

Evaluation of Shape Indexing Methods for Content-Based Retrieval of X-Ray Images

Sameer Antani^a, L. Rodney Long^a, George R. Thoma^a, and D.J. Lee^b

^a Communications Engineering Branch
Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, MD 20894

^b Department of Electrical and Computer Engineering
Brigham Young University
Provo, UT 84602

ABSTRACT

Efficient content-based image retrieval of biomedical images is a challenging problem of growing research interest. 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, a R&D division of the National Library of Medicine, we are developing a content-based image retrieval system for digitized images of a collection of 17,000 cervical and lumbar x-rays taken as a part of the second National Health and Nutrition Examination Survey (NHANES II). Shape is the only feature that effectively describes various pathologies identified by medical experts as being consistently and reliably found in the image collection. In order to determine if the state of the art in shape representation methods is suitable for this application, we have evaluated representative algorithms selected from the literature. The algorithms were tested on a subset of 250 vertebral shapes. In this paper we present the requirements of an ideal algorithm, define the evaluation criteria, and present the results and our analysis of the evaluation. We observe that while the shape methods perform well on visual inspection of the overall shape boundaries, they fall short in meeting the needs of determining similarity between the vertebral shapes based on the pathology.

Keywords: Content-Based Image Retrieval, Medical Image Databases, Shape Representation, Performance Evaluation

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} 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 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. 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, a problem confronting the researcher in image indexing, other than developing robust segmentation algorithms, is developing suitable feature representation

Further author information: (Send correspondence to S.A.)

S.A.: antani@nlm.nih.gov, Tel: +1 (301) 435-3218 L.R.L: long@nlm.nih.gov G.R.T.: thoma@nlm.nih.gov
D.J.L: djlee@ee.byu.edu

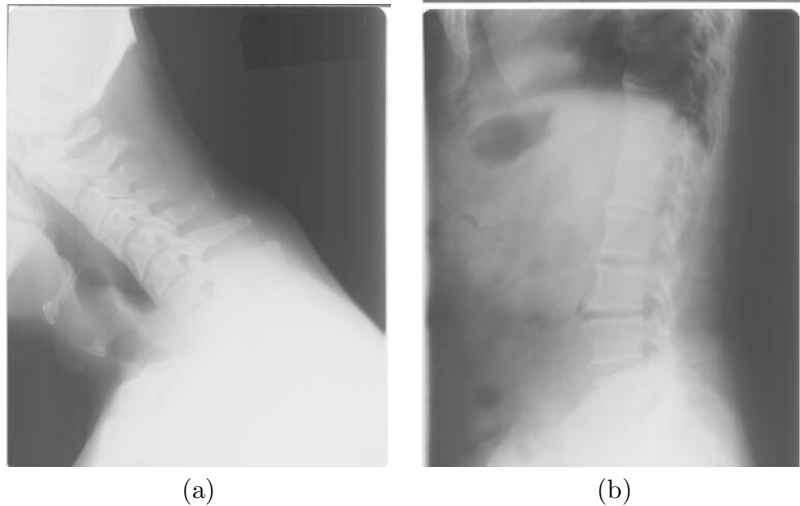


Figure 1. Examples of (a) cervical and (b) lumbar spine x-ray images

and similarity algorithms. An example is in segmented 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,⁴ 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.^{5,6} 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 indexed for these features, which include anterior osteophytes, disc space narrowing, and spondylolisthesis for the cervical spine and spondylolisthesis for the lumbar spine. We are investigating the 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.

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.⁷ The system supports conventional text retrieval, as well as retrieval based on shape similarity to a user-supplied vertebral image or sketch. The shapes are segmented using active contour and human assisted segmentation. An earlier prototype system used expert marked coarse 9-point manually segmented shapes.⁸ The prototype system currently supports 4 shape representation algorithms from the literature. These were implemented to evaluate their effectiveness in retrieving vertebral shapes. The desired technique, in addition to the shape representation properties of *invariance, uniqueness, stability, efficiency, ease of implementation, computation of shape properties, geometric invariance, compact representation, fast matching speed, and high quality image retrieval* identified in,⁹⁻¹¹ should enable

- a) the subtle variations in the pathology to be retained while keeping the dimensionality low,
- b) querying on local regions of the shape which may be pertinent to the pathology.

