2009 10th International Conference on Document Analysis and Recognition

Steerable pyramid based complex documents images segmentation

Mohamed Benjelil [1], Slim Kanoun [1], Rémy Mullot [2], Adel M. Alimi [1]
[1] *REGIM – ENIS, B.P. 1173, 3038, Sfax, Tunisia*
[2] *L3I, University of La Rochelle, Avenue Michel Crépeau, 17042. La Rochelle, France*

Abstract

In this paper, we propose an accurate and suitable designed system for complex documents segmentation. This system is based on steerable pyramid transform. The features extracted from pyramid sub bands serve to locate and classify regions into text and non text in some noise infected, deformed, multilingual, multi script document images. These documents contain tabular structures, logos, stamps, handwritten text blocks, photos etc. The encouraging and promising results obtained on 1000 official complex documents images data set are presented in this research paper.

1. Introduction

Facing the very important mass of information exchanged between the different organizations, the need of systems allowing the recognition, the indexation, the information retrieval and the automatic classification of complex multi-lingual and multi-script document images has grown continuously. Most works of retro-conversion of printed Arabic document images are limited to the textual block recognition without treating complex documents such as letters of information, forms, all type of demands, etc.

In practice, these documents can be noised, skewed, deformed, multi-lingual, multi-script figure(5), with irregular textures and may contain several heterogeneous blocks such as texts (printed and/or handwritten), graphics, pictures, logos, photographs, tabular structures. This situation makes it difficult to analyze and recognize document images.

In this paper, we essentially focus our interest on the segmentation of complex document images on text and non text microstructures by proposing a new texture descriptor based on Steerable Pyramid. Our motivation in using Steerable Pyramids relies not only on the fact that they have demonstrated discrimination properties for texture characterization [1], but also that unlike other image decomposition methods, the feature coefficients are less modified under the presence of image rotations, or even scales.

The remaining of our paper is organized as follows: First we present the existing techniques for the segmentation of document images.

Then, we expose our complex document images segmentation contributions as well as the experimental results obtained on a basis of 1000 complex multi-lingual and multi-script document images. We end up our paper by drawing conclusions and suggesting perspectives.

2. Related works

In the last decade, several works have been proposed for the segmentation of document images. The segmentation process produces a hierarchical structure that captures the physical layout and the logical meaning of the input document image [2]. The top of this structure presents an entire page and the bottom includes all glyphs on the document.

The text blocks, lines, words and characters are placed at different levels in the structure.

Techniques for page segmentation and layout analysis are broadly divided in to three main categories: top-down, bottom-up and hybrid techniques [3].

Top-down techniques start by detecting the highest level of structure (large scale features like images, columns) and proceed by successive splitting until they reach the bottom layer (small scale features like individual characters). For this type of procedures, a priori knowledge about the page layout is necessary. It relies on methods such as Run-length smoothing [4] Projection profile methods [5], white streams [6], Fourier Transform [7], Template [8], Form Definition Language [9], Rule-based systems [5] etc.

Bottom-up methods start with the smallest elements (pixels), merging them recursively in connected components or regions (characters and words), and then in larger structures (columns). They are more flexible but may suffer from accumulation of errors. It makes use of methods like Connected Component Analysis [10], Run-Length smoothing [4], [23], Region-growing methods [10], Neighborhood-Line density [12] and regions classification by neural networks [13]. Most of these methods require high computation.

Hybrid methods combine bottom-up with top-down processing. Among these methods are Texture-based [14], Gabor Filter [15].

Texture analysis has been an active research topic, and numerous methods have been proposed in the open

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.288

literature [16]. A large number of texture features have been proposed. In [17], these features are divided into four major categories, namely, statistical, geometrical, model-based and signal processing features. Model-based methods have been employed in texture analysis [18], including autoregressive model, Gausian Markov random fields, Gibbs random fields, Wold model, wavelet model, multichannel Gabor model and steerable pyramid.

Although several methods have achieved high content-based image retrieval rates [19, 20, 21], some of them were evaluated under controlled scenarios. In this context, the next challenge consists in achieving rotation, and scale-invariant feature representations for non-controlled environments.

3. Steerable pyramid (S.P)

The Steerable Pyramid [22], is a linear multi-scale, multi-orientation image decomposition, that provides a useful front-end for image-processing and computer vision applications. The S.P can capture the variation of a texture in both intensity and orientation.

