Integrating Images to Patient Electronic Medical Records through Content-based Retrieval Techniques

Agma Traina[1], Natália A. Rosa[2], Caetano Traina Jr.[1]
[1] *Computer Science Department - University of São Paulo at São Carlos - Brazil*
[2] *Center for Science of Image and Medical Physics - Medical School of Ribeirão Preto - University of São Paulo at Ribeirão Preto - Brazil*
agma@icmc.usp.br, naty@hcrp.fmrp.usp.br, caetano@icmc.usp.br

Abstract

This paper presents the SRIS-HC – an Patient Electronic Medical Record System with support for Content-based Image Retrieval, developed aiming at demonstrating the benefits of having the ability of similarity retrieval over image datasets based on their contents, at the Clinical Hospital of the Medical School of Ribeirao Preto of the University of Sao Paulo at Ribeirao Preto – Brazil (the HCFMRP/USP). This ability is an additional resource developed over a PACS system, intending to enable the improvement of medical diagnosis by images, as well as to provide a basis to perform analysis of similar medical cases and bibliographic research over them. The SRIS-HC was developed on top of the Radiology Information System (RIS) of the Radio-diagnosis Laboratory of the HCFMRP/USP, which is in turn at the core of the Electronic Patient Record System of the hospital.

1. Introduction

Nowadays the generation of data by hospitals and medical centers is increasing in a very fast pace. This is mainly due to the facility to obtain and record the data from patients, what happens mainly by the decreasing price of the computational equipments and image devices. Therefore, the amount of images obtained by computerized radiography, CT, MRT and ultrasound among others, makes the data management of the hospital a burdensome task. The major part of this increasing volume of data is composed by images, so it is important to have systems specialized in the organization and distribution of the images to the users, i.e., to the physicians and the radiologists involved in the patient care.

With the introduction of the Picture Archival and Communication Systems (PACS) [1] [2], the image management tasks in medical care units have gained a new direction, as PACS are used to capture, store, distribute and display medical images. For diagnostic imaging applications, PACS can be utilized to achieve a nearly filmless environment [3]. However, the organization of large datasets of images, in a way that promotes an easy retrieval of them, requires additional tools.

Hospital Information Systems (HIS) can be broadly divided into two classes: Administrative and Clinical. Administrative functions include room and services reservations, accountability for drugs and supplies, billing, etc., and the data structures deal mainly with numerical and textual information used in the management of the hospital. Clinical functions are usually specialized to help physicians in their day-to-day clinical practice. It includes information describing the physiological characteristics of each patient, its medical historic and exams. Usually, clinical data stored in HIS are also in numerical and textual format. Traditional database management systems are very efficient dealing with this type of data.

Radiology departments stand apart from the other departments of a hospital, regarding the requirements of the information systems they need to manage, because the exams provided are always images. Therefore, Radiology Information Systems (RIS) should be specially tailored to handle images, making heavy use of PACS to store their data [4].

As the information about a patient is spread in the HIS and RIS systems, a Patient Electronic medical Record (PER) system aims at integrating all information about the clinical history of the patients, trying to isolate administrative burdens, so that the medical information can be more easily distributed among different medical institutions taking care of the same patient [5] [6]. Nowadays, existing PER systems are restricted to store in digital media only textual and numerical medical information, keeping the part of exams separated from the patient digital database [7].

However, allowing the integration of a PACS and a Radiology Information Systems to the HIS can bring new perspectives into patient care and treatment, because all the information regarding the medical life of the patients can be instantaneously available to the physicians in charge of them. Therefore, experimental systems are already deployed in some leading medical centers. In these systems, through the textual/numerical data of the PER, health professionals can retrieve every information about the patient, including the images of the exams [8] [9, 10].

Nevertheless, the support that exists for images, even in PACS are beneath the support that exists for textual/numeric data [11]. This is due as textual/numeric data can be used as a key for searching the database, whereas images only can be retrieved based on external keys (textual/numeric attributes) to which the images are linked to. Whereas using a medical image as a key to retrieve one patient medical record is not a needed operation, the ability to retrieve exams based on the similarity of their images is useful for a variety of operations [12] usually asked over exams based on textual/numerical data, as for example: searching for similar medical cases to compare/study/diagnose, statistical studies, decision support, cross-referencing, bibliographic reports and documentation [13].

