International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Comparative Analysis of Text Detection using SVM, KNN and NN

Saurabh Gupta[1] Dr. Sunil Nijhawan[2]

[1]M.Tech Scholar, G.I.E.T, Sonipat

[2]Associate Professor, G.I.E.T, Sonipat

Abstract: ***Text objects occurring in image can provide much useful information for content based information retrieval and counting applications, because they contain much minute information related to the documents contents. However, extracting text from images and videos is a very difficult task due to the varying font, size, color, orientation, and malformation of text objects. Although a large number of text extraction approaches have been reported in the past work, no specific designed text model and character features are presented to capture the unique properties and structure of characters and text objects. We have proposed SVM, KNN and NN Techniques in our research. The thesis was focused on the design of an algorithm to detect the text from images, extract it and convert it into speech for those who are unable to read and its implementation on MATLAB.***

Keywords: Text Extraction, Text Localization, Text Recognition, SVM, KNN, NN, Stroke Width, Text to Speech

1. Introduction

The text elements embedded in born-digital images, prevalenton the Web, carry salient semantic information such as advertisements and security-related information. Therefore, extracting text information from born-digital images enhances the semantic relevance of web content for indexing and retrieval. Usually, a text information extraction system consists of three steps: text localization, text segmentation & text recognition. Text localization is critical to the overall system performance and is suffering from variable image background, text color and layout.

Images are the most convenient means of conveying or transmitting information as they quickly convey information about positions, sizes and inter-relationships between the objects. They depict spatial information such that it can be recognized as an object. Human beings are very good at deriving information from the images, because of our instinctive visual and mental abilities. About 75% of the information received by humans is in the pictorial form. The analysis of pictures that employ an overhead perspective, including the radiation invisible to human eye are considered.

Text detection and text extraction in natural scene images is challenging because of the varied conditions under which the image is taken. Natural scene text understanding aiming at extracting text from daily images is the main concern of this text. The text from the image can also be converted into speech with the help of Microsoft SAPI.

Text-to-speech (TTS) convention transforms linguistic information stored as data or text into speech. It is widely used in audio reading devices for blind people now a day. In the last few years however, the use of text-to-speech conversion technology has grown far beyond the disabled community to become a major adjunct to the rapidly growing use of digital voice storage for voice mail and voice response systems. Also developments in Speech synthesis technology for various languages have already taken place.

In this paper, a methodology is proposed for text region detection and character extraction in natural scene images with complex backgrounds. The proposed hybrid approach involves six steps. First, potential text regions in an image are extracted based on connected component features using Stroke width Transform. In the second step, potential text regions are tested for text content or non-text using SVM classifier, KNN classifier and NN classifier. In the third step, detection of localized text regions is done and same features as training period is extracted. In the fourth step, each character of the image is detected using above classifiers. In the fifth step, characters are merged using bounding box and extracted using OCR. In the last step, the extracted text is converted into speech for the ease of people to recognize the text in the image.

Many methods have been proposed for text localization in images, and they roughly fall into two categories: connected component (CC)-based and texture-based.CC-based methods have achieved state-of-the-art results on several public datasets [2]. Extracting candidate text CCs is the most critical step for those methods. On one hand, as many text CCs as possible should be recalled. On the other hand, fewer non-text CCs should be generated so that text CCs are more easily identified and grouped into text lines. Researchers frequently use color, stroke width transform (SWT) and maximally stable extremal regions (MSERs) to cluster pixels into CCs. Color clustering based methods [3,4] adopt strategies such as k-means to segment an image in the hope that pixels belonging to the same text CCs are segmented into the same sub-image. As expected, deciding cluster number is the main difficulty. SWT based methods [5, 6] rely on the existence of two edge pixels with roughly opposite gradient directions in seeking strokes, and merge strokes with about the same stroke width into CCs. Those methods are sensitive to the defects of the edge images. MSERs based methods assume text CCs correspond to MSERs. To reduce the missing of text CCs, MSERs based methods [7–9] generate tremendous non-text CCs, including many ambiguous ones. Besides, as [10] pointed out, some text elements correspond to extremal regions (ERs) instead

of MSERs. The proposed method is CC-based with SWT. Unlike that many methods have been proposed for scene text detection, few works have been published specifically for born-digital images. it is not necessarily true that methods developed for scene images are appropriate for born-digital images. Text strokes in born-digital images usually have complete contours and pixels on the contours have high divergence compared with the adjacent non-text pixels. This is often not true for text in scene images due to non-ideal camera-capturing environment.

