Wavelet-based Texture Classification of Tissues in Computed Tomography

Lindsay Semler, Lucia Dettori, Jacob Furst
Intelligent Multimedia Processing Laboratory
School of Computer Science, Telecommunications, and Information Systems
DePaul University, Chicago, Illinois, 60604, USA
{lsemler, ldettori, jfurst}@cs.depaul.edu

Abstract

The research presented in this article is aimed at developing an automated imaging system for classification of tissues in medical images. The article focuses on using texture analysis for the classification of tissues from CT scans. The approach consists of two steps: automatic extraction of the most discriminative texture features of regions of interest in the CT medical images and creation of a classifier that will automatically identify the various tissues. A comparative study of wavelets-based texture descriptors from three families of wavelets (Haar, Daubechies, Coiflets), coupled with the implementation of a decision tree classifier based on the Classification and Regression Tree (C&RT) approach is carried on. Preliminary results for a 3D data set from normal chest and abdomen CT scans are presented.

1. Introduction

The research presented in this article is part of an ongoing project [1]-[4], aimed at developing an automated imaging system for classification of tissues in medical images obtained by Computed Tomography (CT) scans. Classification of human organs in CT scans using shape or gray level information is particularly challenging due to the changing shape of organs in a stack of slices in the 3D medical image and the gray level intensity overlap in soft tissues. On the other hand, healthy organs are expected to have a consistent texture within tissues across multiple slices. Therefore, this research focuses on using texture analysis for the classification of tissues from normal chest and abdomen CT scans. The approach consists of two steps: automatic extraction of the most discriminative texture features of regions of interest in the CT medical images and creation of a classifier that will automatically identify the various tissues. This paper focuses on a comparison of wavelet-based texture descriptors from three families of wavelets, coupled with the implementation of a decision tree classifier based on the Classification and Regression Tree (C&RT) approach. A similar study using texture descriptors based on the co-occurrence matrix model and the run-length encoding model was carried out by other members of the research group in [3], [4].

Texture is a commonly used feature in the analysis and interpretation of images. Texture can be characterized by a set of local statistical properties of the pixel gray level intensity. It measures the variations in a surface, looking at properties like smoothness, coarseness and regularity. Typically, texture can be described by statistical, structural, or spectral techniques such as: wavelets, run-length statistics, spectral measures, fractal dimensions, statistical moments, and co-occurrence matrices. The discrete wavelet transform maps the image onto a low-resolution image and a series of detail images, providing a multi-scale representation of the image. The low-resolution image carries little energy and was not included in the texture analysis. First and second order statistics of the wavelet detail coefficients provide texture descriptors that can discriminate contrasting intensity properties spatially distributed throughout the image, according to various levels of resolution (see [5], [6] for a similar approach).

2. Methodology

2.1. Wavelets

In this article, we analyze and compare texture classification based on three families of wavelets: Haar (H), Daubechies 4 (D4), and Coiflet (C6). A wavelet is a mathematical function that can decompose a signal or an image with a series of averaging and differencing calculations. Wavelets are typically used in image decomposition and compression (both lossless and lossy), since the image can be decomposed and then reconstructed by simply reversing the decomposition process. Wavelets calculate average intensity properties as well as several detailed contrast levels distributed throughout the image. Wavelets can be calculated according to various levels of resolution (or blurring) depending on how many levels of averages are calculated. They are sensitive to the spatial distribution of grey level pixels, but also are able to differentiate and preserve details at various scales or resolutions. This multi-resolution quality allows for the analysis of gray level pixels regardless of the size of the neighborhood. These properties lead to the idea that wavelets could guide researchers to better texture classification of human organs in CT scans.

The three wavelet transforms investigated were: Haar, Daubechies, and Coiflet. The Haar wavelet is the oldest and simplest orthonormal wavelet. It conserves the energy of signals, while compressing this energy into a more compact form. The Haar wavelet is conceptually simple, memory efficient, and exactly reversible without the edge effects characteristic of other wavelets. The Haar transform does not have overlapping windows, and reflects only changes between adjacent pixel pairs. The Haar wavelet uses only two scaling and wavelet function coefficients, thus calculates pair wise averages and differences.

