A Prototype Content-based Image Retrieval System for Spine X-rays

L. Rodney Long, Sameer K. Antani, George R. Thoma
National Library of Medicine, Bethesda, MD 20894
{long, antani, thoma}@nlm.nih.gov

Abstract

At the Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine, we are engaged in an effort in content-based image retrieval (CBIR) for biomedical image databases. Toward the goal of developing a functional and significant CBIR capability, we have created a prototype system for image indexing and retrieval which operates on a collection of spine x-rays and associated health survey data. In this paper, we present our prototype system functionality, performance results, ongoing research, and outstanding technical issues.

1. Content-based image retrieval (CBIR)

Our work in CBIR is the latest phase of research and development into the use of technology for the dissemination of biomedical multimedia information; this work has previously resulted in the development of a biomedical multimedia database system, a digital atlas of the cervical and lumbar spine, and an Internet archive of digitized x-ray images [1]. For our purposes, we consider CBIR to be composed of two broad categories of functions for *indexing* and *retrieval*, which we describe in the following.

Indexing - the computer-assisted data reduction of images into mathematical features. For the spine images, the features capture the shape information for vertebrae. Indexing consists of the steps of segmentation of the objects of interest (the vertebrae) and extraction of feature vectors (shape representation, in a data-reduced fashion) from the raw segmentations. An implicit requirement for indexing is that the feature vectors are organized for efficient search and retrieval. A step that we also propose to carry out at indexing time is the classification of the shape data (raw segmentations or feature vectors) into categories of interest at a semantic level: namely, the categories of "normal" or "abnormal" for particular biomedical characteristics associated with osteoarthritis and degenerative disk disease, such as anterior osteophytes, disc space narrowing, subluxation, and spondylolisthesis. Finally, we propose to store any text data that may be associated with our images as additional indexing information.

Retrieval - the user interaction to obtain desired images from the database. We break retrieval into the steps of user query formulation, user query feature vector extraction, query search, and similarity matching. At retrieval time, a feature vector q is derived from the user's query, and the database of feature vectors is navigated to locate feature vectors similar to q. Efficient organization of the database is required to avoid searches that are prohibitively expensive in search time. For example, if the database is organized as a tree, an efficient organization will allow a search to quickly rule out nodes too distant from q, and to localize the search to nodes that are computed to lie within an acceptable search radius, with respect to the similarity metric being used. A characteristic that we also desire for our retrieval system is the capability to

support *hybrid* queries composed of conventional text queries and direct image queries. `Figure 1` is a conceptual diagram of CBIR indexing and retrieval.
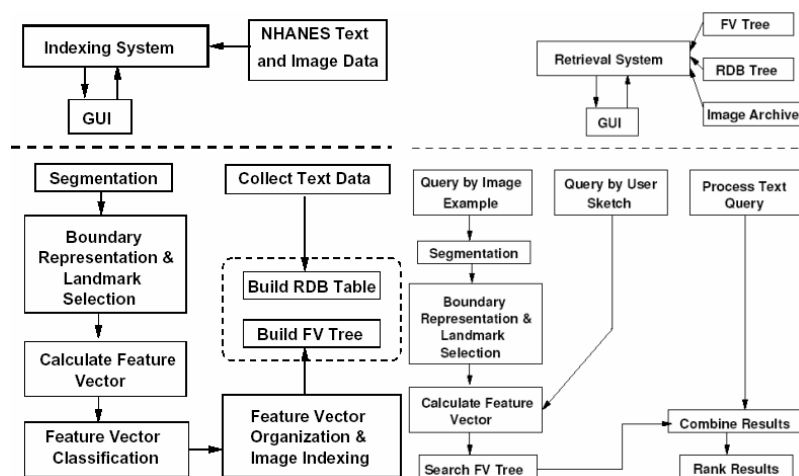

Figure 1. Conceptual view of the CBIR system for indexing and retrieval.

2. Prototype CBIR system

Our prototype system implements parts of both of the indexing and retrieval capabilities described above. Our indexing software provides: manual and computer-assisted segmentation; four methods for extracting feature vectors (i.e., representing shape); and simple "flat" storage for each of these four feature vector representations. For retrieval, the software provides: support for user-specified graphical queries; support for user selection of which of the four feature vector types to use; functionality for computing a feature vector q from the user's graphical query; and functionality for linearly searching our database for feature vectors similar to q, using a similarity matching method specific to the selected feature vector type. The retrieval system also provides the interface and the methods for retrieving the image and text data through combined, hybrid queries. User query specification for image queries allows query by sketch, and query by image example. The implementation is in the MATLAB language, and the indexing and retrieval functions work as an integrated system, except that the feature vector extraction is currently run as a separate script execution step.

