

# Population-Based Incremental Interactive Concept Learning for Image Retrieval by Stochastic String Segmentations

Sennay Ghebreab\*, C. Carl Jaffe, and Arnold W. M. Smeulders

**Abstract**—We propose a method for concept-based medical image retrieval that is a superset of existing semantic-based image retrieval methods. We conceive of a concept as an incremental and interactive formalization of the user's conception of an object in an image. The premise is that such a concept is closely related to a user's specific preferences and subjectivity and, thus, allows to deal with the complexity and content-dependency of medical image content. We describe an object in terms of multiple continuous boundary features and represent an object concept by the stochastic characteristics of an object population. A population-based incrementally learning technique, in combination with relevance feedback, is then used for concept customization. The user determines the speed and direction of concept customization using a single parameter that defines the degree of exploration and exploitation of the search space. Images are retrieved from a database in a limited number of steps based upon the customized concept. To demonstrate our method we have performed concept-based image retrieval on a database of 292 digitized X-ray images of cervical vertebrae with a variety of abnormalities. The results show that our method produces precise and accurate results when doing a direct search. In an open-ended search our method efficiently and effectively explores the search space.

**Index Terms**—Content-based image retrieval, multifeature object description, population-based incremental learning, relevance feedback, visual concept learning.

## I. INTRODUCTION

CONTENT-BASED image retrieval answers the desire to explore large repositories of digital images, which are ubiquitously growing into statistically large numbers. Finding a visually particular image by means other than the verbal description of that image in the written medical record is by no means a trivial task, especially when the differences between normal and abnormal manifest themselves by subtle visual dissimilarities. The medical record contains a condensed representation of the essential aspects of the image and,

therefore, it almost always disregards subtle visual patterns. When all information of the written record is exhausted can we still utilize subtle visual patterns to relate or discover small phenomena? In this paper, we strive to design an adequate content-based image retrieval method that takes into account the demands created by medical images.

Tagare *et al.* [1] postulate that 1) medical images need to be indexed by object rather than by image features, 2) object features need to be formalized iteratively rather than at only one instance in order to deal with the complex and imprecise content of medical images, and 3) object features need to be formalized interactively rather than automatically to deal with the subjectivity and context-relatedness of medical image content. We agree, as 1) objects are the target of attention, 2) the process of defining features should be rich enough to incrementally converge on the phenomena under study, and 3) interactive retrieval and search satisfaction are the primary purpose of nontext based image retrieval. Hence, we suggest an approach that first learns the user's conception of an object in an image and then tries to find images depicting similar objects.

This work addresses the problem of how to incrementally and interactively learn the user's conception of an object in a query image in order to retrieve objects from a database that best match this conception or are in some way related. The paper is organized as follows. In Section II, we elaborate on existing medical image retrieval methods. Section III is devoted to a method for concept-based medical image retrieval by browsing population-based incrementally learned image segmentations. The key features of the method are discussed step by step using schematic illustrations. Concept-based image retrieval is demonstrated in Section IV. The conclusion follows in Section V.

## II. RELATED WORK

A large number of content-based image retrieval methods have been proposed in the literature, ranging from methods based on low-level visual features to more advanced methods based on the semantics of visual data (see [2] for an extensive review). Increasingly, methods tackle the problem of concept-based image retrieval, i.e., retrieval on the basis of the user's perceptual subjectivity and specific preferences within a given semantic framework. Bhanu *et al.* [3], for example, aim at visual concept learning from prior experience with previous queries by various users. Fuzzy clustering and relevance feedback are their main mechanism for visual concept-learning. Other approaches that exploit relevance feedback and some sort

Manuscript received August 8, 2003; revised February 12, 2004. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M.A. Viergever. *Asterisk indicates corresponding author.*

\*S. Ghebreab is with the Biomedical Imaging Group Rotterdam, Room Ee2167, Departments of Radiology and Medical Informatics, Erasmus MC Rotterdam, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands (e-mail: s.ghebreab@erasmusmc.nl).

C. C. Jaffe is with the Cardiovascular Medicine Section, Yale University, School of Medicine, New Haven, CT 06510 USA (e-mail: carl.jaffe@yale.edu).

A. W. M. Smeulders is with the Intelligent Sensory Information Systems Group, Informatics Institute, Faculty of Science, University of Amsterdam, 1098 SJ Amsterdam, The Netherlands (e-mail: smeulders@science.uva.nl).

Digital Object Identifier 10.1109/TMI.2004.826942

of clustering of image categories in order to personalize image semantics include [4]–[6]. Most concept learning and relevance feedback methods (see [7] for a review) have been designed for real-world images and/or their text annotations [8], without addressing the specific characteristics of medical images. As a consequence, they often lack the capability to selectively focus on subtle visual (dis)similarities between objects, which is essential for medical image retrieval.

In the remainder of this section, we will briefly discuss a number of methods that do take into account the demands created by medical images or that exploit opportunities provided by them. Following [1], we do this along the three dimensions of the “content understanding–query completion–user interaction space”: the extent of understanding and reasoning about the image content, the ease with which the query mechanism allows the user to specify what the user wants and the extent of interaction required at data entry or during image retrieval.

