Shape-Based Indexing in a Medical Image Database

Wei Zhang
Department of Biomedical Engineering
Rutgers University, Piscataway, NJ
zhangwei@occlusal.rutgers.edu

Sven Dickinson
Department of Computer Science
Rutgers University, New Brunswick, NJ
sven@cs.rutgers.edu

Stanley Sclaroff
Department of Computer Science
Boston University, Boston, MA
sclaroff@cs.bu.edu

Jacob Feldman
Department of Psychology
Rutgers University, New Brunswick, NJ
jacob@ruccs.rutgers.edu

Stanley Dunn
Department of Biomedical Engineering
Rutgers University, Piscataway, NJ
smd@occlusal.rutgers.edu

Abstract

In this work, we report on a prototype of a clinical radiograph image database to be indexed by image content. The underlying content-based search engine is based on the modal shape description method for characterizing the shape of a 2-D image region [24]. Using a similarity metric defined in a modal vector space, our system successfully classified radiographic images according to the dental pathologies they depicted. This successful classification demonstrates that the proposed similarity metric effectively captures clinical similarity between images in the database. The prototype is implemented in a Web-based environment, allowing remote users in the field to search a central repository of images. Examples of classification performance and typical queries are provided.

1. Introduction

1.1. Background on Content-based Image Databases

Computers, mass storage, and digital imaging systems provide the opportunity to create multimedia electronic databases of patient records instead of the traditional paper and film records. Currently, most contemporary medical image databases can only be indexed by text fields or keywords that have been entered manually. These methods do not directly capture the visual or image properties of the underlying data. Some visual properties are very difficult or nearly impossible to describe with text, such as certain textures or shapes. Also, relations among image objects are very difficult to define or describe by using textual annotation.

Content-based image retrieval refers to image retrieval techniques that are based on visual properties of image objects rather than textual annotation. For example, "given an object in image A, find the most similar one in the database". Such queries would confer enormous advantages: similarity-based matching, interactive search with relevance feedback, a rich query language with which to pose complex queries, and a user-friendly interface. Unfortunately, content-based retrieval is much more challenging than text-based retrieval, and requires the solution of a number of difficult problems in computer vision, databases, and human-computer interaction.

Some automatic image indexing methods have been proposed, based on shape [3, 8, 9, 13, 21, 20], color [2, 7, 25] , or combinations of such indices [15]. The general approach is to calculate some approximate invariant statistic like a color histogram, invariants of shape moments, parametric curve distance, or frequency subband decomposition, and use these to categorize the image database. For example, IBM's Query By Image Content system (QBIC) [15], one of the most advanced 2D image database systems to date, uses color, texture, and shape of image objects and regions. Unfortunately, these methods lack effective models that are robust to noise, scale, and sampling, and models that roughly

correspond to a human's notion of perceptual similarity.

1.2. A Dental Radiograph Database

In this work, we develop a prototype of an interactive system for querying a medical image database by image content, as well as the traditional text query.

In practice, pathological change is progressive, and many considerations are needed to help doctors make a diagnosis, such as radiograph records, clinical history, signs, and symptoms. Our ultimate goal is to build a multimedia database that can provide a fast and convenient mechanism to keep and retrieve both textual and visual information.

In the prototype database described herein, we have considered a sufficiently rich but restricted domain of dentistry and diagnoses based on dental radiographs. The dental radiograph is the most common radiographic examination in this country; the radiographic presentation of disease is based on well-defined image cues (Figure 1). The general features that describe lesions include shape, radiodensity, size, borders, location, and architecture, out of which shape is an important feature for diagnosis and staging disease progression. Accordingly, shape is selected as our primary feature in indexing the image database.

1.3. Shape-based Image Retrieval

Retrieval by shape has been one of the most challenging aspects of content-based image retrieval. No mathematically rigorous definition of shape similarity has been established to account for the various semantic qualities that humans assign to shapes [6, 14]. Shape matching is complicated by the fact that humans will report that shapes are similar even when the two shapes are non-rigidly deformed versions of each other. Especially in the medical domain, challenges arise in representing both the static and dynamic properties of anatomical shapes. In terms of their static shape, anatomical structures can have branches, folds, wrinkles, holes, pores, hairs, etc. In terms of their dynamic shape, anatomical shapes can undergo nonrigid deformation and/or articulated motion; the associated image data sets can depict organs whose shape deforms over time, over the progression of a disease, or from patient to patient. Some examples of shape deformation over time can be found in Figure 2.

While a wide array of shape-based description and matching methods have been proposed, few can explicitly accommodate the nonrigid deformations that occur frequently in the types of objects appearing in medical images. For example, in IBM's QBIC system [15], several shape features are utilized, including area, circularity, eccentricity, major axis of inertia, and higher-order algebraic moment invariants. These shape features are combined into

Figure 1. A Template Used to Describe a Lesion Appearing in a Dental Radiograph.

one feature vector, and shape similarity is measured using a weighted Euclidean distance metric. However, the shape-based search used in this system cannot differentiate between noise and a class-preserving deformation.