This paper presents a system that implements image similarity search in a HIS/RIS of a large university hospital. The existing Patient Electronic medical Record (PER) management system was therefore extended with an integrated **Similarity Retrieval of Images System** (**SRIS**), making it possible to browse on all the patient data, enabling health care personnel to perform similarity queries over the image exams of the patient. For example, "*given the Jane Doe´s abdomen X-Ray taken on January 15th, 2003, retrieve the 10 most similar exams to it from the database*". Therefore, using a HIS/RIS/SRIS integrated system, the physician can access not only the description of the exams and patient's information, but also to cross and compare the present analysis done with similar cases already treated and stored in the database.

2. The similarity retrieval of images sub-system

The system presented herein is a SRIS called the *SRIS-HC* sub-system (the SRIS from the Clinical Hospital - HC in Portuguese), which was developed as an extension of the HIS/RIS already existing in this hospital, a large Brazilian university hospital of the University of São Paulo at Ribeirão Preto. The resources of the original system were completely maintained, including the user interface and the data structure of the existing patient PER management system. The HIS system is the hub where every other sub-systems of the hospital departments are already integrated, so attaching the SRIS sub-system to it make the RIS/SRIS available to every other sub-system. Keeping the original structure of the user interface is important due to the training of the health care personnel regarding the use of the system. Therefore, the existing operational procedures are maintained, and the access to images can be progressively enabled just where it is needed, with the gradual expansion of image support at each department.

The *SRIS-HC* sub-system was developed in C++ using the Borland C++Builder compiler, accessing an Oracle 8i database manager. The three main phases of its development were the definition of *how to prepare* the images to be stored, the definition of *how to store* the images in the database, and the *implementation* of the software modules that define the *SRIS-HC* architecture.

The first phase in the development of the *SRIS-HC* started with the analysis of the images gathered by the imaging equipment of the hospital. They are generated in the DICOM 3.0 format, the standard format used nowadays to obtain, store and distribute medical images, by many modalities of equipments, such as X-rays, CT and MR tomographies and ultrasound. DICOM files are composed by one image and tags describing the image and the process used to obtain it. Tags are textual or numerical sequences of <attribute, value> pairs. The set of tags enables images to be self-contained, being autonomous objects, that can keep together the data describing the image, like where, when and how it was obtained, as well as some data regarding the exam and patient. This monolithic structure provides a strong support to maintain the integrity of the information as the images are transferred from the image acquiring device to the storage media and to the many analysis and viewing stations.

However, there are two problems when DICOM files are used in an HIS/RIS environment with support to image similarity retrieval capabilities. The first problem originates from the fact that the textual/numerical information contained in a PER system is stored separated from the images, and therefore it must exist a way to link each image to its corresponding record in the PER system. The second problem originates from the fact that a similarity search operation needs to scan the full set of images to answer any given query.

The first problem must be addressed to automate the process of storing images in databases as they are being generated by the medical exam device, to identifying the images in a unique and understandable way, so that the link with the corresponding medical record can be maintained even if the image is transferred to viewing workstations detached from the PER system. To solve this problem, a protocol was developed to be followed by the medical staff of the hospital when a new image is acquired, corresponding to the definition of the tag [(0008,0050) - *Accession N umber*] of the DICOM format, that is made available in the images generated by every equipment of the hospital. Although not mandatory by the DICOM specification, this tag can be (and was) used to associate a unique identification to each generated image. A small application software, running across the whole hospital information system generates this unique identification. As the acquisition software of the imaging equipment does not enforce the definition of this tag, the medical staff was trained to define it whenever a new image is generated. This operating protocol constitutes the first phase of the development of the system: how to prepare the images to be stored.

The second problem was targeted in the second phase of the development of the *SRIS-HC*: the definition of how to store the images in the database. As the similarity search operations need to scan every image in the dataset, storing each image in a separated file turn out to be a rather slow operation. Therefore, we opted to store each image inside the database, speeding up the whole process. This procedure also assures a stronger consistence in the data, as the images are hold by a relational database management system, and common problems like moving the images to different directories or renaming the files are not allowed.

However, the database management systems (DBMS) available from-the-shelf store and retrieve images based on their numeric and textual description. Images are seen as large binary objects (blobs), and no image processing or image comparisons inside the databases are allowed. Therefore, it is not possible to compare two images by their content using any commercial DBMS. However, it is possible to integrate new facilities over DBMS to achieve this intent. That is, the DBMS can be expanded with new software modules to perform content-based image retrieval and allow answering similarity queries over medical images. Two types of similarity queries must be supported by a SRIS subsystem: range queries, which retrieves every image that is similar to a given "image reference" up to a given similarity degree; and *k*-nearest neighbors, that retrieves the *k* images most similar to the given "image reference." Notice that every similarity query require a reference image as part of it.

