Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Vis. Commun. Image R. 15 (2004) 285–302
JOURNAL OF VISUAL Communication & IMAGE Representation
www.elsevier.com/locate/jvci

Evaluation of shape similarity measurement methods for spine X-ray images

Sameer Antani[a,*], D.J. Lee[b], L. Rodney Long[a], George R. Thoma[a]

[a] *Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD 20894, USA*
[b] *Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, USA*

Received 5 June 2003; accepted 5 April 2004
Available online 23 August 2004

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 prototype 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 low dimension, be invariant to rotation, translation, and scale transforms, and retain relevant pathology. Additionally, supported similarity algorithms must be useful in retrieving images that are relevant to the queries posed by the intended target community, viz. medical researchers, physicians, etc. This paper describes an evaluation of two popular shape similarity methods from the literature on a set of 250 vertebra boundary shapes. The polygon approximation method achieved a performance score of 55.94% and bettered the Fourier descriptor algorithm which had a performance score of 46.96%.
© 2004 Elsevier Inc. All rights reserved.

* Corresponding author.
E-mail address: antani@nlm.nih.gov (S. Antani).

1047-3203/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2004.04.005

Keywords: Content-based image retrieval; Medical image databases; Shape representation; Performance evaluation

1. Introduction

Research into content-based image retrieval (CBIR) algorithms (Antani et al., 2002) has attracted much interest in recent years. In particular, there has been growing interest in indexing images for biomedical content (Lehmann et al., 2003,; Long et al., 2003; Tagare et al., 1997). Manual indexing of images for content-based retrieval is cumbersome, error prone, and prohibitively expensive. However, due to the lack of effective automated methods, biomedical images are typically annotated manually and retrieved using a text keyword-based search. A common drawback of 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 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, it would be necessary to develop suitable feature representation and similarity algorithms. An example is the collection of vertebra boundary shapes segmented from the digitized images of film X-rays of the human cervical and lumbar spines.

A relevant application of CBIR in the clinical domain has been carried out and reported by Kuo et al. (2002). In this work, the image data consisted of breast sonograms. A total of 263 breast tumors were represented; these were classified into 129 malignant and 134 benign cases by medical experts. The images were indexed by manually locating regions-of-interest (ROIs) and then using computer algorithms to calculate image features within each ROI. Three alternative texture features were used: contrast, covariance, and dissimilarity. After the indexing, the images were presented to the CBIR system and a ranking of similar images, using one of the texture measures and a weighted Euclidean distance, was produced. The system interfaced to a computer-aided diagnosis (CAD) function that used the top K similar images (have had known, ''truth'' classifications), to classify the input tumor image as malignant or benign. The reported sensitivity of this work was 94%, and specificity 90%. In addition, an experimental system for electronic patient records with CBIR capability is being developed and put to use at a hospital in Brazil (Traina et al., 2003). CBIR is also being applied to retrieval of brain MRI images (Traina et al., 2003). These suggest the type of applications that CBIR might be put to in the clinical world. Such systems can also enable medical researchers, and educators retrieve images relevant to a particular pathology.

The Lister Hill National Center for Biomedical Communications, a research and development division of the US 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 (Long and Thoma, 2001), and collaborators at NCHS and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). Also, the capability to retrieve images based on geometric characteristics of the vertebral structures is of interest to the vertebral morphometry community. Automated or computer-assisted classification and retrieval methods are highly desirable to address these goals since such methods offset the high cost of manual classification and manipulation by medical experts. Medical experts 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 for the cervical and lumbar spine, subluxation for the cervical spine, and spondylolisthesis for the lumbar spine. We are investigating automated or computer-assisted methods that use image features for indexing and retrieval of these images in a manner acceptable to the biomedical community.

To provide the intended users with an effective tool that enables retrieval of vertebra shapes significant to the pathology indicated in the query, we evaluated several shape representation methods from the literature (Antani et al., 2003). The methods included geometric shape properties, invariant moments, token description, polygon approximation, and Fourier descriptors. We also considered the curvature scale space (CSS) technique used for describing shapes in MPEG-7 architecture (Bober, 2001). The Gaussian smoothing step in the CSS process, however, removed the characteristics of the pathology in the vertebrae which are critical to the effective retrieval of these shapes. Hence, the method was considered undesirable and not considered for evaluation. Unlike methods designed for trademark or other shape data retrieval found in the literature, medical image retrieval methods need to find shapes that are specific to the query. Additionally, the shape representation cannot approximate the shape data since many pathologies are indicated by subtle variations in shape. Desirable characteristics and challenges of CBIR of medical images are further discussed in Section 2.1.

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 associated with these X-rays (Antani et al., 2002). 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 reliable indexing and retrieval.

In this paper, we present an evaluation of two shape representation algorithms selected from our earlier evaluation (Antani et al., 2003) and make note of the critical issues that need to be addressed for further development of shape similarity methods for 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. The data and evaluation strategy are described in Section 4. We present the results, analysis, and critical comments in Section 5 and conclude with our future research plans in Section 6.

