Comparative Study of Spine Vertebra Shape Retrieval using Learning-based Feature Selection

Haiying Guan, Sameer Antani, L. Rodney Long, and George R. Thoma
U.S. National Library of Medicine
National Institutes of Health, Bethesda, MD, USA
guanh,santani,rlong,gthoma@mail.nih.gov

Abstract

Feature extraction and selection are two important steps for shape retrieval. Given a data set, a set of features which describe the shape property from different aspects are extracted. Our goal is a learning-based methodology to select the features for improving retrieval performance. Our approach uses both global and local feature descriptors. The global shape features include geometric ones (elongation, eccentricity, roughness, and compactness), Fourier descriptors with complex coordinates, Fourier descriptors with Centroid Contour Distance Curve, Coefficients of Fourier Expansion of Bent function, moment invariants, and local shape features that include turn angle and Distance Across the Shape. We propose a learning-based feature selection algorithm as a strategy for optimizing retrieval performance. We provide results from the vertebra shape dataset created from our database containing spine x-rays from the National Health and Nutrition Examination Survey (NHANES II). Finally, we compare the retrieval performances of feature descriptors on "whole shape" and "corner shape" datasets. The experimental results show that various feature descriptors perform differently on different datasets, and that feature selection schemes improve the retrieval performance significantly.

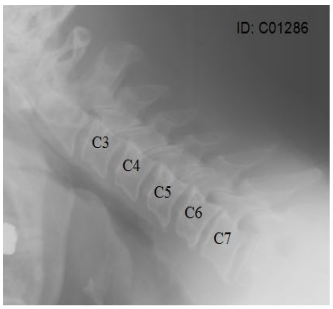
1. Introduction

Shape retrieval has been studied for decades. A common problem in shape retrieval involves features that effectively capture the semantic meaning of the shape and support retrieval with good sensitivity and specificity.

A particular problem of interest is the retrieval of images having the same category label as that of a query image. In this case feature selection must be done to optimize an objective function that is defined, in some sense, in terms of the category labels returned.

In this paper we focus on the vertebra shape retrieval of the cervical spine x-ray images collected in the second National Health and Nutrition Examination Survey (NHANES II) conducted by the National Center for Health Statistics (NCHS). In this set of x-ray images, osteophytes (bone "spurs") are a visual feature of interest to the osteoarthritis community. In our work, we have specifically focused on anterior osteophytes (AO), rather than those located on the vertebra posterior, which are less visible in the image. The x-ray image collection consists of 10,000 cervical spine images and 7,000 lumbar spine images, all digitized from film. From these images, a subset of vertebra shapes has been segmented by manual or computer-assisted methods, to create a vertebra "whole shape" database, containing the boundary of most of the vertebral body; and a "corner shape" database, containing the boundaries of the vertebral anterior corners only. Abnormal vertebrae exhibiting AO are identified by shape variation in the anterior "corner" of the vertebral outline seen in the sagittal view. For the segmented vertebra shape, Macnab's osteophyte classification [5] (claw and traction) in addition to three severity grades (slight, moderate, and severe) has been used for classifying vertebrae [17] into ten classes: normal, slight claw, slight traction, slight claw and traction, moderate claw, moderate traction, moderate claw and traction, severe claw, severe traction, severe claw and traction. The resulting databases of labeled, segmented shapes constitute the "ground truth" for our work. For computing precision and recall, we consider a retrieved image a "match" to the query images, if both images have the same class label.

Content-Based Image Retrieval (CBIR) methods that use shape have been applied to image collections such as trademark data, fish images, and silhouettes [13, 16, 19]. However, not all shape retrieval algorithms are suitable for biomedical shapes such as vertebral outlines due to (i) high similarity across anatomical shapes, (ii) inadequate separability in extracted features in representing subtle boundary differences (which indicate pathology), and (iii) inadequacy of shape description for supporting pathology-



	Expert 1	Expert 2	Expert 3
C3	Normal	Normal	Moderate, Claw
C4*	Normal	Normal	normal
C5	Severe, Claw	Severe, Claw , Traction	Severe, Claw
C6	Severe, Traction	Moderate , Traction	Moderate, Traction
C7*	Severe, Claw	Severe, Claw	Severe, Claw