The rest of the paper is organized as follows. Section II outlines literature survey. Proposed Algorithm to analyzed F- Score is discussed in Section III. The Section IV is concentrated on the simulated result of The conclusions are given in Section V.

2. Literature Review

This section will provide the brief description and highlights the contribution, remarks and factors of the work done by the researchers.

Xiaoming Huang, Tao Shen, Run Wang, Chenqiang Gao [1] split the recognition problem into detection and recognition procedure. Firstly, in the detection stage, in order to extract potential text as much as possible, we use MSER and color clustering to extract connected component. Then, for the obtained candidate connected component, we use visual saliency and some prior information to filter non-text regions. Finally, we can obtain word image by text line generation. In the recognition stage, we use vertical projection to segment word images, then recognize character in SVM based framework. The experiment results evaluated on standard dataset show that with a small amount of prior information and simple segment strategy, the proposed method has a better performance compared to conventional text detection and recognition method.

Kai Chen [2] proposes a new CC based method for text localization in born-digital images. The proposed method generates character candidates effectively by first detecting text contours and stroke interior regions separately and then combining them. The CCs undergo CC filtering, line grouping and line classification to give the final result. Their method has achieved state-of-the-art performance on the born-digital dataset of ICDAR2013 Competition, convincingly demonstrating the effectiveness.

Yuanyuan Feng [3] presents a text detection method based on Extremal Regions (ERs) and Corner-HOG feature. Local Histogram of Oriented Gradient extracted through corners is used to effectively prune the non-text components. Experimental results show that the Corner Histogram oriented Gradient based pruning method can discard an average of 73.06% of all ERs in an image while preserving a recall of 80.51% of the text components.

Kampong Woraratpanya [4] tells Thai Text localization and extraction in natural scenes is still a major challenge in current applications. However, the efficiency of recognition rates depends on text localization, i.e., the higher purity of text-background decomposition leads to the higher accuracy rate of character recognition. In order to achieve this purpose, the text-background decomposition methods, namely adaptive boundary clustering (ABC) and n-point boundary clustering (n-PBC), are proposed to improve a precision of text localization.

Khalid Iqbal [5] tells about text localization .Text localization in scene images is an essential task to analyze the digital image contents. In this work, a Bayesian network scores using K2 algorithm in conjunction with the features based effective text localization method with the help of maximally stable extremal regions (MSERs).

Lukas Neumann [5] gives an end-to-end real-time text localization and recognition method is presented. Its performance is achieved by posing the character detection and localization problem as an efficient sequential selection from the set of Extremal Regions.

Mohammad Shorif Uddin, Madeena Sultana, Tanzila Rahman, and Umme Saym [6] their main objective is to extract text from images .In this method, author discussed a approach for detecting and localization text from scene images based on morphological features. Many researchers have been working on the development to f techniques to extract texts from a scene images.

Nirmala Shivananda and Naga bhushan [7] proposed a hybrid method for separating text from color document images. But this technique can't extract text from difficult graphics.

Pratim Roy JosepL lad'os and Umapada Pal [8] proposed a method for differentiate text from color map based on CC analysis and grouping of characters in a string. These methods can detect the characters connected to graphics and separate them. But some of the characters can't be differentiate through connected component analysis.

The algorithm of Antani and R. Kasturi [9] works well for text separation from mixed text/ graphics image, but it makes an assumption which is not practical that character components in texts are aligned straight and does not touch or overlap with graphics. For improving accuracy they used modified morphological filter and also proposed a clustering methods .Moreover, their method reduces noise in the result ant image. Experimental result confirms the dominance of their approach. With this method F-Score of 92.8% is achieved.

This section has provided the brief review of the work done in past. It also highlighted the factors, contribution and remarks on the achievement.

