TEXT LOCALIZATION IN SCENE IMAGES BY BENDELET TRANSFORM

Ajay Kumar Gupta,
Department of Computer Science and Engineering,
IIMT College of Engineering,
Greater Noida, Uttar Pradesh, India
ajayguptagorakhpur@gmail.com

Dr Rama Krishna K,
Department of Computer Science and Engineering,
IIMT College of Engineering,
Greater Noida, Uttar Pradesh, India
tenaliramak@gmail.com

Abstract: In an automated text recognition system, one of the prerequisites is the localization of text. It is a challenging task in scene image due to their background and non uniform size of characters in the images. In this study, an efficient text localization system using bendlet transform is presented. Among the various multi-resolution and multi-directional analysis, bendlet transform has superior property that they classify the curvature precisely. To achieve this property, it uses an addition parameter than shearlets called bending operator. The system decomposes the scene images by bendlet transform and then reconstructs using the bands which contains only the edge information. Then, a series of post processing is applied to locate the text region in a scene image. Results show the robustness of the text localization system by successfully locating the text region in the scene images with different background and non-uniform text sizes.

Keywords: Text localization system, text recognition system, scene image, bendlet transform, multi-directional.

I. INTRODUCTION

The localization of text plays an important role in many applications such as license plate recognition, content based image or video indexing and document image segmentation. As the text in scene image provides much useful information, the exact localization of text is very important for many image based applications. Text localization using natural scene images based on tensor voting is described in [1]. Initially, the text inputs are preprocessed by using grayscale transformation, bilateral filtering and image pyramid generation. The text is localized by vertical edge detection, text region extraction and text line transformation using tensor voting.

Text localization and detection in natural scene images by a hybrid approach is presented in [2]. The input text images are preprocessed by using text region detector, image segmentation and text confidence and scale maps. Then connected component analysis is obtained for component labeling. The texts are localized and grouped by minimum spanning tree clustering and word partition. Text recognition and localization for integrated natural scene text is described in [3]. At first, the input text images are preprocessed by converting into gray scale image and filtered by median filter. The edges are detected by canny edge detector. The adaptive thresholding is applied for the segmentation and also the unwanted regions are removed by morphological operations. The features are extracted by

bounding box, area, Euler number, perimeter and horizontal crossings and classified by Support Vector Machine (SVM).

Texture feature based localization and detection of text from natural scene images is described in [4]. The first and second order statistics of texture features are extracted. The text regions are detected using texture features. The non-text regions are filtered out by discriminative functions. Finally, the detected text regions are merged and localized. Localization and recognition of English text using natural scene image is described in [5]. The input natural scene images are preprocessed by using median filter. Then thresholding and morphological operations is used for segmentation. Euler number, perimeter, area and horizontal crossings features are extracted and classified by SVM.

Fast guided filter and Maximal Stable Extremal Regions (MSER) based text localization and detection is described in [6]. The input text images are preprocessed by edge smoothing MSER and constituent filtering. The attributes on constituents are extracted. Bayesian classifier is used for classification. The texts are labeled by using markov random field. Mean shift clustering is used for text line integration. Automatic recognition of natural scene images with vertical texts is discussed in [7]. Initially, the natural scene images are preprocessed by grayscale conversion and MSER determination. Then the binarization technique is applied for segmentation and morphological operations are used for the text localization and segmentation.

Arbitrary text extraction using Stroke Width Transform (SWT) in natural scene images is described in [8]. The input natural scene images are preprocessed by edge mapping method. SWT is used for the decomposition of images. The identified character candidates are grouped and detected. Two mask filtering based text detection in natural scene images is described in [9]. Initially, the input natural scene images are preprocessed by text confidence map. The Gabor filter is used for the preprocessing and image enhancement. The edges are extracted by sobel operation and images are merged to form text localization.

Text recognition and localization using natural scene images is discussed in [10]. The input images are preprocessed by MSER and deep convolution neural network is used for classification to predict text or non-text regions. Linear spatial filter based text localization is described in [11]. The input natural scene images are preprocessed and extracted by multi-scale pyramid, connected components and spatial filtering. The images are segmented by double thresholding. The classification is made by multilayer perceptron and neighborhood classifier. Super pixel and stroke width histogram based text localization in natural scene images is described in [12]. The input natural scene images are given to SWT then the stroke pixels are extracted. Then the stroke specific edge is taken. Topological skeleton is identified. The accurate stroke width is computed. Finally, the text line is formed

In this study, bendlet transform based text localization approach for scene image is presented. The rest of the paper is as follows: Section 2 discusses the procedure of bendlet transform for decomposing the scene image and various post processing techniques applied to locate the text regions. Section 3 gives the results of text localization on a vast number of scene images. Section 4 concludes the text localization system and gives the future work which is to be carried out.

