Rough-Fuzzy Clustering and M-Band Wavelet Packet for Text-Graphics Segmentation

Pradipta Maji, Shaswati Roy, and Malay K. Kundu

Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
{pmaji,shaswatiroy_t,malay}@isical.ac.in

Abstract. This paper presents a segmentation method, integrating judiciously the merits of rough-fuzzy computing and multiresolution image analysis technique, for documents having both text and graphics regions. It assumes that the text and non-text regions of a given document are considered to have different textural properties. The M-band wavelet packet is used to extract the scale-space features, which is able to zoom it onto narrow band high frequency components of a signal. A scale-space feature vector is thus derived, taken at different scales for each pixel in an image. Finally, the rough-fuzzy-possibilistic c-means algorithm is used to address the uncertainty problem of document segmentation. The performance of the proposed technique, along with a comparison with related approaches, is demonstrated on a set of real life document images.

1 Introduction

With the advances in information technology, automated processing of documents has become an imperative need. The documents in digitized form require a large amount of storage space, after being compressed using advanced techniques. Text-graphics segmentation partitions a document image into distinct regions corresponding to the text and non-text parts facilitating efficient searching and storage of the text parts in documents.

Many techniques have been proposed to segment the document image into text and non-text regions in the literature [1]. Recently, wavelet techniques have become powerful tools in this domain. Li and Gray [2] have used features based on distribution characteristics of wavelet coefficients in high frequency bands to segment document images into four classes, namely, background, photograph, text, and graph. Kundu and Acharyya [3] proposed a scheme for text-graphics segmentation based on wavelet scale-space features followed by k-means clustering. Lee et al. [4] used an algorithm based on local energy estimation in wavelet packet domain and k-means clustering.

In this paper, a text-graphics segmentation method is proposed, which integrates the principles of rough-fuzzy computing and multiresolution image analysis technique. This approach is based on the assumption that the text portion of the document image is comprised of one texture class and the non-text part of the other. The M-band wavelet packet (MWP) is used to extract the scale-space features, which offers a richer range of possibilities for document image

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 530–538, 2013.
© Springer-Verlag Berlin Heidelberg 2013

and is able to zoom it onto narrow band high frequency components of a signal. It yields a large number of subbands which are required for good-quality segmentation. Subsequently features are computed by using nonlinear energy estimation followed by a smoothing filter. However, one of the main problems in document image segmentation analysis is uncertainty. It includes incompleteness and vagueness in class definitions. To address this issue, an unsupervised clustering algorithm, termed as rough-fuzzy-possibilistic c-means (RFPCM), is used to segment the feature vectors. This approach needs not to assume any a priori information regarding the font size, scanning resolution and type of layout.

2 Rough-Fuzzy-Possibilistic C-Means Algorithm

Let $X = \{x_1, \cdots, x_j, \cdots, x_n\}$ be the set of n objects and $V = \{v_1, \cdots, v_i, \cdots, v_c\}$ be the set of c centroids and β_i be the ith cluster, where $x_j \in \mathbb{R}^m$ and $v_i \in \mathbb{R}^m$. Each cluster in the RFPCM algorithm [5] is represented by three parameters, namely, a cluster centroid, a crisp lower approximation and a fuzzy boundary. The centroid is calculated based on the weighting average of the crisp lower approximation $\underline{A}(\beta_i)$ and fuzzy boundary $B(\beta_i)$ as follows:

$$v_i = w \times \mathcal{C}_1 + \tilde{w} \times \mathcal{D}_1; \text{ where } \mathcal{C}_1 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j, \tag{1}$$

$$\mathcal{D}_1 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{m\acute{}_1} + b(\nu_{ij})^{m\acute{}_2}\} x_j, n_i = \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{m\acute{}_1} + b(\nu_{ij})^{m\acute{}_2}\}.$$