2. Background

In our study of the spine X-rays, examples shown in Fig. 1, 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. The many shape representation and similarity techniques found in the literature (Adoram and Lew, 1999; Günsel and Tekalp, 1998; Hoffman and Wong, 2000; Kliot and Rivlin, 1998; Mehrotra and Gray, 1995; Mokhtarian and Mackworth, 1986; Quddus and Gabbouj, 2002) adopt different approaches for representing shapes. These techniques may be grouped under the following categories:

- Shape geometry based methods. These use shape properties (Ang et al., 1995) 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 (Hu, 1962), generalized complex moments (Kim and Kim, 1997), affine moments, and Zernike moments (Eakins et al., 2001; Ip et al., 1997). Multi-stage modification using invariant moments has yielded very good results (Jain and Vailaya, 1998).
- Polygon approximation methods. Methods that remove small variations and less significant features and then represent the curve in tangent space (Arkin et al., 1991; Latecki and Lakämper, 2001, 2002). Matching is done using the turn angle function.

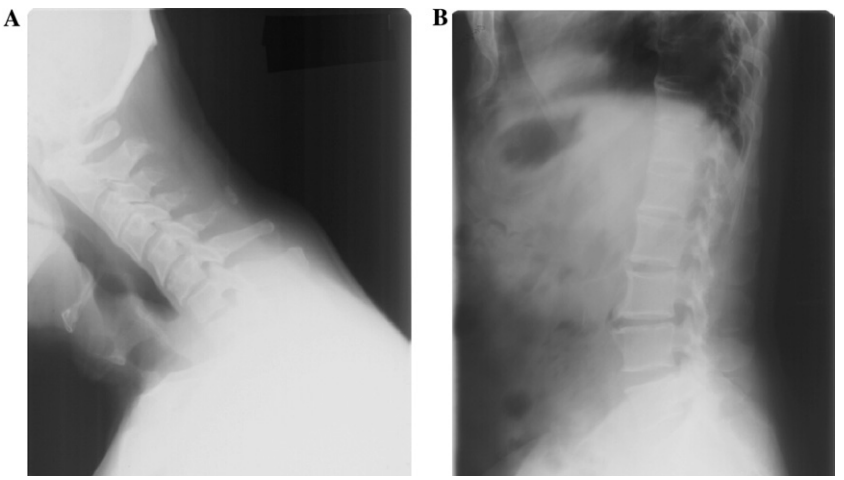

Fig. 1. Examples of (A) cervical and (B) lumbar spine X-ray images.

- Deformable shape based methods. Methods that employ elastic deformation of templates (Jain et al., 1996). Multi-scale shape representation has been used to smooth and simplify the contours (Bengtsson and Eklundh, 1991; DelBimbo and Pala, 1999; Mokhtarian and Mackworth, 1992).
- Fourier transform based methods. Representing the cumulative shape boundary as a function of its normalized length (Zahn and Roskie, 1972). Shapes or contour points have also been described in the frequency domain (Gonzalez and Woods, 2002; Sonka et al., 1999).

These shape representation methods exhibit and retain different shape characteristics. In turn, this affects the reliability 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 particular shapes that populate the database. Evaluations published in the literature (Eakins et al., 2001; Jain and Vailaya, 1998) have been for shape retrieval methods applied to trademark image databases or general objects (e.g., silhouette shapes of a hammer, hand, pliers, etc.)

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 follow a typical shape and structure, e.g., for 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 lengths of 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 exhibit many variations while still belonging to the same semantic notion, such as a lesion. This makes it challenging to conceive a measure for similarity in 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 *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 individually be considered 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 have investigated several shape representation and similarity methods in the literature and evaluated them for applicability to the vertebral shapes (Antani et al., 2003). Based on performance, efficiency, and other factors, Fourier descriptor based methods and polygon approximation based methods (Lee et al., 2003) show promise. These are described below. Results from our evaluation experiments and analysis follow.

3. Selected methods

In this section we briefly describe the two selected methods, followed by a description of the test data set and the evaluation strategy.

3.1. Polygon approximation

Polygon approximation or curve evolution is a process that eliminates insignificant shape features and reduces the number of data points. The resultant representation is one that uniquely describes the shape. The approximated curve is then converted to tangent space for similarity measurement.

- Curve evolution. Curve evolution is used to reduce the influence of noise and to simplify the shapes by removing irrelevant and keeping relevant shape features. This is achieved by iteratively comparing the relevance measure of all vertices on the polygon. Higher relevance value means that the vertex has larger contribution to the shape of the curve. For each iteration, the vertex that has the lowest relevance measure is removed and a new segment is established by connecting the two adjacent vertices. The relevance measure is calculated as

$$K(s_1, s_2) = \frac{\beta(s_1, s_2) l(s_1) l(s_2)}{l(s_1) + l(s_2)}, \tag{1}$$

  where β is the turn angle and l is the normalized length for shapes s_1 and s_2. The relevance measure is in direct proportion to the turn angle and the length of the curve segment.
- Tangent space. The smoothed curve is represented by the turn function, which is the turn angle as a function of the normalized length. Representing shape in tangent space meets the invariance requirements for shape-based retrieval. It is translation invariant because the turn angles and length do not contain information about the shape location. Use of normalized length provides for scale invariance. In case of rotation, the turn function is shifted vertically, and it is translated horizontally when there is a shift in starting point.

