Text Region Segmentation From Heterogeneous Images

Chitrakala Gopalan[†] and Manjula[††]

[†]Dept. of Computer Science &Engineering, Easwari Engineering College, Chennai, India
[††] Dept. of Computer Science &Engineering, College of Engineering, Anna University Chennai, India

Summary
Text in images contains useful information which can be used to fully understand images .This paper proposes an unified method to segment a text region from images such as Scene text images , Caption text & Document images using Contourlet transform . Contourlets not only possess the main features of wavelets (namely, multiscale and time-frequency localization), but also offer a high degree of directionality and anisotropy. It decomposes the image into set of directional sub bands with texture details capture in different orientations at various scales. As the contourlet transform is not shift-invariant , non subsampled contourlet transform NSCT is used here to extract text regions from heterogeneous images. Proposed system is tested on various kind of images & shown promising result .

Key words :
Caption text images, Directional filter bank, Laplacian Pyramid, Non subsampled Contourlet transform, Scene text images, Text region extraction

1. Introduction

Text segmentation / extraction from images is an active research area as it is contributing more to Content based image indexing and Text & database retrieval . Text embedded / inserted in images is used to describe the contents of an image & also it can be easily extracted compared to other semantic contents and it facilitates applications such as keyword-based image search, automated processing & reading of documents , text-based image indexing, Pre processing for OCR technique and multimedia processing . Variation in Font style, size, Orientation, alignment & complexity of background makes the text segmentation as a challenging task. Text in images/video images are classified into Caption text and Scene text [5]. Various kinds of images are shown in Figs.1-3.

Fig 1. a) Caption text b) Scene text

Manuscript received October 5, 2008
Manuscript revised October 20, 2008

Caption text is the one which is inserted text & otherwise called as superimposed/artificial text. Natural images /embedded are called as Scene text/graphics text .Bottom-up , Top-down & Hybrid methods are adopted to extract the text from images. Our proposed system employs Contourlet transform as a texture based method (Top-down) to extract text from Caption text, Scene text & document images. The paper is organized as follows. Section 2 deals with the related work, Section 3 illustrates our method with various modules in Section 4-6. In Section 7 experimental results are reported & conclusions and future works are summarized in Section 8.

2. Previous Work

Xiaoqing and Jagath [2][3] proposed edge based method with edge strength , density and the orientation variance as distinguishing characteristics of text embedded in images which can handle printed document & scene text images. This method used multiscale edge detector for text detection & dilation operator for text localization stages. Julinda [4] proposed Connected component based method which uses color reduction technique followed by Horizontal projection profile analysis which can extract text from Caption text images.[1] proposed Globally matched wavelet filters with Fisher classifiers for text extraction from Document images & scene text images.[6] proposed a modification of two existing texture based techniques such as Gabor feature based method & Log polar wavelet signature method with the inclusion of Harris corner detectors for Document images. [7] employed Haar wavelet transform to detect edges of candidate text regions which is followed by the application of dilation operators & this method can handle Caption text images.[8] used Connected component analysis in

c) Document image

binarized images to segment non text areas based on the size information of the connected regions. A Gabor function based filter bank is used to separate the text & non text areas of comparable size.[9] applied delaunay triangulation for the extraction of text areas in document page by representing the location of connected components in a document image with their centroids. Here text regions in the delaunay triangulation will have distinguishing triangular features from image. In this paper, a system is proposed to handle all kinds of images as such single methodology is not proposed so far.

3. Proposed Methodology

Non subsampled Contourlet transform decomposes the image into set of directional sub bands with texture details capture in different orientations at various scales. Multi – oriented Texture details at high frequency component produces text region whereas at low frequency gives rise to non-text region. NSCT is applied to an input image. It produces 2^n sub bands for n level specified. Energy is computed for the sub bands & sub bands are categorized into Strong and Weak bands. Boosting level is applied to weak so as to bring it to the level of strong bands. Then edge detection followed by suitable dilation operator is applied . Strong & boosted edges after dilation are combined with addition followed by AND operation which forms the text region. Finally, remaining Non text regions are identified & eliminated. The flow chart of the proposed methodology is shown in Fig 2. The proposed method consists of the following stages, Candidate text region detection ,Energy Computation, ,Text region localization & Extraction. Following Abbreviations are used in the paper:

NSCT: Non subsampled contourlet transform, **NSP**: Non subsampled Pyramid , **NSDFB**: Non subsampled directional filter bank, **CC**: Connected component

4. Candidate Text Region Detection

4.1 Non sub sampled Contourlet Transform (NSCT)

The contourlet transform is an extension of the wavelet transform which uses multiscale and directional filter banks. Here images are oriented at various directions in multiple scales, with flexible aspect ratios. The contourlet transform effectively captures smooth contours images that are the dominant feature in natural images. The main difference between contourlets and other multiscale directional systems is that the contourlet transform allows for different and flexible number of directions at each scale, while achieving nearly critical sampling.In addition, the contourlet transform uses iterated filter banks, which

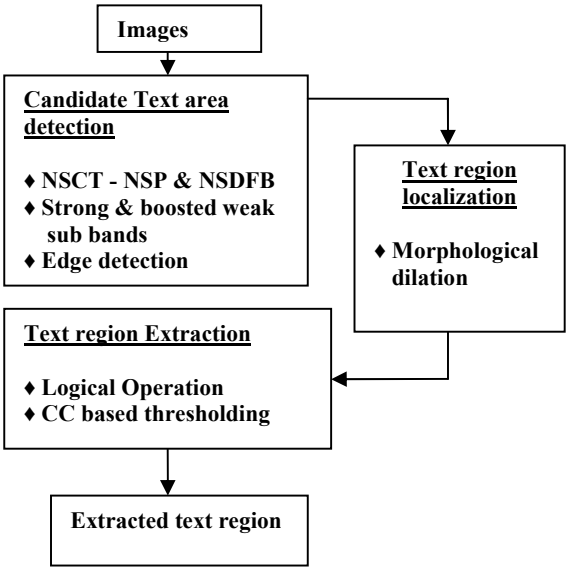


Fig.2 Flow chart of Proposed system

makes it computationally efficient; specifically, it requires O(N) operations for an N-pixel image.The contourlet transform [11] is a multidirectional and multiscale transform that is constructed by combining the Laplacian pyramid [12], [13] with the directional filter bank (DFB) proposed in [14]. Due to downsamplers and upsamplers present in both the Laplacian pyramid and the DFB, the contourlet transform is not shift-invariant. An over complete transform, the Nonsubsampled contourlet transform (NSCT) has been proposed in [10] & NSCT has been applied in our proposed system . The NSCT is a fully shift-invariant, multiscale, and multidirection expansion that has a fast implementation. Here filters are designed with better frequency selectivity thereby achieving better sub band decomposition. Fig. 3(a) displays an overview of the NSCT[10]. The structure consists in a bank of filters that splits the 2-D frequency plane in the sub bands illustrated in Fig. 3(b).This transform can thus be divided into two shift-invariant parts: 1) a non subsampled pyramid structure that ensures the multiscale property and 2) a non subsampled DFB structure that gives directionality. 1) Nonsubsampled Pyramid (NSP): The multiscale property of the NSCT is obtained from a shift-invariant filtering structure that achieves a sub band decomposition similar to that of the Laplacian pyramid. This is achieved by using two-channel non subsampled 2-D filter banks. 2) Nonsubsampled Directional Filter Bank (NSDFB): The directional filter bank of Bamberger and Smith [14] is constructed by combining critically-sampled two-channel fan filter banks and resampling operations. The result is a tree-structured filter bank that splits the 2-D frequency plane into directional wedges. A shift-invariant

directional expansion is obtained with a non subsampled DFB (NSDFB). The NSDFB is constructed by eliminating the downsamplers and upsamplers in the DFB .This is done by switching off the downsamplers/ upsamplers in each two-channel filter bank in the DFB tree structure and upsampling the filters accordingly. This results in a tree composed of two-channel NSFBs. The NSCT is flexible in that it allows any number of directions in each scale. In particular, it can satisfy the anisotropic scaling law. This property is ensured by doubling the number of directions in the NSDFB expansion at every other scale. The NSCT[10] is constructed by combining the NSP and the NSDFB as shown in Fig. 3(a). Eight sub bands have been produced & some of the Non subsampled contourlet coefficients are shown in Fig 4a. for Scene text images.