`Figure 2` illustrates the user interface. The leftmost screen shows the interface to the segmentation function. This allows segmentation by manually marking points on the boundary of a vertebra, or by invoking an active contour segmentation routine. The raw segmentation (x,y) coordinate values are saved in an XML file along with descriptive data such as the segmentor's identification, date, qualifications (medical expert or engineer), and automatically-computed numerical data such as the coordinate values converted to an object-centered coordinate system. The center screen in `Figure 2` shows the query-by-example screen, where the user selects a representative image from the database to use for a query. The rightmost top screen illustrates query by sketch. The user may select from among several pre-stored shape templates and may edit these templates by dragging points to create the desired search shape. The rightmost bottom screen shows results of an image query: the leftmost shape is the query itself; the shapes to the right are the query results, in descending similarity, from left to right, to

the query shape. Below these shapes are the numerical similarity scores assigned by the CBIR system. A system description can be found in the papers of Antani et al [2,3].

The system supports hybrid text/image queries, where the text data associated with the images is health survey data collected in the second National Health and Nutrition Examination Survey (NHANES II) [4]. (Text query capability is not illustrated in `Figure 2`.) Text queries may use age, race, sex, height, weight, and presence/absence of self-reported neck or back pain. The text data and the shape data is stored in a MySQL database.

The four feature vector representations used are *invariant moments*, as defined by Hu [5] (four moments computed from the raw boundary); *scale-space filtering* [6] to calculate a simplified polygon, and representation of the resulting polygon as "shape tokens", each of which describes the polygon shape in a local region; *polygon approximation by curve evolution* using an iterative point-deletion scheme [7], and representation of the resulting polygon as bend angle versus normalized length in tangent space; and *Fourier descriptors* [8], which may be calculated on any vector representation of the shape at any level of detail (hence may be combined with any of the above methods). `Figure 3` illustrates a vertebra boundary simplified by curve evolution (i.e., some of the original raw boundary vertices have been evaluated as being insignificant and discarded). Similarity metrics used are representation-specific and are given in [9].

As separate research into the capability to incorporate computer-assisted semantic classification into the indexing step, we have conducted experiments in the classification of shape data into normal/abnormal categories for osteoarthritis-related conditions by the use of artificial neural networks. (See the Results section for details.)
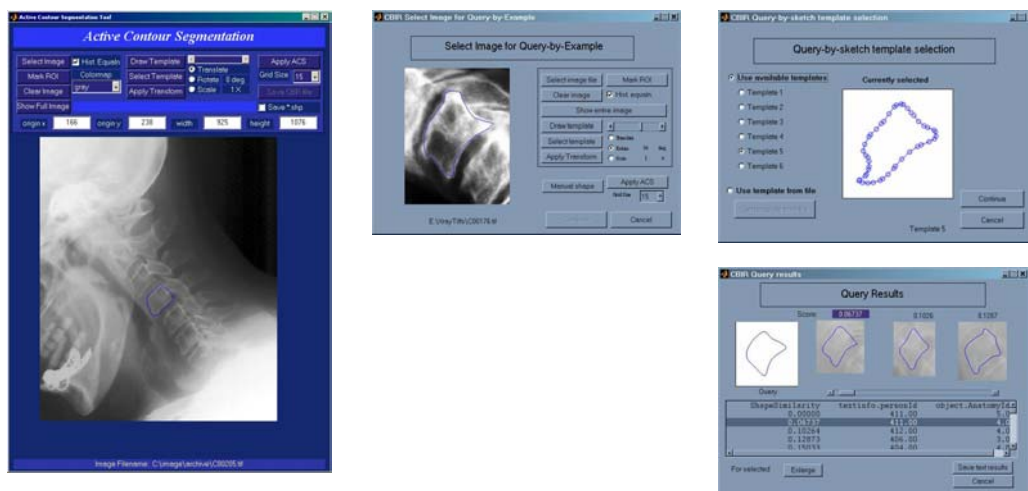

Figure 2. The prototype CBIR system. At left is the segmentation screen, for indexing. Center is the query-by-example screen; top right is query-by-sketch, and bottom right shows ranks results of one query-by-sketch.
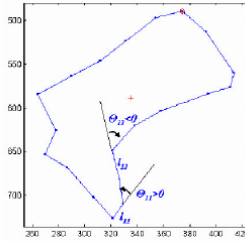
Figure 3. Vertebra shape simplified by curve evolution. Bend angles are illustrated.