A software module was developed to act as a "DICOM handler" between the images stored in DICOM files and the database. This handler opens a DICOM file already linked with a PER, and extracts the image and every relevant tags in the images (such as the equipment that produced the image, the time of the scanning, among others), storing them as a set of attributes, following a predefined structure, that can be imported into the database. The image is stored in an image format file as an attribute of type "blob". Two formats are currently supported: jpeg and a proprietary format, used in a local image processing system. The image and the attributes are then imported into the database. It is required that each image already have the (0008,0050) tag defined. Thus the images can be linked to the other patient information already stored in the database of the HIS/RIS. The handler can also perform the opposite operation, exporting an image already stored inside the database and its attributes, reconstructing the DICOM file, including the original tags stored in the database, and also tags that describe the patient, obtained from the PER of the patient to whom the image is linked to.

The third phase in the development of the *SRIS-HC* was the implementation of the software modules that define the *SRIS-HC* architecture. This architecture is described in the next section.

3. The system architecture

The general architecture of the SRIS system includes six modules: (a) Image storing, (b) Feature extraction, (c) Indexing, (d) Similarity query answering, (e) Image and exams visualization, and (f) Statistics management. There are three main points of internal interaction: (1) through the database, where every data in the system is maintained; (2) through the application programing interface (API) of the similarity query answering module; and (3) through the API of the visualization module. Figure 1 shows a schematic representation of the architecture of the SRIS-HC, and how it interacts with the existing HIS/RIS system.

The data in the SRIS-HC architecture flows as will be detailed next. New images are submitted to the image storing module, through the DICOM handler. This module stores every new image in the database, linking it to the corresponding exam data, which must have been previously stored in the database by the existing HIS/RIS system. For each new image, the image storing module makes a request for the feature extraction module to obtain the image feature vector. The feature vector is then sent to the indexing module, which links the features with the image stored in the database, and includes the features in the indexing structure. When the images are already stored, the system can

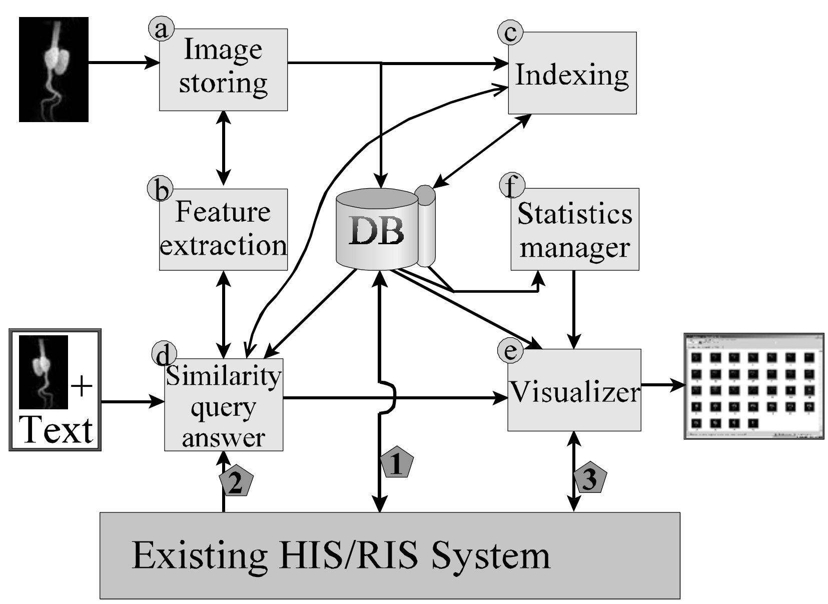

Figure 1. Architecture of the SRIS-HC sub-system.

answer similarity queries, which are posed through the similarity query answer module. Whenever a similarity query is received, this module calls the feature extraction module to obtain the feature vector of the query image, and requests the indexing module to find the images in the database that are the answer for the posed query. It then calls the visualizer to display the resulting images. The visualizer can also prepare the attributes of a image, so the image and its attributes can be submitted to the DICOM handler, to be distributed to every existing workstation already prepared to support this kind of data.

