Texture-Based Image Retrieval for Computerized Tomography Databases

Andrew Corboy, Winnie Tsang, Daniela Raicu, Jacob Furst
Intelligent Multimedia Processing Laboratory
School of Computer Science, Telecommunications, and Information Systems
DePaul University, Chicago, Illinois, 60604
{acorboy, wtsang@students.depaul.edu}, {draicu, jfurst}@cti.depaul.edu

Abstract

In this paper we propose a CBIR system for retrieval of normal anatomical regions present in Computed Tomography studies of the chest and abdomen. We implement and compare eight similarity measures using local and global co-occurrence texture descriptors. The preliminary results are obtained using a CT database consisting of 344 CT images representing the segmented heart and great vessels, liver, renal and splenic parenchyma, and backbone from two different patients. We evaluate the results with respect to the retrieval precision metric for each of the similarity measures when calculated at both levels, per organ and overall.

1. Introduction

In medicine to date, virtually all Picture Archiving and Communications Systems (PACS) retrieve images simply by indices based on patient name, technique, or some-observer-coded text of diagnostic findings [11]. Fields of text tags, such as patient demographics, diagnostic codes (ICD-9, American College of Radiology diagnostic codes), image view-plane (saggital, coronal, etc) and so on usually are the first handles on this process. This textual approach, however, fails to fully account for quantitative and shape relationships of medically relevant structures within an image that are visible to a trained observer but not codable in conventional database terms. Suitable database structures addressing the visual/spatial properties of medical images and more effective techniques to deal with different types of knowledge are necessary.

In this paper we propose a CBIR system for normal anatomical regions (heart and great vessels, liver, renal and splenic parenchyma, and backbone) present in CT studies of the chest and abdomen. The proposed system automatically extracts co-occurrence texture features both at the global (organ) level and local (pixel) level, and then uses these features to measure the similarity between various organ images of a CT organ image database. One of the major challenges in building such a type of system is to determine the best similarity metric to be used in the context of texture features for CT image databases. In our approach, we will investigate the effectiveness of several metrics in performing similarity retrieval based on both pixel- and global-based co-occurrence texture features.

2. Background

General CBIR systems that extract automatically low-level image features from pixel data have been intensively explored by several researchers (Stan et al. [10], Niblack et al. [7], Mehrotra et al. [6], Pentland et al. [9]); although anatomic information rests on visual appearances which makes it a natural feature to use in retrieval, there has been little work done to build medical CBIR systems.

Glatard et al. [3] introduced a CBIR system which uses Gabor filters extracted from segmented cardiac Magnetic Resonance Imaging (MRI)'s to perform clinically relevant queries on large image databases that do not require user supervision. Brodley et al. [2] introduced a CBIR system for the retrieval of CT lung images; their proposed system uses several features (such as co-occurrence texture features, Fourier descriptors and moments), and relies on expert interaction with the system in addition to various machine learning and computer vision techniques. Wei et al. [13] proposed a CBIR system for the mammography imaging modality using the co-occurrence texture signatures as global features.

In our approach, we use both global and local-level co-occurrence texture features to retrieve normal anatomical regions produced by Computed Tomography imaging modality. Since there is no similarity measure known to perform the best for the CT modality, we compare eight similarity measures and show how the selection of a similarity measure affects the retrieval precision.

3. Methodology and Experimental Results

Our preliminary results are based on data extracted from two normal CT studies from Northwestern Memorial Hospital. The data consisted of multiple, serial, axial computed tomography images derived from helical, multi-detector CT acquisitions. The images were provided in DICOM (Digital Imaging and Communications in Medicine) format of size 512 by 512 and having 12-bit gray level resolution. Using the Active Mapping Contour (ACM) algorithm [5], we segmented five organs from 340 axial slices: heart and great vessels, liver, renal and splenic parenchyma, and backbone.

Figure 1 gives an overview of our proposed system. For each segmented organ, we calculated a set of ten Haralick texture features at both global and pixel-level; therefore, each organ or pixel (depending on the considered level) will be represented as a vector with ten elements that will be further used for comparing the similarity among the images/organs. A preprocessing step is also applied before calculating the similarity among images: the texture descriptors are normalized (using a min-max approach) such that their differences in their scales do not influence the similarity results. After the normalization step, eight similarity measures are calculated between the query image and all the other images from the database. Figure 2 shows an example of the query image (liver) and the retrieved results.

