

MULTISCALE PAGE SEGMENTATION USING WAVELET PACKET ANALYSIS

P. GÓRECKI[†], L. CAPONETTI* AND C. CASTIELLO*

[†] *Wydział Matematyki i Informatyki, Uniwersytet Warmińsko-Mazurski
ul. Oczapowskiego 2, 10-719 Olsztyn, Poland
E-mail: pgorecki@matman.uwm.edu.pl*

**Dipartimento di Informatica, Università degli Studi di Bari,
Via E. Orabona 4, 70126 Bari, Italy
E-mail: [laura, castiello]@di.uniba.it*

Abstract.

In this paper, a novel method for document page segmentation using Wavelet Packet analysis is proposed. To discriminate between text and non-text regions, the image is represented by means of a wavelet packet analysis tree. Successively a feature image is introduced to synthesize the information related to some nodes selected from the quadtree. The most discriminant nodes are derived using an optimality criterion and a genetic algorithm. Finally the selected feature image is segmented by means of a Fuzzy C-Means clustering. The approach provides good segmentation results and shows to be invariant to page skew and font variations.

Keywords: Document Image, Segmentation, Wavelet Packet Analysis, Genetic Algorithm, Fuzzy C-Means.

1. Introduction

The task of document image segmentation is to partition document page images into separated regions corresponding to text blocks, graphical images, charts, drawings, and so on. Successively, the regions extracted from a document page can be processed to represent the page in a more suitable form for efficient compression, text querying and re-editing.

Until now, many different techniques for document image segmentation have been proposed in the literature.²³ In general, they can be divided into top-down^{18,14,15,10,24} and bottom-up^{12,20,9,1} approaches. Classical top down approaches are based on run-length encoding and projection profiles: with these methods a page is first split into blocks, which are successively identified and subdivided into paragraphs, text lines, words and finally letters. These methods are sensitive to skewed text and perform well with highly structured page layouts. On the contrary, bottom-up approaches start from the pixel level analysis in order to merge similar pixels into higher level components like letters, words, text lines and paragraphs. The drawback of this last category of approaches is their sensitivity to font size, scanning resolution, interline and inter-character spacing.

To reduce errors in page segmentation due to different resolutions, font sizes and spacings, some authors introduced multi-resolution techniques based on wavelet analysis. For example Deng et al.⁸ propose the use of polynomial spline wavelets and Xi et al.²⁵ present

a multi-resolution approach for the extraction of reference lines and items from document forms.

In this paper we propose a methodology for page segmentation into text and non-text regions based on Discrete Wavelet Packet Transforms (DWPT).²² We assume that regions containing text and non-text have different frequency characteristics. In general, text regions are characterized by high frequencies due to recurring intensity alterations between background and characters. On the other hand, non-text regions – like halftones or background – are characterized by lower frequencies. DWPT analysis provides a powerful tool to locate signals in frequency and in time domains simultaneously. The aim of this work is to select the most discriminative features for the problem of image segmentation, by analyzing the features obtained by the packet wavelet transforms and represented by a quadtree called wavelet packet subspace analysis tree. Features selected from some nodes of the analysis quadtree, are synthesized in an image, called feature image. The most discriminant nodes are derived applying a genetic algorithm to the corresponding feature image. The fitness function of the genetic algorithm is based on an optimality criterion derived from the Fisher’s criterion.²¹

This paper is organized as following. Sections 2 briefly introduces wavelet packet analysis, which is used for the feature extraction process presented in section 3. Section 4 describes the segmentation process of a page and section 5 presents details about the experimental session and the obtained results. Section 6 closes the paper with some conclusive remarks.