3.   Results

We have conducted a preliminary evaluation on the recall and precision performance of the above four shape-representation methods on 250 C-spine and L-spine vertebra shapes. Our evaluation strategy required the development of a "ground truth" similarity ranking for any given query shape, a distance measure for calculating the distance of a similarity ranking returned by the CBIR system from the "ground truth" similarity ranking; and definitions of precision and recall calculable from our distance measure. The "ground truth" rankings were based on the Procrustes distances between 9-point segmentations of the vertebrae made by a board-certified radiologist: given any query shape, the nearest 25 shapes in the database were found and ranked in order of distance from the query shape. The approach to calculating the distance between a similarity ranking returned by the CBIR system and the corresponding "ground truth" similarity ranking, and a detailed description of the work, are given in [2]. The best results were achieved for polygon approximation by curve evolution, with 58% precision and 58% recall rates. Analysis determined that these low scores are due to incorrect ranking of similar shapes. We have since developed a new shape representation algorithm that appears to more accurately retain small variations in the vertebra shapes [10], but large-scale testing of the new approach remains to be done.

The computer-assisted segmentation method implemented in our CBIR indexing system is based on an active contour method. However, we have conducted parallel segmentation research in collaboration with H. Sari-sarraf of Texas Tech University into Active Shape Modeling (ASM). This work has illuminated both the prospects and shortcomings of that method for segmentation of these images. A significant product of this work is the implementation of a standalone version of ASM in the MATLAB language, to support continuing ASM segmentation research.

The biomedical features investigated for the feasibility of automated classification at indexing time are among those specifically identified as of interest by two workshops convened by the National Institute of Arthritis, Musculoskeletal, and Skin Diseases (NIAMS). To date our classification work has included the anterior osteophytes and disc space narrowing, for both the cervical and the lumbar spine, and has been carried out in collaboration with R. J. Stanley of the University of Missouri-Rolla.

For this work, "truth" classifications for vertebrae were obtained from a board-certified radiologist, who classified vertebrae as "normal" or "abnormal" for anterior osteophytes; vertebrae were segmented using a manual method assisted by Kirsch edge detection to guide the manual marking, spline curve fitting to the manual marks, and automatic curve sampling to create dense, uniformly-sampled segmentations. Then the vertebrae were classified as normal/abnormal using a one-hidden-layer artificial neural network trained with back propagation. For the cervical spine, 704 vertebrae were used: 352 were classified as normal and

352, as abnormal, by the radiologist. For each single vertebra shape, 32 features were derived for use in classifying the shape as normal/abnormal in the neural network. These features included radius of curvature and gradient measures along the shape, and a feature derived by mathematical morphology operations that measures how much the shape protrudes from its average local neighborhood. The vertebrae data was divided into training, validation and test sets. On the test data, an overall agreement score of 85% was achieved, relative to the radiologist "truth". For the lumbar spine, 782 vertebrae were used, with 391 classified as normal and 391 as abnormal, by the radiologist. The same feature set and test procedure was used for the lumbar spine as for the cervical spine. An overall agreement score of 71% was achieved, as compared to the given "truth", for the lumbar spine. The poorer performance of the network for the lumbar spine is possibly due to the lower contrast in these images and the resulting ambiguities in segmentation.

For disc space narrowing, again, "truth" classifications were obtained from board-certified radiologists that categorized disc space narrowing as "normal" or "abnormal": 50 cervical spine images were interpreted for disc space narrowing. Vertebra levels C3/4, C4/5, C5/6, and C6/7 were interpreted in each of these images. Two radiologists carried out the interpretation independently. Similarly, 50 lumbar images were interpreted for narrowing, at L3/4, L4/5, and L5/S1, by one radiologist. An algorithm for assessing disc spacing was developed. The algorithm operates on an image region containing two adjacent, segmented vertebrae (and the space between them). The algorithm computes a "vertebrae separator", a curve with points lying equidistant between the adjacent vertebrae boundaries; for each point on this separator, the "distance to a vertebra" is taken to be the Euclidean distance to the closest neighboring point on one of the vertebra. `Figure 4` illustrates the separator created by this algorithm. Using statistics (e.g., mean, standard deviation, minimum, and maximum) of these pointwise distances, the "disc space distance" may be characterized. Using these four statistics as input features to a one-hidden-layer neural network, and a data set of 159 vertebrae, a correct classification rate of 86% was achieved, relative to radiologist "truth". This classification work is described in references [9,11].
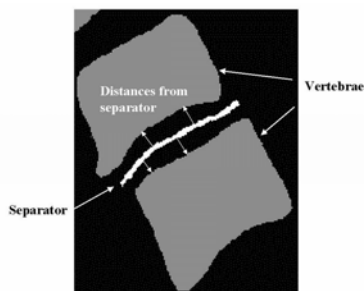