The synoptic diagram for the decomposition (both analysis and synthesis) is shown in (Figure.1). Initially, the image is separated into low and high pass sub bands, using filters L_0 and H_0. The low pass sub band is then divided into a set of oriented band pass sub bands and a lower pass sub band. This lower pass sub band is sub sampled by a factor of 2 in the X and Y directions. The recursive (pyramid) construction of a pyramid is achieved by inserting a copy of the shaded portion of the diagram at the location of the solid circle.

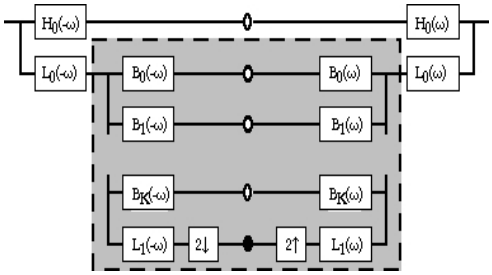

Figure 1.First level of steerable pyramid decomposition [22]

The basic functions of the steerable pyramid are directional derivative operators that come in different sizes and orientations. The necessary conditions for a filter basis to be steerable, is the ability to synthesize a filter of any orientation from a linear combination of filters at fixed orientations (Figure 2). The simplest example of this is oriented first derivative of Gaussian filters, at 0^o and 90^o:

$\alpha_1 = 0^o, \alpha_2 = 90^o$

The steering equation:

$$G_1^{\alpha}(x,y) = \cos(\alpha)G_1^{0^\circ}(x,y) + \sin(\alpha)G_1^{90^\circ}(x,y)$$

We can synthesize a filter at any orientation by linear combination of filters $G_1^{0^\circ}$ and $G_1^{90^\circ}$. Wecan synthesize an image at any orientation by linear combination of the convolution of that image with the filters $G_1^{0^\circ}$ and $G_1^{90^\circ}$:

For $R_1^{0^\circ} = G_1^{0^\circ} * I$ and $R_1^{90^\circ} = G_1^{90^\circ} * I$

The resulting image is

$$R_1^{\alpha}(x,y) = \cos(\alpha)G_1^{0^\circ}(x,y) + \sin(\alpha)G_1^{90^\circ}(x,y)$$

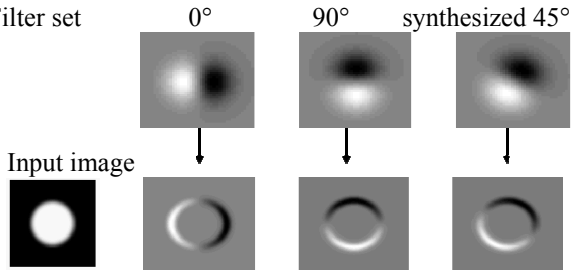

Figure 2.Filters combination [1]

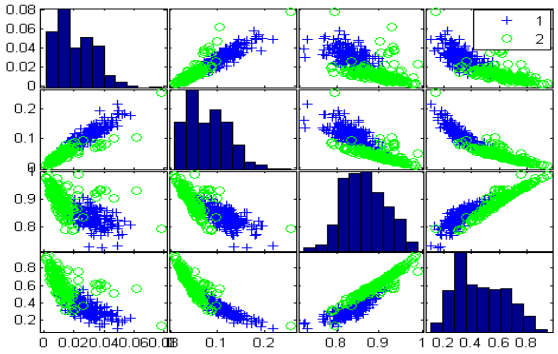

Figure 3. Training data set features scatter plot 1- Latin text bloc 2-Non text bloc

We tested the S.P on 300 printed Latin text bloc and 300 non text objects. This particular steerable pyramid contains 4 orientation sub bands, at 2 scales each. For each image sub bands, we calculated the variance, the mean, the homogeneity and energy. The scatter plots, (Figure 3), show clearly the effectiveness and the distinguishing capability of the S.P. Since complex document images are basically constituted by such types of textures, we decided to use the S.P to segment them.

4. S.P based complex document images segmentation strategy

We consider text and non text as regions with different textures. Since the distinguishing characteristics of text are frequency information, orientation, approximately with the same size and line thickness, located at a regular distance from each other, we can use them to characterize text regions with steerable pyramid decomposition. (Figure 4) shows the synoptic diagram of proposed system.

