C O N T E N T S

Computer Science

Workshop on Grid Computing

New Complex Approaches for Mining Medical Data

Camelia Vidrighin Bratu, Rodica Potolea
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Camelia.Vidrighin@cs.utcluj.ro
Rodica.Potolea@cs.utcluj.ro

Bogdan Petrut
"Iuliu Hateganu" University of Medicine
and Pharmacy
Cluj-Napoca, Romania

Abstract

The medical field has recently become one of the most challenging application areas for data mining techniques. The particularities involved in mining medical problems, such as domain knowledge, ethical and social issues, data quality, complexity and quantity, or cost, have lead to the necessity for more complex approaches. This paper tries to tackle some of the main issues, by introducing three new systems. ProICET adopts a hybrid approach to reduce the costs involved in the diagnosis process and help avoid "dangerous" errors. A second system, based on a PANE method, is introduced to reduce the misdiagnoses while keeping the transparency of the decision process. Finally, a system which combines different classifier predictions is presented. The benefits of such an approach include the possibility of establishing the baseline accuracy for any dataset, and the capability to consider data coming from different sources, with different structures.

1 Introduction

The medical domain is considered to be one of the most challenging areas of application in knowledge discovery. The main difficulties are related to the complex nature of the data involved (heterogeneous, hierarchical, time series), its quality (possibly many missing values) and quantity. Although hospitals hold huge amounts of records belonging to past treated patients, part of this data is not in electronic form, and the idea of transferring it to a database system is usually regarded as time consuming. Also, it is common for the physicians in different hospitals to have slightly different investigation methods. This results in different structures for the data coming from different sources, making it impossible to combine it in most cases. Moreover, since human life is at stake, accurate diagnosis is crucial. Establishing the baseline

accuracy for a given dataset is therefore very important as well. Domain knowledge or ethical and social issues are also of great significance.

An essential particularity of medical problems is the concept of *cost*, which is addressed by cost-sensitive classification. This idea will be developed further in Section 2.

This paper tries to tackle several issues involved in medical data mining. Section 2 introduces ProICET, a cost sensitive approach to the medical diagnosis process. The system introduced in Section 3 tries to improve the accuracy of symbolic classifiers, while keeping the diagnosis process transparent to the physician. Section 4 presents a system for classifier combination, based on the Dempster-Shafer Theory of evidence combination. The system can be employed to establish the baseline accuracy of a dataset, and to combine medical data coming from different sources, having different structure.

2 ProICET – A Cost-Sensitive System for Medical Diagnosis

2.1 Theoretical Aspects

When mining a medical problem, the concept of cost interferes in several key points. First of all, a doctor must always consider the potential consequences of a misdiagnosis. In this field, misclassification costs may not have a direct monetary quantification, but they represent a more general measure of the impact each particular misclassification may have on human life. These costs are non-uniform (diagnosing a sick patient as healthy carries a higher cost than diagnosing a healthy patient as sick). Another particularity of the medical diagnosis problem is that medical tests are usually costly. Moreover, collecting test results may be time-consuming; arguably time may not be a 'real' cost, but it does have some implication for the decision whether it is practical to take a certain

Exploring the textural parameters of ultrasound images to build an imagistic model for prostatic adenocarcinoma (ADKP)

Delia Mitrea, Sergiu Nedevschi
Department of Computer-Science
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Delia.Mitrea@cs.utcluj.ro
Sergiu.Nedevschi@cs.utcluj.ro

Monica Lupsor, Radu Badea, Ioan Coman
University of Medicine and Pharmacy
Cluj-Napoca, Romania

mmlupsor@yahoo.com
rbadea2003@yahoo.com

Abstract

The purpose of this paper is to present a method elaborated in order to build the imagistic model of prostatic adenocarcinoma, focused on the exhaustive set of textural features and the associated textural parameters. Methods like density modeling, regressions and Bayesian Belief Network, are used in order to learn and evaluate the relevance of the textural features. Thus, the relevant textural features are selected and their specific parameters are determined. The possibility of automatic recognition is also studied by applying classification on the textural features and by computing the recognition rates of various classifiers. We aim, in this way, to put in evidence the fundamental imagistic, textural properties of the prostatic adenocarcinoma, to correlate them with its visual properties, perceived by the human eye, and to select the best features for the non-invasive detection of this disease through semi-automatic or automatic, computer-assisted diagnosis.

1. Introduction

The prostatic adenocarcinoma (ADKP), the malignant tumor that causes prostatic cancer, is the most frequently met neoplasy at men. The incidence of prostatic adenocarcinoma increases continuously, being rarely present at men younger than 50. The most trusted method for ADKP diagnosis is, at this moment, the prostatic needle biopsy, but this is often dangerous, so a non-invasive method, based only on image analysis, would be preferred instead. We aim to elaborate such a method in our work by using texture-based analysis and recognition of tumors in ultrasound images.