1.1. Motivation and Background

In our study of the spine x-rays (examples shown in Figure 1), we observe that only shape features appear promising for indexing the images. The images are gray scale and offer very little in terms of texture for the anatomy of interest. In this paper, we focus our attention on evaluating suitable shape description methods, published in the literature, for use with indexing and similarity retrieval of biomedical images in general and vertebral shapes in particular. We observe the following characteristics in vertebra images:

1. In general, there is high similarity among vertebrae. Except for minor characteristics and some exceptions, the outlines of the sagittal view of vertebrae from different regions on the spine are very similar.
2. Minor changes to outlines of the anatomy could be significant. In general, this is true for most biomedical images.
3. Local shape properties indicative of pathology are likely to be queried upon rather than the entire shape. The desired shape indexing methods should be able to support such localized queries.

Biomedical images, as identified above, have conflicting goals of low dimensionality for efficient indexing and matching while requiring the feature representation methods to retain the subtleties in the pathology. The purpose of this evaluation is to determine the effectiveness of shape representation methods by evaluating the retrieval against an expert marked database.

Many shape representation and similarity techniques can be found in the literature,^{11–18} to cite a few. These techniques adopt different approaches for representing shapes which can be broadly grouped under these four categories:

- Shape geometry based methods: Methods that use shape properties 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, Affine moments, and Zernike moments.¹⁹
- Polygon approximation methods: Methods that reduce the number of boundary points to as few as possible while maintaining the desired shape properties. Matching is done using the turn angle function.¹⁷
- Deformable shapes based methods: Methods that employ elastic deformation of templates.²⁰
- Fourier transform based methods: Representing the cumulative shape boundary as a function of its normalized length.²¹

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 if an algorithm is suitable for a CBIR system application can only be done after an evaluation of the shape methods on the shapes that populate the database. Several evaluations have been published in the literature.^{19, 22} However almost all these evaluations have been for shape retrieval methods applied to trademark image databases.

2. APPROACH

A representative method selected from each class of shape representation and similarity methods is tested for the properties desired on the vertebral shapes. In this section we briefly describe the selected methods, followed by a description of the test data set and the evaluation strategy.

2.1. Selected Methods

I Global Shape Properties and Invariant Moments. Global shape properties and moments are features intrinsic to a shape. The properties include area, center of mass, perimeter, axis of rotation, etc. In order to compute these, the shape contour must be converted to a binary image for processing which will give the same weight for each pixel inside the shape contour.

A Global Shape Properties Several global shape properties were calculated for selection. They include:

- *Center of gravity:* The average of x and y coordinates of all pixels inside the shape contour. This is calculated as the first order moment along each axis.
- *Area:* The total number of pixels inside the object contour.
- *Perimeter:* The total length of the contour in number of pixels.
- *Convex perimeter:* An approximation of the perimeter of the convex hull of an object.
- *Major axis length and angle:* The major axis of the result of least-squared error fit of an ellipse.
- *Minor axis length and angle:* The minor axis of the result of least-squared error fit of an ellipse.
- *Compactness:* This is defined as $perimeter^2/4\pi \cdot Area$
- *Roughness:* Defined as $roughness = perimeter/convexperimeter$. A smooth convex object, such as a perfect circle, will have the roughness of 1.0
- *Elongation:* The ratio of major axis length to minor axis length.

Besides the center of gravity and area which do not meet the invariant requirements for shape-based retrieval, these shape properties can be used to create a multi-dimensional feature space for calculating the distance between two shapes (in the feature space). The distance between two points in the feature space is a measure of shape similarity. The farther the two points are, the less similarity between the two shapes.

