

Ontology of Gaps in Content-Based Image Retrieval

Thomas M. Deserno,^{1,2} Sameer Antani,² and Rodney Long²

Content-based image retrieval (CBIR) is a promising technology to enrich the core functionality of picture archiving and communication systems (PACS). CBIR has a potential for making a strong impact in diagnostics, research, and education. Research as reported in the scientific literature, however, has not made significant inroads as medical CBIR applications incorporated into routine clinical medicine or medical research. The cause is often attributed (without supporting analysis) to the inability of these applications in overcoming the “semantic gap.” The semantic gap divides the high-level scene understanding and interpretation available with human cognitive capabilities from the low-level pixel analysis of computers, based on mathematical processing and artificial intelligence methods. In this paper, we suggest a more systematic and comprehensive view of the concept of “gaps” in medical CBIR research. In particular, we define an ontology of 14 gaps that addresses the image content and features, as well as system performance and usability. In addition to these gaps, we identify seven system characteristics that impact CBIR applicability and performance. The framework we have created can be used a posteriori to compare medical CBIR systems and approaches for specific biomedical image domains and goals and a priori during the design phase of a medical CBIR application, as the systematic analysis of gaps provides detailed insight in system comparison and helps to direct future research.

KEY WORDS: Content-based image retrieval (CBIR), pattern recognition, picture archiving and communication systems (PACS), information system integration, data mining, information retrieval, semantic gap

INTRODUCTION

Content-based image retrieval (CBIR) is a technology that accesses pictures by image patterns rather than by alphanumeric-based indices.¹ Using various visual query mechanisms, such as the query-by-example (QBE) paradigm,² the user presents a sample image, image region of interest (ROI), or pattern to the system, which responds with images

similar to the given example. Although this approach was originally developed for multimedia repositories such as those on the World Wide Web, techniques for content-based access to medical image repositories are a subject of high interest in recent research, and remarkable efforts have been reported.^{3–5} In particular, CBIR for picture archiving and communication systems (PACS) discussed in Qi and Snyder,⁶ Müller et al.,⁷ and Lehmann et al.⁸ can make a significant impact on health informatics and health care. In spite of these innovations, however, routine use of CBIR in PACS has not yet been established. The reasons are manifold; some of these obstacles, as well as extensive and detailed discussion of many characteristics of CBIR systems, can be found in the papers of Müller et al.,^{5,7,9}. Our goal is to organize these characteristics in a more formal and systematic approach. We seek to provide an organizational and conceptual framework for analysis of a particular CBIR system or for comparative analysis among CBIR systems, with respect to critical technical or implementation issues. We have organized the framework by generalizing and extending a concept that already has a somewhat restricted use in CBIR research, namely, the concept of a *gap*: a disparity, break, or dis-

¹From the Department of Medical Informatics, Aachen University of Technology (RWTH), Pauwelsstr. 30, 52057 Aachen, Germany.

²From the US National Library of Medicine, US National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA.

Correspondence to: Thomas M. Deserno, Department of Medical Informatics, Aachen University of Technology (RWTH), Pauwelsstr. 30, 52057 Aachen, Germany; tel: +49-241-8088793; fax: +49-241-803388793; e-mail: deserno@ieee.org

Copyright © 2007 by Society for Imaging Informatics in Medicine

Online publication 1 February 2008

doi: 10.1007/s10278-007-9092-x

continuity in some important aspect or characteristic between the potential and the actual realization of that characteristic. We attempt to improve on the published conceptual approaches to thinking comparatively about CBIR systems by:

- Using as one of our main organizing principles a concept (the gap) that highlights both potential deficiencies and how those deficiencies can be addressed
- Presenting both the gaps that we have identified, as well as important system characteristics that are not exposed by the gap ontology, as a hierarchical structure of related attributes rather than as a purely descriptive exposition

Searching for visual similarity by simply comparing large sets of pixels (comparing a query image to images in the database, for example) is not only computationally expensive but is also very sensitive to and often adversely impacted by noise in the image or changes in views of the imaged content. Therefore, to achieve rapid response and to ameliorate the sensitivity to image noise or view changes in position, orientation, and scale of the imaged content, frequent data reduction is carried out as follows: First, discriminant numerical features that serve as identifying signatures are extracted from each image in the repository. Second, the images are indexed on these precomputed signatures. Third, at query time, the signature extracted from the query example is compared with these indices of the images in the database (in this paper, we use the term *signature* to denote the (usually ordered) set of all feature values, also called the *feature vector*, which is used to characterize a particular image). This abstraction, while serving purposes of rapid computation and adding robustness to above-mentioned variations in imaged content, can potentially introduce a disparity (or a gap) between the expected result and the computed result. This gap could be caused by a variety of factors, which include discriminant potential in the extracted signature in general or for the intended query, and the extent to which it was applied to the imaged data, among others. It is, therefore, valuable to consider characterizing CBIR systems through such an itemization of gaps and characteristics.

In the published literature, two gaps have been identified in CBIR techniques: (1) the *semantic gap*^{1,5,10} between the low-level features that are automatically extracted by machine and the high-level concepts of human vision and image under-

standing, and (2) a *sensory gap* defined by Smeulders et al.¹ between the object in the world and the information in a (computational) description derived from a recording of that scene. However, in our view, there are many other gaps that hinder the use of CBIR techniques in routine medical image management. For instance, there is a highly significant gap in the *level of integration* of CBIR into the general patient care information system. As another example, there is a gap in the *automation of feature extraction*. By means of the concept of gaps, we present a systematic analysis of required system features and properties. The paper classifies some prominent CBIR approaches in an effort to spur a more comprehensive view of the concept of gaps in medical CBIR research. We also attempt to show how our approach can be applied to characterize and distinguish prominent medical CBIR methods that have been published in the literature.

METHODS

There are several gaps that one can define to explain the discrepancy between the proliferation of CBIR systems in the literature and the lack of their use in daily routine in the departments of diagnostic radiology at healthcare institutions, for example. It is insufficient, however, to merely define these gaps. To benefit from the concept of gaps, it is imperative to analyze systems presented in the literature with respect to their capability to close or minimize these gaps. In addition to the gaps, it is also important to be aware of other system characteristics that, although not resulting in a gap, might be critical for CBIR system analysis and classification. In this section, we address these points systematically.