Furthermore, it is difficult to objectively validate similarity measures used for indexing, because intuitive similarity (as perceived by the system designers) may not correspond well with meaningful similarity in the database. Hence we conduct an experiment evaluating the ability of our proposed similarity measure to classify shapes into diagnostic categories as judged by expert clinicians in the field. Good performance in this experiment provides objective evidence that the similarity measure corresponds well with clinical similarity in this medical domain.

Energy-based model formulations, e.g., [27], have provided a framework for underconstrained shape recovery. They provide a useful method for interpolating and smoothing raw data, while also capturing information about nonrigid deformations. In theory, these models also offer the advantage that they can incorporate prior knowledge about the physical material properties for a specific anatomical
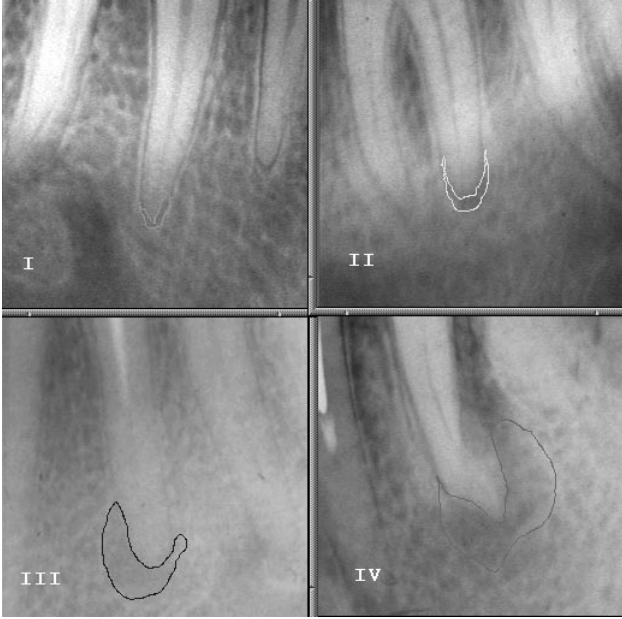
Figure 2. The Four Stages of Periapical Disease: (I) A normal periodontal area; (II) A widened periodontal ligament space; (III) Interruption of lamina Dura; and (IV) Bone resorption

structure. Unfortunately, many of these physical properties are not available, and most models assume isotropic material properties.

Also, current energy-based and statistical shape modeling methods are computationally intensive and are therefore not fast enough to meet the interactivity needs of our problem domain. In addition, many of these deformable shape representations cannot be easily used for recognition, because they are mesh-based. The only general method for determining if two mesh surfaces are equivalent is to generate a number of sample points at corresponding positions on the two surfaces, and observe the distances between the two sets of sample points. Such methods are costly and can easily fail when two equivalent surfaces have very different parameterizations.

1.4. Efficient Shape Description

In formulating anatomic shape representations, we need to simultaneously ensure that the generated shape descriptions can be efficiently recovered *and* that they can be efficiently employed in shape alignment, comparison, and recognition. In our system, we draw on modal shape description [17, 24] to address the above problems. This is a physically-motivated method for computing canoni-

cal shape descriptions that embodies two main techniques: the finite element method (FEM) and eigendecomposition. The FEM provides greater robustness to sampling error and noise, while the eigendecomposition method results in an orthogonal and frequency-ordered shape description which enables efficient shape alignment, comparison, and recognition. The underlying physics of the model makes it particularly capable in dealing with different types of nonrigid deformations. Moreover, the deformation parameters correspond qualitatively with those used by humans in describing shape similarity, and thus provide a meaningful interface for database query.

1.5. Overview of the Paper

This paper is organized as follows: Section 2 is a review of the finite element method, a physically-based mechanism to better deal with sampling and scale problems that exist in general physical models. Modal shape description, which provides robust and unique shape representation, is also introduced. In Section 3, we describe several alternatives for the statistical characterization of the modal shape description of lesions. Several types of similarity (or dissimilarity) measures are introduced to characterize non-rigid shape deformations at various stages of dental disease.

In Section 4, we summarize experiments in using modal shape descriptors to classify (nonrigid) shape deformations due to pathological change, as presented on the dental radiograph. The successful classification performance means that the modal shape descriptors can be used to index the database based on content. Section 5 outlines the system design and implementation based on WWW technologies, including the mechanisms for interactive search. Results are shown, in the form of a set of snapshots, for a typical shape query. Section 6 gives a brief conclusion and discusses future work.

2. Modal Shape Description

2.1. Overview

To better model nonrigid deformation of real-world objects, their underlying physical properties must be accounted for. This has motivated the use of physically-based, deformable models to exploit a priori knowledge about physics to interpolate and smooth raw data based on prior assumptions about noise, material properties, etc.

