A METHOD FOR MICROCALISIFICATIONS DIAGNOSIS IN BREAST MAMMOGRAMS

Abbas Hanon AL-Asadi
Abbashh2002@yahoo.com
Ahmed Kazim Hamed Al-Saadi
ahmed_kazim2007r@yahoo.com
Computer Science Department, Science College, Basrah University,Basrah, Iraq

Abstract

Breast cancer is the main cause of death for women above age 40. In this paper a method is proposed to develop a Computer-Aided Diagnosis (CADx), this method provide a second opinion in microcalcifications diagnostic and making decisions (classify microcalcifications as benign or malignant). The proposed method for microcalcifications diagnosis splits into three steps: The first step extracts the region of interest (ROI). The second step is the features extraction, where we used a set of features from (ROI) by applying wavelet decomposition. The third step is the classification process where discrimination between benign and malignant is performed using a Nearest Neighbor Classifier. The proposed method was evaluated using the Mammographic Image Analysis Society (MIAS) mammographic databases. The proposed method has achieved satisfactory results.

Keywords: Breast Cancer, computer aided diagnosis, Microcalcifications, Mammography, Wavelet Transform

الملخص :

سرطان الثدي هو أحد الأمراض الخطيرة التي تهدّد حياة النساء أكثر من الأمراض السرطانية الأخرى، ويعد الكشف والمعالجة المبكرة لهذا المرض من الطرق الفعّالة لتقليل عدد الوفيات الحاصل بسببه لأن المعالجة قبل أن يتطور المرض وينتشر سوف لا تتطلب تدخل جراحي هائل. التصوير الشعاعي للثدي (Mammography) هو تقنية تستخدم جرعة صغيرة من الأشعة السينية (x-ray) من أجل تصوير نوعي للثدي والمساعدة في الكشف المبكر لسرطان الثدي وعادةً ما تكون النساء لا تشعرن بأي أعراض.شكل ظهور التكلسات المايكروية (Microcalcifications) في صور الثدي الشعاعية (Mammograms) أحد المؤشرات المبكرة لمرض سرطان الثدي. تظهر التكلسات المايكروية على شكل جزئيات لامعة مختلفة الأشكال والأحجام وقد توضعت على الخلفية الشعاعية لنسيج الثدي.

على الرغم من أن تقنية التصوير الشعاعي بأشعة (x-ray) أثبتت بأنها طريقة التصوير الشعاعية الأنجع من اجل الكشف المبكر عن سرطان الثدي، إلا أن هناك بعض المشاكل والصعوبات الناجمة عن تراكب الأنسجة المختلفة في صورة الماموكرام مما يؤدي إلى أخطاء في كشف هذا المرض وتشخيصه لذلك يمكن لتقنية الحاسوب أن تحسن دقة الكشف عن هذا المرض . أن الكشف والتشخيص بمساعدة الحاسوب (CADx) هي تقنيات توفر قارئ ثاني لتنبيه الطبيب إلى مناطق قد يغفل عنها واتخاذ قرار التشخيص بتصنيف التكلسات الميكروية على أنها خبيثة أو حميدة. في هذا البحث تم اقتراح طريقة للكشف والتشخيص المبكر لسرطان الثدي. هذه الدراسة تقترح طريقة لتشخيص التكلسات المايكروية تتألف من ثلاثة خطوات .الخطوة الأولى ويتم فيها استخراج المنطقة المهمة من الصورة (ROI)، أما الخطوة الثانية فهي للحصول على الخصائص (Features) باستخدام التحليل الموجي (wavelet decomposition) وإمرارها إلى خطوة التصنيف وهي الخطوة الثالثة ويتم فيها التمييز بين التكلسات المايكروية الخبيثة والتكلسات المايكروية الحميدة باستخدام المصنف الجار الأقرب Nearest Neighbor Classifier (NN).

1. INTRODUCTION

Breast cancer is a malignant tumor that starts in the cells of the breast [1]. In modern countries, medical statistics have estimated that one woman on eight will contract a breast cancer. Malign tumors are more likely to appear in breast tissues of women above age 40, and they represent 35% of the abnormalities detected in women breasts. Breast cancer is currently responsible for more than 30% of death by cancer in women, which is about 1% of all deaths worldwide. Men are also concerned by breast cancers since it represents 1.5% of all cancer death in men [2, 3].