Defining an Ontology

We aim at defining a classification scheme, which we will call an ontology, by means of individual criteria, i.e., the so-called gaps. According to Lehmann,¹¹ such an ontology must satisfy several requirements regarding the entities (gaps), the catalog (ontology), and the applications of the ontology.

Requirements for the entities

Any ontology is an abstract complex of terms, and concrete criteria for requirements of the entities

must be defined on a meta-level of abstraction. In particular, such terms must be

- *Abstract*. They are formulated in a general manner that allows their instantiation to any approach of a medical CBIR system that has been published in the literature.
- *Applicable*. They are formulated in such a way that they can be used in a variety of semantic contexts of medicine, where CBIR systems are applied. In particular, the instantiation of the entities of the ontology should not be affected by the person using the ontology.
- *Verifiable*. They are formulated in such a way that there exists a method to evaluate each individual criterion.

Requirements for the Catalog

A system of abstract, applicable, and verifiable entities is called an ontology. However, in addition to the characteristics that are required for the entities of the ontology, the ontology itself must satisfy certain criteria. In particular, the collection of criteria must be

- *Complete*. The ontology covers all characteristics of medical CBIR systems and can be mapped to any situation and context of use. In particular, if two systems are characterized by the instances of the entities of the ontology, these instances must differ for different systems.
- *Unique*. The ontology is well defined. In other words, if a system is characterized by means of the ontology, the same system always results in the same instantiation.
- *Sorted*. The entities of the ontology are ordered semantically. For instance, they are grouped to support their unique assignment.
- *Efficient*. The application of the ontology is possible within finite time and effort, and all criteria can be decided without additional devices or computer programs.

Requirements for the Application

With respect to CBIR in medicine, an ontology characterizes existing system approaches or assists the conceptualization and design of a novel system. Hence, there are two basic uses of an ontology:

- *A priori*. The ontology is used as a guideline for system design.

- *A posteriori*. The ontology is used as a catalog of criteria for system analysis and weak-point detection.

The Concept of Gaps

In this paper, we aim to build an ontology of gaps. The concept of gaps has often been used in CBIR literature, and the semantic gap is one of the prominent examples. To elaborate on what we have previously mentioned, the semantic gap is the disparity or discontinuity between human understanding of images and the “understanding” that is obtainable from computer algorithms. This gap has a direct effect on the evaluation of images as “similar,” as judged by humans, versus the same images being judged as similar by algorithms. Image similarity is defined by a human observer in a particular context on a high semantic level. On the other hand, for algorithms, image similarity is defined by computational analyses of pixel values with respect to characteristics such as color, texture, or shape. The semantic gap is closely connected to not only the content (objects) of the image but also to (1) the features used for the signature and (2) the effectiveness of the algorithms that are used to infer the image content. The semantic gap is of high importance as a factor affecting the usefulness of CBIR systems and is frequently cited by CBIR researchers. Three examples are given in this paper: First, the work of Enser and Sandom¹² who have provided a detailed analysis of the semantic gap and created a classification of image types and user types to further understand categories of semantic gaps; second, Eakins and Graham¹⁰ who have used the idea of semantic content as a way to categorize types of CBIR queries—specifically, Eakins defines three types of CBIR queries according to their respective levels of semantic content; and, third, the recent work of Bosch et al.¹³ who have created a classification of published strategies that attempt to bridge the semantic gap by automated methods and have illustrated them for the domain of natural scene images.

In this paper, we extend this concept of “gap” to apply to other facets or aspects of CBIR systems. We may consider the semantic gap to be a break or discontinuity in the aspect of *image understanding*, with “human understanding” on one side of the gap and “machine understanding” on the other. Simi-

larly, we may identify breaks or discontinuities in other aspects of CBIR systems, including

- *The level of automation of feature extraction*, with full automation on one side and completely manual extraction on the other
- *The level of support for fast image database searching*, with optimized algorithms and data structures, supported by parallelized hardware on one side and exhaustive, linear database searching with no specialized hardware support on the other
- *The level to which the system helps the user to refine and improve query results*, with “intelligent” query refinement algorithms based on user identification of “good” and “bad” results, on one side and no refinement capability at all on the other

Each gap (1) corresponds to an aspect of a CBIR system that is explicitly or implicitly addressed during implementation, (2) divides that aspect between what is *potentially* a fuller or more powerful implementation of that aspect from a less powerful implementation, and (3) has associated with it methods to bridge or reduce the gap. We note that a gap, as applied to a particular system, may or may not be significant for achieving the goals of that system and, when bridged, may or may not add value to the system for the particular system purpose. For example, a stand-alone CBIR system operating on a small database may respond to queries perfectly well with an exhaustive, linear search of its database and have no need for search optimization, let alone hardware parallelization. However, it appears highly likely that the use of CBIR systems within clinical routine in large treatment centers will require features such as the ability to handle multiple image modalities for multiple treatment purposes, efficient extraction and indexing of clinical-content-rich features, capability to exchange information with the patient information system, and optimized retrieval algorithms, data organization, and hardware support; in other words, many of the gaps that we identify will need to be bridged for practical application to clinical routine.

We have created four broad categories of gaps, as follows:

- *Content*. The user’s view of modeling and understanding images
- *Features*. The computational point of view regarding numerical features and their limitations

- *Performance*. The implementation and the quality of integration and evaluation
- *Usability*. Ease of use of the system in routine applications

The gap names, the CBIR system aspect to which the gaps apply, and the “sides” of the gaps are given in Table 1.

CBIR Characteristics

In addition to the gaps, additional characteristics are useful to specify and distinguish medical CBIR systems. Because we aim at an a posteriori application of the gap ontology, we additionally characterize the

- *Intent and data*. The goal or intent of the medical CBIR approach and the data that is used with it
- *Input and output (I/O)*. The level of input and output data that is required to communicate with the CBIR system
- *Feature and similarity*. The kind of features and distance measures applied by the system

RESULTS

Figure 1 summarizes the overall results. In total, we defined 14 entities in the four groups of CBIR system gaps, and six entities in the three groups of CBIR system characteristics. The notation “xxx” means that the entity description requires additional information that depends on the medical context and/or system.