The Finite Element Method (FEM) is a standard engineering technique for simulating the dynamic behavior of an object. It effectively deals with the sampling and scale problems that exist in the general physical models. This approach was employed in modeling deformable superquadrics by Pentland and Sclaroff [16, 18] and later by

Terzopoulos and Metaxas [26]. It has also been applied in other models, e.g., snakes, blob models, etc. [5, 10, 4, 5].

The Finite Element Method, however, can not be directly used for comparing objects due to the non-uniqueness problem. Virtually all spline, thin-plate, and polynomial methods suffer from an inability to obtain canonical descriptions because the parameters for these surfaces can be arbitrarily defined, and therefore are not invariant to changes in viewpoint, occlusion, or nonrigid deformations [24].

In the *modal analysis* introduced by Pentland and Sclaroff [18], the FEM is simplified by posing the dynamic equations in terms of the equation's eigenvectors. These eigenvectors are known as the object's deformation modes, or *eigenmodes*, and together form a frequency-ordered orthonormal basis set for shape. Modal analysis subsequently was applied to shape recovery and recognition [16, 18, 19, 23] and nonrigid motion tracking [17].

The eigenmodes may be thought of as the shape's *generalized symmetry axes* [23], in that they describe the principal axes of deformation over a training set of shapes. This generalized coordinate system provides a robust representation for establishing correspondence between similar shapes. Most importantly, the orthogonality of the eigen-representation ensures that the recovered descriptions will be unique, thus making recognition problems tractable.

The solution to the problem of nonuniqueness is to discard enough of the high-frequency modes, so that an over-constrained estimate of shape is obtained. The modes form an orthonormal basis set; therefore, there is only one way to represent an object that is in canonical position. Also, because irregularities in local sampling and measurement noise tend to primarily affect the high-frequency modes rather than the low-frequency modes, the reduced-basis modal representation results in a unique representation of shape . The modal parameters are also physically meaningful, corresponding to how much bending, tapering, pinching, etc., is needed to represent an object. The mathematical formulations of FEM and modal analysis described in the following sections are taken from [1, 24].

2.2. Finite Element Method

In the FEM, interpolation functions are developed that allow continuous material properties, such as mass and stiffness, to be integrated across the region of interest. In general, the polynomial shape function for each element is written in vector form as:

$$u(x) = H(x)U \tag{1}$$

where H is the interpolation matrix, x is the local coordinate of a point in the element where we want to know the displacement, and U denotes a vector of displacement components at each element node.

The polynomial shape function can be used to calculate the strain due to deformation. Strain ϵ is defined as the ratio of displacement to the actual length and can be obtained by:

$$\epsilon(x) = B(x)U \tag{2}$$

where B is the strain displacement matrix. The rows of B are obtained by appropriately differentiating and combining the rows of the element interpolation matrix H.

To deform an elastic body to match the feature points requires solving the *dynamic equilibrium equation:*

$$M\ddot{U} + D\dot{U} + KU = R \tag{3}$$

where R is the load vector and M, D, K are the element mass, damping, and stiffness matrices, respectively.

Both the mass and stiffness matrices are computed directly:

$$M = \int_V \rho H^T H dV \tag{4}$$

and

$$K = \int_V B^T C B dV \tag{5}$$

where ρ is the mass density, and C is the *material matrix* that expresses the material's particular stress-strain law (see [24] for explicit formulas).

Under the assumption of Rayleigh damping, the damping matrix is simply a linear combination of the mass and stiffness matrices:

$$D = \alpha M + \beta K \tag{6}$$

where α and β are constants determined by the desired critical damping [1].

2.3. Modal Analysis

It turns out that the dynamic equation can be decoupled by posing the equation in a basis defined by the M-orthonormalized eigenvectors of $M^{-1}K$ [24]. These eigenvectors and eigenvalues are the solution (ϕ_i, ω_i^2) to the following generalized eigenvalue problem:

$$K\Phi = M\Phi\Omega^2 \tag{7}$$

where

$$\Phi = [\phi_1 | \cdots | \phi_m]$$

$$\Omega^2 = \begin{bmatrix} \omega_1^2 & & \\ & \ddots & \\ & & \omega_m^2 \end{bmatrix}$$

The vector ϕ_i is called the ith mode shape vector, and ω_i is the corresponding frequency of vibration. The mode shape vectors can be thought of as describing the object's

generalized (nonlinear) axes of symmetry. Each mode shape vector ϕ_i is **M**-orthonormal,

$$\Phi^T K \Phi = \Omega^2, \quad \Phi^T M \Phi = I \qquad (8)$$

This generalized coordinate transform Φ is then used to define displacements $\tilde{U}$:

$$U = \Phi \tilde{U} \qquad (9)$$

If we rewrite equation 3 in terms of these generalized or modal displacements, we obtain a decoupled system of equations:

$$\ddot{\tilde{U}}_t + \tilde{D}\dot{\tilde{U}} + \Omega_t^2 \tilde{U}_t = \phi_t^T R \qquad (10)$$

where $\tilde{D}$ is the diagonal modal damping matrix. By decoupling these equations, a closed-form solution to the equilibrium problem is obtained [18].