Figure 4. Equidistant separator line used for computing disc spacing

4. Current research

Segmentation research is continuing, particularly with Active Shape Modeling (ASM). Previous research identified shortcomings in the ASM method for these images. Specific failure modes have been traceable to inadequacies in (1) pose initialization of the ASM search template, (2) deformation of the search template by the ASM grayscale model, and (3) convergence to acceptable tolerances in the highly deformed corners of some vertebrae, even when convergence on the main vertebral body may be acceptable. To address these problems, respectively, a method is being researched that combines Generalized Hough Transform (GHT) segmentation, ASM with edge information used along with grayscale, and active contour segmentation (which we denote here simply by "DM" for "deformable model"). For this unified segmentation method some initial performance results are available. For this evaluation, "ground truth" was taken as manual segmentations done by the R&D team, using for each vertebra 9 boundary points marked by a radiologist as a guide for the segmentations. 100 C-spine images and 100 L-spine images were tested. For the C-spines, 80 landmark points were used, spanning the vertebral range from the bottom of C2 through C6. For the L-spines, 200 landmark points were used, spanning L1-L5. Segmentation error was calculated as the root sum square of the landmark point distances between the converged shape and the "ground truth". Overall results were that, for the C-spines, GHT+ASM+DM achieved an error less than 3 mm in 75% of the cases; for the L-spines, GHT+ASM+DM improvement achieved an error of less than 6.4 mm in 49% of the cases [12]. Work is continuing on the improvement of this unified segmentation approach. Procedures and detailed results of the analysis may be found in Zamora [13].

In the area of semantic classification by automatic methods, work on classifying cervical and lumbar spine vertebrae as normal/abnormal for subluxation and spondylolisthesis, respectively, by the use of artificial neural networks is ongoing. In the area of database organization, we currently store feature vectors in a simple linear order and perform an exhaustive search at query time. Work is underway by H. Tagare to study methods for not only organizing the vertebral shape data in a spatial data tree, but in optimizing the node structure. See [14] for an example of this type approach.

An outstanding problem in the extraction of feature vectors from the raw boundary data is, achieving a significant data reduction while simultaneously preserving the shape characteristics essential for the end use of the database. For the vertebrae, a specific key shape characteristic to be preserved is the vertebra corner shape, where osteophytes typically occur. If this shape is not adequately preserved in the feature vector calculation, the extent, or even the presence, of the osteophyte may not be detectable for retrieval. This has been a particular shortcoming observed in the polygon approximation method by curve evolution, where the original shape boundary is frequently evolved to such an extent that essential corner shape information is lost. However, the recent collaborative work published by Lee [10] has introduced a new "relevance measure" to apply to boundary points as the shape is evolved. (In the curve evolution method, progressively simpler curves are computed in an iterative manner; at each step, vertex points are tested for their shape "relevance" and discarded if this measure falls below a threshold.) In the cases tested, the new relevance measure was observed to result in preserving corner shapes to a significantly greater extent, as compared to the original method. Further, in the same work, Lee applies Fourier transformation to the bend and length coefficients commonly used to represent curve-evolved shape, and uses L2 distance as a similarity metric between the final (Fourier coefficient) shape representations. His retrieval results are discussed below.

A major problem for CBIR in general is the validation of retrieval results. For shape retrieval by a particular shape representation/similarity measure, how can we justify calling one set of results better than another? How can we compare results among different shape representations/similarity measures? For a set of 20 vertebral shapes, Lee used each shape in turn as the query shape to find similar shapes in the whole set and demonstrated that very similar (in our subjective evaluation) shapes are returned in the upper part of the returned similarity ranking in a number of cases. Yet, problematic results are observed: for example, some shapes that are close in the ranking produced by one shape query are far apart in the ranking produced when a different shape is used for the query. The validation of these query results in either a quantitative sense or with a non-quantitative approach that will justify confidence in the returned results remains a critical issue for this work. A further problem that we face is the user may actually want to retrieve shapes that are similar with respect to local shape properties (such as corners) and may, for some purposes, consider large portions of the global shape to be irrelevant.

