

Bridging the integration gap between imaging and information systems: a uniform data concept for content-based image retrieval in computer-aided diagnosis

Petra Welter,¹ Jörg Riesmeier,² Benedikt Fischer,¹ Christoph Grouls,³ Christiane Kuhl,³ Thomas M Deserno (né Lehmann)¹

► An additional appendix is published online only. To view this file please visit the journal online (www.jamia.org).

¹Department of Medical Informatics, RWTH Aachen University of Technology, Aachen, Germany

²ICSMED AG, Oldenburg, Germany

³Department of Diagnostic and Interventional Radiology, RWTH Aachen University Hospital, Aachen, Germany

Correspondence to

Petra Welter, Department of Medical Informatics, RWTH Aachen University, Pauwelsstraße 30, D-52074 Aachen, Germany; pwelter@mi.rwth-aachen.de

Received 25 August 2010

Accepted 30 April 2011

ABSTRACT

It is widely accepted that content-based image retrieval (CBIR) can be extremely useful for computer-aided diagnosis (CAD). However, CBIR has not been established in clinical practice yet. As a widely unattended gap of integration, a unified data concept for CBIR-based CAD results and reporting is lacking. Picture archiving and communication systems and the workflow of radiologists must be considered for successful data integration to be achieved. We suggest that CBIR systems applied to CAD should integrate their results in a picture archiving and communication systems environment such as Digital Imaging and Communications in Medicine (DICOM) structured reporting documents. A sample DICOM structured reporting template adaptable to CBIR and an appropriate integration scheme is presented. The proposed CBIR data concept may foster the promulgation of CBIR systems in clinical environments and, thereby, improve the diagnostic process.

INTRODUCTION

Modern data processing technologies facilitate the convenient transfer and storage of digital imaging data. The challenge today is to successfully retrieve the relevant details from the enormous amount of existing data. In clinical practice, archives hosting images from medical examinations are usually based on the Digital Imaging and Communications in Medicine (DICOM) standard.^{1,2} Traditionally, alphanumeric attributes (eg, the patient's name, examined body part, or imaging modality) are used to query images within the DICOM database. Such image annotations are based frequently on manually entered information, a method that has several disadvantages, including possible misspellings and insufficient description of the image content, leading to substantial costs due to failed queries.

Content-based image retrieval (CBIR) identifies images by their content, utilizing visual information for retrieval.^{3,4} This method has led to increasing interest in medical applications of CBIR techniques.⁴⁻⁶ In the context of computer-aided diagnosis (CAD), the typical application of CBIR supports differential diagnosis, that is, distinction between two or more diseases by systematic comparison to known categories in pathology. The aim of CBIR is to provide the radiologist with an easily accessible diagnostic aid. Relevant past cases are automatically identified along with their

diagnosis and other suitable information, providing the radiologist with additional evidence for the diagnostic assessment.

CBIR has the potential to improve the quality and efficiency of clinical care processes.⁵ Many CBIR prototypes have demonstrated the manifold benefits of CBIR for medical systems.^{6,7} It is expected that CBIR will gain even more acceptance when clinical CBIR applications become part of the radiologist's everyday routine.⁵ However, CBIR systems commonly fail to bridge the gap between imaging and medical information systems. Hence, integration into clinical practice remains hampered.⁸ From our perspective, one reason for this deficiency is a missing data concept for CBIR results that supports integration into the radiologist's workflow with picture archiving and communication systems (PACS).

DRAWBACKS OF CURRENT CBIR RESULTS MANAGEMENT

Currently, most CBIR systems do not integrate with PACS or any other information system used in the intended domain of hospital information systems (HIS).⁹⁻¹¹ CBIR is operated exclusively as a dedicated stand-alone system or as a component incorporated into PACS.¹²

Stand-alone workstations cannot exchange information with other components of HIS and radiology information systems. For instance, the IDEM system¹³ runs on a workstation that retrieves images similar to those pre-selected, which are imported from the PACS. Manually labeled images can be exported to PACS in DICOM format. The ASSERT system¹⁴ uses a physician-in-the-loop approach to obtain images of a pre-configured database. In addition, medGIFT/GIFT¹⁵ is an open-source framework for content-based searching in medical image databases that require application-specific implementation.

