MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-746

Thermal Signal Analysis for Breast Cancer Risk Verification

Lincoln F. Silva[a], Giomar O. Sequeiros[a], Maria Lúcia O. Santos[b], Cristina A. P. Fontes[b], Débora C. Muchaluat-Saade[a] and Aura Conci[a]

[a] *Institute of Computing, Fluminense Federal University, Niterói, RJ, Brazil*
[b] *Department of Radiology, Fluminense Federal University, Niterói, RJ, Brazil*

Abstract

Breast cancer is the second most common cancer in the world. Currently, there are no effective methods to prevent this disease. However, early diagnosis increases chances of remission. Breast thermography is an option to be considered in screening strategies. This paper proposes a new dynamic breast thermography analysis technique in order to identify patients at risk for breast cancer. Thermal signals from patients of the Antonio Pedro University Hospital (HUAP), available at the Mastology Database for Research with Infrared Image - DMR-IR were used to validate the study. First, each patient´s images are registered. Then, the breast region is divided into subregions of 3x3 pixels and the average temperature from each of these regions is observed in all images of the same patient. Features of the thermal signals of such subregions are calculated. Then, the k-means algorithm is applied over feature vectors building two clusters. Silhouette index, Davies-Bouldin index and Calinski-Harabasz index are applied to evaluate the clustering. The test results showed that the methodology presented in this paper is able to identify patients with breast cancer. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved.

Keywords:

Breast cancer; Thermal signals; Clustering; Validity indexes; Classification.

Introduction

Breast cancer is the most common cancer among women worldwide. However, when diagnosed and treated in early stages, this cancer type has relatively good prognosis [1].

Screening is a strategy adopted by health authorities in order to identify women who are at initial stages of breast disease. Thus, it is necessary to develop methods and techniques in order to improve screening procedures because even mammography, which is considered the gold standard for cancer detection, has its limitations, such as high false positive classification rate, insufficient effectiveness in dense breasts and use of ionizing radiation to form breast images [2][3].

Since the cancerous tissue temperature is generally higher than healthy surrounding tissues, thermography has been considered a promising screening method for breast cancer detection by generating images that reveal the heat distribution on the breast surface [3].

The thermal signals used in this work originated from dynamic thermography, which is a method for monitoring the dynamic response of the skin temperature after thermal stimulus. In other methods for detecting breast cancer, the thermal stimulus most utilized is the application of air flow directed to the breasts by an electric fan [4][5][6]. The cooling of the breasts, theoretically, improves the thermal contrast between healthy and unhealthy tissues in the image, because blood vessels promoted by cancerous tumors do not have muscular layer and neural regulation like embryological vessels. These vessels are only endothelial tubes and therefore do not contract in response to sympathetic stimulation. For that reason breast region with cancerous tumors remain with unchanged temperature while the healthy part of the breast is cooled down [7].

When compared to static thermography, dynamic thermography is faster and more robust. Indeed, static thermography requires rigid environmental conditions and significantly long time for acclimatization of the patient to examining room conditions. On the other hand, dynamic thermography is much less dependent of the conditions and temperature of the examining room [8].

This paper proposes a new dynamic breast thermography analysis technique in order to identify patients at risk for breast cancer. Dynamic thermal signals from patients of the HUAP, available at the Mastology Database for Research with Infrared Image - DMR-IR [9] were used for validating this hypothesis. First, each patient has her images registered. Second, the breast region is divided into regions of 3x3 pixels and average temperature from each of these regions is observed in all images of the same patient. Features of the thermal signals of such regions are calculated. Then, the *k*-means algorithm is applied over the feature vectors building two clusters. Silhouette, Davies-Bouldin and Calinski-Harabasz indexes are applied to evaluate clustering. The test results demonstrated that the methodology presented in this paper is able to identify these patients. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved.

The remainder of the paper is organized as follows: related work section contains main work identified in the literature; the methodology proposed in this work is detailed in the section of same name; in the section tests and results, results are presented and discussed; and the conclusion section concludes this paper and indicates future work.

Related work

In Gerasimova *et al.* [10], dynamic thermography has been performed in 46 histopathologically proven breast malignant and benign tumor cases before surgery. From regions with and without tumor, thermal signals have been generated and Fourier Analysis, Wavelet Transform and phase diagram have been applied in order to examine the behavior of these signals. The authors have concluded that chaotic phase diagrams correspond to the healthy tissue; while for cancerous tissues, irregular shape in phase space is typical. Furthermore, results indicate that for healthy tissue, the thermal signals are anti-

correlated, whereas in cancerous tissue, thermal noise correlation has been observed. According to these authors, these results indicate that abnormal tissue has no ability to adapt to external influences. In addition, according to the authors, the results are consistent to the golden rule in biomedical sciences. This rule states that healthy and normal biomedical systems are often very complex and that the complexity decreases when an abnormality or disease occurs.

