33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Hybrid Cosine and Radon Transform-based processing for Digital Mammogram Feature Extraction and Classification with SVM

Salim Lahmiri and Mounir Boukadoum, *Senior Member, IEEE*

Abstract—A new methodology to automatically extract features from mammograms and classify them is presented. It relies on a hybrid processing system that sequentially uses the discrete cosine transform (DCT) to obtain the high frequency component of the mammogram and then applies the Radon transform to the obtained DCT image in order to extract its directional features. The features are subsequently fed to a support vector machine for classification. The approach was tested on a database of one hundred images and shows improved classification accuracy in comparison to using the discrete cosine transform or the Radon transform alone, as done in others works.

I. Introduction

BREAST cancer is one of the leading causes of death for women. Currently, the detection of microcalcifications (MC) is the breast tissue is one important mean of detecting it. Microcalcifications appear in the mammogram as small bright spots that are either scattered or grouped in clusters, and their presence may be indicative of early breast cancer. In the last decade, many Computer-Aided Diagnosis (CAD) Systems have been proposed to help radiologists determine the type of MC in a mammogram and potentially reduce the number of diagnostic errors [1, 2]. In general, a CAD system for breast cancer screening from mammograms operates in three stages: a) Specification of regions of interest (ROI); b) image processing for feature extraction and selection; d) ROI classification. The ROI specification step corresponds to image segments that may contain suspicious clusters as determined by a radiologist or an automated algorithm. Then, the image processing step performs filtering of the ROI, followed by feature extraction and selection to enable the distinction of benign and malignant tissue. This is the role of the classification task.

Many signal processing techniques have been employed to process mammograms for feature extraction and suspicious MC detection. They include multi-resolution analysis tools such as the discrete wavelet transform (DWT) [1], the wavelet packet transform (WPT) [2], Gabor filter banks [3], and the dual-tree complex wavelet transform (DT-CWT) [4]. There have also been uses of alternative techniques with lesser computational costs and easier parameter setting. This is the case of frequency domain transforms such as the discrete cosine transform (DCT) [5]-[9] and spatial transforms such as the Radon transform (RT) [10]-[12]. The discrete cosine transform (DCT) allows representing an image as a sum of sinusoids with varying magnitudes and frequencies, thereby converting the spatial information of the image into a frequency spectrum with real-valued coefficients. Most often, only a few of the obtained frequency components are needed to provide an accurate and compressed description of the image, hence the usefulness of DCT for mammogram processing. As for RT, it changes an image representation from Cartesian coordinates to intensity averages along various orientations in the image plane, using a polar representation [10]. Because of this, RT offers readily-available and valuable information about oriented patterns in mammograms [11]. Other useful features of RT are its invariance to image rotation and scaling [11].

Two other transforms that offer directional information are the Gabor filter banks and dual-tree complex wavelet transform. However, they have drawbacks in comparison to RT. For instance, the outputs of Gabor filter banks are not mutually orthogonal and possible significant correlation between texture features can be found as a result. There is also the need for an optimal tuning of their parameters for different frequencies and orientations. On the other hand, the DT-CWT needs a priori knowledge of the appropriate decomposition level and the type of filter to be employed. Finally, both Gabor filter banks and DT-CWT carry a high computational cost as opposed to the Radon transform.

Several efforts have been made to use DCT or RT for mammogram processing. In [8], the five most significant coefficients of a 64-point DCT, as determined by the fisher criterion, were fed to a three-layer feed forward neural network with error back-propagation training (MLP-BP) to detect MCs. A 100% classification accuracy was reported after studying a database of 40 mammograms. In [9], a 61-feature vector is formed to represent textural, spatial and spectral properties of small ROIs; the spectral domain information consisted of the block average and spectral entropy of a 16×16 DCT. The feature vectors were fed to both support vector machines (SVM) and generalized regression neural networks (GRNN), and the obtained performance was 98% with SVM and 97.80% with GRNN when processing a database composed of 7531 image blocks. In [10], the authors used RT do determine the block energy along 8 orientations and fed the resulting feature vector to a MLP-BP network to classify normal versus cancer images. The obtained correct detection rate was 88% for a database of 200 image blocks. In [11], the mean, variance, skewness, kurtosis, and entropy computed from RT directions were fed to a Kohonen self-organizing map to

Manuscript received March 25, 2011.