Beyond the important issue of what we have called "engineering validation" of results, there remains the further issue of biomedical validation, for the biomedical community is the system end user. We have made a beginning in the direction of creating "truth" data sets, by the creation of a set of 9-point boundary data collected by one radiologist, and by the collection of radiologist interpretations for anterior osteophytes and disc space narrowing. There is a critical need for more extensive expert data, specifically for: (1) detailed vertebral segmentations; (2) more extensive classifications for osteoarthritis and degenerative disc disease; and (3) similarity rankings, chosen from candidate images, for a given query image. *A key requirement of these data sets is that they should be collected by multiple expert observers; only then can the performance of computerized methods relative to variance in human performance be evaluated.* (As an example of performance evaluation of computer methods relative to multiple human observers, Chalana [15] and others have developed methods for segmentation evaluation.)

5. Summary

Our CBIR prototype for cervical and lumbar spine x-ray images and associated health survey data integrates indexing and retrieval in a MATLAB/MySQL system. Manual and computer-assisted segmentation are supported to generate raw vertebral boundaries, which are stored along with feature vectors for these shapes in four alternative representations. Vertebra may be retrieved using query by image example or query by sketch and specifying which shape representation to use. The system uses similarity metrics appropriate to the shape representation, and presents the results in a graphical display that ranks the returned shapes according to similarity to the query. The prototype also supports hybrid text/image queries. Research is ongoing to improve segmentation, evaluate the effectiveness of the four shape methods, determine the capability of classifying shapes into semantic categories by artificial neural networks, and determine an optimized feature vector storage and search scheme.

6. Acknowledgements

7.   References

[1] Long LR, Antani S, Lee DJ, Krainak DM, Thoma GR. Biomedical information from a national collection of spine x-rays: film to content-based retrieval. *Proc. of SPIE Medical Imaging 2003: PACS and Integrated Medical Systems*, Vol. 5033, San Diego, CA, February 15-20, 2003 (forthcoming).

[2] Antani S, Long LR, Thoma GR, Lee DJ. Evaluation of shape indexing methods for content-based retrieval of x-ray images. *Proc. of IS&T/SPIE 15th Annual Symposium on Electronic Imaging, Storage and Retrieval for Media Databases*, Vol. 5021, January 22-23, 2003, 405-416.

[3] Antani S, Long LR, Thoma GR. A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information, *3rd Indian Conference on Computer Vision, Graphics, and Image Processing (ICVGIP '02)*, Ahemdabad, India. December 16-18, 2002, 242-47.

[4] *Plan and Operation of the second National Health and Nutrition Examination Survey 1976-80*, Programs and Collection Procedures, Series 1, No. 15, DHHS Publication No. (PHS) 81-1317, National Center for Health Statistics, Hyattsville, MD, July 1981.

[5] Hu MK. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 1962, 8, 179-187..

[6] Del Bimbo A, Pala P. Shape indexing by multi-scale representation. *Image and Vision Computing*, 1999, **17**(3-4), 245-261.

[7] Latecki LJ, Lakämper R. Shape description and search for similar objects in image databases. In R. C. Veltkamp, H. Burkhardt, and H. P. Kriegel, editors, *State-of-the-Art in Content-Based Image and Video Retrieval*, Vol. 22 of *Computational Imaging and Vision*, Kluwer Academic Publishers, 2001, 69-96.

[8] Zahn C, Roskie R. Fourier descriptors for plane closed curves. *IEEE Computer*, 1972,C-21(3), 269-281.

[9] *Biomedical Imaging Research & Development*. A report to the National Library of Medicine Board of Scientific Counselors, September 27-27, 2002. Available at http://archive.nlm.nih.gov.

[10] Lee D, Antani S, Long LR. Similarity measurement using polygon curve representation and Fourier descriptors for shape-based vertebral image retrieval. *Proc. of SPIE Medical Imaging 2003: Image Processing*, 5032, San Diego, CA, February 15-20, 2003 (forthcoming).

[11] Stanley RJ, Long LR. A radius of curvature-based approach to cervical spine vertebra image analysis. 38th Annual Rocky Mountain Bioengineering Symposium, Vol. 37, 2001, 385-390.

[12] Zamora G, Sari-sarraf, Long R. Hierarchical segmentation of vertebrae from x-ray images. *Proc. of SPIE Medical Imaging 2003: Image Processing*, 5032, San Diego, CA, February 15-20, 2003 (forthcoming).

[13] Zamora G. *Dissertation for the degree of Doctor of Philosophy*. Department of Electrical Engineering, Texas Tech University, Lubbock, TX, December 2002.

[14] Tagare H. Increasing retrieval efficiency by index tree adaption. *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries, IEEE CVPR '97*.

[15] Chalana V, Kim Y. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging*, Vol. 16, 5, Oct 1997, 642-652.