In another work, Gerasimova *et al.* [3] have performed multifractal analysis over breast thermal signals produced by dynamic thermography to detect differences in behavior between tissue with malignant tumor and healthy tissue. The modulus maxima wavelet transform method has been applied to characterize the multifractal properties of the one-dimensional thermal signals of cancerous and healthy breasts. The authors have concluded that complex scalar multifractal properties of the signals over autonomic regulation are drastically altered when the disease is present. In such a study, both breasts of 9 women were imaged, 6 with cancer and 3 healthy. A photovoltaic detector camera *InSb* was used. During acquisition of the images, the patient remains seated with the arms down in order to avoid discomfort. Frontal images were captured at 1 meter distance in a controlled temperature environment between 20ºC to 22ºC. Each set contains 30,000 image frames acquired during 10 minutes. Skin surface markers were used as landmarks for image registration phase in order to eliminate motion artifacts and do not hinder the analysis. In sick patients, the tumor region and the symmetrically positioned region in the other breast are delimited by square regions of 8 x 8 pixels. The analysis has been performed only within these regions. The methodology is able to discriminate between healthy region and region with tumor. Healthy regions exhibited signals with multifractal dimension as complexity signature and tumor regions exhibited signs with monofractal dimension as evidence of loss of complexity.

Recently, Gerasimova *et al.* published another work [11] using a much larger database, with 33 patients with histopathologically proven cancer and 14 healthy volunteers for control. The findings reaffirm the results of the prior study [3].

Scully *et al.*[12] also conducted analysis of skin thermal signal in patients with skin cancer. The temperature monitoring was performed by dynamic thermography. 1,500 thermograms, containing the region with disease, were collected using a camera FLIR SC7700, during 25 minutes. The collected images have been registered in relation to the first image and Wavelet Transform was used to analyze the thermal signals. They found significant differences between the region of control and tumor regions.

Herman [8] claims that dynamic thermography is able to measure the difference in infrared emission between healthy tissue and melanoma during the temperature recovery process after removal from cold stress. Test results suggest that the temperature is higher in cancerous lesion than in non-cancerous lesion during the first 45-60 seconds of thermal recovery. In this methodology, the region with tumor is surrounded by a bounding rectangle. Then, a digital photograph and an infrared image are captured. The infrared image observes the situation of steady state under ambient conditions. After, the skin is cooled by cold airflow per 1 minute. Next, infrared images are captured during 200 seconds. The methodology was applied in 37 patients, 3 with histopathologically confirmed cancerous lesions. The method achieved 100% accuracy.

Liu *et al.* [13] used thermograms to observe the forearm temperature variation and perform classification between three tissues: with micro-vascularity, with large veins and no veins. 3,000 thermographs were captured during 12 minutes and were registered afterwards with reference to the first image of the sequence. Thermal signals of each pixel were built, totaling 81,000 signs. Clustering using *k*-means algorithm and *Short Time Series (STS)* distance were applied over the signs to identify tissues with similar behavior in time. The authors demonstrated differences between three tissue types in terms of temperature change over time (via temporal profiles), magnitude of frequency response (via FFT), and coherency (via wavelet phase coherence and power spectrum correlation).

Unlike previous work, our methodology applies machine learning to decide whether or not a given patient is at risk of breast cancer.

Methodology

Our methodology steps are: image registration; thermal signal construction; thermal signal feature extraction; clustering; clustering evaluation and classification model building.

Image Registration

During examination by dynamic thermography, the patient may perform involuntary movements from breathing and momentary imbalance. These movements cause differences from one image to another and consequently generate thermal noise in the formed signs. Figure 1 shows (a) the first image of a particular patient, (b) the seventeenth image of the same patient, and (c) the result of the subtraction of (b) and (a) images. It is possible to see the difference between the images by observing image (c) which illustrates the effect of the movement of the patient during the examination.

