A new Content-Based Image Retrieval methodology based on Image Registration algorithms processed in Grid Computing

Marcelo Costa Oliveira[1], Walfredo Cirne[2], José Flávio M. V.Junior[2], Paulo M. de Azevedo-Marques[3]

[1]*Departamento de Ciência da Computação/Universidade Federal de Alagoas (UFAL) – Campus Arapiraca, Arapiraca-AL, Brasil*

[2]*Departamente de Sistemas e Computação/Universidade Federal de Campina Grande (UFCG), Campina Grande – PB, Brasil*

[3]*Departamento de Clínica Médica/Universidade de São Paulo (USP), Ribeirão Preto – SP, Brasil*

Email: oliveiramc@gmail.com

Abstract

The Content-Based Image Retrieval (CBIR) has received great attention in the medical community because it is capable of retrieving similar images that have known pathologies. This work proposes a new methodology based on higher processing provided by the Grid Computing (GC) technology to achieve the CBIR using Registration algorithms. We evaluate the accuracy in retrieving images of the same plane, weighted and acquisition sequence related to a reference image. This works uses a heterogeneous image database and uses two anatomic regions (breast and head) as reference. The IR procedure uses Mean Square Metric (MSM) and Cross Correlation (CC). Both metrics showed a higher efficiency, MSM obtained 83% (breast) and 94% (head) precision average, the CC showed 81% (breast) and 98% (head). The higher computational cost related to the Image Registration were amortized by the GC, showing a potential low cost solution for secure data interchanging and integrating multiple hospitals and clinics.

1. Introduction

In the last 10 years, the CBIR has been one of the most studied computational vision technique. Through CBIR, it is possible to use a reference image to find similar diagnosed images [1]. Although part of the information about the diagnosis can be in the DICOM (Digital Imaging and Communication in Medicine) header, this textual tag frequently shows high error rate. Literature has reported cases with 16% of error [1]. This way, the medical community has emphasized the adoption of alternative methods in relation to the traditional methods based on the data manual insertion into the DICOM files.

The Image Registration (IR) is made of methods capable of find spatial transformations necessary to map homologous points between two images. The transformations are done in an iterative way until the best match is obtained between the reference and target images. However, all this iterative process results in a high computational cost that makes intractable to use the IR to find the best match between a reference image and a big image server [2].

The Grid Computing (GC) technology represents the most recent and promising tool in distributed computing. GC is the integration of many computers distributed geographically, making it possible to create a virtual computing platform, giving to users and institutions a virtually unlimited capacity to solve problems related to the storage and access of data, and also to process applications with high computational costs [3].

The main goal of this work is to develop an algorithm capable of using the GC to make viable the retrieval of similar images through the adoption of the IR techniques based on the Mean Squares Metric (MSM) and Cross Correlation (CC) techniques.

2. Materials and Methods

The application was developed in the operation system GNU/Linux Debian using the Java

1.5 programming language. The database image was developed with PostgreSQL-8.1-3 and Hibernate 3.0 and it followed the DICOM compliances. For its assessment, a heterogeneous image database was used, with 3000 magnetic resonance images from different anatomic regions. The application has two CBIR modules. The first module uses the second-order Texture Analysis (TA), suggested by Halarick [4], to filter the 1000 most similar images into the second module. The second module uses the IR algorithms to find the similarity between an image defined by the user as a reference and the images filtered by the first module. The application starts after the specialist select a reference image. When the first module is finished, the images are classified according to the smallest value of Manhatann Distance between the TA of reference image and the database images. Then, the 1000 most similar images are filtered to be used by the second module. The second module starts when the specialists select one of the two IR techniques implemented in this work: MSM and CC. The second module is processed on the OurGrid computational grid. OurGrid assumes that the parallel applications that run on it are Bag-of-Tasks (BoT) [3]. The application sorts in a list the most similar images according to the MSM or CC and show to the user.

3. Results

The efficacy of the IR algorithms used in CBIR was measured by the Precision versus Recall curves between the reference image and the images selected by the first module (Figure 1 and Figure 2). In all the experiments, the results produced by the IR algorithms were better than the traditional TA. The GC makes affordable for the application the high processing time of the IR algorithms (Table 1).

4. Discussion and Conclusion

Nowadays, the CBIR methods are still limited to detect specific diseases. Therefore, the CBIR methods need that new techniques of image processing be integrated into the traditional methods. The GC was fundamental to amortize the total processing time of the IR algorithms that permitted this work to show a new methodology to evolve CBIR's state of art techniques. The IR applied to the CBIR showed high precision, moreover, their results were better than the ones shown by the traditional TA. This work presents initial results in applying the IR as a CBIR technique where the main focus was to analyze the precision in retrieving different anatomies. Therefore, a deeper study still is necessary to analyze the real capacity of the RI methods in retrieval images from a pathological image database. A study in this way was already started.

	MSM		CC	
	Local	GC	Local	GC
Breast	118.39	5.31	115.88	4.78
Head	115.09	4.36	106.64	3.56

Table 1 - Compare time (minutes) between the local processing and the processing using the GC.

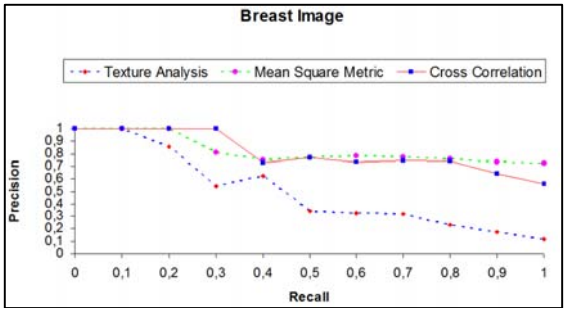

Figure 1 - Precision vs. Recall graph showing the retrieval techniques behavior in retrieve breast images.

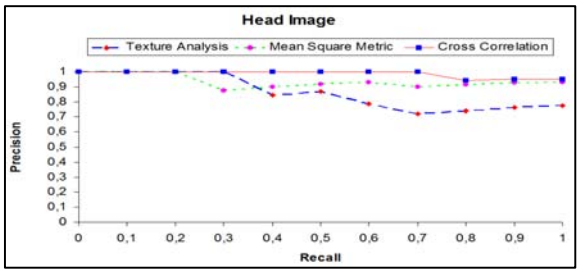

Figure 2 - Precision vs. Recall graph showing the retrieval techniques behavior in head images.

5. References

[1] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1-23, 2004.

[2] M. C. Oliveira, W. Cirne, and P. M. D. Marques, "Towards applying content-based image retrieval in the clinical routine," *Future Generation Computer Systems*, vol. 23, pp. 466-474, 2007.

[3] W. Cirne, F. Brasileiro, N. Andrade, L. Costa, A. Andrade, R. Novaes, and M. Mowbray, "Labs of the World, Unite!!!," *Journal of Grid Computing*, vol. 4, pp. 225-246, 2006.

[4] R. M. Haralick, Shanmuga.K, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems Man and Cybernetics*, vol. SMC3, pp. 610-621, 1973.