Consider the eigenvectors, ϕ_i's, and eigenvalues, ω_i's, of the matrix $M^{-1}K$. The ω_i corresponds to frequency of vibration, allowing us to re-order the eigenvectors by increasing eigenvalue,

$$\Phi = [\phi_1 | \ldots | \phi_{2m}] = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \\ v_1^T \\ \vdots \\ v_m^T \end{bmatrix} \qquad (11)$$

where m is the number of nodes used to build the finite element model. ϕ_i is called the ith *mode*, while the row vectors u_i and v_i are called the ith *generalized feature vectors*. Together, they describe the feature's location in the modal coordinate system [23].

The low-order eigenvectors correspond to the low-frequency modes. For example, in the two-dimensional problem, the lowest three modes correspond to the rigid modes of two translations and one rotation of shape. Also, because the higher-frequency modes are more sensitive to noise, we can truncate them in certain circumstances. After mode truncation, a unique, over-constrained canonical shape description is obtained. This shape matrix preserves many desirable properties, including rotation and scale invariance, insensitivity to sampling errors, parameterization errors, etc.

3. Classifying Shape Using Modal Features

To make content-based search meaningful in a particular medical domain, the features used in a database query must have a statistically significant relationship to the pathologies reflected in the images. In our experiment, four groups

of images, each reflecting one state of periapical disease as judged by an expert clinician, are characterized by modal description. We study a similarity measure based on deformation alignment energy, and design a classifier to evaluate the distinction of modal features of the four image groups. The successful performance of this classifier suggests that the proposed modal similarity measure is useful for indexing the database based on clinical shape content.

3.1. Shape Similarity Measure

Because the modal descriptions are physically-based, we can compute and compare the amount of deformation energy needed to align one object with another. All shapes can be represented as deformations from a standard or prototypical shape; therefore it is possible to compare two shapes by comparing their respective deformations from a common prototype.

We begin by studying the displacement distance between two shapes. When the correspondence of points from the two images has been established, the modal deformation parameters $\tilde{U}$ can be solved by *nodal* displacements U that align corresponding features on both shapes:

$$u_i = x_{1i} - x_{2i} \qquad (12)$$

where x_{1i} is the ith node on the first shape and x_{2i} is its matching node on the second shape.

Recalling that $U = \Phi \tilde{U}$, and using Equation 9, we get:

$$\tilde{U} = \Phi^{-1} U = \Phi^T M U \qquad (13)$$

When strong one-to-one correspondences between features cannot be established, we adopt the "physical simulation" method introduced in [24, section 5.1.3]. Basically, the finite element equations are integrated over time until equilibrium is achieved. This method drives the unmatched nodes to move in a manner consistent with the material properties and the forces at the matched nodes.

Suppose $\tilde{U}_1$ and $\tilde{U}_2$ are two modal descriptions that align the prototype with each candidate object. The normalized dot product of two vectors can be used as a shape similarity metric [18]:

$$\epsilon(\tilde{U}_1, \tilde{U}_2) = \frac{\tilde{U}_1 \tilde{U}_2}{\|\tilde{U}_1\| \|\tilde{U}_2\|} \qquad (14)$$

The similarity measure between two group of images, say $G1$ and $G2$, is defined as the average point-to-point shape similarity metric between two groups:

$$S(G1, G2) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \epsilon(\tilde{U}_i, \tilde{U}_j) \qquad (15)$$

where the U_i's belong to group $G1$, the U_j's to group $G2$, and M and N are the number of images in groups $G1$ and $G2$, respectively.

3.2. Classification Error Probability

Fundamentally, the statistical characteristics of the modal shape description are reflected in the separability of feature distributions, which is independent of the the classifier. Such problastic separation measures include Kullback distance and divergence, among others [11, 28]. Those measures usually require a very large amount of sample data, and indicate only the asymptotics or the limitation of how fast a classifier can improve its performance given a large number of data. It is always desirable to find out how a classifier can perform using the modal features, particularly with a limited number of data or, if possible, find out what the optimal classifier would be. Next, we present a general-purpose classfier using the minimum Mahalanobis distance criterion and then evaluate classifier performance using the "leave-one-out" method.

We assume the modal displacement vectors are normally distributed in modal space. The general multivariate normal density, with dimension d, is written as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{x}-\mu)^{\mathbf{t}}\Sigma^{-1}(\mathbf{x}-\mu)] \quad (16)$$

where x is a d-component column vector, μ is the d-component *mean vector*, and Σ is the d-by-d *covariance matrix*. This equation is abbreviated as $p(\mathbf{x}) \sim N(\mu, \Sigma)$.

