

Using Hurst Coefficient and Lacunarity to diagnosis early breast diseases

Rodrigo Carvalho Serrano

Jesuliana Ulysses

Sildenir Ribeiro

Aura Conci

Computer Institute

Federal Fluminense University - UFF

Niterói, Brazil

rcarvalho@ic.uff.br

Rita Lima

Department of Mechanical Engineering

Federal University of Pernambuco - UFPE

Recife, Brazil

Abstract— This paper presents a study about diagnosis of breast diseases in early stages using Thermal Images, Hurst Coefficient and Lacunarity. It is based on analysis of symmetry of temperatures in breasts. The first step is extraction of region on interest: right and left breast. With these two images, a new image (image Subtraction) is generated. The second step is extraction of features. For this, Hurst Coefficient and Lacunarity were used. Using Hurst Coefficient were extracted 36 features. Others 97 features were obtained by Lacunarity. Therefore, the proposed method characterizes thermal images by 133 features. The last step is classification. With this objective, techniques of learning machine were used to classify the patients in two groups: healthy or sick. The proposed method presented excellent results (ROC Area = 0,958).

Keywords-component; Thermal images, Medical Images, Image processing, Fractal Geometry, Hurst Coefficient, Lacunarity.

I. INTRODUCTION

Thermography is used in order to obtain information on the operational condition of a component, equipment or process. This is a technique that uses infrared radiation that is emitted by bodies in order to measure temperatures or even notice some differences in the pattern of distribution of heat. It does not emit ionizing radiation and is a noninvasive technique. Its application in medicine was initiated after it was discovered that in the region where there is a breast cancer, skin temperature is higher than the temperature of the normal tissue [1]. Furthermore, it was also discovered that poison blood responsible for draining the cancer has temperature higher than the arterial blood. The temperature distribution in healthy breasts occurs in a symmetrical manner and the temperature changes are related to a possible pregnancy and the menstrual cycle, [2]. This study showed that these changes somehow affect both breasts. If the symmetry of the temperature of the breasts no longer exists, then a breast has some type of disease. Patients with breast tumors do not present symmetry, because the cancer cells produce excessive nitric oxide (NO), which carries a new vasculature (angiogenesis) around the tumor. This new vasculature is

responsible for increasing blood flow causing a rise in temperature in the region of the tumor. This variation in temperature between a normal region and one that has some tumor is 2 to 3°C [3]. Hence the importance of thermography, because it captures this increase in temperature on the skin surface. Thus, while mammography and ultrasound tests are anatomical, thermography is a physiological test. In addition, thermography can detect smaller tumors than mammography (the average size of tumor detected on a mammogram is 1.66 cm, while in a thermography is 1.28 cm), and diagnosing tumors around 8 to 10 years before a mammogram [4]. Due to the incidence of breast cancer in women younger, some doctors have recommended mammography from age 35. But as younger breasts are dense it is difficult to detect using mammography. Moreover, as thermal images acquire the temperature of the breast, then it can be used in young patients enables early diagnosis of breast cancer.

II. IMAGES

In this study were used 28 thermal images available at <http://200.20.11.171/proeng/>. In these images each pixel corresponds to a temperature of the acquired scene. These images belong to two groups of patients: healthy and sick. The group of sick patients has the following conditions: cyst, fibroadenoma (benign tumor), carcinoma (malignant tumor). There are many ways to represent a thermal image, for example: Rain, Rain10, Grey, Grey10, Iron, Medical, Midgreen, Midgrey. In this paper were used thermal images in shades of grey. Then, the range of temperatures is associated with the range of gray-scale. This is the simplest way to represent a thermal image. As example of figures used in this study, we present Figure 1 and Figure 2.