Here $1 \leq m\acute{}_1 < \infty$ and $1 \leq m\acute{}_2 < \infty$ are the fuzzifiers, $\mu_{ij} \in [0, 1]$ and $\nu_{ij} \in [0, 1]$ are the probabilistic and possibilistic membership functions, respectively, of the object x_j to the cluster β_i. The parameters w and $\tilde{w}(= 1 - w)$ correspond to the relative importance of lower approximation and boundary region, respectively. The constants a and $b(= 1 - a)$ define the relative importance of probabilistic and possibilistic memberships, respectively. The probabilistic and possibilistic membership values of an object x_j are calculated as

$$\mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}^2}{d_{kj}^2} \right)^{\frac{1}{m\acute{}_1 - 1}} \right)^{-1}; \text{ where } d_{ij}^2 = ||x_j - v_i||^2, \tag{2}$$

$$\text{and } \nu_{ij} = \frac{1}{1 + E}; \text{ where} E = \left\{ \frac{b||x_j - v_i||^2}{\eta_i} \right\}^{\frac{1}{m\acute{}_2 - 1}}. \tag{3}$$

The scale parameter η_i represents zone of influence of cluster β_i. In the RFPCM, the membership values of objects in lower approximation are $\mu_{ij} = \nu_{ij} = 1$, while those in boundary region are the same as fuzzy-possibilistic c-means. The main steps of the RFPCM algorithm are as follows

1. Assign initial centroids v_i, $i = 1, 2, \cdots, c$ and the fuzzifiers $\acute{m}_1$ and $\acute{m}_2$.
2. Compute μ_{ij} and ν_{ij} by (2) and (3), and finally, u_{ij} for c clusters and n objects where $u_{ij} = \{a\mu_{ij} + b\nu_{ij}\}$.
3. Calculate threshold δ, which determines the class labels of all objects as
$$\delta = \frac{1}{n}\sum_{j=1}^{n}(u_{ij} - u_{kj})$$
where n is the total number of objects, u_{ij} and u_{kj} are the highest and second highest memberships of x_j.
4. If $(u_{ij} - u_{kj}) \leq \delta$, then $x_j \in \overline{A}(\beta_i)$ and $x_j \in \overline{A}(\beta_k)$, where $\overline{A}(\beta_i)$ represents the upper approximation of cluster β_i. Furthermore, x_j is not part of any lower approximation region.
5. Otherwise, $x_j \in \underline{A}(\beta_i)$. In addition, by properties of rough sets, $x_j \in \overline{A}(\beta_i)$.
6. Modify μ_{ij} and ν_{ij} as 1 for the objects in lower approximations, while those in boundary regions are remain unchanged.
7. Compute new centroid as per (1).
8. Repeat steps 2 to 7, until no more new assignments can be made.

3 Feature Extraction

This section presents the feature extraction methodology that includes multi-channel filtering using the MWP with adaptive basis selection and subsequently local energy estimation and smoothing.

3.1 *M*-Band Wavelet Packet

In dyadic wavelet (2W) [6], scaling ϕ and wavelet functions ψ are defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k); \text{ and } \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{4}$$

for all $j, k \in \mathbb{Z}$. Here k determines the position along x-axis; j determines function's width, that is, how broad or narrow it is along x-axis.

The M-band orthonormal wavelet bases are constructed as a direct generalization of the 2W. The 2W decomposes a signal into frequency subbands that have the same bandwidth on a logarithmic scale, whereas M-band wavelet (MW), in addition, focuses on narrow band high frequency components of a signal, thereby simultaneously having a logarithmic and a linear decomposition of frequency channels. Let $\phi(t)$ be the scaling function satisfying

$$\phi(t) = \sum_k h_\phi(k)\sqrt{M}\phi(Mt - k). \tag{5}$$

Additionally, the $M - 1$ wavelets can be expressed as

$$\psi_l(t) = \sum_k h_\psi^l(k)\sqrt{M}\psi(Mt - k); l = 1, \cdots, M - 1, \tag{6}$$

