

# About the feasibility of Hurst coefficient in thermal images for early diagnosis of breast diseases

**Rodrigo Carvalho Serrano, rcarvalho@ic.uff.br**

**Aura Conci, aconci@ic.uff.br**

**Marcelo Zamith, mzamith@ic.uff.br**

Universidade Federal Fluminense, Instituto de Computação

Rua Passo da Pátria 156 - Bloco E - 3º andar, CEP: 24210-240, Niterói - RJ – Brasil

**Rita C. F. Lima, ritalima@ufpe.br**

Universidade Federal de Pernambuco, Departamento de Engenharia Mecânica

Rua Acadêmico Hélio Ramos S/Nº, CEP 50740-530, Recife - PE – Brasil

***Abstract.** This paper presents a study for diagnosis of breast diseases in early stages (cysts, fibroadenomas, tumors, etc.). This work is based on analysis of asymmetry between the patient's breasts using thermal images and considering that above certain values this is an indication of some abnormalities. Initially, for each thermal image, each patient's breast is segmented using a square window of the same size. The Hurst coefficient is used to characterize the breast's texture. For each computation, a movable (pixel by pixel for each row and column of the image) window computes the values that will result in the Hurst coefficient. Square windows of sides  $w$  equal to 5, 7, 9, 11, 13 and 15 were used. Consequently, each image of  $N \times N$  pixels has the values of the Hurst coefficient calculated  $\sum (N - (2i + 3))^2$  times with  $i$  ranging from 1 to 6. Then, the average and standard deviation of Hurst coefficient for each window are used to form a feature vector. This is done in two ways. First, using the image of each breast. Second, using a new image obtained by the difference of temperature at vertically symmetric positions of each breast. This methodology is applied considering groups of patients who have confirmed diseases and healthy patients. Then, the possibility of obtaining a correct breast diagnose through the thermal images using combinations of this features are evaluated using machine learning techniques. The rate of correct diagnosis achieve 95% when Classification Via Regression is used.*

**Keywords:** Processing of medical images, Thermography, Hurst coefficient.

## 1. INTRODUCTION

The diagnosis of mastology diseases through the use of thermal imaging is a noninvasive method that makes use of infrared radiation from the body and is encouraged due to the fact of not using ionizing radiation, breast compression or intravenous contrast injection as the other types of exams. However, it is a physiological test, while other forms of diagnosis such as mammography and ultrasound, are anatomical tests (Love, 1985).

It is an experimental fact that a modification in breast temperatures is associated with tissue modification. Malignant tumors always change the temperature of the breast (Yahara et al., 2003). Breast tumors, in general, need a constant flow of nutrients to be developed, for this its cells produce substances responsible for angiogenesis (creation of new vessels) around the tumor. These new vessels increase the blood flow in the region, thereby causing an increase in temperature at the site. In breast tumors, this elevation of local temperature can be observed on the surface of the breast through the capture of thermal images. The diagnosis by thermography therefore is closely linked to the infusion of fluid (lymph and blood) of patients, reflecting the dynamics of the circulatory fluid of the skin surface.

Another advantage of the thermal mammography is the possibility of early detection of any problem in the breast (Ng and Sudharsan, 2001).

## 2. IMAGES

In thermal images each pixel corresponds to a temperature of the acquired scene (Menezes et al., 2009). The simplest form of thermal imaging is represented such temperature in shades of gray. In this case the range of these temperatures may be associated with the image gray-level. Thermal images in gray scale are presented as shown in Fig. 1, where each point of the image between 26.0 and 34.5 degrees Celsius will be associated with a level gray between 0 and 255.

