A New Approach for Segmentation of Image and Text in Natural and Commercial Color Documents

Malay K. Kundu
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
malay@isical.ac.in

Soumyadip Dhar
RCC Institute of Information Technology
Kolkata, India
rccsoumya@gmail.com

Minakshi Banerjee
RCC Institute of Information Technology
Kolkata, India
mbanerjee23@gmail.com

Abstract

This paper presents an efficient method for segmenting text and non text parts of natural real life images and colored document images using M-band wavelet packet frames. Various combinations of band pass channels of M-band wavelet packet frames represent the image at different scale and orientations in the frequency planes of YCbCr components of color images. The scale space feature vector comprises of the local energy around each pixel at different scales and segmentation is achieved using fuzzy C-means clustering. No information regarding font size, scaling representation, type of layout etc. of the images are considered in our algorithm.

Keywords-color document segmentation; M-band packet wavelet ;adaptive basis selection; texture segmentation.

1. Introduction

Text data present in images contain useful information for various application domains like automatic annotation, indexing and structuring of images. Extraction of text from image having complex background is a challenging task as the text may be of different size, language, orientation, color etc.

The common approaches for text detection focus on text string by using distinct features of text characters. Enyedi et. al.[1] uses edge detection algorithm and histogram to localize character coordinates within the image. Shivkumar et. al[2] applied different edge detectors to search for blocks containing most apparent edges of the text character at a fixed scale and location. Yi el al. [3] developed an algorithm for the text string detection from natural scenes.

Wavelets have been widely used for text and non text extraction as it provides a representation of an image that allows information from each scale to be analyzed separately. Haar discrete wavelet was used by Lian and Chen [4] for text localization. Acharyya. et. al. [5] proposed a method using M-band wavelet transform to segment the text from gray images. Kumar et. al.[6] used matched wavelet and MRF model for document image segmentation.

Real life images captured under different imaging conditions exhibit variations due to illuminations and scanned documents contain text of different size, font, languages, orientations etc. Intitutively, the multiscale nature of the document constituents (characters, lines etc.) may be captured by using multiscale representation of M-band wavelet packet transform.

In view of the above facts we propose a method to extract text and non text part from real life and scanned document images. Various combinations of the M-band packet filters decompose the image at different scales and orientations in the frequency plane. The filter extracts the local frequencies of the image, which in essence gives a measure of local energies of the image over small windows around each pixel. The algorithm is capable of isolating text and non text regions in spite of differences in font size, column layout, orientations, colored text regions with complex background and overlapping regions.

2. M-band Wavelet packet transform

The purpose of filtering is to transform the edges in a texture image (comprising of text and non text regions) into measurable discontinuities. In M-band wavelet decomposition the signal is decomposed in a set of independent, spatially oriented frequency channels.

978-1-4673-4700-6/12/$31.00 ©2012 IEEE

The discrete normalized scaling ($\phi_{j,k}$) and wavelet ($\psi_{j,k,l}$) basis functions are defined in terms of their filter responses as

$$\phi_{j,k}(x) = M^{j/2} h_j(M^j x - k) \quad (1)$$

and

$$\psi_{j,k,l}(x) = M^{j/2} g_{j,l}(M^j x - k),$$

$$l = 1,2.....M-1$$

where j and k are the dilation and translation parameters and l = 1,2,…,M -1 is the number of wavelet functions. h_j and $g_{j,l}$ are respectively the lowpass and bandpass filters of increasing width indexed by j, which are expanded to satisfy the quadrature mirror condition (QMF).

In the standard wavelet decomposition method down sampling by a factor M at each scale is essential [7]. But these decompositions are not translation invariant which is required for image analysis tasks. In the following we give a discrete M-band wavelet frame (DMbWF) decomposition [9], which is same as the discrete M-band wavelet transforms (DMbWT), except that no down sampling is done between levels of decomposition, resulting into set of bands having same size of the input image. It is worth mentioning that there are other alternatives to alleviate this problem of shift (translation) invariance by using complex wavelets [8].

Let $I(x,y) \in l^2(R)$ be a 2D image .Its M-band wavelet frames can be represented as

$$\begin{aligned} c_j(x,y) &= [h_{j,x} * [h_{j,y} * c_{j-1}]](x,y) \\ d_{j,l}^{s_l^1}(x,y) &= [h_{j,x} * [g_{j,l,y} * c_{j-1}]](x,y) \\ d_{j,l}^{s_l^2}(x,y) &= [g_{j,l,x} * [h_{i,y} * c_{i-1}]](x,y) \\ d_{j,l}^{s_l^3}(x,y) &= [g_{j,l,x} * [g_{j,l,y} * c_{i-1}]](x,y) \quad for\, l = 1,2,3 \end{aligned} \quad (2)$$

where $c_0 = I(x,y)$ is the original image, * denotes the convolution operator, $h_{j,x}$ $(g_{j,l,x})$ and $h_{j,y}$ ($g_{j,l,y}$) are the low pass (band pass) filter coefficients along x and y direction respectively. The number of frequency channels resulting from the decomposition using the above equations is 16, at the first level, 4 bands (M=4) along the x-direction and 4 bands (M=4) along the y –direction. $c_j(x,y)$ denotes low pass or approximate band whereas $d_{j,l}^{s_l^i}$ corresponds to detail frequency bands at scale j obtained by band pass filtering along a specific direction.

