustering and MSER Technique is given as input for getting detection result in form of text.

5. CONCLUSION

This paper presents a new MSER-based scene text detection method with several novel techniques. First, a fast and accurate MSERs pruning algorithm that enables us to detect most characters even when the image is in low quality is proposed. Second, a novel self-training distance metric learning algorithm that can learn distance weights and clustering threshold simultaneously text candidates are constructed by clustering character candidates by the single-link algorithm using the learned parameters is proposed. Third, a character classifier to estimate the posterior probability of text candidate corresponding to non-text and eliminate text candidates with high nontext probability, which helps to build a more powerful text classifier, is proposed. Finally, by integrating the above new technique, a robust scene text detection system that exhibits superior performance over state-of-the-art methods on a variety of public datasets is proposed.

5.1 Future Work

Text Detection can be robustly automated and one can improve the accuracy of detected text in Multi-orientation scenarios in low resolution and blurry images and cursive text in scene images.

6. REFERENCES

[1] Xu-Cheng Yin, Wei-Yi Pei , Jun Zhang and Hong-Wei Hao, “Multi- Orientation Scene Text Detection With Adaptive Clustering”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Sept 2015.

[2] Q. Ye and D. Doermann, “Text detection and recognition in imagery: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., 2014. DOI: 10.1109.

[3] Y.-F. Pan, X. Hou, and C.-L. Liu, “A hybrid approach to detect and localize texts in natural scene images,” IEEE Trans. Image Process., vol. 20, no. 3, pp. 800–813, Mar. 2011

[4] X. Chen and A. Yuille, “Detecting and reading text in natural scenes,” in Proc. Int. Conf. Comput. Vis. Pattern Recognit., 2004, pp. 366–373.

[5] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, “Adaboost for text detection in natural scene,” in Proc. Int. Conf. Document Anal. Recognit.2011, pp. 429 –434.

[6] K. Kim, K. Jung, and J. Kim, “Texture-based approach for text detection in images using svm vector machines,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[7] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. Int. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2963–2970.

[8] C. Yi and Y. Tian, “Localizing text in scene images by boundary clustering, stroke segmentation” IEEE Trans. Image Process., vol. 21, no. 9, pp. 4256–4268, Sep. 2012.

[9]X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, “Robust text detection in natural scene images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, pp. 970– 983, May 2014.

[10] Anhar Risnumawan, Palaiahankote Shivakumara, Chee Seng Chan and Chew Lim Tan, "A Robust Arbitrary Text Detection System For Natural Scene Images 41(2014) 8027-8048.

[11] X. Yin, X.-C. Yin, H.-W. Hao, and K. Iqbal, "Effective text localization in natural scene images with MSER, and AdaBoost," in Proc. Int. Conf. Pattern Recognit., 2012, pp. 725–728.

[12] Chucai Yi and Yingli Tian, K. Huang, and H.-W. Hao, "Accurate and robust text detection: text retrieval in natural scene images," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2013

[13] L. Neumann and J. Matas, "Scene text localization and recognition with oriented stroke detection," in Proc. Int. Conf. Comput. Vis., 2013.

[14] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," Pattern Recognit., 2012, pp. 1083–1090.

[15] C. Yao, X. Bai, and W. Liu, "A unified framework for multi-oriented text detection and recognition," IEEE Image Process., vol. 23, no. 11, 4737–4749, Nov. 2014.

[16] P. Shivakumara, T. Q. Phan, S. Lu, and C. L. Tan, "Gradient vector flow and grouping-based method for arbitrarily oriented scene text detection in video images," IEEE Trans.vol. 23, no. 10, pp. 1729–1739, Oct. 2013.

[19] Vaibhav G Baviskar and Prof Jyoti R Mankar, "Text Detection for Multi-Orientation Scene Images using Adaptive Clustering", cPGCON 2016, Fifth Post Graduate Conference of Computer Engineering, March 2016.