A Survey on Text Recognition from Natural Scene Images and Videos

Usha Yadav, Nilmani Verma

Abstract*—Text recognition from any natural scenes images and videos is application of image processing technique. Basically text recognition is belongs to the pattern recognition which is part of image processing techniques. Now these days text recognition from natural scene images and videos is very difficult task.For make it easy four basic steps must be apply that approaches are (i) Text image pre processing (ii) character segmentation (iii) character recognition and (iv) Text recognition. In the state of art methods , character segmentation having two major approaches that is Segmentation –based approaches which segment the text into individual character before recognizing and segmentation-free approaches which recognizes character directly from whole text images without any segmentation. Character can also be recognized with two approach that is pattern matching methods in that particular method character are usually identified by a set of features and machine learning methods in which the methods are designed that are learn automatically from the image or after extracting feature. Various method has been applied earlier for extracting text from images and videos. These all methods are trying to provide better result .Various paper use printed text images for recognition their we never required any preprocessing for extracting text . Here is the name of some methods that are used for text recognition that are are Specific directed acyclic graph techniques, scalable feature learning algorithm, k nearest neighbour technique and back propagation algorithm. All those method which has been applying for text recognition , they all provide accuracy in result or we can say that the recognized text are nearly matched with the original one.

Keywords*—Neural based OCR , Character segmentation ,character recognition, Back propagation neural network model, Unsupervised learning.

I. INTRODUCTION

Optical character recognition, usually abbreviated to OCR is the mechanical or electronic representation of images .Which is type of handwritten, typewritten or printed text into machine-editable text. OCR is a field of research in pattern recognition, artificial intelligence and machine learning vision. Though academic research in the field continues, the focus on OCR has shifted to implementation of proven techniques. Optical character recognition (using optical techniques such as mirrors and lenses) and digital character recognition (using scanners and computer algorithms) were originally considered separate fields. Because very few applications survive that use true optical techniques, the OCR term has now been broadened to include digital image processing as well.

Manuscript Received on February 16, 2015.

Er. Usha Yadav received her BE (Computer Scieence) from M P Christian College of Engg. And Technology, Bhilai

Er. Nilmani Verma is working as Assistant Professor in Department of Computer Science & Engineering, School of Engineering & IT, MATS University, Raipur

Early systems required training (the provision of known samples of each character) to read a specific font. "Intelligent" systems with a high degree of recognition accuracy for most fonts are now common. Some systems are even capable of reproducing formatted output that closely approximates the original scanned page including images, columns and other non-textual components. The document image analysis community has shown a huge interest in the problem of scene text understanding in recent years [6] , [20] , [21]. Printed or scanned image are mostly used for text recognition. Text recognition from image is very difficult task because of low resolution, complex background, non uniform lighting, occlusions and blurring effects. All these things make the task of text recognition a challenging problem that has raised a growing interest in recent research activities [3, 4, 5, 10]. For text recognition we can apply two approaches that is text recognition with character segmentation and without character segmentation.

Many research has been done by the researchers in [14] the text detection and recognition in images and video frame is addressed by text segmentation step followed by an traditional OCR algorithm within a multi hypotheses framework relying on multiple segments, language modeling and OCR statistics [3]. In [3] the image text recognition graph (iTRG) used some module that are based on convolution neural network and used ICDAR 2003 dataset[4]. In [15] SMT (Surface Mount Technology) product character recognition has done with the help of back propagation algorithm. For segmentation projection analysis are used. When they apply 20 hidden layer then they found accurate result. In [2] text recognition from natural scene images and videos using natural language processing and unsupervised learning method [5],[9],[10]. In [5] text recognition from images are used convolution neural network technique and multi scale with linguistic knowledge and top down and bottom bottom up approach for recognized text cues[6], [7], [8]. In [1] they represent the scene text image and synthetic images generated from lexicon words using gradient-based features. We then recognize the text in the image by matching the scene and synthetic image features with our novel weighted Dynamic Time Warping (wDTW) approach [1]. Their they used 30 cluster to compute the weights. With using weighted Dynamic Time Warping (wDTW) it is provide high accuracy result. They also used dynamic k-nearest neighbor having an initial value k=3 for their all experiment. All steps provided better result. The remainder of this paper is organized as follows. Section II contain related work of text recognition. Different methodology which are used for text recognition is mention in section III. Section IV shows

the outcome of the different methods. Finally section V conclude the over all text recognition steps.