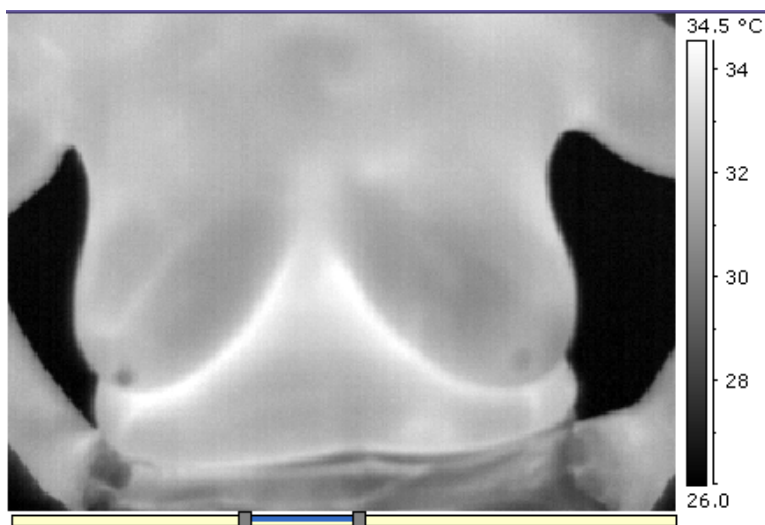


Figure 1. Sample of an original thermography of the used database

Initially, for each thermography, our approach selects each patient's breast using a square window of the same size as in Fig. 2. The window's size depends on the size of patient's breast (in the images used in this study, we had windows between 90x90 and 110x110 pixels). Therefore, for each patient, segmented images like those on Fig. 2 (Serrano et al. 2009) will be used in next steps of the proposed method.

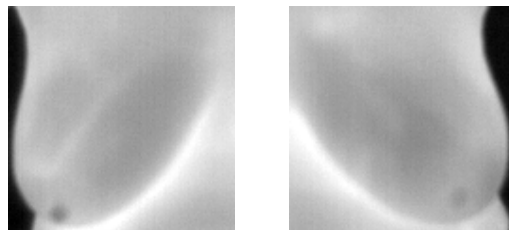


Figure 2. Segmented right and left breast from figure 1

In this paper, nine thermal images were used (five images of patients with some kind of breast problem and four images of healthy patients). Such images were acquired as part of the PROENG (<http://200.20.11.171/proeng/>) project for construct of a thermal image data base (Castro et al. 2009).

### 3. PROPOSED APPROACH

#### 3.1. Features Extraction

Fractal geometry has been used for the analysis of complex natural patterns like landscapes, remote sensing or texture of blood vessels (Barros and Sobreira, 2008; Valous et al., 2009). Three fractal parameters have been considered: fractal dimension, lacunarity and succolarity (Mandelbrot, 1983). The fractal dimension quantifies the density of fractals (or any images) on its metric space. The fractal dimensions are then an objective way to compare one fractal (or figure) to another. Hurst coefficient (H), method used in this paper (<http://200.20.11.171/proeng/>), is an approximation of the fractal dimension for images in gray levels. In our approach H was extracted from the right and left breast's images of each patient as mentioned in previous section. For each computation, a movable (pixel by pixel for each row and column of the image) window computes the values that will result in the Hurst coefficient. Square windows of size  $w=5, 7, 9, 11, 13$  and  $15$  were used. So each image of  $N \times N$  pixel has the values of the Hurst coefficient calculated  $\sum (N - (2i + 3))^2$  times with  $i$  ranging from 1 to 6. Then, the average and

standard deviation of Hurst coefficient for each window are used to form the first four characteristics used on our feature vector. The diagram in Fig. 3 summarizes the previous steps.