Medical image content understanding facilitates the identification of semantically meaningful features that are suitable for image retrieval. Human knowledge, consolidated in knowledge bases, is often used to elevate content-understanding. A number of methods exploit consolidated human delineation of objects or regions in an image. Thoma *et al.* [9] use manual delineations of vertebral boundaries in X-ray images, Euripides [10] uses *a priori* specified spatial relationships between boundaries and Brodley *et al.* use delineations of pathology bearing regions in HRCT lung images. A different approach is exploitation of human image classification rather than object delineation. Liu *et al.* [11] propose a classification-driven semantic-based retrieval method that aims at identifying meaningful features by evaluating how well these features perform on the task of classifying medical images according to predefined pathological cases. Mojsilovic [12] *et al.* also follow this principle. Keyser *et al.* [13], go one step further and classify images according to image modality, body orientation, anatomic region and biological system. Other methods exploit elaborate contextual knowledge bases, e.g., [14]–[16]. We stress that in contemporary medical image retrieval, content understanding is reached at the level of semantics, with semantically meaningful features being identified at data entry and being employed for visual matching during image retrieval.

Query completion is crucial for understanding the semantically meaningful features that are of interest to the user in a particular context. In most content-based medical image retrieval methods, query completion is limited to the specification of a single sketch [9], query object, image region of interest [17] or image property. A few methods go beyond a single query specification. In [16], Chbeir *et al.* propose a method for meta-data-based and content-based medical image retrieval with an ambiguity resolver. After the user has posed a query in terms of conceptual, physical, spatial or semantical features, ambiguities can be discerned to force the user to formulate better queries. Liu *et al.* [11] offer the user the possibility to indicate which of the preselected features is best suited for image retrieval by weighting a subset of potentially discriminating features. Brodely *et al.* [18] propose a more interactive approach to query completion. In their method, the user can elicit feedback on each of the retrieved images. If the user disagrees with

the system, that information can be used to alter the weights of the factors in their similarity metric. We conclude that contemporary medical image retrieval methods generally offer a low degree of query completion, with rudimentary versions of relevance feedback being the most advanced mechanism to iteratively define an interactive query.

In medical image retrieval, interaction is almost always required at data entry or during image retrieval. In [11] the user only selects an example image; no image segmentation is required to compute regional or object features and there is no opportunity for user feedback during retrieval to increase the degree of query completion. In [9] the aim with regard to user interaction is to minimize the frequency and the complexity of user assistance, so that the method may be used with the only human interaction required being expert input at data entry or query specification. The method described in [13] requires a high degree of user interaction at data entry. The user also acts to commence image retrieval, but no user interaction is performed during the image retrieval process. The method in [18] supports the user-in-the-loop paradigm. The user delineates regions in the image at data entry and to query the image. Apart from this, the user is actively engaged in the relevance feedback process. In short, even though user interaction in medical image retrieval is almost always present and occurs at different stages during the retrieval process, only few methods aim at integrating interaction and computation in order to effectively and efficiently exploit human knowledge.

We conclude that the majority of methods concentrate on content-understanding, addressing query completion as a side issue or completely disregard it, and on minimizing user interaction. As a consequence, the use of content-based image retrieval methods has been limited to applications where anatomical structures and differences between different classes of anatomical structures are easily captured by a limited set of predefined or automatically extracted visual features.

### III. A METHOD FOR CONCEPT-BASED IMAGE RETRIEVAL

We depart from the view that content understanding, query completion and user-interaction are complementary, i.e., user interaction is mandatory for query completion and query completion is mandatory for content-understanding. In the following sections we focus on an integral approach to content-understanding, query completion and user interaction to allow image retrieval on the basis of object concepts, an object concept being an incremental and interactive formalization of the user’s conception of the object in the query image. In the remainder, we invariably use the term “object” to refer to an object that is depicted in an image.

Fig. 1 shows a schematic overview of our proposed incremental interactive image retrieval method. Each iteration step consists of the following actions: 1) obtaining an object description; 2) building an object concept; 3) reorganizing the database objects with respect to this object concept; 4) construction of a density function for similarity matching; 5) retrieving those database objects that have smallest distance to the object concept; 6) competitive learning using the database object the user elicits feedback on; and 7) customizing the object concept. Each

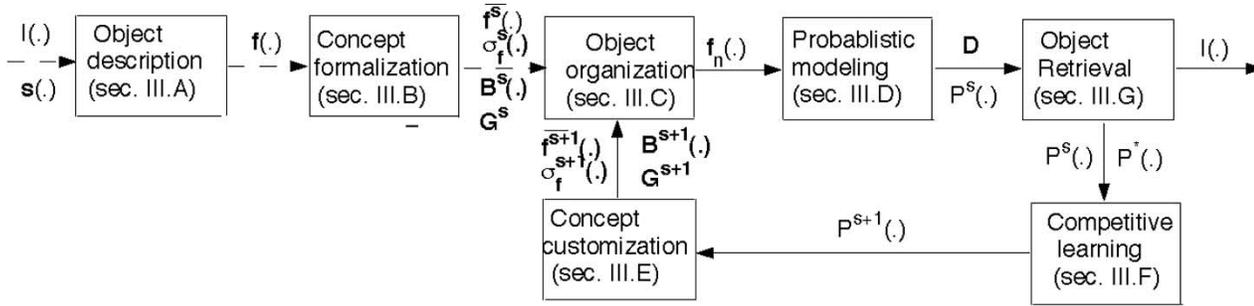


Fig. 1. A schematic overview of our concept-based image retrieval approach. Invariably, in this paper bold face upper case indicates a matrix of functions or of scalars, bold face lower case indicates a vector of functions or scalars, and regular lower case indicates a function or a scalar.

