

Gabor Filter Based Block Energy Analysis for Text extraction from Digital Document Images

Sabari Raju S, Peeta Basa Pati and A G Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science
Bangalore, INDIA – 560 012.
e-mail:{sabari,pati,ramkiag}@ee.iisc.ernet.in

Abstract

Extraction of text areas is a necessary first step for taking a complex document image for character recognition task. In digital libraries, such OCR'ed text facilitates access to the image of document page through keyword search. Gabor filters, known to be simulating certain characteristics of the Human Visual System (HVS), have been employed for this task by a large number of scientists, in scanned document images. Adapting such a scheme for camera based document images is a relatively new approach. Moreover, design of the appropriate filters to separate text areas, which are assumed to be rich in high frequency components, from non-text areas is a difficult task. The difficulty increases if the clutter is also rich in high frequency components. Other reported works, on separating text from non-text areas, have used geometrical/structural information like shape and size of the regions in binarized document images.

In this work, we have used a combination of the above mentioned approaches for the purpose. We have used connected component analysis (CCA), in 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 and the non-text areas of comparable size. The technique is shown to work efficiently on different kinds of scanned document images, camera captured document images and sometimes on scenic images.

Key Words: Gabor filter, connected component analysis, document image, multi-channel filtering.

1. Introduction

The advancement in Science and Technology, has increased the need for information from the pa-

per document. Automatic conversion of paper documents into electronic documents simplifies and hence speeds up storage, retrieval, interpretation and updating processes. However, an automation of this kind involves lot of complicated processes combined in a distributed fashion. Ordinarily, a document contains different classes of information such as text, picture and graphics. The graphics could be sketches or line drawings. In this paper, all such kinds of non-text classes have been collectively referred to as *clutter*. Appropriate processing is needed to detect and extract text regions from document images. This separation of text and non-text regions, in the document image, finds many useful applications in document processing [1]. Moreover, performance of a document understanding system, such as an *Optical Character Recognizer* (OCR), greatly depends on the separating task, which precedes the OCR module.

Automatic page layout analysis is a non-trivial task. Numerous approaches to page layout segmentation have been reported in the literature. Wang and Srihari reported the maiden work with texture analysis in this field in 1989 [2]. Their work, a texture based method, assumed rectangularity of page blocks. They also assumed that such blocks are well separated, by sufficient horizontal and vertical space, from each other. Besides, their documents were assumed to be aligned with the world co-ordinate system, *i.e.*, without any skew.

Jain and Bhattacharjee [3] have reported a Gabor filter based technique for separating the text and non-text regions in a document image. They have assumed the non-text region to be consisting of only natural images and empty background areas. Thus they do not deal with line based drawings. Chan and Coghill have taken a similar approach for text analysis as well as script identification [4].

The technique reported by Antonacopoulos *et al.* [5] segments and classifies a document image based on white tiles. This is in contrast to other approaches, which rely on foreground information. White tiles provide a flexible data representation combining advantages of both rectangles and polygons. Besides, the approach efficiently deals with arbitrary layouts and limited page skew. This system gives robust representation of complex shaped regions and can only process binary image.

Neural networks have been used for document layout classification by Jain and Zhong [6]. They train a two layered network, with back propagation algorithm, for generating a set of masks. They use these masks to best discriminate between text, background, line-drawing and pictures. The training is accomplished by using sample data from each of these classes. Convolutioning these masks with the input image produces textural features. These features are used by the neural network for classification of each image pixel, into one of three classes: (i) text and graphics, (ii) picture, and (iii) background. The regions containing text and graphics, after binarization and a connected component analysis (CCA), are segmented into separate regions of text and graphics. This method is robust across different scripts. The necessity of a large training set is the major disadvantage of this system, as it consumes a large amount of time to generate such a set.

A stationary Hidden Markov Model (HMM) based scheme has been reported by Chen [7]. A model is generated for each texture type and each model is trained independent of the models corresponding to other classes. If a new texture class is added, a new model is created only on samples of the new class. This is in contrast to neural network based approaches, where a complete *re-training* of the network is necessary for all samples including the ones for which it has already been trained.

Traditional approaches for page segmentation and layout analysis are either bottom-up [8] or top-down [9]. In bottom-up technique connected components are grouped together to form progressively higher level descriptions. Top-down approach looks for global information in the page and splits the page from column level to word level. Some techniques combine both the bottom-up and top-down approaches. Other techniques for page layout analysis are based on run length smoothing [10], run length smearing [11], and fractal signature [12].

