Anatomical Shape Representation in Spine X-ray Images

Sameer Antani, L. Rodney Long, George R. Thoma
Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine
Bethesda, MD 20894 USA
http://archive.nlm.nih.gov

D.J. Lee
Department of Electrical and Computer Engineering
Brigham Young University
Provo, UT 84602 USA
djlee@ee.byu.edu

ABSTRACT
Efficient content-based image retrieval (CBIR) of biomedical images is a challenging problem. Feature representation algorithms used in indexing medical images on the pathology of interest have to address conflicting goals of reducing feature dimensionality while retaining important and often subtle biomedical features. At the Lister Hill National Center for Biomedical Communications, an intramural R&D division of the U.S. National Library of Medicine, we are developing CBIR for digitized images of a collection of 17,000 cervical and lumbar spine x-rays taken as a part of the second National Health and Nutrition Examination Survey (NHANES II). The vertebra shape effectively describes various pathologies identified by medical experts as being consistently and reliably found in the image collection. A suitable shape algorithm must represent shapes in a low dimension, be invariant to rotation, translation, and scale transforms, and retain relevant pathology. Additionally, supported similarity algorithms must be useful to the intended target community, viz. medical researchers, physicians, etc. This paper describes our research in the development of such a method and a comparison with the state of the art from the literature.

KEY WORDS
CBIR, Shape, Similarity, Vertebra, Digitized x-ray

1 Introduction

There has been much research interest in recent years in developing content-based retrieval algorithms for images and video [1]. In particular, there has been growing interest in indexing images for biomedical content [2, 3, 4]. In general, manual indexing of images for content-based retrieval is a cumbersome, error prone, and prohibitively expensive task. However, due to the lack of an effective automated method, biomedical images are very often annotated manually and retrieved using a text keyword based search on the disease or pathology described in the medical expert diagnosis. A common complaint of medical professionals using such systems is that the annotations are imprecise with reference to image locations and text is often insufficient in enabling efficient image retrieval. Even such retrieval is impossible for large collections of images that have not been annotated or indexed. Additionally, the retrieval of interesting cases, especially for medical education or building atlases, is a cumbersome task. Content-based image retrieval methods developed specifically for biomedical images could offer a solution to such problems. However, for any class of biomedical images developing suitable feature representation and similarity algorithms are essential tasks. An example is in the representation of vertebra boundary shapes segmented from the images created by digitizing film x-rays of the human cervical and lumbar spines.

The Lister Hill National Center for Biomedical Communications, a research and development division of the National Library of Medicine (NLM) maintains a digital archive of 17,000 cervical and lumbar spine images collected in the second National Health and Nutrition Examination Survey (NHANES II) conducted by the National Center for Health Statistics (NCHS). Classification of the images for the osteoarthritis research community has been a long-standing goal of researchers at the NLM [5], collaborators at NCHS, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), and capability to retrieve images based on geometric characteristics of the vertebral bodies is of interest to the vertebral morphometry community [3]. Automated or computer-assisted classification and retrieval methods are highly desirable to offset the high cost of manual classification and manipulation by medical experts. Two National Institutes of Health (NIH) workshops have identified visual features of the images specifically related to osteoarthritis, but the images have never been manually indexed for these features, which include anterior osteophytes, disc space narrowing, and subluxation for the cervical spine and spondylolisthesis for the lumbar spine. We are investigating automated or computer-assisted methods by which the indexing and retrieval of the images using these particular features may be achieved, in a validated manner acceptable to the biomedical community.

As an initial step, we have implemented a modular prototype system for content-based image retrieval for a subset of the spine x-rays and health survey text data related to these x-rays [6]. The system supports retrieval based on shape similarity to a sketch or example vertebral image, as well as conventional text retrieval. The shapes are segmented using active contour segmentation with human assistance where necessary. An outstanding problem in the extraction of feature vectors from the raw boundary data is development of an effective shape representation and similarity method that simultaneously provides for data reduction while preserving the shape characteristics that are essential for the end use of the database.

