Scene Text Recognition in Mobile Application using K-Mean Clustering and Support Vector Machine

Priyanka N Guttedar, Pushpalata S

***Abstract*— In natural scene image, text characters and strings can provide valuable information for many applications. Extracting text is challenging task in natural scene image because of diverse text patterns and variant background interferences. This proposes a method of scene text recognition from detected text regions. In scene text detection, first we perform layout analysis of color conversion and K-mean clustering respectively to search for image regions of text strings. In scene text recognition, Support Vector Machine is proposed to extract text from surrounding environment. Our algorithm design is compatible with smart mobile devices. An Android based demo system is developed in scene text information extraction to show the effectiveness of our proposed method. The demo system also provides some insight into algorithm design and performance improvement of scene text extraction.**

***Keywords*— Character Segmentation, Clustering, Feature Extraction, Pre-processing, Scene Text Detection, Scene Text Recognition, Support Vector Machine.**

I. Introduction

With the wide use of smart phones and rapid development of the mobile internet, it has become a living style for people to capture information by using of cameras embedded in mobile terminals. Camera-based text information serves as effective tags or clues for many mobile applications associated with media analysis, content retrieval, scene understanding, and assistant navigation. In natural scene images and videos, text characters and string usually appear in nearby sign boards and hand-held objects provide significant knowledge of surrounding environment and objects.

Text-based tags are much more applicable than barcode or quick response code because latter techniques contain limited information and require pre-installed marks. Consequently, extracting text information from natural scene by mobile devices, automatic and efficient scene text detection and recognition algorithms are essential and it has become one of the hottest topics in the area of document analysis and recognition. However extracting scene text is a challenging task due to main factors:

- Cluttered backgrounds with noise and non-text outliers and
- Diverse text patterns such as character types, fonts, sizes.

The frequency of occurrence of text is very low in natural scene, a limited number of text characters are embedded into complex non-text background outliers. Background textures such as grid, window, and brick, even resemble text characters and strings. Although these challenging factors exit in face and car, many state-of-the-art algorithms have demonstrated effectiveness on those applications, because face and car have relatively stable features. For example, a frontal face normally contains a mouth, a nose, two eyes, and two brows as prior knowledge. However it is difficult to model the structure of text characters in scene images due to the lack of discriminative pixel-level appearance and structure features from non-text background outliers. Usually text may consist of different words where each word may contain different characters in various fonts, styles, and sizes, resulting in large intra-variations of text patterns. To solve these challenging problems, scene text extraction is divided into two main processes: text detection and text recognition. Text detection is to localize image regions containing characters and strings. It aims to remove most non-text background outliers. Text recognition is to transform pixel-based text into readable code. It aims to accurately distinguish different text characters and properly compose text words. This will focus on text recognition method. It involves 62 identity categories of text characters, including 10 digits [0-9] and 26 English letters in upper case [A-Z] and lower case [a-z].

II. EXISTING SYSTEM

In this section, we present a general review of previous work on scene text recognition respectively. While text detection aims to localize text regions in images by filtering out non-text outliers from cluttered background. Text recognition is to transform image-based text information in the detected regions into readable text codes.

Scene text recognition is still an open topic to be addressed. In the Robust Reading Competition of International Conference on Document Analysis and Recognition (ICDAR) 2011, the best word recognition rate for scene images was only about 41.2%.

[1] Neumann proposed a real time scene text localization and recognition method based on external regions. Text localization and recognition in real-word (scene) images is an open problem which has been receiving a significant attention since it is a critical component in a number of computer vision applications like searching images by their textual component, reading labels on business in map applications or assisting visually impaired.

[2] J. J. Weinman proposed scene text recognition using similarity and a lexicon with sparse belief propagation. The problem of optical character recognition (OCR), or the recognition of text in machine-printed documents, has a long history and is one of the most successful applications of computer vision, image processing, and machine learning techniques. In this, we focus on scene text

Manuscript received June, 2015.

Priyanka N Guttedar, Computer Science And Engineering from Godutai Engineering College for Women, Gulbarga, Karnataka, India, Mobile no-8951414275.