3. Frame Work for Implementation

The objective of this paper is to develop an automated system that will be able to detect the location of text in the Natural Scene Images and then extract the text from the images. As mentioned above, different methods have different its own advantages and followed by its disadvantages. Region based method lower computational cost but more false positive result. Connected component

based detection method can detect and localize accurately but speed is problem. Learning based methods give more accurate results but difficult to realize and storage is problem. Clustering based methods are faster but computational complexity is the bottleneck. In this paper, Text understanding systems include three main phases: training the classifier, Text detection using classifier and Comparing the result. The overall system is shown in

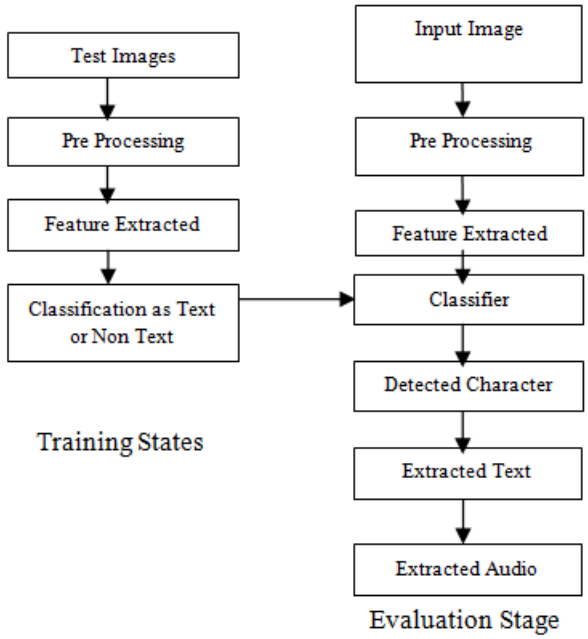

Figure 1: Architecture of Text Detection and Extraction System using Proposed Method

A. Training state:

We have taken 20 different images for training having 2264 Region which are classified as Text or Non text manually.

Figure 2: Training images

A.1 Pre-processing

In order to train the classifier, test image required to be preprocessed. First image is converted into gray scale, then binarization of image is done and later smaller components are removed then edge detection method is applied on the images, then dilation is applied. After that image is segmented into various region using connected components.

Figure 3: Detected Regions

A.2 Feature Extraction

Various features of the segmented region are extracted such as Area, Eccentricity, Euler Number, Orientation, Bounding Box, Centroid, Projection, Moment, Skewness, Kurtosis, Stroke width.

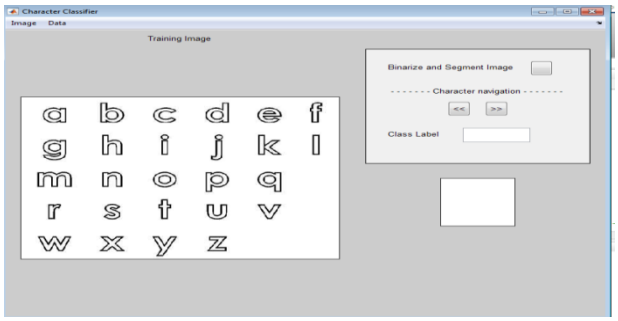

Figure 4: Character Classifier

A.3 Classification as Text or Non Text

The segmented region is shown in the Character Detected Box, then manually we define it as Text or Non Text detected region. For SVM (Support Vector Machine) and KNN (K-Nearest Neighbors), a variable Class is classified into Text and Non Text for all the detected region whereas for NN(Neural Network) two variable are used namely Text and Non text which have binary value 0 or 1 based on its Classification.

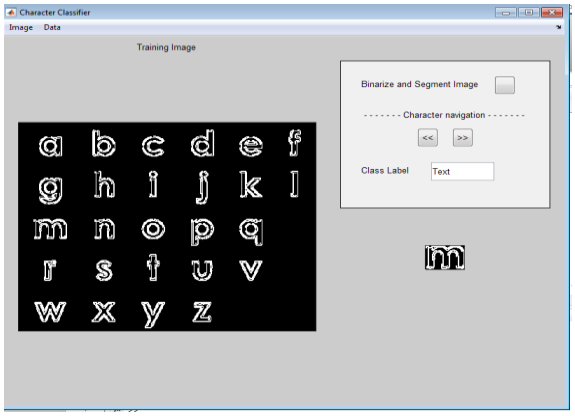

Figure 5: Detected Character and Labeling

A.4 Stroke Width Filter

Before applying it to classifier, Stroke width Transform is applied to the image with a threshold value of 0.4.With the help of SWT, we are able to filter 46 region getting the 2218 samples having 11 Feature for each of them which is saved in a Dataset and applied to classifier.