4.2 Energy computation & Edge detection

The original image is decomposed into eight directional sub band outputs using the DFB at three different scales and the energy of each sub band can be obtained from the decomposed image. The energy of the image block associated with sub band is defined as

$$E_k = \sum_{x=1}^{m_k} \sum_{y=1}^{n_k} |I(x,y)^{\omega}| \qquad (1)$$

Here I (x,y) denote the image intensity corresponding to sub band. Normalized energy value is used instead of energy value to avoid threshold inaccuracies due to spatial intensity variations across the image.The Sub bands are categorized as Strong & weak based on the value of the Computed Energy . To extract dominant directional energy, it is necessary to select a threshold. Now, difference between each sub band with the maximum energy band is determined as in eqn (2) & (3) and Sub bands are arranged in ascending order based on the difference in the energy as in eqn (4).The sub bands occupying first few places from the beginning of the list will have minimum difference & will become candidate for Strong bands. Then proper threshold is applied to separate this sorted list into two sets as Strong & Weak sub bands as in eqn (5) & (6). Energy levels of identified weak sub bands are boosted so as to bring out the proper edges in edge detection stage as in eqn (7). The Sobel operator is used as an edge detector. Edges have been detected for the multidirectional sub bands & some of the edges are shown in Fig 4b.

$M = \text{Max}(E_i)$ (2)
$i = 1,2,\ldots..2^j$, where j= 1,2,….n levels

$d_{kj} = M - E_i$, (3)
$i = 1,2,\ldots..N$ sub bands, $j = 1,2,\ldots..N$

$l = (d_{k1}, d_{k2}, \ldots\ldots\ldots d_{kN})$, (4)
where $d_{k1} < d_{k2} < \ldots\ldots\ldots < d_{kN}$,

ie) set $l_j = d_{kj}$, where j= 1,2,...T….N

Set Threshold T,

$W_i = l_1, l_2, \ldots\ldots\ldots\ldots l_{T-1}$ (5)

$S_P = l_{T+1}, l_{T+1}, \ldots\ldots\ldots l_N$ (6)

$W_i \longrightarrow$ Boosted to $\longrightarrow \text{Min}(S_P)$

as $W_B = W_i * \text{Min}(S_P)$ (7)
where p = T ,T+1,…….. N
i = 1 ,2,…….. ….T-1

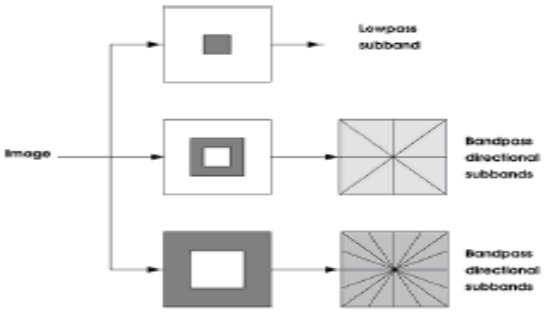
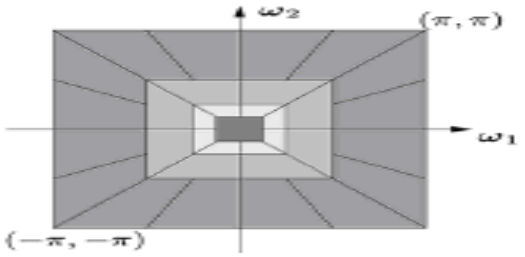