- Similarity measurement. The distance (dissimilarity) between two turn functions Θ_A and Θ_B for shapes A and B can be measured as

$$\delta_2(A,B) = \|\Theta_A - \Theta_B\|_2 = \sqrt{\left(\int_0^1 |\Theta_A - \Theta_B|^2 \mathrm{d}s\right)}$$
$$= \sqrt{\min_{\theta \epsilon R, t \epsilon [0,1]} \left(\int_0^1 |\Theta_A(s+t) - \Theta_B(s) + \theta|^2 \mathrm{d}s\right)}. \quad (2)$$

To measure the distance, the two turn functions must be aligned first. In most cases, the turn functions are not identical because of difference in shape. The alignment can only be achieved through minimizing the distance while shifting one turn function. In other words, the distance between two turn functions is obtained by performing a two-dimensional search to find the minimum distance. Another approach is to reduce the search to one dimension by calculating the best value of θ (Arkin et al., 1991). The best value of θ is a function of length shift t in the X axis that minimizes

$$\begin{aligned} h(t,\theta) &= \int_0^1 |\Theta_A(s+t) - \Theta_B(s) + \theta|^2 \mathrm{d}s, \\ \text{where } \theta'(t) &= \int_0^1 (\Theta_A(s+t) - \Theta_B(s) + \theta)\mathrm{d}s \\ &= \alpha - 2\pi t, \\ \text{where } \alpha &= \int_0^1 \Theta_B(s)\mathrm{d}s - \Theta_A(s)\mathrm{d}s. \end{aligned} \quad (3)$$

- Enhancements. While Eq. (1) works well for describing shapes at different levels of detail, it starts losing the significant pathology of the shapes as the number of data points decreases. Fig. 2 shows the results of curve evolution on a vertebra outline using the method described above. The original shape contour has 172 data points as shown in Fig. 2A. It was reduced to 30 and 20 points as shown in Figs. 2B and C, respectively. The remaining vertices do not contribute to representing the original shape correctly.

An enhanced relevance measure equation was developed to remove short and straight line segments so that the critical points can be detected and preserved

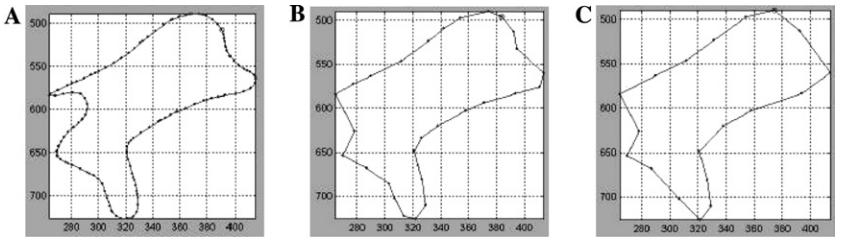

Fig. 2. (A) Original contour with 172 data points, reduced to (B) 30 points and (C) 20 points using original method.

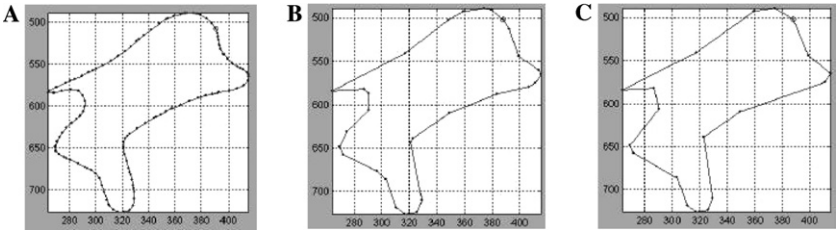

Fig. 3. (A) Original contour with 172 data points, reduced to (B) 30 points and (C) 20 points using enhanced method.

(Lee et al., 2003). It removes the vertices that have short length and/or that have turn angles close to 180° (straight line). This is expressed as

$$K(s_1, s_2) = \frac{|(\beta(s_1, s_2) - 180)|l(s_1)l(s_2)}{l(s_1) + l(s_2)}, \tag{4}$$

where β is the turn angle and l is the normalized length for shapes s_1 and s_2. The relevance measure is in direct proportion to the turn angle and the length of the curve segment. Corresponding results from applying this method are shown in Fig. 3.

3.2. Fourier descriptors

The position of a point on a closed contour is a periodic function. Thus, the Fourier series may be used to approximate the contour. The resolution of the approximating contour is determined by the number of terms in the Fourier series. Since simple operations such as scaling and translation are related to simple operations of the boundary's Fourier descriptors, they are attractive for use with boundary matching (Zahn and Roskie, 1972). Rotation however requires the bend angle function to be computed.

- Bend angle. The bend angle versus normalized length function was calculated so that the shape representation meets the invariance requirements. The bend angle is calculated such that a clockwise turn gives a negative angle whereas a counter clockwise turn gives a positive angle. This method represents a closed polygon curve C (m vertices) as $\Theta(t)$, i.e., the bend angle as a function of length t. The parameter t is the normalized accumulated length. Because it does not contain orientation information, this representation meets the rotation invariance requirement. Normalized length makes it independent of the polygon size. Starting point shift invariance requirement is satisfied by the shift invariance property of the power spectrum.