Asst.prof. Pushpalata.S, *Computer Science and Engineering* working in Godutai Engineering College for Women, Gulbarga, Karnataka, India.

recognition (STR), the recognition of text anywhere in the environment, such as on store fronts, traffic signs, movie marquees, or parade banners.

[3] C. L. Tan proposed document image retrieval through word shape coding. This presents a document retrieval technique that is capable of searching document images without OCR (optical character recognition). The proposed technique retrieves document images by a new word shape coding scheme, which captures the document content annotating each word image by a word shape code.

[4] T. De Campos proposed character recognition in natural images. This tackles the problem of recognizing characters in images of natural scenes. In particular, we focus on recognizing characters in situations that would traditionally not be handled well by OCR techniques.

[5] B. Epshtein proposed detecting text in natural scenes with stroke width transform. Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of computer vision applications, such as computerized aid for visually impaired, automatic geo-coding of business, and robotic navigation in urban environments.

[6] Smith proposed enforcing similarity constraints with integer programming for better scene text recognition. The scene text recognition task frequently requires interpreting versions of letters and digits that are significantly different than those seen in training.

III. PROPOSED SYSTEM

A. The proposed method includes two main phases as and five modules shown by the figure below.

1) Scene Text Detection- includes Pre-processing, clustering, character segmentation.

2) Scene Text Recognition- includes feature extraction, Support vector machine.

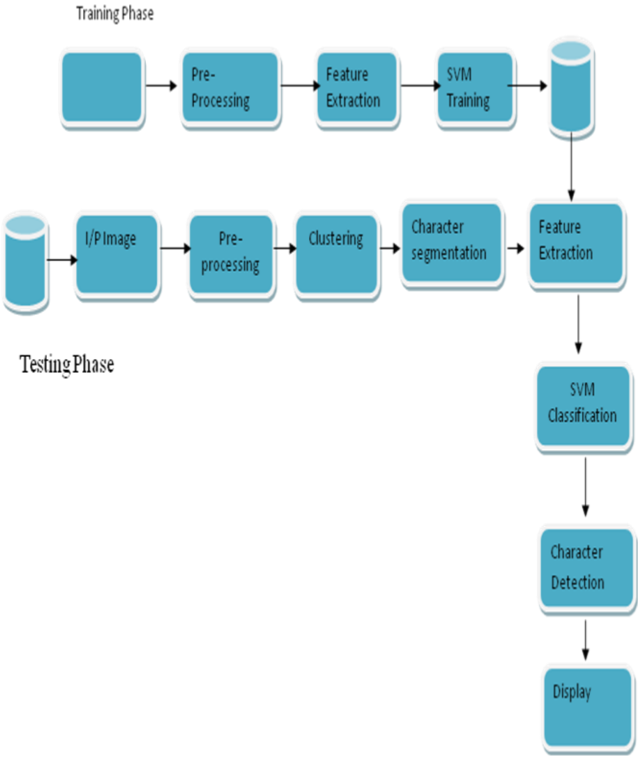


Figure 1: Overall block diagram of proposed method

The first phase is training phase where phases is acquired from user and feature extraction is performed. One thing we should notice here is that recognition of text in scene image is from trained database, which means we trained some of the images in a database.

The second phase is testing phase where the test phases should be classified against the training sample to recognize the text in scene images.

We use Support Vector Machine for the classification as it is based on binary classifier. As the training samples are low and test phases are expected to be significantly different from those training phases, it is mainly depend upon feature matching rather than any prediction.

For extraction of feature we need a scene text image sample. These scene text images contain a some sort of background noise and diverse text pattern. Therefore we include pre-processing step is to normalize the size of all the images by changing their size to some pixels.

A first step in pre-processing is to convert the image to gray levels shown in figure 4. To convert any color image to gray scale, initially obtain the values of red, green, and blue. Then, adding together 30% of the red value, 59% of the green value, and 11% of the blue value. Then the input RGB image is converted to a grey scale image. Therefore, images are captured under natural conditions and which may contain different type of noise. A pre-processing stage is required for smoothing of background texture and contrast enhancement between background and text areas. We applied a low-pass filter to gray scale image.