where $h_\phi(n)$ and $h_\psi^l(n)$ are scaling and wavelet function coefficients, respectively. Scaling and translating the functions, $\phi(t)$ and $\psi_l(t)$, the $\phi_{j,k}(t)$ and $\psi_{l,j,k}(t)$ are obtained, respectively, as

$$\phi_{j,k}(t) = M^{j/2}\phi(M^j t - k), \tag{7}$$

$$\psi_{l,j,k}(t) = M^{j/2}\psi_l(M^j t - k); l = 1, \cdots, M-1. \tag{8}$$

For any given function $f(t) \in L^2(\mathbb{R})$, it can be shown that

$$f(t) = \sum_{k\in\mathbb{Z}} \langle f(t)\phi_{j,k}(t)\rangle \, \phi_{j,k}(t) + \sum_{l=1}^{M-1} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \langle f(t)\psi_{l,j,k}(t)\rangle \, \psi_{l,j,k}(t) \tag{9}$$

where $\mathbb{Z}$ represents the set of integers, $l = 1, \cdots, M-1$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}$ and $< ., . >$ represents the inner product.

An MW decomposes an image into M^2 subbands. Since in the MWP, at each decomposition level every subband is further decomposed, each of these M^2 subbands gives rise to another M^2 number of bases. So if the decomposition depth is p, then an MWP decomposition results in M^{2p} number of subbands, and this large number of bases are required for good quality segmentation.

3.2 Filtering and Adaptive Basis Selection

In the filtering stage, an eight-tap four-band orthogonal wavelet [7], as shown in Table 1, is used to decompose the document images into $M \times M$ channels without downsampling. The filter length is increased with the increasing level of decomposition. The filters are expanded by inserting appropriate number of zeros between taps of filters, thereby satisfying the quadrature mirror filter condition.

Table 1. Filter Coefficients for Eight-Tap Four-Band Wavelet

# of Taps (n)	$\phi(n)$	$\psi_1(n)$	$\psi_2(n)$	$\psi_3(n)$
0	-0.067371764	-0.094195111	-0.094195111	-0.067371764
1	0.094195111	0.067371764	-0.067371764	-0.094195111
2	0.40580489	0.56737176	0.56737176	0.40580489
3	0.56737176	0.40580489	-0.40580489	-0.56737176
4	0.56737176	-0.40580489	-0.40580489	0.56737176
5	0.40580489	-0.56737176	0.56737176	-0.40580489
6	0.094195111	-0.067371764	-0.067371764	0.094195111
7	-0.067371764	0.094195111	-0.094195111	0.067371764

In order to find out a suitable basis without going for a full decomposition, an adaptive decomposition algorithm using a maximum entropy or information content criterion extracted from each of the subbands is used [8]. After first level decomposition of the image into $M \times M$ channels, energy for each subband is evaluated. Among these subbands, those for which energy values exceed a threshold value (ϵ_1) of the parent band energy, are considered and decomposed further. A subband at second decomposition level is further decomposed if its energy value is more than another threshold value (ϵ_2) of the total energy of all the subbands at current scale. Hence, the number of subbands can be generated in this scheme in the range of 16 to 4096. Empirically it is seen that $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.10$ are good choice for the images considered in the experiment.

3.3 Local Energy Estimation and Smoothing

After the selection of significant bases, a local estimator, which constitutes a nonlinear operator followed by a smoothing filter, is applied to each subbands [3]. It gives high energy value for the regions in each subbands where frequency components are strong, otherwise low energy value is obtained where it is weak.