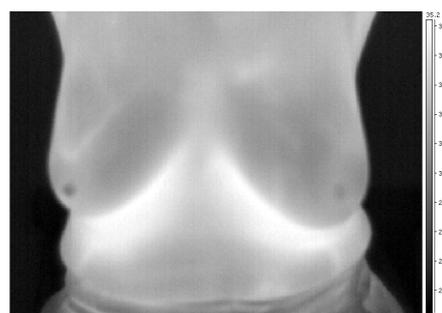


Figure 1. Thermal image of healthy patient

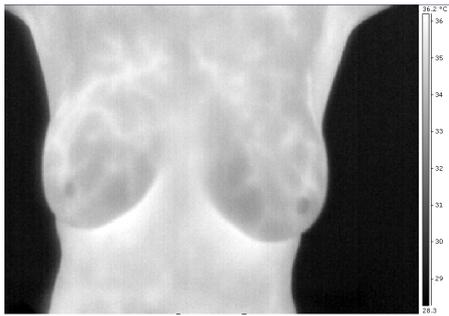


Figure 2. Thermal Image of sick patient

At first figure, each point of the image between 25.2 and 35.2 degrees Celsius will be associated with a level gray between 0 and 255. Similarly, in the second figure each point of the image between 28.3 and 36.2 degrees Celsius will be associated with a level gray between 0 and 255.

A. Preprocessing

The first step of this approach is preprocessing the thermal image. This is done to extract the region on interest (ROI). In this study, for each thermal image, each breast is selected using a square window of the same size [5]. Then, the preprocessing will generate two images: right and left breast. These images will be used in next steps of this approach. As example of preprocessing, Figure 3 and Figure 4 present the results of this step for Figure 1 and Figure 2, respectively.

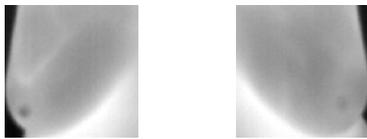


Figure 3. Region on interest (right and left breast) from Figure 1.

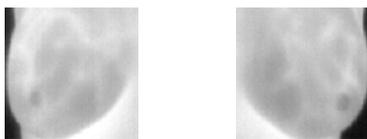


Figure 4. Region on interest (right and left breast) from Figure 2.

The image corresponding to the right breast is rotated 180 degrees to a vertical axis. Using this image (rotated) and the image corresponding to the left breast is generated a new image (image Subtraction) by subtraction pixel to pixel. With objective of illustrating, Figure 5 and Figure 6 present the results to Figure 3 and Figure 4, respectively.



Figure 5. Image Subtraction from images in Figure 3

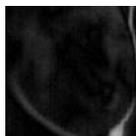


Figure 6. Image Subtraction from images in Figure 4

Therefore, the preprocessing will generate three images: right breast, left breast and image Subtraction.

III. PROPOSED METHOD

A. Feature extraction

Feature extraction aims to collect relevant data from the image [6], [7]. Fractal Geometry had been used to characterize textures. This is used to link each texture to a numerical measure and, with the data generated, making the classification of textures. In this study were used two fractal measures: Hurst Coefficient [8], [12] and Lacunarity [9], [10], [11], [12]. Hurst Coefficient is related to the density of the image or object, that is, how much the image or object occupies the space that contains it. There are various applications of Hurst Coefficient, including: image classification, wear and erosion, detection of bands expectral noisy, corrosion, analysis of the diversity of the landscape, analysis of fractured surfaces, determining the roughness of surface, determining the operational scale of natural phenomena in digital imaging, scaling applied to spatial extensions in remote sensing, the distinction between landscape types, analysis of the effects of converting data into geographic information systems. The lacunarity quantifies the amount of gaps or holes in the image or object. This measure has been used to characterize mammograms images. In this study, Hurst Coefficient (H) and Lacunarity are used to characterize thermal images. Hurst Coefficient will be used in two ways: the first is using the images of the right breast and left breast. For this, a movable (pixel by pixel for each column and row of the image) window computes the values of Hurst coefficient. With this, each image is characterized by various values of Hurst coefficient for a window of size w. Square windows of size w=5, 7, 9, 11, 13 and 15 were used. Then, the average and standard deviation of Hurst coefficient for each window w are used to form the first four characteristics used on our feature vector. Were used 6 different sizes of windows (w = 5,7,9,11,13,15), then this first way provides us 24 features. Figure 7 summarizes the previous steps.

