A study of multi-oriented text recognition in natural scene images

Mona Saudagar[1], S. V. Jain[2]

Student, Computer Science and Eng., Shri Ramdeobaba College of Engineering and Management, Nagpur, India [1]

Assistant Professor, Computer Science and Eng., Shri Ramdeobaba College of Eng. and Management, Nagpur, India [2]

Abstract: The extraction of text in an image is a classical problem in the computer vision. With the increasing popularity of practical vision system text recognition in natural scenes becomes a critical task. Text data present in images and video contain useful information for automatic annotation, image indexing. But as natural scene images contain complex background, multiple fonts and orientations, and different alignments make the problem of automatic text extraction extremely challenging. This paper gives overview of work done for automatic detection of text from images and explains the methodology to extract and recognize multioriented text in natural scene images.

Keywords: Text Extraction, Text Recognition, K-means Clustering, Binarization, Sliding Window, OCR, Hidden Markov Model, Neural Networks.

I. INTRODUCTION

Among all the contents in images, text information is very important, since it can be easily understood by humans and computers. Text in image contains useful information which helps to acquire the overall idea behind the image. Lot of text detection and recognition systems are considered for horizontal or near horizontal texts but detecting texts of random orientations from images have become a challenging task. Detecting texts of random orientations from images is a challenging problem due to the multiple fonts, different sizes, various orientations and alignment, reflections, shadows, the complexity of image background.

Text detection and segmentation from natural scene images are useful in many applications. Text in images can be basically distinguished from is background on the basis of its high contrast or colour. But in natural scene images due to the low contrast or uniform colour it is difficult to segment text from non-text content. And due to this, normal document OCR does not give accurate recognition results.

Extraction and recognition of text from various types of images are very effectual in text based application like: Video and image database retrieval, Image annotation, Data mining, Detection of vehicle license plate. Indexing images or videos requires information about textual content in it. In the same way, natural scene based text explores automatic detection of street name, location, traffic warning and name of commercial goods. Texts in an image directly carry high-level semantic information about a scene, which can be used to assist a wide variety of applications, such as image search and indexing, navigation, and human computer interaction.

II. LITERATURE REVIEW

The methods which are available for text detection in natural scene images are mainly based on texture or region or on both texture and region which is named as hybrid methods.

Texture based methods [3, 6, 7, 8, 9] usually treat the pattern of text appearance as a special texture. Techniques used in these methods include Fourier transform, wavelet decomposition, and combination of wavelet and moments with the help of a classifier to classify text and non-text candidates. The main problems of texture based methods lie in the training of large number of samples and features that heavily depend on the classifier in use. Many of the methods works well for horizontal and near-horizontal straight lines but not curved lines.
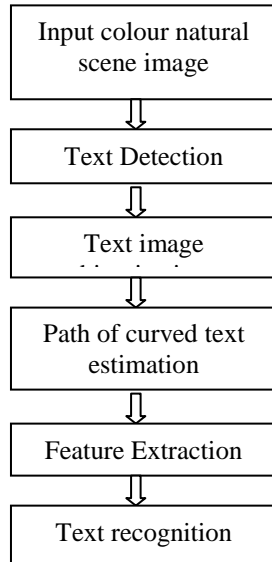
The next category is region-based methods [2, 4, 8, 9]. These methods first identify text regions through edge detection or clustering followed by some heuristics to classify text components followed by a process to eliminate false positives. It requires heuristics and parameters setting. This method is affected by the size of the text and does not provide good results when background and textual contents are closer.

The third category, namely hybrid methods, [5] proposes both texture and region based methods for text detection. Many of the Hybrid methods are suited to horizontal and some near-horizontal straight text lines

images. Such methods work well when all text line components have the same orientations. Any arbitrary text lines such as arc, circular, S, or Z shaped text lines characteristics is no truer as character components on a curved line. Due to such multiple orientation styles text detection has become more challenging problem and demands a new robust text detection technique.

III. METHODOLOGY

Following flow diagram explains the methodology to extract and recognise the text from natural scene images.

```
┌─────────────────────┐
│ Input colour natural│
│     scene image     │
└─────────────────────┘
          ⇓
┌─────────────────────┐
│   Text Detection    │
└─────────────────────┘
          ⇓
┌─────────────────────┐
│     Text image      │
└─────────────────────┘
          ⇓
┌─────────────────────┐
│ Path of curved text │
│     estimation      │
└─────────────────────┘
          ⇓
┌─────────────────────┐
│ Feature Extraction  │
└─────────────────────┘
          ⇓
┌─────────────────────┐
│  Text recognition   │
└─────────────────────┘
```


Figure 1(b)

Figure 2: Binarized image

A. *Text Detection*

The major categories of text detection method are (a) connected component-based [17, 18] (b), texture-based methods [19, 20], and edge and gradient based methods [21, 22]. Connected component based methods expect that the character must be of proper shape; therefore the method may not be suitable for scene text detection in images with complex background or low contrast images.