Ohya *et al.* have tried to extract characters from scene images, which are camera captured [13]. They have employed a scheme for segmenting characters in a given image, based on adaptive thresholding, con-

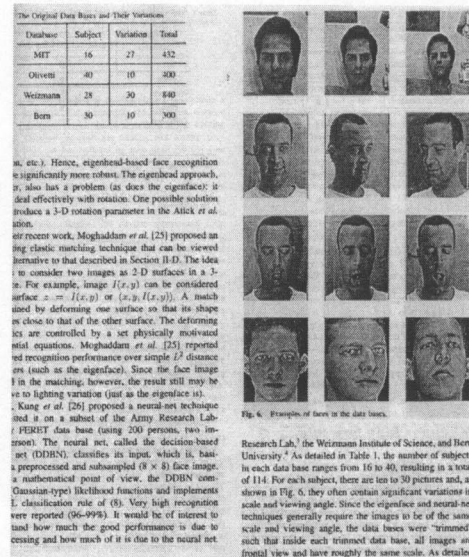


Figure 1. A sample image showing a document containing text, natural images and table.

trast information and region merging based on the contrast information. Wang *et al.* [14] have tried to employ a connected connected analysis based approach, on color images, to localize characters present in scenic images after clustering an image into uniform color regions. Then at each of the color layers, they localize the characters based on a Block Adjacency Graph (BAG) technique along with a aligning-and-merging-analysis (AMA) scheme. Finally, a four step method using, (a) contrast information, (b) structural information, (c) alignment of character limbs, and (d) distance from actual characters in the library, is adopted to retain the valid characters and to eliminate the false ones.

In this paper, a robust technique for layout analysis is proposed. The proposed technique involves both texture-based and non-texture based approaches. The texture-based methods consider a document image to be consisting of various textures of different classes [2, 6]. The lack of ability to discriminate between line-drawings and text regions is a serious disadvantage of methods based on texture properties of the document. On the other hand, non-texture based techniques deal with graphic objects present in the image quite well. This technique relies on connected component analysis [6, 15, 16] through examination of the relationship amongst neighboring pixels.

2. Data Description

Various document images are scanned with, (a) **Hewlett Packard Scanjet 2200c**, and (b) **UMAX ASTRA 5400**, scanners. The spatial scanning resolution is set at 300 dots per inch (dpi) in both the horizontal and vertical directions. The images are scanned in black & white photograph mode (gray level, 8 bits) and stored as Windows Device Independent Bitmap (BMP) format. The database contains about 30 such scanned document images, with variations in layouts and scripts. Figure 1 shows a sample scanned document image containing both text and clutter.

In addition some document images have been captured by a **Minolta Dimage 2330 Zoom** camera. The distance between the camera lens and the surface, on which the document is placed, ranges from 0.4 - 0.7 meters. The illumination of the room is typical study room condition, *i.e.*, 160 watts of tube light (4×40 watts) placed at a distance of 4.5 mts from the surface the document is kept in room of size 12×5 mts. The focus of the zoom lens camera is auto adjusted. The captured color images are stored in JPEG image format. About 20 such camera captured images are contained in our database.

The document images are collected from a variety of books, newspapers and documents printed from the web. Documents are selected such that they contain various kinds of clutter (natural images, graphics, line drawings, hand made sketches etc.) (refer Fig.1). We have assumed that the fonts of the text regions vary both in size and style. We have considered cases for accommodating fonts of size that vary between 8 to 22. The font size we have mentioned are as per the definitions of Microsoft Inc. and are used in MS Word.

3. System Description

The proposed document page text segmentation scheme involves Connected Component Analysis (CCA) on a binary image followed by a Gabor filter based local energy computation on the corresponding gray image. A Block Energy Analysis (BEA) is employed on the energy image, the image consisting of total local energy at each pixel, for separating text blocks from the blocks containing clutter. Anisotropic diffusion based filtering is employed as a preprocessing step, to get rid of possible background noise from the document image. Figure 2 shows a schematic representation of the proposed scheme.

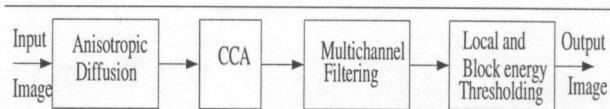


Figure 2. Proposed System for Text Segmentation

3.1. Preprocessing: Anisotropic Diffusion

The primary methods of noise removal are linear low-pass filtering using various filter templates. The price paid for the removal of noise is the decrease of spatial resolution, caused by flattening of sharp edges or blur. When we low pass filter a document image, the edges of the characters are blurred. Many a times, this causes the particular character getting misrecognised, in OCR systems. This drawback of filtering for noise removal is overcome efficiently by anisotropic filtering mechanism by appropriately weighing the diffusion coefficient (blurring coefficient). Such a weight is based on a suitable monotonically decreasing function $g(\|\nabla I\|)$ of the local gradient [17]. Other filters eliminate the high frequency local features.

$$I_t(x, y, t) = \text{div}(c(x, y, t) \nabla I(x, y, t)) \quad (1)$$

where,

$$c(x, y, t) = g(\|\nabla I(x, y, t)\|)$$

$I(x, y, 0)$ is the initial image

$I(x, y, t)$ is the filtered scale-space image.

The anisotropic diffusion based filtering removes noise at pixel locations where gradient doesn't have large oscillations. This could be noticed in figure 3 (b) and (d). In the case of figure 3 (b) which is the filtered output of Fig.3 (a), it can be clearly noticed that the background noise, contributed due to the *show-through* effect, has been removed while the structural elements of the characters are preserved.

3.2. Gabor Filter Bank

In a given text document, consisting of text and non-text regions, the text regions are quite rich in high frequency components. This is clearly demonstrated in figure 4. It can be noted that the high frequency regions of the text image (Fig. 4(c)) are brighter in comparison to the corresponding high frequency elements in the Fourier magnitude spectrum of the non-text image (Fig. 4(d)). This is owing to the fact that the regions consisting of natural images are relatively smooth.