*Used for training and testing

Figure 1. Cervical spine x-ray with AO labeled by three experts for C_3-C_7

specified queries. A key issue is the high inter-object similarity, with only subtle differences at pathology-specific locations. Figure 1 shows an example cervical spine x-ray image. A pathological or abnormal vertebra is identified by the abnormal, subtle variation of the shape on the anterior osteophytes corner. The table on the right describes the AO types and severity levels of C_3-C_7, as labeled by three board-certified radiologists. Due to the inter- and intra-personal variability of the medical experts, the "ground truth" of the vertebra corner type and grade is problematic to some degree. Thus, retrieving a semantically meaningful shape, considering the subtle pathological variations among many similar shapes, is a very challenging problem. Successful similarity retrieval requires that the feature descriptors not only generalize beyond the training set, but also have strong discriminative capability. This combination is difficult to obtain using only a single shape feature.

In this paper, we provide an extensive comparison of CBIR shape descriptors for vertebra shape, using both global and local shape features. The global shape features include those related to geometry (elongation, eccentricity, roughness, and compactness), as well as Fourier descriptors with complex coordinates, Fourier descriptors with Centroid Contour Distance Curve (CCDC), Coefficients of Fourier Expansion of Bent function, and moment invariants. The local shape features are turn angle and Distance Across the Shape (DAS). We analyze the performance of these features for retrieval from our two vertebra databases. Given a shape descriptor defined by a feature vector with alternative components, we propose a learning-based feature component selection algorithm for improving retrieval performance. We then compare the performance of alternative features using the whole shape and the corner shape databases. The experimental results show differential performance across these two databases, and show that our feature selection schemes could improve retrieval performance significantly. Our learning-based feature selection approach is readily generalizable to other shape databases.

2. Shape Features

Shape retrieval involves three primary factors: shape representation, shape similarity measure, and shape indexing. Among these, shape representation is possibly the most significant for shape retrieval. In this section, we focus on shape-based image retrieval features. For a given vertebra shape contour, we use both global and local shape features. We choose features that are invariant with respect to rotation, scale, and translation. These features are described and summarized in the following subsections.

2.1. Segmentation and Preprocessing

The vertebra shapes were segmented using an active contours method modified to constrain evolving contour points to follow "orthogonal curves" [18], to avoid convergence to a self-intersecting solution contour at vertebra corners [9]. The solution contours have 36 points. Nine of these 36 points were distinguished as geometrical or anatomical reference points, with relative locations that are approximately constant across the veterbra shapes. The nine points, shown in Figure 2 were either manually marked by experts, or extracted automatically or semi-automatically by specialized algorithms [9].

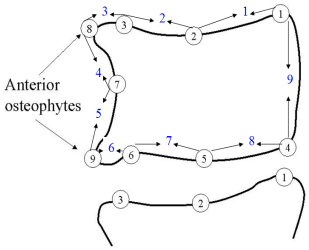
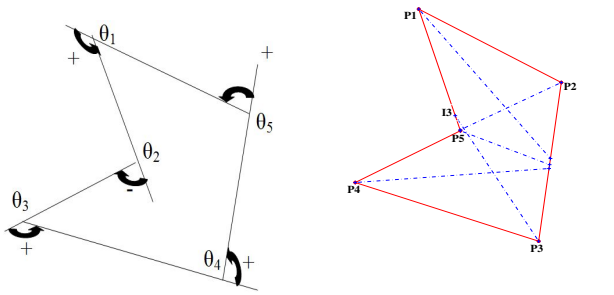


Figure 2. Nine reference points marked on each vertebra

For the current work, we preprocess these segmented shapes by curve smoothing (to reduce noise), fitting (for smoothness), interpolation, and re-sampling (for larger number of evenly distributed points) to obtain the final shape contour description. The curve fitting and interpolation are done with the natural cubic spline algorithm. Then the shape contour is resampled by equal arc length sampling. Finally, the vertebra whole shape is represented by two boundary point sets with different resolutions. The "dense sampling set" contains 180 points, and the superior and the inferior anterior corners are represented by 60 points, respectively. The "sparse sampling set" contains 72 points, with the superior and the inferior anterior corners represented by 25 points, respectively.