Computerized methods, which extract information from medical images, are being widely studied nowadays in order to replace the old methods through a new, non-invasive technology. Thus, virtual biopsy tends to replace the old, traditional one, and this requires the development of adequate, computer-vision based methods, in order to obtain an accurate imagistic model of the malignant tumors. Texture is a fundamental visual property, which is essential especially in tissue characterization and in pathological structures recognition. Texture-based methods like the Grey Level Co-occurrence Matrix (GLCM) and its second-order parameters [2], [3], [16] fractal-based methods [4], the Wavelet Transform [18] and the Gabor Transform [11] are implemented in association with classifiers like the k-Nearest Neighbor (k-nn), Artificial Neural Networks, Support Vector Machines or Decision Trees [6] in order to perform automatic tumor differentiation for various kinds of organs. However, the relevance of the textural features, at multiple resolutions, the correlations between them and their statistical values for malignant prostatic tumors are not determined systematically through scientific, specific, automatic methods. We do this in our research, by using statistical pattern recognition specific methods like density modeling, regressions and Bayesian Belief Networks. The textural features computed on the original image are also being compared with features computed at various resolutions, on sub-images obtained by using the Wavelet Transform, in order to compare the relevance in visual information on various levels of resolution. The elaboration of the imagistic textural model of ADKP required three phases: (1) *the specification phase*, implying the *definition* of the data models and steps used for model generation; (2) *the model generation (implementation) phase* and (3) the *model validation phase*. Concerning the model generation phase, two steps are necessary: an *image analysis step*, consisting in the computation of the textural features and in the identification of the set of relevant textural features; then a *learning step*, in which we obtain some specific features like the mean, the variance, the maximum, the minimum value of the textural

Steatohepatitis Detection from Ultrasound Images Using Attenuation and Backscattering Coefficients

Cristian Vicas, Sergiu Nedevschi

Technical University of Cluj-Napoca
{cristian.vicas, sergiu.nedevschi}@cs.utcluj.ro

Monica Lupsor, Radu Badea
Mircea Grigorescu

3'rd Medical Clinic, Cluj-Napoca
{mmlupsor, rbadea2003}@yahoo.com

Abstract

Diagnosing liver nonalcoholic steatohepatitis (NASH) using non invasive procedures is challenging because the visual aspects in US imaging between healthy and affected liver are very much alike. Attenuation and backscattering coefficients are employed to determine the behavior of the US beams into the liver tissue. The values of these coefficients generate features for a classifier. We try to determine if is possible to discriminate between healthy and steatohepatitis patients and if is possible to assess the steatohepatitis grade.

Support vector machine is used as a classifier and the Cohen's kappa statistic evaluates the performance of the classifier. Experiments performed on US images acquired from a lot of 51 biopsied patients shown that these coefficients can successfully be used to distinguish between healthy and moderate/severe steatohepatitis. The role of fibrosis stage and fatty is also investigated.

Brunt score is used to assess the steatohepatitis grades. Fibrosis stage and steatosis grade are also determined.

1. Introduction

Nonalcoholic Steatohepatitis (NASH) is an important disease because of its prevalence in general population (over 20%) and because it could have a progressive course leading to cirrhosis and hepatic failure [1] [2]. Simplicity and noninvasiveness make ultrasonography an excellent diagnostic tool, albeit one that depends more on the ability and experience of the examiner and is affected by more unstable factors than any other imaging method. Several studies have discussed tissue characterization, which involves measuring the physical parameters of ultrasonograms to obtain quantitatively objective ultrasound information on which a diagnosis can be based [3].

There are two main directions one based on texture attributes of the generated US image and the other one based on the attenuation of the US beam into the liver tissue. This article focuses on the last approach. The behavior of the ultrasound beam into the liver can be assessed using two coefficients: backscattering and attenuation. Backscattering coefficient measures the quantity of energy that is returned by the tissue back to the receiver. Attenuation measures the ratio of absorption with respect to the depth.

Some of the methods employed to compute these coefficients are based on measuring the speckle distribution [4] directly from the RF signal [5] [6] by spectral shift central frequency method [7] or using a high frequency transductor [8].

B-mode ultrasound imaging implies transforming the amplitude of the received signal into intensity gray levels. Based on this property the behavior of the ultrasound into the liver can be assessed by examining the ultrasound image [1] [3] [9]. These methods are employed in this paper to differentiate between normal liver and steatohepatitis and to distinguish between various steatohepatitis grades. The role of fibrosis stage and fatty grade is investigated. It is shown that these pathologies affect the US propagation and therefore the discrimination between steatohepatitis grades.

In chapter 2 are presented the attenuation and backscattering coefficients, in chapter 3 is a short presentation of the classification algorithm. Chapter 4 presents the experimental results on a lot of 51 patients and in chapter 5 some conclusions are drawn based on the results.

Fibrosis detection from ultrasound imaging. The influence of necro-inflammatory activity and steatosis over the detection rates.

Cristian Vicas, Sergiu Nedevschi

Technical University of Cluj-Napoca
{cristian.vicas, sergiu.nedevschi}@cs.utcluj.ro

Monica Lupsor, Radu Badea
Horia Stefanescu

3'rd Medical Clinic, Cluj-Napoca
{mmlupsor, rbadea2003}@yahoo.com

Abstract

Diagnosing liver fibrosis using non invasive procedures is challenging because the visual aspects in US imaging between healthy and fibrosis liver are very much alike. In this paper texture analysis (texture feature computation and texture classification) are employed in order to increase the diagnosis value of US examination. An overview on fibrosis detection is necessary in order to determine and evaluate the best approach. Biopsy and METAVIR score are used to assess the liver pathology. The influence of steatosis and necro-inflamatory activity over the fibrosis detection is also investigated.

Four feature selection methods based on gain ratio, chi squared statistic, correlation and symmetrical uncertainty are evaluated. The results show that fibrosis, steatosis and activity alter the US image texture and implicitly the texture features.

It seems that the best approach in liver fibrosis identification is to build imagistic models for each fibrosis grade.