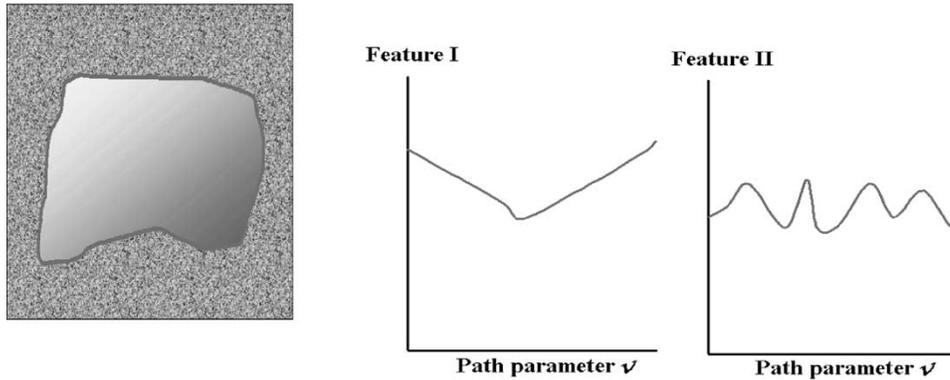


Fig. 2. Schematic illustration of multifeature object description. Left: an object in the image  $I(\mathbf{x})$  and its boundary delineation  $\mathbf{s}(v)$ . Right: gradient magnitude and contour curvature values along the boundary. The function  $\mathbf{f}$  operates on  $I(\mathbf{s}(v))$  to compute local shape, gray-value properties or position, yielding a multiple valued set for each  $v$ . The vertical axis denotes the feature value, the horizontal axis stands for the path parameterization of the boundary curve. Each of the two features, here depicted separately for purpose of clarity, is one dimension of the 2-D functional space, i.e.,  $\mathbf{f}^T(v) = [f_1(v), f_2(v)]$ .

of this components is discussed in more detail in the following sections.

### A. Multifeature Object Description

A natural way of retrieving images is by object features such as edge and curvature properties. For object-based image retrieval, however, it is necessary to use descriptive and discriminative features, which are often not readily available. To facilitate identification of such features we represent an object boundary by a string, a one-dimensional multivariate curve in functional space (see Fig. 2 and [19] for more detail).

Given the image  $I(\mathbf{x})$  with corresponding continuous object outline  $\mathbf{s}(v)$ , e.g., a B-spline curve, a multifeature object description is obtained by relating the curve  $\mathbf{s}(v)$  to the image at points  $I(\mathbf{s}(v))$ . The relation is expressed in terms of  $K$  features derived from the shape as well as from the image (see the Appendix for more detail). The mapping  $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^K$  yields feature function  $\mathbf{f}(v)$  in the  $K$ -dimensional functional space, where each dimension corresponds to one feature, i.e.,

$$\mathbf{f}^T(v) = [f_1(v), \dots, f_K(v)]. \quad (1)$$

The functional  $\mathbf{f}(v)$  captures feature values along the boundary  $\mathbf{s}(v)$ . There is no restriction with regard to the number and the type of features for boundary description. Generally, however, the more application-dependent the features are, the better the descriptive and the discriminative ability of those features.

### B. Population-Based Concept Formalization

We consider an object to be an instance of a concept. An object concept will later be used for image retrieval under the assumption that it is more suitable for capturing the user's retrieval intention because it is a generalization of a single object instance according to a user's subjectivity and preference. In this section we formalize an object concept in terms of the stochastic characteristics of feature values derived from a population of object boundaries. We follow the functional data analysis steps described in [19] and define here only quantities that are relevant for the remainder of this paper.

Given a population of  $M$  objects we build a stochastic description on the basis of the multiple continuous features recorded along the boundaries of these objects, i.e., based on feature functions  $\mathbf{F}^T(v) = [\mathbf{f}_1(v), \dots, \mathbf{f}_M(v)]$ . The feature functions are first aligned to remove variation that is attributed to stretching, shearing and rotation of objects (see the Appendix and [19] for more detail). Then, the average feature function  $\bar{\mathbf{f}}(v)$  and standard deviation  $\sigma_{\mathbf{f}}(v)$  are computed to obtain the elementary statistics of the object population (see Fig. 3). The average feature function is

$$\bar{\mathbf{f}}(v) = \frac{1}{M-1} \sum_{m=1}^M \mathbf{f}_m(v) \quad (2)$$

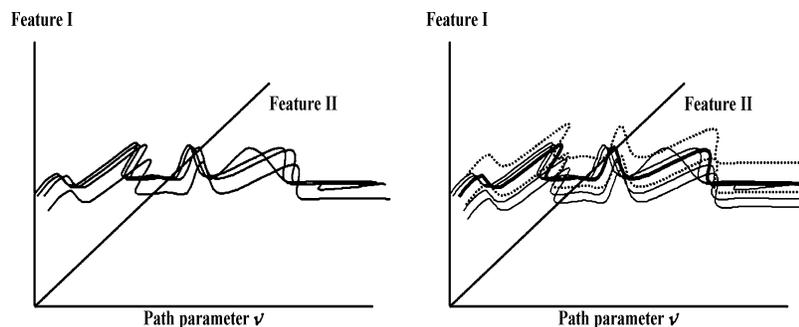


Fig. 3. Schematic illustration of functional data analysis of a collection of feature functions. Left: functional data sets, each representing an object by two features as a functional of  $v$ . Right: the elementary statistics of the feature function collection. Highlighted are the average feature function  $\bar{\mathbf{f}}(v)$  and the standard deviation  $\sigma_{\mathbf{f}}(v)$  (bold dotted line).