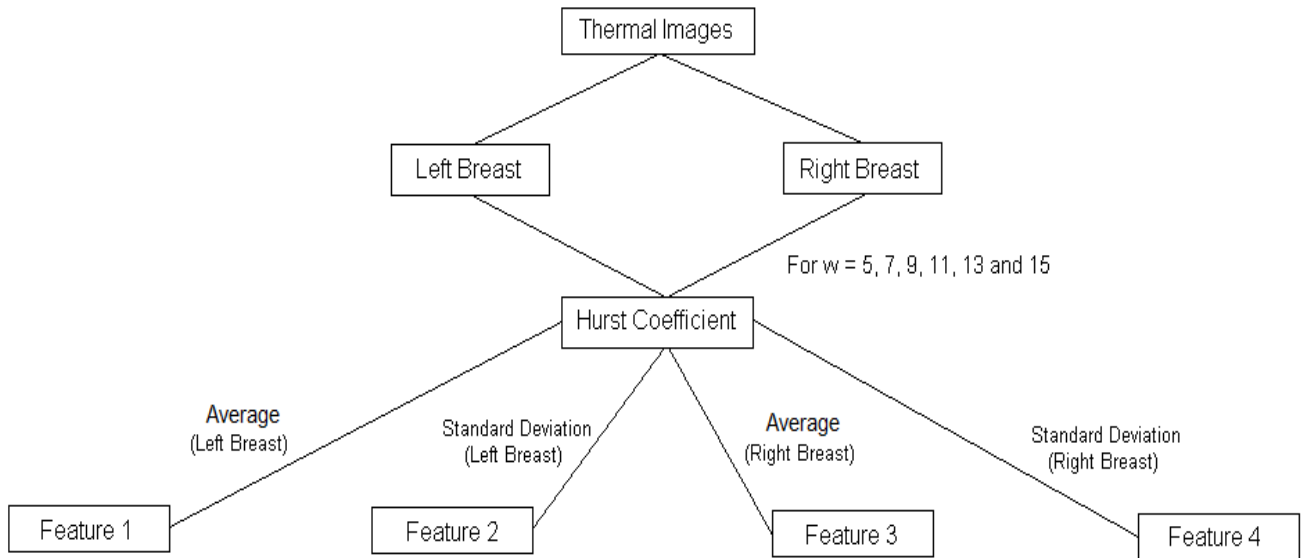


Figure 3. The steps used on this analysis

The average and standard deviation of all possible  $w$  is also computed for a new image obtained by the difference of temperature at positions vertically symmetric of each breast. The diagram below summarizes the steps used to calculate features 5 and 6.

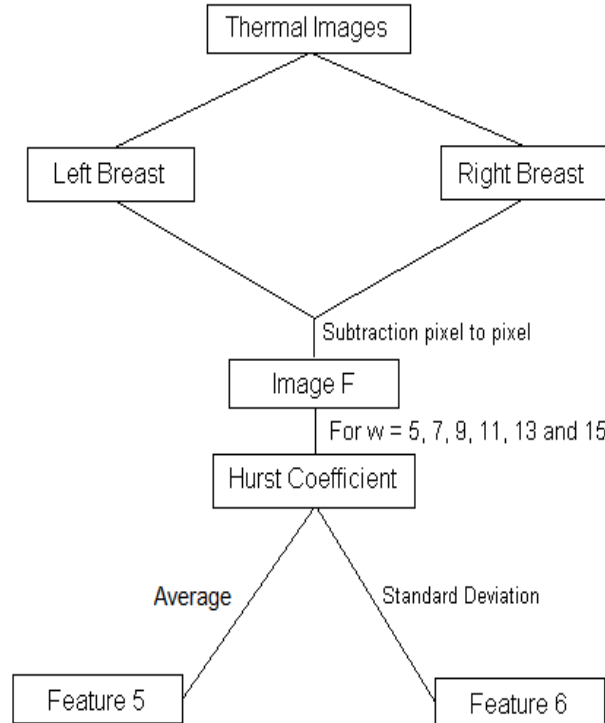


Figure 4. The steps used on this analysis

Therefore, for each patient image were extracted the features  $f$ , with  $f$  ranging from 1 to 6.