Gaps Gaps are characterized in “*Content Gaps*,” “*Feature Gaps*,” “*Performance Gaps*,” and “*Usability Gaps*.” As we discuss each gap, we categorize the ways in which systems attempt to bridge or ameliorate the gap, including the case (not addressed) in which the system does nothing to respond to the gap. We have provided examples for many of these categories, drawing on the characteristics of CBIR systems or technology closely related to CBIR, as reported in the technical literature. For each gap listed in this section, we give

- In italics, the CBIR system aspect to which the gap applies
- A summary overview of the gap
- Categories of methods to bridge or ameliorate the gap and, frequently, examples of the methods

Table 1. Gaps for Various Aspects of CBIR Systems

GAP CATEGORY	CBIR SYSTEM ASPECT		
	LESSER REALIZATION	GAP NAME	FULLER REALIZATION
Content	Image Understanding		
	Computable, "machine understanding"	Semantic	Human understanding of image
	Imaging and/or Clinical Context in which a System May Be Used		
	Narrow, limited context of modalities, protocols, or diagnostic procedures	Use context	General or broad context of modalities, protocols, or diagnostic procedures
Feature	Automation of Feature Extraction		
	Fully manual	Extraction	Fully automated
	Granularity of Image Object Structure Recognized by the System		
	Only global structure is recognized	Structure	Local image objects are recognized, as well as the global image structure
	Granularity of Image Visual Detail Processed by the System		
	Visual detail processed at only one level of scale	Scale	Visual detail processed at multiple levels of scale
	Dimensionality of Spatial and Time Inputs Actually Used to Compute Features		
	Not all spatial and time dimensions available in the input data are used to compute features	Space + time dimension	All spatial and time dimensions available in the input data are used to compute features
	Dimensionality of Channel Inputs Actually Used to Compute Features		
	Not all channel dimensions of input data are used to compute the features	Channel dimension	All channel dimensions of input data are used to compute the features
Performance	Level of Actual Implementation of the System		
	Not implemented; conceptual only	Application	Online executable in the Internet
	Level of Integration into Patient Care Information System		
	System is standalone	Integration	System can interchange information with patient care information system
	Level of Support for Fast Database Searching		
	Simple linear search of entire database; no parallelization of hardware support	Indexing	Data organized in spatial trees and/or clustered, and an optimized search strategy is used; supported by parallelized hardware architecture
Level to Which the System Validity of Retrieval Has Been Evaluated			
No evaluation	Evaluation	Quantitative evaluation with ground truth data	
Usability	Level to Which User May Use and Combine Text/Visual Queries		
	Query by text only	Query	Query by text, numerical specification of features, example, image, selection of structures from restricted sets, sketch, or any combination of these
	Level to Which the System Helps a User Understand Query Results		
	Results are returned without feedback from the system	Feedback	Results include a similarity measure and cue that help the user understand why particular results were returned
Level to Which the System Helps the User to Refine and Improve Query Results			
No refinement capability is provided	Refinement	System "learns" the user's preferences through feedback from the user about which specific results best met the user's goals	

System Characteristics The system characteristics are discussed in "Intent and Data," "Input and Output," and "Feature and Similarity." This system characteristics hierarchy is intended to capture important attributes of CBIR systems that are not well represented by the gap ontology.

Content Gaps

This group of gaps addresses the modeling, understanding, and use of images from the standpoint of a user. We have defined two relevant gaps.

Semantic Gap

Image Understanding The similarity of images defined by a human observer in a particular context is based on a high level of semantics: Terms that are considered useful or "meaningful" for the purpose at hand and that may even be restricted to a controlled vocabulary are assigned to the imaged, human-identified objects. In contrast, computational analysis of image content is based on algorithmic processing of pixel values. In our definition, the semantic gap is bridged if a relation of image