Fig.3. NSCT (a) NSFB structure that implements the NSCT



(b) Idealized frequency partitioning

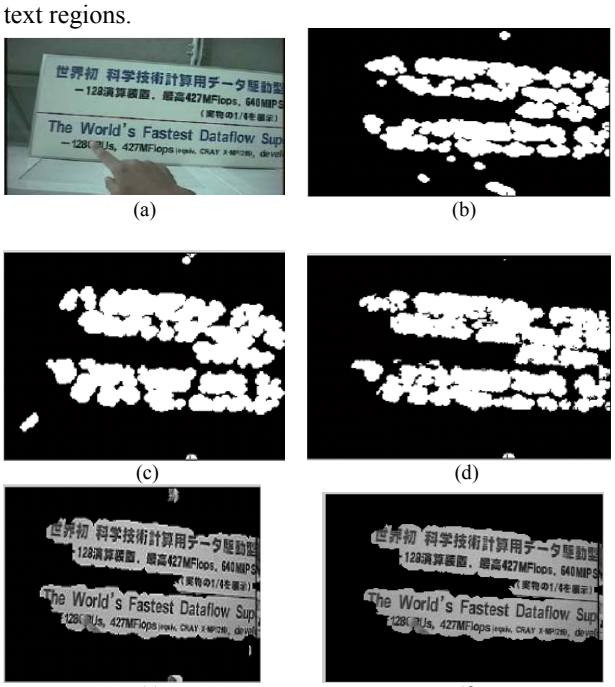
Fig 4a) Non subsampled contourlet coefficients

b) edges

5. Text Region Localization

5.1 Morphological Dilation

Here Detected edges of Strong & boosted weak sub bands are dilated .The basic effect of the dilation operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (*i.e.* white pixels, typically) by adding pixels to the boundaries of the objects in an image. Here dilated images are produced using a disk shaped structuring element of 6 pixels radius. Here Dilation is performed to enlarge or group the identified text regions.



(a) (b) (c) (d) (e) (f)

Fig 5. (a) Scene text (b)&(c) Dilation of boosted sub bands (d)AND operation (e)Mapped to original image (f) After Non-text removal

6. Text Region Extraction

Strong & boosted edges after dilation are combined with addition followed by AND operation which forms the text region as in eqn.8 . Results of Dilation & logical operation and mapping to the original image to get text regions are shown in Fig 5b-5e.

$$O_i = (W_B \cup S_P) \cap IP \quad (8)$$

Remaining Non text regions are identified & eliminated by removing from a binary image all connected components (objects) that have fewer than P pixels, (CC based thresholding)producing another binary image, BW2 as shown in Fig 5f.. The default connectivity is 8 for two dimensions. The basic steps are, 1.Determine the connected components.2. Compute the area of each component. 3. Remove small objects.(CCs having fewer than P pixels).

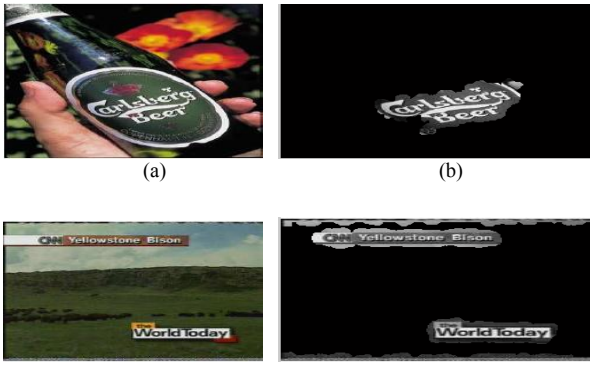
7. Experimental Results & Discussion

40 test images of three types including Caption text, Scene text & document images in which text has different font sizes, colors, orientations, alignments are analyzed to demonstrate the performance of the proposed system. Performance is verified with the oriented text in horizontal & vertical direction with mixed languages (English & Tamil) also .Precision and Recall rates & F-Score have been computed as performance measures . False positives (FP) are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text regions. False negatives (FN)are those regions in the image which are actually text characters, but have not been detected by the algorithm. Correctly detected characters are True Positives (TP).

