Text Localization and Extraction from Complex Gray Images

Farshad Nourbakhsh*, Peeta Basa Pati, and A.G. Ramakrishnan

MILE Laboratory, Department of EE,
Indian Institute of Science, Bangalore, India – 560012
farshad@ragashri.ee.iisc.ernet.in

Abstract. We propose two texture-based approaches, one involving Gabor filters and the other employing log-polar wavelets, for separating text from non-text elements in a document image. Both the proposed algorithms compute local energy at some information-rich points, which are marked by Harris' corner detector. The advantage of this approach is that the algorithm calculates the local energy at selected points and not throughout the image, thus saving a lot of computational time. The algorithm has been tested on a large set of scanned text pages and the results have been seen to be better than the results from the existing algorithms. Among the proposed schemes, the Gabor filter based scheme marginally outperforms the wavelet based scheme.

1 Introduction

The advancement in science and technology has increased the need for information from the document images. Automatic conversion of paper into electronic document simplifies storage, retrieval, interpretation and updating processes. However, prior to such a conversion , we need to separate the text and non-text regions of the page. This enables proper conversion and interpretation of a document image. Besides, such separation of text and non-text regions, finds many other useful applications in document processing [1]. Moreover, the performance of a document understanding system, such as an optical character recognizer, greatly depends on this separation task.

Numerous approaches on text localization have been reported in the literature. Smith [2] uses vertical edge information for localizing caption text in images. Jung [3] used a neural network based filtering scheme to classify the pixels of input image as belonging to text or non-text regions. Jiang et al. [4] have applied merging bounding blocks, which are using special color features, edge features and morphology operator. These features are used to eliminate the false text candidates. However, this method is script dependent and is reported to be working well for Chinese documents. Yuan & Tan [5] have used edge information to extract textual blocks in Manhattan layout. Messelodi et. al. [6] extract connected components (CC) to characterize text objects in book cover

* Corresponding author.

color images. This is based on the (i) size information of the connected blocks, (ii) geometrical features of a single component, and (iii) spatial relationship of the connected block with other such connected components. Jain and Yu [7] extract a set of images by analyzing the color spaces of the input image. They employ connected component analysis (CCA) on each of the derived images to locate possible text regions. Finally, they merge the information so obtained to locate text regions in the original image. Strouthpoulos et. al. [8] have proposed a technique to separate text from non-text elements based on the optimal number of color components present in the input image. In the first step, an unsupervised neural network clusters the color regions. Subsequently, using a tree-search procedure and split-and-merge conditions, they decide whether color classes must be split or merged. They use a page layout analysis technique, on each of the obtained optimal color images. Finally, they merge the information obtained from each of the optimal color images to extract the text regions. Sabari *et. al.* [9] have employed a multi-channel Gabor filter bank approach for separating text from non-text elements in gray images. In the first level, they separate the obviously non-text objects by a statistical analysis of the connected components of the text page. Following this, they extract a Gabor feature vector at each pixel position. Based on these feature vectors, they decide if the pixel belongs to a text region. Antani *et. al.* [10] and Jung *et. al.* [11] present two comprehensive surveys for text separation and its information extraction in document images and videos.

2 System Description

Here, we propose a texture based text extraction scheme. Figure 1 demonstrates a schematic block representation of the scheme. It is assumed that text regions of an image contain more of abrupt changes in the gray values, in various directions. This makes such regions rich in edge information. An ideal feature to discriminate between text and non-text areas should invariably involve directional frequency. So, the idea of this paper is to separate text areas by applying some direction selective functions. Gabor function based filters are well known for their direction-frequency selectivity. Log-polar wavelet energy signature has also been widely used for texture classification tasks [12]. So it is proposed to

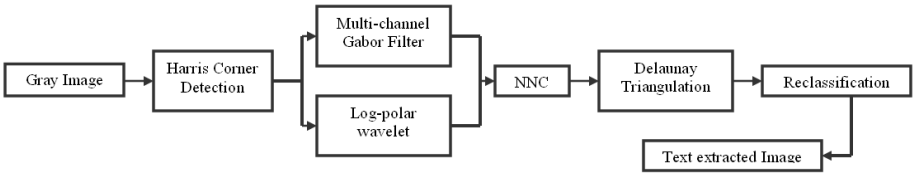

Fig. 1. Figure demonstrating a schematic representation of the proposed text separation algorithms. One of the approaches uses Gabor filters for local energy evaluation while the other uses log-polar wavelets.