Samples drawn from a normal population tend to fall in a single cloud or cluster. The center of the cluster is determined by the mean vector, and the shape of the cluster by the covariance matrix. It follows from Eq. 16 that the loci of points of constant density are hyperellipsoids for which the quadratic form $(x-\mu)^t\Sigma^{-1}(x-\mu)$ is constant. The principal axes of these hyperellipsoids are given by the eigenvectors of Σ, the eigenvalues determining the lengths of these axes. The quantity

$$r^2 = (\mathbf{x}-\mu)^{\mathbf{t}}\Sigma^{-1}(\mathbf{x}-\mu) \quad (17)$$

is sometimes called the squared *Mahalanobis distance* from x to μ. Thus, the contours of constant density are hyperellipsoids of constant Mahalanobis distance to μ. The volume of these hyperellipsoids measures the scatter of the samples about the mean.

We adopt the so-called "leave-one-out" method to maximize usage of data in our experiment. In many classification problems, one must decide how to use sample data to both train and test the classifier. If all data are used to both train and test the classifier, the (estimate of) the performance would be overly optimistic. On the other hand, if only a small number of the samples is used to train the classifier, poor estimates of the classifier parameters may result in inadequate classification performance. Alternatively, if one uses only a small number of samples for testing, the estimated error performance may not be reliable.

The "leave-one-out" method is an alternative approach that yields independent training and test sets, and provides the upper bound. In this method, a classifier is designed using all of the available samples except one, and the remaining sample is used to test the classifier. This procedure is then repeated each time, leaving out a different sample. The basic advantage of this approach is that virtually all of the samples are used in each design, which should lead to a good design, and all of the samples are ultimately used in the test. A Monte Carlo study by Lachenbruch and Mickey [12] gives evidence for the superiority of this method.

4. Evaluating the Modal Shape Description

In this section, we report the test results of the statistical analysis of our modal shape description discussed in Section 3. Our experimental results strongly support the use modal shape vectors as features in classifying pathologies reflected in dental radiographs. Our test results also indicate that, because of orthogonality of modes, the less critical components can be discarded to obtain overconstrained, canonical descriptions in order to provide efficient data reduction, real-time recognition, and robust reconstruction.

4.1. Test Data

There are a wide variety of diseased states which represent different concepts in the dental radiographs. Figure 2 in Section 1 shows examples that have been selected for this study: four states, including one normal case and three levels of progressively diseased states. The first level of disease is evidenced by a widened periapical ligament space, the second level shows destruction of the lamina dura, and the third stage involves resorption of alveolar bone in the periapical area. The first and second stages of disease are generally limited to one root, while the third stage may extend to the roots of the adjacent teeth.

Our test data consist of four groups of radiographs, labeled $G1$ to $G4$, with each group (8 radiographs) representing one of the four categories of periapical disease. $G1$ is normal, $G2$ is widened ligament space, $G3$ is interrupted lamina dura, and $G4$ is bone resorption. Each radiograph is digitized as a 256×256 pixel image. Each lesion contour is drawn by an expert dentist, and then a prototype ellipse is deformed to match the shape of the lesion. The minimum number of modes (eigenvalue of stiffness matrix) in all groups is 28, and the maximum number is 58. We take the first 25 modes, and the corresponding modal displacement vector for each shape is used as the input to our classifier. In Figures 3–6, we show the results of deforming an ellipse to the outlined lesion region.
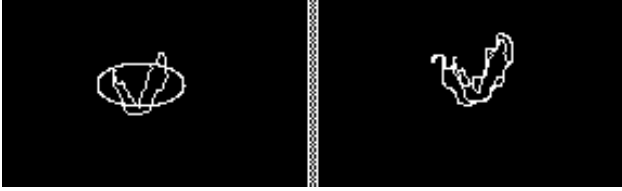
6

Figure 3. Examples of deformation result, Group 1

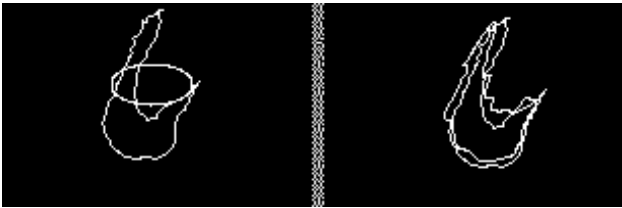**Figure 4. Examples of deformation result, Group 2**

Figure 5. Examples of deformation result, Group 3
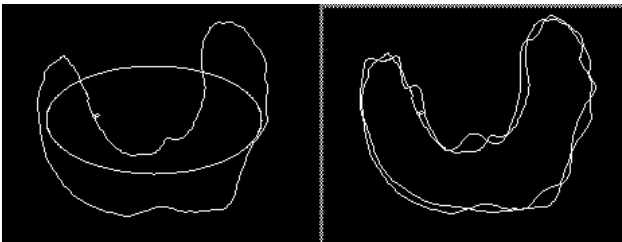

Figure 6. Examples of deformation result, Group 4

We take one sample from each group. Shown in the left column are the prototype shape and the deformation target. In the right column are the shapes after deformation.