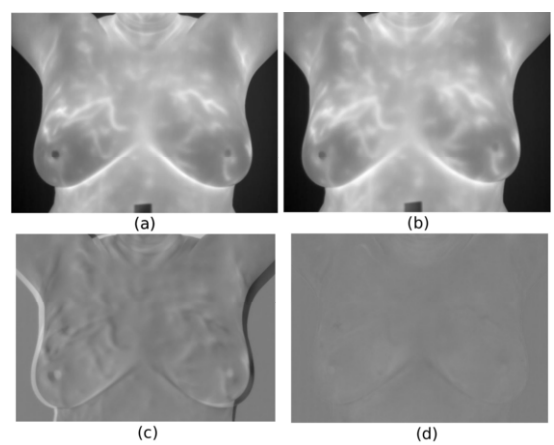

Figure 1- Image registration: (a) first image of sequence; (b) seventeenth image of sequence; (c) subtraction of (b) by (a) before registration; (d) the same subtraction after registration of these images.

In order to decrease the effects of these movements on thermal signals, registration of all images is performed. Registration is a process in which images, roughly speaking, are "matched". Considering two images (reference and sensitive) of the same scene, registration seeks to create a relationship between them to achieve the best overlap possible [14]. In this work, the first image of the sequence was considered the reference (immutable) image and the remaining images have been considered sensitive (transformable) images.

The registration is performed in two steps. In the first stage, a rigid body registration is applied based on intensities with geometric transformation consisting of translation, rotation, and scale. The second step uses non-rigid 2D registration with Residual Complexity (RC) [15].

The result of the registration can be seen in Figure 1(d). That image represents the subtraction of image (b) by image (a). After registration, the difference is much smaller than that one shown in image (c), before registration.

It is noteworthy that the data used in the analysis described in this paper are the temperature matrices of the breast surface, captured by an infrared camera. Thus the temperature matrix is converted to an image in gray tones, then the registration is performed in this image and the transformations are transferred back to the temperature matrix.

Thermal Signal Construction

After registration of the images, the formation of the thermal signal from the dynamic sequence of 20 images of a particular patient comprises the following steps:

1. The region of the breasts of the first image of the sequence is segmented creating a mask as shown in Figure 2(a);
2. The image region corresponding to the created mask is divided into a grid of 3x3 pixel squares R_k(Figure 2(b)), with k=1,...,p , where *p* is the amount of formed squares;
3. The averaged temperature of each square R_k is observed in all twenty images of the sequence producing the $S_k = (t_1,t_2,...,t_{20})$ thermal signal;
4. These twenty points are interpolated by cubic convolution yielding a signal of 1,901 points (in each interval between two points, 100 new points have been inserted, then 100x19 intervals = 1,900 + 1 (last point) = 1,901).

The S_k series values are ordered chronologically, *i.e.,* $t_1$is the mean temperature square R_k in the first image of the sequence, t_2 is the mean temperature square R_k in the second image of the sequence, and so on.

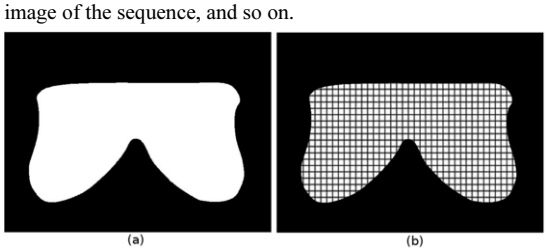

Figure 2 – Mask and grid of squares

Figure 3 shows the thermal signals of a sick patient and Figure 4 shows the thermal signals of a healthy patient. The *x*-axis represents time (the 20 points of each serie). The *y*-axis contains each of S_k (the index) of each thermal signal. The *z*-axis indicates the temperature in Celsius degrees. It can be seen that there is a group of thermal signals with higher temperature (red color) and with sharper increase in the early stages of the temperature recovery after thermal stress for the sick patient (Figure 3) however the same is not true for the healthy patient (Figure 4). The extracted features attempt to emphasize these differences.

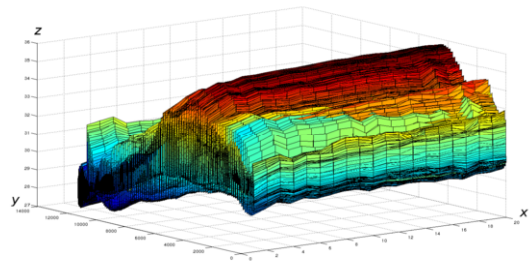

Figure 3 - Thermal signs of a sick patient.