Daubechies is conceptually more complex, and generally has a higher computational overhead. The Daubechies wavelet uses overlapping windows, so the results reflect all changes between pixel intensities. Since Daubechies averages over more pixels, it is smoother than the Haar wavelet. The Daubechies D4 transform has four wavelet and scaling coefficients. The sum of the scaling function coefficients are also one, thus the calculation is averaging over four adjacent pixels. Since the size of the filter is greater than the incoming image, both a mirroring and a periodic extension of the filter were tested. The mirror extension, which involves mirroring the last two pixels of the image, generally proved to be the optimal choice.

Coiflets was originally derived from the Daubechies wavelet. It has an even higher computational overhead and uses windows that overlap more. The Coiflet wavelet uses six scaling and wavelet function coefficients. This increase in pixel averaging and differencing leads to a smoother wavelet and increased capabilities in several image-processing techniques (de-noising images, etc.). The filter follows the same structure as both Haar and Daubechies. It calculates both averages and differences using the same format, only with six adjacent pixels. The Coiflet wavelet also follows the mirror technique.

2.2. The data set

Our preliminary results were obtained on 3D data extracted from two normal chest and abdomen CT studies from Northwestern Medical Hospital. The data consisted of 340 2D DICOM consecutives slices, each slice being 512x512 and having 12-bit gray level resolution. Using an Active Contour Models ("Snake") algorithm, five organs were segmented from the initial data: heart, liver, spleen, kidney, and backbone. An Active Contour Model is a function that recreates the boundary of a particular object when given a set of initial points around the region of interest, as well as values for parameters that determine the boundary's smoothness [3]. The segmentation process generated 140 Backbone slices, 52 Heart, 58 Liver, 54 Kidney, and 40 Spleen. Wavelets are extremely sensitive to contrast in the gray level intensity, therefore, in order to use wavelets-based texture description it was necessary to eliminate all background pixels to avoid mistaking the edge between the artificial background and the organ as a texture feature. Each slice was therefore further cropped, and only square sub-images fully contained in the interior of the segmented area were generated, resulting in 1129 slices of "pure" single-organ tissue (665 Backbone, 103 Heart, 122 Liver, 184 Kidney, 54 Spleen). The data set was then divided into a training set (containing approximately 67% of the images) and a testing set (containing approximately 33% of the images).

2.3. Feature extraction

Once the medical images were prepared, wavelets were used for feature extraction of texture information. Haar, Daubechies and Coiflet wavelet filters were applied to each of the 1128 cropped images, using three different levels of resolution. At each resolution level three detail coefficient matrices were calculated resulting in three matrices representing the vertical, horizontal and diagonal structures of the image. The wavelets were then preprocessed, by taking the absolute value of each coefficient and binning each detail into sixteen bins. Once the preprocessing was completed, the histogram of each of the details coefficient matrix was calculated. First, a histogram was calculated from each wavelet detail. The histogram calculated on wavelet coefficients measures the frequency distribution of contrast levels. Mean and Standard Deviation texture descriptors were then extracted from the histogram of each coefficient matrix. This yields six texture descriptors (two for each detail) for every level of resolution. In addition to this, co-occurrence matrices were also calculated at each detail and level of resolution. A co-occurrence matrix captures the spatial dependence of contrast values, depending on different directions and distances specified. Four co-occurrence matrices were calculated for each detail matrix at each resolution level. A co-occurrence matrix was calculated for four directions, 0, 45, 90, and 135 degrees at a set distance of one. Traditional co-occurrence techniques also consider several distances between pixels. Since the texture descriptors are calculated based on multi-resolution wavelets, the resolution levels act as distances. The following nine Haralick texture descriptors were then extracted from each co-occurrence matrix: energy, entropy, contrast, homogeneity, sum-mean, variance, maximum probability, inverse difference moment, and cluster tendency (see [7]).

The final texture descriptor vector had 132 elements per resolution level, generating a 396-element texture descriptor vector per image. Feature reduction is necessary to reduce the feature space so it is manageable for the decision trees. The feature space was limited by decreasing the number of texture descriptors. The size of the texture description vector was reduced to 99 by averaging over the four co-occurrence directions. Results were then compared with those based on a 33 element texture descriptor vector obtained by averaging over the three wavelet detail directions.