B Invariant Moments For a 2D continuous function $f(x, y)$, the moment of order $(p + q)$ is defined as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (1)$$

The central moments are calculated by shifting the origin to the center of the image given by (\bar{x}, \bar{y}) and defined as:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (2)$$

In the discrete domain, the integral is changed to a summation resulting in:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (3)$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (4)$$

The first and second order of the central moments can be derived from Equation 4 and expressed as:

$$\begin{aligned} \mu_{00} &= m_{00} & \mu_{20} &= m_{20} - \bar{x}m_{10} \\ \mu_{10} &= 0 & \mu_{02} &= m_{02} - \bar{y}m_{01} \\ \mu_{01} &= 0 & \mu_{11} &= m_{11} - \bar{y}m_{10} \end{aligned}$$

The normalized central moments would then be:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}, \text{ where } r = \frac{p+q}{2} + 1 \quad (5)$$

The relevant Hu²³ invariant moments then are given by

$$\begin{aligned}\phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2\end{aligned}$$

II Scale Space Filtering. Scale space filtering reformats the shape boundary points to represent the shape at different levels of detail. It is said to follow human perception of shapes.¹⁴ It provides more detail at scale higher level and progressively reduces the detail level until the shape becomes an oval shape. While capable of shape matching, a problem with this method is that the shape shrinks as it progresses from high detail level toward low detail level making comparison scale sensitive.

- **Curve Smoothing.** A closed planar curve can be parameterized according to its length and be expressed as $c(t) = \{x(t), y(t)\}$, where x and y represent the coordinates for each boundary point and t is the normalized length from a selected starting point. To smooth a curve, functions $x(t)$ and $y(t)$ can be convolved with a one-dimensional Gaussian kernel at different levels of abstraction as expressed below:

$$\begin{aligned}x(t, \sigma) &= \int_{-\infty}^t x(s)g(t-s, \sigma)ds \\ y(t, \sigma) &= \int_{-\infty}^t y(s)g(t-s, \sigma)ds\end{aligned}$$

$g(s, \sigma)$ is a Gaussian function with selectable variance s for different levels of detail.

- **Shape Representation and Token Descriptions.** The curvature function of $c(t, \sigma)$ at $\{x(t, \sigma), y(t, \sigma)\}$ can be expressed as:

$$\Gamma(t, \sigma) = \frac{X'(t, \sigma)Y''(t, \sigma) - X''(t, \sigma)Y'(t, \sigma)}{((X'(t, \sigma)^2 + Y'(t, \sigma)^2)^{3/2}} \quad (6)$$

where X', Y' and X'', Y'' are the first and second derivatives of $y(t, \sigma)$ and $x(t, \sigma)$, the parameterized boundary coordinate functions. The critical points on the polygon can be determined by searching for the minima on the curvature function. Once the critical points are determined, the polygon can be broken down to small segments (shape tokens). The visual aspect of each token is represented by three properties. They are symmetry, length, and orientation. The symmetry is the difference in the length of the segments originating at the two nodes of the token and intersecting at a the center of gravity of the token. The orientation is the angle from horizontal of the line connecting the center of gravity of the token and the mid point of the line connecting two adjacent critical points. The length of token is calculated and normalized by the total length of the polygon. The token orientation cannot be directly used for similarity measurement because it does not meet rotation invariance requirement.

- **Similarity Metric.** The L_2 -norm is used to compute the distance between two shapes. However, a problem with this method is that it is not inherently rotation invariant. The effect of rotation on a shape is that the curvature function shifts to the right or left depending on the direction of rotation. The values at the extreme ends of the plot wrap around and appear at the other end. Rotation invariance can be achieved if a correspondence can be determined between the tokens. This can be done, for example, by selecting a reference token (such as the longest token) and shifting the other so that they match up.

