Available online at www.sciencedirect.com
ScienceDirect
Future Generation Computer Systems 23 (2007) 466–474
FGCS
FUTURE GENERATION COMPUTER SYSTEMS
www.elsevier.com/locate/fgcs

Towards applying content-based image retrieval in the clinical routine

Marcelo Costa Oliveira[a,*], Walfredo Cirne[a,1], Paulo M. de Azevedo Marques[b]

[a] *Laboratório de Sistemas Distribuídos, Departamento de Sistemas e Computação, Universidade Federal de Campina Grande, Campina Grande, Brazil*
[b] *Centro de Ciências das Imagens e Física Médica, Faculdade de Medicina, Universidade de São Paulo, Brazil*

Received 16 December 2005; received in revised form 26 April 2006; accepted 25 June 2006
Available online 10 August 2006

Abstract

Content-based image retrieval (CBIR) has been one the most vivid research areas in the field of computer vision, and substantial progress has been made over the last years. As such, many have argued for the use of CBIR to support medical imaging diagnosis. However, the sheer volume of data produced in radiology centers has precluded the use of CBIR in the daily routine of hospitals and clinics. This paper aims to change this status quo. We here present a solution that applies Computational Grids to significantly speed up the CBIR procedure, while preserving the security of data in the clinical routine. This solution combines texture attributes and registration algorithms that together are capable of retrieving images with greater-than-90% precision, yet running in a few minutes over the Grid, making it usable in the clinical routine.

Keywords: Content-based image retrieval; Texture attributes; Image registration; Grid Computing

1. Introduction

The volume of data produced in hospitals and medical centers is increasing fast. The annual production of the large radiology centers is about ten Terabytes per year [24]. This situation exists due to the ease with which the data of the patients are obtained and stored, resulting principally from the reduction of the cost of the computing equipments and image devices during the last years. However, more and more medical images represent huge quantities of data that need to be safely stored, automatically processed and must be indexed in an intelligent way, because they are fundamental pieces in the clinical diagnosis [23,21].

The increasing use of computers to aid the diagnosis (CAD) produced a fast development of computing algorithms applied to medicine in the last decades. The objective of CAD is to improve the accuracy of the diagnosis, and the consistency of the interpretation of the radiological image [2]. However, some CAD tools that show great results are not used in the clinical routine because they have high computational cost [13].

The difficulties in applying these CAD algorithms in the clinical routine and the limitations that still exist related to the storing, processing, searching and retrieving of images in large databases has been motivating companies and research institutions to find new solutions to solve these problems [3].

The Grid Computing (GC) technology represents the most recent and promising tool in distributed computing. GC is the integration of computers distributed geographically, making it possible to create a virtual computing application to solve problems related to the storing and access of mass data and to the processing of applications with high computing costs.

The shared processing capacity of GC allows the study of large data quantities using images processing algorithms that require high computing power. The use of GC also allows that health regional networks combine resources into one large distributed image database, making it possible for the medical communities to share data as well as medical applications and knowledge, and therefore making possible a greater interaction among the medical centers [18].

However, although very promising, we do not know of the use of GC to support diagnosis as part of the clinical routine of a hospital or clinic. The use of GC to support CAD seems to be yet restricted to medical research. The goal of this work is to analyze and introduce an accessible cost methodology based on the existing techniques, requirements and technologies in

* Corresponding author. Tel.: +55 33101365; fax: +55 33101498.
E-mail addresses: oliveiramc@lsd.ufcg.edu.br (M.C. Oliveira), walfredo@lsd.ufcg.edu.br (W. Cirne), pmarques@fmrp.usp.br (P.M. de Azevedo Marques).
[1] Tel.: +55 33101365; fax: +55 33101498.

0167-739X/$ - see front matter
doi:10.1016/j.future.2006.06.009

Grid Computing. This work used Texture Analysis to select the largest volume of images in the database that need to be submitted to Grid Computing. This capability avoids large data transference and the bandwidth bottleneck. In fact, Texture Analysis was capable of reducing the number of images that need to be processed using costly registration algorithms.