(a) Turning Angle (b) Distance Across the Shape

Figure 3. Example of TA and DAS

2.2. Turn Angle (TA)

To capture the characteristics of shape in local regions, we use two different features. The first is Turn Angle (TA). Turn Angle is also called Turning Angle or Bent Angle. It is defined as follows [3]: if the points on the polygon are ordered in the counterclockwise direction, and the polygon is traversed in this direction, the Turn Angle is the angle between the direction vector for the current polygon segment and the next one; the sense of the Turn Angle is calculated such that a clockwise turn gives a negative angle whereas a counterclockwise turn gives a positive angle. Figure 3 (a) shows an example.

For an arbitrary shape, the Turn Angle feature could be calculated from the approximating polygon for that shape. Turn Angle for a polygon with n vertices is simply a vector in R^n . For example, if the vertebra is represented as a polygon with 72 vertices (our "sparse" representation), the Turn Angle is a 72-element vector. If the polygon has the concept of an initial vertex, similarity computation is straightforward, e.g., with a Euclidean metric. If there is no initial vertex, similarity between two shapes may be computed by a combinatorial comparison of distances between possibly-matching sets of vertices. This computation may be optimized by dynamic programming.

2.3. Distance Across the Shape (DAS)

Distance across the shape [4] is another local shape feature. DAS is defined, for each vertex P in a polygon, as the length of the angle bisector at P, measured as the line segment from P to the intersecting side of the polygon. For Example, the interior bisector of angle $\angle P_2P_3P_4$ in the figure 3 (b) intersects the contour at point I_3. The length of P_3I_3 is the DAS at point $P3$. If the bisector intersects the shape multiple times, the distance to the closest intersection is used. Similarly as for turn angle, if we represent the vertebra shape as a polygon with 72 sample points, the DAS feature may be calculated on those 72 points.

2.4. Geometric polygon features (Geo)

The geometric polygon features that we use are elongation, eccentricity, compactness, and roughness [2]. Elongation is the length of major axis divided by the length of minor axis. Eccentricity is a parameter associated with every conic section that could be treated as a measure of how much the conic section deviates from being circular. Compactness is also called the circularity ratio, which is the ratio of the area of the shape to the area of a circle, mathematically given as $M = 4\pi * area/perimeter^2$. In this work, roughness is defined as the ratio of the perimeter of the shape divided by the perimeter of the convex hull of the shape.

2.5. Hu Moments (Hu)

In addition to the simple geometric polygon features, various global shape representation descriptors are used, which can be classified into two categories: region-based and contour-based. Hu moments [6], a region-based technique, are a family of seven invariant moments, derived from the second and third order normalized central moments. Because moments combine information across an entire object rather than providing information just at a single boundary point, they capture some of the global properties missing from many purely contour-based representations.

2.6. Fourier Descriptor of complex coordinates (CoordFD)

Although, as noted above, region-based shape representation captures more global properties, contour-based shape representation is also widely used. Contour-based shape representation only exploits the shape boundary information. The Fourier descriptor (FD) is a commonly used feature for shape representation. Many FD methods have been reported in the literature [20]. In these methods, various *shape signatures* have been exploited to obtain FD. However, FD derived from different signatures has significantly different performance results for shape retrieval. In this work, we compare shape retrieval using FD derived from three shape signatures. The signatures considered are central distance, complex coordinates, and the Fourier series coefficient of the Bent function.

A complex coordinate function yields a set of complex number generated from the boundary coordinates (x, y):

$$z(t) = x(t) + i * y(t) \quad (1)$$

The discrete Fourier transform of $z(t)$ is given by

$$FD(n) = \frac{1}{N}\sum_{t=0}^{N-1} z(t)exp(\frac{-j2\pi nt}{N}) \quad (2)$$
$$where, n = 0, 1, ..., N-1$$

The Fourier coefficients are calculated by the Fast Fourier Transform (FFT) for each shape in the database. In our experiments, the first 20 Fourier coefficients are considered, and the remaining coefficients are ignored as high frequency noise. The first frequency coefficient FD_0, the "direct current" (DC) coefficient, is related to the position of the shape and is discarded. Each Fourier coefficient has two components: amplitude and phase. In-plane rotation invariance is achieved by using only the amplitude component. Scale invariance is achieved by dividing amplitudes of other descriptors by the amplitude value of the second frequency coefficient FD_1. The final descriptor $CoordFD$, which is the invariant feature vector used to index the shape during the retrieval, is described in Equation 3.