2. Wavelet packet decomposition

Wavelet packet analysis is an important generalization of wavelet analysis.^{5,4,6} Wavelet packet functions are localized in time such as wavelet functions, but offer more flexibility than wavelets in representing different types of signals. Wavelet packet approximators are based on translated and scaled wavelet packet functions $W_{j,b,k}$, which are generated from the following base function:^{7,13}

$$(1) \quad W_{j,b,k}(t) = 2^{j/2}W_b(2^{-j}(t - k)),$$

where j is the resolution level, W_b is wavelet packet functions generated by scaling and translating mother wavelet function ψ , b is the number of oscillations (zero crossings) of W_b and k is the translation shift. In wavelet packet analysis, a signal $x(t)$ is represented as a sum of orthogonal wavelet packet functions $W_{j,b,k}(t)$ at different scales, oscillations and locations:

$$(2) \quad x(t) = \sum_j \sum_b \sum_k w_{j,b,k}W_{j,b,k}(t).$$

where $w_{j,b,k}$ is the wavelet packet coefficient. To compute the wavelet packet coefficients a fast splitting algorithm³ is used, which is an adaptation of the pyramid algorithm¹⁶ for discrete wavelet transform. The splitting algorithm differs from the pyramid algorithm since low-pass and high-pass filters are applied to the detailed coefficients in addition to the approximation coefficients at each stage of the algorithm. Moreover, the splitting algorithm retains all the coefficients, including those at intermediate filtering stages.

If we consider an image $I(x, y)$, the DWPT decomposes the image into a series of band-limited components, called sub-bands. The discrete wavelet transform can be computed

using filter banks that decompose the image into low-frequency components (approximation) and high-frequency components (detail). The filter banks are composed by orthogonal and finite support filters: a low-pass filter L and a high-pass filter H . Given the filters, the image coefficients at the decomposition level $j - 1$ are calculated by passing the image coefficients from level j through a low-pass filter along the rows of the image coefficients. As a consequence, low-pass and high-pass images are obtained which undergo the same process operating along the columns of the image coefficients, producing four decomposition sub-images: the approximation image x_{LL}^{j-1} and the detail images x_{LH}^{j-1} , x_{HL}^{j-1} , x_{HH}^{j-1} . The detail images contain the coefficients related to frequency components of different orientations: horizontal, vertical and diagonal respectively. This process is recursively repeated for all the obtained sub-images.

The entire decomposition process can be represented with a quadtree in which the root node is assigned to the highest scale approximation coefficients, that are the original image itself, while the leaves represent outputs of the LL, LH, HL and HH filters.

Assuming that similar regions of an image have similar frequency characteristics, we infer that these characteristics are captured by some nodes of the quadtree. As a consequence, the proper selection of nodes should allow for localization of similar regions in the image.

3. Feature extraction

Wavelet packet decomposition divides the frequency space into various sub-bands and allows better frequency localization of signals than standard Wavelet decomposition. In the context of page segmentation the goal of the feature extraction process is to obtain the basis of the wavelet sub-bands with the highest discrimination power between text and non-text regions. In this section we propose our methodology for obtaining such a basis by the analysis of the quadtree obtained applying the wavelet packet transform to a given image. In particular, we describe a method for selecting the most discriminative nodes n_i of the subspace analysis tree τ . Successively, the selected nodes will be used in the page segmentation stage.

The initial step consists in representing the given image as a tree τ , using the DWPT. An example of decomposition is depicted in figure 1: the wavelet coefficients of the nodes in the tree are displayed as sub-images. From the analysis of the figure it can be observed that some of the sub-images display a better discriminating configuration of text and non-text regions.

To quantitatively evaluate the effectiveness of the node $n_i \in \tau$ in discriminating between text and non-text, the following procedure is performed. Firstly, the wavelet coefficients of the i -th node c_i are represented in terms of absolute values $|c_i|$, because discrimination power does not depend on the coefficient signs. Then, the coefficients are divided into the sets T_i (text coefficients) and N_i (non-text coefficients), on the basis of the known ground truth segmentation of the given image. Successively, for each set T_i and N_i , the mean and variance values are calculated, denoted as μ_i^T and σ_i^T for text and μ_i^N and σ_i^N for non-text, respectively. Next, the discrimination power F_i of the node n_i is evaluated using an optimality criterion (based on the Fisher's criterion):²¹

$$(1) \quad F_i = \frac{(\mu_i^T - \mu_i^N)^2}{\sigma_i^T + \sigma_i^N}.$$

To a certain extent, the value provided by equation (1) represents a measure of the

signal-to-noise ratio for the text and non-text classes. The nodes with maximum distance between two classes and minimum variance within each class have the highest separation value.