Moreover, the text regions contain a lot of abrupt changes in the gray values, in various directions. The

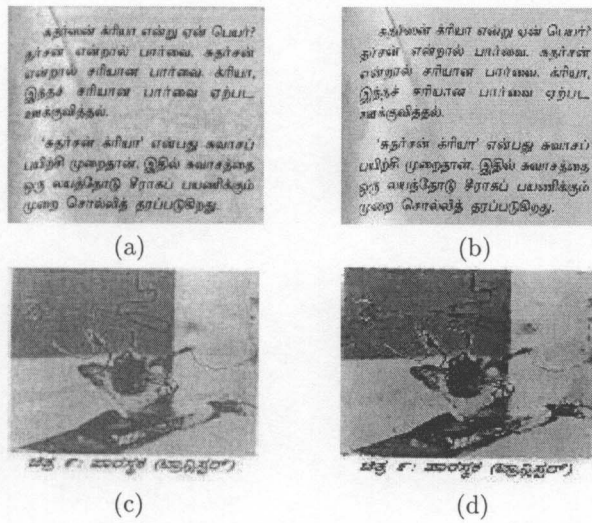


Figure 3. Diffusion Result (a) Poor text document image with noise and background reflection (b) Edges are preserved and background is smoothed (c) Image document with text and image (d) Image is smoothed and the text is preserved

directions of the above mentioned abrupt gray value changes are dependent on the properties of the script of the text for regular kinds of fonts (refer Sec. 2 for the variations on font sizes). This leads to the fact that given a block of the image, if it contains text regions, the block is rich in edge information. So an ideal feature to discriminate such kinds of text and non-text areas would invariably involve directional frequency information.

Biological visual systems are said to be involving a set of parallel quasi-independent filtering mechanisms at the cortical level [18, 19]. This supports the theory of multi-channel filtering of signals in systems that model biological vision [20]. Such a filter scheme employs a filter-bank with each filter modeling a single channel. Moreover, Morrone and Burr [21] have reported that the features, used in HVS, responsible for discriminating different kinds of texture information are direction dependent. Thus, modeling of the HVS is best done by Gabor functions [22, 23]. This scheme is popular in texture segmentation [24] and researchers have tried to use it for text page segmentation [4].

The technique involving multi-channel filtering with Gabor function based filters, for page layout analysis, is developed due to convergence of biological and machine learning approaches. Such a method, meant to detect text regardless of the type of script, size of font

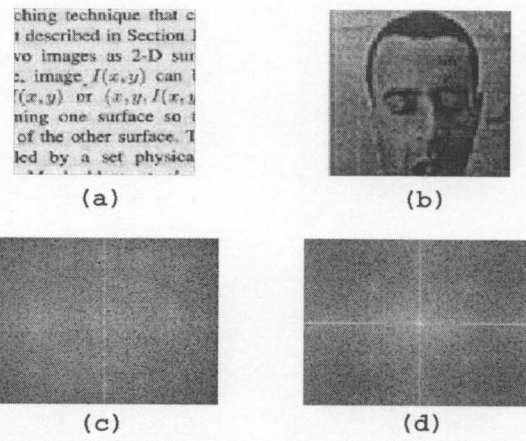


Figure 4. The Fourier magnitude spectra (magnitude is scaled logarithmically) for text and non-text images taken from figure 1. (a) Text region (b) non-text region (c) Fourier magnitude spectra for (a), and (d) Fourier magnitude spectra for (b).

or the layout it is embedded in and is more robust than other kinds of feature detection models. Besides, it has a low sensitivity to various kinds of noise.

A bank of Gabor filters are chosen for extraction of the above mentioned features. This is because of the inherent advantages of the Gabor function. They are: (i) it is the only function for which the lower bound of space bandwidth product is achieved, (ii) the shapes of Gabor filters resemble the receptive field profiles of the simple cells in the visual pathway, and (iii) they are direction specific band-pass filters. Gabor introduced the function in 1946 in one-D case which was later extended to 2D case by Daugman [23]. A Gabor function is a Gaussian modulated by a complex sinusoid. In a 2D case, we put it mathematically,

$$h(x, y) = g(x', y') \exp[j2\pi Ux] \quad (2)$$

where x' and y' are the rotated components of the x and y co-ordinates in the rectangular co-ordinate system. U is the radial frequency in cycles/image width.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (3)$$

$$g(x, y) = \frac{1}{(2\pi\sigma_x\sigma_y)} \exp\left(-\frac{1}{2} \left[\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2 \right]\right) \quad (4)$$

σ_x and σ_y explain the spatial spread and the bandwidth of the filter function $h(x, y)$. If B_r is the radial

frequency bandwidth in octaves and B_θ is the angular bandwidth in degrees, then,

$$\begin{aligned}\sigma_x &= \frac{\sqrt{2}}{2\pi U} \frac{2^{B_r} + 1}{2^{B_r} - 1} \\ \sigma_y &= \frac{\sqrt{2}}{2\pi U \tan\left(\frac{B_\theta}{2}\right)}\end{aligned}\quad (5)$$

The power of discrimination of the different filters are dependent on the values of B_r and B_θ . The aspect ratio of $g(x, y)$ defined as:

$$\lambda = \frac{\sigma_y}{\sigma_x} \quad (6)$$

is the symmetry measure of the filter.

