31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Spatial Feature Extraction Techniques For the Analysis of Ductal Tree Structures

Angeliki Skoura[1], Michael Barnathan[2] and Vasileios Megalooikonomou[1, 2]
[1]Computer Engineering and Informatics Department, University of Patras, Patras, Greece
[2]Data Engineering Lab, Center for Information Science and Technology, Temple University, Philadelphia, USA

Abstract— **The objective of this study is the classification of the breast ductal tree structures appearing in galactograms in order to investigate the relation between topological properties of the tree-like parenchymal networks and radiological findings regarding breast cancer. We present two different methods to characterize the spatial distribution of branching points; a variation of Sholl's analysis and a sectoring technique. Similarity searches and k-nearest neighbor classification of the trees are performed using the cosine similarity metric. We applied our approach to clinical x-ray galactograms to distinguish between cases with reported radiological findings and cases with no reported findings. Experimental results illustrate the effectiveness of the proposed methods, indicating that our analysis can potentially aid in early breast cancer diagnosis.**

I. Introduction

Breast cancer is the most prevalent cancer among women and affects approximately one million women worldwide [1]. Early stage detection, which should allow cancer diagnosis when curative treatment is possible, can be achieved with proper screening. The current screening protocol is based on mammography to identify suspicious regions of the breast. Although the use of this x-ray imaging has become an indispensable tool in diagnosis of cancer masses, it cannot visualize the ductal network. Such a visualization is important as several types of cancerous lessions originate from the epithelial tissue of the breast ducts and spread along the ductal pathway [2]. The x-ray examination which is able to depict the inside of the ductal structure is galactography. Galactography uses mammography and a contrast material to obtain pictures, called galactograms, and was shown to be crucial in reducing mortality rates [3]. The galactogram (Fig. 1.a) visualizes the breast ductal network whose topological structure is like a binary tree.

In the field of medical image analysis, a challenging issue is the extraction of descriptive features which correlate the morphological variability of the various structures with the function and the underlying pathology of the corresponding organs. In particular for tree-like structures, these features characterize and represent the original trees capturing properties such as the branching frequency, the tortuosity and the asymmetry index. In order to evaluate the topology of these structures, a pre-processing step is usually necessary: the trees are traced on the images and extracted using manual or automated procedures. The quantification of the extracted descriptors can be combined with classification techniques to assist medical diagnosis.

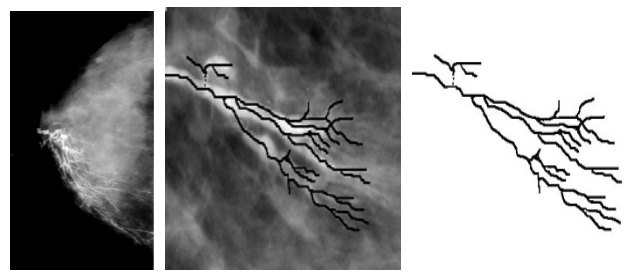

Fig. 1. (a) A real x-ray galactogram, (b) a ductal tree magnified and delineated and (c) the manually extracted ductal tree.

In this paper, we are interested in characterizing the topology of the ducts and its association to radiological findings. For this purpose, we propose a multi-step approach for characterizing and classifying the ductal tree structures. After an initial data acquisition procedure, we use a quantitative scheme of the branching points distribution which is based on two methods; a variation of Sholl's analysis and a sectoring technique. Both of these methods represent the topological patterns with vectors and thus, the problem of classifying trees is reduced to the classification of numeric vectors. We employ the cosine similarity metric to compare the vectors. We perform similarity searches and classification using the k-nearest neighbors technique. The experiments were performed on real clinical x-ray galactograms in order to distinguish among cases of reported and no reported galactographic findings. The results demonstrate the effectiveness of the proposed methods in automatically characterizing and classifying the tree-like structures of galactograms outperforming previous state-of-the art methods in classification accuracy and precision of similarity searches.

II. Background

Several studies in literature have demonstrated that examining the morphology of the ductal network can provide valuable insight to the development of breast cancer and assist in diagnosing pathological breast tissue. Towards this direction, Bakic et al. proposed a quantitative method based on R-matrices (Ramification matrices) to classify galactograms regarding radiological findings [4].

More recently, Megalooikonomou et al. proposed tree encoding techniques such as the depth-first string encoding and the Prüfer encoding to obtain a symbolic string representation of the tree's branching topology [5]. In other approaches, tree asymmetry index and fractal analysis were employed to characterize the topology of the ductal network in order to classify galactograms [6, 7]. Finally, a combination of texture and branching descriptors was investigated to deepen the understanding of the relationships among the morphology, function and pathology of the ductal tree [8].