The early detection of breast cancer is very important because the treatment of an undeveloped and non-metastasized tumor will not require massive surgical interventions.

Microcalcifications are deposits of calcium ($Ca_5 (PO_4)_3OH$) in breast tissue [4] .It is small size lesions, typically range in size from 0.05 to 1 mm in diameter. With these dimensions, microcalcifications are relatively difficult to detect [1, 5].

There is a high correlation between the presence of microcalcifications and breast cancer, particularly when a number of microcalcifications grouped together is termed a cluster and it may be a strong indication of cancer. A cluster is defined as at least three microcalcifications within a 1 cm^2 area. Therefore, an accurate detection of microcalcifications is essential to any early detection of the majority of breast cancers [6, 7].

In the literature, various numbers of techniques are described to detect and classify the presence of microcalcifications in digital mammograms as benign or malignant.

Chan et al. [8- 10] investigated a computer-based method for the detection of microcalcifications in digital mammograms. The method is based on a difference image technique in which a signal suppressed image is subtracted from a signal enhanced image to remove structured background in the mammogram. Global and local thresholding techniques are then used to extract potential microcalcifications signals. Subsequently, signal extraction criteria are imposed on the potential microcalcifications to distinguish true positives from noise and artifacts.

Karssemeijer [11- 13] developed a statistical method for detection of microcalcifications in digital mammograms. The method is based on the use of statistical models and the general framework of Bayesian image analysis.

Strickland et al. [14-19] developed a method based on undecimated biorthogonal wavelet transforms and optimal subband weighting for detecting and segmenting clustered microcalcifications.

Yoshida et al. [20, 21] used decimated wavelet transform and supervised learning for the detection of microcalcifications.

Cheng et al. [22] proposed an approach using fuzzy logic for the detection of microcalcifications.

Yu and Guan [23] developed a CAD system for the automatic detection of clustered microcalcifications through two steps. The first one is to segment potential microcalcifications pixels by using wavelet and gray level statistical features and to connect them into potential individual microcalcifications objects. The second step is to

check these potential objects by using 31 statistical features. Neural network classifiers were used.

Jiang et al. [24] Proposed Genetic Algorithm (GA) technique which is characterized by transforming input images into a feature domain, where each pixel is represented by its mean and standard deviation inside a surrounding window of size 9x9 pixe1. In the feature domain, chromosomes are constructed to populate the initial generation and further features are extracted to enable the proposed GA to search for optimized classification and detection of microcalcifications clusters via regions of 128x 128 pixels.

2. MAMMOGRAM DATABASE

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database (ftp://peipa.essex.ac.uk).

The database used in experiments of this work was taken from the MIAS because it contains complete information about abnormalities of each mammographic image. The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Program and digitized with a Joyce-Lobel scanning microdensitometer to a resolution of 50 μm × 50 μm, 8bits represent each pixel with 1024×1024 pixel size and at 256 gray levels.

The images are in the grayscale file format (.pgm) (Portable Graymap). The used database contains left and right breast images for 161 patients. Its quantity consists of 322 images, which belong to three types such as Normal, benign and malignant. There are 208 normal, 63 benign and 51 malignant (abnormal) images.

3. WAVELET ANALYSIS

An important branch of CADx methods in mammography employs wavelet transforms for feature enhancement. This work uses the Two-Dimensional Discrete Wavelet Transform (2D-DWT), which can be defined as:

$$C(a,b)=C(j,k)=\sum_{x\in Z}\sum_{y\in Z} f(x,y)g_{jk}(x,y) \quad (1)$$

With **a** = 2^j , **b** = k.2^j , j,k ∈ N, where f is the original image, g is the wavelet function, **a** is a scale factor of the wavelet function, **b** is a location parameter of the wavelet function, and C(a,b) is the set of obtained coefficients. Inverse process is calculated by:

$$f(x,y)=\sum_{x\in Z}\sum_{y\in Z} C(j,k)\psi_{jk}(x,y) \quad (2)$$

Where ψ is the wavelet function used to reconstruct the image.