In this work, we also show the higher efficiency of registration algorithms in the retrieval of medical images. Moreover, we present the effective use of Grid Computing to run the images registration algorithm applied to the CBIR in the clinical routine.

2. Content-based medical images retrieval

Among the various CAD techniques, Content-Based Medical Image Retrieval Applications (CBIR) are a major beneficiary of the Grid Computing technology due to their characteristics and necessities: high processing and great need of large storage [24].

The objective of CBIR in radiology information systems is to give the right information to the specialist in the appropriate time, in order to improve the quality and efficiency of the diagnoses. In the process of clinical decisions, CBIR offers great benefits, being able to retrieve images in databases of the same form, anatomical region and pathology [25].

The use of image registration techniques applied to CBIR obtained considerable success in the medical images processing community due to the capacity of offering easy comparison between two images. This technique seeks for a space transformation able to map points of an image in the corresponding points of another image. The image registration can be done by means of a 2D rigid coordinate transformation, which combines rotation, translation and scale, in seeking for the maximum matching between a reference image and a target image. The rigid transformation is based on the squared error minimization between structures' outlines using similarity measure algorithms between the intensity of two images. The non-rigid transformation techniques are out of the scope of this paper and are described in [31,15].

However, to make a comparison between a reference image and a large medical image database using only one computer demands much more time than is reasonable for computer-aided diagnosis. Grid Computing makes feasible the use of the images registration technique applied to the CBIR, employing parallelism to reduce the time used to process the algorithm, so making this technique a promising tool to apply in the clinical routine [13].

In [24], there is presented a prototype to make a CBIR using medical image registration by means of Grid Computing resources. The application consists of a hybrid technique involving metadata and rigid registration algorithms. From a reference image, the application makes a selection based only on the metadata of the image and then these selected images are processed using the images' registration in the Grid. However, the authors says that the methodology applied to the selection resulted in a large volume of data to be sent and processed in the Grid, and that, in the clinical routine, this volume of data would be a lot bigger. Besides, the information contained in the medical images metadata can have a high error rate, there being related cases of up to 16% error [11].

The development of a CAD application must offer an effective and fast answer to the user. Therefore, solutions to optimize the volume of data that will be distributed and processed with assurance in the Grid are necessary.

3. Necessary requirements to the development of CAD applications using the Grid

Since the first project able to identify the potential of the Grid in biomedicine [23], a growing number of papers have shown the importance of the use of Grid Computing in the Life Sciences [21,23,26,20]. However, the state of art of the use of Grid Computing in CAD is still in prototype or for research only.

A great difficulty in introducing a CAD application using Grid Computing resources in the clinical routine is mainly related to the integrity and safety in the manipulation of the medical data in the Grid. We can also mention the complexity and little maturity of current Grid technology, bandwidth network problems that slow data transfers down, lack of a friendly and intuitive interface for the specialist (in the majority of cases laical in computing) and the development of a consistent application with the technologies already established of existing medical data management and medical images.

The main concern related to the distribution of medical data in the Grid is privacy. The major goal for any application that manipulates medical information is the respect for the privacy of the patient. The information and the images of the patients are confidential and should only be accessible by the medical team involved and the patient himself. This way, a medical Grid open to an institution or to a hospital federation must have a restricted access control and must ensure safe transference and safe data storing. The absence of a safe data integration is one of the greatest challenges to be solved in Grid Computing applications in medicine [13].

Data transfers are one of the main components in the total running time of CAD applications in the Grid. The Grid access to the database images involves concurrent access and the transference of a great volume of data. The solution proposed in [22] was to pre-replicate the data in the sites of the Grid, which are connected via a LAN. However, the storing of the distributed data in the Grid introduces data control issues that are more complex than in closed application. In order to assure this functionality, the applications should consider the safety requirements, as already pointed out in this paper.

The solution used in [26] to optimize the Magnetoencephalography data transference in a WAN was the use of a 1 GB/s network. However, this is a high cost solution and often not applicable to small clinics and public hospitals.

A procedure involving a CAD tool normally involves not only one algorithm, but a set of algorithms that use techniques coming from two knowledge areas: computing vision (which involves image processing to enhance, window/level, zoom,