The simplest approach to obtain the best nodes subset $v \subset \tau$ is to select the small number of nodes from τ with the highest discrimination power. To employ the information encompassed in v , we define a feature image $f(x, y)$ by combining coefficients from the subset v . This process involves rescaling and adding together wavelet coefficients from the selected nodes:

$$(2) \quad f(x, y) = \sum_{i \in v} c'_i(x, y),$$

where $c'_i(x, y)$ denotes the $|c_i|$ values rescaled to match the size of the original image $I(x, y)$. Even if the approach for obtaining v is fast and simple, it is not an optimal technique to maximize signal-to-noise ratio between text and non-text classes. Moreover, the optimal number of nodes in v to be chosen is unknown and it must be selected manually.

The problem of selecting the best nodes from all nodes available is a combinatorial problem, producing an exponential explosion of possible solutions. We solve this problem adopting a genetic algorithm.¹⁷ For this aim, each node $n_i \in \tau$ is associated with a binary weight $w_i \in \{0, 1\}$, so the tree τ is associated with a vector of weights $W = [w_1, \dots, w_i, \dots, w_{|\tau|}]$. Consequently, the subset of the best nodes is defined as $v = \{n_i \in \tau : w_i = 1\}$.

Given the weight of each node, the feature image is calculated as following:

$$(3) \quad f(x, y, W) = \sum_{i=1}^{|\tau|} w_i c'_i(x, y).$$

The discrimination power F of the subset v can be computed extending the equation (1), by evaluating the mean values μ^T , μ^N and the deviation values σ^T , σ^N of the coefficients corresponding to text regions and non-text regions:

$$(4) \quad F = \frac{(\mu^T - \mu^N)^2}{\sigma^T + \sigma^N}.$$

To find the optimal subset v by means of equation (3), a genetic algorithm is applied in order to maximize the cost function F . Initially a random population of K weight vectors $\{W_i : i = 1, \dots, K\}$, represented as binary strings, is created. Successively, for each weight vector the feature image is calculated and its cost function is evaluated using the equation (4). The best individuals are subject to crossover and mutation operators in order to produce the next generation of weight vectors. Finally, the optimal subset v is found from the best individuals in the evolved vector population.

4. Page segmentation

The subset v of the nodes obtained during the feature extraction process is used to partition a page into different regions, labelled as text or non-text regions. This task is performed by decomposing the page into a sub-space tree τ using the wavelet functions and the tree depth described in section 3. Successively, the feature image $f(x, y)$ is evaluated by merging the node coefficients using the equation (2).