For complex background or low contrast images texture based methods provide better results as compared to connected component based methods. But it is very difficult to define texture property for scene text detection because sometimes background has the texture similar to text. For scene text detection the edge and gradient feature based method are good in terms of efficiency and some extent to complex background. But these methods suffer from setting threshold values at several stages of the algorithms.

Multi-oriented text has only been partially addressed in [23, 24] where the algorithm is limited to caption text and a few selected directions. Recently, Shivakumara et al. [6] have addressed this multi-oriented issue which is based on Laplacian and skeletonization methods. But this method cannot detect text of arbitrary orientation. That method does not work well for an image with curve text. Method addressed in [10] works well for arbitrary oriented text where method is proposed for

Figure 1(a)

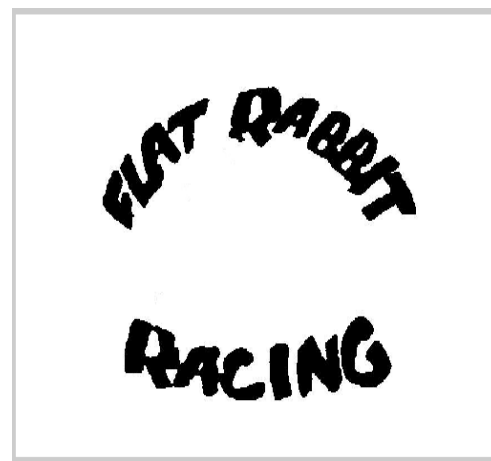dominant text pixel selection based on pixel gradient direction and its magnitude. For each dominant text pixels, text representatives are obtained from the Sobel edge map of the input scene text colour image. To detect the text of arbitrary orientation, region growing phenomenon is introduced to find text direction and then grouping to get entire text. Input colour scene image is shown in figure 1(a) in which we need to detect the text part only as shown in figure 1(b).

A. *Binarization*

If we improve binarization for the segmented text lines in video and natural scene images, recognition rate can be improved with the available OCRs. There various methods for binarization; Niblack, Sauvola, OTSU, SWT, Wavelet-Gradient fusion [13]. Among all binarization methods Wavelet-Gradient fusion has provided the better results as described in [14]. We get the binarized image in figure 2 of figure 1(b).

B. *Determination of Path of Curved Text*

From the binary text line images the foreground pixels of text lines are chosen and fed to a curved fitting algorithm. For feature analysis, the fixed width sliding window is placed at the left most position of the curved line and is moved to the next positions in steps. Since text line is the input for this work, the method finds starting position of text line easily.

C. Feature Extraction

For feature extraction, a sliding window of fixed width can be used to extract a sequence of frames from the curved text. Before the feature extraction stage, the frames are normalized to a pre-defined height. Various features are extracted in this phase. Marti and Bunke [15] explain the approach where sliding window of 1 pixel width has used. In this approach window slides from left to right and at feature extraction is carried out at each successive step.

Another feature extraction approach [16] is based on the calculation of the local gradient histogram (LGH). Here, a sliding window traverses the image and each window is sub-divided into 4×4 (4 rows and 4 columns) regular cells. From all pixels in each cell a histogram of gradient orientations is calculated.

D. Recognition

For recognition purpose Neural Network or Hidden Markov Models [14] can be considered. In [14] recognition is performed using the Viterbi algorithm. It finds the character sequence having the best likelihood using the text line feature vector sequence. Input to this phase is the sequence of feature vectors. In recognition phase, classification errors can be found and those errors can be caused due to ambiguous characters, such as {L, I}, {O, D}, {h, n}, {e, c} etc. It is very difficult to recognise these ambiguous characters correctly. Therefore, to recognise them correctly improvements can be made by considering more robust features.

IV. CONCLUSION

Detecting and recognizing text from natural scene images is still an unsolved problem because sometimes images are of low contrast or of complex background containing text of various orientation styles. As we find classification errors mainly caused by ambiguous characters we can make further improvements by making features of ambiguous characters more robust.