### 3.2. Classification

In this step, we use the features extracted in 3.1 to identify if the patient has a breast disease. With this objective, we used the software WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) and techniques of machine learning, among them: Bayes Logistic Regression, Bayes Net, Complement Naive Bayes, DMNB Text, Naive Bayes, Naive Bayes Multinomial, Naive Bayes Multinomial Updateable, Naive Bayes Simple, Naive Bayes Updateable, Lib Linear, Lib SVM, Logistic, Multilayer Perceptron, RBF Network, Simple Logistic, SMO, Voted Perceptron, IB1, IBK, K Star, LWL, Ada Boost M1, Attribute Selected Classifier, Bagging, Classification Via Clustering, Classification Via Regression, Cost Sensitive Classifier, CV Parameter Selection, Dagging, Decorate, END, Ensemble Selection, Filtered Classifier, Grading, Logit Boost, Meta Cost, Multi Boost AB, Multi Class Classifier, Multi Scheme, Ordinal Class Classifier, Raced Incremental Logit Boost, Random Committee, Random Sub Space, Rotation Forest, Stacking, Stacking C, Threshold Selection Vote, Citation KNN, MI Boost, MISMO, MIW rapper, Simple MI, FLR, Hyper Pipes, VFI, ADTree, BFTree, Decision Stump, FT, J48, J48graft, LADTree, LMT, NBTree, Random Forest, Random, Tree, Rep Tree, Simple Cart, Conjunctive Rule, Decision Table, DTNB, JRip, NNge, OneR, PART, Ridor, ZeroR (Witten and Eibe, 2005). The rules had better results were: Naive Bayes, IB1, Classification Via Regression (CVR), PART.

### 3.3. Results

The performance of the method was evaluated using cross validation and receiver operating characteristic (ROC) analysis. In cross-validation performance is evaluated as follows: Before the initialization of the training step some data are removed. After the ending of the training, the data that were removed can be used to test the performance of the learned model on the "new" data. ROC analysis is a graphical plot of the (1 - specificity) vs. sensitivity. In this case, the specificity measures the percentage of healthy people who are identified as not having breast disease; and the sensitivity measures the percentage of people who are identified as having breast problem. To facilitate the understanding of the results, the features were divided into several groups showed in Figure 5. Table 1 summarizes the results.

Group1: Composed of all the features.

Group2: Composed of the features  $f = 1, 2, 3$  and  $4$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

Group3: Composed of the features  $f = 2$  and  $4$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

Group4: Composed of the features  $f = 1$  and  $3$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

Group5: Composed of the features  $f = 5$  and  $6$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

Group6: Composed of the feature  $f = 6$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

Group7: Composed of the feature  $f = 5$  for  $w = 5, 7, 9, 11, 13$  and  $15$ .

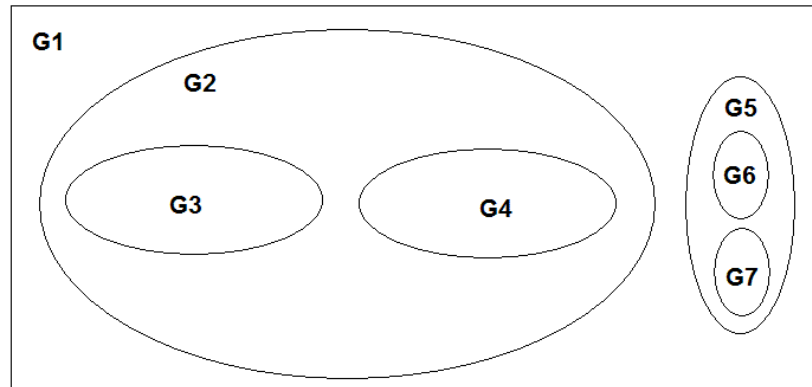


Figure 5. Groups of features

Table 1. Area (A) under receiver operating characteristic (ROC)

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Naive Bayes	0,875	0,8	0,75	0,6	0,825	0,875	0,825
IB1	0,875	0,7	0,675	0,675	0,875	0,875	0,775
CVR	0,775	0,775	0,775	0,45	0,75	0,95	0,75
PART	0,675	0,675	0,675	0,45	0,875	0,7	0,875

Best results are obtained for group 1 using Naive Bayes and IB1; for group 5 using IB1 and Part; for group 6 using Naive Bayes, IB1 and CVR; for group 7 using Naive Bayes and Part.