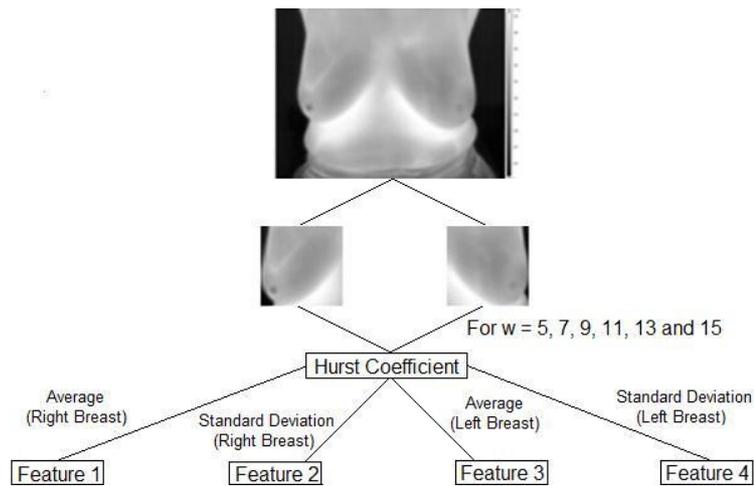


Figure 7. The steps used on this analysis

The second way is using the image Subtraction. Then, the average and standard deviation of Hurst coefficient for each window are used to form two others characteristics for our feature vector. With this, the second way will provide us with 12 features (2 features for each of 6 different windows). Figure 8 summarizes the steps used to calculate the two previous features.

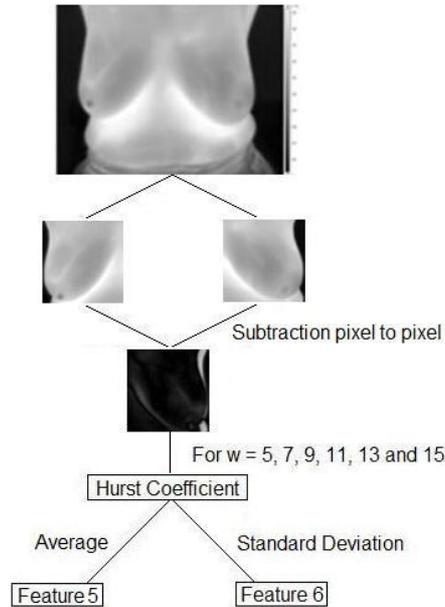


Figure 8. The steps used on this analysis

Therefore, Hurst coefficient characterized each thermal image by 36 features. Others 97 characteristics were obtained from the Lacunarity of the images pre-processed. In this algorithm, the third coordinate (gray scale) is used and the images can be seen as collection of voxels. Then, boxes of size s , $2 \leq s \leq 25$ go through the collection of voxels calculating the values of Lacunarity. So, each box of size s will be associated to a Lacunarity value $\Lambda(s)$. Therefore, 24 Lacunarity values will be generated for each image (image Subtraction, right and left breast). The notations $\Lambda_R(s)$ and $\Lambda_L(s)$ will be used to refer to Lacunarity values of right and left breast respectively. Calculating $|\Lambda_R(s) - \Lambda_L(s)|$ with s ranging $2 \leq s \leq 25$, other 24 features are generated. The last feature is extracted by calculating the standard deviation of $|\Lambda_R(s) - \Lambda_L(s)|$ with s ranging $2 \leq s \leq 25$. Then, this study proposes to use the 72 features obtained through of Lacunarity algorithm (24 features of each image generated in the preprocessing) others 24 features obtained through of $|\Lambda_R(s) - \Lambda_L(s)|$ with s ranging $2 \leq s \leq 25$ and another feature obtained by the standard deviation of $|\Lambda_R(s) - \Lambda_L(s)|$ with s ranging $2 \leq s \leq 25$.