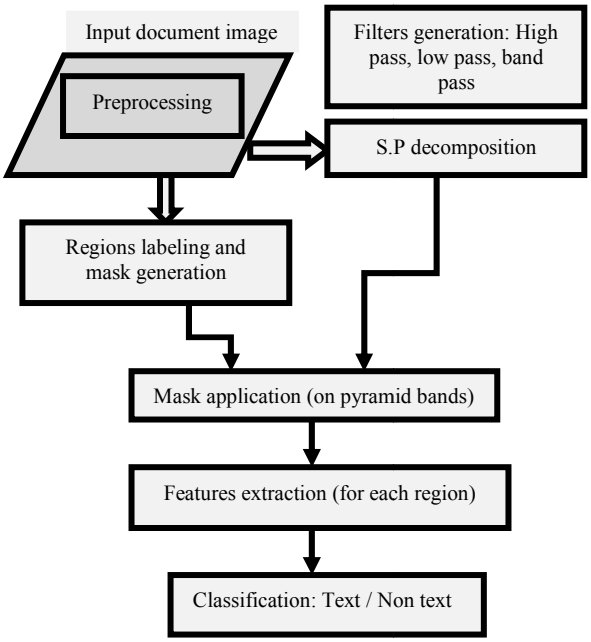

Figure 4.Synoptic diagram of proposed system

Preprocessing: This is a crucial step because of the poor quality of treated document images. The basic idea is the use of morphological operators that remove isolated small objects without removing texts diacritic points (Figure 5a. input image), (Figure 5b. filtered image).

Regions labeling, mask generation and application: Since text shows spatial cohesion (characters appear in clusters at a regular distance aligned to a virtual line). By merging characters inside each cluster by image dilatation with suitable horizontal and vertical structuring elements we can create a document mask.

This mask will serve to extract corresponding regions from steerable pyramid sub bands. Each region will be classified as text or non text depending on its feature values (Figure 5c. S.P decomposition), (Figure 5d. Mask application bounding boxes).

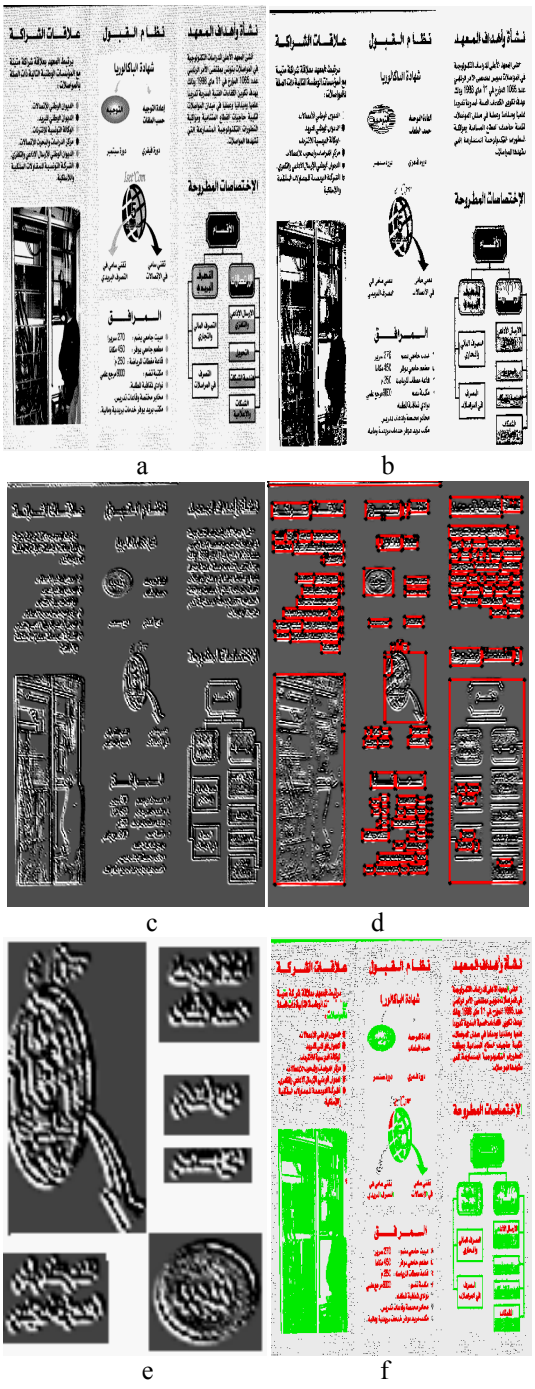

Figure 5.Segmentation process a) Input image b) Filtered image c) S.P decomposition d) Mask application bounding boxes e) Samples of extracted regions f) Final result with Red as text, Green as non text and gray as noise.

Features extraction: These features are defined as statistics of steerable pyramid and form the feature vector as follows:

- Mean of each sub band
- Variance of each sub band
- Homogeneity of each sub band
- Energy of each sub band

Classification: To classify the extracted regions into text or non text classes, we test two classification methods.