ACKNOWLEDGMENT

REFERENCES

[1]   K. Jung, K. I. Kim and A. K. Jain, Text information extraction in images and video: a survey, Pattern Recognition, 2004, 977 - 997.
[2]   T. Q. Phan, P. Shivakumara and C. L. Tan, Detecting text in the real world, In Proc. ACM Multimedia, 2012, 765-768.
[3] X. Chen and A. Yuille, Detecting and reading text in natural scenes, In Proc. CVPR, 2004, 366 - 373.
[4]   L. Neumann and J. Matas, Real-time scene text localization and recognition, In Proc. CVPR, 2012, 3538 - 3545.
[5]   Y.-F. Pan, X. Hou and C.-L. Liu, A hybrid approach to detect and localize texts in natural scene images, IEEE Trans. IP, 2011, 800 - 813.
[6]   P. Shivakumara, T. Q. Phan and C. L. Tan, A Laplacian approach to multi-oriented text detection in video, IEEE Trans. PAMI, 2011, 412 - 419.
[7]   C. Yi and Y. Tian, Text string detection from natural scenes by structure-based partition and grouping, IEEE Trans IP, 2011, 2594 - 2605.
[8]   C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, Detecting texts of arbitrary   orientations in natural images, In Proc. CVPR, 2012, 1083 - 1090.
[9]   P. Shivakumara, A. Dutta, C. L. Tan and U. Pal, Multi-Oriented Scene Text Detection in Video based on Wavelet and Angle Projection Boundary Growing, Multimedia Tools and Applications, 2013, 1 – 25.
[10] N. Sharma, P. Shivakumara, U. Pal, M. Blumenstein and C. L. Tan, A new method for arbitrarily-oriented text detection in video, In Proc. DAS, 74 - 78.
[11] P. Shivakumara, T. Q. Phan, L. Shijian and C. L. Tan, Gradient Vector Flow and Grouping Based for Arbitrarily-Oriented Scene Text Detection in Video Images, IEEE Trans. CSVT, 2013, pp 1729-1739.
[12] H. I. Koo, "Scene text detection via connected component clustering and nontext filtering," IEEE TIP, 2013, 2296 – 2305.
[13] S. Roy, P. Shivakumara, P. P. Roy and C. L. Tan, "Wavelet-Gradient-Fusion for Video Text Binarization" In Proc. ICPR, 2012, pp. 3300-3303.
[14] S. Roy, P. P. Roy, P. Shivakumara, G. Louloudis, C. L. Tan, HMM-based Multi Oriented Text Recognition in Natural Scene Image©2013 IEEE.
[15] V. Marti and H. Bunke, Using a Statistical Language Model to Improve the Performance of an HMM-based Cursive Handwriting Recognition System.
[16] Jos´e A. Rodr´ıguez and Florent Perronnin, Local gradient histogram features for word spotting in unconstrained handwritten documents.
[17] A. K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition, 1998, pp. 2055-2076.
[18] K. Jung and J. H. Han, "Hybrid Approach to Efficient Text Extraction in Complex Colour Images", Pattern Recognition Letters, 2004 pp. 679–699.
[19] K. L Kim, K. Jung and J. H. Kim. "Texture-Based Approach for Text Detection in Images using Support Vector Machines and Continuous Adaptive Mean Shift Algorithm" IEEE Trans. on PAMI, 2003, pp. 1631-1639.
[20] M. Anthimopoulos, B. Gatos and I. Pratikakis, "A Two-Stage Scheme for Text Detection in Video Images", Image and Vision Computing, 2010, pp. 1413-1426.
[21] D. Chen, J.M. Odobez and J.P. Thiran, "A Localization/Verification Scheme for Finding Text in Images and Video Frames based on Contrast Independent Features and Machine Learning", Signal Processing: Image Communication, 2004, pp. 205-217.
[22] A. Jamil, I. Siddiqi, F. Arif and A. Raza, "Edge-based Features for Localization of Artificial Urdu Text in Video Images", ICDAR, 2011, pp. 1120-1124.
[23] H. Tran, A. Lux, H. L. T. Nguyen and A. Boucher, "A Novel Approach for Text Detection in Images using Structural Features", ICAPR, 2005, pp. 627-635.
[24] J. Zhou, L. Xu, B. Xiao and R. Dai, "A Robust System for Text Extraction in Video", ICMV, 2007, pp. 119-124.

BIOGRAPHIES

Mona Saudagar has received her B.E, degree in computer science and engineering in 2012. She is a pursuing Masters in Technology in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management, Nagpur-440013. Her areas of interest include Image Processing, Pattern Recognition and Artificial Intelligence.

Professor Shweta Jain received the Masters in Technology in Computer Science and Engineering from Nagpur University in 2009 as a first merit holder. She is currently Assistant professor in computer science and engineering department at Shri Ramdeobaba College of Engineering and Management Nagpur. She has a total teaching experience of around 13 Years. Her research interests include Pattern Recognition, Digital Image Processing and Machine Learning.