Document Page Layout Analysis Using Harris Corner Points

Farshad Nourbakhsh[1], Peeta Basa Pati [2], A. G. Ramakrishnan [3]
MILE Laboratory, Department of EE, Indian Institute of Science
Bangalore, INDIA – 560012,
Farshad @ ragashri.ee.iisc.ernet.in

Abstract

Extraction of text areas from the document images with complex content and layout is one of the challenging tasks. Few texture based techniques have already been proposed for extraction of such text blocks. Most of such techniques are greedy for computation time and hence are far from being realizable for real time implementation. In this work, we propose a modification to two of the existing texture based techniques to reduce the computation. This is accomplished with Harris corner detectors. The efficiency of these two textures based algorithms, one based on Gabor filters and other on log-polar wavelet signature, are compared. A combination of Gabor feature based texture classification performed on a smaller set of Harris corner detected points is observed to deliver the accuracy and efficiency.

Key words: Gabor Filters, Log-polar Wavelet Signature, Harris Corner Detector, Text Extraction, Page Layout Analysis, Manhattan Layout.

1. INTRODUCTION

Automatic conversion of paper into electronic document simplifies storage, retrieval, interpretation and updating processes. But better than conversion we need to separate the text and non-text regions in a page. This separation of text and non-text regions finds many useful applications in document processing [1]. The performance of a document understanding system, such as an optical character recognizer, greatly depends on this separation task.

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, using special color features, edge features and morphology operator. These features are used to eliminate the false text candidates. Yuan & Tan [5] have used edge information to extract textual blocks in documents with Manhattan layout. Messelodi et. al. [6] extract connected components (CC) to characterize text objects in book cover 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. Initially, an unsupervised neural network clusters the color regions followed by a tree-search procedure and split-and-merge conditions 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.

2. SYSTEM DESCRIPTION

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, making such regions rich in edge information. Gabor function based filters are used for this task. Log-polar wavelet energy signature is used for texture classification tasks [10]. We compare the efficacy of these two techniques in distinguishing text and non-text regions of a document image.

Earlier works have used the above mentioned texture descriptors in the following ways: (i) extract the local energy signatures at each pixel 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. Such as approach makes the algorithm computationally intense and hence consume a lot of time. This makes these algorithms far from being real time implementable. Our intention in this proposed work is to reduce the computational load of the algorithm but not compromise in the quality of output. This is achieved by obtaining the filter responses at a selected fewer information rich points rather than at all pixels.

Such information rich locations are marked by Harris' corner detector [11]. 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). Subsequently, Delaunay triangles are formed using these labeled corner points. 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 … Pn, by the Delaunay triangles. All these points are already labeled as text/non-text for the texture approach. 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.

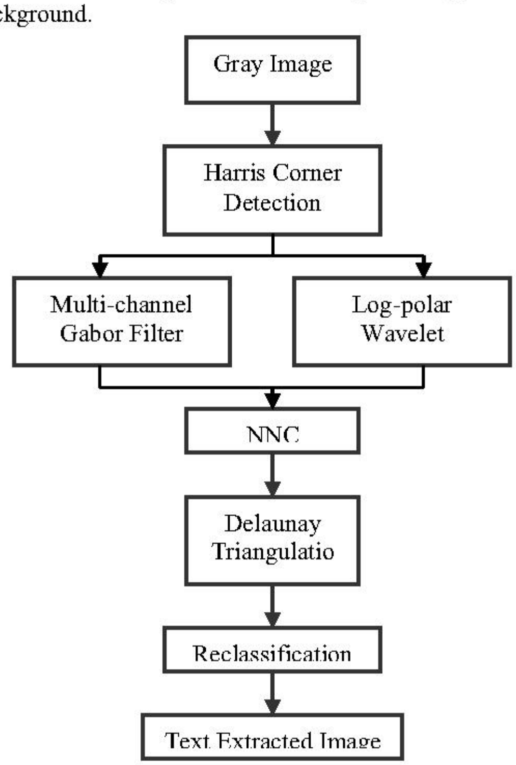


Fig. 1: Figure demonstrating a schematic representation of the proposed text separation algorithms. One of the approaches uses Gabor filters for local energy evaluation while the other uses log-polar wavelets.