PACS components limit the radiologist's options to particular software add-ons provided by a specific system's manufacturer. For instance, the cbPACS¹⁶ embeds CBIR functionalities into the relational database management of PACS. This process is undertaken by providing an interpreter for SQL commands that is extended by CBIR. The COBRA model¹⁷ is an architecture that extends arbitrary PACS with CBIR capabilities. Grid-solutions for the storage of CBIR data have also been proposed.¹⁸

However, none of these CBIR systems include a concept for storing CBIR results. Thus, evidence cannot be recorded, hindering the reproduction of diagnoses based on CBIR results as well as the discussion of critical findings with other physicians.

MODEL DESCRIPTION

To identify the information required to support diagnoses, we applied CBIR to CAD, including input and output data as well as all information required for interpreting and reproducing the CBIR response. The following items should be included in a CBIR result document:

- ▶ CBIR system: In environments with different, exchanged, or updated CBIR systems, information about the used CBIR system and its parameterization is essential to the precise understanding of the query response.
- ▶ Image or region of interest (ROI) of query: According to the query by example paradigm, the input to the CBIR system may be an examination image or an ROI (ie, used to focus on a specific area or pattern) that is predefined by the radiologist.
- ▶ Reference database: Every new examination allows the feature database to grow, leading to varying CBIR results based on when individual queries are processed. Reproducibility is established by logging the exact date and time of every reference set created in the system. Furthermore, any restrictions to the reference database require monitoring. For example, a CBIR request on a mammographic pattern is usually compared to other mammograms only.
- ▶ Result images: The report should include all the information that directly addresses the corresponding DICOM images in the PACS (ie, Study, Series, and Simple Object Pair (SOP) Instance UID).
- ▶ Similarity scores: In addition to the image pattern, the similarity score computed by the CBIR system is of particular interest to the radiologist.

We suggest using the DICOM standard for reliable data integration (ie, DICOM SR and DICOM SR templates) (see online supplementary appendix). The DICOM SR templates have been developed and approved by experts from a variety of professional backgrounds, which enhances their value. Approved SR sub-trees are reused in the CBIR domain. In particular, our templates are derived from the corresponding mammography CAD SR (TID 4000) and the chest CAD SR (TID 4100) templates due to their direct relation to CBIR.

Five templates (figure 1)¹⁹ are combined as an integration scheme for CBIR response data. The main or root template

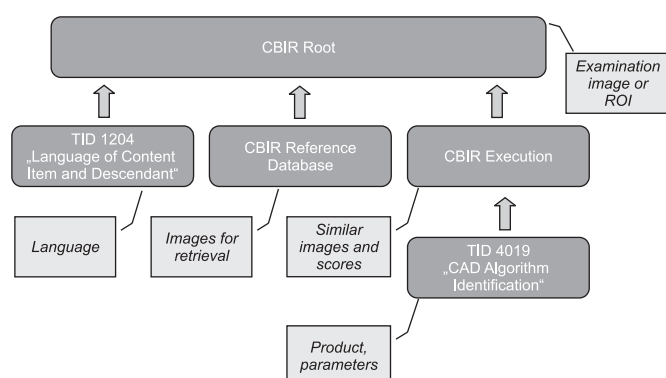


Figure 1 Composition of the content-based image retrieval Structured Reporting template.

‘CBIR Root’ contains the query image or the query ROI. If an ROI has been specified, the source image is listed. ‘CBIR Root’ includes the standard template TID 1204 ‘Language of Content Item and Descendant’ to specify the selected language. ‘CBIR Root’ further includes two sub-templates:

- ▶ ‘CBIR Reference Database’ holds a concrete list of image IDs or the abstract search criteria defining a special view within the image database. Search criteria are given as DICOM attribute tags, divided into group and element number. Alternatively, a search string may be specified. For example, to restrict the view on data on male smokers, an entry for patient gender with DICOM group number ‘0010’ and element number ‘0040’ is denoted with the value ‘M.’ Smoking status is given by a second entry holding the group and element numbers ‘0010’ and ‘21A0,’ respectively, and the value ‘YES.’
- ▶ ‘CBIR Execution’ includes the standard template TID 4019 ‘Algorithm Identification,’ which specifies the CBIR algorithm with the particular product version and parameters. ‘CBIR Execution’ further contains the retrieved images by their DICOM SOP Class UID and their similarity scores.