$$CoordFD = \{\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, ..., \frac{|FD_{N-1}|}{|FD_1|}\} \quad (3)$$

2.7. Fourier Descriptor of Centroid Contour Distance Curve (CCDCFD)

The centroid distance function is expressed by the distance of the boundary points from the centroid (x_c, y_c) of the shape shown in Eq. 4 [20].

$$r(t) = [x(t) - x_c]^2 + [y(t) - y_c]^2 \quad (4)$$

After the subtraction of the centroid center, which represents the position of the shape, from boundary coordinates, the centroid distance representation is also invariant to translation. The first N coefficients of the Fourier transform of the CCDC are the resulting Fourier descriptors. These remaining coefficients are scaled to the interval $[0, 1]$ by the first coefficient. This makes the Fourier descriptors scale invariant.

2.8. The coefficient of the Fourier Series expansion of the Bent function (BFFS)

The Bent function measures the angle change of the counterclockwise tangent as a function of arc length t. As shown in Figure 3, the points on the polygon are ordered in the counterclockwise direction. $\theta_1 \cdots \theta_5$ are the turn angles. The Bent function is Turn Angle as a function of segment length. When the path is polygonal, the Bent function is piecewise-constant, with jump points corresponding to the vertices. By definition, the Bent function is invariant under translation, rotation, and scaling of the polygon. Figure 4 shows the corresponding Bent function of the polygon shown in Figure 3.

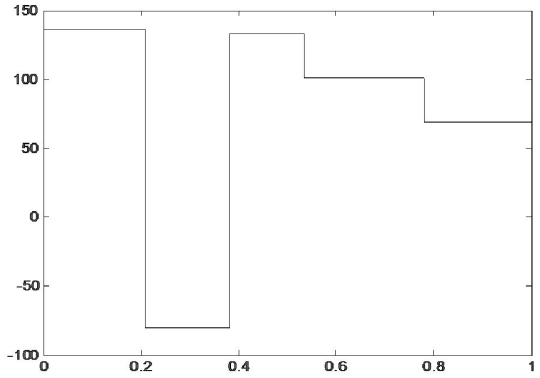


Figure 4. Example of the Bent function of the polygon shown in Figure 3.

The Fourier feature, BFFS, is defined as the amplitude of the coefficient of the discrete Fourier series expansion of the Bent function [8]. An important practical consideration is that the Bent function is noise-sensitive. For this reason, the curve samples must be smoothed before the calculation of the Bent function. If there are noisy points on the curve, the shape of the Bent function will change greatly, and so will the values of the Fourier series coefficients derived from this function.

3. Learning-based feature selection

It has been pointed out that for a fixed sample size, increasing the number of features (with a corresponding increase in the unknown parameters) adversely affects the reliability of parameter estimates [7]. This is commonly known as the "curse of dimensionality" and may result in degraded classifier performance. For CBIR applications, this suggests that, whenever a pool of features is available to index the image database, the best choice of features to use maybe a subset, rather than all, of the feature in the pool. The 'best' performance features from the training phase may yield better retrieval performance on the testing data set.

In this section, we propose a feature component selection scheme, which takes the performance of individual feature components in the training set into account and accumulates their contributions. We use Manhattan distance to rank the images in our database with respect to similarity to a query image, for a single feature. First, according to their performance on the training data, the features with high retrieval accuracy are selected. Let MAP^m denote

the *mean average precision* (MAP) [10] achieved for training queries for feature m ($m = 1, 2, 3, ..., M$, where M is the number of features); then after sorting all the MAP^m, we choose the first K features with highest precision rates, $MAP^{m_k}(k = 1, ..., K)$, where k corresponds to the kth largest mean average precision of a single feature. Then, to measure similarity between shapes represented by more than one feature, we define the *fused score* f_{qi}^K for the query shape q and retrieved shape i with K number of features as follows:

$$f_{qi}^K = \sum_{k=1}^{K} |n_q^k - n_i^k| \tag{5}$$

where n_q^k is the value of feature m_k (with performance MAP^{m_k}) of the query, and n_i^k is the value of feature m_k of the ith item in the data set. K is the number of features selected, and may affect the retrieval performance greatly. If the retrieval performance (MAP) of the training data using the fusion score in Eq. 5 is denoted as MAP^{f_K}, then the optimal $K_{optimal}$ is determined by the following equation,