In this paper we describe our work to date on developing effective shape representation algorithms and make note of the critical issues that need to be addressed in selecting necessary boundary points in medical images. Section 2 provides a brief background on the requirements of the shape representation and similarity algorithms. In Section 3 we present a brief description of the methods and results from comparative evaluations. Summary of results in presented in Section 4. We conclude with an analysis of the results, critical comments and our future research plans in Section 5.

2 Background

In our study of the spine x-rays we observe that only shape features appear promising for indexing the images, since the images are gray scale and offer very little in terms of texture for the anatomy of interest. Many shape representation and similarity techniques can be found in the literature [7, 8, 9, 10, 11, 12], to cite a few. These techniques adopt different approaches for representing shapes which can be broadly grouped under these categories:

- Shape geometry based methods: Methods that use shape properties [13] such as area, perimeter, convexity, elongation, orientation, etc.

- Invariant moments: Several forms of invariant moments are seen in the literature such as Hu invariant moments [14], generalized complex moments [15], affine moments, and Zernike moments [16, 17]. Multi-stage modification using invariant moments has yielded very good results [18].

- Polygon approximation methods: Methods that remove small variations and less significant features and then represent the curve in tangent space [19, 20, 21]. Matching is done using the turn angle function.

- Deformable shape based methods: Methods that employ elastic deformation of templates [22]. Multi-scale shape representation has been used to smooth and simplify the contours [23, 24, 25].

- Fourier transform based methods: Representing the cumulative shape boundary as a function of its normalized length [26]. Shapes or contour points have also been described in the frequency domain [27, 28].

Each shape representation method exhibits and retains different shape characteristics. In turn, this affects the quality of query and retrieval in a CBIR system. Determining the suitability of an algorithm for a CBIR system application can only be done after an evaluation of the shape methods on the shapes that populate the database. Evaluations published in the literature [16, 18] have been for shape retrieval methods applied to trademark image databases or general objects.

2.1 Shape Representation and Similarity: Challenges

Boundary data is extracted as (x, y) coordinates in the image space and needs to be represented in a form suitable for archiving, indexing, and similarity matching. A shape representation method converts a dense 2D representation of a boundary, i.e., the (x, y) coordinates of boundary points, into a form that has certain properties, which include *uniqueness*, *stability*, *geometric invariance*, and *compact representation*. In addition, the representation should retain properties of the shape that are meaningful to the application. These requirements may be extended to include matching of partial boundaries or specific local regions in the boundary.

In archiving biomedical images for content-based retrieval, one has to address conflicting goals of maintaining low feature dimensionality for efficient indexing and matching while requiring the feature representation methods to retain the subtleties in the pathology. Shapes found in biomedical images express different characteristics for different anatomy. Some have a typical shape and structure, e.g., bones, heart, lungs, etc., while others can be arbitrary, e.g., lesions on tissue. Each shape type presents its own challenges in representation. One can consider significant areas on the boundary of structured shapes which must be retained in the representation, and also, number and position of boundary points. In contrast, for arbitrary shapes, such as lesions, it is difficult to determine *significant* aspects of the boundary shape. Such shapes can have many variations while still belonging to the same semantic notion, such as a lesion. This makes it challenging to conceive a notion of similarity for content-based retrieval. Content-based retrieval relies on differences between features to determine the notion of dissimilarity. However, with medical images, variations can also occur among *semantically similar* shapes, e.g. shapes of *pathologically normal* anatomy which can vary significantly over the population. This is a problem since very often differences between normal and pathological conditions are subtle, at least in the early stages of disease. Other challenges for shape representation methods are in representing anatomical structures that tend to be very similar to each other but can be individually considered to be unique, e.g., vertebrae. This makes it challenging to select shape methods that retain sufficient information and provide the necessary notions of similarity. Additionally, it is necessary for the designers of such systems to determine how much information is necessary and what is sufficient for purposes of indexing.