Any combination of B_r , B_θ and U involves two filters, one corresponding to the sine function and the other corresponding to the cosine function in the exponential term in Eqn. 2. The cosine filter (refer figure 5(a)), also known as the real part of the filter function, is an even-symmetric filter and acts more like a low pass filter, while the sine part being odd-symmetric acts like a high pass filter (refer Fig. 5(b)). Gabor filters of $B_r = 1$ octave and $B_\theta = 45^\circ$ at four different orientations ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) have been used for the reported work. Three different radial frequency values ($U = 0.2, 0.3$ and 0.5) have been chosen as they have been observed to be working well for all kinds of documents mentioned in section 2.

3.3. Connected Component Analysis

While designing Gabor filter bank we have assumed that the text regions are more rich in high frequency components than the regions containing natural images. However, this assumption of ours is violated by the presence of line based clutter. Line based clutter are also equally rich in high frequency components as there are abrupt change of the gray levels in the neighborhood of a pixel. This led to the proposition of adopting a different technique to separate these clutter. We have employed connected component analysis (CCA) to accomplish the job. CCA is able to discriminate the text and graphics portions based on component size information (graphics are assumed to be much bigger in size than character symbol blocks).

The document image is binarized prior to feeding it to CCA. This binarization is done by selection of a local threshold value. The gray image is divided into non-overlapping windows. The image is divided into 30 such windowed images (based on experience), 6 blocks along the direction of height and 5 divisions along the width direction. This is accomplished in order to compensate for the variations in the gray levels of the text

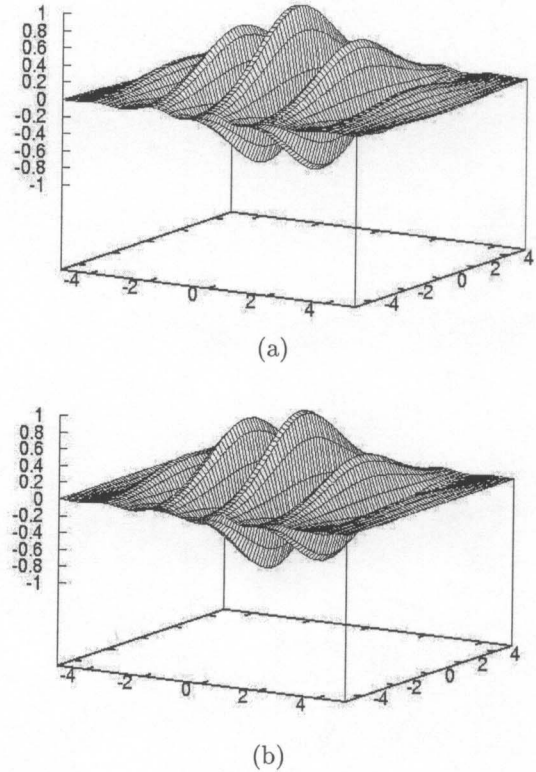


Figure 5. Gabor Function: $U = 0.5$ CPI, $\theta = 0^\circ$ (a) Even Symmetric Gabor Function (b) Odd Symmetric Gabor Function

regions, across different sections of the document image. Such a variation occurs owing to varying illumination intensity and quality of the printing as well the paper. This helps in adapting the threshold for binarization based on the local image statistics. Within each block, the image is assumed to be having a bi-modal histogram, from which the threshold for binarization is appropriately chosen.

The CCA algorithm is run through the whole document image (binarized one) where each connected component is measured for its size. Let the area, A_i , of a connected component, C_i , be the total number of on pixels present in C_i . h_i and w_i are the height and the width of the rectangular bounding box that encloses the component C_i , respectively.

$$\begin{aligned}T_A &= \frac{3}{M} \sum_{i=1}^M A_i \\ H &= \frac{1}{M} \sum_{i=1}^M h_i\end{aligned}$$

$$W = \frac{1}{M} \sum_{i=1}^M w_i$$

where M is the total number of connected components present in the binarized document page. H and W are the average height and width of the bounding boxes to C_i 's respectively. T_A is thrice the average area of the C_i 's, a threshold chosen to discriminate graphics from text, based on area of components.

A given C_i is taken to be a clutter if any one of the conditions mentioned below are satisfied.

- (i) $A_i \geq T_A$ AND, either h_i or w_i is greater than the H or W respectively
- (ii) $h_i \geq 5 \times H$ AND $w_i \leq 0.1 \times W$
- (iii) $w_i \geq 5 \times W$ AND $h_i \leq 0.1 \times H$
- (iv) $h_i \leq 3$ AND $w_i \leq 3$

where condition (ii) and (iii) are meant for taking care of lines. Condition (i) takes care of obvious bigger objects while (iv) ensures the noisy dots present in the image are dealt with appropriately. The on pixels in the clutter, detected by the above method, are set to the gray value equal to the threshold used for binarization of the document in that locality. Such a replacement is accomplished on the original image so as to avoid abrupt changes in the gray levels in the locality, which would defeat the purpose of lowering the energy content in that locality.