III Polygon Approximation. Polygon approximation or curve evolution is a process that eliminates insignificant shape features and reduce the number of data points. The resultant representation is one that uniquely describes the shape. The approximated curve was 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

$$K(s_1, s_2) = \frac{\beta(s_1, s_2)l(s_1)l(s_2)}{l(s_1) + l(s_2)} \quad (7)$$

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 then represented by the turn function, which is the function of turn angle versus the normalized length. Representing shape in tangent space meets the invariant requirements for shape-based retrieval. It is translation invariant because the turn angles and length do not contain information about the shape location. It is scaling invariant because it uses normalized length. For rotation and starting point shift, the turn function remains the same shape expect shifting vertically when there is a rotation and moving horizontally when there is a shift in starting point.
- **Similarity Measurement** The distance (dissimilarity) between two turn functions Θ_A and Θ_B can be measured as

$$\begin{aligned} \delta_2(A, B) &= \|\Theta_A - \Theta_B\|_2 \\ &= \sqrt{\int_0^1 |\Theta_A - \Theta_B|^2 ds} \\ &= \sqrt{\min_{\theta \in \mathbb{R}, t \in [0,1]} \int_0^1 |\Theta_A(s+t) - \Theta_B(s) + \theta|^2 ds} \end{aligned} \quad (8)$$

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 (query or database). 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 .²⁴ The best value of θ is a function of length shift t in the X axis to minimize

$$\begin{aligned} h(t, \theta) &= \int_0^1 |\Theta_A(s+t) - \Theta_B(s) + \theta|^2 ds \\ \text{when } \theta'(t) &= \int_0^1 (\Theta_A(s+t) - \Theta_B(s) + \theta) ds \\ &= \alpha - 2\pi t, \\ \text{where } \alpha &= \int_0^1 \Theta_B(s) ds - \Theta_A(s) ds \end{aligned} \quad (9)$$

IV 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 contour is approximation 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.²¹ 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 the clockwise turn gives a negative angle whereas a counter clockwise turn gives a positive angle. This method represents a closed

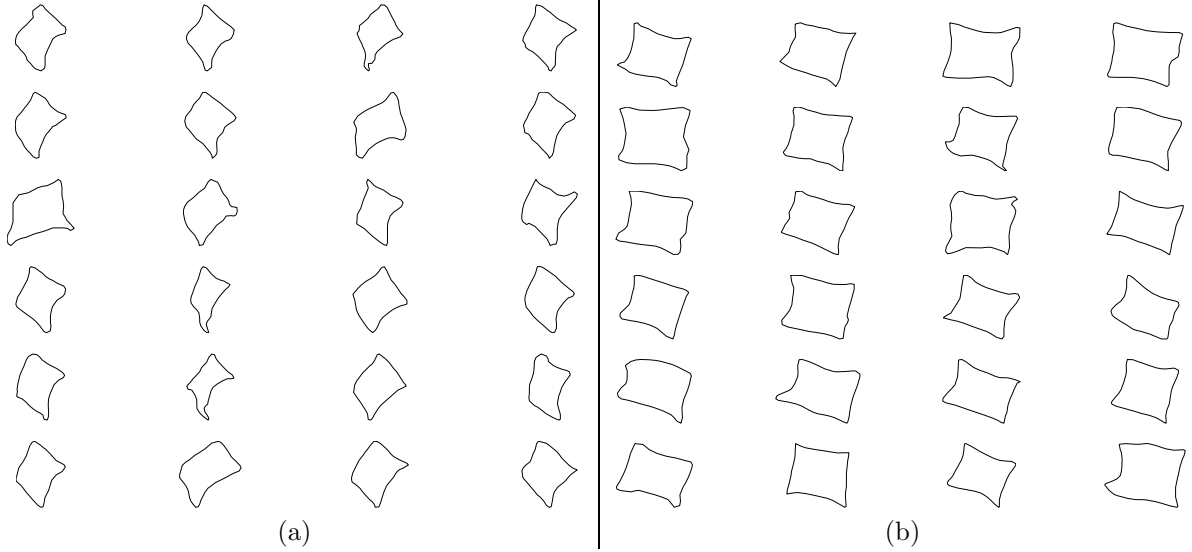


Figure 2. Examples of segmented (a) cervical and (b) lumbar vertebra shapes in sagittal view

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 on the polygon size. The only issue left is the starting point shift invariance requirement which is taken care of by the shift invariance property of the power spectrum.