After converting image to gray-scale image, then we have to find edged image from the gray-scale image. For this we have many different methods like Roberts Edge Detection, Sobel Edge Detection, Prewitt edge detection, Canny Edge Detection. We have used canny operator to find the edged image. The canny operator performs a 2-D spatial gradient measurement on image. The canny method applies two thresholds to the gradient: a high threshold for low edge sensitivity and a low threshold for high edge sensitivity. Edge starts with the low sensitivity result and then grows it to connected edge pixels from the high sensitivity result. This help to fill the gaps in the detected edges. In general the purpose of edge detection is to significantly reduce the amount of data in an image.

A. CLUSTERING

For clustering we have used K-mean clustering technique. The K-mean is basically clustering algorithm which partition a data set into cluster according to some defined distance measure. In this direction of analyzing data within the image, segmentation is the first phase to estimate quantity of the object present in an object. K-mean clustering algorithm is an unsupervised clustering protocol which categorizes the input data points into multiple types based on their inherent distance from each other. The protocol considers that the data features create a vector space and tries to locate normal clustering in them.

CHARACTER SEGMENTATION

Character regions located in images are required to be binarized before recognition. Generally, characters can be distinguished from background with color, thus color clustering is common for binarization. So that characters in natural scene images often suffer from uneven light, reflex and shadow, which likely to make stroke broken and bring about some isolate noise, hence the performance of color clustering is greatly affected. Hence we proposed a character segmentation method.

Here we perform a morphological operation like dilation and noise removal. The non-text regions are removed using morphological operations. Various types of boundaries like vertical, horizontal, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. Dilation is designed to fit user-defined input of text based image with various types of characteristics.

Noise removal can be done by using weighted average Filter. The median filter replaces each pixel in the input image by the median or gray levels in the neighborhood. Thus it leads to smoothing and hence reduces noises. If the number of pixels, K in a window is odd, the median is said to be the (K+1)/2 largest value. The number of comparison needed to find median in this case: N= 3(K2-1)/8

FEATURE EXTRACTION

The feature is defined as function of one or more measurements, each of which specifies some quantifiable property of an object, and is computed such that it quantifies some significant characteristics of the object. Extraction of feature is based on the color, shape, and texture. The various features currently employed as follows:

- General features: Application independent features such as color, texture, and shape. According to the abstraction level, they can be further divided into:
  1. Pixel-level features: Features calculated at each pixel, e.g. color, location.
  2. Local features: Features calculated over the results of subdivision of the image band on image segmentation or edge detection.
  3. Global features: Features calculated over the entire image or just regular sub-area of an image.
- Domain-specific features: Application dependent features such as human faces, fingerprints, and conceptual features. These features are often a synthesis of low-level features for a specific domain.

SUPPORT VECTOR MACHINE

Support Vector Machines (SVM) recently became one of the most popular classification methods. They have been used in a wide variety of applications such as text classification, facial expression recognition, gene analysis and many others. Support vector Machines can be thought of as a method for constructing a special kind of rule called a linear classifier that produces classifiers with theoretical guarantees of good predictive performance (the quality of classification on unseen data). SVM initially developed for classification and it has been extended for regression and preference (or rank) learning.

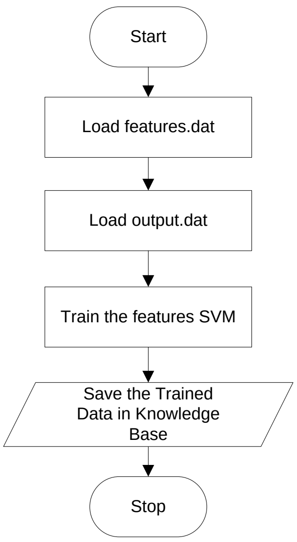


Figure 2: Figure shows training the SVM features

The initial form of SVMs is a binary classifier where the output of learned function is either positive or negative. Binary SVMs are classifiers which discriminate data points of two categories. Each data object or data point is represented by n-dimensional vector. Each of these data points belongs to only one of two classes. A multiclass classification can be implemented by combining multiple binary classifiers using pair-wise coupling method.

B. Algorithms

We mainly use two algorithms for our project: K-mean clustering and SVM. K-Mean clustering is used for representing the features from scene image sample. SVM used to classify the features to classes.