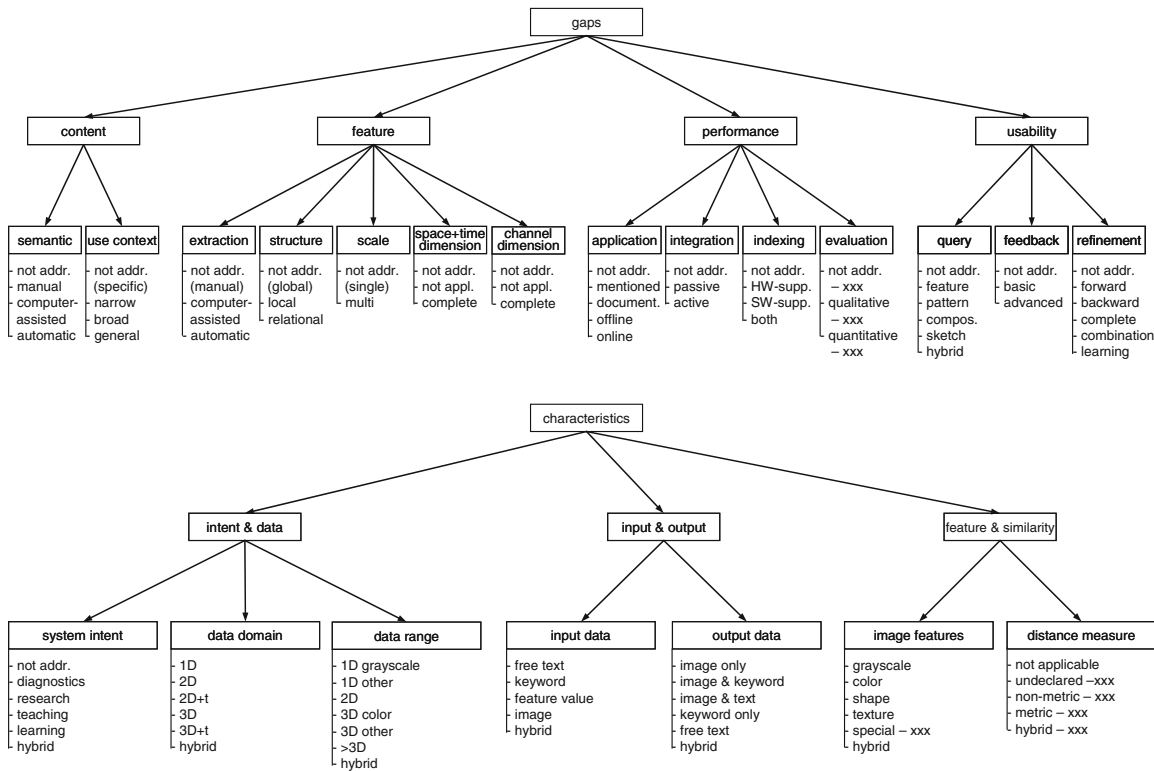


Fig. 1. Results. Gap and system characteristics in medical CBIR research.

structures to medical meaning is established. Categories and examples for this gap:

- *Not addressed.* Meaningful terms are not assigned to images or ROIs; images are indexed by strictly mathematical measures, such as measures of color, texture, and shape.
- *Manual.* Meaningful terms are manually assigned; for X-ray images of the cervical spine, a human operator may use interactive software to assign vertebrae labels “C1,” “C2,” ... “C7” to image regions.
- *Computer assisted.* A semi-automatic process is used to assign meaningful terms; in the above example, a computer algorithm may assign the labels to the regions on the image; a human operator then reviews and corrects them.
- *Automatic.* Meaningful terms are automatically assigned; in this case, a computer algorithm would assign the region labels with no human intervention; some experimental work toward developing methods to automatically extract and associate low-level features to meaningful medical semantics has been reported in some

limited domains, such as the mapping of shape, size, intensity, and texture features to radiologist semantics used for lung nodules (lobulation, malignancy, margin, sphericity, and others) in thoracic CT images.¹⁴

Use Context Gap

Imaging and/or Clinical Context in which a System May Be Used By “use context” in this study, we refer to the modality of the images, the type of gross or fine anatomy captured, the presentation, and/or the particular medical intent for which they were acquired (“context” might also reasonably be used to refer to the associated patient history; we treat this latter factor in “Integration Gap”). The context in which a CBIR system can be used is usually restricted. Medical CBIR systems frequently are designed to support queries on a certain imaging modality or within a certain clinical context, such as a particular medical protocol or diagnostic procedure. These restrictions allow the use of medical a priori knowledge of the

imaging modality or context: Otherwise, the CBIR problem may be difficult to formulate so that it is computable. Ideally, of course, the system should support generalized use with minimal or no user limitation and would automatically determine modality, anatomy, and presentation directly from image contents. Categories and examples for this gap:

- *Not addressed.* The system is *specific* to a certain context, and the context gap is wide; for example, the system may be tailored to the retrieval needs of a database of gastrointestinal (GI) tract histology images.
- *Narrow.* The system operates only for a small number of modalities or protocols or diagnostic procedures or on a small number of combinations of these; for example, a cancer-oriented system may be designed to operate on histology images of breast, lung, and uterine cervix, and may support labeling from controlled vocabularies for each of these domains only.
- *Broad.* The system operates for a large number of modalities or protocols or diagnostic procedures, or on a large number of combinations of these; for example, a system may allow the user to store segmented shapes from any types of digital imaging and communications in medicine images into a database and to query the database by sketches of these shapes.
- *General.* No restrictions apply at all, neither to the modality, the protocol, nor the diagnostics.

Feature Gaps

When we consider the implementation steps that must occur to derive characterizations of images that are computable, we discover feature-related gaps. These gaps correspond (1) to the inadequacies of the chosen numerical features to characterize the image content or (2) to the practical difficulties of extracting these features from the images.

Extraction Gap

Automation of Feature Extraction Not all medical CBIR systems automatically extract the features. Some are based on manual indexing of images, such as manually marking the boundaries of vertebrae in spine X-rays. This manual process is usually labor-intensive and prone to error. This gap may be bridged

by computer-assisted or automatic feature extraction methods. Categories and examples for this gap:

- *Not addressed.* Feature extraction is completely interactive or *manual*, e.g., manually outlined shapes, such as cardiac anatomical features (atria, ventricles, ascending aorta, and pulmonary artery).¹⁵
- *Computer-assisted.* Feature extraction is partly interactive, e.g., shapes segmented with the “live-wire” algorithm,¹⁶ which completes shape segmentation, such as for vertebrae on spine X-ray images, based on a few user-supplied “guiding points”; another example is interactive region segmentation on histology images by region-growing or *K-means* clustering algorithms.¹⁷
- *Automatic.* There is no human interaction in the feature extraction; examples would be extraction of color or grayscale histograms, Gabor wavelet coefficients, or object counts, computed from an image with no human intervention.¹⁸

Structure Gap

Granularity of Image Object Structure Recognized by the System The extraction of global parameters that describe the entire image is frequently insufficient for medical applications. Hence, ROIs that describe only a certain part of an image must be identified and characterized by appropriate parameters. Categories and examples for this gap:

- *Not addressed.* Features are extracted for the entire image (*global case*); examples would include grayscale histograms computed from all of the pixels in the image.¹⁸
- *Local.* Features are extracted for individual ROIs; examples include color and texture measures computed from the interiors of tissue regions of known type, such as from the cervix region on images that contain the uterine cervix and surrounding anatomy.¹⁹
- *Relational.* Features are extracted for a certain composition of individual ROIs or objects; an example is the characterization of the relative spatial relationships of cardiac chambers (atria and ventricles) on tomography images.¹⁵