A.5 Applied to Classifier

A.5.1 SVM

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

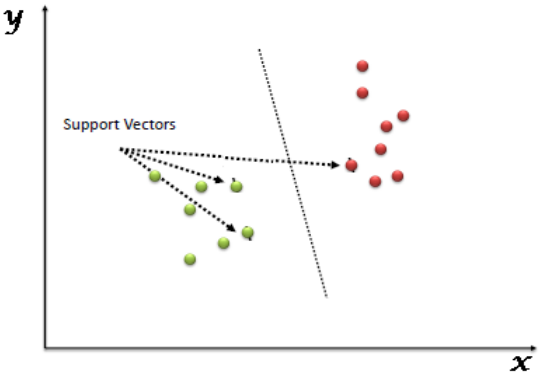

Figure 6: Linear SVM

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).Using Classifier Learner We are able to train SVM for test detection which gives us the best accuracy of 87.2% using Cubic SVM. Then the trained classifier is stored as Variable in the Workspace. Confusion matrix is shown in the figure.

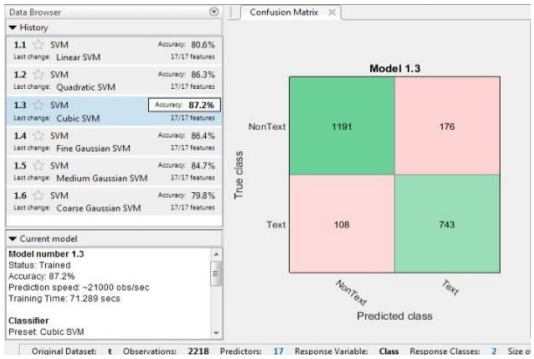

Figure 7: Confusion Matrix for SVM

A.5.2 k-Nearest Neighbors (KNN)

STATISTICA k-Nearest Neighbors (KNN) is a memory-based model defined by a set of objects known as examples (also known as instances) for which the outcome are known (i.e., the examples are labeled). Each example consists of a data case having a set of independent values labeled by a set of dependent outcomes. The independent and dependent variables can be either continuous or categorical. For continuous dependent variables, the task is regression; otherwise it is a classification. Thus, STATISTICA KNN can handle both regression and classification tasks.

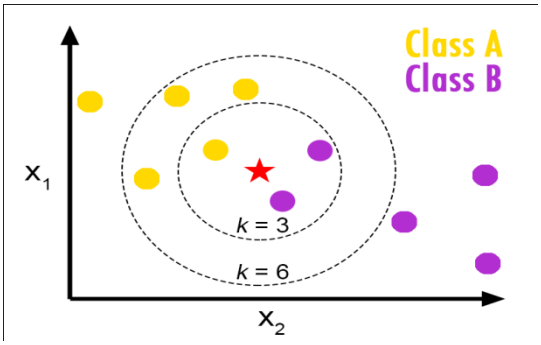

Figure 8: Linear KNN

Using Classifier Learner We are able to train *KNN* for test detection which gives us the best accuracy of 86.8% using Weighted KNN .Then the trained classifier is stored as Variable in the Workspace. Confusion matrix is shown in the figure

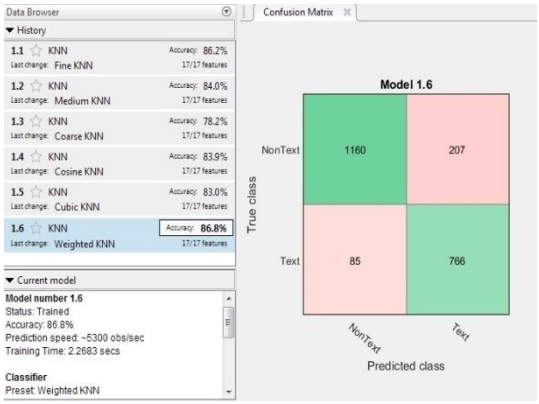

Figure 9: Confusion Matrix for KNN

A.5.3 Neural networks (NN)

Artificial neural networks (ANNs) or connectionist systems are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn (progressively improve performance) to do tasks by considering examples, generally without task-specific programming. They have found most use in applications difficult to express in a traditional computer algorithm using rule-based programming. The original goal of the neural network approach was to solve problems in the same way that a human brain would.