In this feature-extraction scheme, standard deviation is used as the nonlinear operator, calculated over small overlapping windows around each pixel. The local energy $E_b(x, y)$ around the (x, y)th pixel of bth subband is given as

$$E_b(x, y) = \sqrt{\frac{1}{R} \sum_{m=1}^{w} \sum_{n=1}^{w} \left| F_b(m, n)^2 - \overline{F}_b(x, y)^2 \right|} \tag{10}$$

where w is the window size and $R = w \times w$. $\overline{F}_b(x, y)$ is the mean around the (x, y)th pixel and $F_b(m, n)$ is the filtered image. Gaussian low-pass filter used as the smoothing filter is of the form

$$H_G(u, v) = \frac{1}{2\pi\sqrt{\sigma}} e^{-\frac{1}{2\sigma^2}(u^2+v^2)} \tag{11}$$

where σ determines the passband width of the averaging filter. Formally, the feature image $Feat_b(x, y)$ corresponding to subband image $F_b(x, y)$ is given by:

$$Feat_b(x, y) = \frac{1}{G^2} \sum_{(m,n) \in G_{xy}} \Gamma(F_b(m, n)) H_G(x - m, y - n) \tag{12}$$

where $\Gamma(.)$ gives the energy measure and G_{xy} is a $G \times G$ window centered at pixel with coordinates (x, y). It is found that an averaging window size of 9×9 to be appropriate in most of the segmentation experiment.

3.4 Choice of Energy Window Size

A nonlinear energy estimator is used in order to discriminate texture pairs as the unprocessed wavelet coefficients do not convey enough information for efficient representation of texture cues. In the present work, the energy window size is decided based on the measure of the edge density of the image as follows:

$$Den_e = \frac{no.\ of\ edge\ pixels}{total\ no.\ of\ pixels\ in\ image} \tag{13}$$

It gives the measure of overall image busyness. Sobel edge detector is used here to extract the edges. Den_e has a dynamic range of values between $[0, 1]$. For highly active image, Den_e is close to 1, so a smaller window for nonlinear operation is required. Moderately active images having the value of Den_e within $[0, 1]$ would require a moderate window size for good feature extraction. It has been found experimentally that the energy window size, for these different categories of image activities, ranges from 5×5 to 19×19.

4 Experimental Results

In this section, the performance of proposed text-graphics segmentation methodology is extensively compared with wavelet based different feature extraction techniques and several clustering algorithms. In the current study, features are extracted by the 2W decomposed upto third level, dyadic wavelet packet (2WP) decomposed upto third level, the MW as implemented in [3] and the proposed MWP. Daubechies 6 filter has been used for the feature extraction using the 2W and 2WP decomposition. The filter coefficients are shown in Table 2. The MW and MWP decompositions use the filter coefficients as depicted in Table 1. The clustering algorithms involve hard c-means (HCM), fuzzy c-means (FCM) [9], possibilistic c-means (PCM) [10], fuzzy-possibilistic c-means (FPCM), rough-fuzzy c-means (RFCM) [5], rough-possibilistic c-means (RPCM) [5] and RFPCM [5]. The values of parameters $\acute{m}_1 = \acute{m}_2 = 2.00$, $a = 0.50$ and $w = 0.95$.

Table 2. Filter Coefficients for Six-Tap Daubechies Wavelet

# of Taps (n)	$\phi(n)$	$\psi(n)$
0	0.3326705530	-0.0352262919
1	0.8068915093	-0.0854412739
2	0.4598775021	0.1350110200
3	-0.1350110200	0.4598775021
4	-0.0854412739	-0.8068915093
5	0.0352262919	0.3326705530

Some of the document images analyzed in the experiment are standard documents, others are scanned and taken online from parts of Anandabazar Patrika (www.anandabazar.com) and Times of India (timesofindia.indiatimes.com). Structured document images with nonoverlapping text and graphics regions are shown in Fig. 1(i) having size of 496×496, Fig 1(ii) of 256×256, Fig 1(iii) of 377×431 and Fig 1(iv) of 512×512.