Algorithm 1: K-Mean Clustering

i. The histogram of intensities which highlight estimates of pixels in that specific tone is estimated as shown below

n = $\sum_{i=1}^{k} m_i$

Where, n = total estimates of observations
K = total estimates of tones.

The quantity of the pixels is estimated by the m which has equivalent value. The graph created with the help of this is only the alternative way to represents histogram.

ii. The centroid with K arbitrary as in eq. (2) should be initialized.

iii. The following steps are iterated until the cluster labels of the image do not alters anymore.

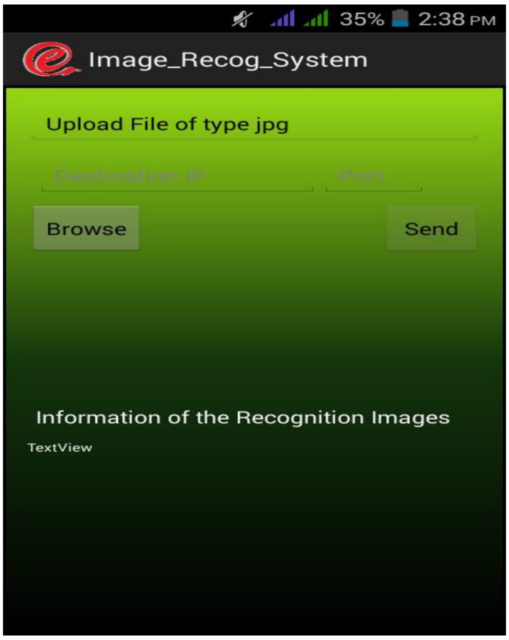
iv. The points based on distance of their intensities from the centroid intensities are clustered.

v. The new centroid for each of the clusters is evaluated.

Algorithm 2: Support Vector Machine

```
        Candidate SV = {closest pair from opposite classes}
        While there are violating points do
        Find a violator
        Candidate SV = candidate SV      violator
        If any    αp< 0 due to addition of c to S then
        Candidate SV = candidate SV \ p
        Repeat till all such points are pruned
        End if
        End while
Result
```