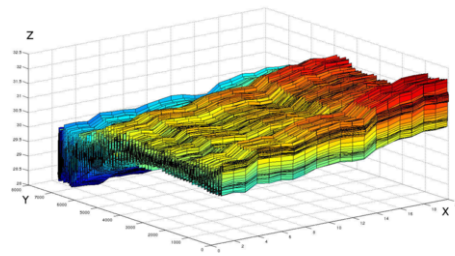

Figure 4 - Thermal signs of a healthy patient.

Thermal Signal Feature Extraction

The S_k time series can be treated as a biological signal of temperature t_i, i=1 ... N, where N=20. In biomedical systems, the evaluation of the complexity (main feature) is an important factor for diagnosis. This is possible through identifying and evaluating the complexity of the signal created by these systems [16].

In this work we used two signal complexity measures as features: the signal mobility which uses normalized first-order variations of the signal; and the signal complexity that uses second-order variations of the signal. The two characteristics measure the degree of variation along the signal. [16]. Considering a particular patient, from each sign t_i first-order variations are calculated, defined by d_i=t$_i - t_{i-1}$, i=2 ... N, and second-order variations, defined by g_i=d$_i - d_{i-1}$, i=3 ... N.

$$S_0 = \sqrt{\frac{\sum_{i=1}^{N} t_i^2}{N}} \quad (1)$$

$$S_1 = \sqrt{\frac{\sum_{i=2}^{N} d_i^2}{N-1}} \quad (2)$$

$$S_2 = \sqrt{\frac{\sum_{i=3}^{N} g_i^2}{N-2}} \quad (3)$$

Measures of complexity and mobility are defined in Equations (4) and (5), respectively:

$$C = \sqrt{\frac{S_2^2}{S_1^2} - \frac{S_1^2}{S_0^2}} \quad (4) \qquad M = \frac{S_1}{S_0} \quad (5)$$

Clustering, Evaluation of Clusters and Classification

The calculated features for each signal are stored in a vector. Clustering is performed over the set formed by feature vectors. In this work, we applied the *k*-means algorithm for clustering the vector set into two clusters. The two expected clusters are: thermal signals generated from diseased tissue and thermal signals generated from healthy tissue. We believe that for healthy patients the two clusters formed are very similar (healthy patients have no diseased tissues and, therefore, all the signs of both breasts are very similar), whereas for sick

patients the two clusters formed are more compact and well separated. The *k*-means algorithm has been performed by applying the correlation distance and three repeated steps, where each iteration has a different set of initial cluster centroid positions. The formed clusters are evaluated by clustering validity indexes. In this work, we applied: Silhouette index [17], Davies-Bouldin [18], Krzanowski-Lai index [19] and Calinski-Harabasz index [20].

The Silhouette index indicates the number of clusters that best separate the data set, *i.e.*, the maximum value of this index indicates the optimal number of compact and well separated clusters. It is defined by:

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (6)$$

where $a_i$is the average dissimilarity (*i.e.* how far away two elements are from each other using, *e.g.*, Euclidean distance) of i with all other data within the same cluster, b_i is the lowest average dissimilarity of i to any other cluster of which i is not a member. Note that $-1 \ S \ 1$.

The Calinski-Harabasz index is defined in terms of the traces of the between-clusters and within-cluster scatter matrices. The trace is defined to be the sum of the elements on the main diagonal of the matrix. Calculated for each possible cluster solution, the maximal achieved index value indicates the best data clustering. It is calculated using the following equation:

$$\mathrm{CH}=\frac{[\mathrm{traceB}/K-1]}{[\mathrm{traceW}/N-k]} \text{for } \mathrm{K} \in \mathbb{N} \quad (7)$$

where B denotes the error sum of squares between different clusters (inter-cluster) and W the squared differences of all objects in a cluster from their respective cluster center (intra-cluster).

The Krzanowski-Lai is based on the square differences of all feature vectors in a cluster from their respective cluster center. It is calculated using the following equation:

$$\mathrm{KL}(K) = \left|\frac{\mathrm{DIFF}(K)}{\mathrm{DIFF(K+1)}}\right| \quad (8)$$

where

$$\mathrm{DIFF}(K) = (K-1)^{\frac{J}{2}} \cdot \mathrm{traceW}_{K-1} - K^{\frac{2}{J}} \cdot \mathrm{traceW}_K \quad (9)$$

which is the difference between a clustering in K and a clustering in $K-1$ clusters. J is the number of variables that have been measured on each $x_i \in X$and *trace* W_K the sum of square function that corresponds to the clustering in K clusters.