1. Introduction

The fibrosis is the scarring response formed in the chronic injury of any cause. It is a dynamic process, with a possibility of reversibility. For the moment, the golden standard in evaluating fibrosis is the liver biopsy. Using the liver biopsy one can establish with certainty the diagnosis, one can assess the severity of necroinflamation and fibrosis and one can distinguish the simultaneous liver diseases. On the other hand, it is an invasive procedure, with possible side-effects [1],[2]

An alternative examination procedure is ultrasound imaging. In fibrosis the visual aspects of healthy/affected liver are very similar so, the diagnosis value of the ultrasound imaging is relatively low in these diseases. Image processing techniques are used

to study ultrasound images and to improve the diagnosis value of ultrasound in diffuse liver diseases.

Features are computed over image textures and the images are classified based on the computed feature vectors.

One fractal dimension based algorithm (local multifractal morphological exponents) is used for the first time in fibrosis level detection field.

This paper is structured in the following manner: in chapter 2 are presented shortly the feature computation algorithms and the support vector machine classifier, in chapter 3 are presented some experimental results and chapter 4 contains discussions and conclusions.

2. Texture features and classification algorithm

Eight feature computation algorithms are employed: histogram statistics, grey tone difference matrix, grey level co-occurrence matrix, multifractal differential box counting, morphological multifractal exponents, multi resolution fractal dimension, Law's energy measures and wavelet transform.

Most of these algorithms are common in texture analysis field so they will be presented briefly.

2.1 First and second order statistics

Histogram statistics.[3]
The shape of the gray level histogram can provide clues about the image. Central moments are derived from the histogram in order to characterize the texture. These are mean, variance, skewness, kurtosis, energy and entropy.

Grey-Tone difference matrix [4]. A Grey-Tone Difference Matrix a column vector containing G elements where G is the total number of gray levels. Its entries are computed measuring the difference between the intensity level of a pixel and the average intensity computed over a square window centered at the pixel.

Ultrasound Elastography: from Physical Principles to Computer-Aided Image Analysis and Quantification

Sorin M. Dudea, Carolina Botar-Jid,
Dana Dumitriu
*"Iuliu Hatieganu" University of Medicine
and Pharmacy, Cluj-Napoca*

Sergiu Nedevschi, Tiberiu Marita,
Cosmin Pantilie
The Technical University of Cluj Napoca

Abstract

Elastography is a new development in the field of diagnostic ultrasound. The aim of the paper is to present the fundamental physics of the method, its advantages and limitations and the concept underlying a software product developed for improving the efficiency of the technique.

The method, as it is commercially available, uses speckle tracking to provide information on tissue strain. The specific information is displayed in color hues superimposed on a grayscale sonographic image. The extremely high motion sensitivity of the technique, although extremely useful, induces diagnostic limitations, to the point where the method has only qualitative value but no real quantification means.

To bypass some of the limitations, a software product has been developed with the specific aim to provide quantification means by image analysis. In this product, hue analysis is combined with area tracking, continuous area recognition and dynamic analysis of motion sequences.

1. Introduction

The assessment of mechanical tissue properties through palpation has been used in medicine, for ages. The main property assessed is elasticity or, as an opposite, tissue stiffness. Although information obtained by palpation has always been considered of extreme value, it is, by its nature, largely subjective and nonreproducible.

Pathologic changes in tissue structure are, by large, associated with changes in stiffness. Typically, cancer nodules are stiff while fat is elastic. On the other hand, many disease processes, although altering tissue structure and stiffness, are not accompanied by specific changes of the corresponding ultrasonographic image [1, 2].

The last 20 years witnessed tremendous efforts of the scientific community aimed at developing imaging methods for the visual characterization of tissue elasticity. Nowadays, the technique commercially available on ultrasound medical scanners is called elastography.

The purpose of this paper is to present the physical principles of elastography, to review its advantages and limitations and to introduce concepts of image analysis aimed at surpassing some of the method's drawbacks.

2. Physical principles of ultrasound elastography

The elastographic information is obtained by applying longitudinal pressure over an area of the body and measuring, on the corresponding ultrasound image, the displacement of speckle nuclei in the tissue, as a consequence of the applied external pressure [1, 2].

Essentially, elastography implies two distinct steps: tissue excitation and information analysis, to generate the specific image [3].

Excitation is a dynamic process, achieved by permanent vibration of the ultrasound transducer, applied with the "free-hand" technique. External excitation applies the deforming force on the surface of the skin, aiming to displace the underlying tissues. Unfortunately, free-hand technique is associated with transducer translation and lateral displacement becomes significant [4].

The **information analysis** is based on the assumption that a stiff area in the tissue will undergo less strain than elastic one. Observing the displacement of individual tissue speckle units allows the assessment of the distribution of tissue strain (figure 1). Therefore, elastography produces a strain profile of the tissues. The method measures local axial tissue strain changes

Computer Aided Diagnosis Tool for Cytological Slides

Ramona Galatus, Tiberiu Marita
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
ramona@bel.utcluj.ro, Tiberiu.Marita@cs.utcluj.ro

Abstract

Cervical cancer is one of the leading causes of women's' death, but if discovered early, the chances of a cure are as high as 90 percent. To reduce the risk of high stage cervical cancer and its incidence and mortality, women are advised to be screened with Pap test at least every two to three years. Because of high number of cytological slides needed to be processed in a cytological laboratory, a computer aided diagnosis tool is very helpful. This paper deals with an automatic system and method for detecting and classifying diagnostic cells. The main goal in this case is the recognition of Cervical Intraepithelial Neoplasia (CIN) grade. First a robust segmentation method is used. Then set of relevant features characterizing atypical cells are extracted. Finally a method for the CIN grade prediction is proposed based on the performance analysis of different statistical classifiers.