Precision rate (P) = [TP / (TP)+FP] *100%

Recall rate (R) = [TP / (TP)+FN] *100%

F-score is the harmonic mean of recall and precision equal to :2 * P * R / (P + R)



(a) (b) (c) (d)

Fig 6 a) Scene text. b)Extracted text c) Caption text d) Extracted text

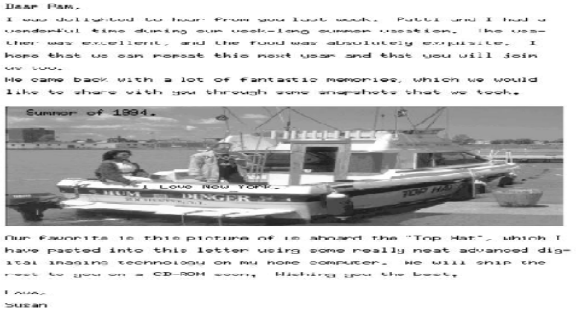


Fig 7a . Document image

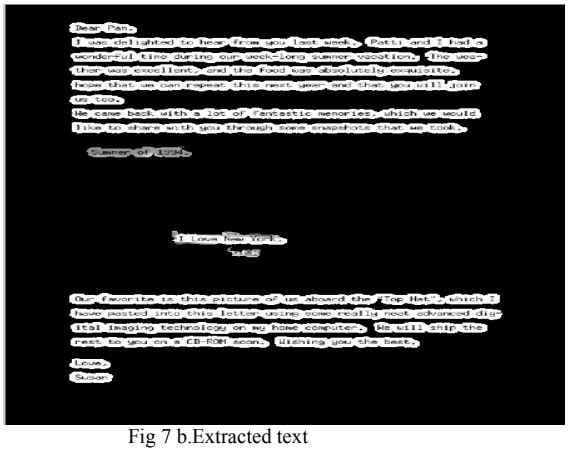


Fig 7 b.Extracted text

Table .I Average Performance measures (for a dataset with 40 images)

Image type	Avg. TP	Avg FP	Avg FN	Avg P	Avg R	Avg F-Score
Caption text images	37	8.9	5.7	81.5	91.7	86.6
Scene text images	31.4	7.5	13.6	72.5	89.7	77.88
Document images	257.6	3.66	18.5	97	87	90.91

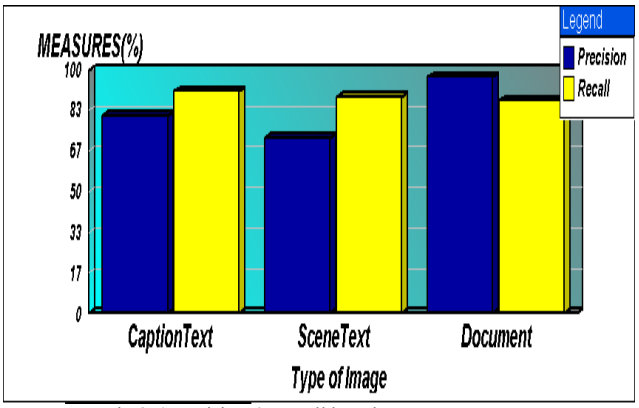
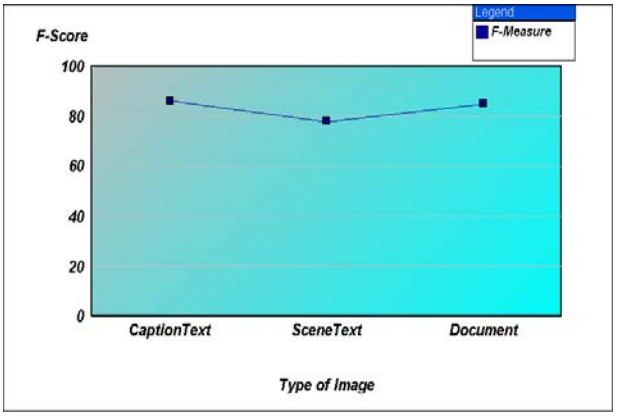