In the first method, we use a non-supervised k-means classifier with two classes based on regions detected in a single document, one for the text and the other for the non-text. We consider the class that includes more items as the text class. This is justified by the fact that in a document, the number of textual regions is greater than the number non-textual regions.

In the second method, we use a supervised K nearest neighbor's classifier (KNN) with different values of K. In this framework, starting with 4664 text block images and 1258 graphic blocks, we choose 600 text block images and 600 graphic blocks as training data set (Figure .5e samples of extracted regions), (Figure .5f final result).

5 Results and discussion

To validate the proposed system, we create a data set of 1000 complex document images (350 Dispatch notes, 350 Forms and 300 magazines). The choice of this kind of documents is justified by the fact that our system is designed to work in real official environment.

The S.P parameters tested are sp0filter, sp3filter, sp5filter with respectively 2, 4, 6 orientations and 1, 2, 3 and 4 levels [22].

We used a k-means classifier with two classes and KNN classifier with K=3, 5 and 7.

In order to evaluate our algorithm, we adopt a quantitative method based on expert report according to the following formulas:

$$R = cdo/eo$$
$$P = cdo/do$$
$$CSR = (p.r)/(p + r)$$

With: cdo: Number of correctly detected objects, eo: Estimated number of objects, do: Number of detected objects, R: Recall, P: Precision, CSR: Correct Segmentation Rate.

We found that the best segmentation rates are obtained by sp3filter with 4 orientations and 2 levels. The Table 1, Table 2 and Table 3 show that the segmentation rate obtained by k-means classifier is better than the one obtained by the KNN classifier. KNN classifier gives best results with K=5. The segmentation rate is about 93.44 %. This is reasonable at this stage and we can afford adding some preprocessing and more improved feature selection. The quality and the reliability of our segmentation method are sensitive to some kinds of acquired noises that form a certain texture which confounds itself with the texture of a textual zone of the document image.

Table 1.Correct classification rates

Classifier	R	P	CSR
k-means	94.83 %	92.33 %	93.44%
knn	89.26 %	86.79 %	88.56%

Table 2.Correct segmentation rates by type of document with k-means classifier

Documents	Nb	R	P	CSR
Dispatch notes	350	93.80%	94.60%	94.16%
Forms	350	99.20%	93.90%	96.46%
magazines	300	91.50%	88.50%	89.70%

Table 3.Correct segmentation rates by type of document with KNN classifier with K=5.

Documents	Nb	R	P	CSR
Dispatch notes	350	88.98%	88.92%	88.51%
Forms	350	93.24%	88.26%	91.04%
magazines	300	85.58%	83.19%	84.31%

Limits: One shortcoming of our algorithm is that it fails to locate text blocks that are not well separated from the background or linked to graphics. The classification failed when some words or text blocks have size and shape which look like graphics. This is the case of titles, with large font size, which confound themselves with graphics.

Comparison: In order to evaluate the performance of our system, we compared it with Gabor filter bank designed for image texture segmentation. In this comparison, we used three datasets namely: dispatch notes, forms and magazines. The classification results for the two multi-resolution decompositions for the three datasets are presented in Table 4.

From Table 4, we can see that the steerable pyramid with sp3filters orientations achieves the highest accuracy for the second dataset. The Gabor wavelet follows it with Sx=4, Sy= 8, f=18, thêta=0, 45, 90,135 on the same data set. For dataset 1 and 3, the steerable pyramids with sp3filters have the best classification accuracy.

Table 4.Steerable pyramid Vs Gabor filter bank

Methods	dispatch notes	forms	magazines
S.P sp1Filter , level 2	86.51%	84.51%	85.56%
S.P sp3Filter , level 2	94.16%	96.46%	89.70%
S.P sp5Filter , level 2	92.08%	92.08%	87.62%
Gabor Sx=4, Sy= 8, f=6, thêta=0,45,90,135	83.60%	85.64%	82.70%
Gabor Sx=4, Sy= 8, f=18, thêta=0,45,90,135	89.87%	90.87%	86.72%
Gabor Sx=4, Sy= 8, f=30, thêta=0,45,90,135	86.30%	89.30%	84.88%

8. Conclusion and future work

The work developed in this paper aims at setting up a system of segmentation of complex multilingual multi script document images. Thus, we began with a study from the existing systems of document images segmentation. Within this framework, we showed a few systems that treated complex multilingual multi script document images. We presented the proposed system which is based on steerable pyramid. Lastly, we exposed the results obtained on a data set of 1000 documents. The rate of correct segmentation obtained is about 93.44 %.

