Med-e-Tel 2013

Electronic Proceedings

of

The International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business

Editors
Malina Jordanova, Frank Lievens

April 10-12, 2013
Luxembourg, G. D. of Luxembourg

Publisher

International Society for Telemedicine & eHealth (ISfTeH)
Coordinating Office
c/o Frank Lievens
Waardbeekdreef 1
1850 Grimbergen
Belgium
Phone: +32 2 269 8456
Fax: +32 2 269 7953
E-mail: info@isfteh.org
www.isft.org

Med-e-Tel 2013
Electronic Proceedings: The International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business
Editors: Malina Jordanova, Frank Lievens
ISSN 1818 - 9334

With the publication of this Electronic Proceedings as well as with the sixth book from the series "Global Telemedicine and eHealth Updates: Knowledge Resources", Med-e-Tel strengthens its position as a widely recognized International Educational, Networking and Business Forum for eHealth, Telemedicine and Health ICT.

Enjoy your reading!

Editors
Malina Jordanova, MD, PhD
Educational Program Coordinator, Med-e-Tel
Space Research & Technology Institute, Bulgarian Academy of Sciences
Bulgaria

Frank Lievens
Director, Med-e-Tel
Board Member, Secretary and Treasurer of International Society for Telemedicine & eHealth (ISfTeH)
Belgium & Switzerland

Speed-up the Image Retrieval of Lung Nodules in the BigData Age

M. C. Oliveira[1], D. C. da Silva[1], K. B. C. dos Santos[1]
[1]Instituto de Computação(IC) e Laboratório de Telemedicina e Informática Médica do Hospital Universitário(HUPAA) da Universidade Federal de Alagoas (UFAL), Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, Maceió, 57072-900, Brazil, oliveiramc@ic.ufal.br

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. However, the sheer volume of data produced in radiology centers has precluded the use of CBIR in the daily routine of hospitals. The volume of medical images produced in medical centers has increased fast. The annual data produced from exams in the big radiology centers is greater than 10 Terabytes. Therefore, we have reached to an unprecedented age of "BigData". We here present a bag-of task approach to speed up the images retrieval of lung nodules stored in a large medical images database. This solution combines texture attributes and registration algorithms that together were capable of retrieving images of benign lung nodules with greater-than-72% precision and greater-than-67% in malignant cases, yet running in a few minutes over the Grid, making it usable in the clinical routine.**

Introduction

The volume of data produced in medical centers has increasing fast. The annual production of the big radiology centers is about 10 Terabytes. This situation exists due to the ease that the data of the patients are obtained and stored, resulting mainly from the reduction of the cost of the equipment during the last years.

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. However, the sheer volume of images produced in radiology centers has precluded the use of CBIR in the daily routine of hospitals. Therefore, we have reached to an unprecedented age of "BigData" and it has been motivating research and companies to find new solutions.

The Grid Computing (GC) technology represent one of 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. Techniques focused on the medical image retrieval are a major beneficiary of the GC technology due to their characteristics and necessities: high processing and large storage. Besides, GC is a low cost solution for public hospitals and small clinics, because, it is able to use the idle recourses of computers [1]. This paper presents a Bag-of-Task GC approach to speed up the images retrieval of lung nodules stored in a big medical images database.

Application Description

The overall application is described in Figure 1 and a detailed description is given in the sections below. All the images inserted in the PACS have removed the patients` information in respect to the HIPAA [2]. We have used 20,000 images from the Lung Image Database Consortium (LIDC), which is a BigData of lung cancer [3]. In the LIDC each nodule is manually segmented and then classified in benign or malignant by physicians. To use as reference we selected 100 benign nodules and 100 malignant nodules

The application has two CBIR modules. The first module uses the second-order Texture Analysis (TA) (co-occurrence matrix) to filter the 1000 most similar images into the second module. The second module uses the Image Registration (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 second module starts when the specialists select the registration module. The IR algorithm is executed in parallel on the grid machines using the reference image and the images classified by the first module. The application sorts in a list the most similar nodules according to the Cross Correlation values. Based on the DICOM protocol information, the application can also retrieve the information stored on the LIDC to aid the physician.

Because of the high computational cost related to the IR algorithms, the second module is processed on the OurGrid computational grid (www.ourgrid.org). This grid is free-to-join and cooperative, where sites share their idle computing resources and, when necessary, receive idle resources from other sites. OurGrid assumes that the parallel applications that run on it are a Bag-of-Tasks (BoT), i.e., those tasks are independent from each other. Currently, the OurGrid is composed of nearly 500 computers [4].