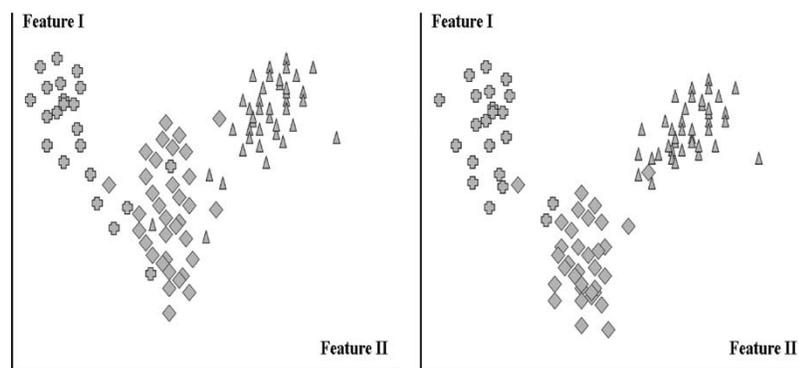


Fig. 4. Concept-dependent database organization. A projection of a database object onto the vector-valued search space reduces to estimating the principal components scores  $\mathbf{g}_n$  of the database objects using regression function  $\mathbf{B}^s(v)$ , which defines a concept at step  $s$ . Where an object is projected depends on how it deviates from a given concept. Left: projection with respect to  $\mathbf{B}^s(v)$ . Right: projection with respect to  $\mathbf{B}^{s+1}(v)$ . Note that objects belonging to the same category (different shapes) tend to occupy the same location of the search space, implying that a distinction between the categories can be made on the basis of a single object concept.

and the standard deviation

$$\sigma_{\mathbf{f}}(v) = \left( \frac{1}{M} \sum_{m=1}^M \|\mathbf{f}_m(v) - \bar{\mathbf{f}}(v)\|^2 \right)^{\frac{1}{2}}. \quad (3)$$

To obtain a more detailed description of the variation in the object population, the original feature functions are centered and scaled to unit variance, then subjected to functional principal components analysis (see the Appendix and [19] for a detailed description and [20] for background). The result is the  $M \times Q$  principal component scores matrix  $\mathbf{G}$ , composed of the  $Q$ -dimensional principal component scores  $\mathbf{g}_m = [g_{m1}, \dots, g_{mQ}]$ , for  $m = 1, \dots, M$ . The matrix  $\mathbf{G}$  summarizes the most important independent variation found in the object population. The scores are subsequently used to build an underlying functional principle components regression model of the feature functions. The matrix of orthonormal regression functions  $\mathbf{B}(v) = [\beta_1(v), \dots, \beta_Q(v)]^T$  is computed by least squares minimization such that

$$\mathbf{B}(v) = \arg \min_{\mathbf{B}^*(v)} \sum_{m=1}^M \int_v \|\mathbf{f}_m(v) - \mathbf{g}_m \mathbf{B}^*(v)\|^2 dv. \quad (4)$$

The matrix  $\mathbf{B}(v)$  indicates where each of the  $K$  measured features along an object boundary contributes to the principal component functions. Hence, the regression functions indicate which boundary features are locally most important to characterize the object population at hand.

Recapitulating, the following visual characteristics of an object population are used for concept formalization: average feature function  $\bar{\mathbf{f}}(v)$ , the normalized standard deviation  $\sigma_{\mathbf{f}}(v)$ , the matrix of orthonormal regression functions  $\mathbf{B}(v)$  and the principal component scores matrix  $\mathbf{G}$ . In Section III-G we will explain how we get an object population for concept formalization.

### C. Concept-Dependent Object Organization

Given an image database containing different classes of objects we organize the objects with respect to a concept. Concept-dependent object organization implies the grouping of objects that resemble a concept as well as those that have a common deviation from that concept. This way, the user can create a particular view of the database that permits him/her to easily retrieve objects in terms of closeness to the concept the user has in mind. This is schematically illustrated in Fig. 4.

To organize the database objects with respect to a given object concept we use of the matrix of regression functions  $\mathbf{B}(v)$ . For the  $n$ th database object with corresponding feature function  $\mathbf{f}_n(v)$ , the principal component score  $\mathbf{g}_n$  is computed by solving

$$\mathbf{f}_n(v) = \mathbf{g}_n \mathbf{B}(v)^T + \epsilon_n(v) \quad (5)$$

using least squares minimization of  $\epsilon_n(v)$  (see the Appendix for more detail). The principal component score vectors for the  $N$  database objects are all computed this way, resulting in a distribution of points in a multidimensional vector-valued concept space.

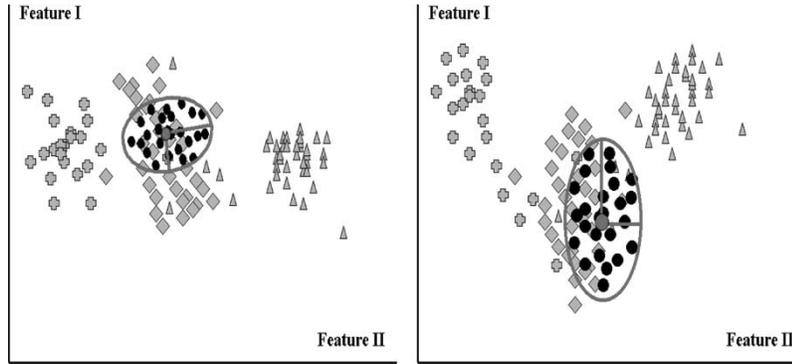


Fig. 5. Two different views on the database objects. The views are determined by the object concept, expressed by the Gaussian-shaped probabilistic model with its constituents (black dots). Note that the constituents may be virtual in the sense that objects with such a visual appearance may not exist. Left: a focused view on objects from a single category (diamonds). The database objects closest to the center of the probabilistic model best resemble the concept. This subset will be presented to the user for relevance feedback. Right: another more wide view on the database objects. The probabilistic model reduces investigation of the search space to finding a refined average and standard deviation. Note that a different concept automatically leads to a different database organization.