References

[1] W. T. Freeman, E. H. Adelson The design and use of steerable filters. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 13, no. 9, pp. 891-906, September,1991.

[2] J. Liang, I.T. Phillips, R.M. Haralick *A Statically Based, Highly Accurate Text-line Segmentation Method.* Proc. 5th Intl. Conf. on Document Analysis and Recognition. ICDAR'99, Bangalore, India, pp. 551.

[3] O. Okun, D. Doermann, Matti P. *Page Segmentation and zone classification.* The State of the Art, Nov 1999.

[4] J. Kanai, M.S. Krishnamoorthy, T. Spencer *Algorithm for Manipulating nested block represented images.*SPSE's 26th Fall Symposium, Arlington VA, USA, Oct 1986, pp.190-193.

[5] K.H. Lee, Y.C. Choy, S. Cho *Geometric Structure Analysis of Document Images: A Knowledge-Based Approach*. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, Nov 2000, pp. 1224-1240.

[6] T. Pavlidis, J. Zhou *Segmentation by White Streams.*Proc. Intl. Conf. on Document Analysis and Recongition, ICDAR'91, St-Malo, France, pp. 945-953.

[7] M. Hose, Y. Hoshino *Segmentation method of document images by two-dimensional Fourier transformation.* System and Computers in Japan, Vol. 16, No. 3, 1985, pp. 38-47.

[8] A. Dengel and G. Barth. *Document description and analysis by cuts.* Proc. Conference on Computer- Assisted Information Retrieval , MIT USA, 1988.

[9] H. Fujisawa and Y. Nakano. *A top-down approach for the analysis of document images.* Proc. of Workshop on Syntactic and Structural Pattern Recognition (SSPR' 90), 1990, pp. 113-122.

[10] J. P. Bixler. *Tracking text in mixed-mode document.*Proc. ACMConference on Document Processing System, 1998, pp. 177-185.

[11] A.K. Jain, Fundamentals of Digital Image Processing. Prentice Hall USA, 1989.

[12] O. Iwaki, H. Kida and H. Arakawa. *A character / graphics segmentation method using neighborhood line density.* Trans. of Institute of Electronics and Communication Engineers of Japan, 1985, Part D J68D, 4, pp. 821 828.

[13] C.L. Tan, Z. Zhang *Text block segmentation using pyramid structure.* SPIE Document Recognition and Retrieval, Vol. 8, January 24-25, 2001, San Jose, USA, pp. 297-306.

[14] D. Chetverikov, J. Liang, J. Komuves, R.M. Haralick. *Zone classification using texture features.* Proc. Of Intl. Conf. on Pattern Recognition, Vol. 3, 1996, pp. 676-680.

[15]A. K. Jain and S. Bhattacharjee, *Text SegmentationUsing Gabor Filters for Automatic Document Processing,* Machine Vision and Applications, Vol. 5, No. 3, 1992, pp. 169-184.

[16] Reed, T.R., Du Buf, J.M.H., 1993. A review of recent texture segmentation and feature extraction techniques. CVGIP: Image Understanding 57 (3), 359–372.

[17] Tuceryan, M., Jain, A.K., 1998. Texture analysis. In: Chen, L.F., Pau, L.F., Wang, P.S.P. (Eds.), Handbook of Pattern Recognition and Computer Vision, second ed. World Scientific, Singapore.

[18] Zhang, J.G., Tan, T.N., 2002. Brief review of invariant texture analysis methods. Pattern Recognition 35 (2), 735–747.

[19] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. IEEE Transactions on Image Processing, 11(2):146– 158, 2002.

[20] P. W. Huang, S. K. Dai, and P. L. Lin. Texture image retrieval and image segmentation using composite sub-band gradient vectors. J. Visual Communication and Image Representation, 17(5):947–957, 2006.

[21] M. Kokare, B. N. Chatterji, and P. K. Biswas. Cosinemodulated wavelet based texture features for content-based image retrieval. Pattern Recognition Letters, 25(4):391–398, 2004.

[22] Simoncelli, E.P., Freeman, W.T., 1995. The steerable pyramid: A flexible architecture for multi-scale derivative computationIn: Proc. IEEE Second Internat. Conf. on Image Process. Washington, DC, pp. 444–447.

[23] M. Benjlaiel, S. Kanoun, A. Alimi, « Une méthode de segmentation d'Images de Documents Composites », 9*ème Colloque International Francophone sur l'Ecrit et le Document : CIFED'2006, Semaine du Document Numérique : SDN'2006,* pp. 121-126, 18-21 Septembre 2006, Fribourg, Suisse.