In light of these requirements, we investigated several shape representation and similarity methods in the litera-

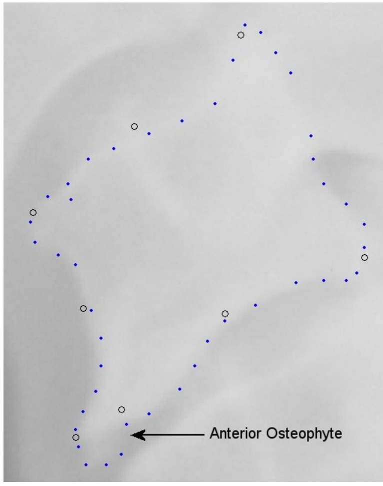


Figure 1. Vertebra image with 9-points (o) and segmented boundary data (.) superimposed.

ture and evaluated them for applicability to the vertebral shapes [29]. Based on performance, efficiency, and other reasons, Fourier Descriptor based methods and polygon approximation based methods [30] show promise. These issues, other approaches and results from our experiments are discussed below.

3 Representing Vertebral Shape

The bone morphology community frequently describes the vertebral shape using 6, 8, 9, and 10 point models [31]. These landmark points are placed at pathologically significant points on the vertebra boundary. A small portion of the NHANES II spine image collection was marked using the 9-point model by a board certified radiologist. The locations of these points, shown in Figure 1, are indicative of the pathology found to be consistently and reliably detectable in the image collection.

Our initial efforts in representing the vertebra used the 9-point model [32]. The similarity between vertebrae was determined using the Procrustes distance. However, while the 9-point model is sufficient in describing the pathology to an expert human observer, it falls short of providing a sufficiently rich description of the shape for purposes of content-based retrieval.

It was realized that the shape representation method would need to serve the dual purpose of providing a rich description of the vertebra shape while being acceptable to the end user community consisting of medical professionals, educators, researchers, and epidemiologists who typically employ coarser models. In addition, the similarity methods that support the shape description would need to be stable and be able to meet the semantic notions of the user. Stability of a method is defined as its ability to respond to changes in the query shape in a predictable and meaningful manner. The following questions need to be addressed in this respect, viz.,

i Should the number of boundary points used to represent the vertebra be fixed or variable?

ii How should pathological conditions be represented?

iii How would a similarity method be developed to allow localized, open-contour, matching?

iv How can the sensitivity of a similarity method be used to match its results to the query semantic?

3.1 Evaluating Candidate Methods

In our effort to determine answers to these questions, we implemented and evaluated candidate algorithms in various categories of shape representation methods [29]. Our implementation of these methods adopted a hybrid approach with the number of boundary points. The vertebral shapes were segmented using our implementation of Active Contour Segmentation [33] with manual intervention where necessary. This resulted in approximately 130 boundary points. To improve computational efficiency, the dimensionality of the vertebra boundary was reduced to 40 boundary points using the polygon approximation algorithm [20]. This approach is considered a hybrid approach because, unlike the case where a fixed number of points define the vertebra boundary, the positions of the 40 points are dependent on the shape of the vertebra and the algorithm. The methods were evaluated on 250 shapes comprising an equal number of cervical and lumbar vertebrae. To normalize for variations in query sketches, each shape from this data set was used as a query shape. The retrieved results were compared against the retrieval results from the Procrustes distance based matching for 9 point description of these shapes, which was considered as the ground truth. Fourier Descriptors algorithm found 54.71% of the shapes in the top 25 ranked by the Procrustes method. While this result appears unimpressive, a closer study revealed that the method had been heavily penalized for incorrect ordering of the shapes. That is, the method found the relevant shapes, but not in the same order as that determined by the 9-point method. The average displacement was 23.86. The displacement is the difference between the expected similarity rank and that determined by the retrieval algorithm. While the evaluation indicated that the algorithm was able to retrieve the relevant shapes, there were some potential unresolved issues regarding the approximated shape and its comparison with the 9-point method.

3.2 Automatic 9-Point Selection

We studied the criteria used for selected boundary points in the polygon approximation algorithm. It was realized that some critical corners were being eliminated in the approximation process which resulted in vertebral boundary shapes that were not well represented. Our modification [30] corrected the problem. Further, a method was developed that divided the vertebra shape into 7 regions. 9 points were automatically selected from these regions at locations that best matched those picked by the board certified radiologist. These 9-point representations were then indexed using the Fourier Descriptor algorithm. The method was now able to find 70% of the relevant shapes from the data set with an average displacement of 8.99, a more promising result.