M-band wavelet packet transform (DMbWPT) is a direct generalization of Discrete M-band wavelet transform (DMbWT). DMbWPT recursively decomposes the higher frequency bands and results in tree structured multiband extension of wavelet transform. At scale $j = J$, the input image is first decomposed into $M \times M$ bands using all the filters h_j and $g_{j,l}$ with $l = 1,2,3.....$ and without downsampling. The process is repeated recursively resulting a tree structure of frequency bands.

3. Proposed methodology

The color image is transformed into YCbCr planes and each plane is decomposed into a number of frequency bands using M-band packet wavelet transform without downsampling (oversampled). Energy for each channel is then computed followed by an adaptive basis selection to reduce the dimensions of the bands. Among various channels, those for which energy values exceed ε_1 percent of the energy of the parent band and ε_2 percent of the total energy of all the subbands at the current scale are considered for further decomposition. The analysis is performed up to the third level of decomposition and this result in a set of wavelet packet bases. Empirically, we have seen that a value of ε_1 =5 percent and ε_2 =70 -80 percent are good choices for the images we have considered here. From the selected bands feature image is generated computing local energy around each pixel followed by smoothing function. The size of the window for computing local energy is determined using a spectral flatness measure (SFM) which is defined as the ratio of arithmetic mean and the geometric mean of the Fourier coefficients of the image. It has been reported in literature that the size of the neighborhood for computation of localized energy range from 11×11 to 31×31, while SFM varies from 1 to 0. From the computed feature image feature vectors are generated. Finally we have used fuzzy C-means clustering to segment the feature vectors into two clusters. The detailed algorithm is shown in the figure 1.

4. Implementation details

M-band packet wavelet color and texture feature extraction:

We use a 1-D, 16 tap 4 band wavelet filters with linear phase and perfect reconstruction property for the wavelet packet transform of the image. Prior to the packet transform we decomposed the image into Y, Cb and Cr planes. This is done to take into consideration both the color and texture property of the image.

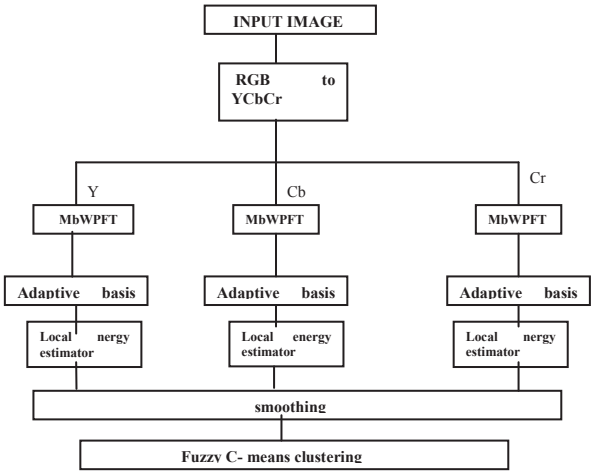


fig 1.

M-band Wavelet packet decomposition over the intensity plane characterizes the texture information, while the packet decomposition over chromaticity planes characterizes color.

Local energy estimator:

Local energy of each pixel is computed to capture the areas in each band where the band pass frequency components are strong resulting in a high energy values and the areas where it is weak into a low energy values. The local energy $eng_{k_i}(x,y)$ around the each pixel (x,y) for nonlinearities is given below.

$$eng_{k_i}(x,y)=\sum_{m=1}^{w}\sum_{n=1}^{w}h_{k_i}(x,y)^2 \qquad (3)$$

where w is the window size and $R=w\times w$ and $h_{k_i}(x,y)$ is the filtered image. The nonlinear operation is followed by a Gaussian low pass (smoothing) filter of the form

$$H_G(u,v)=\frac{1}{2\pi\sqrt{\sigma}}e^{\frac{1}{2\sigma}(u^2+v^2)} \qquad (4)$$

where σ determines the passband width of the averaging window. Formally the feature image $Feat_{k_i}(x,y)$ corresponding to the transformed image is given by

$$Feat_{k_i}(x,y)=\frac{1}{G^2}\sum_{(m,n)\in G_{x,y}}|\Gamma(h_{k_i}(m,n))| \qquad (5)$$

where $\Gamma(\cdot)$ is the local energy estimator and $G_{x,y}$ is a $G\times G$ window centered at the pixel with the coordinates (x,y).