This concept space is our search space. The point distribution in this search space gives a particular view of the database objects. The user has the possibility to alter the view on the database objects by customizing the object concept at hand.

#### D. Probabilistic Modeling of Search Space

At this point we have defined an object instance in  $\mathbf{f}(v)$  and  $\mathbf{g}$ , an object concept in  $\mathbf{G}$  and object organization in  $\mathbf{B}$ . We now determine the probability that a database object is an instance of a given object concept. We compute the probability on the basis of a Mahalanobis distance model derived from the matrix  $\mathbf{G}$ . The Mahalanobis distance matrix is defined as

$$\mathbf{D} = \frac{\mathbf{G}^T \mathbf{G}}{(M+1)}. \quad (6)$$

The Mahalanobis distance matrix indicates the distance of an object  $\mathbf{f}_n(v)$  to the population average, taking into account the variation seen in the object population. This Mahalanobis distance matrix acts as quality measure. For  $\mathbf{f}_n(v)$  the Mahalanobis distance to the population average is computed using  $\mathbf{g}_n$  in the following manner:

$$D^2(\mathbf{f}_n(v), \bar{\mathbf{f}}(v) | \mathbf{g}_n) = \mathbf{g}_n \mathbf{D}^{-1} \mathbf{g}_n^T. \quad (7)$$

The Mahalanobis distance matrix is used for probabilistic modeling of the search space. The probabilistic model  $P(\cdot)$  gives the probability that  $\mathbf{f}_n(v)$  is an instance of the concept at hand. We define the probabilistic model as a Gaussian with average  $\bar{\mathbf{f}}(v)$  and standard deviation  $\sigma_{\mathbf{f}}(v)$  as we expect a Gaussian distribution after the feature functions in functional feature space have been mapped onto the vector-valued concept space. The probability of  $\mathbf{f}_n(v)$  is computed on the basis of its principal component scores as follows:

$$P(\mathbf{f}_n | \mathbf{g}_n) = |\mathbf{D}|^{-\frac{1}{2}} (2\pi)^{-\frac{Q}{2}} e^{-\frac{1}{2} \mathbf{g}_n \mathbf{D}^{-1} \mathbf{g}_n^T}. \quad (8)$$

With help of  $P(\cdot)$  we maintain a probabilistic model of the promising regions of the search space, as illustrated in Fig. 5. The aim is to refine  $P(\cdot)$  by updating its parameters  $\bar{\mathbf{f}}(v)$  and  $\sigma_{\mathbf{f}}(v)$  such that at the end of the refinement, objects that are an instance of the concept the user has in mind have high probability according to this probabilistic model.

Note that the adaptive nature of  $P(\cdot)$  allows the method to deal with both Gaussian, and to a certain extent non-Gaussian feature distributions since it will try to find a Gaussian distributions at any scale. For example, depending on the user's search behavior  $P(\cdot)$  will focus in on a small Gaussian distribution in a larger non-Gaussian distribution.

#### E. Example-Driven Concept Customization

Concept customization is required as the object search initially relies on the user's unformalized perception of a query object or on some other default concept. The aim of concept customization is to gradually transform the user's initial object perception into a concept formalization that is precise enough for robust and reliable image retrieval.

Concept customization is achieved in a number of steps by relevance feedback. When we formulate the object concept at step  $s$  by the stochastic description of  $\mathbf{F}^s(v)$ , i.e., by  $\bar{\mathbf{f}}^s(v)$ ,  $\sigma_{\mathbf{f}^s}(v)$ ,  $\mathbf{B}^s$  and  $\mathbf{G}^s$ , and corresponding probabilistic model  $P^s(\cdot)$ , one step in the concept customization process reduces to finding  $\bar{\mathbf{f}}^*(v)$  and  $\sigma_{\mathbf{f}^*}(v)$  that define the new probabilistic model  $P^*(\cdot)$ . The new average feature function  $\bar{\mathbf{f}}^*(v)$  is determined with help of the Euclidean distance measure

$$D_E(\mathbf{f}_m^s(v), \mathbf{f}_l^s(v) | \mathbf{g}_m^s, \mathbf{g}_l^s) = \|\mathbf{g}_m^s - \mathbf{g}_l^s\| \quad (9)$$

where  $\mathbf{f}_l^s(v)$  and  $\mathbf{g}_l^s$  represent the initial query object or, after the first step, the object the user elicits feedback on (we call it the pilot object in the remainder). In an elitist approach, that object from the population  $\mathbf{F}^s(v)$  is selected that has smallest Euclidean distance to  $\mathbf{f}_l^s(v)$ , or in case of negative feedback, the largest Euclidean distance. In case of the former, we have

$$\bar{\mathbf{f}}^*(v) = \arg \min_{1 \leq m^* \leq M} D_E(\mathbf{f}_{m^*}^s(v), \mathbf{f}_l^s(v)). \quad (10)$$

This object forms the center of the new concept and presumably better corresponds to the sought concept than does  $\bar{\mathbf{f}}^s(v)$ . The probabilistic model is further defined by the new standard deviation  $\sigma_{\mathbf{f}^*}^*(v)$ , which is computed as the squared difference between the previous average feature function  $\bar{\mathbf{f}}^s(v)$  and the feature function corresponding to the pilot object, i.e.,

$$\sigma_{\mathbf{f}^*}^*(v) = \left( \|\mathbf{f}_l^s(v) - \bar{\mathbf{f}}^s(v)\|^2 \right)^{\frac{1}{2}}. \quad (11)$$