4.2. Separation of Test Groups in Modal Space

For the modal space to be a sufficiently powerful representation and basis for the proposed content based search engine, the representations of the classes of pathology must be well separated. That is, the inter-class variations between all pairs must be greater than the individual intra-class variations; this notion is captured by the similarity metric in Equation 15. The results are shown in Figure 7, in which S_{ij} denotes the similarity measure between the ith and j^{th} image groups. The inter-group shape similarity measure distinctively decreases from group $G1$ to $G4$, which indicates that the modal shape description could serve as a good feature in reflecting the four stages of periapical disease. This trend is also evidenced in Table 1, where we choose a mode subset, modes 6 through 9, to calculate the inter-group similarity measure.
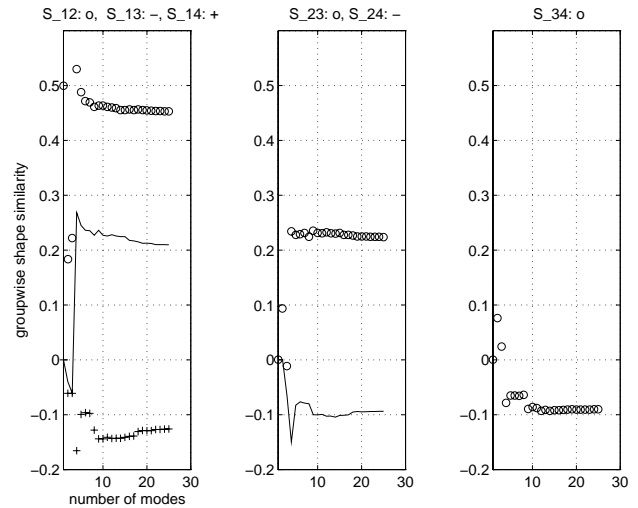
Figure 7. Inter-group shape similarity varying with number of modes

	group $G1$	group $G2$	group $G3$	group $G4$
$G1$		0.2974	0.1720	-0.3674
$G2$			0.3274	-0.2791
$G3$				-0.2423

Table 1. Inter-group similarity measure with $6-9$ mode subset

4.3. Classification Error Probability

We built a *Minimum Mahalanobis Distance Classifier* following the discussion in section 3.2. First, we estimate the mean and covariance of each group using the Maximum Likelihood [29] method. Then, for each test sample, we find the group that is closest to that sample using the Mahalanobis Distance, and classify the test sample into that group. The "leave-one-out" method is used to maximize usage of the test sample into that group. Following the discussion on retaining only low-order modes, we discard the first 5 modes and take the next 6 modes (6 through 11). Calculating more modes would have resulted in a singular covariance matrix estimate due to insufficient sample data.

We computed the performance of the classifier against the number of modes. Let P_{ji} be the probability of an image in group i being classified into group j. We found that the modal subsets in Table 2 optimize the classification probability (P_{ii}'s), and that the subset that maximizes ($\sum_{i=1}^{4} P_{ii}$) is set $\{6, 7, 8, 9\}$, as shown in Table 3. If we aggregate the 4 groups into normal (group $G1$) and abnormal (group $G2$, $G3$, $G4$) groups, then the classification performance is shown in Figure 8, with the best performance again from subset $\{6, 7, 8, 9\}$. All these results indicate that because modal shape descriptions decompose shape information into an ordered basis of orthogonal principal components, the less critical and often noisy high-order components can be discarded to obtain overconstrained, canonical descriptions. This allows for the selection of only the most important components to be used for efficient data reduction, real-time recognition, and robust reconstruction.

	characteristic mode set	$\max P_{ii}$
$G1$	7-9	87.5%
$G2$	8-9	75.0%
$G3$	7-9	62.5%
$G4$	6-8	62.5%

Table 2. Classification Performance based on individually optimal mode subset

4.4. Summary

We investigated the statistical characteristics of the modal shape description in classifying nonrigid shape deformations due to various pathologies reflected in dental images. Good classification performance based on modal features is achieved. There exists some characteristic mode subset that best characterizes a particular disease stage or abnormality. Modal shape descriptors seem to provide a

	group $G1$	group $G2$	group $G3$	group $G4$
$G1$	87.5%	12.5%	0	0
$G2$	25%	62.5%	12.5%	0
$G3$	25%	12.5%	62.5%	0
$G4$	12.5%	12.5%	12.5%	62.5%

