THE UNIVERSITY OF WARWICK

Original citation:
Rajpoot, K. and Rajpoot, Nasir M. (Nasir Mahmood) (2003) Wavelet based segmentation of hyperspectral colon tissue imagery. In: 7th IEEE International Multi Topic Conference (INMIC 2003), Islamabad, Pakistan, 9 Dec 2003. Published in: 7th International Multi Topic Conference, 2003. INMIC 2003. pp. 38-43.

Permanent WRAP url:
http://wrap.warwick.ac.uk/61279

Wavelet Based Segmentation of Hyperspectral Colon Tissue Imagery

Kashif M. Rajpoot[*], *Nasir M. Rajpoot*[†]

[*]Faculty of Computing Sciences & Engineering, De Montfort University
Leicester, UK
[†]Department of Computer Science, University of Warwick
Coventry, UK
Emails: iamkashi@yahoo.com, nasir@dcs.warwick.ac.uk

Abstract — *Segmentation is an early stage for the automated classification of tissue cells between normal and malignant types. We present an algorithm for unsupervised segmentation of hyperspectral human colon tissue cell images into its constituent parts by exploiting the spatial relationship between these constituent parts. This is done by employing a modification of the conventional wavelet based texture analysis, on the projection of hyperspectral image data in the first principal component direction. Results show that our algorithm is comparable to other more computationally intensive methods which exploit spectral characteristics of the hyperspectral imagery data.*

Keywords: *Hyperspectral imaging, colon tissue cell classification,, dimensionality reduction, image segmentation, wavelet transform*

1. INTRODUCTION

It has been shown through experiments that hyperspectral imaging can be successfully utilized to distinguish normal vs. malignant cells of the same cellular lineage [1]. Diagnostically important spectral features can be subtle and not easily assessed by the naked eye. Also, the high spectral resolution characteristics of hyperspectral sensors preserve important aspects of the spectrum [2]. This eventually makes segmentation of different materials possible.

According to a recent publication [3], 34,000 new cases of colorectal cancer are diagnosed each year. During the year 2000, there were 16,250 deaths from colorectal cancer in the UK alone. Colorectal cancer is the third most commonly diagnosed cancer in the UK after lung and breast cancer. The UK had one of the worst detection rates for colorectal cancer in Europe. Yet 80% of colorectal cancer cases can be treated if caught at an early stage. The limited availability of specialist pathological staff and the huge amount of data provided by the hyperspectral sensors means that the user fatigue is a significant obstruction in the examination of these images and the identification of colon cancer in early stages. New improved screening and diagnosis methods could potentially save thousands more lives each year.

The reliable detection of malignant cells in stained tissue samples is still one of the most demanding and time-consuming tasks in pathology and is a typical example of a *pattern recognition* problem [1]. Generally, the task of pattern recognition in images consists of three independent steps, which can be applied to the tissue classification problem as follows:

(i) *Image segmentation:* The objects contained in the image scene are separated from the background. This is the separation of constituent parts of tissue cells.

(ii) *Feature extraction:* The characteristics of each object are quantified. These features, of course, should contain enough discriminant information sufficient to distinguish a normal tissue from a malignant tissue.

(iii) *Classification:* Each object is assigned to a generic target class. The extracted features from the segmentation labels are utilized to discriminate between normal and malignant tissue cells.

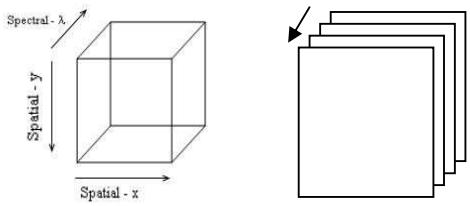
This paper reports on work related to the above first stage of segmentation. The whole process of classification is described in the MSc thesis of the primary author [4]. The input hyperspectral image cubes were taken from micro-array archival colon tissue sections between 450-650nm wavelength and of dimensions 1024x1024x20.

This paper starts with a presentation of the essential background knowledge for the work, followed by a brief description of the dimensionality reduction phase, an important step during hyperspectral image segmentation process. The wavelet multiresolution analysis procedure to perform the segmentation by exploiting the spatial relationship is presented. We also illustrate, in brief, another experimented method of segmentation which utilizes the spectral characteristics. Finally, we conclude that the results using the former algorithm are visually comparable to those using latter one but with certain limitations.

