International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-18, May 2019 E-ISSN: 2347-2693

Matrix Method for Distinction between Text and Non-Text Images

P. Karmakar[1*], C. Md. Mizan[2], S. Jana[3], S. Dasgupta[4], S. Paul[5], R. Das[6], S. Das[7]

[1,2,3,4,5,6,7]Dept. of Information Technology, Guru Nanak Institute of technology, Kolkata, India

[*]*Corresponding Author: pradipta23karmakar@gmail.com, Tel.: 8902781676*

Available online at: www.ijcseonline.org

Abstract— Recognition of text and non-text images is a major challenge in the field of computer vision so as to efficiently extract the text from that image. The algorithm used for the extraction of the text from the images would have a higher efficiency if it is known beforehand that the image is a text image or a non-text image. However, there are many images such as old manuscripts where the extraction of the text becomes very difficult. In that case, the algorithm for the distinction between the text and non-text becomes very easy for text detection and have high accuracy and fast in detecting the text from the image. This method can also be applied to detect and extract the text from the signboards also. In our approach, we had built a system that takes any sort of image as an input. After the input of the image, it is then processed and converted into a binary image. Distance transform method is then applied and the measure of the distance between the various points in the image are then calculated. From the calculated points, duplicate points are merged into one point and are sorted in ascending order. The total area of the binary image is then calculated and also the image corresponding to each of the distance transform points are then calculated. The total area of the binary image is then divided by each of the area value of the corresponding distance transform points are the value extracted is known as the feature values. After getting all the feature values the whole value is then divided into small intervals and is then processed through the classifier. The accuracy of the classifier is then calculated and evaluated for the distinction between text and non-text images. This method is a very simple and accurate method for the distinction between the text and the non-text images and also helps in the extraction of the text from the image. Experiment have been done with simple text and non-text image dataset and the efficiency of the proposed method is then demonstrated.

Keywords—text recognition, distance transform, classifier

I. INTRODUCTION

The text is an important part of the image. In some cases, it is clearly understandable, while in others it becomes very difficult to extract the text from that image. The work for the extraction of the text becomes more difficult if it is the case with the video. A video is nothing but a collection of image frames together which are displayed at a regular interval of time so as to create the persistence of vision in the human mind. As a result, the work becomes tedious for extraction of the text from the video frames. So it would much easier for one if they are able to distinguish between an image containing text and an image that does not contain any text before the extraction of the text from that image. This process makes it less complex as it knows beforehand that the image contains any sort of text or not.

We know that there are many functions and algorithms which can detect a text from an image. But sometimes it happens that there efficiency is not as good as it is expected to be. This is because some structures in natural images seem to represent a text. The algorithm tries to extract that text from that image. But unfortunately, it is unable to find any sort of text and at last, after a long time, it fails to extract any text. As a result, it takes a lot amount of time for the extraction of the text from that image which actually does not contain any sort of text. Also in some old manuscripts of ancient times the ancient symbols sometimes represent a text but later it came to be known that it does not represent any text.

In our approach we present a new solution which is very effective and also very accurate in determining the type of image i.e., it is a text or non-text. This solution would help in improving the accuracy for extraction of the text from the image. In our approach, we have obtained a graph for both text-based and non-text based image and it has been observed that both the image represented a different graph. For a set of a text-based image the data is being collected and fetched into a classifier and the average accuracy have been determined. The same test was performed for the non-text image also. On the basis of the accuracy of the classifier, the image has been distinguished as the text image or as a non-text image. In the following sections, we have

reviewed the old work, propose the new classification scheme and discuss the results that are obtained.

II. REVIEW WORK

There are many methods which have been proposed both for the purpose of extraction of the text from images and distinction between the text-based image and non-text based image. However, among them we present a few studies as follows:-

1. In the year 2016, authors Najwa-Maria Chidiac, Pascal Damien and Charles Yaacoub presented a paper in which they have applied the technique of applying both the method of MSER and OCR for the extraction of the text from the image.
2. In the year 2015, authors Radhika Patel and Suman K. Mitra presented a paper in which they have used the intensity values for separation of the text and the non-text area and then recognize the text from the image. The algorithm produces the best results on the Gujarati degraded document images.
3. In the year 1999, authors R. Malik and SeongAh Chin presented a paper in which they they have used an algorithm which is used to extract the text blocks from the whole image and then extract the text from that image. Besides this the stroke width factor is also used for extraction of the text from the image.
4. In the year 2017, authors Sezer Karaoglu, Ran Tao, Theo Gevers and Arnold W M Smoulders, presented a paper in which they assigned the images to different classes and then extract the text according to the scene and is the most successful method in both the fields of classification of the image and text retrieval.
5. In the year 2015, authors Chengquan Zhang, Cong Yao, Baoguang Shi, and Xiang Bai presented a paper in which they have combined the three techniques MSER, CNN and BOW and the experiment have been performed in natural images in many varieties of scenario.