2.4. Classification

The classification step was carried out using a decision tree classifier based on the Classification and Regression Tree (C&RT) approach. A decision tree predicts the class of an object (organ) from values of predictor variables (wavelet-based texture descriptors in this case). The most relevant texture descriptors are found for each specific organ, and based on those selected descriptors, a set of decision rules are generated. This set of rules is then used for the classification of the each region. Out of 1128 cropped medical slices, approximately 67% of the data were then used for training and 33% were used for testing. Using the C&RT approach, each tree's parameters were optimized, including depth of tree, number of parent nodes, and number of child nodes. The parameters were considered optimal when the highest possible rate of accuracy was found. From the semi-optimal decision trees, a misclassification matrix was calculated for each Haar, Daubechies, and Coiflet wavelet to evaluate the performance of each classifier.

A misclassification matrix is a table that lists each organ and its true positives, true negatives, false positives and false negatives (Table 1). The number of true positives is the number of organs that are correctly classified as that organ. The number of false positives is the number of organs that are incorrectly classified as that organ. The number of true negatives is the number of other organs that are incorrectly classified as the organ. The number of false negatives is the number of other organs that are correctly classified as other organs. From the misclassification matrix specificity, sensitivity, precision, and accuracy statistics were computed. Specificity measures the accuracy among positive instances, and is calculated by dividing the true negatives by the number of all other organ slices. Sensitivity measures the accuracy among negative instances, and is calculated by dividing the number of true positives by the total number of that specific organ slices. Precision measures how consistent the results can be reproduced. Accuracy reflects the overall correctness of the classifier, and is calculated by adding the true positives and negatives together and dividing by the entire number of organ slices.

Measure	Definition
Sensitivity	True Positives / Total Positives
Specificity	True Negatives / Total Negatives
Precision	True Positives / (True Positives + False Positives)
Accuracy	(True Positives + True Negatives) / Total Samples

Table 1: Measures of classification performance

3. Results and future work

For the testing set, the accuracy performance was in the 88-96% range, with the 33 Haar-based descriptors outperforming the others on each organ [see Table 2]. The specificity performance was in the 89-99% range, with Daubechies-based descriptors outperforming the others everywhere except for heart and kidney [see Table 3]. Sensitivity and Precision performances are in the 25-94% and 29-97% range respectively [see Table 4, and 5]. Spleen and heart are the organs responsible for the low values at the left side of the spectrum. The lower end of the spectrum when spleen and heart are eliminated jumps to 68% and 82% respectively. We conjecture that these results are negatively affected by the small number of spleen and heart slices available to train and test the decision tree. The author will be receiving additional radiologist-labeled slices of "pure single-organ" tissue from Northwestern Hospital and new tests will be generated to confirm this conjecture. Except for kidney, the Haar based 33 descriptors are the ones providing the best sensitivity performance. These results indicate that, in general, a reduction of the number of descriptors improves the discriminative power of wavelet-based texture analysis. This was confirmed by a preliminary study of the descriptors vector done use the Principle Component Analysis, which indicated that the most discriminative descriptors are the first order statistics descriptors mean and standard deviation obtained from the histogram of details matrices, along with contrast and maximum probability.

Further area of investigation include using a more sophisticated system of feature reduction based on a combined principle component analysis (PCA) on all descriptors calculated from the three wavelets, as well as other non wavelet based descriptors like run-length and co-occurrence based descriptors. Limiting the feature vectors to the component selected by the PCA should lead to an increase in accuracy rates.

Results obtained in this article will also be validated on a new data set based on radiologist-labeled single organ slices. This will eliminate the possible contamination of texture descriptors by background and different organ regions which might have been overlooked by the initial automatic segmentation. This should result in a more accurate training of the decision tree and ultimately improved accuracy rates. Also underway is a more comprehensive study of additional families of wavelets.