1. Introduction

Mass screening programs of the population to identify seemingly healthy individuals has been a growing trend for over 30 years, to prevent the incidence of some diseases which represent a substantial public health burden. Cervical Intraepithelial Neoplasia (cervical cancer, called CIN) is one of the most common form of cancer among women worldwide. As all the tumors, cervical intraepithelial neoplasia changes continuously in time. Over longer periods of time, its physical and functional characteristics change significantly. If it is discovered early, in the first stage of disease, the chances of a cure are as high as 90 percent. CIN has three grades of evolution: 1 (mild), 2 (moderate) and 3 (severe) and is characterized by some distinct features [1]: disproportionate nuclear enlargement, which leads to high nuclear-to-cytoplasm ratio, hyperchromasia,

irregularity in form and outline of the nuclei, irregular chromatin distribution, presence of keratinization, abnormalities in the number, size and form of the nucleoli, multi-nucleation.

The traditional screening test is called Pap smear and consists in the examination of cells collected from the uterine cervix in a cytology laboratory under the microscope. Even the best laboratories can miss from 10 up to 30% cancerous cases (false negatives), or made some mistakes (false positives) due to the following reasons:

- Huge number of normal slides being analyzed, each containing a huge number of cells;
- Large number of views that must be explored under microscope for each slide
- Short time allocated to each slide (10-15 minutes).

In this case the "cost" of false negatives is much important then false positive rates.

The impossibility for the human eye to detect all cases of cancer, lead to the need of creating some automated methods of cancer detection. The most significant researches done so far are: the Papnet (a system based on neural networks) [2], a fractal analysis-based approach [3], an optics-based approach [4], visual inspection detection [5], direct visual inspection or self-adaptive methods based on the existence of localized group of discriminatory elements [6] or image-processing based approaches applied on single cell images [7],[8]. Unfortunately, most of these approaches are not completely satisfactory and none detects perfectly the tumor, because of the complex CIN cell-patterns involved, with arbitrary orientation, location, features and scale.

The performance of a CIN grade recognition system depends on many factors:

a). *Slide preparation methods*: Having perfectly prepared smear slides is an expensive process because it implies technological resources like ThinPrep, AutoCytePrep, Sure-Path System [9] etc. The slides prepared with traditional techniques present some

Structured DICOM echocardiographic reporting for improved diagnosis, teaching and research in cardiology

Calin Homorodean, Dan Olinic
Nour Olinic
Medical Clinic no. 1
University of Medicine and Pharmacy
Cluj-Napoca, Romania

Sergiu Nedevschi

Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania

Abstract:

We developed a DICOM compatible, software environment and a data structuring methodology that was experienced in our echocardiography department, with the main purpose of improving the quality of diagnosis. An archiving system was generated, including a DICOM image server, a DICOM client application and structured reporting tools. In a first version of our structuring method, data was structured following four steps: investigation context, compulsory keyword (corresponding to anatomical structures), attributes and qualifiers. Using this reporting method we generated a database containing 420 echocardiographic images for 80 patients. At a later stage of our research, an advanced structured diagnosis reporting methodology was implemented, including concepts, contexts and templates that cover the diagnosis for all echocardiographic entities. The benefits of inhospital implementation of the structured reporting method and databases developed by us will be evident in the improvement of the quality of echocardiographic diagnosis, in ensuring a more scientific based data evidence in healthcare environment and it will be an extremely powerful tool, both for medical teaching and research., in an university environment,

1. Background:

1.1 The need for organizing medical information

The primary function of a clinical information system is to allow health care practitioners to report and retrieve clinical findings about their patients. In a medical imaging department without video films, with the information stored in a digital format, the interpersonal contact between the ultrasonographer and the clinician is very much diminished (2). The report is the final product by which the ultrasonographer makes contact with the clinician.

The lack of an intelligible indexing for the medical resources makes difficult the processes of looking up and retrieval of medical images, in a verifiable and ordered way. Such an images indexing is not possible in the absence of standardized protocols.

In most systems, the clinical findings are stored in the patient record database as free text. The problem with free text is that it lacks structure: information is trapped in the language of the report, making it difficult to compare reports or find a specific detail without reading through the text (3). In spite of the importance of the imagistic report, there are very few lingvistic guidings for their construction.

1.3 Combining images with text

The most important solution both for an efficient archiving and reporting of medical imaging data is the concept of combining images with the text and other image-related data (1). This concept is possible to be implemented in a digital environment. The physician can select images of interest from an imagistic study and include them in an electronic report, by hypertext links, using keywords. This will allow to the clinician reading the report to "click" on the underlined keyword, in order to visualize the image. These hypertext links are not limited to images from current examination. Images from patient's medical history and images from complementary imagistic modalities are also included in the reporting document, with the aim of comparing them, allowing to the reading physician, to visualize essential images for the diagnosis, together with measurements illustrating pathological findings and their modification in time (1). A large number of imagistic studies include hundred of images. By including the images in the report, the physician is able to identify the essential ones and to provide to the clinician the measurement which assist the decision making process. Thus, specific images may be identified, so that the measurement can be later performed at the same anatomic level.