3.4. Block Energy Analysis (BEA)

A *space-frequency* filter bank, using Gabor filters, has been designed and implemented for separating the text and non-text areas in a gray level document image. The filter response of a complex filter (refer Eqn. 2) at any pixel (i, j) is:

$$E(u_l, \theta_k) = \sqrt{e(u_l, \theta_k)^2 + o(u_l, \theta_k)^2} \quad (7)$$

where $e(u_l, \theta_k)$ and $o(u_l, \theta_k)$ are the outputs of the cosine(even) and sine(odd) filters, with radial frequency u_l and angle θ_k , at that pixel location, respectively. The total local energy is the sum total of the outputs of all the complex filters at that given pixel.

$$E_T = \sum_{l=1}^3 \sum_{k=1}^4 E(u_l, \theta_k) \quad (8)$$

We low pass filter the magnitude response image of the Gabor filter bank (the E_T image) with a Gaussian filter. A 11×11 Gaussian mask ($\sigma_x = \sigma_y = 3$) is employed for this purpose. It is seen that this removes artifacts and irregularities present and, thereby, helps in text detection. We evaluate block energy at

each pixel by considering blocks of 15×15 size. An average of the block energies is evaluated across all the blocks in the image. This average energy is multiplied by a scale factor of 2 (decided heuristically) to set the threshold for block wise text and non-text separation, *i.e.*, if the block energy is found to be above the threshold value, thus calculated, the block is considered to be belonging to text area.

4. Experimental Results

The proposed scheme can be put down in an algorithmic fashion as follows (refer Fig. 2):

1. Remove isolated noise from the image by anisotropic diffusion (refer Sec. 3.1).
2. Binarize the image using local thresholds (refer Sec. 3.3).
3. Remove the line based clutter with CCA (See Sec. 3.3).
4. Generate the E_T image from the output of the Gabor filter bank (See Sec. 3.2 & 3.4).
5. Separate the text areas based on BEA (refer section 3.4).

The proposed segmentation scheme is applied on the document images described in section 2. Figure 6 deals with the comparison of results obtained by various text separation schemes. Figure 6 (a) shows the input original bilingual newspaper document. In this document, the shown image is printed at a very low resolution. This creates a lot of variations in the gray values, contributing substantial high frequency components in that locality. Hence, multi-channel filtering scheme alone does not work here. However, filtering the document image with anisotropic diffusion (refer Fig.6 (b)), helps in getting rid of this undesirable effect. The results obtained by our scheme is presented in Fig. 6(c). Fig. 6(d) shows the result obtained by a text separation scheme involving multichannel Gabor filtering followed by BEA. It could be noted that such a scheme on this bilingual document has not been able to produce the desired effect. This is because the regions in a natural image contain frequency components in the band of text regions. So patches of the natural image has been detected to be text regions. Fig. 6(e) contains the result of the scheme reported by Chan and Coghil [4]. On closer examination of the Fig. 6(e), it can be noted that some of the characters are smeared, broken and even missing while non-text portions of images have come up as characters. The result of our proposed scheme with-out an anisotropic diffusion on the original image is presented in figure 6(f). The result seems to be reasonably acceptable. However,

we have observed that there are certain cases of documents where our scheme with Anisotropic diffusion performs better than the one without it. At the same time we have not come across a case which proves the contrary, *i.e.*, the result of the scheme without Anisotropic diffusion is better than the one with it.

Figure 7(b) illustrates the results after connected component analysis employed on the binary image of Fig7 (a). It can be noted that the identified non-text objects are homogeneously set to an intensity, almost similar to the background (refer section 3.3). Figure 7(c) shows the E_T image, the image comprising of pixel values corresponding to the E_T at that pixel location, after a low pass filtering with a Gaussian mask. It can be noted that the E_T values in the edge locations in the text areas have taken significantly high values compared to other pixel locations. The final result, after BEA, is shown in Fig. 7(d).

The ability of our scheme to detect even handwritten text areas in images of cards has been demonstrated with figure 8. In this figure, (a) shows the input image while (b) illustrates the well detected hand written text image, which is the output of our proposed scheme.

Figure 9(a) shows a document image captured with the digital camera (mentioned in section 2). Figure 9(b) shows the E_T image for Fig. 9(a) while the output for the same is presented in Fig. 9(c). Please note that the proposed scheme works quite well for camera captured document images also and is independent of the skew of the document image.

Figure 10(a) shows the input to the proposed scheme while Fig. 10(b) presents the output of our system. Closer observation illustrates the detection of the text present on the surface of the airplane image has also been detected. Thus our scheme is not only efficient in separating text and non-text regions but also could be employed for detecting regions containing textual information in natural images.

5. Conclusion and Discussion

In the reported work, we have presented a scheme for selection and separation of text elements in a document image. The document image could either be obtained by flatbed scanner or with a hand held device like camera. The scheme has been illustrated to be working well for all such kinds of documents, irrespective of script, kind of clutter or mode of acquisition.

This model mostly relies on separating the text and non-text elements based on the spatial frequencies of the locality. This has been efficiently exploited by a Gabor filter bank. Since Gabor filter banks have been

shown to be modeling the HVS system the best way, it is employed for the design of the system. The hypothesis that a system designed on the working model of HVS would out perform all other corresponding models has been amply supported here.