Fig. 1(v) - Fig. 4 show the comparative analysis among text-graphics segmentation algorithms using different feature extraction methods to prove the efficacy of the proposed algorithm. From the results reported in Fig. 1(v) - Fig. 4, it is seen that there is a significant improvement in the segmentation results using the MWP compared to the classical 2W and 2WP, where $M = 2$. This may be explained by the significance of intermediate frequency bands obtained using the MWP decomposition in characterizing the textural features. Here, in these figures, the performance of different clustering techniques is also analyzed. It is found that rough set based clustering approaches are yielding good segmentation results over non-rough set based algorithms, irrespective of feature extraction techniques and images. Among all rough set based clustering methods, the RFPCM gives excellent results as far as text identification is concerned.

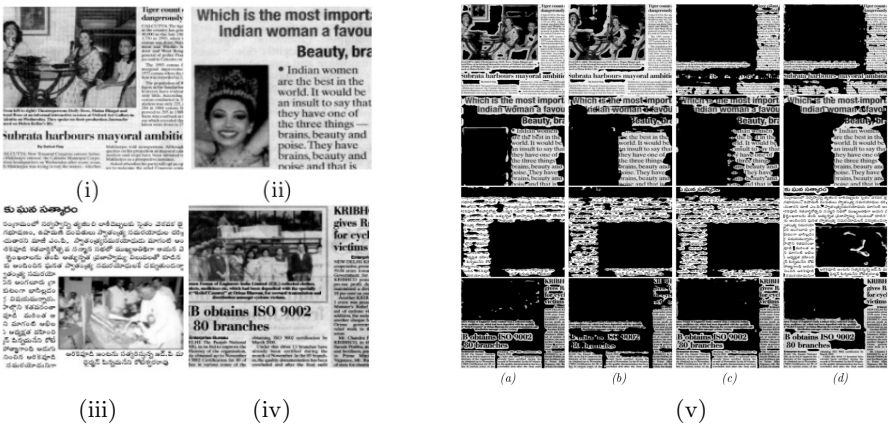

Fig. 1. (i)-(iv) Input data set for document image segmentation. (v) Text obtained using HCM: (a) 2W (b) 2WP (c) MW (d) MWP.