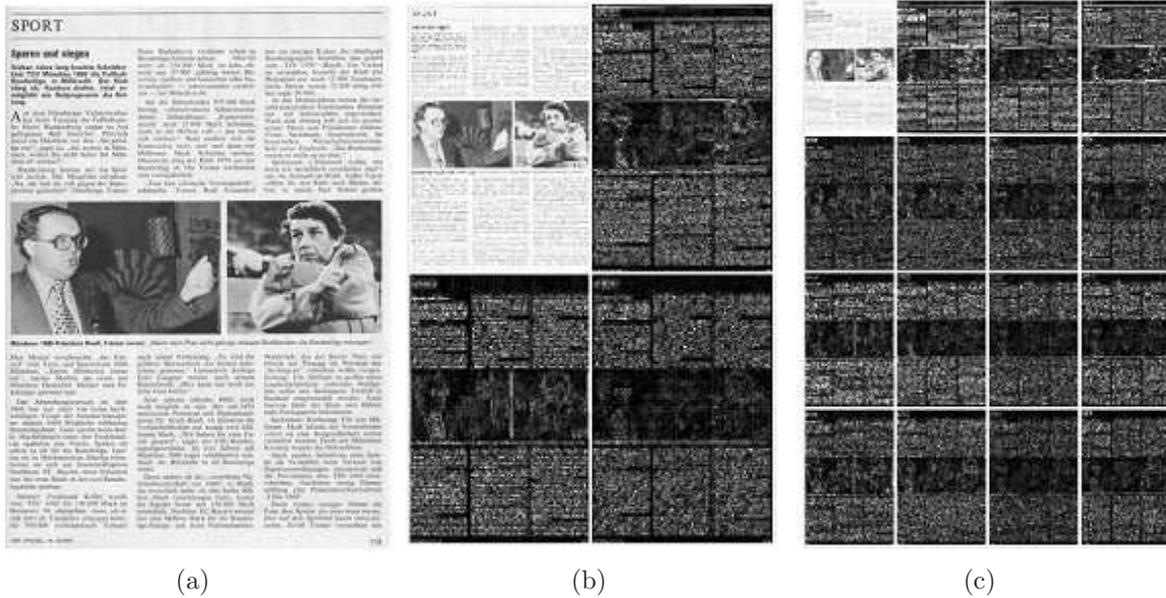


Fig. 1. DWPT decomposition of the image (a) at levels 1–2 (b–c). Each subimage in (b–c) is a different node of the DWPT tree.

In order to segment a document image, its feature image can be exploited in various ways. One of the simplest approaches is image thresholding; another possibility is to apply a clustering algorithm. In our approach we propose the use of a Fuzzy C-Means algorithm² to group pixels of $f(x, y)$ into two clusters corresponding to text and non-text regions, respectively. Replacing each pixel of $f(x, y)$ with its cluster label leads to the segmented image.

As the clustering is not performed in the image space but not in the feature space, additional post processing is necessary to refine segmentation, i.e. removing isolated pixels or filling small holes. This is achieved by de-noising the segmented image with median filter, followed by a morphological operator.¹¹

5. Experimental session

The proposed approach was tested on the Document Image Database available from the University of Oulu.¹⁹ This database includes a collection of document images, scanned from magazines, newspapers, books and manuals. The images vary both in quality and contents: some of them contain text paragraphs only (with Latin and Cyrillic fonts of different sizes), while others contain mixtures of text, pictures, photos, graphs and charts. Moreover, only part of the set of documents is characterized by Manhattan page layout.

To obtain the DWPT tree, each image was decomposed with Daubechies db2 wavelet functions to the depth of three levels. One of the images was segmented manually and was employed to extract the best nodes of a tree. It should be noted that more than one image can be combined into one larger image for the purpose of node selection.

The best nodes were selected by means of the standard genetic algorithm with a population of 20 weight vectors, represented as binary strings initialized randomly. New generations of vector population were produced by crossover (80%) and mutation operator (20%). After 50 generations, the best subset of nodes was obtained, containing 39 out of all 85 nodes.