II. METHODS AND MATERIALS

The text localization system presented in this study is designed based on bendlet transform [13] which is a frequency domain analysis that classify the curvature precisely than other analyzes such as wavelet transform [14], contourlet

transform [15], and shearlet transform [16]. Figure 1 shows the overview of text localization approach using bendlet transform.
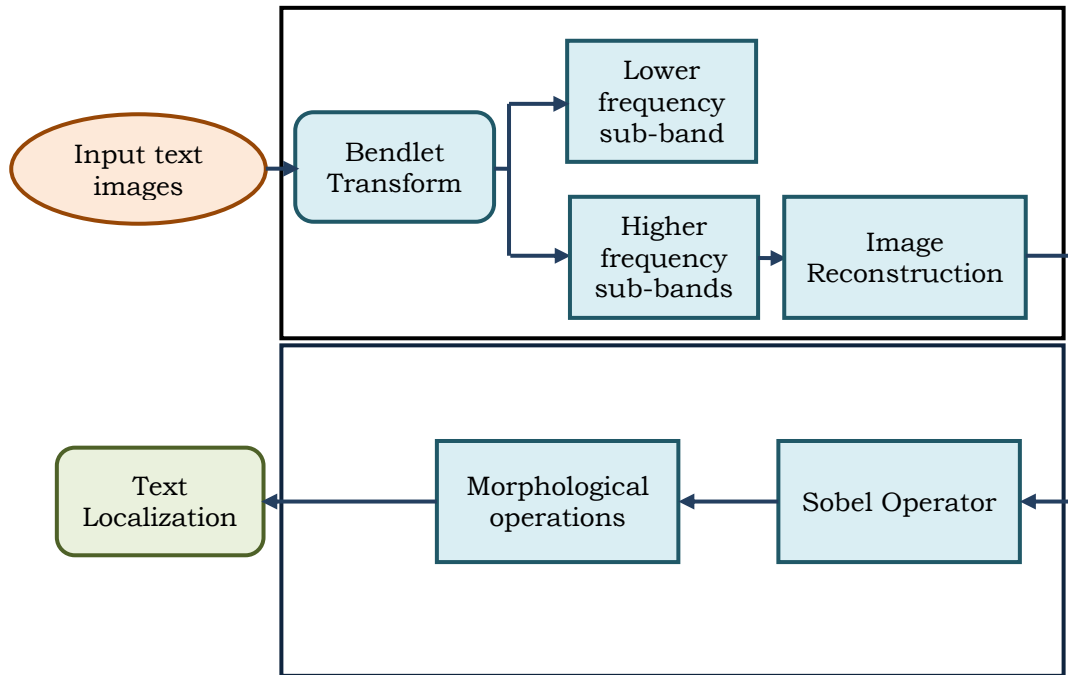
Fig. 1 Bendlet transform based text localization approach

A.   *Bendlet Transform*

Generally, images consist of regions and their contours are piecewise smooth curves. Each region corresponds to one of the objects in that image and their occlusion provides edges. The location of objects in an image facilitates many image processing tasks. Wavelet transform is an efficient approach to extract and characterize the boundaries in a piecewise smooth images. The directional representation systems such as contourlet and curvelet [17] provide information about the normal directions of the boundary curves also. Shearlets detect the non smooth corner points along with the boundaries. However, all these frequency domain transformations are unable to classify the curvature precisely. To overcome this bendlet transform is introduced.

The construction of bendlet transform differs from shearlet in the method of scaling only. It uses alpha-scaling method instead of parabolic method and this parameter is called as bending parameter. It is defined by

$$A_{a,\alpha} = \begin{pmatrix} a & 0 \\ 0 & a^{\alpha} \end{pmatrix} \; where \; a > 0 \, and \, \alpha \in [0,1] \tag{1}$$

where α determines the scaling anisotropy. When α is equal to zero, it becomes directional scaling. The parabolic scaling is achieved by assigning 0.5 to α and for isotropic scaling, it is set to 1. The *l*-th order shearing operator is defined by

$$S_r^l(x) := \begin{pmatrix} 1 & \sum_{m=1}^{l} r_m x_2^{m-1} \\ 0 & 1 \end{pmatrix} (x_1, x_2)^T \tag{2}$$