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

$$\Theta(t) = \mu_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt), \quad (10)$$

where 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 invariant 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.

2.2. Data

The boundary data that forms the test set consists of vertebra outlines segmented from the digitized NHANES II spine x-ray images. The quality of the digitized NHANES II spine x-rays is fairly poor making automated segmentation a challenging task. We are exploring active contour models²⁵ and active shape modelling²⁶ techniques for automated segmentation. For purposes of the evaluation we have adopted a manually assisted segmentation approach by fitting splines to manually identified coarse boundary and active contour segmentation. This approach was selected in order to address the problem of shape representation while research is conducted on other CBIR components. Figure 2 shows a few sample manually segmented shapes from the data set. The methods are evaluated on 250 shapes which include 25 vertebra shapes each for cervical C3-C7 and lumbar L1-L5 vertebrae. The ground truth is based on the coarse radiologist marked 9 point data defining the vertebra outline,⁸ shown in Figure 3(a). The 9 point model was chosen because of the semantic relevance of the marked points and the fact that they were marked by an expert. The semantic relevance is as follows:

- Points 1 and 4 mark the upper and lower posterior corners of the vertebra, respectively.

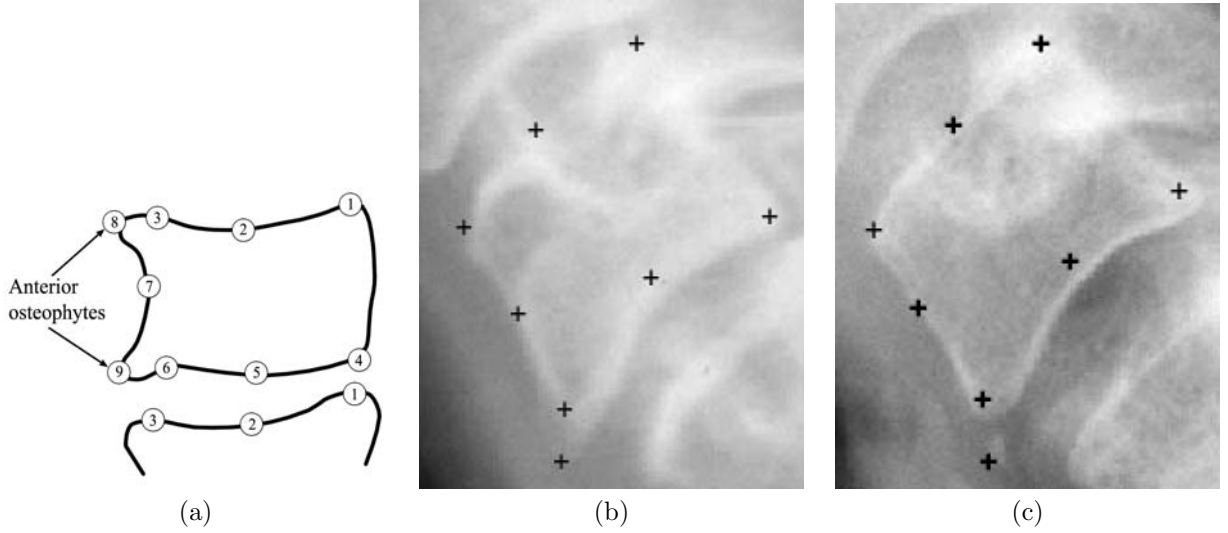


Figure 3. (a) Radiologist marked 9 points, (b) Example query image with 9 points superimposed. (c) Procrustes matching of most similar different image with 9 points superimposed.