The multiresolution representation carried out by 2D-DWT fragment the frequency spectrum of an image f(x,y) into a low pass subband image cA^j and a set of band-pass subband images cDH^j, cDV^j, cDD^j, j = 1,…,L, where L denote the number of levels for a representation, cDH^j is formed by low pass filtering the rows followed

by high pass filtering the columns, and is therefore sensitive to horizontally oriented features. In the same way the cDV^j contains vertically oriented structure, and cDD^j contains primarily diagonal structure [25].

Generally speaking, multiresolution representations are implemented by a cascade of analysis/synthesis (A/S) filter banks. The discrete wavelet transform uses two different wavelet mothers: h(x) for multiresolution decomposition (analysis) and g(x) for reconstruction (synthesis) of the original image from its multiresolution representation. An efficient way to implement discrete wavelet transform using filters was developed by Mallat [26].

Figure (1) shows the implementation of a one-level (L = 1) multiresolution representation of the discrete wavelet transform, which divides orientations into three bands.

As seen in Figure (1), the forward 2D wavelet transform is implemented using a bank of 1D low pass ($h_1(x)$) and high pass ($h_2(x)$) analysis filters. The reconstruction process, or inverse wavelet transform, is likewise computed via 1D synthesis filters, $g_1(x)$ and $g_2(x)$.

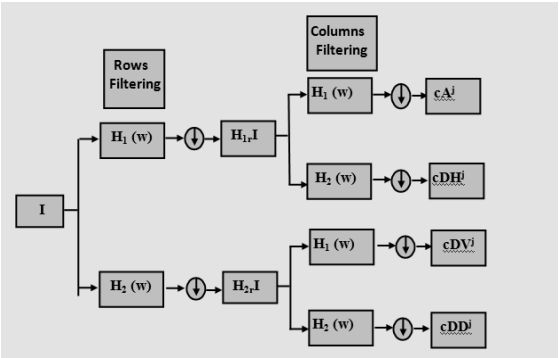

Figure (1): Wavelet decomposition algorithm

Wavelet-based image decomposition can be interpreted as an image filtering process. For a given image I of size 2^n x 2^n, wavelet-based subband decomposition can be performed as follows: The wavelet filters $h_1(x)$ and $h_2(x)$ are applied to the rows of the image I. The filter $h_1(x)$ is a low-pass filter with frequency response $H_1(w)$ and $h_2(x)$ is a high pass filter with frequency response $H_2(w)$. By filtering the image I with $H_1(w)$, low-frequency is obtained information (background).

By filtering the image with $H_2(w)$, the high-frequency information is obtained (edges). After downsampling by a factor of two, two subbands are obtained: H_1rI and H_2rI (the subscript r suggests that the filters are applied to rows of the image I). The filters $H_1(w)$ and $H_2(w)$ are then applied to the columns of the subbands H_1rI and H_2rI, followed by downsampling by a factor of two, and the following four subbands are obtained: cA^j, cDH^j, cDV^j and cDD^j. The subband cA^j contains the smooth information of the image, and the subbands cDH^j, $cDV^j$and cDD^j contain the detail information of the image. Then the cA^j subband of the frequency domain is segmented into four subbands at the second level decomposition, and so on.

4. MICROCALCIFICATIONS DIAGNOSIS METHOD

In this method, the first step is to apply a preprocessing to improve the edge of breast and then segmentation process (Region of interest) for eliminating some regions in the image, which are not useful for the mammographic interpretation.. Then features are extracted after applied DWT and finally classification process is performed for classifying the microcalcifications. Flowchart in Figure (2) depicts the steps of this method.

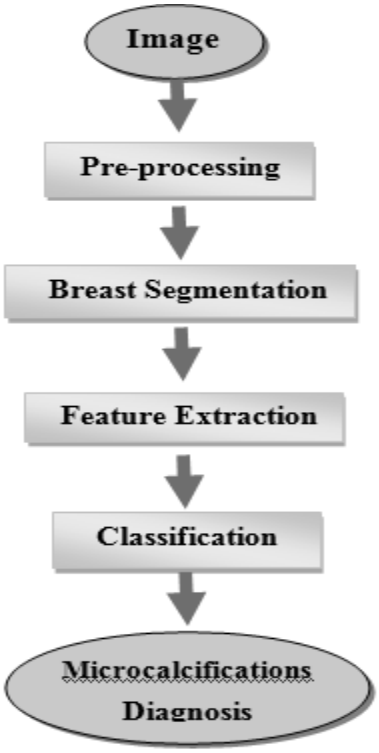

Figure (2): Flowchart of the steps of Microcalcifications Diagnosis method.