IV. RESULT AND DISCUSSION



35% 2:38 PM

Image_Recog_System

Upload File of type jpg

Destination IP

Port

Browse

Send

Information of the Recognition Images

TextView

Figure 3 shows character recognition in mobile application system

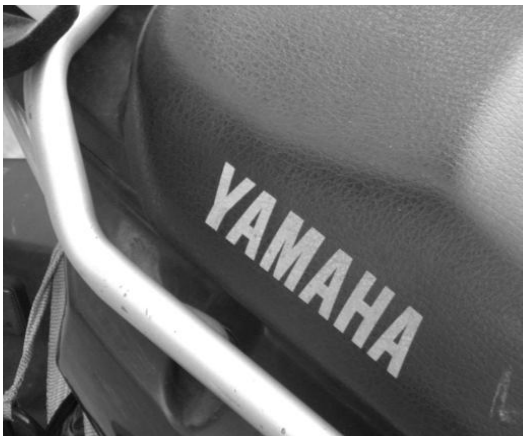


Figure 4: Shows Gray-Scale Image

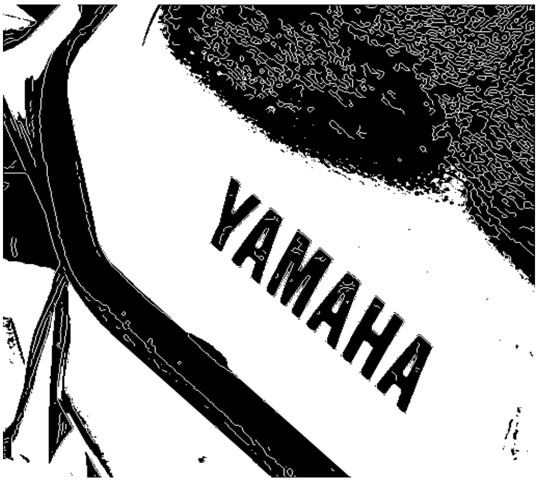


Figure 5: Shows segmented binary image

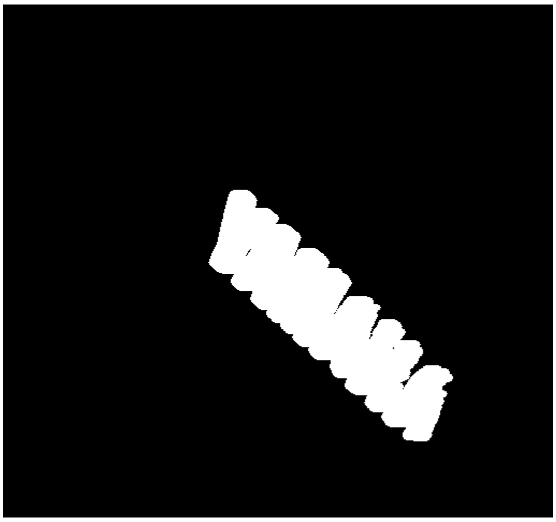


Figure 6: Shows the noise removal using median filter

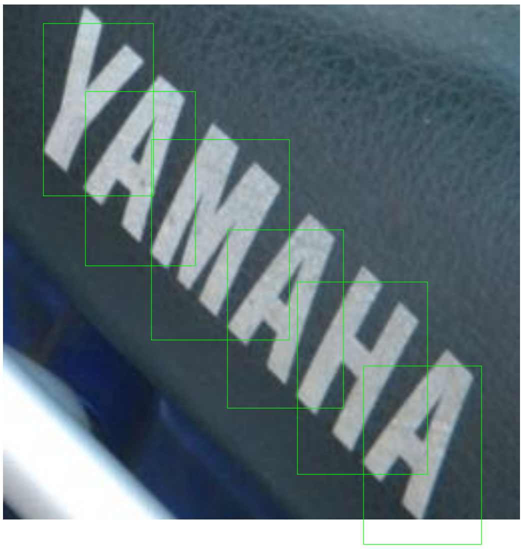
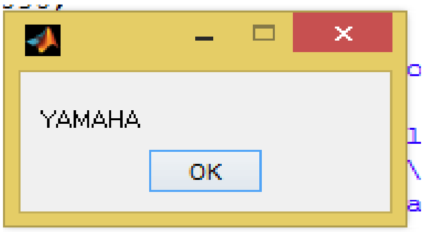


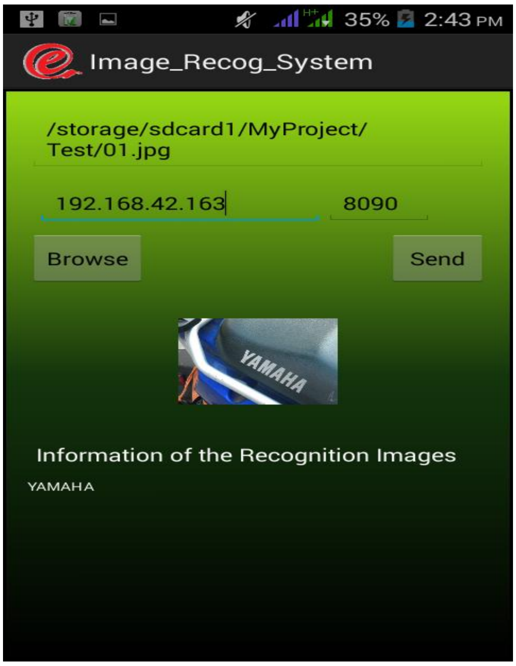
Figure 7: Shows Character Segmentation



YAMAHA

OK

Figure 8: Shows Detected Result Using SVM



Image_Recog_System

/storage/sdcard1/MyProject/Test/01.jpg

192.168.42.163 8090

Information of the Recognition Images

YAMAHA

Figure 9: Shows the confirmation sent to mobile client

V. CONCLUSION

With the rapid growth of camera based application readily available on mobile phones, understanding scene text is very important. By comparing and categorizing different challenging task such as cluttered background and diverse text patterns, we analyzed the advantage and disadvantage of five popular schemes existed in the field, namely Pre-processing, Clustering, Character segmentation, Feature Extraction, and Support Vector Machine. However, existing approaches face the challenging task. In this, we have presented a method of scene text recognition from detected text regions, which is compatible with mobile application. To solve these challenging tasks, scene text extraction is divided into two main processes: text detection and text recognition. . In our application, maximum effort is made to identify the text from the image and to produce more accurate results than the existing systems. In future work, we will improve the accuracy rate of text detection and recognize text in videos also.