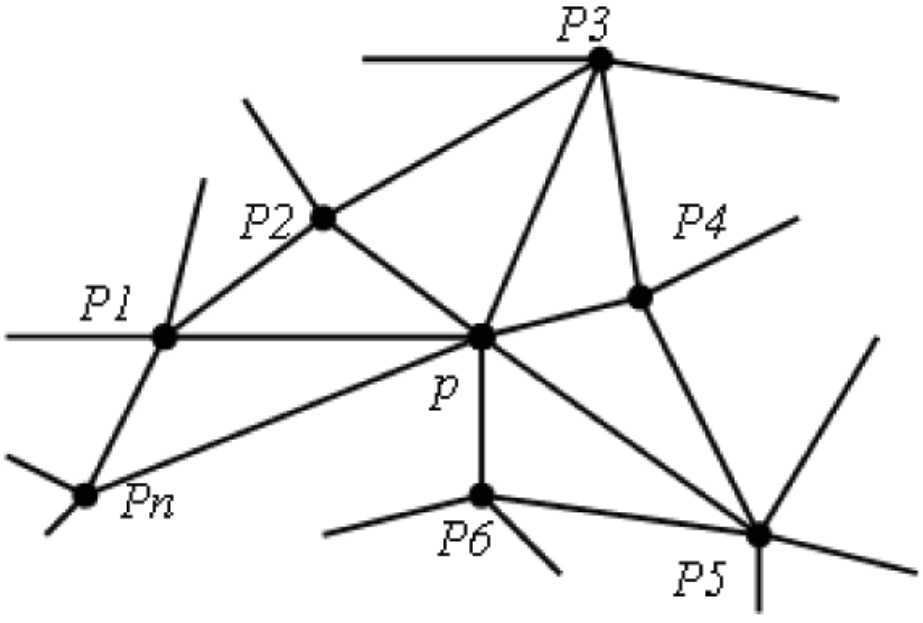

Fig. 2. Figure demonstrating a point p and its neighbors the number of neighbors are n and the Figure shows the connectivity between point p and its neighbors

test its efficacy in distinguishing text and non-text regions of a document image and compare it against that of the Gabor functions.

Previously reported works have used the above mentioned texture descriptors in the following ways: (i) extract the local energy signatures at a pixel level throughout the image, (ii) divide the image into smaller uniform blocks and apply the technique at such a block level. The local energy computation is done uniformly throughout the image in both of the above mentioned ways. However, in the present work, we propose to apply these texture descriptors in a non-uniform fashion, *i.e.*, in some selected information rich points of the image. Such information rich locations are marked by Harris' corner detector [13]. Each of this information rich Harris' corner points are classified as either a text point or non-text point by a Nearest Neighbor classifier (NNC)in the feature domain. The database consists of two classes, text and non-text, formed from english document images. The size of each class is 2000. The extraction of the features at selected points, for text/non-text separation, reduces the computational complexity of the algorithm by many times.

Subsequently, Delaunay triangles are formed using these labeled corner points in the image domain. Lets consider a corner point P in the image, as shown in figure 2. A number of triangles originate from P. Thus, P is associated with a number of other such corner points, P1, P2 $\cdots$ Pn, by the Delaunay triangles. All such points, P1, P2, $\cdots$ Pn are connected to the point P by the Delaunay

triangles are said to be the neighbor points of the point P. All these points are already labeled as text/non-text points. For any given P, the labels of all such neighbor points are considered. For any P, if 75% of the neighboring corner points have a different label, the label of P is altered. Finally, a windowed portion of the original image is retained around all the points which have text label. The rest portions of the image are suppressed to the background.