III. METHODOLOGY

The proposed methodology is based on branching points density quantification schemes. It starts with preprocessing of the images to segment the tree-like structures from the background of the galactograms. Then, the characterization of the extracted tree topologies is performed by using both a variation of Sholl's analysis and a sectoring technique. In order to evaluate our approach, we perform similarity searches to find the trees that are most similar to a query, given a collection of tree structures and a query tree structure. We also use the k-nearest neighbor classification scheme to evaluate the effect of these approaches in classification accuracy.

A. Preprocessing

Certain preprocessing steps are necessary before the tree-like structures are available for analysis. At first, the boundaries of these structures need to be traced out of the background of the image. Since here we concentrate on the analysis of branching points distribution, we use a thinning process on the branches. The tree structures are reconstructed by identifying points of branching and resolving potential defects (such as anastomoses, occurring mostly as a result of 2D acquisition artifacts). In the application presented in this paper, we perform tracing and reconstruction of the breast ductal tree, using nipple as the root (Fig. 1.c). Since segmentation is beyond the scope of this paper, the steps described above were performed manually. Afterwards, the segmented trees are rotated into a standard position, so that the root is the leftmost node of the tree and the nearest one to the bottom of the image. This rotation works as a baseline orientation in order to perform the spatial node characterization techniques.

B. Characterization

In this section, we present two different descriptors for spatial branching points distribution of tree-like structures: a variations of Sholl's analysis and a sectoring technique. These characterization methods capture properties of the topological structure of the corresponding tree.

1) The Sholl Analysis

According to Sholl's methodology, a series of n equidistantly arranged concentric circles, which are centered in the root of the tree, is constructed [9]. The distance between the radii of two respective circles is constant. Let y the function that quantifies the surface density of the branches. Then, y is defined as

$$y(r_i)=\frac{N_i}{r_i^2\pi} \qquad (1),$$

where N_i is the number of intersections between the tree and the $i-th$ circle with radius r_i. The expression in the denominator of Eq.1 is the area S of the circle of radius r_i. The $y(r_i)$ is an inverse function of the circle area. Since in this study, we concentrate on the distribution of branching points, we use a variation of Sholl's method; instead of counting the number of intersections, we let N_i be the number of branching points included within the ring between the i-th and the (i+1)-th circle (Fig. 2). Finally, the branch distribution of a tree-like structure is encoded by a vector $\overrightarrow{V_{Sholl}} = [y(r_1), y(r_2), \dots, y(r_n)]$.

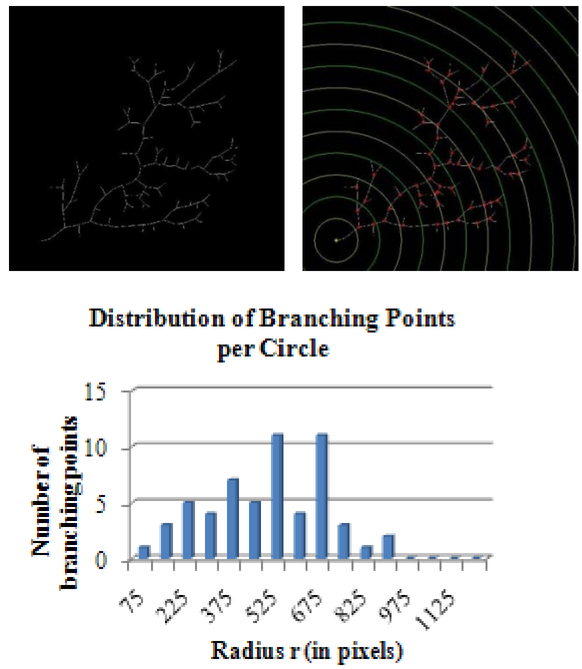

Fig. 2. (a) A segmented ductal tree network, (b) Application of the proposed Sholl variation over the tree structure, (c) Distribution of the tree branching points per circle of Sholl analysis.

2) The Sectoring Technique

In this technique the plane is divided into *n* neighboring sectors (i.e., triangular areas) originating from the root of the tree. A sector is a plane figure bounded by two radii and the included arc of a circle. The branching point distribution of a tree-like structure is encoded by the vector $\overrightarrow{V_{Sectoring}} = [y(n_1), y(n_2), \dots, y(n_n)]$, where $y(n_i)$ counts the number of branching points that are included in the area n_i (Fig. 3).