1.4 The need for standardization protocols

In 1997, the "Integrating the Healthcare Enterprise" (IHE) project was launched in order to encourage

Tele-screening and Tele-monitoring System for the Surveillance of Hepatocellular Carcinoma

Ovidiu Dancea
Dorina Capatana
Rares Cazan
IPA SA - R&D
Institute for Automation
odancea@automation.ro

Horia Stefanescu
Radu Badea
Monica Lupsor
"Iuliu Hatieganu"
University of Medicine and
Pharmacy, Ultrasonography
Research Centre
hstefanescu@umcluj.ro

Luciana Neamtiu
Ofelia Suteu
Oncological Institute
Cluj-Napoca, Romania
luciana@iocn.ro

Abstract

Hepatocellular carcinoma (HCC) is one of the neoplasic diseases that meets the requirements to become the target of a close surveillance in a screening program: it is a frequent condition; it has a high rate of mortality; it affects a very well defined at risk population (chronic liver diseases patients). This population can be closely followed-up using a sensitive and easily accepted method (ultrasonography – US). Under these circumstances, the improvement of the screening and surveillance system for the patients at risk for HCC is needed, so the early diagnosis is increased, as well as the chance for an efficient and curative therapeutic intervention. A possible choice would be a telemedicine network for the US screening of a large number of chronic liver disease patients (especially liver cirrhosis) for the early detection of HCC.

1. Introduction

Nowadays, the hepatocellular carcinoma (HCC) is recognized as the fifth cancer worldwide, in terms of both incidence and mortality rate. A telemedicine-supported process that will monitor this disease into the northwestern part of Romania is both practical and advantageous. The objective of this paper is to describe the development and implementation of a Web application that allows efficient communication between general practitioners and ultrasound specialists, in order to support a telescreening programme for patients with HCC.
as cost-effective. The criteria that define an efficient screening program are: (1) high incidence and high mortality disease; (2) high sensitivity and specificity of

The screening process is performed according to the following steps: the patients are presenting to the general practitioner (GP) and undergo a specific medical consultation comprising, inter alia, ultrasound examination. The ultrasound images, acquired according to a specific protocol, described in section 1.2, are sent together with the other medical data regarding HCC screening to an expert diagnosis center. At this level, the specialists analyze the images and the additional information and establish the existence of cancerous formations. Based on the specialist evaluation, transposed into an electronic form, the system establishes a new date for the patients to present themselves to the GP, in order to undergo a new examination. The term for the new examination date appears to the GP when he/she accesses the patient's records. The application provides dedicated functionality for each user group already identified: GPs, imaging specialists, database administrators.

1.1 Epidemiology and risk factors

The incidence of HCC has a great geographical variability, which is related mostly to different risk factors incidence in general population [1]. Any agent that produces liver suffering and leads to cirrhosis is considerate to be a risk factor for HCC. The most important ones are hepatitis B virus (HBV), hepatitis C virus (HCV) and alcohol. The screening means testing of a large number of individuals designed to identify those with a particular genetic trait, characteristic, or biological condition [2]. This has been proven to be the most efficient method in diagnosis of cancers, as well the screening test; (3) easily accepted by the target population; (4) efficient curative treatment offered.

Applications of Telemedicine in Ultrasonography

Adrian Tarţa, Daniela Constantin

Numeris Com, 18/31 Moldovei St,
Baia Mare, Romania
TELEMED distributor for Romania

Radu Badea, Horia Ştefănescu
Sorana Iancu
Medical Imaging Department
"Iuliu Haţieganu" University of Medicine
and Pharmacy
Cluj- Napoca, Romania

Abstract

Telemedicine is generally understood as usage of telecommunications for improving health services and is used in various ways, including medical assistance for remote areas. Ultrasonography is a sensitive, reproducible, real time, non-invasive and cheap examination, being also operator dependent, fact that decreases its performances. The most important facility that telemedicine adds to ultrasound examination is the possibility of distance ultrasound consultation.

Telemed company found technical solutions developing new PC based ultrasound scanner – EchoBlaster - which offers high image quality and Telemedicine ready concept. This led us to possibility to acquiring, processing and transmitting ultrasound images via LAN, Internet or E-mail.

Our intention is to present few of the possible solutions for remote ultrasound images transfer in order to enhance the unlimited options that are available for users of PC based ultrasound scanners and show that in the modern world of communication technology PC based ultrasound systems are Telemedicine ready.

1. Introduction

Remote diagnosis using ultrasound imaging has received considerable attention in recent years. Until now, because of the limitation of technical resources it has been difficult to improve the diagnostic accuracy to

higher percentages. But this limitation is ready now to be overdrawn due to technology progress.

Two important steps were made in this direction and these steps refer to the main features that support a telemedicine ultrasound system:

Communication networks – now a wide range of fast connections for data transfer is available (Internet high speed connections, LAN&WAN, mobile communications), which can assure support for ultrasound imaging transfer.

Ultrasound devices – quality of ultrasound scanners has considerably increased in last years and systems for archiving and transfer ultrasound images to distance were developed.

Most usual configuration for ultrasound telemedicine consists of ultrasound scanner with video output, PC with video capture board, communication network, remote PC for remote ultrasound imaging display (Figure 1).
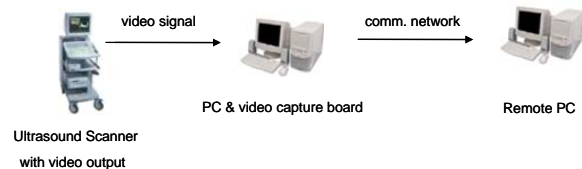
Figure 1 Usual configuration for ultrasound telemedicine

In order to simplify this complicate chain, Telemed company found technical solutions developing new PC based ultrasound scanner Echo Blaster 128 which offers high image quality and "Telemedicine ready" concept.