2.1 Harris Corner Detector

The Harris corners [13] are located as follows:

1. For each pixel located at (x, y) in the image, I, calculate the autocorrelation matrix M.

$$\mathbf{M(x,y)} = \mathbf{G}_\sigma * \begin{pmatrix} (\frac{\partial \mathbf{I}}{\partial x})^2 & (\frac{\partial \mathbf{I}}{\partial x})(\frac{\partial \mathbf{I}}{\partial y}) \\ (\frac{\partial \mathbf{I}}{\partial x})(\frac{\partial \mathbf{I}}{\partial y}) & (\frac{\partial \mathbf{I}}{\partial x})^2 \end{pmatrix} \tag{1}$$

where $\mathbf{G}$ is a Gaussian blurring function with variance σ.
2. Construct a cornerness $\mathbf{C}$(x, y)

$$\mathbf{C}(x, y) = \mathrm{Det}(\mathbf{M}) - k * \mathrm{Trace}(\mathbf{M})^2 \tag{2}$$

Here k is a constant (i.e 0.4-0.6).
3. Threshold the cornerness map $\mathbf{C}$(x, y) – set all values in $\mathbf{C}$(x, y) below a threshold T to zero. Here, the threshold T is taken to be 10 percent of the maximum corner response.
4. Perform non-maximal suppression to find local maxima.
All non-zero points that remain after step 4 are declared as the corners.

2.2 Multi-channel Gabor Filtering

Sabari *et. al.* [9] have used Gabor filter banks for text localization and extraction. The technique involves multi-channel filtering with Gabor function based filters, for page layout analysis. Such a method is reported to detect text regardless of the type of script, size of font or the layout the text is embedded in. It is also claimed to be more robust than other kinds of feature detection models. The bank of Gabor filters reported by Sabari et al. [9], with minor changes to the parameters, is used for the presented work. Here, we are using 8 orientations and 5 different radial frequencies.

2.3 Delaunay Triangulation

The definition of the Delaunay Triangulation [14] is based on the Voronoi diagram through the principle of duality. Given a set of points, the plane can be split in domains for which the first point is closest; the second point is closest, etc. Such a partition is called a Voronoi diagram. If one draws a line between any two points whose Voronoi domains touch, a set of triangles is obtained, known as the Delaunay triangles. Generally, this triangulation is unique. One of its properties of this triangulation rule is that the enclosing circle of a Delaunay triangle does not contain another point. This is demonstrated in figure 3.

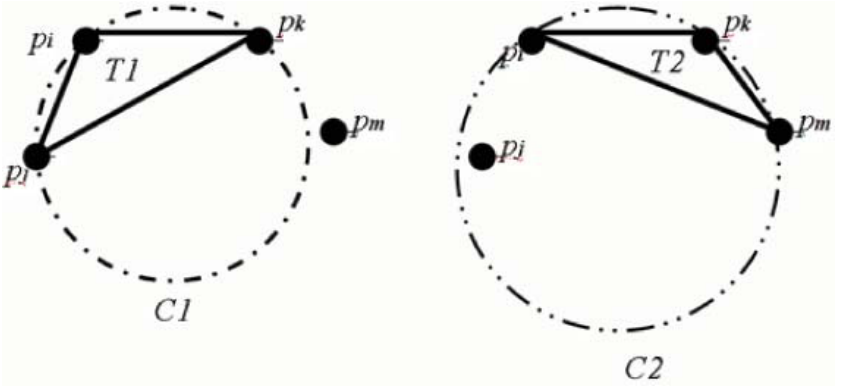

Fig. 3. Figure demonstrating the definition of the Delaunay triangulation between the 3 points pi, pj, pk. Here, the triangle T1 is a Delaunay triangle while the triangle T2 is not because the point Pj lies inside enclosing circle C2. This violates the Delaunay triangulation criteria.