The line based clutters have analytically been found to be equally rich in higher frequency components. Another method based on the geometry of the elements of the image is used. This is accomplished by the CCA algorithm and has been shown to be adequately complementing the filterbank approach, enhancing the performance of the system. However, it may be noted from Fig.8 that though handwritten lines are similar to line drawings, they have been classified as text elements. Anisotropic diffusion reduces additive noise efficiently. This is illustrated in Fig. 6(b). Generally, additive noise is found in all kinds of old documents, like old newspapers and books. Thus the choice of anisotropic diffusion has paid off well.

Comparing the scheme with the reported work of Chan and Coghil [4], the BEA is seen to be enhancing the performance of the proposed system. This is mostly because, while our scheme is taking the relationship of a given pixel with its neighboring pixels, such a concept is missing in previously reported Gabor filter based works. However, our experience tells that some post processing with morphological operators, if properly designed and chosen, would enhance the performance of the layout analysis system reported by Chan and Coghil.

Chaudhuri *et. al.* [25] have proposed a scheme for separating text areas in documents mainly consisting of north Indian scripts (*e.g.* Devanagari), taking in to account the connectivity of characters in words. They have employed a scheme based on the horizontal projection profile of the image, where they look for the line connecting the characters, named as *shiro-rekha*. However, our scheme is seen to be working equally well for Devanagari script based documents also. This is mostly because we are employing a scheme which is global in nature and is independent of the script characteristics.

The major advantages of our algorithm are:

1. handles multi-script documents,
2. invariant to skew,
3. works fairly well on camera captured digital document images too, and
4. accommodates complex layout, involving all kinds of clutter.

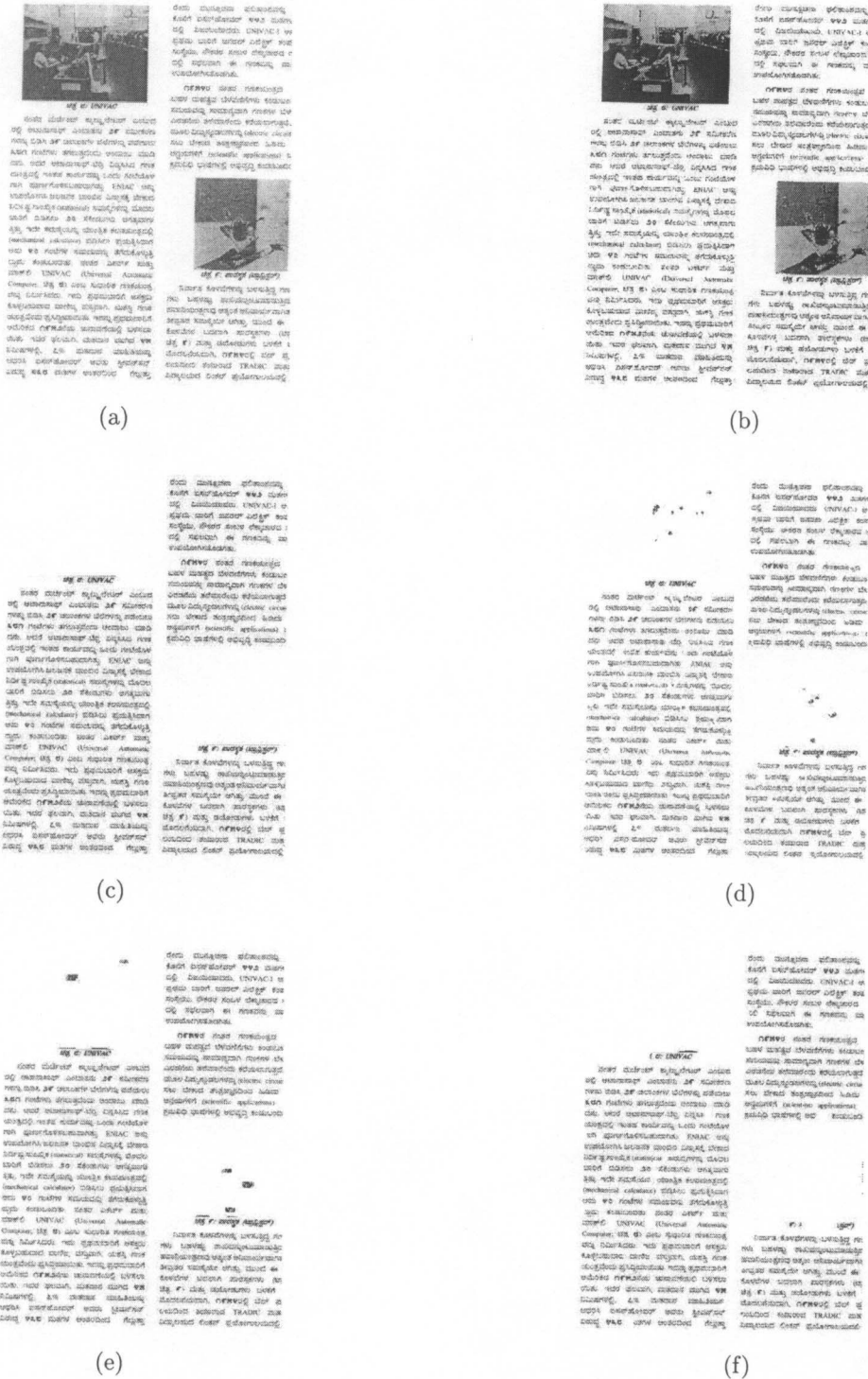
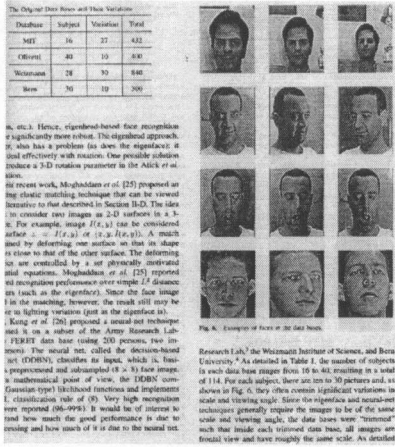
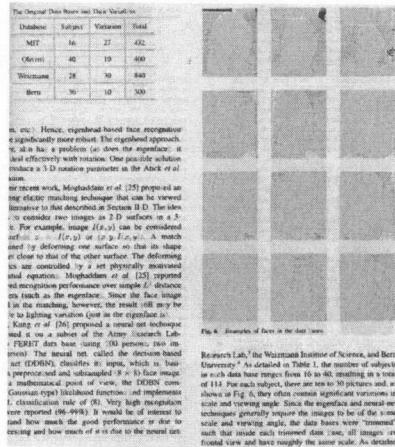