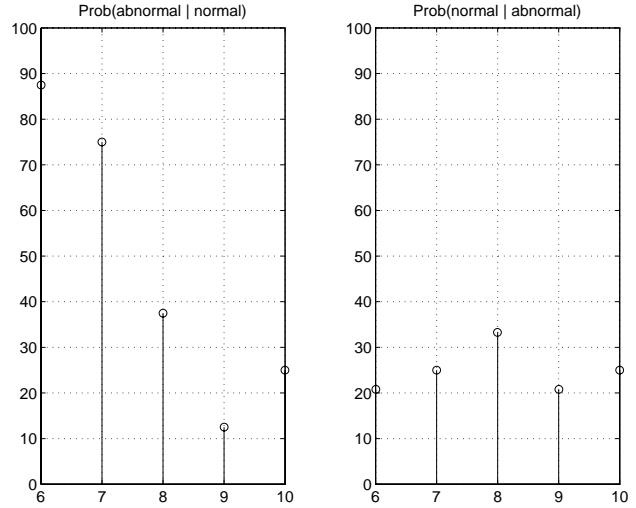**Table 3. Classification Performance based on the overall optimal mode subset**

Figure 8. Error probabilities for aggregated normal and abnormal group.

possible search index for retrieving similar lesions from our dental image database.

5. Prototype Application

5.1. System Implementation

Our prototype image retrieval system is designed for Internet access. Both text-based and content-based search are supported; For a text-based search, a user can fill in keyword information, such as patient name, to obtain multimedia patient archive records that contain personal information, clinical history, and radiographs. For content-based image search, a user can select (or load) an image for study, interactively draw a contour around a lesion to be interpreted, and request images (and their corresponding patient records) containing lesions whose shapes are similar to the query lesion.

The interactive operation is as follows: An outline of a

lesion contour from the user is tracked, recorded and sent to a server. The server passes the contour to the content search engine where the target lesion shape is compared to those in the database using the modal shape description. The search engine returns the indicies of the "most similar" lesions to the server, which are used to produce a display of the most similar lesions.

An alternative method for comparing two shapes is to compute the amount of deformation energy needed to align an object [23]. When correspondence between two modal shape features is established, the modal displacement can be obtained using Equation 13, and the strain energy is simply:

$$E = \frac{1}{2}\tilde{U}^T \Omega^2 \tilde{U} \qquad (18)$$

Although also not yet implemented, we can utilize the strain-energy distance metric to order objects based on their similarity to a prototype [22]. In such a scenario, a few prototypes are selected to span the variation of shape in each category. Every shape in the database is then aligned with each of the prototypes using modal matching, and the resulting modal strain energy is stored as an m-tuple, where m is the number of prototypes. Each shape in the database now has a coordinate in this low-dimensional space, and shapes can be compared in terms of their Euclidean distance in this space.

We now present a few snapshots of our system. Figure 9 shows the graphical user interface. A user can select a particular image from the list on the top and then use the mouse to outline the target lesion in the image. This drawing is recorded and sent to the database server which handles lesion search, including the computation of its modal description. The query results, in the form of image icons, are sent back to the user, as shown in Figure 10. Detailed lesion characterization can be displayed by clicking on a particular image icon, as shown in Figure 11. From here, the user can track other related clinical information about the patient, including clinical history, other radiographs, diagnosis notes, etc., although not shown in this paper.

6. Conclusion and Future Work

In this work, we successfully developed a prototype of a clinical radiograph image database retrieval system indexed by image content. The underlying content-based search engine is based on the modal shape description method. We investigated the statistical characteristics of modal shape description via a number of approaches. Using a similarity metric defined in the modal vector space, we successfully classified (nonrigid) shape deformations due to pathological changes. This indicates that the modal shape descriptors can be used to index the database based on content.

Our (more near-term) goal is to build specialized, deformable shape models that can learn priors for a particular organ or tumor type. These "organ agents" would then be used to segment and track anatomical structures. It would be of interest to investigate combining other image description methods (such as texture, radiodensity, etc.) with shape descriptors in order to index the database more effectively.

References

[1] K. Bathe. *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, 1982.

[2] E. Binaghi, I. Gagliardi, and R. Schettini. Indexing and fuzzy logic-based retrieval of color images. *Proc. Visual Database Systems II, IFIP Transactions A-7*, pages 79–92, 1990.

[3] Z. Chen and S. Y. Ho. Computer vision for robust 3d aircraft recognition with fast library search. *Pattern Recognition*, 24(5):375–390, 1991.

[4] I. Cohen, N. Ayache, and P. Sulger. Tracking points on deformable objects. In *European Conference on Computer Vision*, Santa Margherita Ligure, Italy, May 1992.

[5] L. Cohen and I. Cohen. A finite element method applied to the new active contour models and 3-d reconstruction from cross sections. In *Third International Conference on Computer Vision*, Dec. 1990.

[6] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3:231–262, 1994.

[7] M. Ioka. A method of defining the similarity of images on the basis of color information. Technical Report RT-003 0, IBM Tokyo, 1989.

