PATHOLOGY-BASED VERTEBRAL IMAGE RETRIEVAL

Zhiyun Xue, L. Rodney Long, Sameer Antani, George R. Thoma

U.S. National Library of Medicine, NIH

ABSTRACT

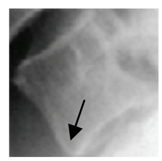
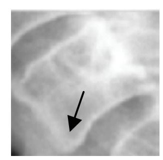
Searching for vertebrae in a large collection of spine X-ray images that are relevant to pathology is potentially important for providing assistance to radiologists and bone morphometrists. Developing appropriate methods for such searching tasks is very challenging due to the high similarities among vertebral shapes in contrast to the subtle dissimilarities that characterize the pathology. In this paper, we target two aspects of this problem: first, we develop mathematical features that can effectively represent the biomedical characteristics of interest; second, we exploit similarity learning to enhance and try to optimize the retrieval performance. We evaluate our proposed method on an expert-annotated dataset of 856 vertebrae and demonstrate its retrieval performance by precision-recall and average-precision graphs. We also demonstrate how we have integrated our method into our Web-accessible spine X-ray image retrieval system.

Index Terms— Content Based Image Retrieval, Spine X-ray Biomedical Database, Anterior Osteophytes, Partial Shape Matching

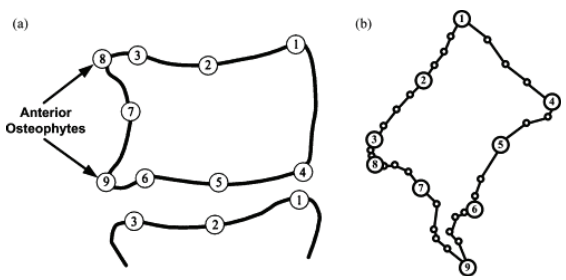
1. BACKGROUND

Osteophytes are bony spurs that grow on the vertebrae. They often develop as a person ages and are a sign of degeneration in the spine. Ostephytes can cause pain and limit movement if the bone protuberances come into contact with nerves and muscles. Anterior osteophytes (AOs) refer to those bony outgrowths that develop on the front (usually on the two anterior "corners" in the sagittal view) of the vertebrae. The degree of severity for an AO can be classified into three grades: slight, moderate, and severe, based on its appearance as described in Table 1 [1]. If present, AOs are usually visible on X-rays, and many examples can be seen in the spine X-ray image collection maintained by the National Library of Medicine. This image collection contains 17,000 cervical and lumbar spine X-ray images that were collected in the second National Health and Nutrition Examination Survey (NHANES II) by the National Center for Health Statistics. Besides images, textual data such as demographical information, anthropometric data, health history, and medical history were also collected in this survey. A subset of the collection was reviewed by a panel of medical experts, and AO was identified as one of several biomedical visual features that can be reliably detected using vertebral shape characteristics. In this paper, we focus on the task of retrieving spine vertebrae that are visually similar to the query vertebra with respect to the AO pathology feature. The challenges this task face include the high shape similarity exhibited across vertebrae and the difficulty in effectively representing the subtle differences that indicate AO pathology.

Table 1 Criteria used to classify the degree of severity for an AO.

Severity	Slight	Moderate	Severe
Feature	No narrowing or a <15° angle by the AO from the expected normal anterior face of the vertebra or protrusion length <1/5 of the vertebra width or height.	Mild narrowing or a [15–45°] angle by the AO from the expected normal anterior face of the vertebra or protrusion length (1/5–1/3) of the vertebra width or height.	Sharp/severe narrowing or a ≥45° angle by the AO from the expected normal anterior face of the vertebra or protrusion length >1/3 of the vertebra width or height.
Image			

To deal with these difficulties, we propose a new partial shape retrieval method. It targets the local area where the AO typically appear and limits the shape matching to this area: namely, the anterior corner segments along the vertebral boundary. This method uses the landmark model used by radiologists for marking AO to develop a new set of features. Although the definitions of this set of features are straightforward, they are very effective for representing the AO pathology and differentiating the subtle differences among them. Besides the effort towards designing "smart" features for better similarity matching, we also investigated methods for improving the retrieval performance through *distance measure learning*. Specifically, we implemented and tested the unsupervised learning method proposed recently in [2]. For a given similarity measure, this method learns a new similarity through graph transduction in which the similarity of a shape to the query is also influenced by the neighbors of the shape. We applied this learning approach to our *partial shape retrieval* method, where we use simple Euclidean distance for comparing the query feature vector with the feature vectors of the vertebrae in the database. We quantitatively evaluated this partial shape retrieval method with and without the distance measure learning on a “ground truth” dataset of 856 vertebrae graded by one medical expert (while caution must be taken in considering this set as a *gold standard*, it can be considered a good test bed for evaluating our algorithms). Each vertebra in the “ground truth” dataset was used as a query, and the retrieval performance was assessed using precision-recall and average precision graphs. We have also integrated our proposed partial shape retrieval method into the Spine Pathology & Image Retrieval System (SPIRS), a Web-based image retrieval system we developed for exploring the NHANES II spine X-ray database.