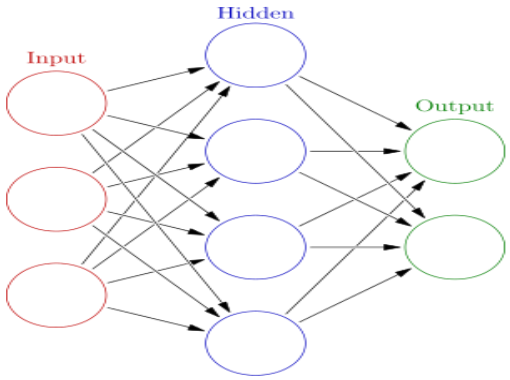

Figure 10: Basic Neural Network

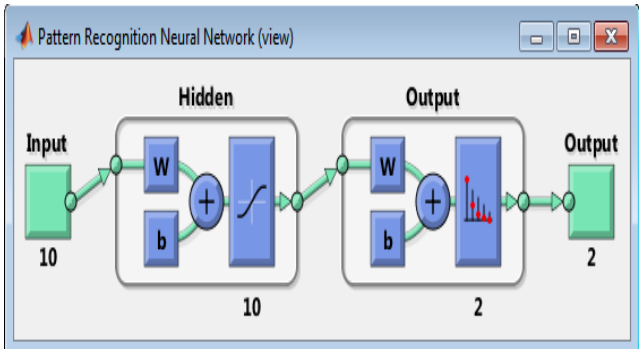

Figure 11: Our Neural Network

Using Neural Network We are able to train *NN* for test detection which gives us the best accuracy of 83.6% using NN .Then the trained classifier is stored as Network in the Workspace. Confusion matrix is shown in the figure 11.

Figure 12: Confusion Matrix for NN

Evaluation Stage

We have taken 10 test images for evaluation of the classifiers .Each image is segmented and then their features are extracted .After that they are passed to the Classifier for Detection. Later the Non-text Candidates are discarded and text candidates are shown in fig

Figure 13: Test Images

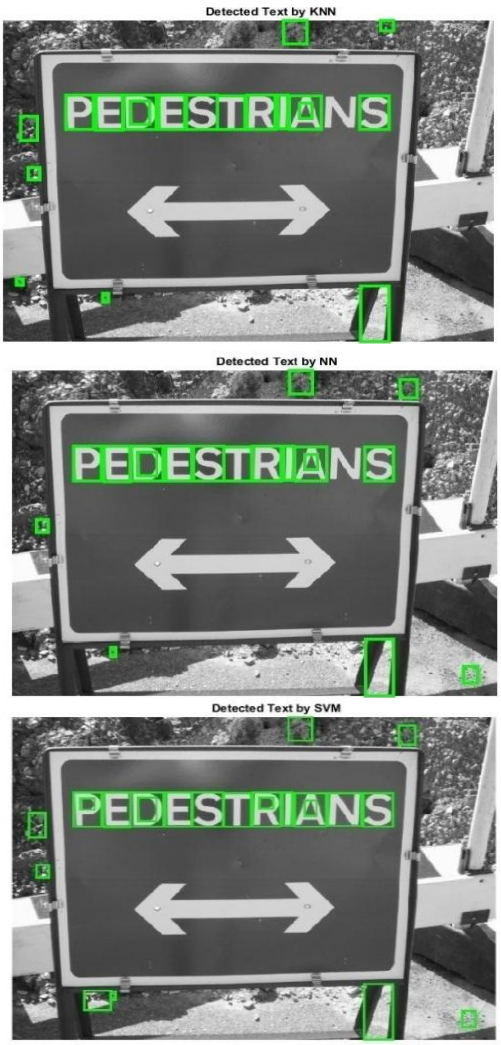

Figure 14: Detected Text by Various Classifier

Later the Detected characters are merged using Bounding Box and extracted by OCR.

Figure 15: Extracted Text By Various Classifier

Later the Extracted Text is converted into speech for the people by using Microsoft Speech API used in Matlab.