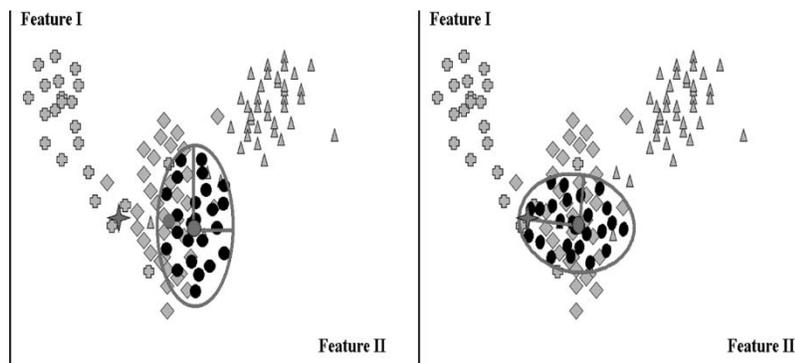


Fig. 6. Schematic illustration of example-driven concept customization. Left: the initial concept, i.e., probabilistic model, the projection of the pilot object (star) and the projection of the best matching object (gray dot) that is a constituent of the current object population underlying the concept. Right: the updated probabilistic model with the new average and standard deviation. Note that the probabilistic model has shifted and that the degree of standard deviation depends on the difference between average object and pilot object, allowing not only movement in search space but also zooming in and out.

The object population associated with the new probability model is obtained by sampling feature functions. The sampled feature functions are not always realistic because they do not necessarily relate to real objects. Rather they are virtual functions arising from exploring parts of the search space that correspond to not yet seen examples or to counter example objects. In the first case one has achieved the goal of the query by finding new supportive evidence for the object concept. In the latter case the query needs refinement, still to be considered a useful experience. The virtual feature functions are obtained by sampling  $M$  times from  $P^*(\cdot)$ , i.e.,  $\mathbf{f}_m^*(v) \sim P^*(\mathbf{f}^*, \sigma_{\mathbf{f}}^*(v))$ .

The customized object concept is represented by the stochastic characteristics of the newly sampled set of feature functions  $\mathbf{F}^*(v)$ , computed by functional data analysis in the same way as described in (2)–(4) and by probabilistic model  $P^*(\cdot)$ . The database objects that have highest probability of belonging to this concept are eligible for relevance feedback. This means, that the pilot object for the next refinement step is selected among the objects with feature functions that are projected onto the part of the search space spanned by  $P^*(\cdot)$ . This is illustrated in Fig. 6.

#### F. User-Controlled Search Strategy

In the previous section we described how a refined probabilistic model  $P^*(\cdot)$  is obtained with help of a pilot object. In this section we discuss how we can exploit more information from the prior probability model  $P^s(\cdot)$  to offer the user more control on the search strategy. The desire for a user-controlled search stems from the fact that user usually exhibits a variety of search behaviors, ranging from exact match, where the user precisely knows what he/she is looking for, to browsing, where the user has a less well-formed idea of interesting objects and is, therefore, willing to inspect a larger, more wide-ranging subset of objects [1].

We use a population-based incremental learning algorithm [21] that allows explicit control of the user's search by means of a single parameter. Using this algorithm, the probability model  $P^{s+1}(\cdot)$  for the next concept customization step is obtained using information gained from  $P^s(\cdot)$  and  $P^*(\cdot)$ . The probability update rule employed here is similar to weight update rule in competitive learning when an output is moved toward a

particular sample feature function [21]. The probability model  $P^{s+1}(\cdot)$  is defined by average

$$\bar{\mathbf{f}}^{s+1}(v) = (1 - \gamma)\bar{\mathbf{f}}^s(v) + \gamma\bar{\mathbf{f}}^*(v) \quad (12)$$

and deviation

$$\sigma_{\mathbf{f}}^{s+1}(v) = (1 - \gamma)\sigma_{\mathbf{f}}^s(v) + \gamma\sigma_{\mathbf{f}}^*(v). \quad (13)$$

The innovation parameter  $\gamma$  offers explicit control of how fast concept customization should be realized. As the probability model is used to generate the next population, the innovation parameter also effects which portion of the search space will be explored. The innovation parameter refines the probability model in the direction of the best member of the current population governed by  $\gamma$  (see Fig. 7). When  $\gamma = 0$ , there is no exploration of large portions of the search space. As  $\gamma$  increases, the amount of exploration increases and the ability to exploit the information gained from the previous search steps diminishes.

#### G. Concept-Based Image Retrieval

To commence concept-based image retrieval, an initial object concept is required. There are two options for bootstrapping image retrieval: a low-level and a high-level concept. The first is a default object concept that corresponds to a specific class of objects. In this case, the user needs to gradually change the object concept from default to user-specific in a relatively large number of steps, but with the possibility to inspect a wide range of database objects. The second option is the use of an object concept that is a generalization of a query object. In this case, virtual feature functions are sampled in the same way as described in (12), using the query object as the average and using an *a priori* defined standard deviation. This option converges quicker, but at the cost of missing potentially relevant objects that initially seem to deviate much from that concept. In absence of relevance feedback, this option is the equivalence of searching the best match in a single step.