II. RELATED WORK

A. What is character recognition

Character recognition is divide in two main approaches that is: pattern matching methods and machine learning methods.

In the first category, characters are usually identified by its features. First, a database of models of features is generated. Then for each image corresponding to a character, features are extracted and matched against the database in order to recognize the character class. In [11] edges and contours are considered as features characterizing characters, some cases for each binaries image character, four side profiles are extracted and matched to recognize characters. Side-profiles are obtained by counting the white pixels in each direction (left, right, up and down) until encountering black pixels. We also used a projection profile technique to recognize character. However as in any pattern recognition problem, the major issue is to define the robust features that represent characters independently of the image resolution and the background complexity. Therefore performance of these methods may be very variable depending on the chosen features and the image conditions.

In the second category, methods are designed to learn automatically how to classify characters either directly from their images or after extracting features. In [12] Sadane and Garcia have presented an automatic method for scene character recognition based on a convolutional neural classification approach. The system is able to deal directly with the raw pixels of extremely variable characters and appears to be particularly robust to different image distortions. Another method we can use SVM classifier which learns how to recognize characters from image pixels and which also obtains good results. Voting method was chosen by Kusachi et al. [13] to identify characters with recognition dictionaries obtained by patterns learning. Recently a method based on unsupervised features learning was proposed to detect and recognize characters in natural scene images[2].

III. METHODOLOGY

Here we discuss the methods which are used for text recognition.

A. K Nearest Neighbour

In [1] there we have a scene text that is related to wild and it is a challenging problem and required great attention for recognizing text from it. For it we take a different framework that is holistic word recognition technique. They firstly take scene text images and then generated synthetic images from lexical words with the help of gradient based features. A list of word is available that is called lexical driven word according to that list the ranked list of matched synthetic words found (each corresponding to one of the lexicon words), our goal is to find the text label. Their resize the word image to a width of 300 pixel with the respective aspect ratio. Divide the word image into vertical strip and in each strip we compute histogram of gradient orientation at edges. With the help of overlapping vertical strip features are computed. For providing training two public data sets is used that is street view text (SVT) and ICDAR 2003.

These paper choose an alternative path and use holistic word recognition technique for finding text from images which are taken from wild. The retrieving framework introduced is similar to [17] related the area of handwritten and printed word spotting. Following approach make this paper differ from [17].(1) Their matching score is based on a novel set ,which improve the performance than the profile feature in [17] .(2) They formulate the problem of finding the best word from a maximum likelihood framework and also maximize the probability of that are generated two feature sequence from same word.(3) They propose a robust way for finding the word match that is the value of K in K-NN is decided dynamically from the top retrievals. They starting with generate various font and style of synthetic images for the word from the lexicon. Then gradient based features are computed for all images as well as the test images. Then recognizing the text in the images by matching the scene and synthetic images features with weighted Dynamic Time Wrapping (wDTW). With the help of weighted dynamic time wrapping (wDTW) they achieve high recognition accuracy (used 30 cluster to compute the weights).Then they apply k-nearest neighbor approach for finding the good value of k. This parameter is often set manually their we use dynamic k-NN to avoid this selection. With an initial value of k we measure the randomness of the top k retrievals. When all the top k retrievals are different words, randomness is maximum, and is minimum (i.e. zero) when all the top k retrieval are same. We increment k by 1 until this randomness decreases. There we assign the label of the most frequently occurring synthetic word to a given scene text. In summarized way given a scene text word and a set of lexicon words, we transformed each lexicon into a collection of synthetic images, and then represent each image as a sequence of features. We find candidate optimal matches for a scene text image in a maximum likelihood framework and solve it using weighted DTW. The weighted DTW scheme provides a set of candidate optimal matches. Finally with the dynamic k-NN method it provide the optimal word in a given scene text image.