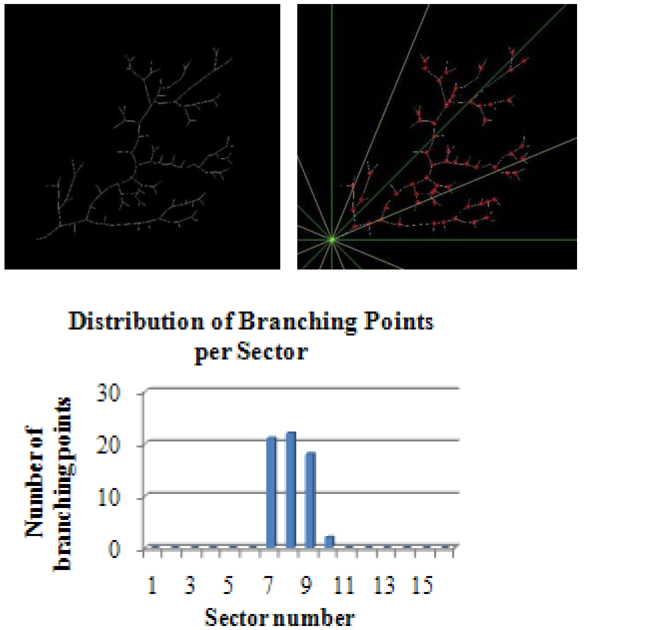

Fig. 3. (a) A segmented ductal tree network, (b) Application of the sectoring method over the tree structure, (c) Density of the number of branching points per sector.

A. *Similarity searches and Classification*

In order to evaluate our methods, we perform similarity searches and classification experiments. In similarity searches, we consider each tree as the query subject and we retrieve the k most similar trees based on a similarity metric. We calculate the pairwise similarity between all the trees of the dataset with both of our characterization methods. There are many ways to compute the similarity between two vectors; here, we use the cosine similarity metric:

$$\text{CosSim}\left(t_j, q\right) = \frac{\vec{t_j} \cdot \vec{q}}{|\vec{t_j}| \cdot |\vec{q}|} \qquad (2),$$

where q is the representative vector of the query tree and t_j is the corresponding vector of the tree j of the dataset.

As far the classification is concerned, we perform leave-one-out k-nearest neighbor experiments. The classes we consider are two: NF (no reported galactographic findings) images and RF (reported galactographic findings) images. For each test tree we retrieve the k closest neighbor trees according to cosine similarity metric. We assign the test tree the class that appears most frequently among its neighbors.

IV. RESULTS

In our study, we considered 21 x-ray galactograms, acquired at Thomas Jefferson University Hospital from a total of 15 women. From these images, 10 corresponded to women with no reported galactographic findings (NF) and 11 to women with reported findings (RF). The ductal trees were manually delineated and extracted (Fig. 1.b-c).

For the similarity experiments, we report the percentage of relevant trees among the retrieved ones (precision) averaged over all the similarity queries performed. As relevant trees we consider the trees belonging to the same class (NF or RF) as the query tree. Table 1 illustrates the similarity search results obtained when using the Sholl analysis, the Sectoring technique and two state-of-the-art methods; the tree asymmetry method and the Prüfer/tf-idf approach which are the best past methodologies in this field [5-6]. Precision was calculated as the proportion of neighboring images belonging to the same class as the query, averaged over the entire dataset. Considering the small size of our dataset, the parameter k ranges from 1 to 5. Table 2 illustrates the classification accuracy obtained when using the two proposed methods and the previous one. As in similarity cases, the parameter k takes values from 1 to 5.

Our results outperform pervious experimental results reported in the literature for the same dataset, where methods of tree encoding and the tf-idf text mining technique were used to classify the galactograms [5]. Furthermore, our results outperform the method in which the tree asymmetry index of the breast ductal trees was used to distinguish among two classes (NF vs. RF) in the same dataset of galactographic images [6].

V. CONCLUSION AND DISCUSSION

In this paper, we proposed an efficient way to detect early radiological findings in galactograms in order to assist the early diagnosis of breast cancer. We suggested two techniques for quantifying the similarity among tree-like structures such as the ductal trees appearing in the breast. The first method employs a variation of Sholl's analysis to characterize the spatial distribution of branching points across the various levels of the tree as one moves from the root of the tree toward the leaves. The second method uses a sectoring technique to characterize the spatial distribution of branching points as one scans the tree along the direction of the branches moving in a direction which is perpendicular to that of Sholl's analysis. Our methodology was applied to breast ductal trees extracted from clinical x-rays galactograms. The images were divided into two groups; those with no reported galactographic findings (NF) and those with reported findings (RF). We performed similarity searches and classification experiments with both techniques. The experimental results illustrate the potential of the proposed tree characterization and classification frameworks for the analysis of tree-like structures in medical images. The proposed schemes outperform in classification accuracy and similarity search precision the competitors. Concluding, one possible direction for future work is the integration of the two methods and the development of features that represent other branching properties, such as branch length, angle, and tortuosity to further increase the classification accuracy and precision in similarity searches. The proposed methodology can potentially increase the diagnostic accuracy and assist the medical doctors in the interpretation of galactographic images by correlating the branching points distribution of ductal network with the radiological findings regarding breast cancer.