Having defined an initial concept, the user interactively and incrementally searches the database until the user finally settles on a concept formalization that is robust and reliable enough for retrieval. If the user decides to stop after one relevance feedback step, for example, the final concept is formulated

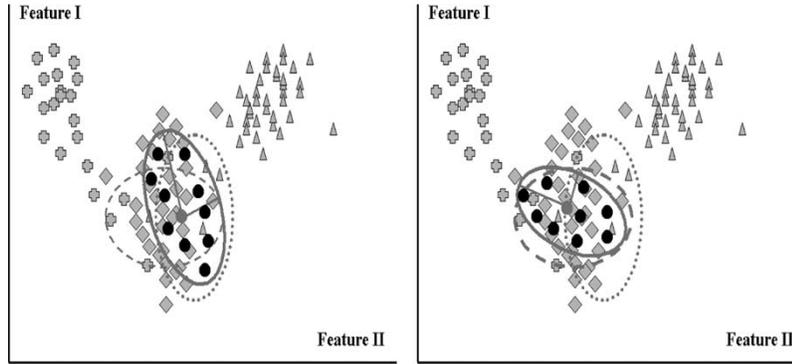


Fig. 7. The parameter  $\gamma$  determines the relative importance of past feedback cycles. Left: a high  $\gamma$  results in  $P^{s+1}(\cdot)$  (continuous ellipse) closer to  $P^*(\cdot)$  (dashed ellipse) than to  $P^s(\cdot)$  (dotted ellipse). Consequently, a wide range of objects will be explored. However, subtle differences may be missed. Right: a low  $\gamma$  results in  $P^{s+1}(\cdot)$  (continuous ellipse) closer to  $P^s(\cdot)$  (dotted ellipse) than to  $P^*(\cdot)$  (dashed ellipse). Consequently, no significantly different object will be explored. Rather subtle differences are highlighted. Hence,  $\gamma$  allows fine tuning.

by the stochastic characteristics of  $\mathbf{F}^{s+1}(v)$ , i.e., by  $\bar{\mathbf{f}}^{s+1}(v)$ ,  $\boldsymbol{\sigma}_{\mathbf{f}^{s+1}}(v)$ ,  $\mathbf{B}^{s+1}(v)$  and  $\mathbf{G}^{s+1}$  and corresponding probabilistic model  $P^{s+1}(\cdot)$ . In that case, the database objects are reorganized with respect to this final concept, which means that, for all  $n = 1, \dots, N$ , the principal component score  $\mathbf{g}_n^{s+1}$  are computed with help of  $\mathbf{B}^{s+1}(v)$  by solving

$$\mathbf{f}_n(v) = \mathbf{g}_n^{s+1} \mathbf{B}^{s+1}(v)^T + \boldsymbol{\epsilon}^{s+1}(v) \quad (14)$$

using least squares minimization of the residual  $\boldsymbol{\epsilon}^{s+1}(v)$ . Then, the Mahalanobis distance of each score  $\mathbf{g}_n^{s+1}$  to the average of the evolved population is computed on the basis of the updated distance model  $\mathbf{D}^{s+1}$

$$D_m^2(\mathbf{f}_n(v), \bar{\mathbf{f}}^{s+1}(v)) = \mathbf{g}_n^{s+1} \mathbf{D}^{s+1} \mathbf{g}_n^{s+1T}. \quad (15)$$

and the ones with minimal Mahalanobis distance are considered those that satisfy the final object concept. The database objects are ranked by their Mahalanobis distance and presented to the user as the final image retrieval result. The most interesting objects are placed first and the least ones at the last place. The best match is, thus

$$\mathbf{f}_n(v) = \arg \min_{1 \leq n^* \leq N} D_m^2(\mathbf{f}_{n^*}(v), \bar{\mathbf{f}}^{s+1}(v)). \quad (16)$$

The process of incrementally and interactively refining the object concept and the database organization, is repeated a number of times, until the user has narrowed the search subspace such that a sufficient number of relevant objects from the database are fetched. At the end of the search, image retrieval reduces to recovering objects from the database best matching the concept, ranked by the degree of content matching.

#### IV. EXPERIMENTS AND RESULTS

To demonstrate our concept-based image retrieval method we use an image database consisting of four different classes of cervical vertebrae: normal cervical vertebrae, vertebrae with lower anterior osteophyte, vertebrae with upper anterior osteophyte and vertebrae with both lower and upper osteophyte (see Fig. 8). An osteophyte is characterized by bony outgrowths on the anterior corners of the vertebral body. For example, the shape of the lower anterior of the C5 vertebra in the second image of



Fig. 8. Digital images of cervical vertebrae. From left to right: normal, lower osteophyte, upper osteophyte, lower and upper osteophyte. Note the limited detail and the complexity of the boundary.

Fig. 8 clearly extends from the body of the vertebra. In the image, the spurs are furthermore associated with the structural image boundary as opposed to histogram or texture characteristics [22].

The image material is acquired from the approximately 17 000 X-ray films collected during the Second National Health and Nutrition Examination Survey (NHANES II) conducted by the NCHS [23]. In this cross-sectional population survey, X-rays were taken of persons aged between 25 and 74. Two X-rays of the spine, PA and lateral, were made except of pregnant women and women under 50 years of age, to provide evidence of osteophyte and degenerative disc diseases. The films were subsequently digitized at a horizontal and vertical sampling rate of 146 dpi using Lumisys laser scanning equipment [23]. A medically certified diagnosis is attached to each image. We use a subset of the diagnosed NHANES II images that includes an expert delineation of the vertebral boundaries in the images, consisting of seven discrete points at a fixed position along the vertebral boundary. The delineations have been done by a single radiologist a single time, with expected intra-observer variability of approximately 5 pixels.