2.4 Log-Polar Wavelet Energy Signature

Pun and Lee [12] have proposed log-polar wavelet energy signatures for rotation and scale invariant texture classification. Their scheme applies a log-polar transform to attain rotation and scale invariance. This produces a row shifted log-polar image which is then passed to an adaptive row shift invariant wavelet packet transform. This is done to eliminate the row shift effects. Thus, the output wavelet coefficients are both rotation and scale invariant. A feature vector consisting of the most dominant log-polar wavelet energy signatures, extracted from each sub-band of wavelet coefficients, is constructed. This feature vector is used for texture classification. We use 25 most significant coefficients to form the feature vector at each point for text/non-text separation.

3 Experimental Results

Document images are scanned using (a) Hewlett Packard Scanjet 2200c, and (b) UMAX ASTRA 5400 scanners and stored with Windows Device Independence Bitmap (BMP) format. The database contains about 100 such scanned document images. The images have variation in document layout and scripts (3 scripts Persian, English and Kannada) and have been taken from newspapers, journals and books. Some images are downloaded from the www net. The input to all the algorithms are gray images. However, we have presented the results with the corresponding color images for better visibility.

Figure 4 shows the outputs of the proposed algorithm at various stages. The colored version of the original input image is presented in (a). The output of the Harris' corner detection algorithm is presented in (b). Here, the detected corner points are marked on the original image. Each of these corner points are classified as either text or non-text points, using a nearest neighbor classifier (NNC)

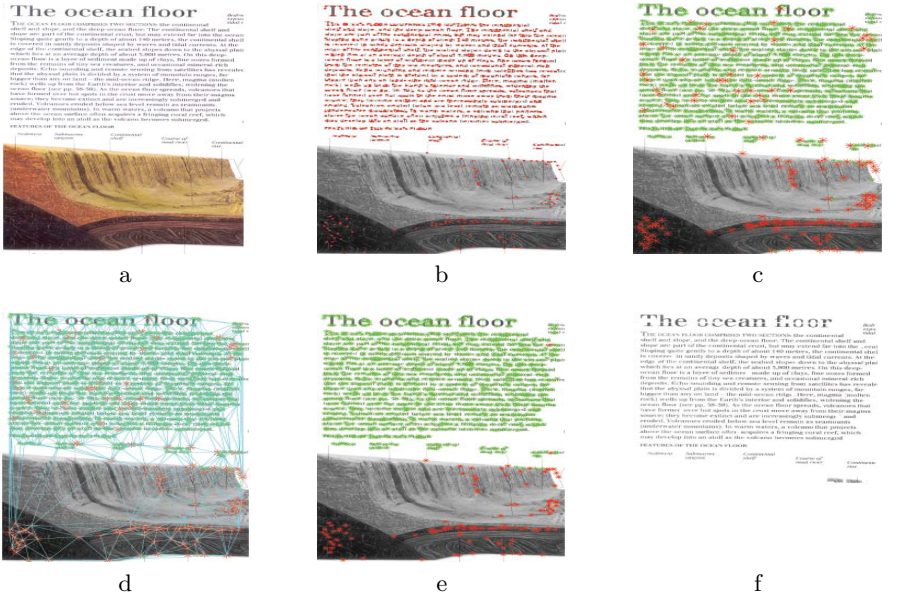

Fig. 4. Figure demonstrating the various stages of the proposed algorithms. (a) Original input image, (b) output of the Harris' corner detection algorithm, (c) labeled text/non-text points - labeling done by NNC on either the Gabor features, (d) Delaunay triangulation using the Harris corner points, (e) classification of the document page using the Delaunay triangles, and (f) the final output.