Mounir Boukadoum is full professor in the department of computer science at the University of Québec at Montréal. C. P. 8888, Succursale C.-V., Montréal (Québec), H3C 3P8 Canada (e-mail: boukadoum.mounir@uqam.ca).

Salim. Lahmiri is a PhD candidate in the same department (e-mail: lahmiri.salim@courrier.uqam.ca).

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

classify mammograms. The achieved correct recognition rate was 83% using a database of 1080 mammograms.

This paper proposes a mammogram processing and classification system that uses both the DCT and Radon transforms, and statistical features derived from them for the classification task with SVM. A desirable feature of the system is that no prior mammogram decomposition into image blocks is used, leading to a reduced computational load. As stated earlier, DCT is powerful at providing a compressed image representation with real-valued coefficients. Also, its high frequency component coefficients can play a dominant role in isolating suspicious MCs [6], and it enhances the scale invariance properties of feature sets from biomedical images [13]. However, DCT lacks directional information capability, an important feature for diagnosing medical images [14]. As a result, it is expected that a DCT-Radon processing system should provide both efficient high frequency mammogram information thanks to the DCT, and directional representation thanks to the RT, all with reasonable computational cost. Sequentially applying the two transforms prior to feature extraction should thus improve mammogram processing and lead to more accurate MC classification in comparison to using DCT or RT alone. A SVM is used for the classification task thanks to its demonstrated efficiency [9]; among the interesting features of a SVM with respect to other classifiers are the ability to avoid local minima and scalability.

The paper is organized as follows: The proposed methodology is described in Section II; the experimental results of using it and alternative approaches are presented in Section III. Finally, a conclusion follows in Section IV.

II. METHODOLOGY

The proposed methodology is as follows:

1. The DCT is applied to a mammogram to obtain a high frequency component image since MC are found in dense biological tissue, which corresponds to high frequencies in the image spectrum [2][6].
2. The Radon transform with different orientations is applied to the obtained DCT high frequency image to obtain RT-filtered images that provide directional information.
3. Statistical features are computed from the Radon-processed images.
4. SVM is used to classify the extracted feature for final diagnosis.

The schematic diagram of the proposed system is shown in Fig.1, and those of only using DCT [7][8] or RT [10][11] are summarized in Fig. 2 and Fig. 3, respectively, for comparison. Finally, in order to study the impact of using one transform before the other, experiments were also conducted where RT is applied first, and its resulting directional images processed by DCT (Fig. 4). The Matlab Image Processing Toolbox was used for computations.

A. Discrete cosine transform

As mentioned before, the discrete cosine transform converts the spatial representation of an image into the frequency domain. The two-dimensional DCT of an $M \times N$ image $I(m,n)$ is defined as follows:

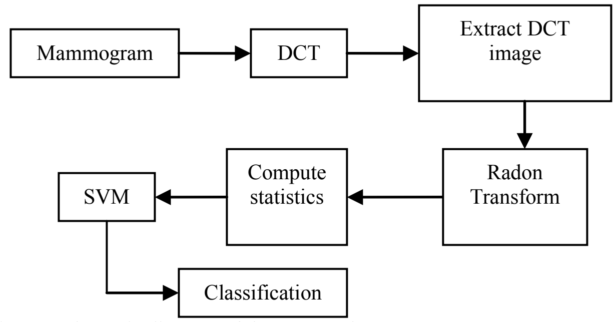

Fig. 1. Schematic diagram of the proposed DCT-RT system

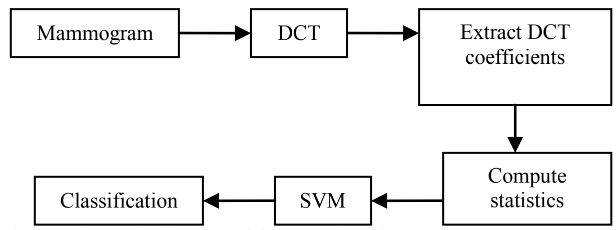

Fig. 2. Schematic diagram of the DCT-based system

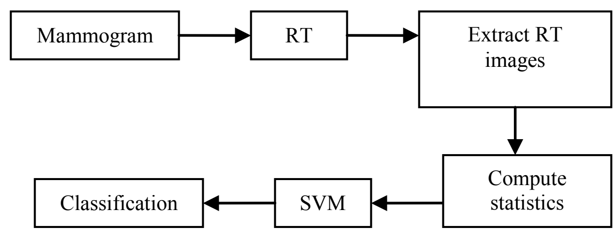

Fig. 3. Schematic diagram of the RT-based system

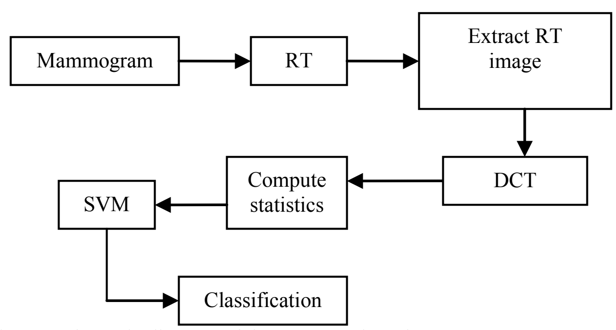

Fig. 4. Schematic diagram of the RT-DCT-based system

$$C_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m,n) \cos\left(\frac{\pi(2m+1)p}{2M}\right) \cos\left(\frac{\pi(2n+1)q}{2N}\right), \quad \begin{array}{l} 0 \le p \le M-1 \\ 0 \le q \le N-1 \end{array} \tag{1}$$

where, C_{pq} are the coefficients of the discrete cosine transform, and the parameters α_p and α_q are defined by:

$$\alpha_p = \left\{ \frac{1}{\sqrt{M}} \text{ if } p = 0; \ \sqrt{\frac{2}{M}} \text{ if } 1 \le p \le M-1 \right\} \tag{2}$$

$$\alpha_q = \left\{ \frac{1}{\sqrt{N}} \text{ if } q = 0; \ \sqrt{\frac{2}{N}} \text{ if } 1 \le q \le M-1 \right\} \tag{3}$$

B. Radon transform

The Radon transform [10]-[12] describes an image in terms of the sum of pixel intensities along lines pointing in various directions. The continuous radon transform of an image $I(x,y)$ is given by:

$$R(\rho,\theta) = \int_{y=-\infty}^{y=+\infty}\int_{x=-\infty}^{y=+\infty} I(x,y)\delta(\rho - x\cos(\theta) - y\sin(\theta)).dx.dy \quad (4)$$

where δ(.) is the Dirac delta function and $\rho = x.\cos(\theta) + y.\sin(\theta)$ defines the perpendicular distance of all lines in the image plane which form an angle $\theta \in [0,\pi]$ with respect to the x-axis. Consequently, $R(\rho,\theta)$ represents scans of $I(x,y)$ over the infinite set of lines defined by $\rho - x.\cos(\theta) + y.\sin(\theta) = 0$. In this study, four orientations are used as in [3]: $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$ so that the application of the Radon transform to a DCT-filtered image leads to four separate one-dimensional results.

C. Features extraction

Two commonly used statistics were extracted from the DCT and RT-processed images. Entropy helps distinguish homogenous and heterogeneous biological tissue [15] and energy is suitable to detect topological changes and asymmetry in cancer images [16]. Thus, for each DCT image processed by the Radon transform, an 8-element feature vector is formed (four DCT-RT orientations, each one with one value of entropy and one of energy) and written as:

$$\boldsymbol{x}_{DCT\text{-}RT} = [e_{\text{DCT-RT},1}, e_{\text{DCT-RT},2}, e_{\text{DCT-RT},3}, e_{\text{DCT-RT},4}, E_{\text{DCT-RT},1}, E_{\text{DCT-RT},2}, E_{\text{DCT-RT},3}, E_{\text{DCT-RT},4}] \quad (5)$$