Echo Blaster 128 is based on the latest digital technology and provides high performance ultrasound imaging combined with advantages of a PC. Using Telemed equipment we can obtain next simplified telemedicine system (Figure 2):

Mathematical and Experimental Models of the Heat Transfer in Tissues

Dorian Popa, Octavian Mircia Gurzău
Tiberiu Mariţa
Technical University of Cluj-Napoca
{Popa.Dorian,gurzau}@math.utcluj.ro,
Tiberiu.Marita@cs.utcluj.ro

Mihai Lucan
Institute of Urology and Renal Transplants
Cluj-Napoca, Romania
Mihai.Lucan@renaltransplant.ro

Abstract

In this paper we construct a mathematical model of the heat transfer on an artificial structure based on TWMB - thermal wave model of bioheat transfer- which is an extension of the classical Pennes equation. A numerical method is used to solve this equation in one dimensional case. The numerical solution is compared with an experiment made to the Clinical Institute of Urology Cluj-Napoca.

1. Introduction

Cryosurgery is known as a surgical procedure from 1945 where James Arnott exposed a malign tumor of the skin to a frozen saline solution at a temperature of -22 Cº. As an effective method of treatment it was introduced by Cooper and Lee in 1961 when they presented the first cryostat for medical application.

The concept of cryosurgery as a minimal invasive technique, using multiple cryoprobes, was introduced in the middle of '80-'90, as a consequence of the development of the imagine technology in medicine, based on ultrasound and MNR (Onik-1985, Rubinsky-1993, Gilbert-1998). The first cryosurgical procedure that became a routine of treatment was in prostate cancer. The efficiency of the treatment depends on the precise placement of the cryoprobes into the tissue, which usually is irregular in shape, with different dimensions in each individual. For the planning of the cryosurgical procedure and the evaluation of the treatment effects is necessary to have a mathematical model and a computer program that can determine the dimensions of the ice ball and the distribution of the temperature in the tissue.

Modern methods of treatment in medicine as cryosurgery, cryopreservesion, therapy by radiation need understanding of thermal phenomena and the variation of temperature of living tissue. Mathematical analysis and prediction of temperature distribution in living tissues, in heating or freezing processes, are used in the study and optimization of cryosurgical procedures or irradiation procedures. In the last years, one of the most frequently approaches of the heat transfer in living tissue are based on Pennes equation (bio-heat equation):

$$C\rho \frac{\partial T}{\partial t} = \nabla(k\nabla T) - C_b \rho_b Jf(T - T_b) + S \qquad (1)$$

where T is the tissue temperature, T_b is the artery temperature, C is the specific heat, ρ is the density of the tissue, C_b and ρ_b are the specific heat and the density of the blood, J is the blood flow rate per unit tissue volume, K is the thermal conductivity of the tissue and S is the rate of metabolic heat generation.

Cofactor f, ranging from 0 to 1, is due to heat exchange between arterial and venous blood flowing through the nearest vessels.

The equation of Harry H. Pennes had a great influence on the study of the heat transfer in living tissue in the last fifty years.

Propagation of the freezing front is usually described by Pennes equation with the following boundary conditions:

$$v_n \rho L = -(k\nabla_n T)|_{\Gamma +} + (k\nabla_n T)|_{\Gamma -}$$
$$T|_{\Gamma +} = T|_{\Gamma -} = T_f \qquad (2)$$

where L is the latent heat of the fusion, Γ_+ and Γ_- are the boundaries of the freezing front of the living part and the frozen part of the tissue and T_f is the freezing temperature.

Analytical solution of Pennes one-dimensional equation was obtained by J.Liu.X.Chen and X.Xu [4]. Some important results on the problem (1)+(2) using different numerical methods was obtained in the papers [2], [3], [5], [7] where are presented some mathematical models of the heat transfer and computer programs for the simulation of the cryosurgical procedure.

Semi-Static Algorithms for Data Replication and Scheduling Over the Grid

Frédéric Desprez
INRIA - GRAAL Project/LIP Laboratory
UMR CNRS, ENS Lyon, INRIA, UCBL 5668
F-69364 Lyon Cedex 07

Antoine Vernois
LBP Laboratory / PMC Team
UMR 6023 CNRS U. B. Pascal Clermont-Ferrand II
F-63177 Aubière cedex

Abstract— **Managing large datasets has become one major application of grids. Life science applications usually manage large databases that should be replicated to let applications scale. Grid middlewares are asked to manage data and schedule computation tasks at the same time. These two important operations have to be tightly coupled to get the best results.**

This paper presents several heuristics that combines data management and scheduling using a steady-state approach. Starting from an intial distribution given by the resolution of a linear program, we optimize the results at run-time to improve the quality of the data mapping and the scheduling of computation requests when the request distribution evolves. Our theoretical results are validated using simulation and logs from a large life science application.

I. Introduction

An early and important usage of grid environments comes from applications managing large data sets [1], [2] in fields such as High Energy Physics [3] or life sciences [4]. To improve the global throughput of software environments, replicas are usually put at carefully selected sites. Moreover, computation requests have to be scheduled among the available resources. To get the best performance, scheduling and data replication have to be tightly coupled which is not always the case in existing approaches. In existing grid computing environments, data replication and scheduling are usually two independent tasks. In some cases, replication managers are used to find the best replicas in terms of access costs. However the choice of the best replication scheme should be done at the same time as the scheduling of computation requests.