$$K_{optimal} = \arg \max_{K=1}^{M} MAP^{f_K} \tag{6}$$

where M is the total number of features. Figure 5 shows the average precision-recall graph [10] of retrieval performance with 1-7 ($K = 1, ..., 7$) 'best' Hu moment features using Manhattan distance for similarity measurement (the fifth and sixth curves almost overlap with each other, and slightly different from the seventh curve). Note that the retrieval performance with four features is best ($K_{optimal} = 4$), rather than using all seven features. These results suggest a strategy of selecting a 'best' subset of features by evaluating alternative subsets on the training data set.

4. Evaluation, results, and discussion

The dataset for this experiment was generated from 200 cervical spinal x-ray images selected from the NHANES II collection. As noted earlier, the ground truth was marked by three board-certified radiologists and divided into ten classes: normal, slight claw, slight traction, slight claw and traction, moderate claw, moderate traction, moderate claw and traction, severe claw, severe traction, severe claw and traction. Due to the inter- and intra-observer variability of the medical experts, the ground truth of the vertebra corner type and grade confounds the class categories to some degree. Only 117 vertebrae that had complete agreement among the radiologists were selected for training and testing using a leave-one-out algorithm. Two data sets were generated for the experiments: one is the whole vertebral shape data set, and the other is the vertebral corner shape data set.

In general, it is difficult to compare different image retrieval systems from results reported in the literature [12]. For example, some medical image retrieval systems that do report evaluations often only provide screenshots of example results to queries [14], [15]. However, several query example results do not reveal a great deal about the real performance of the system and are highly subject to bias, as the best possible query result can be chosen arbitrarily by the author. This problem in comparative retrieval system evaluation is described in detail in [11].

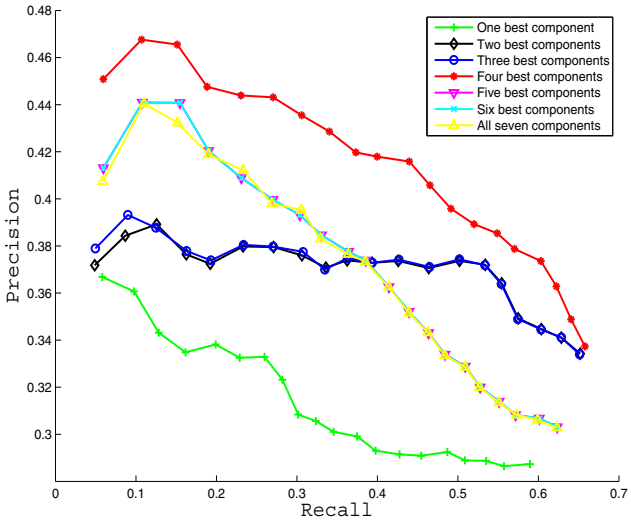


Figure 5. The average precision-recall graph for retrieval using Hu Moment features

In this paper, we adopt two quantitative evaluation criteria. First, we used the average precision-recall graph. Retrieval $precision$ is defined as the proportion of the images among all those retrieved that are truly relevant to a given query; $recall$ is defined as the proportion of the images that are actually retrieved among all the relevant images to a query. The average precision-recall graph is a plot of the average retrieval precision vs. the average recall over the precision and recall operating ranges of interest. We considered an image to be truly relevant to a query if the retrieved images were in the same class (both have the same type and grade) as the query image. This is a very strict criterion for the retrieval quality evaluation, because the determination of ground truth of the vertebra corner type and grade is subject to variability among the medical experts due to the subtle variations of the corner shapes. Second, we used the F1 measure [1], which is the weighted harmonic mean of precision and recall, and plot it against the number of retrieved items. The definition of F_1 is given by:

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{7}$$

4.1.Comparisons on whole shape data set

We compare the retrieval performance of the shape feature descriptors described in Section 2 on the whole vertebral shape data set, by plotting the average precision-recall metric and F1 measure [1] as shown in Figure 6. The retrieved vertebrae are considered to be relevant matches only if both the pathology type and grade match with the query. The results show that different feature descriptors perform differently on the data set. For the whole shape data set, feature BFFS outperforms other features, which is followed by Turn Angle, DAS, and CoordFD.