Fig. 5. Comparison of results of the algorithm using the Gabor features (b) and the log-polar wavelet features (c) for the input image (a)

on either the Gabor features or the log-polar wavelet features. Here, this classification task has employed Gabor features and the results are presented in (c). (d) presents the Delaunnay triangulation using the corner points. (e) demonstrates the results of re-classifying the corner points after the Delaunay triangulation and class label consideration of the neighboring corner points. Finally, the output of the proposed algorithm, with the detected text areas, is presented in (f).

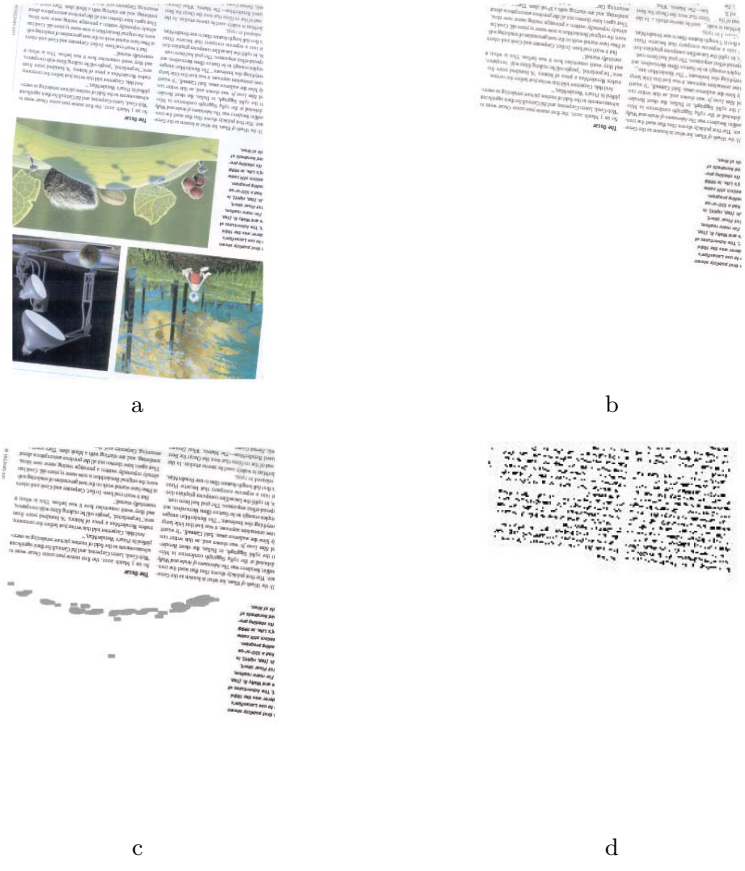

Fig. 6. Figure demonstrating the comparison between result of the algorithm and two previously existing techniques.(a) Original image, (b) the output of the proposed scheme, (c) output of scheme proposed by Sabari *et. al.*, and (d) output of scheme proposed by Xiao *et. al.*

A comparison of the results generated by use of (a) the Gabor features and (b) the Log-polar wavelet features has been demonstrated in figure 5. The above mentioned features have been extracted at the corner points and a near-neighbor classifier has been employed to classify such points to text and non-text points, as has been described in the section 2. Here, it could be observed that the result of employing the Gabor feature has yielded a better output than the ones employing the wavelet features. This has been observed with consistency when tested on other images as well. So, we have used Gabor features for generating the output of our proposed algorithm, in all cases of reported results.

The comparison of the results of the proposed algorithm with those two of the previously proposed algorithms have been demonstrated in figure 6. The two previously proposed algorithms are: (i) page layout analysis using Delaunay Tesselations [15], and (ii) the layout analysis technique proposed by Sabari *et. al.* in [9].

Fig. 7. Figure demonstrating the result of the algorithm for different documents. (a) input image with text in Kannada script, (b) output of the input image in (a), (c) web text image of Persian script and (d) the output of (c).