4.1.Preprocessing

This step enhances the breast's edge before segmenting the region of interest (ROI) from the mammographic images; this process can be performed by using the Logarithm transformations tool for dynamic range manipulation. Logarithm transformations are implemented using the expression:

$$g = c^* \log (1+ \text{double } (f))$$

where g is output image, f is input image, and c is a constant.

4.2. Segmentation

The aim of the segmenting of the image is to reduce the data amount consequently, reduce the computational time and the process searching for Regions of Interest that include a lesion with high probability**.** This step eliminates the parts in the image that are not useful for the mammographic interpretation and extract only the region of the image that corresponds to the breast. We performed this step by

binarization the image with automatic threshold for obtaining a binary image and Morphological operations for the extract the region of breast.

4.3. Features Extraction

In an automated analysis system an important goal is to extract features that are able to "summarize" meaningful information in the mammographic image; information that is otherwise distributed among a large number of pixels. Features here are extracted from the ROI based on the wavelet decomposition process. These features are passed to the classification stage. There are three processing steps in the features extraction stage. Features, in our method, are extracted from the coefficients that were produced by the wavelet analysis decomposition. The following steps discuss in details the extracted features.

i. Wavelet Decomposition

In this work, the wavelet decomposition applied on the region of interest. The outputs of wavelet analysis are the decomposition vector C and corresponding book keeping matrix S. The vector C consists of horizontal, vertical, and diagonal detail coefficients and one approximation.

ii. Coefficients Extraction

The horizontal, vertical, and diagonal detail and approximation was extracted from the wavelet decomposition structure [C, S]. These vectors were extracted at level one of wavelet decomposition.

iii. Energy, Mean, and Standard Deviation Computation

The energy, mean, and standard deviation have been computed for extracted coefficients. The produced values are considered as features (12 features) for the classification process.

4.4. Classification

The classification process is divided into the training phase and the testing phase. In the training phase, labeled data are given. Separately, the data on a candidate region which has already been decided as microcalcifications or as normal is given and the classifier is trained. In the testing phase, unknown data are given and the classification is performed using the classifier after training. The number of images which were used in training and testing sets is shown in Table(1).This step can be divided into two processes:

Table (1): Number of training and testing images

Category	No. of image	No. of training sets	No. of testing sets
Normal	40	20	20
Abnormal	25	17	8
Benign	12	8	4
Malignant	13	9	4

1. **Features Reduction**

After computing the features for every image in training set we reduce the number of features by estimated the mean for each wavelet coefficient at each feature vector. Table (2) and Table (3) are examples of training database that used in classification.

Table (2): Energy Features for Malignant Class

Approximation	Horizontal	Vertical	Diagonal
99.9327	0.0094	0.0558	0.002
99.9728	0.0128	0.013	0.0014
99.9709	0.0099	0.0178	0.0014
99.9879	0.0032	0.008	0.00098003
99.942	0.0138	0.0417	0.0025
99.9675	0.0132	0.0174	0.0019
99.9141	0.0091	0.0746	0.0023
99.965	0.004	0.0302	0.00082275
99.9782	0.0076	0.013	0.0011

Table (3): Energy Feature for Benign Class

Approximation	**Horizontal**	**Vertical**	**Diagonal**
99.9402	0.0075	0.0509	0.0014
99.9006	0.01	0.0878	0.0016
99.9674	0.0069	0.0241	0.0016
99.9398	0.0123	0.0462	0.0016
99.9299	0.0064	0.0624	0.0013
99.9649	0.0103	0.0231	0.0018
99.9305	0.007	0.0606	0.0019
99.9757	0.0076	0.0151	0.0016

2. Classify

The Nearest Neighbor (NN) classifier has been used to classify classes normal, malignant, and benign. The classifier used the Euclidean distance as a metric between the correspondents normalized wavelet coefficients, as shown in Equation (3).

$$D_{Euclidian} = \sqrt{\sum_{i=1}^{n}[(x[i]-x_j[i])]^2} \qquad (3)$$

where **x** is the feature vector of unknown pattern, $\mathbf{x}_j$ is the mean vector of class j, and n is the dimensionality of the feature space. Then the Nearest Neighbor states that the vector x is to be assigned to the class which has the minimum distance.