#### 4. CONCLUSIONS

The proposed approach can classify correctly up to 95% of the cases using CVR classifiers and only the standard deviation of a subtracted image of the patient breasts using the Hurst coefficient computed of different window sizes. Next works will apply this method to a larger image data base of thermal images (Castro et al. 2009) and new techniques using other fractal geometry features (Silveira et al. 2009).

#### 5. ACKNOWLEDGEMENTS

The authors thanks to sponsors of CNPq and CAPES.

#### 6. REFERENCES

- Barros Filho, M. N. M., Sobreira, F. J. A., "Accuracy of lacunarity algorithms in texture classification of high spatial resolution images from urban areas". In: XXI Congress of International Society of Photogrammetry and Remote Sensing, 2008, Beijing, China.
- Castro, F.J., Vasconcelos S., Serrano, R.C., Motta, L.S., Menezes, P.M., Bezerra, L.A., Lima, R.C.F., Conci, A., Um sistema para pré processamento de imagens térmicas e modelagem tridimensional aplicadas à Mastologia. - trab 104- II Encontro Nacional de Engenharia Biomecânica: ENEBI 6 a 8 de maio 2009 - Florianópolis- SC.
- Love, T.J., Analysis and application of thermography in medical diagnosis, in Shitzer, A.; Eberhart, R. C.: Heat transfer in medicine and biology, New York: Plenum Publishing Corporation, 1985.
- Mandelbrot, B., The fractal geometry of nature, Freeman, New York, 1983.
- Menezes, P.M., Serrano, R.C., Zamith, M., Knechtel, M., Montenegro, A., Clua, E.W.G, Conci, A., Bezerra, L.A., Lima, R.C.F., Reconstruindo as imagens termográficas a partir dos arquivos JPEG em false color para auxílio no diagnóstico mastológico - trab 83- II Encontro Nacional de Engenharia Biomecânica: ENEBI 6 a 8 de maio 2009 - Florianópolis- SC.
- Ng, E.Y.K., Sudharsan, N.M., Effect of blood flow, tumor and cold stress in a female breast: A novel time-accurate computer simulation, Proceedings of the Institution of Mechanical Engineering in Medicine, 215 (4), pp.392-404, 2001.
- Serrano, R.C., Motta, L.S., Batista, M., Conci, A., Using a new method in thermal images to diagnoses early breast diseases - XXII Brazilian Symposium on Computer Graphics and Image Processing: Sibgrapi 2009 – Rio de Janeiro – RJ.
- Silveira, O.T.F., Conci, A., Serrano, R.C., Mello, R., Lima, R.C.F., Paper ID: 549374 On Using Lacunarity for Diagnosis of Breast Diseases Considering Thermal Images, presented in Session: Image processing for medical applications, 4pp. IWSSIP 2009: 16th International Workshop on Systems, Signals and Image Processing for the year 2009, organised by the Technological Educational Institute of Chalkida, 18-20 June.
- Valous, N. A., Mendoza, F., Sun, Da-Wen, Allen, P., "Texture appearance characterization of pre-sliced pork ham images using fractal metrics: Fourier analysis dimension and lacunarity", Food Research International, vol. 42, Issue 3, 2009, pp. 353-362, ISSN 0963-9969.
- Witten, I.H., Eibe, F., 2005. "Data mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
- Yahara, T., Koga, T., Yoshida, S., Nakagawa, S., Deguchi, H., Relationship between micro vessel density and thermo graphic hot areas in breast cancer, Surgery Today, 33 (4), pp. 243-248, 2003.

#### 7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.