Similarly, the entropy and energy are extracted from each RT image processed by the DCT in the RT-DCT-based approach shown in Fig. 4, leading to the following feature vector:

$$\boldsymbol{x}_{RT\text{-}DCT} = [e_{\text{RT-DCT},1}, e_{\text{RT-DCT},2}, e_{\text{RT-DCT},3}, e_{\text{RT-DCT},4}, E_{\text{RT-DCT},1}, E_{\text{RT-DCT},2}, E_{\text{RT-DCT},3}, E_{\text{RT-DCT},4}] \quad (6)$$

For the standard DCT-based (Fig. 2) and RT-based approaches (Fig. 3), the extracted feature vectors are respectively:

$$\boldsymbol{x}_{DCT} = [e_{\text{DCT}}, E_{\text{DCT}}] \quad (7)$$

$$\boldsymbol{x}_{RT} = [e_{\text{RT},1}, e_{\text{RT},2}, e_{\text{RT},3}, e_{\text{RT},4}, E_{\text{RT},1}, E_{\text{RT},2}, E_{\text{RT},3}, E_{\text{RT},4}] \quad (8)$$

D. Support Vector Machines

Support Vector Machines (SVM) [17] are employed to distinguish normal from malignant images, using the obtained feature vectors as inputs. The discriminant function of the non-linear SVM for a binary classification problem is given by:

$$g(\boldsymbol{x}) = sign\left(\sum_{i=1}^{S} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \quad (9)$$

where x_i is the training data that belong to either {+1,-1}, S is the training data size, α_i are Lagrange multipliers subject to $0 < \alpha_i < c$, b is a bias weight, $K(.)$ is the kernel function and c is a parameter that influences the tolerance to misclassifications. In this study, a polynomial kernel is used for the SVM since it is a global kernel, thus allowing data points that are far away from each other to have an influence on the kernel values as well. The polynomial kernel is given by:

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = ((\boldsymbol{x}_i \cdot \boldsymbol{x}) + 1)^d \quad (10)$$

where the kernel parameter d is the degree of the polynomial to be used; d is set to 2 in this study.

III. EXPERIMENTAL RESULTS

In order to investigate the performance of the combined DCT and Radon transforms for feature extraction, one hundred digital mammograms were taken from The Digital Database for Screening Mammography (DDSM) [18]. They consisted of fifty normal images and fifty cancer images. An example of a digital mammogram is shown in Fig. 5, and the Radon transform spectra of its DCT signal is given in Fig. 6.

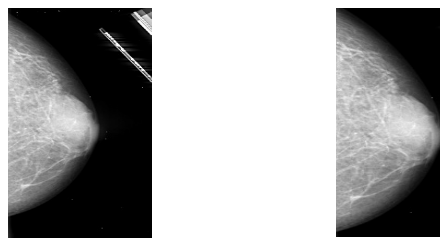

Fig. 5. Example of a normal mammogram. Left: original mammogram; right: ROI image for processing (96×322 pixels).

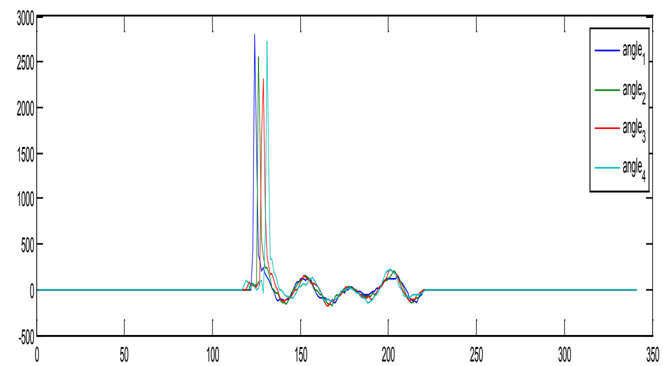

Fig. 6. RT of DCT image as function of perpendicular line distance from the origin X' = $\rho = x.\cos(\theta) + y.\sin(\theta)$, for $\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$