Regarding the Davies-Bouldin index, it requires the dispersion measure and the cluster similarity (how much the elements resemble each other) measure. Thus it is defined as the ratio of the dispersion within the clusters and the separation between clusters and is calculated by:

$$\mathrm{DB}(K) = \frac{1}{K}\sum_{\mathrm{k}=1}^{K} R_k \text{ ,K} \in \mathbb{N} \quad (10)$$

where

$$\left(\frac{S_k+S_J}{d_{\mathrm{k,j}}}\right), \ \mathrm{k} \in [1,\dots,\mathrm{K}] \quad (11)$$

and

$$S_k = \frac{1}{\sum w_{\mathrm{k,i}}}\sum_{\mathrm{i}=1}^{n}\|x_i - \bar{x}_k\| \text{ ,k} \in [1,\dots,\mathrm{K}] \quad (12)$$

as well as

$$d_{\mathrm{k,j}} = \left\|\bar{x}_k - \bar{x}_j\right\| \quad (13)$$

In the case of this index, the minimum observed value indicates the best solution for clustering.

Clustering validity index values have been used as features to generate classification models. These values have been submitted to Weka tool [22] applying the *MultilayerPerceptron* (a neural network), the BayesNet (a bayesian network), and the J48 (a decision tree) algorithms.

Tests and results

The used infrared images are from the *Mastology Database for Research with Infrared Image - DMR-IR*. This database is described by Silva *et al.* [9], where the image acquisition protocol is also described with more details. Briefly, in the execution of the protocol, regions of the breasts and armpits of the patient are cooled by an electric fan for some minutes in an environment with controlled temperature (20ºC to 22ºC). After cooling, 20 images are captured during 5 minutes. Images are captured using a FLIR thermal camera, model SC620 [21]. The sensitivity of the camera is smaller than 0.04ºC, the detectable temperature range is between -40ºC and 500ºC and images are generated with dimension of 640x480 pixels.

Tests have been performed with images of 22 patients including 11 with histopathologically proven cancer and 11 healthy. For each patient, all images obtained by dynamic thermography are submitted to the steps of the methodology described in the previous section and implemented in MatLab. The results are in Table 1, Table 2 and Table 3, respectively. In all tests, the *k*-Folds Cross Validation technique has been applied in order to evaluate the classification results, with *k=3*. In all tables, TP Rate is the true positive rate, FP is the false positive rate and AUC is the area under the ROC.

Table 1: Neural network classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.81	0.00	1.00	1.00
Healthy	1.00	0.18	0.85	1.00
Weighted Avg	0.91	0.09	0.93	1.00
Confusion Matrix			Expected class	
			With cancer	Healthy
Classified as with cancer			9	2
Classified as healthy			0	11
Correctly Classified Instances			20	90.91%

Table 2: Bayesian network classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.73	0.00	1.00	0.93
Healthy	1.00	0.27	0.79	0.93
Weighted Avg	0.86	0.14	0.89	0.93
Confusion Matrix			Expected class	
			With cancer	Healthy
Classified as with cancer			8	3
Classified as healthy			0	11
Correctly Classified Instances			19	86.36%

Table 3: Decision tree classification result

Class	TP Rate	FP Rate	Precision	AUC
With cancer	0.91	0.09	0.91	0.92
Healthy	0.91	0.09	0.91	0.92
Weighted Avg	0.91	0.09	0.91	0.92
Confusion Matrix			Expected class	
			With cancer	Healthy
Classified as with cancer			10	1
Classified as healthy			1	10
Correctly Classified Instances			20	90.91%

The percentage of correctly classified instances ranges from 86.36% to 91.90% considering the results of all classifiers. From the results obtained, we can conclude that the proposed methodology is potentially able to identify patients at risk for breast cancer. However, we will definitely confirm the effectiveness of the methodology when we extend our test database. At the moment, we are working hard to that effect.

Conclusion

Breast cancer has been killing many women around the world. This work has exploited the fact that breast regions with cancer produce thermal signals with complexity alteration when analyzed over time. Thermal signals have been generated by dynamic thermography. The complexity of these signals has been measured by well-known measures described in the literature. Then the *k*-means algorithm has been applied to the extracted features for clustering them into two clusters. Clustering validity indexes were applied to identify patients at risk for breast cancer. The test results demonstrated that the methodology presented in this paper is able to identify these patients. Classification techniques have been applied on the index values and 90.90% hit rate has been achieved. In future work, other features will be extracted as well as other clustering algorithms and clustering validity indexes will be tested.

Acknowledgements

Thanks to Brazilian CAPES, CNPq and FAPERJ agencies for partially funding this work and to FAPEMA for funding the camera.