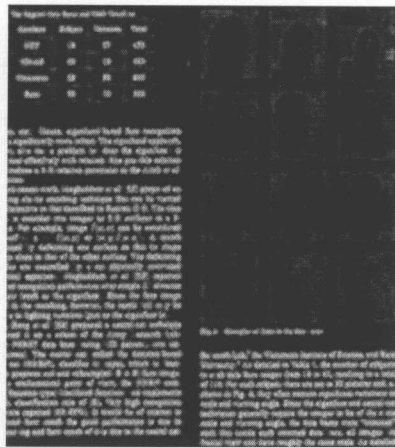
Figure 6. (a) Original bilingual document image, (b) Anisotropic diffusion filtered output, (c) Result with proposed system (Text region bounded), (d) Result of Gabor filter with local energy, (e)Result of Gabor filter + BEA, and (f) Result of CCA + Gabor filter + BEA (without anisotropic diffusion).



(a)



(b)

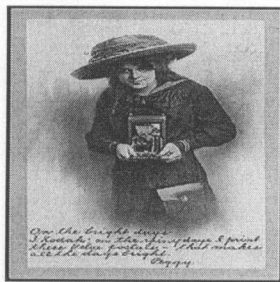


(c)

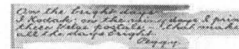


(d)

Figure 7. Experimental results for (a) an ordinary document (1100x1000) with text/image/line-drawing, (b) objects identified as non-text region are homogeneous set, (c) magnitude response(local energy) of multichannel filter, and (d) segmented text using block energy analysis and thresholding.

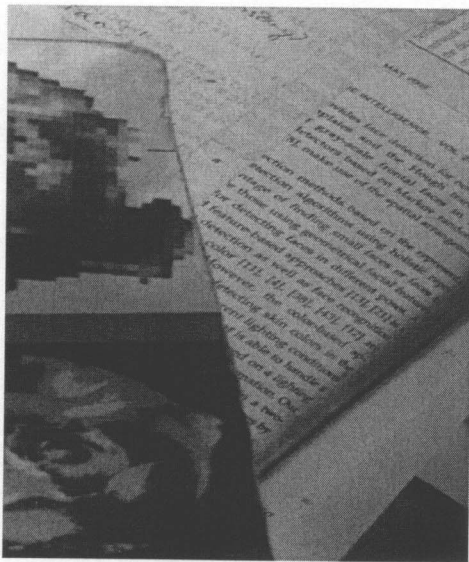


(a)

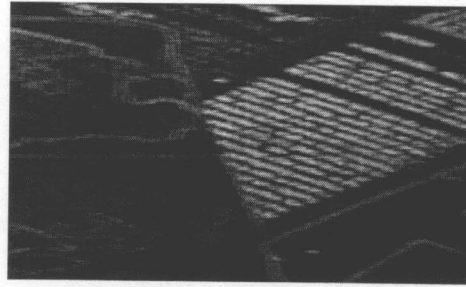


(b)

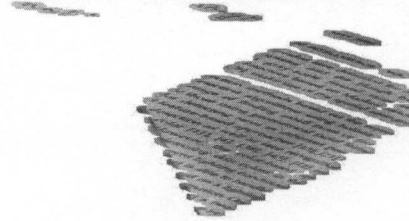
Figure 8. (a) Natural Image (250x250) with Hand-written text, and (b) extracted text region.