- 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 ground truth data is a list of top 25 similar shapes from the database for every shape in the database. 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 Equation 11, where the shapes X and X' have n boundary points whose coordinates are given by (x, y) and (x', y') .

$$P = \sum_{i=1}^n \left| \begin{bmatrix} S \cdot \cos \theta & -\sin \theta & T_x \\ \sin \theta & S \cdot \cos \theta & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}_X - \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}_{X'} \right|^2 \quad (11)$$

The matching process translates shape X by (Tx, Ty) 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 Figure 3(b) and Figure 3(c).

This method finds the closest distance between two shapes. When applied to the 9 point model, it can be considered as the closest semantic distance between the shapes. This metric was chosen over a manually marked ground truth because when determining similarity, human readers tend to focus on local shape features such as elongated osteophytes, flattening of the vertebra, etc. Thus, marking the ground truth in this manner would require normalization of the notion of shape similarity as perceived by different users. Using the Procrustes distance overcomes this hurdle while allowing a fair evaluation of the methods.

2.3. Evaluation Strategy

Each shape method finds the 25 most similar shapes from the data set for every shape in the data set. These are ranked in order of decreasing similarity. The methods also report similarity scores for each of these shapes.

However, the range of these scores, which indicate degree of “goodness”, is different for each method. Comparing the methods based on these similarity scores is not feasible without normalizing them. However, since the degree of “goodness” is relative to the “best shape in the set”, there is a lack of a reference point which makes normalization problematic. 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.

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

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

Table 1. Example of comparing similarity rankings

method is computed by determining the mismatch between the two lists, which is described in the rightmost column of the table. This disparity can be identified as false alarms and missed detections and is expressed in the following four cases:

1. A shape is deemed less similar than it is in the ground-truth: This case is one in which there is a “downward” movement in the similarity rank of a shape. This case can be treated as a missed detection.
2. A shape is deemed more similar than it is in the ground-truth: This case is one in which there is an “upward” movement in the similarity rank of the shape. This case can be treated as a false alarm.
3. A shape that is ranked in the ground-truth is not detected as similar (at least not similar enough to show up in the top 25) by the shape representation method, i.e. the classical missed detection.
4. A shape that is not in the ground-truth is detected as similar, i.e. the classical false alarm.

The last two cases are degenerate forms of the first two. Shapes higher up in the rank are more important and the method should be penalized more if it does not place them in the same slot as in the ground truth. Formally, the penalty should be a function of the rank distance weighted by the importance of the expected rank. A problem with using a single penalty for false alarms and missed detections (cases 1 and 2) is that it hides the fact that for every downward movement in rank there is some corresponding, but not uniform, upward movement. To separate these events and allow a better appreciation of the performance of the shape algorithms we need to penalize false alarms and missed detections differently.

We penalize missed detections three times more heavily than false alarms with a penalty $w_{md} = 0.6$ and $w_{fa} = 0.2$, respectively. Since items at a higher rank are more important, the penalty is linearly decreased with increasing rank with a function $f(a)$ ranging from 25 to 1. This penalty is then associated with the absolute mismatch distance d between expected rank of a shape and its determined rank. Then the performance P is given by Equation 13. Here, d_{fa} and d_{md} are the upward and downward distances for the mismatches. These are computed for each shape and summed over all N shapes in the database.

$$P_{type} = \sum^N f(a) * W * D \tag{12}$$

where,

$$type = \begin{cases} fa & \text{false alarm} \\ md & \text{missed detection} \end{cases}$$

$$\begin{aligned}
W &= \begin{cases} w_{fa} & \text{if } type = fa \text{ false alarm} \\ w_{md} & \text{if } type = md \text{ missed detection} \end{cases} \\
D &= \begin{cases} \sum d_{fa} & \text{if } type = fa \text{ false alarm, upward mismatch distance} \\ \sum d_{md} & \text{if } type = md \text{ missed detection, downward mismatch distance} \end{cases}
\end{aligned}$$

This results in two performance values, P_{fa} and P_{md} for false alarms and missed detections, respectively. For cases 3 and 4 described above, the shapes are considered to be at rank 26. Thus, the penalty for false alarms and missed detections is $|26 - GT_{rank}|$, where GT_{rank} is the ground truth rank.