EXEMPLARY IMPLEMENTATION

Based on our proposal, a working system for CBIR and PACS integration has been established. The system was implemented in C++ using the DICOM toolkit ‘DCMTK’,²⁰ with the DCMTK module ‘dcmshr’ offering methods for accessing, reading, creating, writing, modifying, printing, and viewing DICOM SR documents. To validate system capabilities, we installed a test bed and implemented the core parts of the system:

- ▶ CBIR adapter: This module was implemented with complete functionality. The CBIR adapter²¹ is enabled to receive the CBIR request, to exchange data with PACS, to operate the CBIR system, and to create a CBIR SR document based on the proposed CBIR SR template.
- ▶ CBIR system: The test bed is based on the Image Retrieval in Medical Applications framework.²² Image Retrieval in Medical Applications provides both an invocation by system command and a Simple Object Access Protocol interface for the exchange of parameters and results.²¹
- ▶ PACS: We limited our PACS to the DICOM Application Entity accomplished by DCMTK²⁰; this restriction provides sufficient functionality for retrieving and storing DICOM objects.
- ▶ CBIR trigger: A script transferring DICOM information on the query image to the CBIR adapter initiates the CBIR process.
- ▶ SR viewer: The IRMACON SR viewer was developed²³ to help overcome the general DICOM SR viewer’s limited layout capabilities.²³

These components can be used in any PACS environment, and only the plug-ins for triggering CBIR and viewing SR documents are vendor-specific and dependent on the actual clinical setting. The integration scheme and workflow, including all communication steps, are depicted in figure 2 and box 1, respectively.

USER INTERFACE

The IRMACON SR viewer combines hanging protocols (ie, a consistent arrangement of components on the screen, typically applied for examination images in radiology) with convenient access to meta-information that is retrieved automatically from

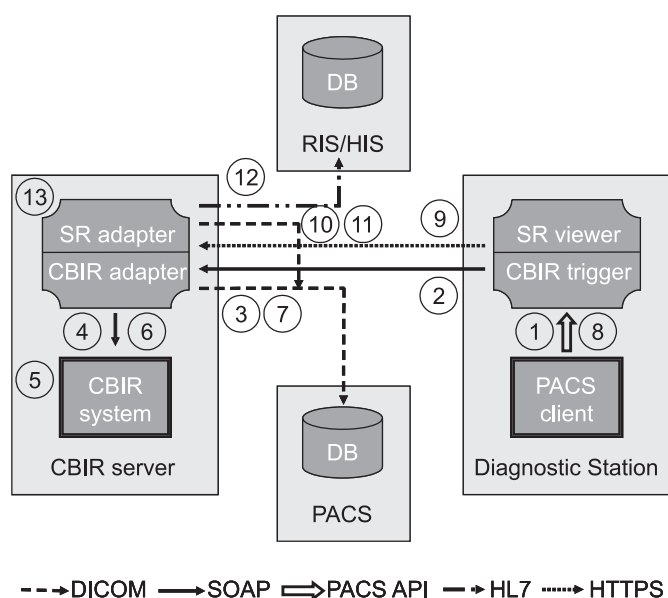


Figure 2 System overview of CBIR integration with PACS/RIS/HIS components (arrows point from the initiator of a step towards the destination). Communication is realized by DICOM network protocol, Health Level 7 (HL7) messages, Simple Object Access Protocol (SOAP), HTTPS, and by utilizing the API of the PACS vendor. PACS, CBIR adapter, and SR adapter are DICOM Application Entities and, therefore, enabled to communicate over the DICOM network protocol.

the radiology information systems. The starting page shows the query and the CBIR response in descending order of similarity (figure 3, top). Detailed diagnostic information on the particular

examination is obtained by double-clicking the displayed image (figure 3, bottom). Available diagnostic reports are listed next to the response image. Selecting a report (double click) launches a specific application (eg, a PDF viewer) and displays the report's content.