5. CADx EVALUATION

This section presents the results achieved in microcalcifications diagnosis method. We obtained results of classification among normal, benign and malignant classes using Nearest Neighbor (NN) classifier.

In the beginning, the method of classification is tested twice: once, between normal and microcalcifications, and once between benign and malignant. The results were good in classifying the malignant and benign, whereas it was bad in classifying the normal images. Also, this method tested with one, two, and three levels of the family wavelet (db4) decomposition and was best results with level one.

The classification results of method depending on Table (1) and using the family wavelet (db4) with level one of decomposition were 100% sensitivity (Equation 4), 100% specificity (Equation 5), 100% accuracy (Equation 6), and 100% precision (Equation 7). Table (4) shows the successful rates of classification for normal, benign and malignant classes.

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (4)$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (5)$$

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true negatives} + \text{true positives} + \text{false positives} + \text{false negatives}} \quad (6)$$

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (7)$$

Table (4): Successful rates of classification

Class	No. of testing sets	No. of successful sets	Successful rates
Normal	20	20	100%
Benign	4	4	100%
Malignant	4	4	100%

6. CONCLUSIONS

In this work, we proposed method for microcalcifications diagnosis in mammograms of breast. After implementing this method, we concluded the following:

1. Despite the abundance of theoretical studies in this field, computerized microcalcifications detection and diagnosis is one of the things that need a lot of study and practice. This belongs to many reasons. Those come from the great variability in the database mammograms, the use of poor resolution microcalcifications mammograms, small number of the available database, some mammograms are still film-based and are read using a light box thus commercial (CADx) systems digitize the film and present markers on a small display or on a separate printout, for the purpose of detection and classification the image is just

decomposed to generate features for a classifier and the task of enhancement is more complex as it also requires an image reconstruction.

2. Microcalcifications are subtle signs of breast cancer and are very difficult to detect and diagnose in the mammographic images because of their small size, low contrast with respect to the normal breast tissue and proximity to the surrounding tissues.
3. Because of the importance of the information in medical images, our goal was to keep this information. Perhaps the uses of wavelet transform characteristics have had a significant impact in achieving this goal.

7. REFERENCES

[1] American Cancer Society (2011). Breast Cancer. Atlanta, Ga: American Cancer Society.

[2] Zimmerman, B. T. (2004). Understanding Breast Cancer Genetics. United States of America: University Press of Mississippi.

[3] Denarie, B. E. (2010).Using SURF imaging for efficient detection of micro-calcifications. M.Sc. thesis, Department of Engineering Cybernetics, Mathematics and Electrical Engineering, Faculty of Information Technology, Norwegian University of Science and Technology, Norway.

[4] Nesbitt, D. (1995). Automated Detection of Microcalcifications in Digitized Mammogram Film Images. M.Sc. thesis, Electrical Engineering, The Faculty of Graduate studies**,** The University of British Columbia.

[5] Sankar, D., and Thomas, T. (2010). A New Fast Fractal Modeling Approach for the Detection of Microcalcifications in Mammograms. J Digit Imaging. 2010 Oct; 23(5):538-46. Epub 2009 Jul 18., 23, No.5, pp. 538-546.

[6] Uchiyama, N., and Nascimento, M. (2012).Mammography- Recent Advances. InTech.

[7] Li, H., Liu, R., and Lo, S. (1997). Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms. IEEE Trans on Medical Imaging, Vol.16, No.6, pp. 785-798.

[8] Chan, H. P., K., Doi, S., Galhotra, Vyborny, C. J., MacMahon, H., and Jokich, P. M. (1987). Image feature analysis and computer-aided diagnosis in digital radiography, 1. Automatic detection of microcalcifications in mammography. Med. Phys., Vol. 14, No. 4, pp. 538-548.