1.1 Background

Hyperspectral imaging sensors capture image scenes in contiguous but narrow spectral bands over visible and near infrared wavelength range of electromagnetic spectrum. In this way, they can potentially capture tens to hundreds of spectral bands covering the narrow spectral

features of the captured material as closely as possible. The image data provided by hyperspectral sensors can be visualized as a 3D cube or a stack of multiple 2D images (Figure 1) because of its intrinsic structure, where the cube face is a function of the spatial coordinates $f(x, y)$ and depth is a function of wavelength $d(\lambda)$.



(a) 3D cube (b) Stack of multiple 2D images

Figure 1: Hyperspectral image data representation

In this case, each spatial point on the face is characterized by its own *spectrum* (often called *spectral signature*). This spectrum is in direct correspondence with the amount of energy in the material represented, as hyperspectral sensors commonly utilize the simple fact that a body with temperature over absolute zero emits light in certain frequency bands. Consequently, the separation of constituent regions in the image scene becomes possible.

Anatomically, human body is made up of thousands of millions of different kind of cells. Our cell reproduction system constantly controls the growth of these cells, and a growth is made only if required by the body organs. When an organ of the body is affected by cancer, tumour cells are created and this reproduction control system becomes ineffective due to the continuous growth of these tumour cells.

The colon is the upper part of the large intestine tube while the rectum is the lower part of this tube (Figure 2, Source: http://cancer.gov/). Practically, colon or rectum cancer is characterized as separate cancer instances. Colorectal or bowel cancer is a composite name for colon and rectum cancer. It is the uncontrolled growth of tissue cells in either colon or rectum which causes the colorectal cancer.

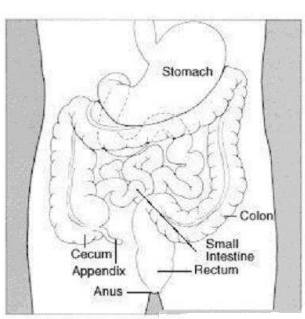


Figure 2: Human colon shown in digestive system

At a microscopic level, human colon tissue cells can be characterised as having 4 constituent parts: nuclei, cytoplasm, lamina propria, and lumen. According to a dictionary of cancer related medical terms [11], these constituent parts are defined as:

(i) *Nuclei*: the core central part of a cell, containing DNA, which controls its growth

(ii) *Cytoplasm*: the fluid inside a cell but outside the cell's nucleus. Most chemical reactions in a cell take place in the cytoplasm.

(iii) *Lamina propria*: a type of connective tissue found under the thin layer of tissues covering a mucous membrane

(iv) *Lumen*: the cavity or channel within a tube or tubular organ such as a blood vessel or the intestine

The aim of this work is to separate a given hyperspectral image data cube into these constituent parts.

2. DIMENSIONALITY REDUCTION

Before the formal process of hyperspectral image segmentation, an intermediate step of dimensionality reduction is often involved. Hyperspectral imagery data provides a wealth of information about an image scene which is potentially very helpful in the segmentation of objects. At the same time, the huge size of hyperspectral image data (with tens to hundreds of spectral bands) normally means high computational complexity. High dimensional vector spaces have been found by mathematicians to have some rather unusual and unintuitive characteristics [5].

This is often recognized as the *curse of dimensionality* in the literature. In this situation, it is usually required to reduce the dimensionality of the data before proceeding to the next essential tasks. The hyperspectral sensors commonly oversample the spectral signal to ensure that narrow band features are adequately represented [2]. The important job here is to eliminate this redundancy while, at the same time, preserving the high-quality features for the segmentation algorithm.

Principal component analysis (PCA) is a statistical multivariate data analysis tool which attempts to find the natural coordinate axes for the multidimensional dataset. It is the representation of the higher-dimensional data into lower-dimensional orthogonal axes such that it is highly decorrelated. This representation can be considered as the transformation of the original data into a new vector space where the basis vectors are actually a linear combination of the original data vectors. We utilized PCA for dimensionality reduction because of its intrinsic simplicity and well-established mathematical groundings.