In this figure, (a) shows the original input image. The output of the proposed algorithm, using the Gabor features, has been presented in (b). (c) and (d) represent the output of the text analysis schemes proposed by Sabari *et. al.* and Xiao *et. al.*, respectively. It could be noted that the result of the proposed algorithm has generated output, which is better than the results generated by the other two techniques.

Figure 7 presents the results of our proposed algorithm, using Gabor features, on some more images. The right column of each row of the figure shows the output of the algorithm for the input image, shown in the left column of the same row.

4 Conclusion and Discussion

The proposed algorithm has been tested on a large set of images. Such images have a lot of variation in the non-text elements present in them. They also have variations in font style, size and script used in the document image. The results have also been verified against two other existing algorithms and the results, in all cases, have either been found to be better than the existing algorithms or as good as their output. Thus, it can be concluded that the proposed algorithm works fine for separation of text from non-text elements in a document image.

A comparison of the results generated by use of Gabor features with those generated by the use of Log-polar wavelet features, has also been done. Here, it is observed that the results of employing Gabor features generates better text/non-text separation in most cases. In some other cases, it is as good as the ones generated using the wavelet features. We never came across a case where result of the wavelet feature was substantially better than the one of Gabor features. This result substantiates the claim made by Sabari *et. al.* that the Gabor functions generate the optimal feature set for text/non-text separation.

The major advantages of the algorithm are:

(i) handles multi-scripted documents,
(ii) invariant to any arbitrary skew,
(iii) accommodates complex layout, and non-Manhattan documents,
(iv) works on poor quality images with acceptable result, and
(v) computationally more efficient than the other proposed techniques.

References

1. K. C. Fan, L. S. Wang and Y. K. Wang, "page segmentation and identification for intelligent signal processing", Signal Processing, 45:329-346, 1995.
2. Smith, M.A., Kanade, T.: "Video skimming for quick browsing based on audio and image characterization", CMU-CS-95-186, Technical report, Carnegie Mellon University, 1995.
3. Jung, K., "Neural network-based text location in color images", Pattern Recognition Letters 22:1503-1515, 2001.
4. Jiang Wu, Shao-Lin Qu, Qing Zhuo, Wen-Yuan Wang, "Automatic text detection in complex color images", Proc. of Intl. Conf. on Machine Learning and Cybernetics, 2002.

5. Yuan, Q., Tan, C. L., "Text Extraction from Gray Scale Document Images Using Edge Information", Proc. of Sixth Intl. Conf. on Document Analysis and Recognition, 2001.
6. Messelodi, S., Modena, C.M., "Automatic identification and skew estimation of text lines in real scene images", Pattern Recognition, 32:791-810, 1999.
7. Jain, A.K., Yu, B., " Automatic text location in images and video frames", Pattern Recognition, 31:2055-2076, 1998.
8. C.Strouthpoulos, N.Papamarkos, Atsalakis, A.E., "Text extraction in complex color Document", Pattern Recognition, 35:1743-1758, 2002.
9. S. Sabari Raju, P.B. Pati, and A.G. Ramakrishnan, "Text Localization and Extraction from Complex Color Images", Int. Sym. on Visual Computing, LNCS – 3804:486-493, 2005.
10. R. Jain, S. Antani and R. Kasturi, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video", Pattern Recognition, 35(4):945-965, 2002.
11. K. Jung, K. I. Kim and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", Pattern Recognition, 37(5): 977-997, 2004.
12. C.M. Pun and M.C. Lee, "Log-polar wavelet energy signature for rotation and scale invariant texture classification", IEEE Trans. PAMI 25(5):590-603, 2003.
13. C. Harris and M. Stephens, "A combined corner and edge detector", proc. 4th Alvey Vision Conf., 147-151, 1988.
14. F. Davoine, et al., "Fractal images compression based on Delaunay triangulation and vector quantization", IEEE Trans. on Image Processing, 5(2):338-346, 1996.
15. Y. Xiao and H. Yan, "Text region extraction in a document image based on the Delaunay tessellation", Pattern Recognition, 36:799-809, 2003.