EVALUATION

Based on the exemplary implementation and the presented example, the requirements of the proposed CBIR SR document are then verified:

- ▶ **CBIR system:** Relevant facts on the CBIR algorithm, its parameters, and the database configuration are available to the radiologist. By employing the CBIR response, the radiologist can use this information to make a diagnosis.
- ▶ **Image or ROI of query:** The SR document contains all information required to identify the particular query image or ROI, which was input for the CBIR request.
- ▶ **Reference database:** The timestamp specifies the database searched. As CBIR systems generally do not hold meta-information such as date of examination, patient gender, etc, additional search criteria are used to filter the response following CBIR execution. This step can be time consuming as each corresponding DICOM header must be retrieved from PACS and processed.
- ▶ **Result images:** The CBIR SR information is sufficient to identify and address the corresponding examinations in the PACS.
- ▶ **Similarity scores:** The SR document refers to the result images in descending order according to their similarity score, providing the radiologist with a quick overview of image relevance.

Box 1 Workflow and communication steps in content-based image retrieval (CBIR) and picture archiving and communication systems (PACS) integration

- Step 1: The radiologist at the PACS client initiates the plug-in 'CBIR trigger,' which is based on the application programming interface (API) of the PACS, and selects the query image or ROI.
- Step 2: The CBIR trigger initializes a connection to the CBIR adapter. A Simple Object Access Protocol (SOAP) message is sent; this message contains all information relevant for CBIR, the DICOM Study, Series, and SOP Instance UID of the query image, and if applicable, the ROI annotated by the physician.
- Step 3: The CBIR adapter checks if the query image has already been processed by the CBIR system. An internal mapping database records all processed images on both PACS and CBIR. If the image is new to the CBIR system, the CBIR adapter retrieves the DICOM file from PACS via DICOM C-MOVE and continues with step 4, else with step 6.
- Step 4: The DICOM file is converted into a CBIR-applicable graphics format by the CBIR adapter. It is then passed to the CBIR system via a SOAP message.
- Step 5: The CBIR system extracts features from the image and stores the data in its internal database, creates a unique identifier for the newly processed image, and returns the image to the CBIR adapter.
- Step 6: The CBIR adapter sends to the CBIR system a SOAP message that includes the CBIR identifier of the query image. The CBIR system subsequently executes the retrieval of similar images found in its reference database and returns them together with their similarity scores.
- Step 7: The CBIR adapter creates a CBIR SR document comprising all references to the corresponding DICOM UIDs of the CBIR response images. The CBIR SR document is stored in the PACS by DICOM C-STORE.
- Step 8: The display on the PACS client is updated. The CBIR SR document is listed now, and the radiologist activates the DICOM SR viewer plug-in.
- Step 9: The SR viewer calls an HTML page from the SR adapter on the CBIR server. The required context information is provided in the URL.
- Step 10: The SR adapter retrieves the DICOM SR document from PACS and converts it into XML for parsing.
- Step 11: The SR adapter retrieves all DICOM images that are listed in the SR document and converts them into a suitable graphics format for presentation in a web browser.
- Step 12: The SR adapter sends a Health Level 7 (HL7) message to the RIS or the HIS, and diagnostic findings of examinations referenced in the SR document are retrieved.
- Step 13: The SR adapter generates HTML output containing the collected information.

Figure 3 IRMA CBIR results presented by the IRMACON viewer, entry page above (A) and detail page below (B) together with an exemplary diagnostic document.



DISCUSSION

The potential impact of CBIR on CAD, evidence-based medicine, or case-based reasoning has been described frequently in the literature, but CBIR has not yet been widely employed in clinical practice. Here, we propose the introduction of a DICOM SR document to interface CBIR with PACS. In general, we assume that future CAD systems will rely on DICOM SR.²⁴ Integrating the Healthcare Enterprise²⁵ is a global initiative by healthcare professionals aimed at addressing interoperability challenges associated with healthcare systems and information resources.