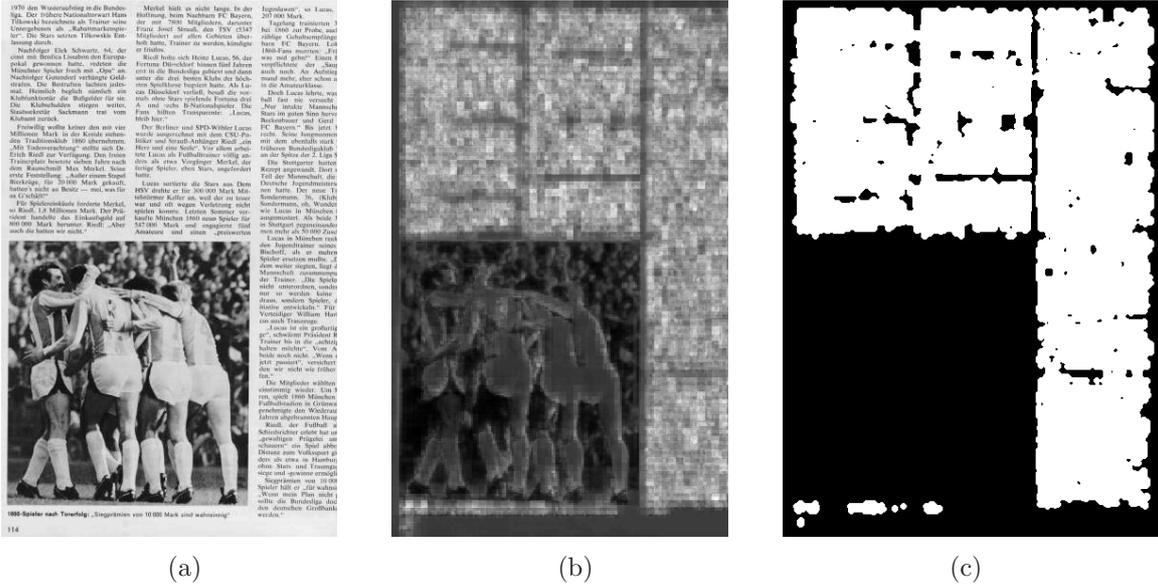


Fig. 1. Document image (a), its corresponding feature image (b) and segmentation result (c).

Using the selected nodes, the other images were segmented by clustering the pixels of the feature image into text and non-text classes. The Fuzzy C-Means (FCM) algorithm was adopted for the clustering process. Additionally, segmentation was refined by median filtering and morphological operators. Figure 1 shows the feature image, obtained from a document page and its segmentation obtained by the FCM algorithm.

To quantitatively measure the effectiveness of the proposed method, a ground truth knowledge was considered. It derives from the correct analysis of the segmentation of the 40 images extracted from the database and employed during the experimental sessions. Some of the images were rotated randomly to test the robustness against page skew. The segmentation results provided by the adopted methodology were compared with the corresponding ground truth, and the effectiveness of the overall process was measured by introducing the concept of “Percentage of segmentation accuracy” P_a :

$$(1) \quad P_a = \frac{\text{Number of correctly class. pixels}}{\text{Total number of pixels in the image}} * 100\%.$$

In our experiments we obtained an average segmentation accuracy of 92.63%. The best result was 97.18% and the worst 84.37%. Some of the results are presented in figure 2

The quantitative measure is useful also for carrying out a comparison with different image segmentation techniques proposed in literature. As an example, we can compare the obtained results with those reported in,⁸ where a polynomial spline wavelet approach has been proposed and the same kind of measure has been employed to quantify the overall accuracy. Particularly, the best results in⁸ achieved an accuracy of 98.29%. Although our methodology produced slightly lower accuracy results, it should be observed that we analyzed a total number of 40 images, instead of the 6 images considered in.⁸

6. Conclusion

In this paper we propose a method for the segmentation of document images into text and non-text regions using a filter based Discrete Wavelet Packet Transform and

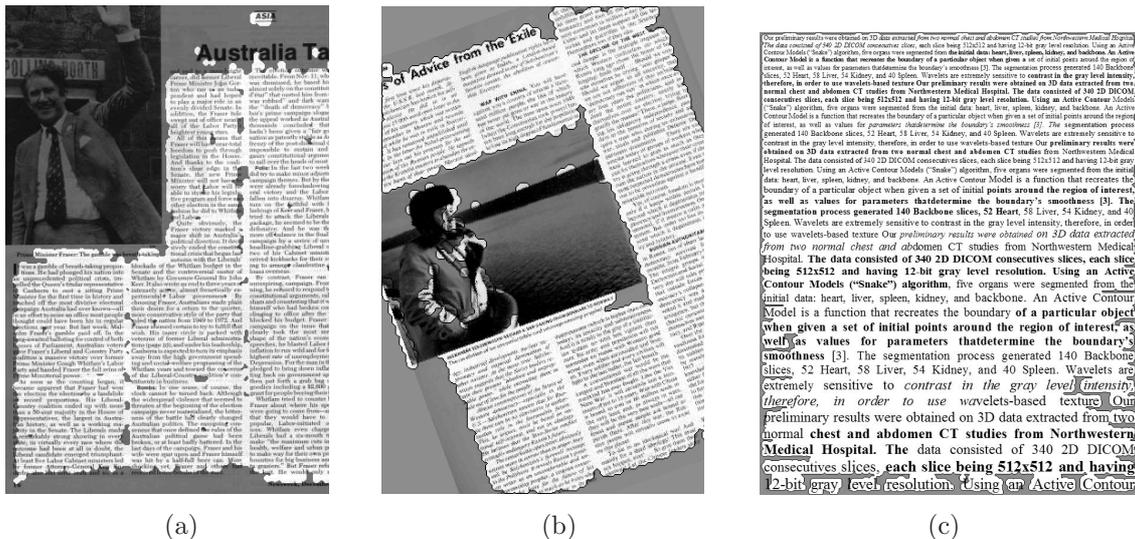


Fig. 2. Segmentation results. Segmentation of the document image (a), invariance to page skew (b) and invariance to font changes (c).