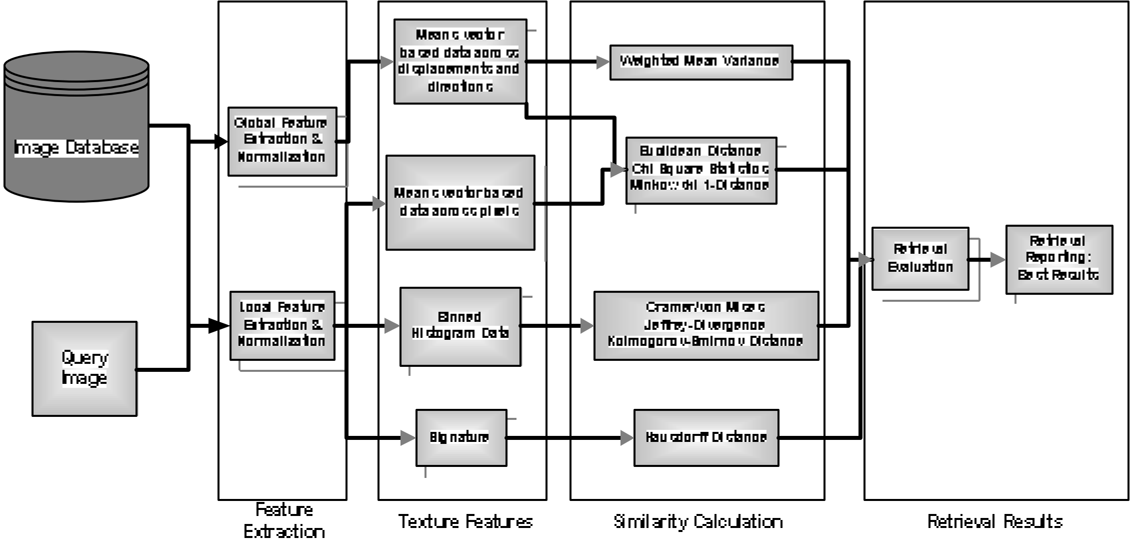

Figure 1: Diagram of the proposed CBIR system

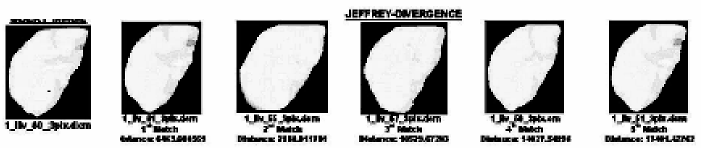

Figure 2: The best retrieval results for the liver query image shown as the image

3.1. Texture Features

In medical image processing, texture is especially important, because it is difficult to classify human body organ tissues using shape or gray level information. Several methods have been applied towards the analysis and characterization of texture within medical images including fractal dimension, run-length encoding, discrete wavelet transform, and co-occurrence matrices.

In our current implementation, we use the Haralick co-occurrence texture model and its texture descriptors that capture the spatial dependence of gray-level values and texture structures within an image [5]. We are using the following ten descriptors: entropy, energy (angular second moment), contrast, homogeneity, sum mean, variance, correlation, maximum probability, inverse difference moment and cluster tendency [8]. These descriptors are calculated at both local (pixel) and global-level depending on the similarity measures to be used and the fundamental structures present in the images.

To compute global-level features, the normalized co-occurrence matrices are calculated in four directions and five displacements generating twenty matrices per segmented image. The ten Haralick features are calculated for each of the twenty matrices and then, the twenty values are averaged and recorded as a *mean feature vector* for the corresponding segmented image [8].

To compute pixel-level features, a 5-by-5 neighborhood is considered for each pixel within the segmented region, generating one co-occurrence matrix per 5-by-5 neighborhood region. While co-occurrence matrices are normally defined for a fixed distance and direction when calculated at the global level, for the pixel-level approach, since the neighborhood of the pixel is small, we do not calculate the co-occurrence along fixed directions and displacements, but instead consider all the pixel pairs within that neighborhood. Thus, our implementation produces a single co-occurrence matrix for each pixel rather than for each choice of distance and direction. Then, for each co-occurrence matrix (each pixel), we calculate ten Haralick features which can be related to specific characteristics in the image.

Since the gray-levels for our images range from 0 to 4096, for reasons of computational efficiency, the number of gray levels can be reduced if one chooses to bin them, thus reducing the size of the co-occurrence matrix. In our approach, before calculating the matrices, we applied a linear binning such that the range [0, 4096] was mapped to the range [0,256].