B. Scalable Feature Learning Algorithm

In [2] main aim of this paper is text detection and character recognition in scene images with unsupervised feature learning approach . They used k mean clustering technique to train a bank of feature similarly to the system in [18].This method is simpler and faster with other feature learning methods. The learning architecture of this system is proceeds in several stages:

1) They use an unsupervised feature learning algorithm to a set of image that produced from the training data to learn a bank of image features.
2) Features are evaluated convolutionally over the training images. The number of features can be reduced by using spatial pooling [19].
3) Train a linear classifier for text detection and character recognition.

They used ICDAR 2003 training dataset, with the help of word bounding box the text and non text are recognized. Feature extraction method converts each image into 9d dimensional vector. This vector and label acquired from bounding box are then used to train a linear SVM. Then both are use for detection of sliding window. With the help of this features learning method the text recognition performance is increased. Like many feature learning schemes, our system works by applying a common recipe [2]:

1) Collect a set of small image patches, ˜x (i) from training data. In [2] case , they use 8x8 grayscale1 patches, so ˜x(i) ∈ R 64 .
2) Apply simple statistical pre-processing (e.g., whitening) to the patches of the input to yield a new dataset x(i) .
3) Run an unsupervised learning algorithm on the x(i) to build a mapping from input patches to a feature vector, z (i) = f (x(i)).

For the unsupervised learning stage, we use a variant of K-means clustering. K-means can be modified and it will be a dictionary D ∈ R 64 ×d of normalized basis vectors. In [2] they have produced a text detection and recognition system based on a scalable feature learning algorithm and applied it to images of text in natural scenes. They demonstrate that with larger banks of features they are able to achieved increasing accuracy with top performance comparable to other systems. It is similar to results observed in other areas of computer vision and machine learning.

C. Specific Directed Acyclic Graph

In [3] construction of the image text construction graph (iTRG) are their. They present a graph based scheme for color text recognition in images and videos, which is particularly robust to complex background, low resolution or video coding artifacts. This scheme is based on a novel method named the image Text Recognition Graph (iTRG) composed of five main modules: an image text segmentation module, a graph connection builder module, a character recognition module, a graph weight calculator module and an optimal path search module. The first two modules are based on convolutional neural networks. So that the proposed system auto learns how to perform segmentation and recognition. The proposed method is evaluated on the public ICDAR 2003 test word dataset. They contribute the state-of-the-art comparison efforts initiated by ICDAR 2003 by evaluating the performance of our method on the public ICDAR 2003 test word set. The over segmentation method are applied. With the help of this graph is build then character is recognized as [16].The graph weigh calculator module takes input with both result of over segmentation module and the recognition module.Then construction of text recognition graph is finished (iTRG).To retrieved the best sequence of edge we used dijkstra algorithm.

The equation [3] governing the edge weights are detailed below:

Edge weight i,j = outReci,j × [Ps(i) × Ps(j)] × dist(i, j)

• outReci,j represents the output of the character recognition system applied on the image segment (i, j)

• Ps (i) represents the probability that the position cor-responding to the vertex i is a frontier position.
• dist (i, j) is the number of columns separating positions of vertex i and vertex j in the text image

D. Back Propagation Algorithm

In [15], present an approach to recognizing characters in surface mount technology (SMT) product. The recognition rate of SMT product character is not very good, it contains lots of unexpected noise in SMT the product character image. An improved SMT product character recognition method is proposed which can improve the recognition rate. Firstly the character image is changed into a gray image. It is color transformed base method. The noise cannot be eliminate in gray processing. In the low-frequency part generally the energy of the image concentrated, while the noise is located mainly in the high-frequency part. At the same time the extracting image information concentrated in the low-frequency part. Low-pass filtering algorithm used to remove the high-frequency part for maintain the low-frequency information. It's theory says that present pixel is replaced by an average value. By analyzing and comparing an ideal template 1/16 [1 2 1; 2 4 2; 1 2 1] is used. With the use of low-pass filter noise the images are smooth but there probably exists some big noise points in the image and low-pass filter cannot removed them completely even a single or several superimposed low-pass filter are use in several times. The median filter is use to remove them. If f(x, y) represents the dealt image and g(x, y) represents the result image then Median Filter can be showed as :