designed with a genetic algorithm. From the analysis of the experimental results the method appears to be invariant to rotation of the page and to changes in font type and size. The proposed approach is in an early stage of development, revealing many opportunities of improvement. Also, the possibility to extend the proposed methodology to the general two-texture image segmentation problem is taken under consideration.

REFERENCES

1. T. Akiyama and N. Hagita. Automated entry system for printed documents. *Pattern Recognition*, 23(11):1141–1154, 1990.
2. J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.
3. A. Bruce and H. Y. Gao. *Applied Wavelet Analysis with S-Plus*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1996. ISBN:0387947140.
4. R. Coifman, Y. Meyer, S. Quake, and V. Wickerhauser. Signal processing and compression with wavelet packets. Technical report, Yale University, 1990.
5. R. Coifman and V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992.
6. R. R. Coifman. Wavelet analysis and signal processing. In Louis Auslander, Tom Kailath, and Sanjoy K. Mitter, editors, *Signal Processing, Part I: Signal Processing Theory*, pages 59–68. Springer-Verlag, New York, NY, 1990.
7. Ingrid Daubechies. *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*. Soc for Industrial & Applied Math, December 1992. ISBN:0898712742.
8. S. Deng, S. Latifi, and E. Regentova. Document segmentation using polynomial spline wavelets. *Pattern Recognition*, 34(12):2533–2545, 2001.
9. L. A. Fletcher and R. Kasturi. A robust algorithm for text string separation from mixed text/graphics images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(6):910–918, 1988.

10. H. Fujisawa and Y. Nakano. A top-down approach for the analysis of document images. In *Proc. SSPR90*, pages 113–122, 1990.
11. R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
12. A.K. Jain and B. Yu. Document representation and its application to page decomposition. *IEEE Trans. Pattern Anal. Machine Intell.*, 20:294–308, 1998.
13. C. L. Jones, G. T. Loneragan, and D. E. Mainwaring. Wavelet packet computation of the Hurst exponent. *Journal of Physics A*, 29(10):2509–2527, 1996.
14. M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan. Synthetic segmentation and labelling of digitized pages from technical journal. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(7):737–747, 1993.
15. K. K. Lau and C. H. Leung. Layout analysis and segmentation of chinese newspaper articles. *Comput. Process. Chinese and Oriental Languages*, 8(8):97–114, 1994.
16. S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):674–693, 1989.
17. M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996. ISBN:0-262-13316-4.
18. G. Nagy, S. Seth, and M. Viswanathan. A prototype document image analysis system for technical journals. *Computer*, 25:10–22, 1992.
19. University of Oulu. Document image database.
<http://www.ee.oulu.fi/research/imag/document/>.
20. L. O’Gorman. The document spectrum for structural page layout analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1162–1173, 1993.
21. J. Sammon. An optimal discriminant plane. *IEEE Transactions on Computers*, C-19:826–829, 1970.
22. E. Stollnitz, A. DeRose, and D. Salesin. *Wavelets for Computer Graphics: Theory and Applications*. Morgan Kaufmann, 1996. ISBN: 1558603751.
23. Y.Y. Tang, S.W. Lee, and C.Y. Suen. Automatic document processing: a survey. *Pattern Recognition*, 29(12):1931–1952, 1996.
24. D. Wang and S.N. Srihari. Classification of newspaper image blocks using texture analysis. *Computer Vision Graphics and Image Processing*, 47:327–352, 1989.
25. D. Xi and S.-W. Lee. Extraction of reference lines and items from form document images with complicated background. *Pattern Recognition*, 38:289–305, 2005.