III. PROPOSED METHOD

In our approach, we have built a system which can take the input of both images and video frames of many different varieties. The output of the system is that it is able to distinguish between a text-based image and a non-text based image. The steps of our proposed method are as follows:-

1. At first, the image or the video frame which is intended for the distinction between text and non-text is taken as input into the system.
2. The image is then converted into a black and white image which makes it easier for the extraction of the feature values from the image as the intensity levels become very much clear.
3. After the conversion into the black and white image, the distance of the separation of the points in the image is then calculated which is also known as the Euclidean Distance Transform.
4. After the calculation of the distance transform values, the values without repetition are selected and are sorted in ascending order.
5. In the next step, two image areas are calculated. At first, the total area of the binary image was calculated and then the area formed by each of the corresponding distance transform points was calculated one by one. After the computation of all the area, the area constituted by the distance transform points were divided by the total area of the binary image and were termed as feature values.
6. After the extraction of the feature values, the values are then divided into various intervals such as 0.0000-0.0999, 0.1000-0.1999 and so on. And a bar chart has been plotted for the maximum values obtained from the specified intervals. The bar chart exhibits that the text-based image has a parabolic curve while the non-text based image has a non-parabolic curve. From here only we can somewhat infer about the type of image. For more interpretation, the analysis was further conducted.
7. After the wrenching of the highest values among the feature values that fit into the corresponding periods, the value is then saved in an excel sheet in order to form a dataset. There were two datasets formed one for an image wholly containing text and another which is wholly a non-text based image.
8. After the formation of the dataset, the whole dataset was fetched into a classifier and trained and then the execution of the classifier was then judged and the consequences were explained.

IV. RESULTS AND DISCUSSION

The analysis was implemented using purely text and non-text based image set. A set of images was fed into the system and the bar chart, which was obtained from the extracted feature values of the images were examined and analyzed. The analysis was further extended to study the average accuracy of both the images for the distinction between the text and non-text based images. Two classifiers Ensemble and SVM were employed to study the matter of the subject so as to draw a clear understanding of the matter. The result from the images obtained are as follows:-

(i). Text- based image

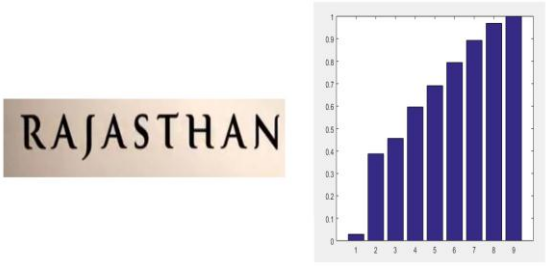

Fig 1:- Bar chart of text based image

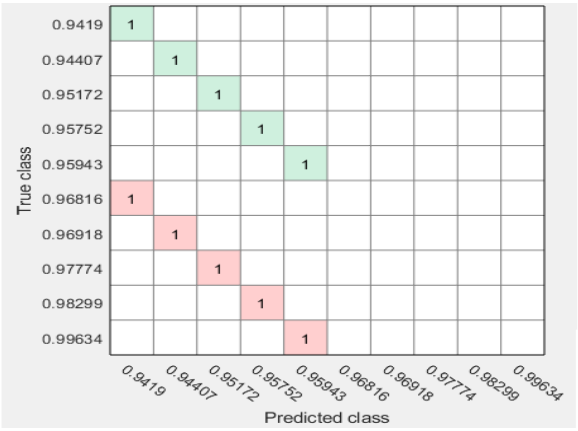

Fig 2:- Confusion matrix of text image in SVM classifier

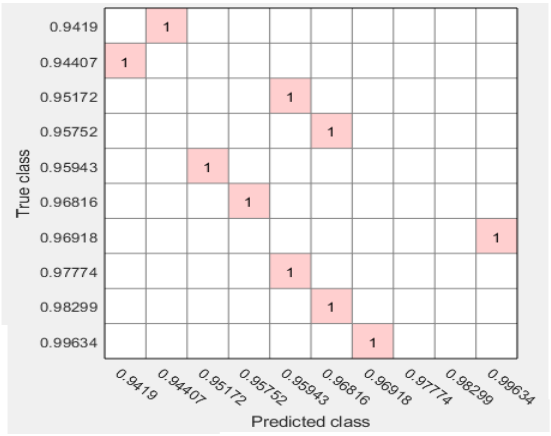

Fig 3:- Confusion matrix of text image in ensemble classifier

From the results obtained from the text based image, we can infer that the pattern of the curve obtained by joining the top points of a bar chart is a parabolic curve. It gradually increases from zero to the highest level. When this data is fed into a classifier, it has been observed that for an ensemble classifier the mean average accuracy is about 50% and for SVM classifier the mean average accuracy is about 0%.