Fig 8 a)Precision & Recall bar chart



b) F-Score line graph

The output image (Fig 6 &7) of the proposed algorithm only consists of detected text regions for Scene text, Caption text & Document images. Average performance measures for the experiments on three types of images are presented in Table I where the average of number of True positives ,number of false alarms/false positives , number of False negatives/Misses and the corresponding average values for precision ,recall and F-score are listed. From Fig.8a, We can see that Our proposed algorithm produces highest recall rate for Caption text image as only some relatively weak texts are missed & highest Precision rate for document images as it has few number of false alarms . Recall rate is comparable for three types of images. From Fig 8 b, we found that even though the Precision rate & in turn average F-score of scene text images comparatively dropped down when compared to caption text images & document images(because of their embedded nature of text inside image), experimental results shown the capability of the proposed system to handle three kinds of images with F-score up to 90% .

8. Conclusion

An unified method is proposed to extract a text region from heterogeneous images such as Scene text images , Caption text images & Document images by using Contourlet transform as it has high degree of directionality . The gradients of the contours at different directions at various scales are used to detect the text regions by combining strong & boosted weak sub bands with suitable logical operation . The proposed method yields better F score upto 90 % . Work is under progress to improve the Precision rate for Scene text images also & We plan to use OCR system to check the Recognition performance for the text images produced by the proposed method.

References

[1] Sunil Kumar, Rajat Gupta, Nitin Khanna, Santanu Chaudhury, and Shiv Dutt Joshi," Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model ", IEEE Transactions On Image Processing, VOL. 16, NO. 8, PP. 2117-2128AUGUST 2007

[2] Xiaoqing Liu and J Samarabandu, Multiscale edge-based Text extraction from Complex images, IEEE, 2006.

[3] Xiaoqing Liu and Jagath Samarabandu, An Edge-based text region extraction algorithm for Indoor mobile robot navigation, Proceedings of the IEEE, July 2005.

[4] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, A Robust algorithm for Text detection in images, Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis, 2003.

[5] Keechul Jung, Kwang In Kim and Anil K. Jain, Text information extraction in images and video: a survey, The

journal of the Pattern Recognition society, 2004.
[6] Nourbakhsh, F. Pati, P.B. Ramakrishnan, A.G. , Document Page Layout Analysis Using Harris Corner Points, Proceedings of ICISIP 2006.
[7] Chung-Wei Liang and Po-Yueh Chen," DWT Based Text Localization", International Journal of Applied Science and Engineering ,2004. 2, 1: 105-116
[8] Sabari Raju,Peeta Basa Patiand A.G.Ramakrishnan," Gabor filter based Block energy analysis for text extraction from digital Document images", Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04)
[9] Yi Xiao,Hong Yan, Text region extraction in Document image based on the Delaunay tessellation, The journal of the Pattern Recognition society, 2003.
[10] Arthur L. da Cunha, Jianping Zhou, and Minh N. Do, The Nonsubsampled Contourlet Transform: Theory, Design, and Applications ,IEEE Transactions On Image Processing, VOL. 15, NO. 10, OCTOBER 2006
[11]. M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," IEEE Trans. Image Proc., IEEE Transactions on ImageProcessing, Dec. 2004.
[12]. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun., vol. 31, no. 4, pp. 532–540, April 1983.
[13] M. N. Do and M. Vetterli, "Framing pyramids," IEEE Trans. Signal Process., vol. 51, no. 9, pp. 2329–2342, Sep. 2003.
[14]. R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," IEEE Trans. Signa Proc., vol. 40, no. 4, pp. 882–893, April 1992

S. Chitrakala is an Assistant professor in Department of Computer science and Engineering, Easwari engineering college . She received the B.E. & M.E degree from Univ. of Madras in 1995 & 2002 respectively & Pursuing Ph.D in Anna university, India . Her research interest includes Image processing, Data mining and Natural language processing .

Dr.D.Manjula is an Assistant professor in Department of Computer science and Engineering , Anna university. She gained her B.E degree from Thiyagarajar College of Engineering in 1983 & M.E & Ph.D degree from Anna university , Chennai in 1987 & 2004 respectively. At present , she teaches & leads research towards language technologies , Data mining , Text mining , Imaging & Networking.