The definition of the Delaunay Triangulation [13] 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.

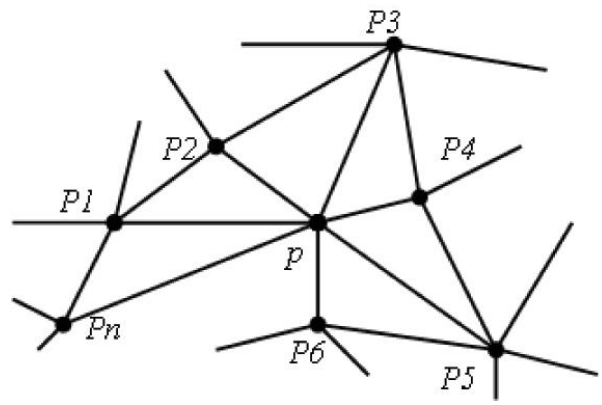


Fig. 2: Figure demonstrating a point p and its neighbors the number of neighbors are n and the Figure shows the connectivity between point p and its neighbors.

Sabari [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.

Pun and Lee [10] 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 Scan-jet 2200c and (b) UMAX ASTRA 5400 scanners and stored with Windows Device Independence Bitmap (BMP) format. The database contains about 100 such scanned 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 World Wide Web. The input to all the algorithms is gray images. However, we have presented the results with the corresponding color images for better visibility.

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 3. 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 comparisons of the results of the proposed algorithm with those two of the previously proposed algorithms have been demonstrated in figure 4. The two previously proposed algorithms are: (i) page layout analysis using Delaunay Tessellations [12], and (ii) the layout analysis technique proposed by Sabari [9]. 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, 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 5 presents the results of our proposed algorithm, using Gabor features, on some more images. The image with different text Scripts (kanara and Persian) for showing the algorithm is script indipendent.

4. CONCLUSION & 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 result 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 that the Gabor functions generate the optimal feature set for text/non-text separation.

The major advantages of the algorithm are:

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

REFERENCES

[1] K. C. Fan, L. S. Wang and Y. K. Wang, “Page segmentation and identification for intelligent signal processing”, *Signal Processing*, vol. 45, 329-346. 1995.

[2] M. A. Smith, Kanade, “Video skimming for quick browsing based on audio and image characterization age segmentation”, *CMU-CS-95-186, Technical report, Caranegie Mellon University*, 1995.

[3] K. Jung, “Neural network-based text location in color images”, *Pattern Recognition Letters*, vol. 22, 1053-1515. 2001.

[4] J. Wu, S. Qu, Q. Zhuo and W. Wang, “Automatic text detection in complex colour images”, *Proc. Of Intl. Conf. on Machine Learning and Cybernetics*, 2002.

[5] Q. Yuan and C. L. Tan, “Text extraction from gray scale document images using edge information”, *Proc. Of Sixth Intl. Conf. on Document Analysis and Recognition*, 2001.

[6] S. Messelodi and C. M. Modena, “Automatic identification and skew estimation of text lines in real scene images”, *Pattern Recognition*, vol. 32, 791-810. 1999.

[7] A. K. Jain and B. Yu, “Automatic text location in images and video frames”, *Pattern Recognition*, vol. 31, 2055-2076. 1998.

[8] C. Strouthpoulos, N. Papamarkos, and A. E. Atsalakis, “Text extraction in complex colour document”, *Pattern Recognition*, vol. 35, 1743-1758. 2002.