All four processing systems illustrated in Fig. 1 to 4 were used in turns, with 10-fold cross-validation used in each experiment. The ten folds corresponded to a random split of the one hundred mammograms into groups of ten, after which the SVM was trained ten times, each once with nine folds used for learning and the remaining one for testing (in other words, one of the folds was kept out from training in rotation). For each fold, the correct classification rate as defined in [3][4][7][10]-[13] and its standard deviation were computed. Fig. 7 provides the obtained results. It shows that the feature extraction approach based on combining the DCT and Radon transform (RT) improved the classification accuracy with respect to using DCT or RT only. For instance, when using only RT-based feature extraction, the achieved average accuracy was 67.76% (±0.04), and when using only DCT-based feature extraction, it was 88.13% (±0.02). On the other hand, applying the Radon transform to the DCT signal before feature extraction allowed an accuracy improvement by more than four percentage points overall: the obtained classification accuracy with DCT-RT was 92.98% (±0.06).

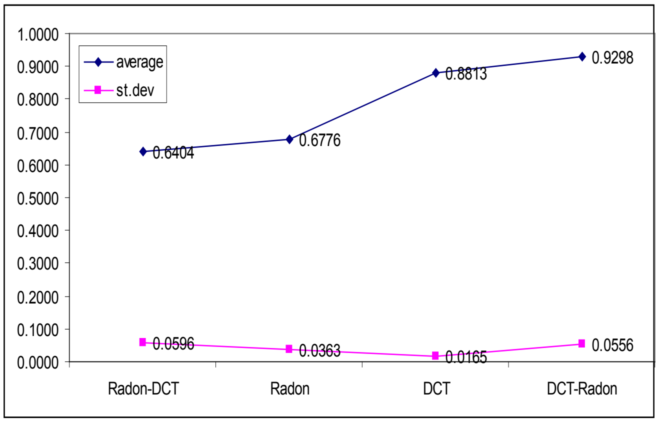

Fig. 7. Correct classification rates given 10-folds cross-validation.

This is no longer true if the order of the DCT and RT operations is reversed; Fig. 7 shows that the Radon-DCT approach led to the worst classification performance since the achieved average detection rate was only 64.04% (±0.06). This is intuitively understandable since the radon data only corresponded to four orientations in this study and performing a spectral analysis of such sparse data does not lead to really useful information. In sum, it is more suitable to transform the mammogram in the frequency domain and then apply Radon transform to obtain directional features from the high frequency representation of the original mammogram.

IV. Discussion and Conclusion

The obtained results are very promising given the relative computational complexity of the DCT-RT approach and the near 93% classification accuracy it achieved. These are to be compared to alternative studies with multiresolution transforms based on the Gabor filter or wavelet transforms (WT). For instance, works based on Gabor filter banks [3], dual-tree continuous wavelet transform [4], Radon transform [10], and complex wavelet transform [13] reported respective correct classification rates of 73%, 88%, 80%, and 87%. However, these results were obtained with different mammogram databases, making it difficult to make objective comparisons between the different techniques. Nevertheless, they can provide a crude estimate of the merit of the proposed DCT-Radon approach.

In summary, we proposed a supervised-learning processing system for mammogram classification that uses statistical features obtained from the sequential combination of the discrete cosine transform and the Radon transform to classify normal and cancer images, using support vector machines as classifiers. Our validation results show that this hybrid processing model allows achieving higher classification accuracy then when using the discrete cosine transform or the Radon transform alone. In other words, directional features regarding frequency domain information appear to help improve the detection of suspicious mammograms. In future work, more features and angles will be examined to investigate this further.

References

[1]B.L. Mann and S. Nithya,"Using Neural Networks with Wavelet Transforms for an Automated Mammographic Mass Classifier,'' IEEE Proc of the 22"d Annual EMBS international Conference, July 23-28, Chicago IL, pp. 985-987, 2000.

[2]M.H.J. Sepehr, R-R. Gholamali and H. Behnam,"A Novel Method for Breast Cancer Prognosis Using Wavelet Packet Based Neural Network,'' Proc. of the IEEE Engineering in Medicine and Biology 27th Annual Conference, September 1-4, Shanghai, China, pp. 3414-3417, 2005.