Computation of feature and Classification:

The vector of features at different scales taken at a single pixel in an image is given by

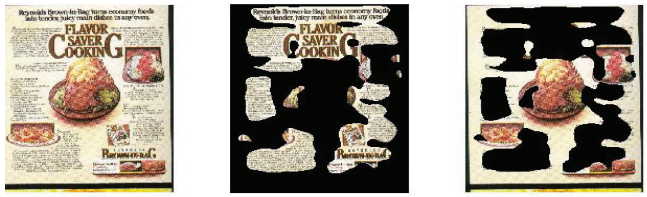
$$Feat(i,j)=[Feat_0(i,j),Feat_1(i,j),.....,Feat_k(i,j)]$$

Where 0…k represents the bands and subjected to fuzzy C-means clustering assuming two clusters one text and another non text regions.

5. Experiments and comparison

We use ICDAR dataset, document images, colored advertisements and container label to demonstrate the performance of our algorithm.

- Structured Document images: Figure 2(a), fig 4(a) show colored document image. Fig 4(a) shows combinations of newspaper at different angles.
- Highly unstructured images with overlapping texts: fig 6(a) and fig 7(a) show the overlapping texts with complex background.
- Natural scene images: fig 3(a), fig 5(a) and fig, 8(a) are natural scene camera captured images. Fig 5(a) and fig 8(a) taken from ICDAR dataset.



(a) (b) (c)

fig. 2 (a) Multicolored document image with text of different sizes. (b) Segmented out text portion. (c)Image portion of the document

(a) (b)

fig 3 (a) signboard image with texts of different sizes.(b) segmented text portion is highlighted

(a) (b)

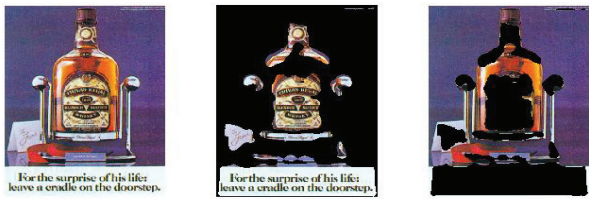
fig 4.(a) image of different newspaper at different angles.(b) segmented out text portion.

(a) (b)

fig 5.(a) Camera captured natural scene image.(b) text portion is segmented out.

(a) (b) (c)

fig 6. (a) colored image with overlapping text (b) image portion segmented out.(c) segmented out text portion.



(a) (b) (c)

fig 7 (a) texts with complex and overlapping background. (b) Segmented out text portion. (b) Image portion.

(a) (b)

fig 8 (a) image taken from ICDAR dataset (b) segmented text.

We have compared our algorithm with the method developed by Yi et. al. [3].[Fig 9] They also reported to have a failure is case ligature and multicolored text. But our algorithm can successfully detect text portion in spite of the above mentioned limitations. The results are shown in the fig 10.

6. Conclusion.

We present a wavelet packet based technique for extracting text having different characteristics. The comparative results suggest that our algorithm gives satisfactory performance in case of images having ligature and multicolored text. We plan to recognize the extracted text as a further scope of research.

(a) (b) (c)

fig 9(a) text in natural scene.(b) segmented text by proposed method(c)Segmented text by method in [3].

(a)

(b)

fig 10. (a) segmented out text in.(a) ligature.(b) multicolored text by proposed method.

7. References

[1] Balazs Enyedi . Lajos Konyha. Kalman Fazekas. Turan Jain: Character Localization in Video Sequences. C*ontent based multimedia indexing, 2008:504*-511 London. U.K 2008.

[2] P. Shivkumar, W. Huang and C. L. Tan : An efficient edge based technique for text detection in video frames. *The eighth IAPR workshop on document analysis systems.* 2008.

[3] Chucai Yi and Ying Li Tian : Text string detection from natural scenes by structure-based partition and grouping. *IEEE Transactions on Image processing,* 20(9):2594-2605. September 2011

[4] C. W. Liang and P. Y. Chen: DWT based text localization .*Int. J. Appl. Sci. Eng.* 2(1): 105-116.

[5] Mausumi Acharyya and Malay K. Kundu: Document Image segmentation using scale-space features. *IEEE Transactions on circuits and systems on video technology.* 12(12):1117-1127. December 2002.

[6] 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.* 16(8):2117-2128. August 2007.

[7] O. Alkin and H. Caglar: Design of efficient m-band coders with linear phase and perfect reconstruction properties. IEEE *Transactions on signal processing,*43(7):1579-1590, 1995.

[8] N. G. Kingsbury . Complex wavelets and shift invariance: In Proceedings IEEE Colloquium on Time-Scale and Time-Frequency Analysis and Applications.2000.

[9] Mausumi Acharyya and Malay K. Kundu: Image Segmentation Using Wavelet packet Frames and Neuro-Fuzzy Tools: *International Journal of Computational Cognition.*5(4):27-43 December 2007.