When an existing HIS sub-system is updated to support images, it is extend to interface with the similarity query and visualizer modules. The extensions are intended to enable the visualization of images the systems now recognize as part of the data they handle, and to enable the submission of similarity queries from inside these sub-systems. The existing HIS/RIS access to the database without modification, as the same existing data structures are maintained in the database. The new structures are dealt only with the SRIS modules, so there are no conflicts that may require implementation modifications in the existing sub-systems. However, the visualizer, statistics manager and similarity query modules can alter the data in the portion of the database that is accessed by the HIS/RIS sub-systems, altering the data they display in return from queries they pose.

In the current version, the features used by the SRIS-HC to compare images are the traditional color histogram (considering only gray levels) and the metric histogram [14]. To enable an efficient and fast retrieval of images, the SRIS-HC uses the Slim-tree [15], a fast metric access method to index very large sets of image of the exams stored at the system. The system was developed in C++, using an Oracle 8i database server. Figure 2 shows two screen shots of the SRIS-HC, The first one shows the preparation of a nearest-neighbors query of images similar to the image shown, and the second screen shot shows the answer of this query.

4. Conclusions

The present stage in the support of images in existing HIS/RIS systems enables just the storage of images, and

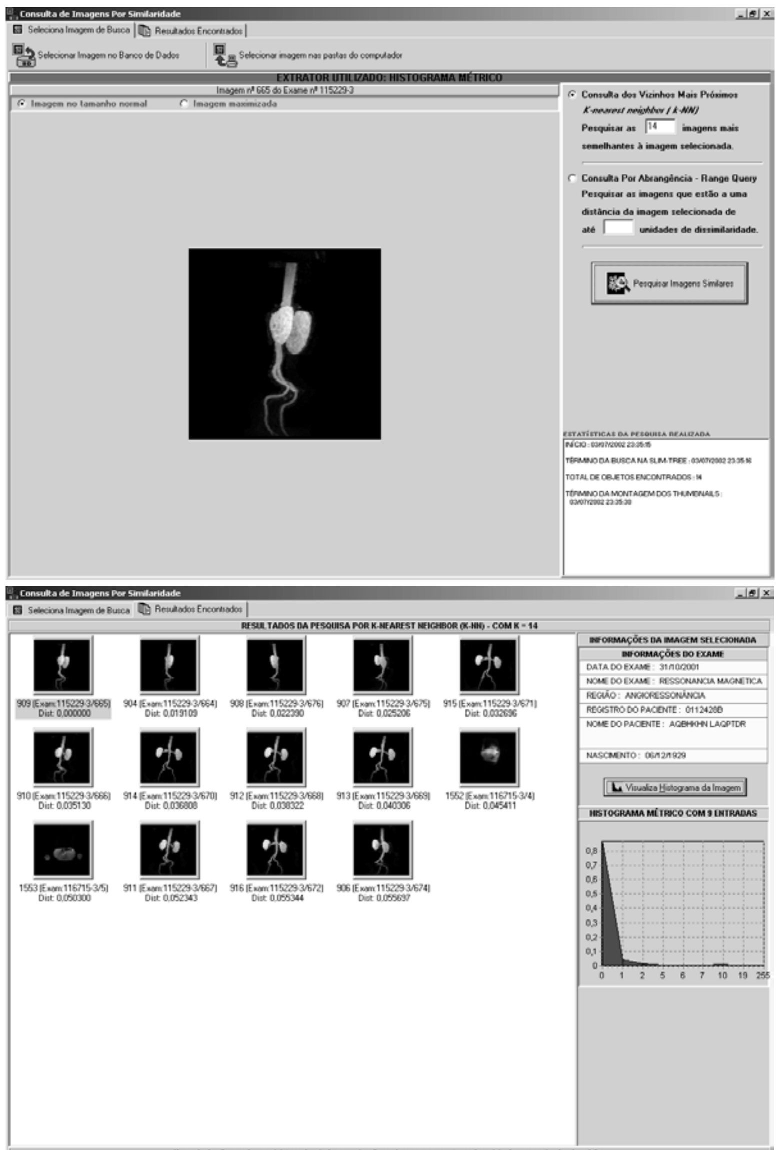

Figure 2. Screen shots of the SRIS-HC module. (a) Query preparation. (b) Query answer.

their retrieval through linked data in textual/numerical format. The next step is the retrieval of images based on their contents. This paper describes the prototype of a system that can perform the retrieval of images through similarity queries based on their contents. This ability enables a better usage of digitalized information to aid in health procedures, as searching for similar medical cases to compare/study/diagnose, perform statistical studies, cross-referencing, bibliographic reports and documentation. This system is running in the university hospital, and the results are very promising.