Healthcare Enterprise uses DICOM SR extensively for document encoding,²⁶ which shows its high relevance and acceptance in healthcare systems. In general, CAD and PACS integration tends to incorporate CAD results as DICOM SR documents.^{12 27-29}

Recently, DICOM has been extended by Supplement 118, Application Hosting,³⁰ which proposes an open, standardized application programming interface for a DICOM-based image processing and analysis system; the aim of this system is to develop portable applications that are capable of plugging into any host. Application Hosting simplifies the CBIR adapter and

supersedes any application programming interface-based plug-in implementation for the PACS. The convenient coupling of such external CAD systems is certain to foster further integration of CBIR with PACS. However, the proposed concept to manage CBIR responses will still be required for adequate bridging of the gap between imaging and information systems.

While this paper addresses the data integration problem of CBIR results, two general paradigms of CBIR for CAD must also be mentioned. (i) The diversity of symptoms known to the reference database has a crucial influence on the retrieval quality, and hence, high-quality CBIR-CAD requires a significant number of already processed PACS examinations, which are not yet considered in the test bed. (ii) At the moment, the performance of CBIR systems do not fulfill the general requirements of radiological routine. Existing solutions are mostly specialized on certain diseases and implemented proprietarily. However, a consistent and reliable data integration scheme supports the further development of CBIR in CAD.

CONCLUSIONS

A gap in CBIR-CAD integration in clinical practice was identified and a possible solution suggested. To obtain the predicted benefits for CAD, CBIR results must be encoded in a standardized format and integrated into PACS and HIS to meet radiologists' workflow requirements. These steps are essential for bridging the gap between imaging and medical systems. As long as physicians are required to incorporate individual CBIR results manually into the diagnostic findings, CBIR will not become part of clinical practice. The proposed system architecture using DICOM SR documents may bridge this gap. This proposed technology follows the general paradigm of healthcare information systems, to provide 'the right information, in the right place, at the right time',^{31 32} and it merges imaging and information systems in healthcare.