In a single sentence, PCA can be formulated as *the projection of the multivariate data on the orthogonal axes which are in fact the eigenvectors of the covariance matrix of the original data*. Thus the new basis set for data is derived from the original data vectors. These orthogonal eigenvectors of the covariance matrix are actually called the principal components. Suppose A represents the

multivariate data, then a mathematical formulation can be given as [6]:

$$\Sigma = \frac{1}{n} AA^T \qquad \textbf{(1)}$$

where Σ is the covariance matrix of A. The PCA problem reduces to the computation of the eigenvectors of this matrix where an eigen-analysis problem for Σ is devised as:

$$\Sigma v = \lambda v \qquad \textbf{(2)}$$

such that v and λ represent the eigenvectors and the corresponding eigenvalues of Σ with values in λ sorted in descending order. The eigenvector corresponding to the highest eigenvalue is the principal component with maximum variation in that direction. If we assume that V is a matrix whose columns are the eigenvectors of Σ, then the projected data in the direction of principal components is given by:

$$Z = B * V \qquad \textbf{(3)}$$

where B is obtained by subtracting mean vector from each vector of A.

Later in the paper, we will come across the concept of the amount / percentage of variance preserved by the projected data in a certain principal component direction. Here, we explain this for the variance preserved in one or more directions. This is directly related to the corresponding eigenvalues and is computed, for k eigenvectors, as follows:

$$vp = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{n} \lambda_j} \qquad \textbf{(4)}$$

where $k < n$, while assuming that the eigenvalues of the covariance matrix are in a sorted order.

3. SEGMENTATION

Two ways of segmenting the hyperspectral image data into its constituent parts are: (1) *spatial analysis*: by exploiting the spatial relationships between these parts, and (2) *spectral analysis*: the exploitation of spectral characteristics. Wavelet multiresolutional analysis technique evaluates the spatial relationship between the objects in an image at multiple scales and thus falls in the former category.

Wavelets are orthogonal basis functions, having compact support in time (space), which can be used to represent a signal (image). Unlike conventional Fourier transform, which utilizes sines and cosines of varying amplitude and frequency as its basis functions, wavelet transform makes use of these wavelet functions which are scalable. A variety of such functions exists and a well suitable wavelet can be selected particular to an application depending on the signal / image characteristics to represent.

Wavelet theory is based on strong mathematical foundations and it employs established tools including pyramidal image processing, subband coding, and quadrature mirror filtering. One of the most striking and powerful applications of wavelet theory is the possibility of multiresolution analysis, shown by Mallat in 1987 [12]. Multiresolution analysis allows us to exploit the signal or image characteristics, matched to a particular scale, which might go undetected in other analysis techniques [7]. This capability of multiresolution processing paved the way to successful analysis of various kinds of texture.

In the following section, we present briefly the conventional wavelet based texture analysis procedure and describe how it was modified to our wavelet based hyperspectral image segmentation.

3.1 Wavelet based texture analysis

Texture can be defined as an attribute representing the spatial arrangement of the gray levels of the pixels in a region [8]. The sole aim of a texture analysis method is to describe different textures present in an image. One of the most important aspects of texture description has been identified as scale [9]. If we can collect these descriptive features corresponding to a texture at various scales, we can distinguish different textures in an image. Wavelet transform, on the other hand, provides a unified way of multiresolution analysis.

A usual sequence of operations performed for wavelet based texture analysis [10] is:

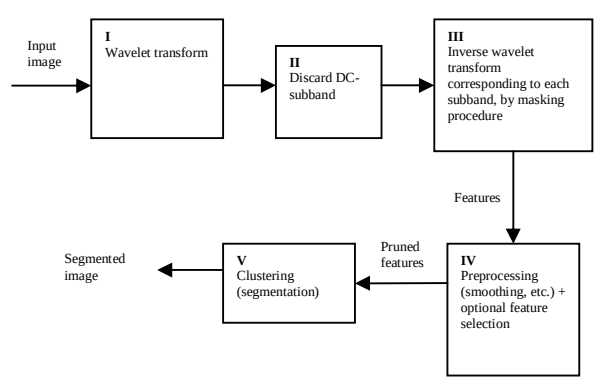
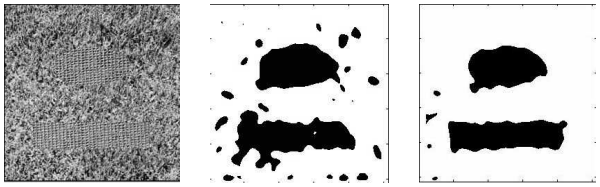


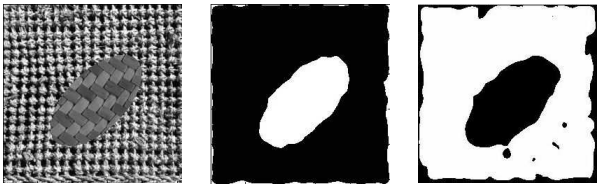
Figure 3: Sequence of operations for wavelet based texture analysis

Texture segmentation results for two image scenes with quite dissimilar textures produced with the above sequence of operations are shown in the following figures.