  The Fourier expansion of $\Theta(t)$ is expressed as

$$\Theta(t) = \mu_0 + \sum_{n=1}^{\infty}(a_n \cos nt + b_n \sin nt). \tag{5}$$

In Eq. (5), a_n and b_n are coefficients for each frequency component. The power spectrum of the bend angle function is invariant to the shift in length (t in this case). Because of this property, Fourier descriptors on a bend angle function meet all invariance requirements for shape-based retrieval. The similarity between shapes is the normalized difference between the Fourier descriptors of the shapes. The lower the difference, greater is the similarity.

4. Evaluation strategy

4.1. Data

The boundary data that forms the test set consists of vertebra outlines segmented from the digitized NHANES II spine X-ray images. The contrast of the digitized NHANES II spine X-rays is fairly poor, making automated segmentation a challenging task. We are exploring active contour models (Kass et al., 1988) and active shape modelling (Cootes and Taylor, 2001) techniques for automated segmentation. For purposes of the evaluation we have adopted a computer-assisted manual segmentation approach by fitting splines to manually-identified coarse boundary. Fig. 4 shows a few sample manually segmented shapes from the data set.

The data set consists of 250 segmented vertebrae which include 25 each for cervical C3–C7 and lumbar L1–L5 vertebrae. These data were used to create ground truth necessary for the experimentation. The ground truth is based on the coarse radiologist-marked 9-point data defining the vertebrae outline (Krainak et al., 2002), shown in Fig. 5A. The 9-point model was chosen because of it is a model used in the vertebra morphometry community, and because they were marked by an expert. The points indicate the following:

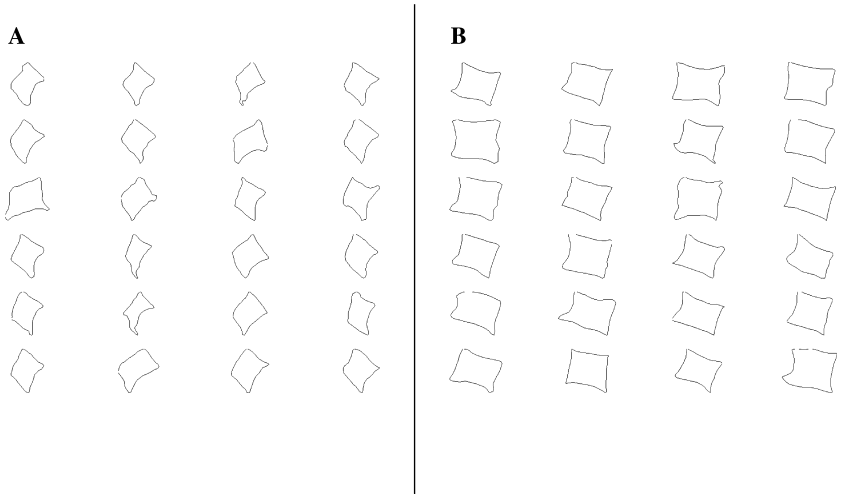

Fig. 4. Examples of segmented (A) cervical and (B) lumbar vertebra shapes in sagittal view.

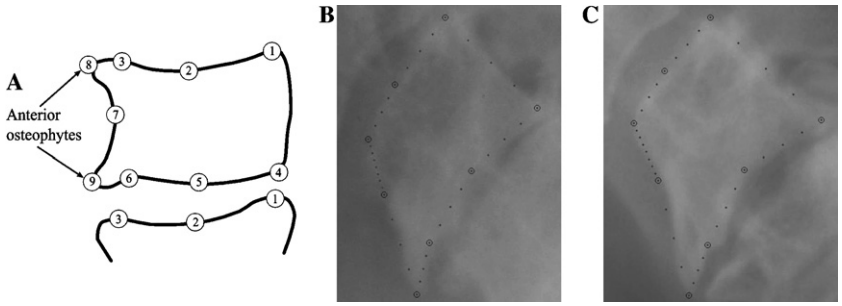

Fig. 5. (A) Radiologist marked 9-points, (B) Example query image with 36 points superimposed. Nine points marked with 'o' (C) Procrustes matching of the most similar different image with 36 points superimposed.

- Points 1 and 4 mark the upper and lower posterior corners of the vertebra, respectively.
- Points 3 and 6 mark the upper and lower anterior corners of the vertebra, respectively.
- Points 2 and 5 are the median along the upper and lower vertebra edge in the sagittal view.
- Point 7 is the median along the anterior vertical edge of the vertebra in the sagittal view.
- Points 8 and 9 mark the upper and lower anterior osteophytes. If osteophyte(s) are not present on the vertebra, then these points coincide with points 3 and 6, respectively.