Scale Gap

Granularity of Image Visual Detail Processed by the System A fundamental characteristic of a digital

image is its resolution or granularity of detail that is available in the image. Some image characteristics may be best captured at reduced resolution (as a human observer may more easily interpret some characteristics “at a distance”). The image may be processed to produce additional images with lesser resolutions. For a particular query task, one resolution level may be more suitable than another; for this reason, multiscale representations of image features are highly desirable. The availability of features extracted at multiple image resolution levels (multiscale features) adds potentially significant flexibility to the system, allowing the user to search for both gross-level and fine-level characteristics of the images (note that scale refers to the *resolution of detail* within the image and is not identical with *structure*, which refers to the composition of the image from regions or objects of interest). Categories and examples for this gap:

- *Not addressed.* Features are extracted for a fixed *single* scale; an example would be calculation of texture features from co-occurrence matrices that are applied to the image only at its original spatial resolution.
- *Multi.* Features are extracted at multiple scales of the image; an example would be a system that applies Gaussian blurring and down sampling to create multiple spatial resolutions for each image, and then applies co-occurrence matrices to the image at each of these resolutions; a variation of this idea is to use the image at its original resolution but to apply mathematical operators that output information about the image contents at multiple levels of detail, as has been done²⁰ for tumor shape, using mathematical morphology operators with multiple sizes of structuring elements; another example is any approach that includes features based on *curvature scale space*, which is inherently a multiscale approach, as has been done to characterize masses in mammography images.²¹

Space + Time Dimension Gap

Dimensionality of Spatial and Time Inputs Actually Used to Compute Features Features may be extracted from data that is different from the original given data. It is convenient to speak of a mapping from the domain space of the original data to the range space containing what we term

the “feature source data” from which features are actually computed. These two spaces may differ in spatial dimension. For example, the original data may be 3D, such as 2D magnetic resonance imaging slices along with information about the third spatial dimension; If features are computed from the individual 2D images while ignoring the third spatial dimension information, there is a spatial dimensionality gap between the original data and the feature source data or, in terms of the mapping described above, a dimensionality gap between the domain and range spaces. For full generality, we speak of a space + time dimension gap to cover the cases where time is also a dimension in the original data. As it appears that 1D and 2D data are always mapped to feature source data of the same dimensionality for feature extraction, we believe that this gap does not apply when the system’s original data is of either of those dimensions. Note that when we speak of dimension gaps and feature computation in this and the following section, the reader should bear in mind that we are not speaking of dimensionality of feature vectors but only dimensionality of the data from which the feature vectors are computed. Categories and examples for this gap:

- *Not addressed.* The dimension of the data range space is less than the dimension of the data domain space.
- *Not applicable.* The system handles 1D or 2D data only.
- *Complete range.* The dimension of the data range space is equal to the dimension of the data domain space; an example is the indexing of functional imaging data consisting of 3D positron emission tomography images plus associated temporal information by including the volumetric characteristics of the data in the indexing.²²

Channel Dimension Gap

Dimensionality of Channel Inputs Actually Used to Compute Features It is also possible that the original system data differs in “channel dimensionality” from the feature source data, where the channels correspond to separate image data planes (typically colors). With respect to channel dimensionality, for example, the original data domain space may have three color planes (such as RGB), but features may be computed from a single-plane intensity image that

is derived from the RGB data by irreversible, dimensionality-reducing transformation, hence, a channel dimension gap exists. If the original data is single channel (e.g., grayscale), this gap does not arise. Categories and examples for this gap:

- *Not addressed.* The dimension of the channel data range space is less than the dimension of the channel data domain space.
- *Not applicable.* The original system data is single channel.
- *Complete range.* The dimension of the channel data range space is equal to the dimension of the channel data domain space; an example is characterizing skin lesions on dermatology images by RGB histograms;²⁰ a variation of this technique is to first transform the image, with a dimensionality-preserving transformation, to a different color space, such as the MPEG7 HDS space, before calculating the histogram.²³

Performance Gaps

Not all systems found in the literature are completely implemented and executable for performance evaluation. For those that are implemented and testable, the performance criteria include quality of integration, level of support for fast database searching, and the extent to which evaluation of the system for acceptable retrieval has been done.

Application Gap

Level of Actual Implementation of the System This is a gap between what is described in the published literature and what is available for test and use. In scientific literature, there is an immense gap between the conceptual level of the described medical CBIR systems and their implementation. Frequently, concepts are published, but a running system is not available or the level of implementation is not clear (Müller et al.⁵ make this same point in the discussion of CBIR in PACS and other medical databases). Categories for this gap:

- *Not addressed.* An implementation is not mentioned at all.
- *Mentioned.* An implementation is described, but no supporting evidence is provided.
- *Documented.* Screen shots are shown in the publication as evidence of the implementation of the system.

- *Offline.* An implementation is available for download and installation.
- *Online.* An implementation is directly accessible and executable via the Internet.

Integration Gap

Level of Integration into Patient Care Information System Eakins and Graham¹⁰ implicitly recognizes this gap in the context of general CBIR systems:

“The experience of all commercial vendors of CBIR software is that system acceptability is heavily influenced by the extent to which image retrieval capabilities can be embedded within user’s overall work tasks.”

This is also true, and may be particularly so, within the medical domain. If a system for medical CBIR exists, it may or may not satisfy the critical need of being integrated with the patient information system and may be purely stand-alone. An “integration gap” may exist, then, which is bridged according to the level of clinical workflow integration. At one end of the integration spectrum, a stand-alone CBIR system would allow queries by image characteristics only; at the other end of this spectrum, a CBIR system that is completely integrated into the patient care databases would also allow queries by any of the patient parameters related to medical history, diagnosis, and treatment, in combination with queries by image characteristics. Categories and examples for this gap:

- *Not addressed.* The application is not interconnected with clinical data; for example, a prototype system for retrieval of cervicography images by color and texture from a small database of uterine cervix images.²⁴
- *Passive.* The patient/image data is passed to the CBIR application.
- *Active.* The application can initiate its own access to clinical data.