Our motivating example comes from an existing life science application. The application tasks search for the signature or functional site of a protein (or protein family) in a database. They have the following characteristics: a large number of independent tasks of small duration and reference database sizes from several MBs to several GBs which are updated on a daily or weekly basis, several computational servers available on the network, and the size of the overall data set is too large to be completely replicated on every computational server. In order to run such applications on the grid, one has to solve two problems related to replication: **finding how and where to replicate the databases** and **choosing wisely the data to be deleted when new data have to be stored**. On the scheduling side, computation requests must be **scheduled on servers by minimizing some performance metric** that takes into account the data location.

In a previous paper [5], we have presented a static algorithm that provides simultaneous data management and scheduling using a steady-state approach. Using a model of the platform, the number of requests as well as their distribution, and the number and size of databases, we define a linear program to satisfy the constraints at every level of the platform in steady-state. The solution of this linear program will give us a placement for the databases on the servers as well as providing, for each kind of job, the server on which they should be executed. However, even if this algorithm is very efficient when the requests frequency and the load of the platform does not change over time, it reaches its limits rapidly on a highly dynamic environment. Thus we have improved our algorithm by adding heuristics that allow some data to be deleted or moved on some specific computing servers and adapt scheduling policies. On the replication side, we need to choose which data should be more replicated and which data must be deleted to gain storage space. On the scheduling side, servers must be chosen according to their load at a given time. Thus this leaves us with different combinations of data and server management. This paper presents a study of the behavior of this semi-static algorithm under dynamic constraints and some solutions to keep the overall system well balanced. Our theoretical results are validated using simulation and logs from a large life science application.

This paper is organized as follows. In the first section, we discuss some previous work in the areas of data replication and data and computation scheduling. In Section III, we present our model of the problem and the algorithm we designed to solve it. In Section IV, we present our semi-static algorithm which improves the data mapping when requests frequencies change. Finally, before some conclusions and our future work, we discuss our experiments using the OptorSim simulator [6] for replica managers.

II. Related Work

In computation grids, some work are focused on replication [7]; an example is Datagrid project from the CERN [8]. OptorSim [9], built within this project, allows one to simulate data replication algorithms over a grid. A modified version of this tool was used in our paper. In [10], several strategies are simulated including unconditional replication (oldest file deleted, LRU) and an economic approach. The target application is data management for the Datagrid physics

Digital libraries on GRID infrastructure

Doina Banciu

Bucharest University
doina.banciu@ici.ro

Gheorghe Sebestyen, Anca Hangan
Ramona Marfievici

Technical University of Cluj-Napoca
gheorghe.sebestyen@cs.utcluj.ro
anca.hangan@cs.utcluj.ro
ramona.marfievici@cs.utcluj.ro

Abstract

Digital libraries require high performance computing, storage and communication resources that may be provided by GRID infrastructures. This paper analyses the requirements of digital library systems and the possibilities of implementing digital library functionalities by using GRID middleware services as support. The authors propose a distributed, service-based model for digital libraries, which takes advantage of existing GRID middleware services.

1. Introduction

In a classical vision a library is a repository of knowledge organized as a huge set of documents preserved mainly on paper. A digital library is not only a digitized version of a classical library, it involves a new set of functionalities and services enabled and enhanced by the latest information and communication technologies. In a future vision digital libraries [1] "should enable any person to access all human knowledge anytime and anywhere, in a friendly, multi-modal, efficient, and effective way, by overcoming barriers of distance, language, and culture and by using multiple Internet-connected devices". In a wider way digital libraries must evolve to a knowledge repository and information exchange infrastructure that allows data generation, processing and seamless access to relevant information, regardless of the geographic distribution of hardware resources, databases or persons.

In order to achieve such goals a digital library must be built on a proper, high performance computing and communication infrastructure. Some recent research [2, 3] showed that GRID infrastructures and services may be a feasible development platform for digital library architectures. Many facilities offered by a GRID middleware are useful building blocks for digital libraries.

This paper analyzes the requirements of a modern digital library and evaluates the possibilities to use GRID middleware services as support for the implementation of digital library functionalities.

2. Recent research on GRID and Digital libraries (DL)

In the last decade, in order to achieve higher computing performance, grater data-storage facilities or an enhanced cooperative research infrastructure, an intense research activity was performed in the direction of GRID infrastructures. At the beginning, GRID computing facilities were developed mainly for scientific purposes. Nuclear and high energy physics, bio-chemistry and aero-spatial research are some examples of fields that require huge amount of storage resources and high performance computing facilities. CERN (Conseil Européen pour la Recherche Nucléaire) is one of the main promoters of GRID research; the experiments which will be made in the next years at CERN will generate a huge amount of data that must be stored in real-time on distributed databases. The EGEE (Enabling Grids for E-sciencE) initiative is intended to develop the necessary GRID computing infrastructure that covers the storage and processing requirements for the high energy physics experiments.

As part of the EGEE initiative, the Diligent project is intended to define and implement digital libraries on GRID infrastructure. In the first stage the research team tried to identify the requirements imposed for a digital library in two very different areas: arts (paintings) and satellite imaging. Based on these requirements, a GRID-based digital library architecture was specified [2]; now a beta version of the library services is in the implementation phase.