The 9-point shapes are made dense to 36 points using linear interpolation. Three points are interpolated between every pair of original 9 points. Additionally, the points are re-ordered so as to better represent the actual shape. The new ordering of boundary points is points 1, 2, 3, (8), 7, (9), 6, 5, and 4. Points 8 and 9 are parenthesized since a vertebra without any osteophytes, would not exhibit points 8 and/or 9. In cases where the osteophyte(s) are absent, the density of points between the adjacent pair is increased to keep the number of points constant. Such an approach is considered valid, since an osteophytes are bony "growths" which could be considered to grow out between points 3 and 7 or 7 and 6. An example of this case is seen in Figs. 5B and C, where 7 points are interpolated between points 3 and 7 due to a lack of the anterior superior osteophyte.

4.2. Ground truth

The ground truth data set is an ordered list of similar shapes from the data set for every shape in the data set. The ground truth data is generated using the Procrustes similarity metric. The Procrustes metric finds the best fitting match between two shapes and is represented by Eq. (6), where (x, y) and (x', y') are n boundary point coordinates of shapes X and X'.

$$P = \sum_{i=1}^{n} \left\| \begin{bmatrix} S \cdot \cos\theta & -\sin\theta & T_x \\ \sin\theta & S \cdot \cos\theta & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}_X - \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}_{X'} \right\|^2. \quad (6)$$

The matching process translates shape X by (T_x, T_y) such that the center of gravity of the two shapes coincide. Next, the shape X is scaled by S and rotated by θ for the minimum sum of squared distances between the boundary points of the two shapes. The $-$ sign indicates the Euclidean distance measure between two 2D points. Results from a matching is shown in Figs. 5B and C. This method finds the closest distance between two shapes. When applied to the 36 point model, it can be considered as the closest semantic distance between the shapes. This metric was chosen over a manually marked ground truth shape because when determining similarity, human readers tend to focus on local shape features such as elongated osteophytes, flattening of the vertebra, etc. Marking the ground truth in this manner would require normalization of the measure of shape similarity as perceived by different users. Using the Procrustes distance overcomes this hurdle while allowing a fair evaluation of the methods.

4.3. Performance metrics

The ground truth data has 36-point shape descriptions. For the purposes of the evaluation, dense segmented vertebra boundary shapes are reduced from approximately 150-point descriptions to 36 points using the enhanced polygon approximation method. These reduced density shapes are then used for similarity matching in the tangent space using polygon approximation similarity matching and using bend angle with Fourier descriptors.

Each shape method determines the similarity distance between every pair of shapes in the data set. The performance of these methods can only be compared if the similarity scores are normalized. However, this is challenging since each method uses a different distance measure. To circumvent this problem and still achieve the desired goal of comparing the methods, we use the similarity rank assigned by the method to each shape. The similarity rank of each shape when compared to a query shape from the data set using a test method is compared with the similarity rank of that shape using the Procrustes distance metric. The mismatch (displacement) in the rank, if any, is used to quantify performance. The greater the displacement, poorer the performance.

To illustrate the comparison, consider the example shown in Table 1. Here, the columns indicate the shapes ranked in decreasing order of similarity. The column on the left is the output of a shape method and that in the middle is the ground truth. The numbers, v3, v5, etc., represent the shapes. The performance of the shape method is computed by determining the mismatch between the two lists, which is described in the rightmost column of the table.

The performance P is given by Eq. (7).

$$P = \frac{T_d - D_m}{T_d} \times 100\%. \quad (7)$$

Table 1
Example of comparing similarity rankings

Shape method	Ground truth	Mismatch displacement
v3	v3	None
v5	v99	v5 has moved 'up' 2 ranks
v99	v21	v99 has moved 'down' 1 rank
v18	v5	v18 has moved 'up' 1 rank
v21	v18	v21 has moved 'down' 2 ranks

Here, D_m is the total mismatch distance for a particular method over all shape queries. It is computed as the mismatch displacement, d_m, for each shape query summed over all N shapes in the database, i.e., $Dm = \sum^{N} d_m$. The maximum possible displacement, T_d, is computed as follows:

$$T_d = 2 \times \sum_{k=0}^{\left|\frac{N}{2}-1\right|} N - (2k+1). \tag{8}$$

This follows from the observation that in a set of N shapes the maximum possible displacement for the first shape is $(N-1)$. The next shape has a maximum displacement of $(N-3)$. This is because the first shape displaces a shape with its maximum displacement and has a fixed position. This reduces the list size to $M = (N-2)$. The second shape can now have a maximum displacement of $(M-1)$, which is $(N-3)$. The third shape has a maximum displacement of $N-5$, and so on. Since there are 250 shapes in the data set, the maximum displacement T_d is computed to be 31,250.

In this evaluation, we have not used standard precision and recall performance metrics because it is being conducted on a finite number of shapes. As such, a false negative at one similarity rank position will result in a false positive at the same rank and similar pair of false readings where these expected shapes appear. This results in nearly equal precision and recall measures. Instead, computing the displacement and its direction is a better way of evaluating performance. Creating a window with, for example, the top 25 similar shapes (Antani et al., 2003), could be used as an approach to apply precision and recall metrics. There is, however, no clear methodology to use in calculating window size that can provide confidence that a complete picture of algorithm performance can be obtained. Using the charts presented in Figs. 6 and 8, one can compute the performance at varying window widths. In addition, by studying the displacement direction, the shape, and the similarity algorithm one could gain greater insight into their relationship.