(a) 9-point model (b) 36-point model

Figure 1. Vertebral shape description

2. PARTIAL SHAPE RETRIEVAL

2.1. Vertebra representation

The vertebrae are first segmented from the spine X-ray images and are represented by a sequence of (x,y)-coordinate boundary points. The details of the segmentation and representation may be found in our previous papers [3,4]. To explain our rationale for developing these new partial shape features, it is useful to review the 9-point model and the 36-point model that we have used for describing vertebral shape. The 9-point model is one of the models used by radiologists to mark critical points on the vertebra boundary for diagnosis purpose. In the 9-point model, as illustrated in Figure 1(a), points 1 and 4 specify the upper and lower posterior corners of the vertebra, respectively; points 3 and 6, the respective upper and lower anterior corners; points 2 and 5 indicate the respective median points along the upper and lower vertebra edges; point 7 is the median point along the anterior vertical boundary; and points 8 and 9 mark the presence of the upper and lower anterior osteophytes, respectively. For normal vertebra (no osteophytes present), points 8 and 9 overlap with points 3 and 6, and the corner angles on the vertebral body are approximately right angles. We have developed an automatic method to identify salient points on the boundary contour and extract 9 landmark points that mimic this 9-point radiologist model. Since the 9-point model is often too sparse for representing vertebra for useful shape retrieval, we created a denser, 36-point model (Figure 1(b)) by 36 salient points including the 9 radiologist landmark points. More details on the landmark point localization can be found in [4]. In the 36-point representation, the points are numbered from 1 to 36 starting from the superior posterior corner of the vertebra and proceeding counterclockwise along the boundary. There are roughly three new boundary points between each pair of the 9 landmark points in the 36-point representation.

2.2. Partial shape feature

Because AOs typically occur on the two anterior “corners” of the vertebral outline (in the sagittal view), we take advantage of this prior knowledge and restrict the similarity matching to the interval of interest. (If the whole shape is used for similarity matching, the AO retrieval results are adversely affected by the similarities of non-AO parts of the shape.) Previously, we developed a partial shape matching method that can be applied to any user-specified segment along the boundary [5]. In this paper, we restrict the segment used for partial shape matching to target the specific localized characteristics of AOs. The new partial shape features are extracted as follows. First, within our algorithm, we specify the segment of interest (which should not be too long, or we will include some points that are not of interest for AO pathology; and should not be too short, or we may exclude some critical points that carry AO information) automatically. After inspection of a number of AOs in various individuals, we derived a heuristic method, for the 36-point model, of selecting the interval between point 17 (P_1) and point 28 (P_2) as the segment of interest for inferior AO characterization. Part of our rationale for this choice is the observation that points 17 and 28 approximate the locations of the median points along the anterior vertical boundary and the lower horizontal boundary of the vertebra (which are points 7 and 5 in the 9-point model), respectively. Similarly, we selected the interval between point 5 (P_1) and point 16 (P_2) as the segment of interest for superior AO. Given these intervals, our algorithm automatically detects the “tip” of the corner (the farthest protruding point on the specified segment), by finding the point (P_0) whose distance to the center of mass of the vertebral polygon is the greatest among all the points in the segment of interest. Please note the numbers of points between these three critical points are usually different among the vertebrae. After the selection of these three critical points (P_1, P_2, and P_0), the angle features are calculated as illustrated in Figure 2. Denote the first three points along the segment between P_0 and P_1 (P_2) that are the closest to P_0 to be P_{11}, P_{12}, and P_{13} (P_{21}, P_{22}, and P_{23}), respectively. For each pair of points (P_{1i} and P_{2i}, $i = 1,2,3$), three angles are computed as follows:

$$\alpha_{0i} = \text{the angle between } \overline{P_0P_{1i}} \text{ and } \overline{P_0P_{2i}}$$
$$\alpha_{1i} = s_{1i} * \text{the angle between } \overline{P_0P_{1i}} \text{ and } \overline{P_0P_1}$$

$$\alpha_{2i} = s_{2i} * \text{the angle between } \overline{P_0P_{2i}} \text{ and } \overline{P_0P_2}$$

Where s_{1i} (s_{2i}) is +1 if $\overline{P_0P_{1i}}$ ($\overline{P_0P_{2i}}$) is on the left of $\overline{P_0P_1}$ ($\overline{P_0P_2}$) and is -1 if $\overline{P_0P_{1i}}$ ($\overline{P_0P_{2i}}$) is on the right of $\overline{P_0P_1}$ ($\overline{P_0P_2}$). This simple scheme creates a feature set (a vector of length nine) that implicitly takes into account the key criteria identified by medical experts for rating the severity levels of the AO pathology as described in Table 1, such as the narrowing of the base, the vertebral protuberance along the vertical or horizontal boundary, and the bending directions of the corner tip, through the various angle measurements and their signs.