The Delos Network of Excellence [1] is a European research project intended to define and implement digital libraries on new computing and communication technologies. The main contribution of this project was

SEE-GRID project - support for sustainable development of Grid oriented infrastructure and applications in the SEE region

Gabriel Neagu, Alexandru Stanciu, Vladimir Florian, Adrian Vataman
National Institute for Research and Development in Informatics - ICI
gneagu@ici.ro, stanciu@grid.ici.ro, vladimir@ici.ro, avataman@ici.ro

Abstract

The paper is devoted to the European project SEE-GRID which started in May 2004 with the vision to pave the way towards the provision to distributed research communities in the region with an independent of geographic location, round-the-clock, common market of computing resources. The second stage of the project is under development aiming at sustainable development of Grid infrastructure at the national level, in line with current orientation of the eInfrastructure development at the European level. The focus is put on the results achieved by the SEE-GRID-2 project during the first year period, with the stress on sustainability policies and solutions, infrastructure development and gridified applications offer.

1. Introduction

The European FP6 project SEE-GRID - South-Eastern European Grid-enabled eInfrastructure Development was conceived as a four-year project structured on two stages. In fact there were two separate projects, as the approval of the second stage was based on the thorough evaluation of results obtained during the first stage. For the first project stage (May 2004 – April 2006), besides the European Center for Nuclear Research (CERN) Geneva, the consortium included organizations from 10 countries in the SEE region: Greece, Albania, Bosnia and Herzegovina, Bulgaria, Croatia, FYROM, Hungary, Romania, Serbia, and Turkey. The project is coordinated by the Greek Research and Technology Network (GRNET) from Athens. ICI Bucharest represents the Romanian partner – a consortium of five RoGrid member organizations including also University "Politehnica" of Bucharest, National Institute for Physics and Nuclear Engineering, National Institute for Aerospace Research, and University of Bucharest.

The main project objectives were the following [1]:
- create a human network in the area of Grids, eScience and eInfrastructures in SE Europe;
- integrate incubating and existing National Grid infrastructures in all SEE-GRID countries, building upon and exploiting the infrastructure provided by the Gigabit Pan-European Research & Education Network (GEANT) and the South-East European Research and Education Networking (SEEREN) initiative in the region;
- ease the digital divide and bring SEE Grid communities closer to the rest of the continent;
- establish a dialogue at the level of policy developments for research and education networking and provide input to the agenda of national governments and funding bodies;
- promote awareness in the region regarding Grid developments through dissemination conferences, training material and demonstrations for hands-on experience;
- migrate and test Grid middleware components and APIs developed by pan-European and national Grid efforts in the regional infrastructure;
- demonstrate Grid application of regional interest;
- expand the operations and support center activities in all participating countries.

The SEE-GRID-2 project started in May 2006, including two additional countries: Moldova and Montenegro. The main project directions may be summarized as follows [2]:
- Policy-focused deployment strategy: shift priority from a "top-down" approach (i.e. from regional project execution to national implementation) towards a "bottom-up" approach (from national priorities, cooperation, and innovation to regional cohesiveness);
- Growth of infrastructure:
  - expand regionally to include new countries/areas and widen the SEE eInfrastructure community;
  - expand nationally to include new sites/institutes and strengthen collaboration in each country;

Cryptographic and Cryptanalitic Algorithms for Grid Applications

Alin Suciu, Rodica Potolea
Technical University of Cluj-Napoca

{Alin.Suciu,Rodica.Potolea}@cs.utcluj.ro

Gabriel Neagu
National Institute for Research and Development in Informatics – ICI
gneagu@ici.ro

Abstract

The concept of grid computing addresses the next evolutionary step of distributed computing. The goal of this computing model is to make a better use of distributed resources, put them together in order to achieve higher throughput and be able to tackle large scale computation problems. In this paper we investigate the particularities of implementing cryptographic and cryptanalitic algorithms for the grid. Special attention is paid on finding ways to increase the performance of these algorithms by exploiting the data parallelism, where possible. The results presented here come from a series of benchmarks carried on our local grid (GridMOSI), using various cryptographic and cryptanalitic algorithms.

1. Introduction

A challenging computational problem is making resources available for those needing them sparingly, without necessary owning them. The birth of grid computing was very often associated with the introduction of the electrical power grid due to certain similarities [1]. Back in the beginning of the 20th century, electric power generation was possible, but the real problem was making it available worldwide without the necessity of each home consumer to possess an electric generator.

What once was the electrical power is now the computational power, and what once were the electrical generators are now high performance computing systems with large storage capabilities. Offering storage and computation capabilities to users not having those resources is the ultimate goal of grid computing. The revolutionary thing to do would be to introduce a grid infrastructure to make computing power and resources available to a greater extent.

Nowadays scientists are more and more concerned with how many floating point operations per month or per year they can extract from a computing environment, rather than considering floating point operations per second [2, 3]. With the introduction of the grid concept, more attention has been devoted to such computing environments known as High Throughput Computing environments.

Parallelization has emerged as a need for improving the response time – the time needed to generate the solution to a problem. To achieve parallelization of a problem there are two main approaches [4]:

- Data parallelism – refers basically to running the same code on different pieces of input data. The major task here is splitting the input into chunks of data on which the same algorithm may be applied. This may be done only when the particularity of the problem allows it, when working on a subset of the initial input does not alter the final solution. Constraints may occur that can limit the overall potential for parallelism.
- Control parallelism – refers to running different sequences of code on the same data in order to obtain the result. The task here is parallelizing the problem at the algorithm level, i.e. create a parallel approach to the solution by splitting the computing responsibilities among several processes. This is a much more challenging approach, as it involves handling communication among processes which might increase the overall running time. In order to obtain a better performance, communication among processes must be kept to a minimum.

The use of cryptographic algorithms is widespread nowadays, but adapting them for grid-based applications is not a trivial problem. Moreover, performance increases can be obtained for some classes of algorithms by exploiting the data parallelism inherent to these algorithms. Obviously there are some