$$g(x, y) = f(x, y) * h(x, y)$$

where h(x, y) is 1/9 (1 1 1;1 1 1; 1 1 1)
Some image processing algorithms are used to remove the noise. Then single character image is obtained after character segmentation and character normalization. The method of character segmentation is based on the projection of binary image in horizontal direction. A three-layer back propagation (BP) artificial neural network module is constructed. In order to improve the convergence rate of the network. It can also avoid oscillation and divergence, for that BP algorithm with momentum item is used. As a result the SMT product character recognition system is developed and its implementation method and steps are introduced. Experimental results indicate that the proposed character recognition can obtain satisfactory character recognition rate and the recognition rate reached over by 98.6%.

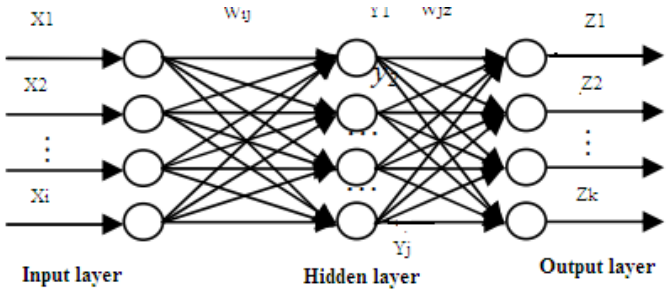

Figure 1 : BP neural network model

They have presented propose approaches to recognizing the SMT product character. By using some appropriate image processing algorithms image noise is eliminated that can affect the character recognition rate. By adding a momentum item in BP algorithm, the character recognition rate is improved and the iteration times in training the neural network is decreased .As figure 1 it shows that three layer back propagation neural network model. Where Xi is input node, Yj is hidden layer and Zk is output node. The hidden layer nodes are too more or too less, the image characters cannot be recognized properly. The character must recognized properly with the help of hidden layer. Maximum number of hidden layer provide proper output. The recognition rate is the best when the hidden layer has maximum nodes. In order to improve the convergence rate of the network and avoid oscillation , the BP algorithm with momentum item is used. The appropriate image processing technology and improved the BP algorithm insure height the recognition rate by 98.6%.

IV. RESULT

In section III we discussed some methodology the previous two methodology is based on without character segmentation segmentation and another two methodology based on character segmentation. The comparison of the result of all above methodology are show in below table.

Table I : TEST RECOGNITION ACCURACY ON VARIOUS DATASET

Method	Dataset		
	SVT word	***ICDAR 2003***	***SMT product***
Non local + Gradient Base Feature+ wDTW	77.28%	89.69%	84.65
Scalable +Feature Learning	72.24	85.5%	89.24
Oversegmentation +GWC+Best Path Search	80.14	92.88%	94.2
Projection Analysis + BPNN	82.35	94.6	98.6%

V. CONCLUSION

In our survey after studied various research paper we found that all different approach for character segmentation in natural scene image provide improved performance. In section IV all the compared result of all those methodology are shown. In future for providing better text recognition we applied genetic algorithm approach.