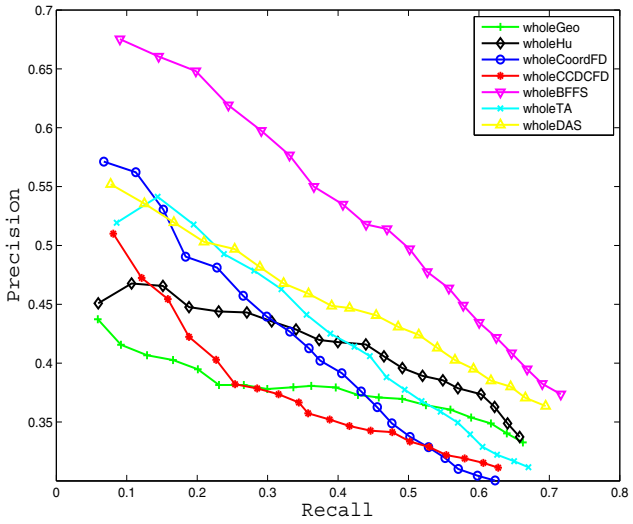


Figure 6. The average precision-recall graph for retrieval using the whole shape

4.2. Comparisons on corner shape data set

We also measure the performance of feature descriptors on the corner shape data set using the average precision-recall and F1 metrics. The results show that for the corner shape data set, feature DAS outperforms other features, which is followed by Turn Angle.

The experiments demonstrate that the retrieval performance of the features are highly dependent on the data set itself. Features should be selected appropriately for a particular shape data set. The proposed learning-based algorithm may contribute to improved retrieval performance by providing a methodology for feature selection that is coupled to the particular types of shapes to be searched.

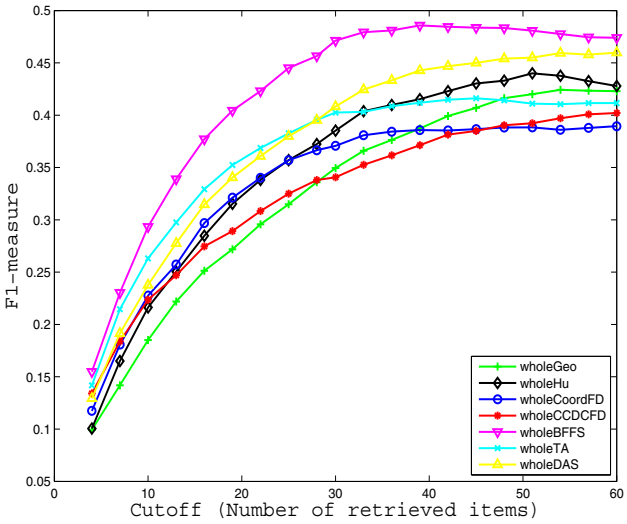


Figure 7. The F1 measure for retrieval using the whole shape

5. Conclusion

A challenge in Content-Based Image Retrieval (CBIR) is selecting features that will result in semantically relevant retrievals. In this paper, we compare the retrieval performance of several feature sets by testing them on two shape databases. We also propose a learning-based feature component selection algorithm to improve retrieval performance, and thus try to minimize the gap between the extracted image features and the pathology semantics. The experimental results show significant improvement with the proposed algorithm.

6. Acknowledgement

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

References

[1] M. Abramowitz and I. Stegun. *Information Retrieval*. Butterworth-Heinemann; 2nd edition, March 1979.

[2] F. Aghdasi, R. Ward, J. Morgan-Parkes, and B. Palcic. Feature selection for classification of mammographic microcalcification clusters. In *Engineering in Medicine and Biology Society, 15th Annual International Conference of the IEEE*, pages 58–59, 1993.