(ii). Non-text based image

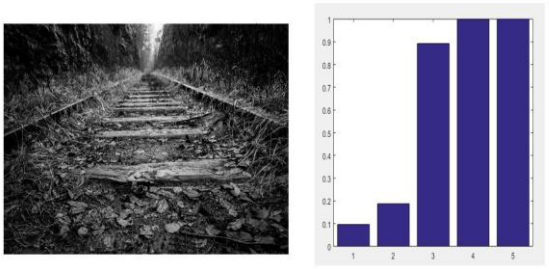

Fig 4:- Bar chart of Non-text image

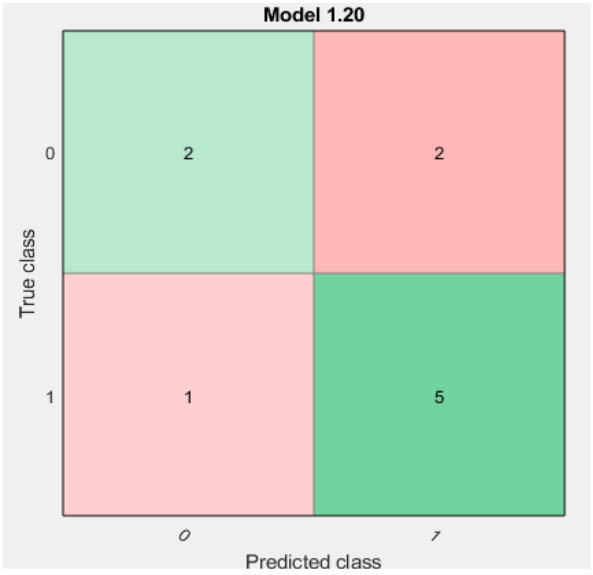

Fig 5:- Ensemble classifier of a non-text image

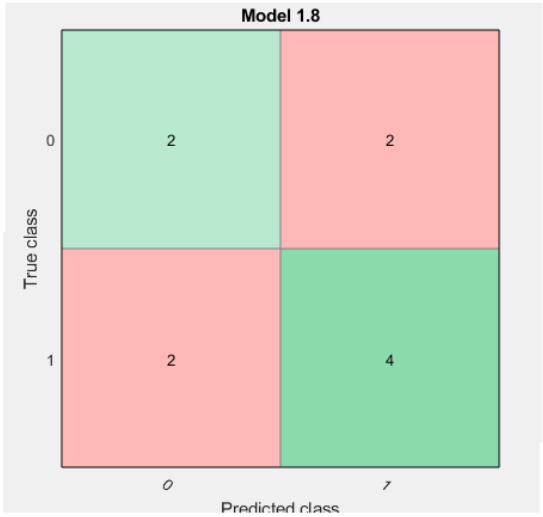

Fig 6:- SVM classifier of a Non-Text image

On the other hand, the results obtained from a non-text image, we can infer that the pattern of the curve obtained by joining the top points of a bar chart is a non-parabolic curve. It starts from zero and rises irregularly to the highest point. When this data is fed into a classifier, it has been observed that for an ensemble classifier the mean average accuracy is about 70% and for SVM classifier the mean average accuracy is about 60%.

In this study, we have presented the approach of distinguishing between the text and the non-text image using the both the ensemble and SVM classifier. This proposal is totally based on extraction of the feature value from the images and then fetching this value into a classifier. After the training of the classifier, the accuracy of the classifier is

then observed and studied. MATLAB has been used to develop the software to improve the efficiency of the method.

REFERENCES

[1]. Najwa Maria Chidiac, Pascal Damein and Charles Yacoub, "A robust algorithm for text extraction from images", 39th International conference on Telecommunication and Signal Processing, 2016.

[2]. Radhika Patel and Suman K Mitra, "Extracting text from degraded documents", 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2015.

years of teaching experience and 4 years of Research Experience.

[3]. R. Malik and SeongAh chin, "Extraction of text in images", Proceedings of International Conference on Information Intelligence and Systems, 1999.

[4]. Sezer Karaoglu, Ran Tao, Theo Gevers and Arnold W. M. Smeulders, "Words matter: Scene Text for Image Clssification and Retrieval", IEEE transactions on multimedia, vol. 19, no. 5, may 2017.

[5]. Chengquan Zhang, Cong Yao, Baoguang Shi and Xiang Bai, "Automatic discrimination of text and non-text natural images", 13th International Conference on Document Analysis and Recognition, 2015.