Indexing Gap

Level of Support for Fast Database Searching Given a query image, a CBIR system searches the database for similar images. A critical performance parameter of a medical CBIR system is the response

time experienced by the user which, in a large database, may depend upon the indexing of multi-scale image descriptions for efficient data access. This indexing is not trivial. Simple strategies like A*-trees or inverse files (a concept borrowed from the document processing community; these files associate lists of image features with images containing those features²⁵) cannot be applied directly, and research is required to cope with large image repositories such as those generated in health care. Specialized hardware architecture may also be critical. Categories and examples for this gap:

- *Not addressed.* The system is based on a brute force approach, where the distance between the query feature vector and every feature vector in the database is computed; this approach is usually feasible only for stand-alone CBIR systems operating on small databases.
- *Hardware supported.* The system is based on the brute force approach, but the database search is supported by specialized hardware architecture, such as a parallel computing environment; an innovation in this area is the use of active disk architecture, where some of the database search intelligence is placed on processors on the disk devices, and an “early discard” strategy is used to discard database entries that do not satisfy query requirements rather than sending them over the system connection to the CPU.²⁶ If the active disks are operated in parallel, this approach has both the advantages of distributed computing and early data discard.
- *Software supported.* The database of feature vectors is organized into clusters or cluster trees; the system uses algorithms tailored to this tree organization for fast access to feature vectors relevant to a particular query; for example, data organization based on clustering in shape space and a search strategy coupled with that organization have been implemented for a database of spine X-rays;²⁷ a second example are the spatial access methods and specialized feature extraction developed for a database of tumor shapes that are reported in Korn et al.²⁸
- *Both.* The system incorporates the indexed approach as described above and supports it with a distributed computing environment.

Evaluation Gap

Level to which the System Validity of Retrieval has been Evaluated In large data bases, the gold standard or ground truth is unknown, i.e., it is impossible to determine the correct answer for a test query. In other words, an expected output of the system answering a certain question is unavailable. Hence, the comparison of competing approaches for global/local feature extraction and distance measures is difficult and inaccurate. Instead of error measures computed from leave-one-out experiments, precision, recall, and the F measure are calculated, where the number of correct answers is not used. Categories and examples for this gap:

- *Not addressed—xxx.* No experiments are described; the database contains xxx images.
- *Qualitative—xxx.* Experiments are described but without expected output or ground truth based on xxx images.
- *Quantitative—xxx:* Experiments are described with expected output or ground truth based on xxx images; for example, Xu et al.²⁹ report results for retrieval of spine vertebrae by shape from a ground truth set of 207 images.

Usability Gaps

This group of gaps addresses the usability of the system. Whereas the performance gaps focus on the area in which the system is used, the usability gaps describe the ease of use of the system from the perspective of the end user.

Query Gap

Level to which User May Use and Combine Text and Visual Queries To use the QBE paradigm, where a visual example is presented to the retrieval system, specialized mechanisms and interfaces are required. Currently, effective tools to assist the user in drawing or composing a search pattern are missing, and QBE is difficult and time-consuming. Categories and examples for this gap:

- *Not addressed.* The user inputs alphanumeric text, disregarding the QBE paradigm.
- *Feature.* The user specifies certain intervals of feature vectors or vector components.

- *Pattern*. The user specifies an example image or a part of an image (ROI); examples include systems like image retrieval in medical applications (IRMA),⁸ which the user may submit an entire image and search for similar images.
- *Composition*. The user interactively selects and places structures from a given set; for example, the uterine cervix CBIR system described in Antani et al.³⁰ allows the selection of ROIs, pre-drawn by medical experts, to be used as part of the query. The user selects properties of these ROIs, such as color and/or texture, to complete the query definition.
- *Sketch*. The user interactively creates example patterns, including the previous options, but without being restricted to choosing from pre-defined pattern sets (for example, the user may create a “freehand drawing” of a query shape); examples include the uterine cervix CBIR system²⁴ referenced above, which also allows freehand drawing of the ROIs to be used in the query; another example is the retrieval of spine vertebrae by sketching the desired shape.²⁹
- *Hybrid*. The user may input text, one of the above visual patterns, or a combination of both.

Feedback Gap

Level to which the System Helps the User to Understand Query Results The result of a CBIR query is usually presented by displaying the most similar images found in the archive. However, it may be difficult to understand how similar the system believes the individual results are to the query and how the query needs to be altered to improve the recall and precision. To close the feedback gap, some rationale or cues for the retrieved results may be provided by the CBIR system. Categories for this gap:

- *Not addressed*. The results returned by the system are not commented at all.
- *Basic*. A similarity or dissimilarity number is given for each returned result; for example, the spine X-ray CBIR system of Antani et al.³⁰ returns a dissimilarity (distance) measure for each result.
- *Advanced*. More sophisticated explanations are provided by the system, such as cues indicating

the relative significance of the various features in the returned results.

Refinement Gap

Level to which the System Helps the User to Refine and Improve Query Results CBIR systems should provide the user options to repeat and modify a query. Sometimes, they also track the refinement process to learn user preferences. Categories and examples for this gap:

- *Not addressed*. Just one request is answered.
- *Forward*. A rudimentary option for query refinement is provided, such as the user being able to provide “relevance feedback” to the system by ranking individual results returned by the system on a scale that ranges from “low relevance” to “high relevance” and resubmitting the query.³¹
- *Backward*. In the refinement loop, the user can step back if results become worse.
- *Complete*. A full history of the interactive session is available for restoration of any intermediate stage.
- *Combination*. Based on the complete history, different queries can be performed, and their results can be combined; for example, the extended query refinement approach by the IRMA framework, which, additionally, supports set combination (such as AND, OR, and NOT) of intermediate query results.³²
- *Learning*. During the usage, the system adapts to the user’s need.

System Characteristics: Intent and Data

Under this heading, we group the intent or goal of the CBIR application, as well as the data domain (input data) and range (data used to compute features) in use.

System Intent

The purpose of a system, as well as the target user group, may vary. A medical CBIR system can assist the user in various clinical and research tasks. Categories for this system characteristic:

- *Not addressed*. No information about the purpose is given.

- *Diagnostics*. For example, the system is intended for case-based reasoning.
- *Research*. For example, the system is intended to collect data to support evidence-based medicine.
- *Teaching*. For example, the system is intended to find examples for sets of case collections.
- *Learning*. For example, the system is intended for the self-exploration of medical cases.
- *Hybrid*. The system is intended for at least two of the previously mentioned cases.