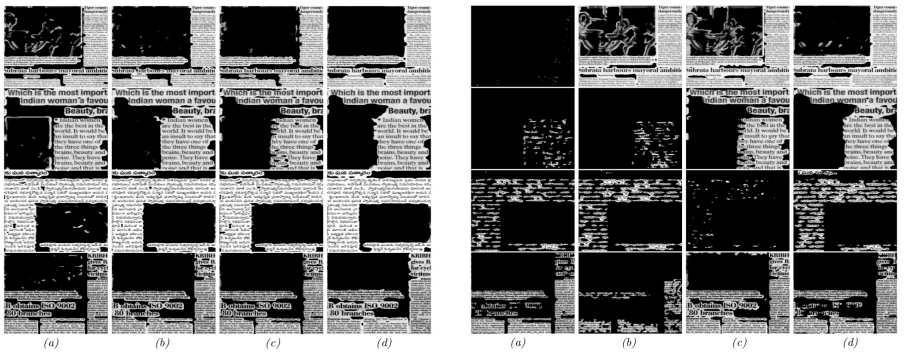

Fig. 2. Text obtained using FCM and PCM: (a) 2W (b) 2WP (c) MW (d) MWP

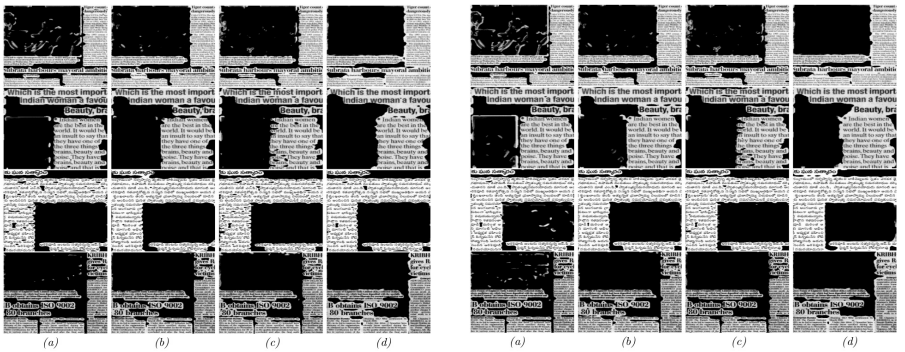

Fig. 3. Text obtained using FPCM and RFCM: (a) 2W (b) 2WP (c) MW (d) MWP

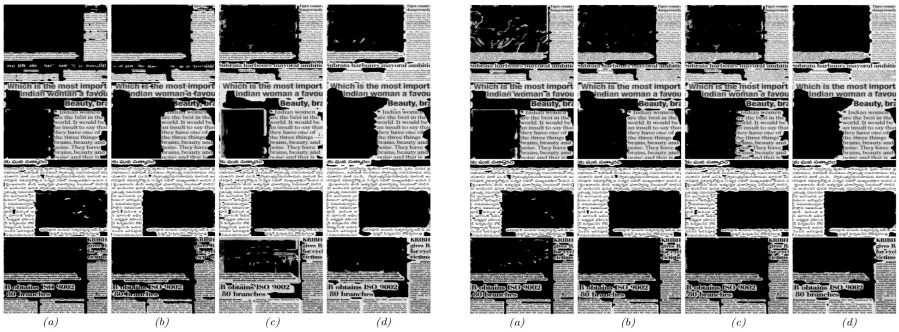

Fig. 4. Text obtained using RPCM and RFPCM: (a) 2W (b) 2WP (c) MW (d) MWP

5 Conclusion

In this paper, a new texture-based methodology is presented for segmenting the text part from the graphics part based on textural cues. The use of wavelet theory via MWP decomposition of images provides a multiscale multidirectional representation of the image and yields a huge number of frequency channels and hence facilitates an improved segmentation of the different class regions. The RFPCM is geared towards maximizing the utility of both rough sets and fuzzy sets with respect to uncertainty handling. An extensive comparative study with different feature extraction techniques and clustering approaches shows that the proposed methodology is indeed effective in characterizing document images in a better way.

References

1. Srihari, S.N.: Document Image Understanding. In: Proc. Fall Joint Computer Conference, pp. 87–96 (1986)
2. Li, J., Gray, R.M.: Context-Based Multiscale Classification of Document Images Using Wavelet Coefficient Distributions. IEEE Trans. Image Processing 9(9), 1604–1616 (2000)
3. Acharyya, M., Kundu, M.K.: Document Image Segmentation Using Wavelet Scale-Space Features. IEEE Trans. Circuits and Systems for Video Technology 12(12), 1117–1127 (2002)
4. Lee, G.B., Odoyo, W.O., Lee, J.H., Chung, Y., Cho, B.J.: Two Texture Segmentation of Document Image Using Wavelet Packet Analysis. In: Proc. 9th Intl. Conf. Advanced Communication Technology, vol. 1, pp. 395–398 (2007)
5. Maji, P., Pal, S.K.: Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices. IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics 37(6), 1529–1540 (2007)
6. Mallat, S.G.: Multifrequency Channel Decompositions of Images and Wavelet Models. IEEE Trans. Acoustics, Speech and Signal Processing 37(12), 2091–2110 (1989)

7. Alkin, O., Caglar, H.: Design of Efficient M-Band Coders With Linear-Phase and Perfect-Reconstruction Properties. IEEE Trans. Signal Processing 43(7), 1579–1590 (1995)
8. Acharyya, M., Kundu, M.K.: Image Segmentation Using Wavelet Packet Frames and Neuro-Fuzzy Tools. Intl. Jrnl. Computational Cognition 5(4), 27–43 (2007)
9. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers Norwell, MA (1981)
10. Krishnapuram, R., Keller, J.M.: A Possibilistic Approach to Clustering. IEEE Trans. Fuzzy Systems 1(2), 98–110 (1993)