(a)



(b)



(c)

Figure 9. Experiment on Camera Capture Image (380×836) (a) Image Captures in Digital Camera at 1m height (b) Gabor response invariant to skew (c) Result text segmentation



(a)



(b)

Figure 10. Experiment on Newspaper Document (800×1100) (a) Input Image with poor text regions (b) Result of text segmentation, observe that text on the picture is also segmented

The major disadvantage of our algorithm is computational cost. The running of the algorithm on the document image shown in figure 1 takes about 30 seconds on a PC with Pentium-IV processor at 2.2 GHz. This is because we have not optimized the implementation of our algorithms. When quality is the criteria for selection of a page layout analyzer, our scheme has been illustrated to be performing well.

6. Acknowledgment

The authors would like to express their heartfelt thanks to Mr. R. Padmaram, Dept. of Instrumentation, for extending his help on programing the anisotropic diffusion portion of the work and explaining the concepts involved therein. They also express their thanks to Mr. Nishikanta Pati and Ms. L. Shradha for collecting and scanning the documents and creating the database.

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] D. Wang and S. N. Srihari, *Classification of newspaper image blocks using texture analysis*, Computer Vision, Graphics, and Image Processing, 47:327-352, 1989.
- [3] A. K. Jain and S. Bhattacharjee, *Text Segmentation Using Gabor Filters for Automatic Document Processing*, Machine Vision and Applications, 5:169-184, 1992.
- [4] W. Chan and G. Coghill, *Text analysis using local energy*, Pattern Recognition, 34:2523-2532, 2001.
- [5] A. Antonacopoulos and R. T. Ritching, *Representation and classification of complex shaped printed regions using white tiles*, Proc. of the Third Intl. Conf. on Document Analysis and Recognition, Montreal, Canada, 1132-1135, 1995.
- [6] A. K. Jain and Y. Zhong, *Page segmentation using texture analysis* Pattern Recognition, 29(5):743-770, 1996.
- [7] J. L. Chen, *A Simplified approach to the HMM based texture analysis and its application to document segmentation*, Pattern Recognition Letters, 18(10):993-1007, 1997.
- [8] A. Simon, J. Pret, and A. Johnson, *A Fast Algorithm for Bottom-Up Document Layout Analysis*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 19:273-276, 1997.
- [9] H. Fujisawa and Y. Nakano, *A Top-Down Approach for the Analysis of Documents*, Proc. 10th Intl. Conf. Pattern Recognition, 113-122, Atlantic City, 1990.
- [10] F. Wahl, K. Wong and R. Casey, *Block Segmentation and Text Extraction in Mixed Text/Image Documents*, Computer Vision, Graphics, and Image Processing, 20:375-390, 1982.
- [11] A. Yamashita, T. Amano, Y. Hirayama, N. Itoh, S. Kato, T. Mano and K. Toyokawa, *A Document Recognition System and Its Applications*, IBM J. Research and Development, 40:341-352, 1996.
- [12] Y. Tang, H. Ma, X. Mao, D. Liu, and C. Suen, *A New Approach to Document Analysis Based on Modified Fractal Signature*, Proc. Third Intl. Conf. Document Analysis and Recognition, 567-570, Montreal, 1995.
- [13] J. Ohya, A. Shio, and S. Akamatsu, *Recognizing characters in scene images*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 16(2):214-220, 1994.
- [14] K. Wang and J. A. Kangas, *Character location in scene images from digital camera*, Pattern Recognition, 36:2287-2299, 2003.
- [15] L. A. Fletcher and R. Kasturi, *A robust algorithm for text string separation from mixed text/graphics images*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 10(6):910-918, 1998.
- [16] S. He, N. Abe and C. L. Tan, *A clustering-based approach to the separation of text strings from mixed text/graphics documents*, Proc. of Thirteenth International Conference on Pattern Recognition, Austria, 25(29):706-710, August 1996.
- [17] Pietro Perona and Jitendra Malik, *Scale-Space and Edge Detection Using Anisotropic Diffusion*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 12(7), 1990.
- [18] F. W. Campbell and J. G. Robson, *Application of Fourier analysis to the visibility of grating*, J. of Physiology, 197:551-556, 1968.
- [19] R. L. De Valois, D. G. Albrecht and L. G. Thorell, *Spatial-frequency selectivity of cells in macaque visual cortex*, Vision Research, 22:545-559, 1982.
- [20] M. Porat and Y. Y. Zeevi, *The generalized gabor scheme of image representation in biological and machine vision*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 10(4):452-467, 1988.
- [21] M. C. Morrone and D. C. Burr, *Feature Detection in Human Vision: a phase dependent energy model*, Proceedings of the Royal Society of London(B), 235:221-245, 1988.
- [22] S. Marcelja, *Mathematical description of the response of simple cortical cell*, J. Opt. Soc. Am., 70:1287-1300, 1980.
- [23] J. G. Daugman, *Two-dimensional spectral analysis of cortical receptive field profiles*, Vision Research, 20:874-856, 1988.
- [24] A. K. Jain and F. Farrokhnia, *Unsupervised texture segmentation using gabor filters*, Pattern Recognition, 24(12):1167-1186, 1991.
- [25] A. Ray Chaudhuri and A. K. Mandal and B. B. Chaudhuri, *Page Layout Analyser for Multilingual Indian Documents*, Proc. of the Language Engineering Conference, pp:24-32, 2002.