REFERENCES

1. Vibhor Goel, Anand Mishra, Karteek Alahari, C. V. Jawahar. Whole is Greater than Sum of Parts: Recognizing Scene Text Words. International Conference on Document Analysis and Recognition, Aug 2013, Washington DC, United states.
2. Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, Andrew Y. Ng. Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning.International conference on document analysis and recognition 2011.
3. Saıdane, Z., Garcia, C., Dugelay, J.: The image text recognition graph
4. (iTRG). In: International Conference on Multimedia and Expo, pp. 266–269 (2009).
5. kaoula Elagouni, Christophe Garcia ,Pascale Sébillot.A Comprehensive Neural-Based Approach for Text. rognition in Videos using Natural Language Processing. ICMR, Trento : Italy (2011).
6. Khaoula Elagouni, Christophe Garcia, Franck Mamalet, Pascale S ebillot.Combining Multi-Scale Character Recognition and Linguistic Knowledge for Natural Scene Text OCR .10th IAPR International Workshop on Document Analysis Systems 2012
7. T. Wang, D. Wu, A. Coates, and A. Ng. End-to-end text recognition
8. with convolutional neural networks. In ICPR, 2012.
9. A. Mishra, K. Alahari, and C. V. Jawahar. Top-down and bottom-up
10. cues for scene text recognition. In CVPR, 2012.
11. A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang,
12. D. J. Wu, and A. Y. Ng. Text detection and character recognition in
13. scene images with unsupervised feature learning. In ICDAR, 2011.
14. Elagouni, K., Garcia, C., Sɐebillot, P.: A comprehensive neural-based approach for text recognition in videos using natural language processing. In: International Conference on Multimedia Retrieval (2011).
15. Khaoula Elagouni, Christophe Garcia, Franck Mamalet,Pascale Sɐebillot. Text Recognition in Multimedia Documents: A Study of two Neural-based OCRs Using and Avoiding Character Segmentation. International Journal on Document Analysis and Recognition, IJDAR, 2014, 17 (1), pp.19-31
16. Kopf, S., Haenselmann, T., Effelsberg, W.: Robust character recognition in low-resolution images and videos.Universitɐat Mannheim/Institut fɐur Informatik (2005)
17. Saıdane, Z., Garcia, C.: Automatic scene text recognition using a convolutional neural network. In: Conference on Computer Vision and Pattern Recognition, pp. 100–106 (2007).
18. Kusachi, Y., Suzuki, A., Ito, N., Arakawa, K.: Kanji recognition in scene images without detection of text fields robust against variation of viewpoint, contrast, and background texture. In: International Conference on Pat tern Recognition, vol. 1, pp. 457–460 (2004).
19. Chen, D., Odobez, J., Bourlard, H.: Text detection and recognition in images and video frames. Pattern Recognition 37(3), 595–608 (2004).
20. Huihuang . Zhao, Dejian. Zhou, Zhaohua. Wu. SMT Product character Recognition Based on BP Neural Network. 2010 Sixth International Conference on Natural Computation (ICNC 2010).
21. Z. Saidane and C. Garcia. Automatic scene text recognition using a convolutional neural network. In Proceedings of the Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), Sept. 2007.
22. T. M. Rath and R. Manmatha. Word image matching using dynamic time warping. In CVPR, 2003.
23. A. Coates, H. Lee, and A. Y. Ng, "An analysis of single-layer networks an unsupervised feature learning," in International Conference on Artificial Intelligence and Statistics, 2011.
24. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Back propagation applied to handwritten zip code recognition," Neural Computation, vol. 1, pp. 541–551, 1989.
25. T. Q. Phan, P. Shivakumara, B. Su, and C. L. Tan. A gradient vector
26. Flow based method for video character segmentation. In ICDAR, 2011.
27. A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang,
28. D. J. Wu, and A. Y. Ng. Text detection and character recognition in
29. scene images with unsupervised feature learning. In ICDAR, 2011.

AUTHORS PROFILE

Er. Usha Yadav received her BE (Computer Scieence) from M P Christian College of Engg. And Technology, Bhilai (Recently known as Christian College of Engg. And Technology) in 2009. M. Tech. pursuing with Computer Science and Engineering from MATS University, Raipur. Her areas of interest is Image Processing.

Er. Nilmani Verma is working as Assistant Professor in Department of Computer Science & Engineering, School of Engineering & IT, MATS University, Raipur (CG). He received his BE (Information Technology) from Govt. Engineering College, Raipur (Recently known as NIT Raipur), M.Tech in Computer Science from Birla Institute of Technology, Mesra in 2009. He is having publication in reputed journals and Conference.His areas of interest include Image Processing, Machine Learning & AI.