5. Results and analysis

This section describes the results for the evaluation experiments described above. The performance of the methods is presented using three experiments. In the first

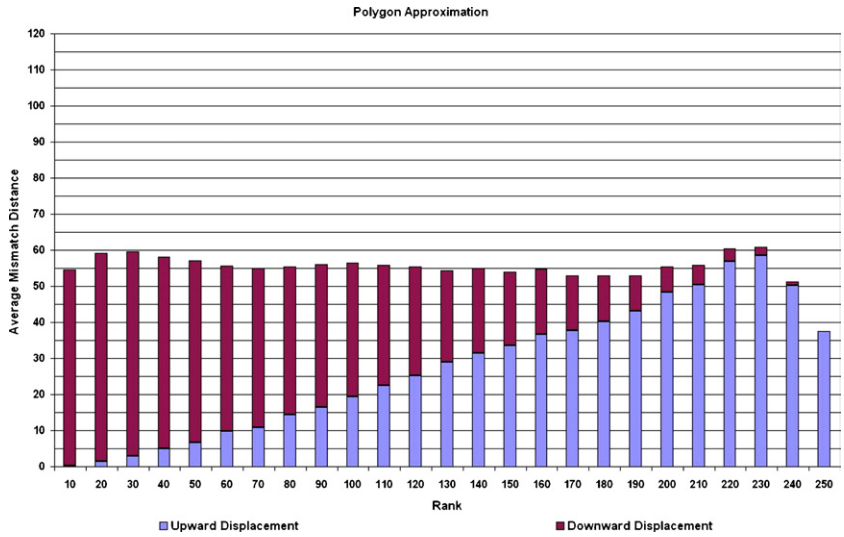

Fig. 6. PA method: mismatch distance in the upward and downward direction averaged over 250 queries. Bins are similarity rank intervals of 10.

experiment, the performance is evaluated by comparing the similarity rank of a shape retrieved using a test method against the ground truth. Each shape of the data set is used as a query and 249 similar shapes, other than the query shape, are ranked. The upward and downward mismatch distance is then computed for every query shape. The mismatch distance is averaged over similarity rank interval of 10. The results for the polygon approximation (PA) and Fourier descriptor (FD) shape similarity method are presented in Figs. 6 and 8. The charts show the amount of upward and downward displacement as a proportion of the overall mismatch displacement. For shapes ranked higher in the ground truth, most of the displacements are inevitably downward movement. For shapes ranked lower, more upward movement and less downward movement can be seen. By comparing the two sets of data, the FD method can be seen to have larger average displacement then the PA method. Figs. 7 and 9 show the accumulated displacement for the PA and FD methods, respectively. The total averaged accumulated displacement in one direction over 250 queries is 8287.6 for the FD method and is 6883.4 for the PA method. This indicates that the FD method incurs larger displacement and hence is less accurate than the PA method. It is important to note that since the queries are exhaustive over a closed set, the total upward displacement is the same as the total downward displacement.

A second experiment was to determine the retrieval precision of the methods. For this, the shapes were compared with the ground truth over an interval regardless of the displacement within the interval. The FD method retrieved 59.2% of the 100 top ranked shapes and 80.34% of the top 200 top ranked shapes accurately. Again, the PA method shows higher retrieval precision returning 62.2 and 86.3% accurate matches, respectively. Obviously, 100% of the shapes are accurately found

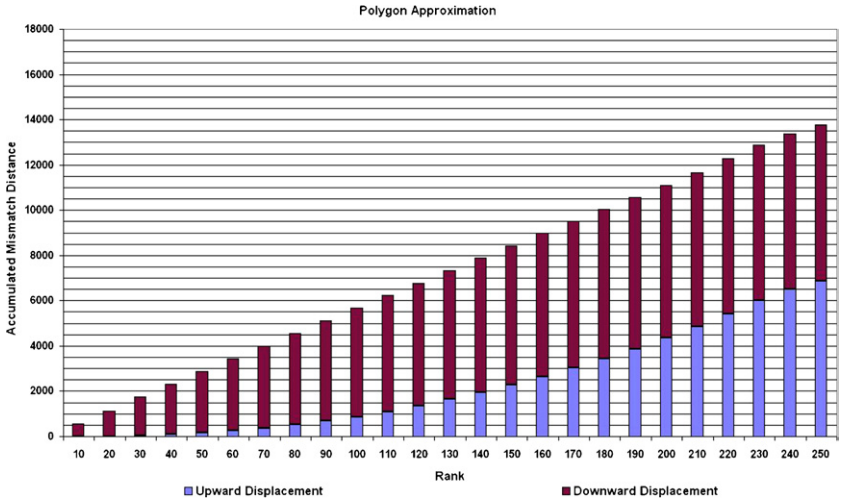

Fig. 7. PA method: mismatch distance accumulated over 250 queries.