As a further development, the support to similarity queries can be extended to handle other types of data, and not only images, for example, combining textual/numerical data with images, or even just textual/numerical.

Acknowledgments

This research was supported, in part, by the Sao Paulo State Research Foundation (FAPESP) under grant 02/07318-1, and by the Brazilian National Research Council (CNPq) under grants 52.1685/98-6, 860.068/00-7 and 35.0852/94-4.

5. Bibliography

[1] X. Cao and H. K. Huang, "Current Status and Future Advances of Digital Radiography and PACS," *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, pp. 80-88, 2000.

[2] C.-S. Lee, H.-J. Tschai, Y.-H. Kuo, W.-T. Ko, Y.-H. Cheng, and C.-C. Wang, "PACS: Construction and Application to Medical Image Enhancement," presented at Third Intl. Conference on Knowledge-Based Intelligent Information Engineering Systems, pp. 246 -249, 1999.

[3] E. L. Siegel and R. M. Kolodner, *Filmless Radiology*. New York City, NY: Springer Verlag, 1999.

[4] M. Krol, H. Fong, M. Gauzza, and D. Reich, "Linking Clinical, Research and Administrative Computer Systems," presented at 13th IEEE Symposium on Computer-Based Medical Systems (CBMS'00), Houston, TX, pp. 287-292, 2000.

[5] S. Khludov, L. Vorwerk, and C. Meinel, "Internet-orientated Medical Information System for DICOM-data Transfer, Visualization and Revision," presented at 13th IEEE Symposium on Computer-Based Medical Systems (CBMS'2000), Houston, TX, pp. 293 -296, 2000.

[6] W. Cai, D. Feng, and R. Fulton, "Web-Based Digital Medical Images," *IEEE Computer Graphics and Applications*, vol. 21, pp. 44-47, 2001.

[7] K. Adelhard, S. Nissen-Meyer, C. Pistitsch, U. Fink, and M. Reiser, "Functional Requirements for a HIS-RIS-PACS Interface Design Including Integration of "Old" Modalities," *Methods of Information in Medicine*, vol. 38, pp. 1-8, 1999.

[8] T. Zrimec, "A Medical Image Information System," presented at VISIM Workshop: Information Retrieval and Exploration in Large Collections of Medical Images, Utrecht, Netherland, pp. 5, 2001.

[9] T. S. Yoo, "Toward Validation Databases for Medical Imaging: Engineering a Scientific Rendezvous," presented at VISIM Workshop: Information Retrieval and Exploration in Large Collections of Medical Images, Utrecht, Netherland, pp. 4, 2001.

[10] T. M. Lehmann, B. Wein, D. Keysers, J. Bredno, M. O. Guld, H. Schubert, and M. Kohnen, "Image Retrieval in Medical Ap-plications: the IRMA-Approach," presented at VISIM Workshop: Information Retrieval and Exploration in Large Collections of Medical Images, Utrecht, Netherland, pp. 4, 2001.

[11] L. Vorwerk and C. Meinel, "A Multimedia-Editor for making Findings in Radiology," presented at 13th IEEE Symposium on Computer-Based Medical Systems (CBMS'2000), Houston, TX, pp. 297-302, 2000.

[12] J. S. Ducan and N. Ayache, "Medical Image Analysis: Progress over Two Decades and the Challenges Ahead," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 85-106, 2000.

[13] R. Haux, E. Ammenwerth, W. Herzog, and P. Knaup, "Health care in the information society. A prognosis for the year 2013," *International Journal of Medical Informatics*, vol. 66, pp. 3-21, 2002.

[14] A. J. M. Traina, C. Traina, Jr., J. M. Bueno, F. J. T. Chino, and P. M. d. Azevedo-Marques, "Efficient Content-based Image Retrieval through Metric Histograms," *World Wide Web Journal, Kluwer Academic Publishing Co.*, vol. to appear, pp. 29 pags, 2003.

[15] C. Traina, Jr., A. J. M. Traina, C. Faloutsos, and B. Seeger, "Fast Indexing and Visualization of Metric Datasets Using Slim-trees," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, pp. 244-260, 2002.