Original image with two different kinds of texture | Textures segmented using wavelet based approach at wavelet decomposition level 2 | Textures segmented using wavelet based approach at wavelet decomposition level 3

Figure 4: Texture segmentation with wavelet based approach - I



Original image with two different kinds of texture | Textures segmented using wavelet based approach at wavelet decomposition level 2 | Textures segmented using wavelet based approach at wavelet decomposition level 3

Figure 5: Texture segmentation with wavelet based approach - II

These results basically highlight the important role of the wavelet decomposition level which sets the detail of scale viewed by processing / analysis method. The hidden fact behind multiresolution processing for texture analysis is to generate a number of homogeneous features that represent the response of a bank of filters at different scales.

3.2 Spatial analysis based hyperspectral image segmentation

The sole purpose of this practice is to exploit the spatial characteristics in the image rather than the spectral features. Before describing the method, we would like to state what the input to this method is. It is not very uncommon in hyperspectral colon tissue imagery to have 80% or even more variance concentrated in the data projected in the first principal component direction (this fact is based on experimental results with image data cubes). Our experiments show that the projected data in the first principal component direction has sufficient spatial information to segment the cell image into constituent parts.

Assuming that each of the constituent parts of the colon tissue cells is a distinct type of texture which may be described by multiresolutional analysis procedure, we experimented with wavelet texture analysis technique on this projected data. Results showed (Figure 7 (b)) that this is perhaps not a suitable segmentation method for our problem. This was the stimulation behind experimenting new methods to exploit the multiresolutional characteristics.

The simple trick we used in our segmentation algorithm was the skipping of steps II and IV of Figure 3. Thus, the sequence of operations for hyperspectral colon tissue image segmentation becomes:

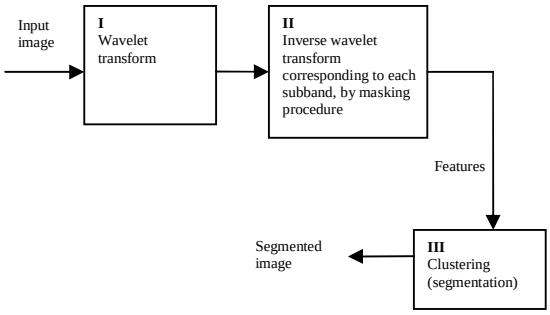
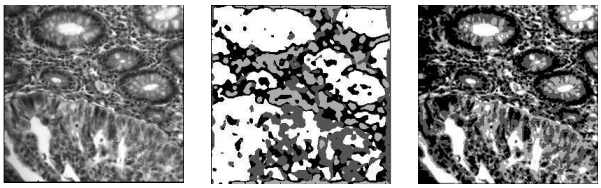


Figure 6: Sequence of operations for wavelet based hyperspectral image segmentation

Hyperspectral colon tissue image segmented into cell-constituent parts with this method is shown in the Figure below:



(a) Projected data in the first principal component direction | (b) Segmentation with conventional wavelet texture analysis approach | (c) Segmentation with our proposed approach

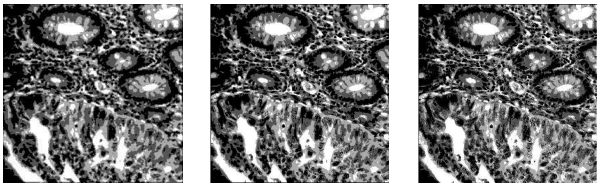
Figure 7: Wavelet based hyperspectral colon tissue image segmentation

The rationale behind this process is that the preprocessing stage (smoothing, etc.) in conventional wavelet texture analysis method loses the necessary discriminant information. Also, the discarded DC subband contains important grey value intensity approximation to the original input image. Therefore, inclusion of the DC subband feature image and avoiding the preprocessing stage actually permits the clustering algorithm to observe the intensity variation in the features and assign the labels based on these differences. Although the experimentation with wavelet decomposition level and selection of wavelet filters is not exhaustive, early attempts show that a

decomposition level 2 and daubechies-8 filters perform well for hyperspectral colon tissue segmentation.