[9] Chan, H. P., Doi, K., Vyborny, C. J., Lam, K. L., and Schmidt, R. A.(1988).Computer-aided detection of microcalcifications in mammograms methodology and preliminary clinical study. Investigative Radiol., Vol. 23, pp.664-671.

[10] Chan, H. P., Doi, K., Vyborny, C. J., Schmidt, R. A., Metz, C., Lam, K. L., Ogura, T., Wu, Y., and Maxmahon, H.(1990). Improvement in radiologists' detection of clustered microcalcifications on mammogram: The potential of computer-aided diagnosis," Investigative Radiol., Vol. 25, pp.1102-1110.

[11] Karssemeijer, N. (1991).A stochastic model for automated detection of calcifications in digital mammograms. in Proc. 12th Int. Conf. Information Processing Medical Imaging, Wye, U.K., pp. 227-238.

[12] Karssemeijer, N. (1993).Recognition of clustered microcalcifications using a random field model, biomedical image processing and biomedical visualization. In SPIE Proc., vol. 1905, San Jose, CA, pp. 776-786.
[13] Karssemeijer, N. (1993).Adaptive noise equalization and recognition of microcalcification clusters in mammograms. Int. J. Pattern Recognit. Artificial Intell., Vol. 7, No. 6, pp. 1357-1376.
[14] Strickland, R. N. and Hahn, H. I. (1994).Wavelet transforms for detecting microcalcifications in mammography. In Proc. Int. Conf. Image Processing, Austin, TX, pp. 402-406.
[15] Strickland, R. N. and Hahn, H. I. (1994). Detection of microcalcifications using wavelets. In Digital Mammography'94, Proc. 2nd Int. Workshop Digital Mammography, York, U.K., pp. 79-88.
[16] Strickland, R. N. and Hahn, H. I. (1995). Wavelet transform matched filters for the detection and classification of microcalcifications in mammography. In Proc. IEEE Int. Conf. Image Processing, Washington, DC, pp. 422-425.
[17] Strickland, R. N., Hahn, H. I., and Baig, L. J.(1996).Wavelet methods for combining CAD with enhancement of mammograms. In Medical Imaging 1996: Image Processing, SPIE Proc., Vol. 2710, pp. 888-903.
[18] Strickland, R. N. and Hahn, H. I. (1996).Wavelet transform for detecting microcalcifications in mammograms. IEEE Trans. Med. Imag., vol. 15, pp. 218–229.
[19] Strickland, R. N. and Hahn, H. I. (1997). Wavelet transform methods for objects detection and recovery. IEEE Trans. Image Processing, Vol. 6, pp. 724-735.
[20] Yoshida, H., Doi, K., and Nishikawa, R. M. (1994). Automated detection of clustered microcalcifications. In Digital Mammograms Using Wavelet Transform Techniques, Medical Imaging 1994: Image Processing, Proc. SPIE, Vol. 2167, Newport Beach, CA, pp. 868-886.
[21] Yoshida, H., Zhang, W., Cai, W., Doi, K., Nishikawa, R. M., and Giger, M. L. (1995).Optimizing wavelet transform based on supervised learning for detection of microcalcifications in digital mammograms. In Proc. IEEE Int. Conf. Image Processing, Vol. 3, Washington, DC, pp. 152-155.
[22] Cheng, H., Lui, Y. M., and Feiimanis, R. I. (1998). A novel approach to microcalcification detection using fuzzy logic techniques. IEEE Trans. Med. Imag., Vol. 17, pp. 442-450.
[23] Yu, S., and Guan, L. (2000). A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films. IEEE transactions on medical imaging, Vol. 19, No. 2, pp.115-126.
[24] Jiang, J., Yao, B.,and Wason, A.M.(2007). A genetic algorithm design for microcalcification detection and classification in digital mammograms. Computerized medical imaging and graph ics, Vol. 31, No.1, pp. 49-61.
[25] Laine, A., Schuler, S., Fan, J. and Huda, W. (1994). Mammographic feature enhancement by multiscale analysis. IEEE Transactions on Medical Imaging, Vol.13, No.4, pp.725-740.
[26] Mallat, S. G. (1989). Multifrequency channel decomposition of images and wavelet models. IEEE Trans. Acoust., Speech, Signal Processing, Vol. 37, pp. 2091-2110..