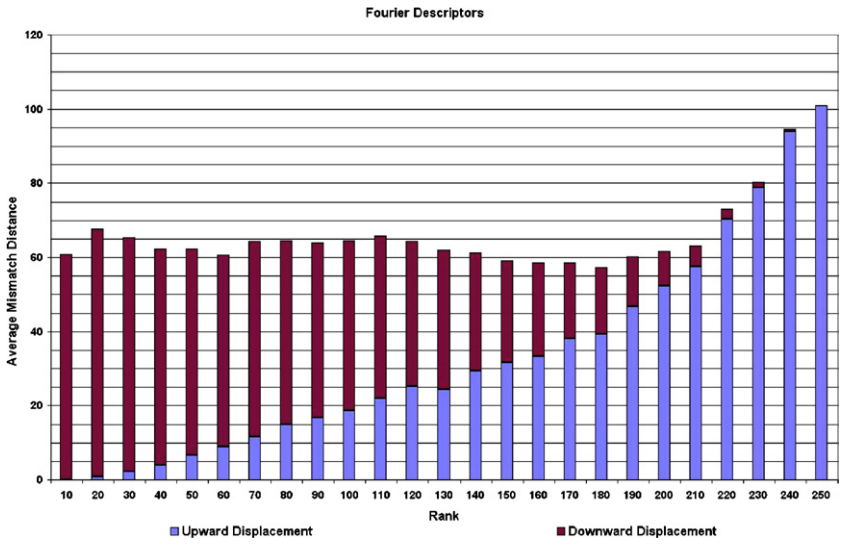

Fig. 8. FD method: mismatch distance in the upward and downward direction averaged over 250 queries. Bins are similarity rank intervals of 10.

over the entire data set. This experiment models the trade-off between sensitivity and specificity in an ROC curve. The intent is to present the researcher with a sense of the performance of the method over a range of intervals. Fig. 10 shows a comparison of the precision rates for the two methods. The PA method has an overall higher retrieval precision rate. This experiment is useful in separating

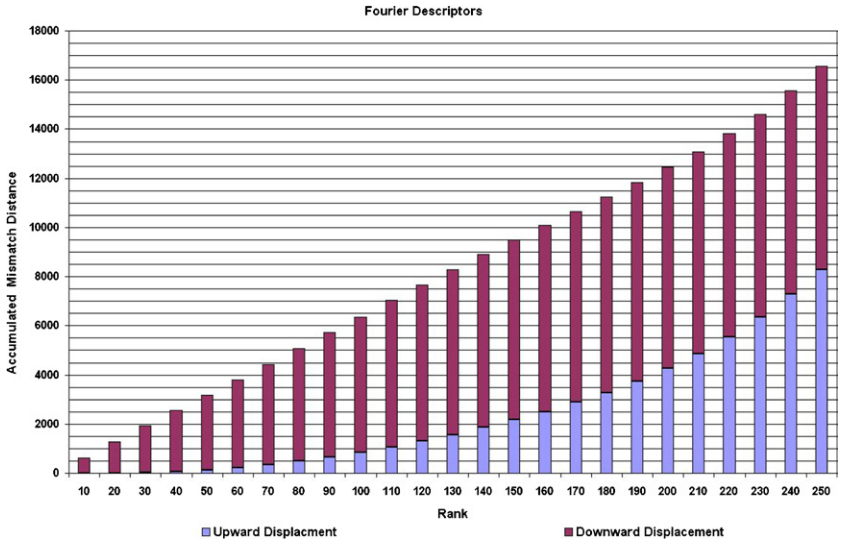

Fig. 9. FD method: mismatch distance accumulated over 250 queries.

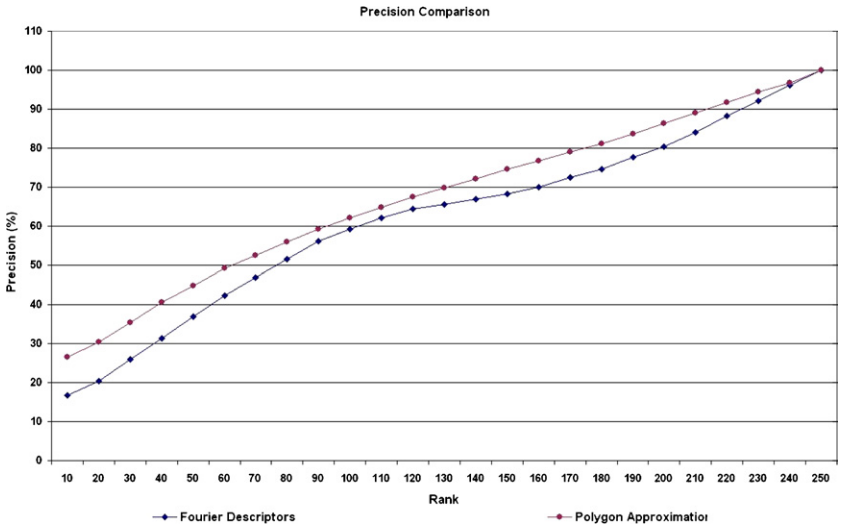

Fig. 10. Comparison of retrieval precision of PA and FD methods.

the performance of the methods with very similar mismatch displacement values. It also allows the researcher to determine the expected retrieval performance for a given interval.