Data Domain

This category defines the input data available to the system. A medical CBIR system usually copes with 2D images, a sequence of images over time ($2D + t$) or 3D volumes. Categories for this system characteristic:

- *1D*. The system data consists of biomedical signals.
- *2D*. The system data consists of images.
- *2D+t*. The system data consists of image sequences.
- *3D*. The system data consists of volumetric datasets.
- *3D+t*. The system data consists of a sequence of volumes.
- *Hybrid*. The system data consists of more than one of the categories above.

Data Range

This category defines the data from which features are computed (the “feature source data” of “[Space + Time Dimension Gap](#)” and “[Channel Dimension Gap](#)”). Medical image feature source data is typically grayscale (1D) or color (3D). However, in multispectral imaging, higher data ranges do exist. Categories for this system characteristic:

- *1D grayscale*. The system data consists of grayscale images or volumes.
- *1D other*. The system data has a 1D range other than grayscale.
- *2D*. The system data has a 2D range.
- *3D color*. The system data consists of color images or volumes.
- *3D other*. The system data has a 3D range other than color.
- *>3D*. The system data consists of a multichannel range.

- *Hybrid*. The system data consists of more than one of the categories above.

System Characteristics: I/O

Content-based image retrieval in medical applications may also be combined with a text-based search in the patient health record. According to Tang et al., different combinations between text and images for input and output might be used.⁴ In general, it is easier to make inferences from text to images than from images to text. The first can be done from text associated with the image (e.g., Google image search), whereas the latter needs semantic concepts.

Input Data

Categories for this system characteristic:

- *Free text*. The system input consists of any alphanumerical wording that requires stemming, etc. for automatic processing.
- *Keyword*. The system input consists of words addressing a concept of special semantics, e.g., as part of a controlled vocabulary.
- *Feature value*. The system input consists of instances of an image-based feature, e.g., a numerical range.
- *Image*. The system input consists of a query image, marked region of interest, drawing, or any other nonalphanumeric data.
- *Hybrid*. The system input consists of more than one of the categories above.

Output Data

Categories for this system characteristic:

- *Image only*. The system returns similar images.
- *Image and keyword*. The system returns similar images and controlled image category information.
- *Image and text*. The system returns similar images and other text, such as in multimedia documents.
- *Keyword only*. The system returns a restricted set of words based on a controlled vocabulary.
- *Free text*. The system returns alphanumerical wording that describes the image.

- *Hybrid*. The system output consists of more than one of the categories above, for example, images, keywords, and free text.

System Characteristics: Features and Similarity

The process of computing the similarity between images is dependent on (1) the particular representation of the image signature (i.e., the numerical features that are used to characterize the image) and (2) the distance or similarity measure that is being used to compute similarity of signatures.

Image Features

The type of features that are used to represent an image for content-based retrieval is one of the most critical system characteristics. These features may be computed from points, lines, or areas. Categories for this system characteristic:

- *Grayscale*. The image features are based on image intensity only.
- *Color*. The image features are based on color and grayscale.
- *Shape*. The image features are based on location or delineation of a region.
- *Texture*. The image features are based on complex visual pattern related to a ROI.
- *Special—xxx*. The image features are based on a context-based feature, where xxx denotes the name of the feature.
- *Hybrid*. The image features are based on more than one of the categories above.

Distance Measure

Besides the type of features, different methods to determine the similarity or dissimilarity between the features must be applied. It is of special interest whether the distance measure forms a metric, as the properties of metric distance functions, in particular, the triangle inequality, can be exploited to optimize searching of database images. For example, Qian and Tagare²⁷ have shown how this can be done for shape similarity searches in a database of spine X-ray images by clustering the database images into nodes, with each node corresponding to groups of similar images, and then implementing database

searches by comparing the query image to nodes (cluster centers) rather than to all of the images in the database. Traina et al. summarize the metric distance function axioms in Traina et al.:¹⁸ a distance function $d(A,B)$ of features $A \neq B \neq C$, which is a metric, must satisfy (1) reflexivity, i.e., $d(A,A)=0$, (2) nonnegativity, i.e., $d(A,B) > 0$, (3) symmetry, i.e., $d(A,B)=d(B,A)$, and (4) the triangle inequality, i.e., $d(A,B)+d(B,C) \geq d(A,C)$. Categories for this system characteristic:

- *Not applicable*. No distance measure is used, e.g., the system does retrieval by intervals of feature values.
- *Undeclared—xxx*. The distance measure is named xxx, but it is not asserted to be a metric.
- *Nonmetric—xxx*. A nonmetric distance measure is used, where xxx denotes the measure.
- *Metric—xxx*. A metric distance measure named xxx is used.
- *Hybrid*. Any combination of the above is used.

DISCUSSION AND CONCLUSION

In this paper, we have proposed a nomenclature and classification scheme for analysis and assessment of medical CBIR systems. We have attempted to address the core features and required functionality of medical CBIR explicitly, systematically, and comprehensively, using the concept of gaps as a unifying idea to highlight potential shortcomings in various aspects of CBIR systems, as well as to illustrate methods for addressing those shortcomings. For important CBIR system characteristics that do not fit into the gaps ontology, we have provided a second, supplementary hierarchical grouping of related attributes. It is our intent that this effort will contribute to the ongoing research and development in medical CBIR by providing a more formal and methodical approach to conceptualizing CBIR systems in terms of their characteristics, their potential shortcomings, and how these shortcomings may be addressed, than has hitherto been available.