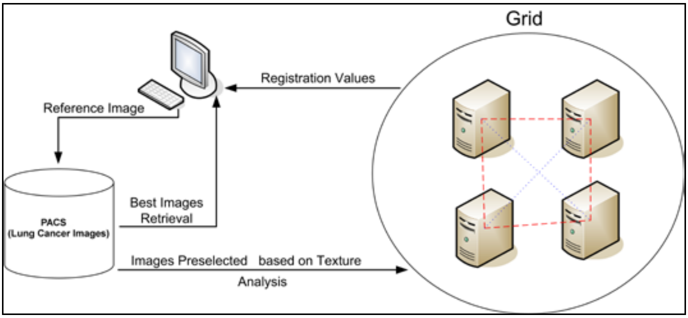

Fig. 1 CBIR lung cancer image retrieval schema using Grids

Results

The results of the CBIR tool developed in this work were assessed by the leave-one-out technic. The tests were repeated five times for each module and we have used different kinds of nodules for each test (ten benign nodules and ten malignant nodules). The nodules were selected in a random way. The algorithm precision was obtained by dividing the type number of nodules retrieved over the number of nodules retrieved.

The first module's time processing average was 1.9 minutes. This time was related to the calculation of the Manhattan Distance between the characteristic vector of each DICOM server's image and the characteristic vector of the reference image. The average precision of retrieving obtained by the first module was 0.73 (benign nodules) and 0.76 (malignant nodules), considering the ten most relevant nodules retrieved. To analyze the image retrieval capacity of the second module, a group of 1000 most similar nodules filtered by the first module has used in the second module. The average precision of retrieving obtained by the second module was 0.72 (benign nodules) and 0.67 (malignant nodules). In all the experiments, the results produced by the IR algorithms were very close to the traditional TA technique.

The Bag-of-Task approach was able to greatly reduce the high processing time of the IR algorithm. The GC was able to amortize the total time of the algorithm in 116.97 minutes, compared to the processing time obtained in one machine. The experiments used 50 Grid processors and 100 MB/s LAN. The application total time, considering the time to calculate the Manhattan Distance and the time to execute the IR, was 5.02 min.

Discussion and Conclusion

A computational algorithm was developed in this work, capable of using the high processing power of the Grid Computing technology to make feasible the CBIR using the Image Registration algorithms against a BigData of lung cancer images. Oliveira MC and colleagues [1] showed a preliminary work using IR as CBIR technique. However, the authors did not focus his attention in a specific disease. To our knowledge the results using IR and TA to retrieve lung nodules have never been published. Furthermore, our results evidenced that IR is precise and effective in retrieve similar lung nodules, besides that, the IR has showed very close results to the traditional TA technique.

The retrieval processing time was more than 100 minutes in only one computer using IR techniques. This time is impracticable for a CBIR to be applied in the clinical routine. In the grid, however, this time was reduced to less than 3 minutes in mean, making it affordable for clinical use. Therefore, a Bag-of-Task approach was fundamental to amortize the total processing time of the Image Registration algorithms against a BigData of lung cancer images. Besides this, we have shown a new methodology to evolve CBIR's state of art techniques through the Image Registration techniques.

Acknowledgment

This study was supported by Fundação de Amparo à Pesquisa do Estado de Alagoas (FAPEAL) and Hospital Universitário Prof. Alberto Antunes (HUPAA/UFAL).

References

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

[2] B. J. Liu Z. Zhou, H. K. Huang, "A HIPAA-Compliant Architecture for Securing Clinical Images," *Journal of Digital Imaging*, vol. 19, no. 2, pp. 172–180, 2006.

[3] H. Lin, Z. Chen, and W. Wang, "A pulmonary nodule view system for the Lung Image Database Consortium (LIDC).," *Academic radiology*, vol. 18, no. 9, pp. 1181–5, Sep. 2011.

[4] W. Cirne, Brasileiro F, Andrade N, Costa L, Andrade A, Novaes R, and M. M., "Labs of the World, Unite!!!," *Journal of Grid Computing*, vol. 4, no. 3, pp. 225–246, 2006.

Marcelo Costa Oliveira is member of the faculty at Computer Institute of the Federal University of Alagoas, in Brazil. Dr. Oliveira holds a PhD. in Medical Informatics from University of São Paulo (USP), M.Sc degree in Physics Applied to Medicine and B.Sc in Computer Science. Dr. Oliveira heads the Research and the Telehealth Project of the Hospital of the University of Alagoas. His current research interests include BigData Health, Cloud Computing and Medical Image Processing.

Darlan Chrystian da Silva is an undergraduate student at Federal University of Alagoas, in Brazil. He is member in the Research and the Telehealth Project of the Hospital of the University of Alagoas. His research interest BigData Health.

Keila Costa Barbosa's degree in Systems Analysis from the University of Northern Paraná, in Brazil. Degree in Systems Analysis and Development of Health State University of Alagoas. He is currently a MS in Computational Modeling Knowledge Federal University of Alagoas and Post graduate in Technologies for Web Applications at the University of Northern Paraná. His research interests include Health BigData, Cloud Computing and electronic medical records.