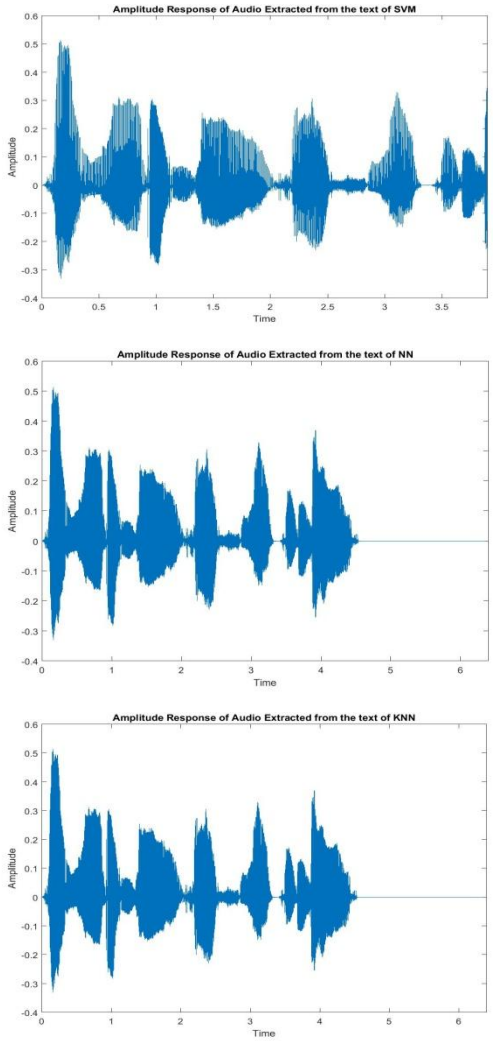

Figure 16: Extracted Audio by Various Classifier

4. Result Analysis

MATLAB platform has been used to evaluate the results. Some assumptions are made to simulate the results as discussed:

It can be done on the basis of following parameters.

1) False Positives (FP) / False alarms are those regions in the image which are actually a text, but have been detected by the algorithm as text.
2) False Negatives (FN)/ these are those regions in the image which are actually texts, but have not been detected by the algorithm.
3) Precision rate (p) is defined as the average ratio of correctly detected characters text without image to the sum of correctly detected characters text plus false positives as represented in equation.
4) p = correctly detected characters/ [Correctly detected characters +FP]
5) RRC (Recall rate) = (No. of extracted characters text in image / No. of characters text in image) x 100.
6) F-score is the H.M of the precision rate and recall rates

Table 1: Comparison of F- Score of different Images using SVM Technique

Images	Precision	Recall	F-Score
Image1	79.4	100	0.88
Image2	100	100	1
Image3	85.7	81.8	.83
Image4	31.2	35.7	.333
Image5	72.7	80	.76
Image6	51.2	83.3	.63
Image7	60	66.6	.63
Image8	47.8	100	.64
Image9	56.8	71.4	.63
Image10	80	100	.888

Table 2: Comparison of F- Score of different Images using NN Technique

Images	Precision	Recall	F-score
Image1	89.2	92.5	0.91
Image2	88.8	80	.88
Image3	88.8	66.6	.76
Image4	55.5	35.7	0.43
Image5	100	80	.88
Image6	45.1	60.8	.517
Image7	75	66.6	.79
Image8	55.5	90.9	.69
Image9	50	71.4	.58
Image10	88.8	100	.94

Table 3: Comparison of F- Score of different Images using KNN Technique

Images	Precision	Recall	F-score
Image1	79.4	100	.88
Image2	100	60	.75
Image3	100	77.2	.87
Image4	35.7	35.7	.35
Image5	80	80	.80
Image6	43.7	63.6	.518
Image7	64.7	61.1	.62
Image8	52.6	90.9	.66
Image9	53.6	68.7	.60
Image10	72.7	100	.84

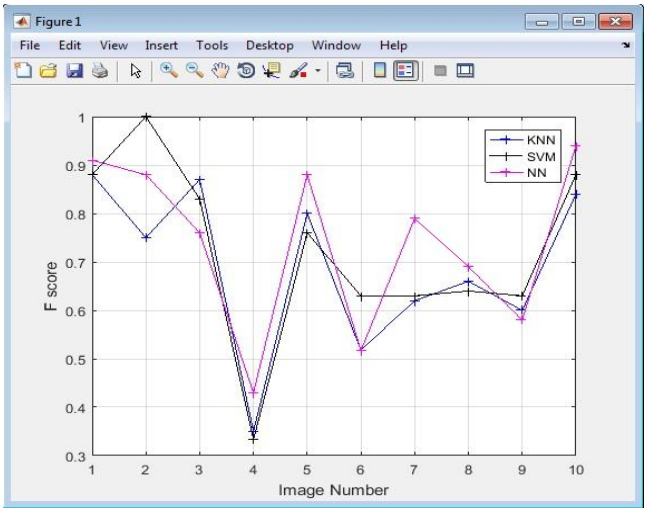

Figure 15: Comparison of F- Score of different Techniques

Table 4: Comparison with the base paper

Method	Hinnerk	Alex	Ashida	HW David	Jiakai	**Proposed Technique**		
						SVM	**NN**	**KNN**
Precision	62	60	55	44	68	66	64	68
Recall	67	60	46	46	65	**86**	**74**	**67**
F-score	0.62	0.6	0.5	0.45	0.66	**0.72**	**0.73**	**0.68**