[8] M. A. Ireton and C. S. Xydeas. Classification of shape for content retrieval of images in a multimedia database. In *Proc. 6th International Conf. on Digital Processing of Signals in Comm.*, pages 111–116, 1990.

[9] H. V. Jagadish. A retrieval technique for similar shapes. In *Proc. International Conf. on Management of Data, ACM SIGMOD 91*, pages 208–217, 1991.

[10] P. Karaolani, G. Sullivan, and K. Baker. Active contours using finite elements to control local scale. In *British Machine Vision Conference*, pages 481–487, 1992.

[11] S. Kullback. *Information Theory and Statistics*. Dover Publications, Inc., 1978.

[12] P. A. Lachenbruch and M. R. Michey. Estimation of error rates in discriminant analysis. *Technometics*, 10:1–11, Feb. 1968.

[13] R. Mehrotra and W. I. Grosky. Shape matching utilizing indexed hypothesis generation and testing. *IEEE Trans. of Robotics and Automation*, 5(1):70–77, 89.

[14] D. Mumford. Mathematical theories of shape: Do they model perception? *Geometric Methods in Computer Vision, SPIE*, 1570:2–10, 1991.

[15] W. Niblack, R. Barber, W.Equitz, M. Flickner, E. Glasman, D. Petkovic, and P. Yanker. The qbic project: Querying images by content using color, texture, and shape. *In Proc.*

SPIE Conf. on Storage and Retrieval of Image and Video Databases, 1908, 1993.

[16] A. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, 4(2):107–126, Mar. 1990.

[17] A. Pentland and B. Horowitz. Recovery of non-rigid motion and structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):730–742, 1991.

[18] A. Pentland and S. Sclaroff. Closed-form solutions for physically-based shape modeling and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):715–729, July 1991.

[19] A. Pentland and J. Williams. Good vibrations: Modal dynamics for graphics and animation. *Computer Graphics*, 23(4):215–222, 1989.

[20] K. Perez-Lopez and A. Sood. Comparison of subband features for automatic indexing of scientific image databases. In *In Proc. SPIE Conf. on Storage and Retrieval of Image and Video Databases*, volume 2185, 1994.

[21] B. Scassellati. Retreving images by 2d shape: a comparison of computation methods with human perceptual judgments. In *In Proc. SPIE Conf. on Storage and Retrieval of Image and Video Databases*, volume 2185, 1994.

[22] S. Sclaroff. Deformable prototypes for encoding shape categories in image databases. *Pattern Recognition*, 30(4):627–642, April 1997.

[23] S. Sclaroff and A. P. Pentland. Modal matching for correspnondence and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.

[24] S. E. Sclaroff. *Modal Matching: A Method for Describing, Comparing, and Manipulating Digital Signals*. PhD thesis, MIT, Feb. 1995.

[25] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[26] D. Terzopoulos and D. Metaxas. Dynamic 3-d models with local and global deformations: Deformable superquadrics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):703–714, July 1991.

[27] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3d shape and nonrigid motion. *Artificial Intelligence*, 36:91–123, 1988.

[28] C. W. Therrien. *Decision Estimation and Classification*. John Wiley and Sons, 1989.

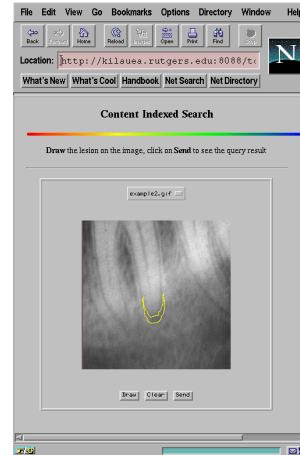[29] H. L. V. Trees. *Detection, Estimation, and Modulation Theory*. John Wiley and Sons, Inc., 1968.
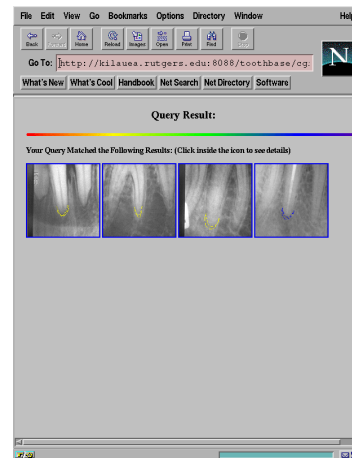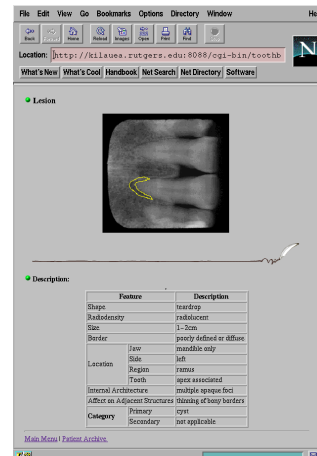
Figure 9. Lesion Query

Figure 10. Lesion Query Results

Figure 11. Lesion Description