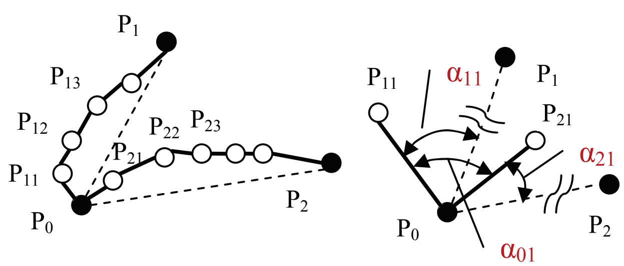


Figure 2. Partial shape features

3. DISTANCE LEARNING

After extracting the partial shape features described in Section 2, the similarity between the query vertebra and the other vertebrae is determined by the Euclidean distance between their feature vectors. To further improve the results, we applied and tested the distance learning method published recently in [2]. We were particularly interested in this method because: i) the method is general; ii) it is an unsupervised learning approach; and iii) its published performance has been validated using shape databases. The key idea of this distance learning method is to "replace the distances in the original distance space with distances induced by geodesic paths in the shape manifold" [2]. For a given similarity measure s_0 and a shape query q, the new similarity between shapes p and q is learned according to the following principle: if neighbors of p are also similar to q, then the new similarity $s(p,q)$ will be high; if neighbors of p are not similar to q, then $s(p,q)$ will be low even if $s_0(p,q)$ is high. Therefore, it is termed "context-sensitive", and the context of a shape is defined as database shapes that are most similar to it. The learning is done iteratively through graph transduction. Because the calculation of the new similarity for each query involves the computation and ranking of the distance between almost all pairs of shapes, this method is suitable for a database where the queries are known in advance so that the similarity measures can be pre-computed and saved. For an online database, given a new query, the computation will likely take too long to be practical. For more details on this algorithm and the pseudo code, see [2].

4. EXPERIMENTAL EVALUATION

We evaluated our method on a dataset of 856 shapes, consisting of 400 cervical (C3-C7) vertebrae and 456 lumbar (L1-L5) vertebrae. They were segmented from 204 spinal X-ray images, and each is represented using 36 boundary points. Both the inferior anterior "corner" and superior anterior "corner" of these vertebrae were expert-labeled with AO severity levels using the grading system described in Table 1. For inferior AO, among 856 vertebrae, 518 of them were graded as "slight", 234 of them were graded as "moderate", and 104 of them were graded as "severe"; for superior AO, 740 of them were "slight", 85, "moderate", and 32, "severe".