[3] E. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell. An efficiently computable metric for comparing

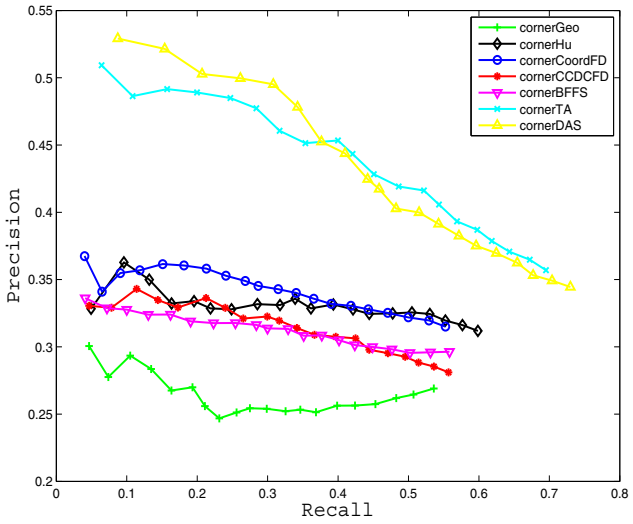


Figure 8. The average precision recall graph on the corner shape database

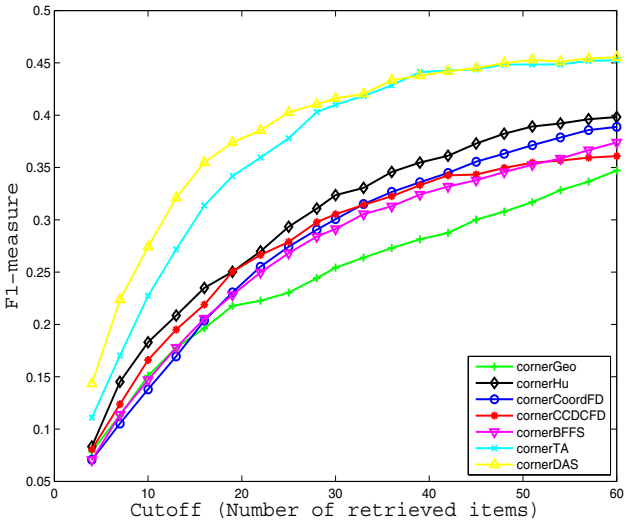


Figure 9. The F1 measure on the corner shape database

polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):209–216, 1991.

[4] L. Chen, R. Feris, and M. Turk. Efficient partial shape matching using Smith-Waterman algorithm. In *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, pages 23–28, Anchorage, Alaska, 2008.

[5] M. Heggeness and B. Doherty. Morphologic study of lumbar vertebral osteophytes. *Southern Medical Journal*, 91(2):187–189, February 1998.

[6] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, February 1962.

[7] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(1):4–37, January 2000.

[8] D.-J. Lee, P. Zhan, A. Thomas, and R. B. Schoenberger. Shape-based human detection for threat assessment. volume 5438, pages 81–91. SPIE, 2004.

[9] L. Long, S. Antani, and G. Thoma. Image informatics at a national research center. *Computerized Medical Imaging and Graphics*, 29:171–193, 2005.

[10] C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[11] H. Mller, W. Mller, D. Squire, S. Marchand-Maillet, and T. Pun. Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters*, 22(5):593 – 601, 2001.

[12] H. Müller, N. Michoux, D. Bandona, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. *Internatinal Journal of Medical Informatics*, 73(1):1–23, 2004.

[13] M. Pelillo, K. Siddiqi, and S. Zucker. Matching hierarchical structures using association graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1105–1120, 1999.

[14] G. Robinson, H. D. Tagare, J. S. Duncan, and C. C. Jaffe. Medical image collection indexing: Shape-based retrieval using kd-trees. *Computerized Medical Imaging and Graphics*, 20(4):209–218, 1996.

[15] F. Schnorrenberg, C. S. Pattichis, C. N. Schizas, and K. Kyriacou. Content-based retrieval of breast cancer biopsy slides. *Technology and Health Care*, 8(5):291–297, 2000.

[16] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571, 2004.

[17] J. Shou, S. Antani, L. Long, and G. R. Thoma. Evaluating partial shape queries for pathology-based retrieval of vertebra. In *Proc. 8th World Multiconference on Systemics, Cybernetics and Informatics*, pages 155–60, July 2004.

[18] H. Tagare. Deformable 2-d template matching using orthogonal curves. *IEEE Transactions on Medical Imaging*, 16(1):108–117, February 1997.

[19] D. Zhang and G. Lu. A comparative study on shape retrieval using fourier descriptors with different shape signatures. 2001.

[20] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, pages 1–19, 2004.