Similarity searches												
k	Precision											
	Sholl			Sectoring			Prüfer/tf-idf			Tree asymmetry		
	Total	NF	RF	Total	NF	RF	Total	NF	RF	Total	NF	RF
1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2	87,1%	76,2%	97,9%	92,1%	86,2%	97,9%	74,5%	84,2%	64,9%	83,6%	87,9%	83,3%
3	71,1%	61,6%	80,5%	77,5%	76,6%	78,4%	63,5%	73,1%	53,9%	71,2%	76%	67,8%
4	63,1%	58,1%	68,2%	70,3%	75%	65,6%	61,9%	68,3%	55,6%	62,1%	67,7%	56,5%
5	58,9%	55%	62,9%	67,9%	75,5%	60,4%	58,1%	64%	52,1%	58,1%	63,1%	53,1%

Table 1. The precision for similarity search experiments when using the proposed characterization methods, i.e., Sholl's analysis and sectoring compared to that obtained by Prüfer encoding with tf-idf and tree asymmetry index.

Classification												
k	Accuracy											
	Sholl			Sectoring			Prüfer/tf-idf			Tree asymmetry		
	Total	NF	RF	Total	NF	RF	Total	NF	RF	Total	NF	RF
1	61,6%	61,5%	61,7%	59,5%	76%	43%	41%	49,1%	32%	61,1%	58,3%	64%
2	75,5%	70%	81%	85,6%	100%	71,2%	67,8%	75,1%	60%	81,3%	72,9%	89,6%
3	38,6%	19%	58,2%	58,7%	82,7%	34,7%	40,1%	49,6%	32%	38,9%	43,9%	33,9%
4	57%	47%	67%	75,5%	93,5%	57,5%	64,5%	71,6%	57,2%	58,1%	68,8%	47,3%
5	20,6%	10,6%	34,7%	56,6%	85,5%	27,7%	43,7%	49,2%	38,2%	44,7%	52,1%	37,3%

Table 2. Comparison of the accuracy for classification experiments amongst the two proposed methods i.e., Sholl's analysis and sectoring and previously proposed Prüfer encoding with tf-idf and tree asymmetry index.

ACKNOWLEDGEMENT

This work was supported in part by NSF under grant IIS-0237921. The funding agency specifically disclaims responsibility for any analyses, interpretations and conclusions. The authors are grateful to Dr. Predrag R. Bakic and Dr. Andrew D.A. Maidment for providing anonymized x-ray galactograms and for assisting in preprocessing the images.

REFERENCES

[1] W.D. Foulkes, "Inherited susceptibility to common cancers," The New England Journal of Medicine, vol. 359(20), pp. 2143-53, 2008.

[2] DB. Kopans, "Breast imaging," 2nd ed., Ed. Philadelphia: Lippincott-Raven, 1998.

[3] J. Cuzick, T. Powles, U. Veronesi, J. Forbes, R. Edwards, S. Ashley and P. Boyle, "Overview of the main outcomes in breast-cancer prevention trials", The Lancet, vol. 361(9354), pp. 296-300.

[4] P.R. Bakic, M. Albert and A.D. Maidment, "Classification of galactograms with ramification matrices: preliminary results," Academic Radiology, vol. 10, pp. 198-204, 2003.

[5] V. Megalooikonomou, M. Barnathan, D. Kontos, P.R. Bakic and A.D. Maidment, "A Representation and Classification Scheme for Tree-like Structures in Medical Images: Analyzing the Branching Pattern of Ductal Trees in X-ray Galactograms," IEEE Transactions in Medical Imaging, (in press).

[6] A. Skoura, M. Barnathan and V. Megalooikonomou, "Classification of ductal tree structures in galactograms," in 2009 Proc. 6th IEEE International Symposium on Biomedical Imaging, (to appear).

[7] D. Kontos, V. Megalooikonomou, A. Javadi, P.R Bakic and A.D. Maidment, "Classification of galactograms using fractal properties of the breast ductal network,", in 2006 Proc. 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1324-1327.

[8] M. Barnathan, J. Zhang, D. Kontos, P. Bakic, A.D. Maidment and V. Megalooikonomou, "Analyzing Tree-Like Structures in Biomedical Images Based on Texture and Branching: An application to Breast Imaging," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5116 LNCS, pp. 25-32

[9] D.A.Sholl, "Dendritic organizations in the neurons of the visual and motor cortices of the cat," J. Anatomy, vol. 87, 1953, pp. 387-406.