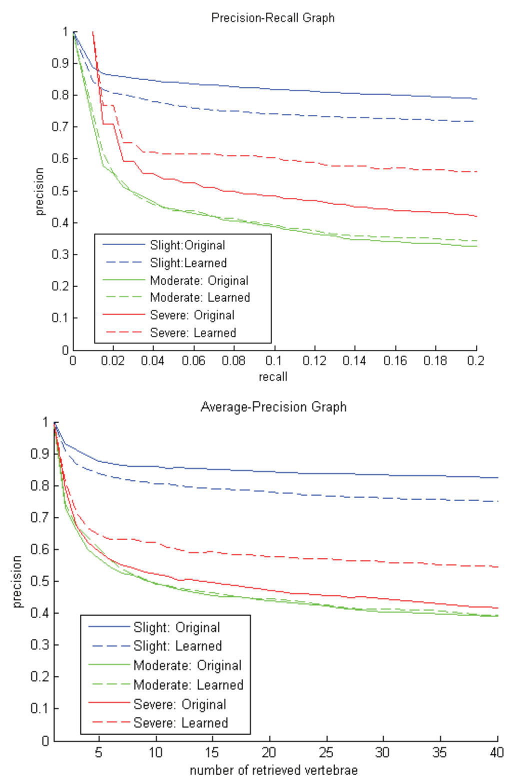


Figure 3. Retrieval performance w.r.t. inferior AO

Precision and *recall* are two commonly-used measures for evaluation of retrieval performance. Precision is defined as the number of relevant images retrieved divided by the total number of retrieved images, and recall is defined as the number of relevant images retrieved divided by the total number of existing relevant images. We considered a retrieved image to be relevant if the corner of interest in the retrieved image had the same AO severity grade as the corresponding corner of interest in the query image. In our evaluation we used both the precision-recall graph and the average-precision graph. Since the numbers of vertebrae are not balanced across all grade categories, we generated both graphs for each grade class. For example, for the vertebrae labeled "slight", we used each one of them as the query vertebra and searched for similar vertebrae in the entire dataset. We then generated the precision-recall graph and the average-precision graph by averaging recall and precision, respectively, over all the query vertebrae that are graded "slight". The results with respect to inferior AO are listed in Figure 3. This figure shows the results of the partial shape retrieval method without, and with, distance measure learning with solid lines and dashed lines, respectively. These results appear to satisfy a common criterion for many users: return relevant results in the top few of the returned images. The performance is better for "slight" grade than the "moderate" and "severe" grades. This may be due to the higher number of cases being "slight" in the dataset. For both the inferior AO and superior AO, the learning method improves the results of the "severe", but not the "slight" class (performance is actually worse) or the "moderate" class (comparable performance). Although the application of the distance learning method does not boost the performance in all severity categories, the fact that it enhances the results in the "severe" class may be significant, since this class is presumably of most biomedical interest. The results with respect to superior AO follow the same patterns as those for the inferior AO shown in Figure 3.

To perform image retrieval on the NHANES II database, we have been developing the Spine Pathology & Image Retrieval System (SPIRS), a working proof-of-concept CBIR system that allows users to pose hybrid visual and text queries [1]. Characteristics of SPIRS include: i) distributed architecture and open standards conformance; ii) automatic shape extraction and representation; iii) feature indexing for fast retrieval; iv) feature description options for comparative vertebral shape retrieval performance; and v) serving as a platform for vertebral content-based image retrieval (CBIR) algorithm evaluation. After our quantitative testing using a large ground truth dataset, we incorporated our partial shape retrieval method into SPIRS which has 4513 indexed vertebrae. The retrieval results of one example visual query (for superior AO) are given in Figure 4. Though systematic subjective evaluation has not been carried out yet, preliminary examination by three research engineers was conducted. Each engineer randomly selected 10 query vertebrae and evaluated each of the top 6 returned results on a subjective similarity scale from 1 (no or little similarity) to 5 (high similarity); the average score (the score for the first returned result was not counted since it is always the same as the query for this algorithm) across the engineers was 3.75, which indicates the retrieval performance is good with respect to visual similarity for most of the examined cases.

5. CONCLUSION

In summary, we present our latest work in developing and evaluating a localized shape querying and retrieving algorithm for vertebral shapes. Our proposed method focuses on a small interval on the vertebral outline pertinent to the pathology of AO. We propose a new, intuitive and effective feature set that appears to capture critical characteristics used by radiologists for grading AO pathology; we also implement and test a new unsupervised similarity learning approach. Our quantitative evaluation is based on an expert "ground truth" dataset established using a standard grading system for vertebral osteophytes. We have incorporated our method into SPIRS, our online spine x-ray image retrieval system and begun evaluation of retrieval performance in that system.

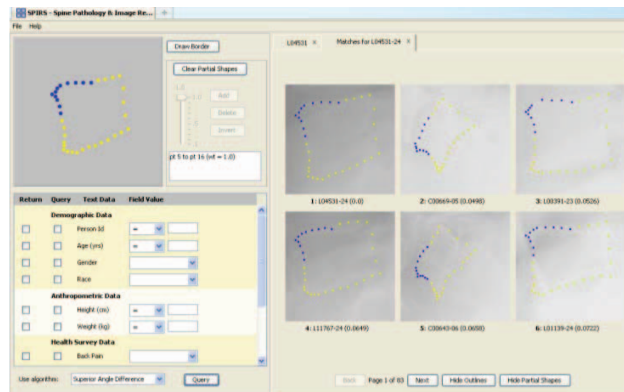


Figure 4. Retrieval results shown in SPIRS (query for superior anterior corner segment)

ACKNOWLEDGEMENT

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

REFERENCES

[1] W. Hsu, S. Antani, L. R. Long, L. Neve, and G.R. Thoma, "SPIRS: a Web-based Image Retrieval System for Large Biomedical Databases," *Int. J. Medical Informatics*, vol. 8, no. 1, pp. S13-S24, 2009.

[2] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu, "Learning Context-Sensitive Shape Similarity by Graph Transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.32, no. 5, pp. 861-874, 2010.

[3] X. Xu, D. J. Lee, S. Antani, and L. R. Long, "Localizing Contour Points for Indexing an X-ray Image Retrieval System," in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, New York, Jun. 26–27, 2003, pp. 169–174.

[4] S. Antani, L. R. Long, and G. R. Thoma, "Applying Vertebral Boundary Semantics to CBIR of Digitized Spine X-Ray Images," *Proceedings of SPIE,* vol. 5682, pp. 98-107, January 2005.

[5] X. Xu, D. Lee, S. Antani, and L. R. Long, "A Spine X-Ray Image Retrieval System Using Partial Shape Matching," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 100-108, 2008.