REFERENCES

[1] L.Neumann and J. Matas, “Real-time scene text localization and recognition,” in *Proc. IEEE Conf. Computer. Vis. Pattern Recognition.*, Jun. 2012, pp. 3538–3545.

[2] J. J. Weinman, E. Learned-Miller, and A. R. Hanson, “Scene text recognition using similarity and a lexicon with sparse belief propagation,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 31, no. 10, pp. 1733-1746, Oct. 2009.

[3] S. Lu, L. Li, and C. L. Tan, “Document image retrieval through word shape coding,” IEEE Trans. Pattern Anal. Mach. Intel, vol. 30, no. 11, pp. 1913-1918, Nov. 2008.

[4] T. De Campos, B. BABU, and M. Varma, “Character recognition in natural images,” in Proc. VISAPP, 2009.

[5] A. Coates, “text detection and character recognition In scene images with unsupervised feature learning,” in Proc. ICDAR, Sep. 2011, pp. 440-445.

[6] D. L. Smith, J. Field, and E. Learned-Miller, “Enforcing similarity constraints with integer programming for better scene text recognition,” in Proc. IEEE Conf. Computer. Vis. Pattern Recognition, Jun. 2011, pp. 73-80.

[7] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. CVPR, Jun. 2010, pp. 2963-2970.

[8] R.Beaufort and C. Mancas-Thillou, “A weighted finite-state framework for correcting errors in natural scene OCR,” in Proc. 9th Int. Conf. Document Anal. Recognition, Sep. 2007, pp. 889-893.

[9] A.Mishra, K. Alahari, and C. V. Jawahar, “Top-down and bottom-up cues for scene text recognition,” in Proc. IEEE Conf. Computer. Vis. Pattern Recognition, Jun. 2012, pp. 1063-6919.

[10] X. Chen, J. Yang, J. Zhang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” IEEE Trans. Image Process., vol. 13, no. 1, pp. 87-99, jan. 2004.

[11] XiaopeiLiu, Zhaoyang Lu, Jing Li, and Wei Jiang, “Detection and segmentation Text from Natural Scene Images based on Graph Model,” IEEE Trans. Signal Processing. 2014.

[12] Sinitha Beevi, Sajeena, “A Novel Method for Character Segmentation of Vehicle License Plates,” Sep. 2011.

[13] Ryszard S. Chora’s, “Image Feature Extraction Techniques and their Applications for CBIR and Biometrics Systems,” in Image retrieval and Indexing. vol. 1. Sep. 2007.

[14] DmitriyFradkin and IlyaMuchnik, “Support Vector Machines for classification” vol. 8. Jan. 2011

[15] Durgesh K, Srivastava, Lekha Bhambhu, “Data Classification using Support Vector Machine” in 2008.

[16] Hwanjo Yu and Sungchul Kim, “SVM Tutorial: Classification, Regression, and Ranking” in 2009.

[17] K. N. Narasimha Murthy, Dr. Y S Kumaraswamy, “A Novel Method for Efficient Text Extraction from Real Time Images with diversified background using Haar Discrete Wavelet Transform and K-Means Clustering,” vol. 8. Sep. 2011.

Priyanka N Guttedar received the B.E degree in computer science and technology from the Visvesvaraya Technological University, Belgaum, India, in 2013, and is currently pursuing the M.Tech degree with the computer science and technology, from the Visvesvaraya Technological University, Belgaum, India. Her research interests include Image Processing, Scene Text Recognition.

Pushpalata S (Asst.Prof) received her Mtech degree (CSE) from Poojya Doddappa Appa Engineering College, Gulbarga, and Presently working as Asst Prof in Godutai Engineering College,Gulbarga.

.