[3]I. Buciu and A. Gacsadi,"Gabor Wavelet Based Features for Medical Image Analysis and Classification," IEEE 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), November 24-27, Bratislava, Slovak Republic, pp. 1–4, 2009.

[4]A. Tirtajaya and D. Santika,"Classification of Microcalcification Using Dual-Tree Complex Wavelet Transform and Support Vector Machine," IEEE Second International Conference on Advances in Computing, Control, and Telecommunication Technologies, December 2-3, Jakarta, Indonesia, pp. 164-166, 2010.

[5] J. I. Arribas, C. Alberola-Upez, M. Mateos-Marcos, and J. Cid-Sueiro, ''Neural Posterior Probabilities for Microcalcification Detection in Breast Cancer Diagnoses,'' Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering, Capri Island, Italy March 20-22, pp.660-663, 2003.

[6] C-M. Tiu, T-L. Jong, and C-W. Hsieh, ''Self Organizing Map Neural Network with Fuzzy Screening for Micro-Calcifications Detection on Mammograms,'' IEEE Conference On Soft Computing in Industrial Applications, June 25-27, Muroran, JAPAN, pp. 421-425, 2008.

[7] T.M. Lehmann, M.O. Güld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B.B. Wein, ''Automatic Categorization of Medical Images for Content-Based Retrieval and Data Mining,'' Computerized Medical Imaging and Graphics, Vol. 29, pp.143-155, 2005.

[8] A. Farag and S. Mashali, ''DCT Based Features for the Detection of Microcalcifications in Digital Mammograms,'' IEEE International Symposium on Micro-Nano Mechatronics and Human Science, December 30, Cairo, Egypte, Vol. 1, pp: 352-355, 2003.

[9] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, and H.K. Wu, '' Image Segmentation Feature Selection and Pattern Classification for Mammographic Microcalcifications,'' Computerized Medical Imaging and Graphics, Vol. 29, pp. 419-429, 2005.

[10] L. Bocchi and J. Nori, ''Shape Analysis of Microcalcifications using Radon Transform,'' Medical Engineering & Physics, Vol. 29, pp. 691-698, 2007.

[11]S.K. Kinoshita, P.M. de Azevedo-Marques, R.R Pereira Jr, J.A.A.H. Rodrigues, and R.M. Rangayyan, ''Content-Based Retrieval of Mammograms Using Visual Features Related to Breast Density Patterns,'' Journal of Digital Imaging. Vol 20, No 2, pp. 172-190, 2007.

[12]X. Zhang, X. Gao, and Y. Wang, ''MCs Detection with Combined Image Features and Twin Support Vector Machines,'' Journal of Computers, Vol. 4, No. 3, pp: 215-221, 2009.

[13]M. Hafner, A. Uhl, A. Vecsei, G. Wimmer, and F. Wrba, ''Complex Wavelet Transform Variants and Discrete Cosine Transform for Scale Invariance in Magnification-Endoscopy Image Classification,'' 10th IEEE International Conference on Information Technology and Applications in Biomedicine, November 3-5., Vienna, Austria, 2010.

[14]S. Banik, R.M. Rangayyan and J.E.L. Desautels, "Detection of Architectural Distortion in Prior Mammograms of Interval-Cancer Cases with Neural Networks," IEEE 31st Annual International Conference of the EMBS, September 2-6, Minneapolis, Minnesota, USA, pp. 6667-6670, 2009.

[15]Z. Chen and R. Ning, "Breast Volume Denoising and Noise Characterization by 3D Wavelet Transform," Computerized Medical Imaging and Graphics, vol.28, pp.235-246, 2004.

[16] J. Tang, R. Rangayyan, J. Yao, Y. Yang, ''Digital Image Processing and Pattern Recognition Techniques for The Detection of Cancer,'' Pattern Recognition, Vol. 42, pp.1015-1016, 2009.

[17]V.N Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, 1995.

[18]http://marathon.csee.usf.edu/Mammography/Database.html