3.3 Spectral analysis based hyperspectral image segmentation

Apart from the spatial analysis, another possibility for the segmentation of hyperspectral data is by doing a spectral analysis. This approach is in correspondence with the spectral signature (or spectrum) of each point on the face of the data cube. In practice, we rarely perform a spectral analysis on the original image cube. Rather, we transform it into lower dimensions (by PCA, ICA, etc.), to remove the spectral redundancy which may hamper the segmentation procedure, and then perform the analysis to differentiate between the objects by labeling each face point. In the case of PCA used for dimensionality reduction, the data is projected in first few principal component directions such that this projection contains possibly over 98% of the variance, calculated according to equation (4). Usually, the projected data in first three to four principal component directions can preserve over 99% variance of the whole data. This is also the case with our data where we fed the projected data into a nearest-centroid *K*-means clustering algorithm for the segmentation, Figure 8 (b). Shown also, in Figure 8 (c), is the segmentation result based on spectral analysis but with a relatively sophisticated approach of ICA (Independent Component Analysis) preprocessed by high-emphasis filtering. This method is described in detail in [4].



(a) Wavelet based segmentation (b) Spectral analysis based segmentation with PCA (c) Spectral analysis based segmentation with ICA

Figure 8: Comparison of segmentation performance

4. CONCLUSIONS

A method for segmentation of hyperspectral cell imagery data is presented with the objective of exploiting the spatial relationships between constituent parts of the tissue cells. This is quite a simple but elegant approach with established mathematical groundings. Experimental results show that although projection of data in only one principal direction was used to segment the image data, consequently saving storage and computational time, our algorithm is comparable to spectral analysis method for segmentation.

However, our wavelet based technique is limited as it will produce fine results only when the projected data in the first principal component direction covers more than 80% of the data variance. Although its performance on the data projected in the first principal component direction with less than 80% of the variance is not tested, we predict that the resulting segmentation may not be a true representation of the regions in the colon imagery. On the other hand, spectral analysis based technique is not merely dependent on first principal component direction and, therefore, it should be relatively more consistent and reliable than wavelet based technique.

The segmentation labels (or the segmented image, in other words) are further utilized for feature extraction and classification tasks. The details of this can be found in [4].

ACKNOWLEDGMENT

We gratefully acknowledge obtaining hyperspectral imagery data used in this work from and having fruitful discussions with Ronald Coifman and Mauro Maggioni of the Applied Mathematics Department of Yale University (USA).

REFERENCES

[1] R. Levenson, C. Hoyt, "Spectral Imaging & Microscopy," *American Laboratory* 2000

[2] G. Shaw, D. Manolakis, "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12-16, January 2002

[3] Cancer Research UK, *Bowel Cancer Factsheet*, May 2002

[4] K. M. Rajpoot, *Hyperspectral Colon Tissue Cell Classification*, MSc DSIP Dissertation, Faculty of Computing Sciences & Engineering, De Montfort University, UK, September 2003

[5] L. Jimenez, D. Landgrebe, "Supervised classification in high dimensional space: geometrical, statistical and asymptotical properties of multivariate data," *IEEE Trans. Syst., Man, Cybernet., C*, vol. 28, pp. 39-54, February 1998

[6] T. Marks, *Principal Component Analysis (PCA)*, Lecture Notes, Cognitive Science Department, University of California at San Diego (UCSD) USA

[7] R. Gonzalez, R. Woods, *Digital Image Processing*, Second Edition, Prentice Hall, 2001

[8] IEEE Gl., *IEEE Standard Glossary of Image Processing & Pattern Recognition terminology*, IEEE Std 610.4-1990, March 1990

[9] P. Scheunders, S. Livens, G. Van de Wouwer, P. Vautrot, P. Van Dyck, "Wavelet-based Texture Analysis ," *IEEE Conference on Image Processing and its Applications*, July 1997

[10] N. M. Rajpoot, "Texture Classification Using Discriminant Wavelet Packet Subbands," *Proceedings of 45th IEEE Midwest Symposium on Circuits and Systems (MWSCAS'2002)*, Tulsa, Oklahoma (USA), Aug. 2002

[11] National Cancer Institute, http://www.cancer.gov/

[12] S. Mallat, "A Compact Multiresolution Representation: The Wavelet Model," *Proceedings IEEE Computer Society Workshop on Computer Vision*, IEEE Computer Society Press, Washington, D.C., pp. 2-7, 1987