Funding This research was partly funded by the German Research Foundation (DFG), Le 1108/9.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **National Electrical Manufacturers Association.** *The DICOM Standard. PS 3.1-PS 3.18.* 2009. <http://medical.nema.org/medical/dicom/2009> (accessed 10 Apr 2011).
2. **Bidgood WD Jr, Horii SC, Prior FW, et al.** Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc* 1997;**4**:199–212.
3. **Smeulders AWM, Worring M, Santini S, et al.** Content-based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell* 2000;**22**:1349–80.
4. **Eakins JP, Graham ME.** *Content-Based Image Retrieval - A Report to the JISC Technology Applications Program. Technical Report, Institute for Image Data Research.* Newcastle: University of Northumbria, 1999. http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc (accessed 10 Apr 2011).
5. **Müller H, Rosset A, Garcia A, et al.** Informatics in radiology (infoRAD): benefits of content-based visual data access in radiology. *Radiographics* 2005;**25**:849–58.
6. **Müller H, Michoux N, Bandon D, et al.** A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int J Med Inform* 2004;**73**:1–23.
7. **Deselaers T, Deserno TM, Müller H.** Automatic medical image annotation in ImageCLEF 2007: overview, results, and discussion. *Pattern Recognit Lett* 2008;**29**:1988–95.
8. **Deserno TM, Antani S, Long R.** Ontology of gaps in content-based image retrieval. *J Digit Imaging* 2009;**22**:202–15.
9. **Mattie ME, Staib L, Stratmann E, et al.** PathMaster: content-based cell image retrieval using automated feature extraction. *J Am Med Inform Assoc* 2000;**7**:404–15.
10. **Ribeiro MX, Bugatti PH, Traina C Jr, et al.** Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques. *Data Knowl Eng* 2009;**68**:1370–82.
11. **Wei L, Yang Y, Nishikawa RM.** Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognit* 2009;**42**:1126–32.
12. **Huang H.** Utilization of medical imaging informatics and biometrics technologies in healthcare delivery. *Int J Comput Assist Radiol Surg* 2008;**3**:27–39.
13. **Le Bozec C, Zapletal E, Jaulet MC, et al.** Towards content-based image retrieval in a HIS-integrated PACS. *Proc AMIA Symp* 2000:477–81.
14. **Aisen AM, Broderick LS, Winer-Muram H, et al.** Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 2003;**228**:265–70.
15. **Müller H, Rosset A, Vallee JP, et al.** Integrating content-based visual access methods into a medical case database. *Stud Health Technol Inform* 2003;**95**:480–5.
16. **Traina C Jr, Traina AJ, Araújo MR, et al.** Using an image-extended relational database to support content-based image retrieval in a PACS. *Comput Methods Programs Biomed* 2005;**80**(Suppl 1):S71–83.
17. **El-Kwae EA, Xu H, Kabuka MR.** Content-based retrieval in picture archiving and communication systems. *J Digit Imaging* 2000;**13**:70–81.
18. **Oliveira MC, Cirne W, De Azevedo Marques PM.** Towards applying content-based image retrieval in the clinical routine. *Future Generat Comput Syst* 2007;**23**:466–74.
19. **Welter P, Deserno TM, Gülpers R, et al.** Exemplary design of a DICOM structured report template for CBIR integration into radiological routine. *Procs SPIE* 2010;**7628**:0B1–10.
20. **OFFIS.** *DICOM Toolkit DCMTK.* <http://dicom.offis.de/dcmk.php.en> (accessed 10 Apr 2011).
21. **Welter P, Hocken C, Deserno TM, et al.** Workflow management of content-based image retrieval for CAD support in PACS environments based on IHE. *Int J Comput Assist Radiol Surg* 2010;**5**:393–400.
22. **Göld MO, Thies C, Fischer B, et al.** A generic concept for the implementation of medical image retrieval systems. *Int J Med Inform* 2007;**76**:252–9.
23. **Welter P, Topal F, Jansen S, et al.** Combined DICOM and HL7 viewer in support of a bridge from content-based image retrieval to computer-assisted diagnosis. In: Pietka E, Kawa J, eds. *Information Technologies in Biomedicine.* Berlin: Springer, 2010:145–52. (Advances in Soft Computing; vol 69).
24. **Oosterwijk H.** *Structured Reporting, a Powerful DICOM Mechanism.* http://www.ringholm.de/docs/02030_en.htm (accessed 10 Apr 2011).
25. **Integrating the Healthcare Enterprise.** *IHE Technical Framework, vol. I: Integration Profiles.* 2008. http://www.ihe.net/Technical_Framework/upload/ihe_tf_rev9-0ft_vol1_2008-06-27.pdf (accessed 10 Apr 2011).
26. **Hussein R, Engelmann U, Schroeter A, et al.** DICOM structured reporting: Part 1. Overview and characteristics. *Radiographics* 2004;**24**:891–6.
27. **Le AH, Liu B, Huang HK.** Integration of computer-assisted diagnosis/detection (CAD) results in a PACS environment using CAD—PACS toolkit and DICOM SR. *Int J Comput Assist Radiol Surg* 2009;**4**:317–29.
28. **Philips Medical Systems.** *Integration of Digital Mammography on PACS.* 2007. http://www.healthcare.philips.com/phpwc/main/shared/assets/documents/about/events/2009/ecr/pdf/hi/integration_of_digital_mammography_on_pacs.pdf (accessed 10 Apr 2011).
29. **Carestream Health, Inc.** *Make Mammography Work Flow.* 2010. http://www.carestreamhealth.com/MI-IT_PACS_for_Mammography.pdf (accessed 10 Apr 2011).
30. **National Electrical Manufacturers Association.** *DICOM Supplement 118: Application Hosting.* 2010. http://medical.nema.org/medical/dicom/final/sup118_ft2.pdf (accessed 10 Apr 2011).
31. **Reichert PL.** Towards systematization. *Methods Inf Med* 1977;**16**:125–30.
32. **Winter A, Lagemann A, Budig B, et al.** Integration of health professional workstation in hospital information systems. In: Brender J, Christensen JP, Scherrer J-R, et al, eds. *Medical Informatics Europe '96.* Amsterdam: IOS Press, 1996:685–89.