References

[1] M. H. Forouzanfar, K. J. Foreman, A. M. Delossantos, R. Lozano, A. D. Lopez, C. J. L. Murray, and M. Naghavi, Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis, *The Lancet*, 378 (9801), 1461–1484, 2011.

[2] C. Bouchardy; G. Fioretta; H. M. Verkooijen; et al. Recent increase of breast cancer incidence among women under the age of forty, *British Journal of Cancer*; 96 (11), 1743-1746, 2007.

[3] E. Gerasimova; B. Audit; S. G. Roux; A. Khalil; F. Argoul; O. Naimark1 and A. Arneodo, *Multifractal analysis of dynamic infrared imaging of breast cancer, EPL (Europhysics Letters)*, 104 (6), 2013.

[4] P. Kapoor and S. Prasad, Image processing for early diagnosis of breast cancer using infrared images, *2nd Int. Conf. on Computer and Automation Engineering*, 13 (1), 564-566, 2010.

[5] J. Koay; C. Herry and M. Frize, Analysis of breast thermography with an artificial neural network, *Engineering in Medicineand Biology Society – IEMBS*, 1 (1), 1159-1162, 2004.

[6] Y. Ohashi and I Uchida, Applying dynamic thermography in the diagnosis of breast cancer, *Engineering in Medicine and Biology Magazine, IEEE*, 19 (3), 42-51, 2000.

[7] W. C. Amalu, Nondestructive Testing of the Human Breast: The Validity of Dynamic Stress Testing in Medical Infrared Breast Imaging, *Engineering in Medicine and Biology Society*, 1, 1174-1177, 2004.

[8] C. Herman, The role of dynamic infrared imaging in melanoma diagnosis, *Expert Rev Dermatol*, 8 (2), 177-184, 2013.

[9] L. F. Silva; D. C. M. Saade; G. O. Sequeiros; A. C. Silva; A. C. Paiva; R. S. Bravo and A. Conci, A New Database for Breast Research with Infrared Image, *Journal of Medical Imaging and Health Informatics*, 4 (1), 2014.

[10] E. Gerasimova; O. Plekhov; Yu.Bayandin; O.Naimark and G. Freynd, Identification of breast cancer using analysis of thermal signals by nonlinear dynamics methods, *International Conference on Quantitative InfraRed Thermography*, 2012.

[11] E. Gerasimova; B. Audit; S. Roux; A. Khalil; O. Gileva; F. Argoul; O. Naimark and A. Arneodo, Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis, *Frontiers in Physiology*, 5 (176), 2014,

[12] C. G. Scully; W. Liu; J. Meyer; A. Dementyev; K.H. Chon; P. Innominato; F. Lévi and A.M. Gorbach, Time-Frequency Analysis of Skin Temperature in a Patient with a Surface Tumor Monitored with Infrared Imaging, *International Conference on Quantitative InfraRed Thermography*, 81-88, 2010.

[13] W. Liu; J. Meyer; C.G. Scully; E. Elster and A. M. Gorbach, Observing Temperature Fluctuations in Humans Using Infrared Imaging, *10th International Conference on Quantitative InfraRed Thermography*, 55-64, 2010.

[14] A. Ardeshir Goshtasby, 2-D and 3-D Image Registration: for Medical, *Remote Sensing, and Industrial Applications, Wiley, New Jersey*, ISBN 0-471-64954, 2005

[15] A. Myronenko and S. Xubo, "Intensity-Based Image Registration by Minimizing Residual Complexity," *Medical Imaging, IEEE Transactions on*, 29 (11), 1882-1891, Nov. 2010.

[16] N. Kayvan and R. Splinter, Biomedical Signal and Image Processing, *CRC Press, edition 1th*, 2005.

[17] P. J. Rousseeuw, Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Computational and Applied Mathematics*, 20, 53–65, 1987.

[18] D. L. Davies and D. W. Bouldin, A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1 (2), pp. 224–227, 1979.

[19] W. Krzanowski and Y. Lai, A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44 (1), pp. 23–34, 1988.

[20] T. Calinski and J. Harabasz, A Dendrite Method for Cluster Analysis, *Communications in Statistics – Theory and Methods*, 3 (1), pp. 1-27, 1974.

[21] Flir, http://www.flir.com/BR/, Accessed on April 2014, 1999.

[22] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten; The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11 (1), 2009.

Address for correspondence

Lincoln F. Silva: lsilva@ic.uff.br