3.3 Evaluating the Query-Semantic

In this approach, a medical expert classified 208 cervical and lumbar vertebra from 256 images on their pathology. Retrieval results on query shapes were scored for their relevance and accuracy. The medical expert generated a set of images for each pathology query. There were 10 queries for the cervical vertebra pathologies and 2 for normal cervical vertebra. Similarly, there were 10 pathology and 2 normal queries for the lumbar vertebra. The set of images corresponding to the expected response for these queries was considered as the ground truth.

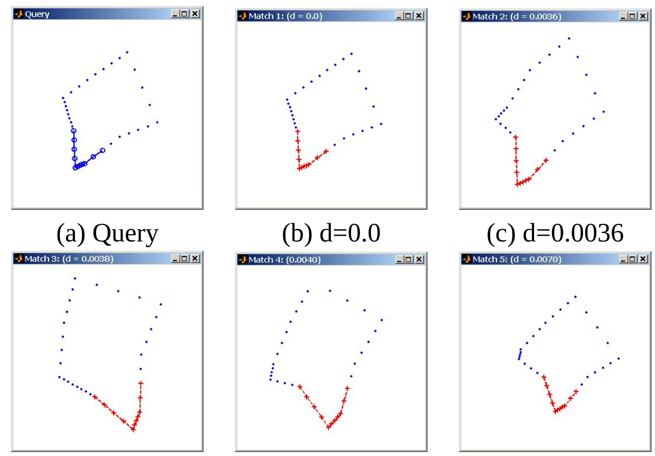
Our CBIR database currently contains 512 cervical vertebrae and 219 lumbar vertebrae. Only a subset of these were viewed by the medical expert. To account for the possibility that the CBIR system may retrieve shapes that were not viewed by the expert, we score the top 20 returned shapes for relevance. The relevance score is an indicator of how closely the returned shapes matched the query semantic. A score of 0.5 is assigned to each vertebra image in the top 20 retrieved images that is considered to depict the query semantic. A score of 0.25 is assigned to the vertebra image if it *probably* exhibits the pathology in the query. Such cases are likely to occur between different grades of pathology. A score of 0 is assigned for a poor and no match. Thus, the maximum relevance score is 10.

To compute the accuracy score, the retrieval result images were pruned to include only those that belonged to the subset of 256 vertebra images viewed by the medical expert. The top 50 ranked images from this new set are then compared with the ground-truth list for that query and an accuracy score is generated. A staggered scoring approach is used. A score of 0.5 is assigned for every image in the top 20 ranked retrieved images that is also in the ground-truth set. A score of 0.25 is assigned to every image in the next 15 ranked retrieved images (rank 21 - 35) that also appear in the ground-truth set. A score 0.1 is assigned to the next 15 ranked retrieved images (rank 36 - 50) that also appear in the ground-truth set. A score of 0 is assigned for any image that does not appear in the ground-truth set. Maximum accuracy score is 10.

The average relevance score across all queries was 5.71. The accuracy scores averaged 3.75. If the matches in the top 50 were considered, on the average, 50.6% of the shapes were found.

4 Summary of Results

A summary of the results from the above experiments are presented in Table 1. In general, the Fourier Descriptor algorithm combined with the Polygon Approximation method is able to find about 60% relevant shapes in the top 50 retrieved items. The accuracy score reveals that it finds half of them in the top 25. In some queries, the relevance scores are as high as 10. Relevance scores in 7's and 8's are not uncommon. However, lower scores are unexplained in certain queries, especially on sketches of pathologically *normal* vertebrae. We are presently investigating correlation between vertebra shape and retrieval results.



(a) Query (b) d=0.0 (c) d=0.0036

(d) d=0.0038 (e) d=0.0040 (f) d=0.0070

Figure 2. An example of partial shape matching of anterior-inferior osteophyte using Procrustes distance. (a) Query shape, (b)-(f) Top 5 matches with normalized *dissimilarity* distances.