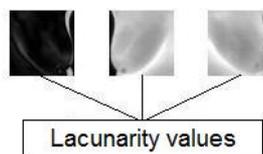


Figure 9. The steps used on this analysis

B. Classification

In order to find out what are the features that best describe the thermal images, the 133 features were divided into 14 groups [12]. The groups are:

Group 1: Composed of all 133 features;

Group 2: Composed of all the features extracted by the Hurst coefficient (36 features);

Group 3: Composed of the features $f = 1, 2, 3$ and 4 for $w = 5, 7, 9, 11, 13$ and 15 (24 features);

Group 4: Composed by features $f = 2$ and 4 for $w = 5, 7, 9, 11, 13$ and 15 (12 features);

Group 5: Composed by the features $f = 1$ and 3 for $w = 5, 7, 9, 11, 13$ and 15 (12 features);

Group 6: Composed by the features $f = 5$ and 6 for $w = 5, 7, 9, 11, 13$ and 15 (12 features);

Group 7: Composed by the feature $f = 5$ for $w = 5, 7, 9, 11, 13$ and 15 (6 features);

Group 8: Composed by the feature $f = 6$ for $w = 5, 7, 9, 11, 13$ and 15 (6 features);

Group 9: Composed of all the features extracted by Lacunarity (97 features);

Group 10: Composed by $\Lambda_R(s)$ and $\Lambda_L(s)$ with $2 \leq s \leq 25$ (48 features);

Group 11: Composed by $|\Lambda_R(s) - \Lambda_L(s)|$ with $2 \leq s \leq 25$ (24 features);

Group 12: Composed by the standard deviation of $|\Lambda_R(s) - \Lambda_L(s)|$ with $2 \leq s \leq 25$ (one feature);

Group 13: Composed by Lacunarity $\Lambda(s)$ of image Subtraction with $2 \leq s \leq 25$ (24 features);

Group 14: Composed by the features in groups 12 and 13 (25 features);

After this, the data generated were used to classify patients in healthy or sick. Then, techniques of learning machine were used.

C. Results

The results were evaluated using Receiver operating characteristic (ROC). The tables below present the best results.

TABLE I. RESULTS OF GROUPS 1 AND 2

Techniques	G1	G2
Naive Bayes Simple	0,792	0,854
Classification Via Clustering	0,771	0,375
Naive Bayes	0,708	0,875
Naive Bayes Updateable	0,708	0,875
Random Committee	0,516	0,865
Nnge	0,5	0,5

TABLE II. RESULTS OF GROUPS 3 AND 4

Techniques	G3	G4
Naive Bayes Simple	0,885	0,927
Classification Via Clustering	0,771	0,792
Naive Bayes	0,927	0,958
Naive Bayes Updateable	0,927	0,958
Random Committee	0,667	0,833

Nnge	0,479	0,458
------	-------	-------

TABLE III. RESULTS OF GROUPS 5 AND 6

Techniques	G5	G6
Naive Bayes Simple	0,667	0,484
Classification Via Clustering	0,646	0,5
Naive Bayes	0,667	0,49
Naive Bayes Updateable	0,667	0,49
Random Committee	0,734	0,411
Nnge	0,75	0,5

TABLE IV. RESULTS OF GROUPS 7 AND 8

Techniques	G7	G8
Naive Bayes Simple	0,625	0,458
Classification Via Clustering	0,708	0,563
Naive Bayes	0,594	0,438
Naive Bayes Updateable	0,594	0,438
Random Committee	0,391	0,745
Nnge	0,458	0,458

TABLE V. RESULTS OF GROUPS 9 AND 10

Techniques	G9	G10
Naive Bayes Simple	0,557	0,542
Classification Via Clustering	0,688	0,625
Naive Bayes	0,49	0,488
Naive Bayes Updateable	0,49	0,488
Random Committee	0,568	0,391
Nnge	0,438	0,5

TABLE VI. RESULTS OF GROUPS 11 AND 12

Techniques	G11	G12
Naive Bayes Simple	0,281	0,698
Classification Via Clustering	0,479	0,813
Naive Bayes	0,313	0,625
Naive Bayes Updateable	0,313	0,625
Random Committee	0,271	0,563
Nnge	0,458	0,563