ACKNOWLEDGMENT

This research was supported [in part] by the Intramural Research Program of the U.S. National Institutes of Health (NIH), U.S. National Library of Medicine (NLM), and the U.S. Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

1. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R: Content-based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell* 22(12):1349–1380, 2000
2. Niblack W, Barber R, Equitz W, Flickner M, Glasman E, Petkovic D, Yanker P, Faloutsos C, Taubin G: The QBIC project: Querying images by content using color, texture, and shape. *Proc SPIE* 1908:173–187, 1993
3. Tagare HD, Jaffe CC, Duncan J: Medical image databases: A content-based retrieval approach. *J Am Med Inform Assoc—JAMIA* 4(3):184–198, 1997
4. Tang LHA, Hanka R, Ip HHS: A review of intelligent content-based indexing and browsing of medical images. *Health Inform J* 1(5):40–49, 1999
5. Müller H, Michoux N, Bandon D, Geissbuhler A: A review of content-based image retrieval systems in medical applications. Clinical benefits and future directions. *Int J Med Inform* 73(1):1–23, 2004
6. Qi H, Snyder WE: Content-based image retrieval in picture archiving and communications systems. *J Digit Imaging* 12(2 Suppl 1):81–83, 1999
7. Müller H, Rosset A, Garcia A, Vallée JP, Geissbuhler A: Informatics in radiology (Inforad): Benefits of content-based visual data accessing radiology. *Radiographics* 25(3):849–858, 2005
8. Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohlen M, Schubert H, Wein BB: Content-based image retrieval in medical applications. *Methods Inf Med* 43(4):354–361, 2004
9. Müller H, Clough P, Hersh W, Deselaers T, Lehmann T, Geissbuhler A: Evaluation axes for medical image retrieval systems—the ImageCLEF experience. *Proc. 13th Annual ACM International Conference on Multimedia* 1014–1022, 2005
10. Eakins J, Graham M: Content-based image retrieval. Report 39, JISC Technology Applications Programme, University of Northumbria at Newcastle, October 1999
11. Lehmann TM: Digitale Bildverarbeitung für Routineanwendungen. Evaluierung und Integration am Beispiel der Medizin. Deutscher Universitäts-Verlag, GWV Fachverlage, Wiesbaden, 2005 (in German)
12. Enser P, Sandom C: Towards a comprehensive survey of the semantic gap in visual image retrieval. *Lect Notes Comput Sci* 2728:291–299, 2003
13. Bosch A, Muñoz X, Martí R: Which is the best way to organize/classify images by content? *Image Vis Comput* 25:778–791, 2007
14. Raicu DS, Varubangkul E, Cisneros JG, Furst JD, Channin DS, Armato III, SG: Semantics and image content integration for pulmonary nodule interpretation in thoracic computer tomography. *Proceedings SPIE* 6512:OS1–OS12, 2007
15. Tagare HD, Vos FM, Jaffe CC, Duncan JS: Arrangement: a spatial relation between parts for evaluating similarity of tomographic section. *IEEE Trans Pattern Anal Mach Intell* 17(9):880–893, 1995
16. Mortensen EN, Barrett WA: Intelligent scissors for image composition. *Proc SIGGRAPH* 191–198, 1995
17. Lee YJ, Bajcsy P: An information gathering system for medical image inspection. *Proc SPIE* 5748:374–381, 2005
18. Traina C, Jr., Traina AJM, Araujo MRB, Bueno JM, Chino FJT, Razente H, Azevedo-Marques PM: Using an image-extended relational database to support content-based image retrieval in a PACS. *Comput Methods Programs Biomed* 80 (Suppl. 1):S71–S83, 2005
19. Xue Z, Antani S, Long R, Thoma G: Comparative performance analysis of cervix ROI extraction and specular reflection removal algorithms for uterine cervix image analysis. *Proc SPIE* 6512:411–419, 2007
20. Stanley RJ, Moss RH, Van Stoecker W, Aggarwal C: A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images. *Comput Med Imaging Graph* 27:387–96, 2003
21. Tao Y, Lo S-CB, Freedman MT, Xuan J: A preliminary study of content-based mammographic masses retrieval. *Proc SPIE* 6514:1Z–1–12, 2007
22. Kim J, Cai W, Feng D, Wu H: A new way for multi-dimensional medical data management: volume of interest (VOI)-based retrieval of medical images with visual and functional features. *IEEE Trans Inf Technol Biomed* 10(3):598–607, 2006
23. Howarth P, Yavlinsky A, Heesch D, Ruger S: Medical image retrieval using texture, locality and colour. *Lect Notes Comput Sci* 3491:740–749, 2005
24. Xue Z, Antani SK, Long LR, Jeronimo J, Thoma GR: Investigating CBIR techniques for cervicographic images. *Proc AMIA* 2007 (in press)
25. Squire DM, Muller W, Muller H, Raki J: Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback. *Proc 10th Scandinavian Conference on Image Analysis (SCIA'99)*
26. Huston L, Sukthankar R, Wickremesinghe R, Satyanarayanan M, Ganger GR, Riedel E, Ailamake A. Diamond: a storage architecture for early discard in interactive search. *Proc 3rd USENIX Conference on File and Storage Technologies (FAST'04)*
27. Qian X, Tagare HD: Optimal embedding for shape indexing in medical image databases. *Lect Notes Comput Sci* 3750:377–84, 2005
28. Korn P, Sidiropoulos N, Faloutsos C, Siegel E, Protopapas Z: Fast and effective retrieval of medical tumor shapes. *IEEE Trans Knowl Data Eng* 10(6):889–904, 1998
29. Xu X, Lee DJ, Antani S, Long LR: Pre-indexing for fast partial shape matching of vertebrae images. *Proc 19th International Symposium on Computer-Based Medical Systems (CBMS 2006)*, 105–110
30. Antani S, Cheng J, Long J, Long LR, Thoma GR: Medical validation and CBIR of spine x-ray images over the Internet. *Proc SPIE* 6061 0J(1–9), 2006
31. Xu X, Lee DJ, Antani SK, Long LR: Relevance feedback for spine x-ray retrieval. *Proc 18th International Symposium on Computer-Based Medical Systems* 197–202, 2005
32. Deserno TM, Güld MO, Plodowski B, Spitzer K, Wein BB, Schubert H, Ney H, Seidl T: Extended query refinement for medical image retrieval. *Journal of Digital Imaging* 2007 (in press). DOI [10.1007/s10278-007-90374](https://doi.org/10.1007/s10278-007-90374)