We define the following performance metrics for each method:

- Precision: By computing the total mismatch distance in upward movements, we can compute Precision and Recall as

$$Precision = \frac{D_{fa} - T_D}{T_D} \quad (13)$$

$$Recall = \frac{D_{md} - T_D}{T_D} \quad (14)$$

where

$$D_{fa} = \sum^N d_{fa}$$

$$D_{md} = \sum^N d_{md}$$

$$T_D = \text{Total possible displacement}$$

Here, N represents total number of shapes. T_D is computed as the maximum possible displacement distance over all the shapes. Since there are 25 shapes in each ranking, each shape could be incorrectly placed in one of 24 incorrect slots or could be at position 26 indicating a missed detection (or, conversely, a false alarm); which makes the maximum possible distance for any ranked shape as 25. There are 25 ranked shapes for each similarity test and 250 tested shapes in all. Thus, $T_D = 250 \times 25 \times 25$.

- Mean (μ) and standard deviation (σ) of the performance scores, P_{fa} and P_{md} , over 250 shapes.
- Total number of displacements.

3. RESULTS

This section describes the results for the five methods described above. The table showing the results over all 250 shapes is It is important to note that the ideal performance score should be 0. So, a lower score is a

Shape Representation Method	Precision %	Recall %	P_{fa}		P_{md}		Number of Displacements
			μ	σ	μ	σ	
Geometric Properties & Invariant Moments	55.06	54.81	2731.60	186.80	918.60	66.60	5949
Polygon Approximation	58.44	57.57	2476.90	304.70	855.70	86.00	5905
Token Description	54.47	53.97	2757.70	173.20	935.20	46.10	5975
Fourier Descriptors	55.50	54.71	2695.30	197.40	919.60	54.10	5965

Table 2. Evaluation results table

better performance. The number of displacements indicates the relationship of the performance to the number of mismatches. Two algorithms performing similarly can be separated by the number of movements.

3.1. Analysis

As seen in Table 2, the overall performance of all methods is very similar and rather unimpressive. The polygon approximation algorithms performs the best from the above. All of the methods have been penalized very heavily for the false alarms. A high score on P_{fa} indicates that these occurred in higher ranked shapes. The total number of movements is also very high. Note that even though the penalty for the missed detections, P_{md} , is lower than P_{fa} (about a third), the recall remains as low as the precision. This is because the downward misplacement distance is very high. This is indicative of a high number of incorrectly ranked shapes.

An analysis of the evaluation points to the lack of sensitivity in these methods to localized shape features, i.e. they do not really provide the same shape information as the 9-point model. Analyzing each method, we see that the Geometric Shape properties and Invariant Moments are an overall description of the shape. They do not provide a high degree of separation for similar shapes. Although, in smaller evaluations, we have found it to work well in separating shapes that are significantly different. The Polygon Approximation method is designed to reduce the number of boundary points while retaining shape uniqueness. However, the decision to discard a point is local and not related to the overall boundary shape. It seems to work better than other methods, however it is possible that points from the dense boundary representation that are coincident with the 9 points representation may have been discarded causing a loss of information. The Token Analysis method uses curvature function analysis to represent shape. However, it keeps only the “valleys” in the curvature representation of the boundary and the tokens shrink the contour during the scale space filtering process. This removes significant local features causing the poor performance. The Fourier Descriptors method provides an overall representation of the shape. However it is likely that in the process of simplifying the computationally intensive comparison process, some high frequency components may have been ignored.