TABLE VII. RESULTS OF GROUPS 13 AND 14

Techniques	G13	G14
Naive Bayes Simple	0,417	0,422
Classification Via Clustering	0,417	0,438
Naive Bayes	0,443	0,453
Naive Bayes Updateable	0,443	0,453

Random Committee	0,542	0,557
Nnge	0,5	0,667

The Groups 6, 9, 10, 11, 13 and 14 did not show good results, however, as we can see in previous tables, the proposed method presented excellent results (ROC Area = 0,958 using the features of group 4) to others groups. The best results was obtained for group 1 using Naive Bayes Simple (0,792); for group 2 using Naive Bayes and Naive Bayes Updateable (0,875); for group 3 using Naive Bayes and Naive Bayes Updateable (0,927); for group 4 using Naive Bayes and Naive Bayes Updateable (0,958); for group 5 using Nnge (0,75); for group 7 using Classification Via Clustering (0,708); for group 8 using Random Committee (0,745); for group 12 using Classification Via Clustering (0,813). In future works will be used these features presented in this study and others fractal measures to diagnosis breast diseases in early stages.

REFERENCES

- [1] Bronzino, J.D., "Infrared imaging of the breast – an overview", The Biomedical Engineering handbook – Medical Systems and Devices, 3° ed., CRC, 2006.
- [2] M. Gautherie. Atlas of breast thermography with specific guidelines for examination and interpretation (Milan, Italy:PAPUSA), 1989.
- [3] Hu, L., Gupta, A., Gore, J.P., Xu, L. X., Effect of Forced Convection on the Skin Thermal Expression of Breast Cancer, Journal of Biomechanical Engineering, vol. 126, pp. 205-211, 2004.
- [4] Keyserlingk, J. R.; Ahlgren, P. D.; Yu, E.; Belliveau, N.; Yassa, M., Functional infrared imaging of the breast. IEEE Engineering in Medicine and Biology, pages 30 to 41, May/June 2000.
- [5] Serrano, R.C., Motta, L., Batista, M., Conci, A., Using a new method in thermal images to diagnose early breast diseases, XXIIInd Brazilian Symposium on Computer Graphics and Image Processing - SIBGRAPI, Rio de Janeiro, Brazil, October 2009.
- [6] Sonka, M., Hlavac, V., Boyle, R., Image Processing Analysis and Machine Vision, 3rd Ed. Thomson, 2008.
- [7] Conci, A., Azevedo, E., Leta, F.R., Computação Gráfica: Teoria e prática, vol. 2 Editora Campus (Elsevier), Rio de Janeiro, Brazil, 2008.
- [8] Serrano, R. C., Conci, A., Zamith, M., Lima, R.C.F., About the feasibility of Hurst coefficient in thermal images for early diagnosis of breast diseases, 11th Pan-American Congress of Applied Mechanics - PACAM XI, Paraná, Brazil, January 2010.
- [9] Melo, R.H.C., Vieira, E.A., Conci, A., Comparing two approaches to compute lacunarity of mammograms, 13th International Conference on Systems, Signals and Image – IWSSIP, Budapest, Hungary, September 2006.
- [10] Melo, R.H.C., Vieira, E. A., Conci, A., Characterizing the lacunarity of objects and image sets and its use as a technique for the analysis of textural patterns, Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science, Publisher Springer Berlin/ Heidelberg, ACIVS - IEEE Benelux Signal Processing Chapter, v. 4179, pp. 208 – 219, Antwerp, Belgium, September 2006.
- [11] Allain, C., Cloitre, M., Characterizing the lacunarity of random and deterministic fractal sets, Physical review A., 44, pp.3552-3558, 1991.
- [12] Serrano, R.C., Análise da viabilidade do uso do Coeficiente de Hurst e da Lacunaridade no auxílio ao diagnóstico precoce de patologias da mama, Dissertação de Mestrado, Instituto de Computação, Universidade Federal Fluminense, 2010.