Finally, in the third experiment, a performance score, as described in Eq. (7) is computed for each method. As shown in Table 2, the performance scores of the methods are

Table 2
Retrieval performance of the PA and FD algorithms

Method	Displacement			Performance	
	Total	Expected	SD	Expected	SD
PA	13767	55.06	4.36	**55.94%**	8.95
FD	16575	66.30	10.66	**46.96%**	9.59

fairly low. However, it should be noted that this score is dependent on the mismatch displacement. These scores may be acceptable in applications where the similarity rank is not as important, e.g., a researcher browsing the database for images pertinent to some pathology. In such cases, the set of acceptable results can be kept fairly large allowing for the expected displacement. In other applications the rank of retrieved images is fairly important, e.g., a physician looking up the database. The accumulated mismatch chart can then be used to determine an appropriate window size. However, these results do not appear meaningful for such applications. Further research is needed for developing a method that meets this and other requirements discussed earlier.

6. Conclusions

This paper presents a performance evaluation of two shape representation and similarity methods. The polygon approximation method was enhanced by improving the point selection criteria to better represent the vertebra shapes. The results indicate that the polygon approximation method performs a better than the Fourier descriptor method. However, the execution time of the PA method is significantly longer than the FD method. Both methods satisfy most requirements in matching medical images. However, they use global shape matching and do not permit queries on the partial shape. A related, and unevaluated, problem is determining their sensitivity to shape characteristics. The methods tend to do well in separating shapes with gross differences. However, they do not perform well with a set of closely related shapes, such as the vertebra collection. Research in these directions is under way and is the focus of our future work.

Our evaluation of these representative shape methods should help researchers seeking to develop or adopt shape representation and similarity methods. We have presented the literature in the field of shape matching, developed ground truth data and performance evaluation criteria that give similarity and rank relevance. We also expect this case study to be very valuable since shape based retrieval techniques for biomedical images have been largely unexplored.

References

Adoram, M., Lew, M.S., 1999. IRUS: Image Retrieval Using Shape. In: IEEE International Conference on Multimedia Computing Systems, vol 2, pp. 597–602.

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

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

Antani, S., Long, L.R., Thoma, G.R., 2002. 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, pp. 242–247.

Antani, S., Long, L.R., Thoma, G., Lee, D.J., 2003. 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, Santa Clara, CA, January 2003, pp. 405–416.

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

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

Bober, M., 2001. Mpeg-7 visual shape descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11 (6), 716–719.

Cootes, T.F., Taylor, C.J., 2001. Statistical models of appearence for computer vision. Technical report, University of Machester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester, Manchester, M12 9PT, UK, February 2001.

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

Eakins, J.P., Edwards, J.D., Riley, J., Rosin, P.L., 2001. 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, pp. 196–207.

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

Gonzalez, R., Woods, R., 2002. Digital Image Processing, second ed. Prentice Hall, Englewood Cliffs, NJ.

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

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

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

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

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

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. International Journal of Computer Vision 1 (4), 321–331.

Kim, Y.S., Kim, W.Y., 1997. Content-based trademark retrieval system by using visually salient feature. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 307–312.

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

Krainak, D.M., Long, L.R., Thoma, G.R., 2002. 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, pp. 108–116.

Kuo, W.-J., Chang, R.-F., Lee, C.C., Moon, W.K., Chen, D.-R., 2002. Retrieval technique for the diagnosis of solid breast tumors on sonogram. Ultrasound in Medicine and Biology 28 (7), 903–909.

Latecki, L.J., Lakämper, R., 2001. Shape description and search for similar objects in image databases. In: Veltkamp, R.C., Burkhardt, H., Kriegel, H.P. (Eds.), State-of-the-Art in Content-Based Image and

Video Retrieval, volume 22 of Computational Imaging and Vision. Kluwer Academic Publishers, Dordrecht, pp. 69–96.

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

Lee, D.J., Antani, S. and Long, L.R., 2003. 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, February 2003, San Diego, CA, pp. 1283–1291.

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

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

Long, L.R., Antani, S., Lee, D.J., Krainak, D.M., Thoma, G.R., 2003. 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, February 2003, San Diego, CA.

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

Mokhtarian, F., Mackworth, A.K., 1986. Scale-based description and recognition of planar curves and two-dimensional shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1), 34–43.

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

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

Sonka, M., Hlavac, V., Boyle, R., 1999. Image Processing, second ed. Brooks/Cole Publishing Company.

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

Traina, A.J.M., Castañón, C.A.B., Traina Jr., C., 2003. MultiWaveMed: A system for medical image retrieval through wavelets transformations. In: Proceedings of 16th IEEE Symposium on Computer-Based Medical Systems, New York, NY, pp.150–155.

Traina, A., Rosa, N.A., Traina Jr., C., 2003. Integrating images to patient electronic medical records through content-based retrieval techniques. In: Proceedings of 16th IEEE Symposium on Computer-Based Medical Systems, New York, NY, pp. 163–168.

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