From the pixel-level data, we derived *1) means vector-based data, 2) binned-histogram data and 3) texture signatures*. 1) The means vector-based data consists of the average of the normalized pixel-level data for each region such that the texture representation of that corresponding region is a vector instead of a set of vectors given by the pixels' vector representation within that region. 2) The binned histogram data is generated using 16 bins. 3) The texture signatures are clustered representations of the normalized local level data obtained using a k-d tree clustering algorithm [1]. The k-d tree algorithm iteratively divides the data space using predefined stopping criteria. In our approach, we implemented two stopping criteria: the first criterion was to establish a minimum variance within the subset to be divided to prevent creating redundant clusters and over-splitting; the second stopping criterion was used to enforce a minimum cluster size as a percentage of the original data set. This was done to maintain significant size within the clusters and to prevent outliers from uncontrollably growing the tree. Multiple signatures were developed by varying both the variance and minimum cluster size and used for the directed Hausdorff distance calculation and retrieval evaluation.

3.2. Similarity Measures

Similarity metrics are measures that describe how similar two images are (organs in our case). We implement eight similarity measures as follows: 1) Euclidean distance, 2) Minkowski 1-distance (city block distance or L_1 norm), 3) Chi-square (c^2) statistics (used to distinguish whether distributions of the descriptors differ from each other), 4) weighted-mean-variance (WMV – uses the means and standard deviations for each of the considered features), 5) Jeffrey-divergence (used to compute the distance between class distributions of two values of the same feature), 6) Cramer-von Mises (similar to the squared Euclidean distance but calculated between the distributions and as the maximal discrepancy between the cumulative distributions) , 7) Kolmogorov-Smirnov distance (used for unbinned data distributions and it is invariant to arbitrary monotonic transformations), and 8) Hausdorff distance (used on texture signatures). For the mathematical definitions of these metrics, we refer the reader to [12].

3.3. Evaluation of results

We evaluate the system's similarity retrieval results using precision as a performance metric. The precision is calculated as the number of relevant retrieved images divided by the total number of retrieved images in return to the query. A retrieved image is considered to be relevant if belongs to the same anatomical region as the query image. In our current implementation, we look only at the best five retrieval results and evaluate the eight similarity measures with respect to these top similarities.

Since several similarity measures are used for both pixel-level and global-level data, we need to compare and find out what are the best similarity measures for pixel-level and global level data. The best similarity metric result for pixel-level data would also be compared with the best result from global-level data. At the pixel-level data, we compared the best similarity metric for means vector-based with the best from binned-histogram and texture signatures.

4. Conclusion and Future Work

Four out of the eight similarity measures gave an overall precision over 85%: the Euclidean distance, c^2 statistic, Minkowski 1-distance and Jeffrey-Divergence. When comparing global (Figure 3) versus local (Figure 4), the pixel-level texture descriptors performed better than the global-level and this might be explained by the fact that they capture better the local information in the data; we also noticed that even though there is not a high difference in the overall performance of the two levels of descriptors, for liver and spleen the performance is 10% to 20% better when using pixel-level descriptors. Furthermore, comparing the best similarity metrics per organ at the pixel level, we notice that Jeffrey divergence performs the best for backbone (100%), heart (89.7%) and kidneys (96%), and Euclidean distance performs the best for liver (73%) and spleen (62.1%). A CBIR system using pixel level data and different similarity measures for different organs will improve the retrieval precision significantly. For our experimental results, if we used the best similarity metric for each class of images (i.e. for each organ), the overall precision of the system would increase to about 89% up from about 87%.

GLOBAL LEVEL		Backbone	Heart	Left Kidney	Right Kidney	Liver	Spleen	OVERALL
VECTOR-BASED PRECISION	Eucliean Distance	98.8%	78.5%	88.2%	92.8%	60.1%	45.8%	81.7%
	Chi Square Statistics	99.5%	85.9%	88.7%	94.2%	58.3%	45.4%	82.9%
	Minkowski 1 Distance	99.2%	80.8%	90.9%	93.5%	60.9%	47.5%	82.8%

Figure 3: Precision at the global-level when each image from the database becomes a query image