where l determines the shearing matrix. It is an ordinary shearlet transform when l=1, and bendlet transformation is achieved when l=2. Thus bendlet transform is called as second order shearlet transform. The above two equations, scaling and shearing matrix used with a translation operation gives the bendlet transform for the given input. More information can be found in Bendlet [16], Shearlet [14] and Wavelets []. After decomposing the scene image by bendlet transform, only the high frequency bands are utilized for reconstruction. By removing the low frequency components allow the reconstruction process to produce the image with only edge information and curvature details. Figure 2 shows the bendlet sub-band images, and reconstructed image (last image in the bottom row) with only high frequency bands.
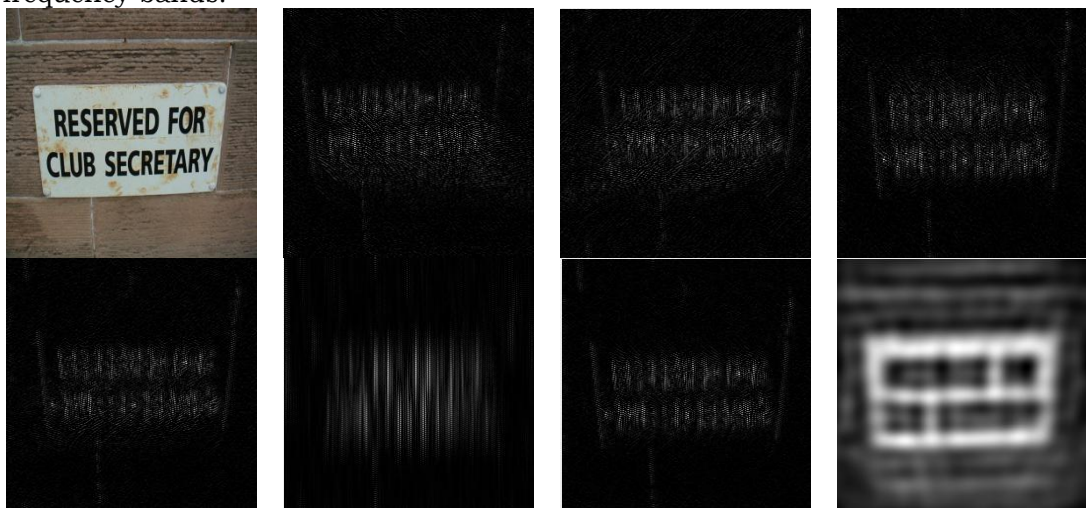
Fig. 2 Bendlet transform sub-band images and reconstructed image

B.   Post Processing

In order to obtain the exact text region, the obtained text region by the bendlet transform is post processed by various image processing approaches. At first, the edge detection operator sobel is employed to detect the edges. It uses one 3x3 kernel in the horizontal direction and one 3x3 kernel in the vertical direction to detect edges in the horizontal and vertical directions respectively. They are defined by

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad and \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3}$$

The convolution of above two kernels with the reconstructed image produces the edges in the right direction and down direction respectively. Then the gradient magnitude is computed as fine edges. It contains text as well as non text region. To remove non text region, morphological operation is used. Also, the connected component analysis is used to extract the information of connected

components such as area, perimeter, major axis length, minor axis length and circularity. From the information, the non text regions are removed.

III. RESULTS AND DISCUSSION

The performance of text localization technique using bendlet transform is evaluated on International Conference on Document Analysis and Recognition (ICDAR) scene images. Figure 3 shows some sample images in the ICDAR databases and ground truth images.
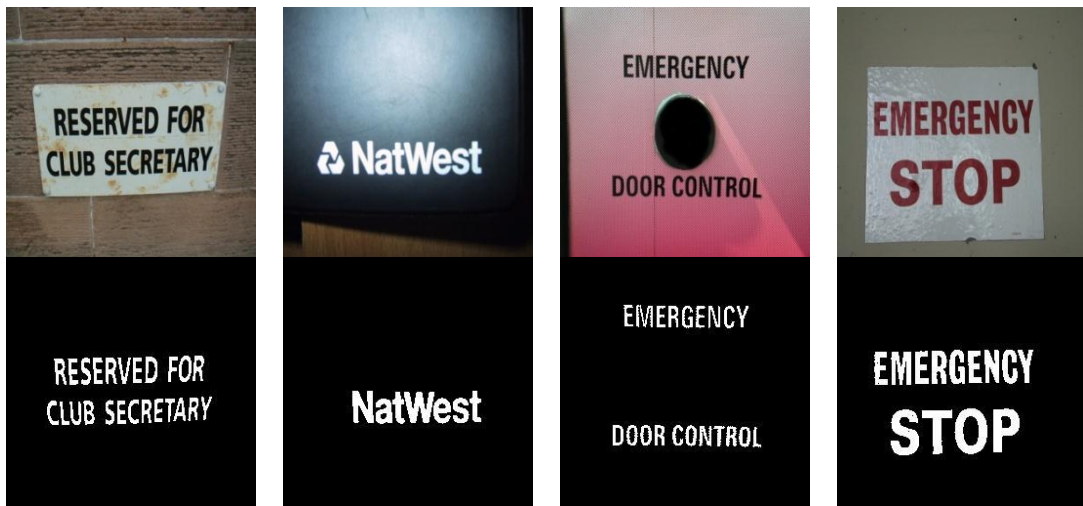
Fig. 3 ICDAR scene images (top row) and ground truth images (bottom row)

At first, the input scene text images are decomposed by bendlet transform with 0.332 as bending parameter. After decomposed, the text regions are initially localized by the high frequency components and then they are fine tuned in the post processing stage. Figure 4 shows the text localized region superimposed with the original image.