5. Conclusion

Detection of text in Natural Scene Images is challenging for complex background. There are many methods available to perform the text detection and character extraction in natural scene images. In this work, SVM,NN and KNN classifier is used to detect the text. For training them eleven features such as Area, Eccentricity, Euler Number, Orientation, Bounding Box, Centroid, Projection, Moment, Skewness, Kurtosis, Stroke width. Experimental results demonstrated the effectiveness of the method by locating most text regions in test images. The selected images are pre-processed, features are extracted and classified using this classifier and characters are extracted. Extracted texts are displayed in a window. Extracted Text is converted into speech for ease of people. This method is working well for large size text only. Overall F-score of this work is SVM: 0.72, NN .736, KNN: .68

References

[1] Kai Chen "Effective Candidate Component Extraction for Text Localization in Born-Digital Images by Combining Text Contours and Stroke Interior Regions"*12th IAPR Workshop on Document Analysis Systems (DAS)*, pp 352 – 357, 11-14 April 2016

[2] Yuanyuan Feng "Scene text localization using extremal regions and Corner-HOG feature" *IEEE International Conference on Robotics and Bio mimetics (ROBIO,)*pp881 – 886 , 6-9 Dec. 2015.

[3] Kuntpong Woraratpanya "Text-background decomposition for thai text localization and recognition in natural scenes" *6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 1-6, 7-8 Oct. 2014.

[4] Khalid Iqbal "Bayesian network scores based text localization in scene images", *International Joint Conference on Neural Networks (IJCNN)* 2218 – 2225 , 6-11 July 2014.

[5] Lukas Neumann "Real-Time Lexicon-Free Scene Text Localization and Recognition" *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 38, Issue 9 pp 1872 – 1885 , Oct 2015

[6] Mohammad ShrifUddin, Madeena Sultana, Tanzila Rahman, and Umme Sayma Busra, "Extraction of texts from a scene image using morphology based approach", *IEEE Transactions on image processing*, Vol.10, pp.306-309, 2012.

[7] Nirmala Shivanandaand P.Nagabhushan, "Separation of Foreground Text from Complex Background in Color Document Images", *IEEE Transactions on Image Processing*, vol.10, pp.306-309, (2009).

[8] Partha Pratim Roy, JosepLl ad´osand Umapada Pal," Text/ Graphics Separation in Color Maps", *IEEE Transactions on Image, vol .7*, pp.545-551, (2007). Processing

[9] D.Crandall, S.Antani, and R.Kasturi " Robust Detection of Stylized Text Events in Digital Video", *Proceedings of International Conference on Document Analysis and Recognition*, 2001, pp. 865-869.

[10] R. Chandrasekaran, R.M. Chandrasekaran, P. Natarajan," Text localization and extraction in images using mathematical morphology and SVM*",IEEE International Conference(ICAESM-2012)* March 30-312012.

[11] J. Canny. A Computational Approach to Edge Detection.*IEEE Trans. PAMI*, 8(6):679–698, November 1986.

[12] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod.Robust text detection in natural images with edge-enhanced maximally stable extremal regions.*In Proc. ICIP*, pages 2609–2612, 2011.

[13] A. Clavelli, D. Karatzas, and J. Llados.A Framework for the Assessment of Text extraction Algorithms on Complex Colour Images. *In Proc. 9th DAS*, pages 19–28, 2010.
[14] T. E. de Campos, B. R. Babu, and M. Varma. Character recognition in natural images.*In Proc. VISAPP*, February 2009.

Author Profile

Saurabh Gupta (M.tech Scholar),Gateway Institute of Enginnering and Technology, Sonipat, Haryana, India

Dr. Sunil Nijawan, Doctorate Degree in Electronics and telecommunication from Baba Mast Nath University, Rohtak. Presently he is working as Associate Professor with Gateway institute of Engineering and Technology, Sonipat from the last four years. He has around 19 years of experience in teaching as well as in industry. For the last eight years he is associated with technical education at different levels.