4. CONCLUSIONS

Overall, in the general sense of the vertebral shape, the visual results from these methods are fairly good. Specific queries on the pathology are likely to be correct only about 55%-60% of the time as indicated by the evaluation. There is a need for a new shape method that addresses the characteristics of biomedical shapes and the issues pertinent to their representation and retrieval.

We hope that this evaluation of representative methods from major classes of shape methods will enable researchers seeking to develop or adopt shape representation and similarity methods. We have defined evaluation criteria that give importance similarity and rank relevance and view the mismatches as false alarms and missed detections. This enables a performance evaluation in terms of standard precision and recall values. We also expect this case study to be very valuable since shape based retrieval techniques for biomedical images have been largely unexplored.

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), pp. 945–965, 2002.
2. T. M. Lehmann, B. B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, and M. Kohnen, “Content-based image retrieval in medical applications: A novel multistep approach,” *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases* **3972**, pp. 312–320, 2000.
3. 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), pp. 184–198, 1997.
4. L. Long and G. Thoma, “Landmarking and feature localization in spine x-rays,” *Journal of Electronic Imaging* **10**(4), pp. 939–956, 2001.
5. L. Long and G. Thoma, “Feature indexing in a database of digitized x-rays,” in *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2001, Vol. SPIE 4315*, pp. 393–403, (San Jose, CA), January 2001.
6. L. Long and G. Thoma, “Indexing of image content in spine x-rays,” in *Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2000, Vol. SPIE 3972*, pp. 55–63, (San Jose, CA), January 2000.

7. 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 *To appear in Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing*, 2002.
8. 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**, pp. 108–116, February 2002.
9. F. Mokhtarian and A. K. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(1), pp. 34–43, 1986.
10. 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), pp. 789–805, 1992.
11. 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**, pp. 86–97, 2000.
12. A. Quddus and M. Gabbouj, "Wavelet-based corner detection technique using optimal scale," *Pattern Recognition Letters* **23**(1-3), pp. 215–220, 2002.
13. M. Adoram and M. S. Lew, "IRUS: Image Retrieval Using Shape," in *IEEE International Conference on Multimedia Computing Systems*, **2**, pp. 597–602, 1999.
14. A. Del Bimbo and P. Pala, "Shape indexing by multi-scale representation," *Image and Vision Computing* **17**(3-4), pp. 245–261, 1999.
15. B. Günsel and A. M. Tekalp, "Shape similarity matching for query-by-example," *Pattern Recognition* **31**(7), pp. 931–944, 1998.
16. M. Kliot and E. Rivlin, "Invariant-based shape retrieval in pictorial databases," *Computer Vision and Image Understanding* **71**(2), pp. 182–197, 1998.
17. L. J. Latecki and R. Lakämper, "Shape description and search for similar objects in image databases," in *State-of-the-Art in Content-Based Image and Video Retrieval*, R. C. Veltkamp, H. Burkhardt, and H. P. Kriegel, eds., *Computational Imaging and Vision* **22**, pp. 69–96, Kluwer Academic Publishers, 2001.
18. R. Mehrotra and J. E. Gray, "Similar-shape retrieval in shape data management," *IEEE Computer* **28**(9), pp. 23–32, 1995.
19. 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," *IS&T/SPIE Conference on Storage and Retrieval for Media Databases* **4315**, pp. 196–207, 2001.
20. A. K. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(3), pp. 267–278, 1996.
21. C. Zahn and R. Roskie, "Fourier descriptors for plane closed curves," *IEEE Computer* **C-21**(3), pp. 269–281, 1972.
22. A. K. Jain and A. Vailaya, "Shape-based retrieval: A case study with trademark image databases," *Pattern Recognition* **31**(9), pp. 1369–1390, 1998.
23. M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory* **8**, pp. 179–187, 1962.
24. 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), pp. 209–216, 1991.
25. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision* **1**(4), pp. 321–331, 1988.
26. T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," tech. rep., University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester, Manchester, M12 9PT, U.K., February 2001.