5 Conclusions and Future Work

The above evaluations have pointed to the need for further work in shape analysis for content-based retrieval in medical images. State of the art shape representation methods are able to retrieve relevant images correctly half the time. This suggests a gap between the shape as depicted and as represented by the method. It is still unclear what aspects of the boundary data and algorithm adversely affect the results. Two important criteria in this are the number of the boundary points and their position with respect to the pathology. In case of the shape representation methods, the Polygon Approximation method with modifications suggested in [30], is very useful in reducing the granularity of the boundary data while keeping the visual aspect of the shape data intact. The Fourier Descriptors algorithm does find perceptually similar shapes and is very quick. However, as the above results show, it cannot be relied upon for accuracy. Additionally, the user may just want to concentrate on a local pathology. Querying on local regions on the shape boundary is difficult using most algorithms found in

Experiment	Evaluated To	Summarized Results
Fourier Descriptor on shape data reduced using Polygon Approx.	Similarity list using Procrustes Dist. on 9-point shape data.	On the average 54.71% of retrieved shapes matched the top 25 ranked shapes in similarity list. Average displacement = 23.86.
Fourier Descriptor on automatic 9-point data from modified Polygon Approx.	Similarity list using Procrustes Dist. on 9-point shape data.	On the average 70% of retrieved shapes matched those in the similarity list. Average displacement = 8.99.
Fourier Descriptor on shape data reduced using Polygon Approx.	Similarity list by medical professional on different grades of pathology.	Average Relevance = 5.71/10. Average Accuracy = 3.75/10 using staggered scoring. (Scores averaged over 48 cervical and lumbar queries.)

Table 1. Summary of Results

the literature, especially the Fourier Descriptors algorithm which considers the entire shape in forming the descriptors. We are now working on methods to retrieve images using Partial Shape Matching (PSM) techniques. An example is shown in Figure 2. With PSM, the user need only sketch the region of interest on the boundary in the query. We believe that this will enable users to retrieve vertebrae by sketching only the pathology of interest and thereby improve the quality of retrieval. We are in the process of developing these techniques further as well as evaluating them.

In general, there is a need for additional research in CBIR methods for medical images. The particular aspects of medical image data and usability by the target user community should be used as guides in designing algorithms for such images. We are continuing our research in developing shape methods that can improve on the above results. We are leaning toward the use of a fixed number of boundary points and working on developing better shape representation and similarity methods. Topics in focus are open-contour matching, localized shape querying and similarity algorithms. We are also working on representing other shapes, such as lesions, that do not have a well-defined structure like a vertebra.

Finally, it is necessary for various shape methods to be evaluated on a standardized a ground-truth data set. Additionally, the results must be validated by a number of medical experts to minimize the effect of inter- and intra-observer variability. We have presented our work to date in shape representation and similarity algorithms for content-based retrieval of digitized vertebral images by their shape and pathology. The retrieval results are evaluated against a machine-generated as well as a human ground-truth data set.

References

[1] S. Antani, R. Kasturi, and R. Jain. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition*, 35(4):945–965, 2002.

[2] T. M. Lehmann, C. Güld, M. O.Thies, B. Fischer, D. Keysers, M. Kohnen, H. Schubert, and B. B. Wein. Content-based image retrieval in medical applications for picture archiving and communication systems. *Proceedings of IS&T/SPIE Medical Imaging 2003: PACS and Integrated Medical Systems, Vol. SPIE 5033*, 5033, February 2003.

[3] L. R. Long, S. Antani, D. J. Lee, D. M. Krainak, and G. R. Thoma. Biomedical information from a national collection of spine x-rays: film to content-based retrieval. In *Proceedings of IS&T/SPIE Medical Imaging 2003: PACS and Integrated Medical Systems, Vol. SPIE 5033*, San Diego, CA, February 2003.

[4] H. D. Tagare, C. C. Jaffe, and J. Duncan. Medical image databases: A content-based approach. *Journal of the American Medical Informatics Association (JAMIA)*, 4(3):184–198, 1997.

[5] L. R. Long and G. R. Thoma. Landmarking and feature localization in spine x-rays. *Journal of Electronic Imaging*, 10(4):939–956, 2001.