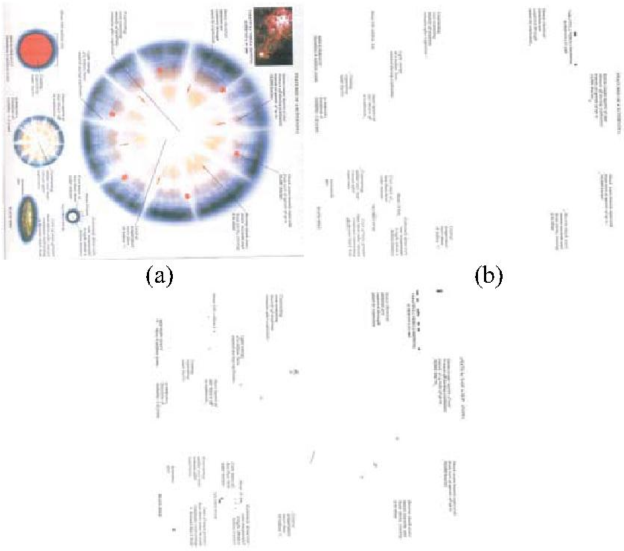
[9] S. S. Raju, P. B. Pati, and A. G. Ramakrishnan, “Text localization and extraction from complex colour images”, *Int. Sym. On Visual Computing, LNCS*, vol. 3804, 486-493. 2005.

[10] C. M. Pun and M. C. Lee, “Log-polar wavelet energy signature for rotation and scale invariant texture classification”, *IEEE Trans. PAMI*, vol. 25(5), 590-603. 2003.

[11] C. Harris and M. Stephens, “A combined corner and edge detector”, *Proc. 4th Alvey Vision Conf.*, 147-151. 1988.

[12] Y. Xiao and H. Yan, “Text region extraction in a document image based on the Delaunay tessellation”, *Pattern Recognition*, vol. 36, 799-809. 2003.

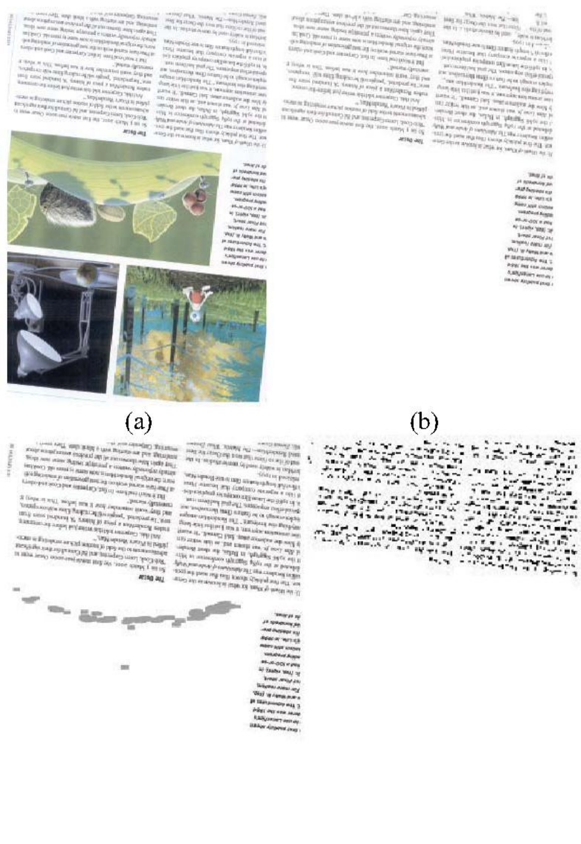
[13] F. Davoine, et., “Fractal images compression based on Delaunay triangulation and vector quantization”, *IEEE Trans. On Image Processing*, vol. 5(2), 338-346. 1996.



(a) (b)

(c)

Fig. 3: Comparison of results of the algorithm using the Gabor features (b) and the log-polar wavelet features (c) for the input image (a).



(a) (b)

(c) (d)

Fig. 4: 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, and (d) output of scheme proposed by Xiao et. al.

(a) (b)

(c) (d)

Fig. 5: Figure demonstrating the result of the algorithm for different documents. (a and c) Original input images layout, (b and d) are the output of the image shown in (a and c), respectively.