PIXEL LEVEL		Backbone	Heart	Left Kidney	Right Kidney	Liver	Spleen	OVERALL
VECTOR-BASED PRECISION	Eucliean Distance	99.8%	84.9%	86.0%	85.5%	73.0%	62.1%	86.4%
	Chi Square Statistics	99.9%	86.5%	83.9%	87.0%	67.8%	57.9%	85.3%
	Minkowski 1 Distance	99.5%	86.5%	83.9%	88.4%	71.3%	59.6%	86.0%
	Weighted-Mean-Variance	98.7%	80.8%	90.9%	85.5%	69.8%	59.6%	85.0%
BINNED HISTOGRAM PRECISION	Cramer/von Mises	100.0%	88.1%	87.1%	87.0%	61.5%	55.8%	84.5%
	Jeffrey-Divergence	100.0%	89.7%	95.7%	96.4%	63.2%	59.2%	86.9%
	Kolmogorov-Smirnov Distance	100.0%	87.8%	88.7%	90.6%	59.8%	57.9%	84.8%
SIGNATURE-BASED PRECISION	Hausdorff 5% v 1% cs	100.0%	86.9%	92.5%	88.4%	58.9%	42.5%	83.0%

Figure 4: Precision at the global-level when each image from the database becomes a query image

As future work, we would like to experiment our system with ‘patches’ of pure tissues instead of imperfectly segmented images produced by the snake algorithm as used in the current implementation. We plan to obtain the pure tissue ‘patches’ selected by experts (i.e. radiologists) in the medical field. Besides evaluating the similarity measures and the proposed system for pure patches, we plan to investigate different window sizes for calculating the pixel level texture and explore other similarity measures. As a long term goal, the system can be further improved and integrated in the standard DICOM Query / Retrieve mechanisms in order to allow texture-based retrieval for the daily medical work flow.

References

[1] J.L. Bentley. *Multidimensional binary search trees used for associative searching*. Communications of the ACM, 18:509-517, 1975.

[2] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, and A.M. Aisen. *Content-Based Retrieval from Medical Image Databases: A Synergy of Human Interaction, Machine Learning and Computer Vision*. In Proc. of the Sixteenth National Conference on Artificial Intelligence (AAAI'99), Orlando Florida, July 1999.

[3] T. Glatard, J. Montagnat, and I.E. Magnin. *Texture based medical image indexing and retrieval: application to cardiac imaging*. ACM SIGMM international workshop on Multimedia Information Retrieval (MIR'04), Proceedings of ACM Multimedia 2004, New-York, USA, October 15-16, 2004.

[4] R.M. Haralick, K. Shanmugam, and I. Dinstein. *Textural Features for Image Classification*. IEEE Transactions on Systems, Man, and Cybernetics, vol. Smc-3, no.6, Nov. 1973. pp. 610-621.

[5] Kass, M., Witkin, A., Terzopoulos, D. (1988). Snakes: Active contour models. Int’l. J. of Comp. Vis. 1(4).

[6] S. Mehrotra, Y. Rui, K. Chakrabarti, M. Ortega, and T.S. Huang, "Multimedia Analysis and Retrieval Sys-tem", Proceed. of the 3rd Int. Workshop on Multimedia Information Systems, Como, Italy, September 25-27, 1997.

[7] W. Niblack, R. Barber, W. Equitz et al. The QBIC project: querying images by content using color, texture, shape. Proc. SPIE: Storage and Retrieval for Images and Video Databases, vol. 1908, pp. 173-187, 1993.

[8] D.S. Raicu, J.D. Furst, D. Channin, D.H. Xu, & A. Kurani, *A Texture Dictionary for Human Organs Tissues' Classification*. Proceed. of the 8th World Multiconf. on Syst., Cyber. and Inform., July 18-21, 2004.

[9] A. Pentland, R. W. Picard, S. Sclaroff, "Photobook: tools for content-based manipulation of image data-bases," Technical Report, Volume 255, MIT, Media Lab., (1993)

[10] Stan D. & Sethi I.K., “eID: A System for Exploration of Image Databases”, Information Processing and Management Journal, Elsevier (Ed.), vol. 39/3, pp. 335-361, May 2003.

[11] Tagare, D.H., Jaffe, C.C., & Duncan, J. Medical Image Databases: A Content-based retrieval approach. *Journal of American Medical Informatics Association*, 4 (3), 1997, pp. 184 – 198.

[12] Y. Rubner and C. Tomasi. *Texture Metrics*. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pages 4601-4607, October 1998.

[13] C.-H. Wei, C.-T. Li and R. Wilson. *A General Framework for Content-Based Medical Image Retrieval with its Application to Mammograms*. in *Proc. SPIE Int’l Symposium on Medical Imaging*, San Diego, February, 2005.