Fig. 4 Post processed images (top row) and text localized image (bottom row)

Experimental results in Fig. 4 show that the text localization method by bendlet transform provides promising results on ICDAR database images. However, the system is unable to locate or locate partial text in the following ICDAR images. This is due to the lighting condition and illumination variations in the text region. Figure 5 shows the images with different conditions.

Fig. 5 Images of ICDAR with different conditions

IV. CONCLUSION

In this study, an efficient text localization approach using bendlet transform is discussed for scene images. The text localization is very difficult as scene images have different background and the sizes of texts differ from each other. To extract the location of text, the high frequency components in the bendlet transformed image is utilized. After initial localization of text by bendlet transform, post processing is employed to fine tune the region by edge detection and morphological operations. Results show that the locations of texts are detected accurately on a vast number of scene images. In future, the system can be extended to detect the text from the text localized image which is known as text binarization.

REFERENCES

[1]. T. D. Nguyen, J. Park, and G. Lee, "Tensor voting based text localization in natural scene images", IEEE Signal Processing Letters, Vol.17, No.7, 2010, pp. 639-642.

[2]. Y.F. Pan, X. Hou, and C.L. Liu, "A hybrid approach to detect and localize texts in natural scene images", IEEE transactions on image processing, Vol.20, No.3, 2010, pp. 800-813.

[3]. K.S. Satwashil, and V.R. Pawar, "Integrated natural scene text localization and recognition", International conference of Electronics, Communication and Aerospace Technology, 2017, pp. 371-374.

[4]. T. Kumuda, and L. Basavaraj, "Detection and localization of text from natural scene images using texture features", IEEE International Conference on Computational Intelligence and Computing Research, 2015, pp. 1-4.

[5]. K.S. Satwashil, and V.R. Pawar, "English text localization and recognition from natural scene image", International Conference on Intelligent Computing and Control Systems, 2017, pp. 555-559.

[6].   R. Soni, B. Kumar, and S. Chand, "Text detection and localization in natural scene images using MSER and fast guided filter", International Conference on Image Information Processing, 2017, pp. 1-6.

[7].   O.Y. Ling, L.B. Theng, A. Chai, and C. McCarthy, "A Model for Automatic Recognition of Vertical Texts in Natural Scene Images", International Conference on Control System, Computing and Engineering, 2018, pp. 170-175.

[8].   J. Jameson, and S.N.H.S. Abdullah, "Extraction of arbitrary text in natural scene image based on stroke width transform", International Conference on Intelligent Systems Design and Applications, 2014, pp. 124-128.

[9].   H. Turki, M.B. Halima, and A.M. Alimi, "Text detection in natural scene images using two masks filtering", International Conference of Computer Systems and Applications, 2016, pp. 1-6.

[10]. A. Ray, A. Shah, and S. Chaudhury, "Recognition based text localization from natural scene images", International Conference on Pattern Recognition, 2016, pp. 1177-1182.

[11]. X. Gironés, and C. Julià, "Real-Time Text Localization in Natural Scene Images Using a Linear Spatial Filter", International Conference on Document Analysis and Recognition, 2017, pp. 1261-1268.

[12]. Y. Zhou, S. Liu, Y. Zhang, Y. Wang, and W. Lin, "Text localization in natural scene images with stroke width histogram and superpixel", In Signal and Information Processing Association Annual Summit and Conference, 2014, pp. 1-4.

[13]. C. Lessig , P. Petersen, and M. Schäfer, "Bendlets. A second-order shearlet transform with bent elements", Applied and Computational Harmonic Analysis, Vol. 46, No. 2, 2017, pp. 384-399.

[14]. S. Mallat, "A wavelet tour of signal processing: the sparse way", Academic Press, 2008.

[15]. M.N. Do, and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation", IEEE Transactions on image processing, Vol, 14, No. 12, 2005, pp. 2091-2106.

[16]. W.Q. Lim, "The discrete shearlet transform: a new directional transform and compactly supported shearlet frames", IEEE Trans. Image Processing, Vol. 19, No. 5, 2010, pp. 1166-1180.

[17]. D. Donoho and E. Candes, "Continuous curvelet transform: II. Discretization and frames", Applied and Computational Harmonic Analysis, Vol. 19, No. 2, 2005, pp. 198-222.