Accuracy							
Wavelet type	**# Descriptors**	**Backbone**	**Heart**	**Liver**	**Kidney**	**Spleen**	**Average**
Haar	33	92.28	94.07	94.07	93.47	96.44	94.07
	99	91.10	93.77	94.96	91.69	93.47	93.00
Daubechies	33 Mirror	91.39	92.88	94.36	90.21	96.14	93.00
	99 Mirror	90.21	92.88	94.36	90.50	94.66	92.52
Coiflets	33 Mirror	91.10	94.07	90.21	88.13	93.18	91.34
	99 Mirror	89.32	93.47	91.69	88.13	93.47	91.22

Table 2: Comparison of accuracy rates for each organ

	Backbone				Heart			
	Sens.	**Spec.**	**Prec.**	**Acc.**	**Sens.**	**Spec.**	**Prec.**	**Acc.**
Haar (33)	93.33	90.85	93.33	92.28	58.06	97.71	72.00	94.07
Haar (99)	91.28	90.85	93.19	91.10	51.61	98.04	72.73	93.77
D4 (33 Mirror)	87.18	97.18	97.70	91.39	70.97	95.10	59.46	92.88
D4 (99 Mirror)	90.26	90.14	92.63	90.21	77.42	94.44	58.54	92.88
C6 (33 Mirror)	91.79	90.14	92.75	91.10	67.74	96.73	67.74	94.07
C6 (99 Mirror)	87.69	91.55	93.44	89.32	64.52	96.41	64.52	93.47

Table 3: Classification performance for backbone and heart

Liver					Kidney			
	Sens.	Spec.	Prec.	Acc.	Sens.	Spec.	Prec.	Acc.
Haar (33)	76.92	96.31	73.17	94.07	87.50	94.66	76.56	93.47
Haar (99)	71.79	97.99	82.35	94.96	85.71	92.88	70.59	91.69
D4 (33 Mirror)	68.42	97.66	78.79	94.36	94.74	89.29	64.29	90.21
D4 (99 Mirror)	71.05	97.32	77.14	94.36	75.44	93.57	70.49	90.50
C6 (33 Mirror)	57.89	94.31	56.41	90.21	64.91	92.86	64.91	88.13
C6 (99 Mirror)	73.68	93.98	60.87	91.69	68.42	92.14	63.93	88.13

Table 4: Classification performance for liver and kidney

Spleen				
	Sens.	Spec.	Prec.	Acc.
Haar (33)	50.00	98.75	66.67	96.44
Haar (99)	50.00	95.64	36.36	93.47
D4 (33 Mirror)	37.50	99.07	66.67	96.14
D4 (99 Mirror)	25.00	98.13	40.00	94.66
C6 (33 Mirror)	31.25	96.26	29.41	93.18
C6 (99 Mirror)	31.25	96.57	31.25	93.47

Table 5: Classification performance for spleen

10. References

[1] D. Xu, J. Lee, D.S. Raicu, J.D. Furst, D. Channin. "Texture Classification of Normal Tissues in Computed Tomography", The 2005 Annual Meeting of the Society for Computer Applications in Radiology, Orlando, Florida, June 2-5, 2005.

[2] D. Channin, D. S. Raicu, J. D. Furst, D. H. Xu, L. Lilly, C. Limpsangsri, *"Classification of Tissues in Computed Tomography using Decision Trees"*, Poster and Demo, The 90th Scientific Assembly and Annual Meeting of Radiology Society of North America (RSNA04), Chicago, IL, USA, November 28 - December 3rd, 2004.

[3] Kurani, A, Xu, D. H., Furst, J. D., Raicu D. S.. "Co-occurrence matrices for volumetric data", The 7th IASTED International Conference on Computer Graphics and Imaging - CGIM 2004, Kauai, Hawaii, USA, in August 16-18, 2004

[4] D. H. Xu, A. Kurani, J. D. Furst, & D. S. Raicu, *"Run-length encoding for volumetric texture"*, The 4th IASTED International Conference on Visualization, Imaging, and Image Processing - VIIP 2004, Marbella, Spain, September 6-8, 2004.

[5] Kara, Bayram, and Nurdal Watsuji. Using Wavelets For Texture Classification. *IJCI Proceedings of International Conference on Signal Processing*. ISSN 1304-2386, Columne:1, Number: 2. September 2003.

[6] Van de Wouwer, G., P. Scheunders, and D. Van Dyck. Statistical Texture Characterization from Discrete Wavelet Representations. University of Antwerp: Antwerpen, Belgium.

[7] Haralic, R.M., Shapiro, L.G., Computer and Robot Vision, Addison_wesley Publishing Co., 1992.