[6] S. Antani, L. R. Long, and G. R. Thoma. A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information. In *Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing*, pages 242–247, 2002.

[7] M. Adoram and M. S. Lew. IRUS: Image Retrieval Using Shape. In *IEEE International Conference on Multimedia Computing Systems*, volume 2, pages 597–602, 1999.

[8] B. Günsel and A. M. Tekalp. Shape similarity matching for query-by-example. *Pattern Recognition*, 31(7):931–944, 1998.

[9] M. E. Hoffman and E. K. Wong. Content-based image retrieval by scale-space object boundary shape representation. *IS&T/SPIE Conference on Storage and Retrieval for Media Databases*, 3972:86–97, 2000.

[10] M. Kliot and E. Rivlin. Invariant-based shape retrieval in pictorial databases. *Computer Vision and Image Understanding*, 71(2):182–197, 1998.

[11] R. Mehrotra and J. E. Gray. Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9):23–32, 1995.

[12] A. Quddus and M. Gabbouj. Wavelet-based corner detection technique using optimal scale. *Pattern Recognition Letters*, 23(1-3):215–220, 2002.

[13] Y. H. Ang, Z. Li, and S. H. Ong. Image retrieval based on multidimensional feature properties. *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases III*, 2420:47–57, 1995.

[14] M. K Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8:179–187, 1962.

[15] Y. S. Kim and W. Y. Kim. Content-based trademark retrieval system by using visually salient feature. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 307–312, 1997.

[16] J. P. Eakins, J. D. Edwards, J. Riley, and P. L. Rosin. A comparison of the effectiveness of alternative feature sets in shape retrieval of multi-component images. In *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2001, Vol. SPIE 4315*, pages 196–207, 2001.

[17] H. H. S. Ip, D. Shen, and K. K. T. Cheung. Affine invariant retrieval of binary patterns using generalized complex moments. In *Second International Conference on Visual Information Systems (VISUAL'97)*, pages 301–308, 1997.

[18] A. K. Jain and A. Vailaya. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9):1369–1390, 1998.

[19] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. B. Mitchell. An efficient computable metric for comparing polygon shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):209–216, 1991.

[20] L. J. Latecki and R. Lakämper. 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*, volume 22 of *Computational Imaging and Vision*, pages 69–96. Kluwer Academic Publishers, 2001.

[21] L. J. Latecki and R. Lakämper. Application of planar shape comparison to object retrieval in image databases. *Pattern Recognition*, 35(1):15–29, 2002.

[22] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1996.

[23] A. Bengtsson and J. O. Eklundh. Shape representation by multiscale contour approximation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):85–93, 1991.

[24] A. Del Bimbo and P. Pala. Shape indexing by multiscale representation. *Image and Vision Computing*, 17(3-4):245–261, 1999.

[25] F. Mokhtarian and A. K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805, 1992.

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

[27] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 2 edition, 2002.

[28] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Brooks/Cole Publishing Company, second edition edition, 1999.

[29] S. Antani, L. R. Long, G. Thoma, and D. J Lee. Evaluation of shape indexing methods for content-based retrieval of x-ray images. In *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2003, Vol. SPIE 5021*, pages 405–416, Santa Clara, CA, January 2003.

[30] D.J. Lee, S. Antani, and L. R. Long. Similarity measurement using polygon curve representation and fourier descriptors for shape-based vertebral image retrieval. In *Proceedings of IS&T/SPIE Medical Imaging 2003: Image Processing, Vol. SPIE 5032*, pages 1283–1291, San Diego, CA, February 2003.

[31] M Jergas and R. S. Valentin. *Techniques for the Assessment of Vertebral Dimensions in Quantitative Morphometry*, chapter 10, pages 163–188. University of California Printing Services, 1995.

[32] D. M. Krainak, L. R. Long, and G. R. Thoma. A method of content-based retrieval for a spinal x-ray image database. *Proceedings of IS&T/SPIE Medical Imaging 2002: PACS and Integrated Medical Systems*, 4685:108–116, February 2002.

[33] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.