We simulate a situation in which a user browses a cervical image database of diagnosed cases to help determine the diagnosis of a new and unknown case suspected of having a lower anterior osteophyte. The user segments the vertebral structure from the image and uses the segmented vertebra as the initial query object. The system returns a number of potentially interesting cases and asks the user to indicate “interesting” or “not interesting” cases. The system uses the cases that the user elicits relevance feedback on as a pilot for concept learning. This way the system brings the user to a set of reference cases best

TABLE I

THE FOUR CATEGORIES OF CERVICAL VERTEBRAE, THE TOTAL NUMBER OF IMAGES IN EACH CATEGORY, AND THE COMPOSITION OF IMAGES FOR THE THREE EXPERIMENTS, INDICATING FOR EACH CATEGORY THE NUMBER OF IMAGES THAT HAVE BEEN USED FOR LEARNING AND FOR TESTING. ALL EXPERIMENTS START WITH AN IMAGE OF A VERTEBRA WITH LOWER OSTEOPHYTE AS THE INITIAL QUERY EXAMPLE. THIS QUERY IMAGE IS EXCLUDED FROM LEARNING AND TESTING SETS. NOTE THAT THE DATABASE IS CONSTANT IN SIZE AND COMPOSITION

Class	Total	Experiment I		Experiment II		Experiment III	
		Learning	Testing	Learning	Testing	Learning	Testing
Normal cervical	145	20	125	0	125	0	125
Lower osteophyte	78	0	58	20	58	0	58
Upper osteophyte	20	0	20	0	20	0	20
Lower and Upper	49	0	49	0	49	0	49

TABLE II

THE DIMENSIONS OF  $\mathbf{f}(v)$ : CONTOUR CURVATURE, ISOPHOTE CURVATURE AND THE DIRECTIONAL CORRESPONDENCE BETWEEN THE NORMAL  $\mathbf{n}(v)$  TO THE SHAPE  $\mathbf{s}$  AT  $v$  AND THE IMAGE GRADIENT AT  $\nabla I(\mathbf{x})$ . THE GRADIENT IMAGE IS OBTAINED BY CONVOLUTION OF THE IMAGE  $I(\mathbf{x})$  WITH A GAUSSIAN DERIVATIVE OPERATOR WITH SPATIAL SCALE  $\sigma = 4$ . NOTE THAT  $\mathbf{x} = \mathbf{s}(v)$

Dimension	Feature	Definition
1	contour curvature	$\frac{\mathbf{s}_x(v)\mathbf{s}_{yy}(v) - \mathbf{s}_{xx}(v)\mathbf{s}_y(v)}{(\mathbf{s}_x^2(v) + \mathbf{s}_y^2(v))^{3/2}}$
2	isophote curvature	$\frac{I_{xx}(\mathbf{x})I_y^2(\mathbf{x}) - 2I_x(\mathbf{x})I_y(\mathbf{x})I_{xy}(\mathbf{x}) + I_{yy}(\mathbf{x})I_x^2(\mathbf{x})}{(I_x^2(\mathbf{x}) + I_y^2(\mathbf{x}))^{3/2}}$
3	directional correspondence	$\nabla I(\mathbf{s}(v)) \cdot \mathbf{n}(v)$

matching the new and unknown case, hence, best suited for supporting the diagnosis. This is a highly challenging task considering the subtle shape and image differences among vertebrae and in this case the marginal image quality.

#### A. Experiments

The system is steered by and relies on the preferences and subjectivity of the user. It is, therefore, natural that retrieval outcomes differ from user to user, making an objective comparison of retrieval outcomes from different sessions by different users difficult. For this reason, we have chosen to perform experiments completely automatically. In a single retrieval session, an image of a vertebra with lower anterior osteophyte is automatically chosen from the database and its given manual delineation used to construct a query feature function. Images are retrieved on the basis of this query, and the highest ranked image of a vertebra with lower anterior osteophyte among the top 10 retrieved images is used to elicit feedback on. In absence of an image of a vertebra with lower anterior osteophyte among the top 10, the top ranked image is used for negative feedback.

1) *Setup*: The goal of the first experiment is to evaluate the performance of our method for different kinds of image seeking behavior. We consider a situation where the user starts *a*) from a low-level concept that corresponds with the normal cervical vertebra, *b*) from an intermediate-level concept corresponding with a vertebra with a typical lower anterior osteophyte and *c*) from a high-level concept that corresponds to the specific abnormality seen in the query image. These three situations reflect the varying types of search behavior that range from open-ended to direct search. In all three situations one vertebra with lower osteophyte is used to bootstrap the retrieval process. The composition of images for concept learning and retrieval for the three situations is listed in Table I.

We perform experiments with three boundary features. We confine ourselves to invariant features as they generalize appli-

cability, but more importantly, as they minimize the need for feature alignment. We use the  $K = 3$  invariant features listed in Table II. In this case, we have  $\mathbf{f}(v) = [f^1(v), f^2(v), f^3(v)]^T$  in a three-dimensional functional space.

Feature functions are obtained by recording features at 50 sample points along a continuous vertebra boundary delineation. The continuous boundary delineation is obtained from the 7 manually marked points by interpolation of a B-spline curve through these points. Feature functions are computed from these curves by sampling the above image and shape features. The feature functions themselves are also represented by B-spline curves and regularized prior to functional data analysis. We note that, due to curve interpolation through the sparse point set, the resulting curves may sometimes be locally inaccurate, resulting in noisy feature functions.

2) *Measurements*: To assess how relevant the retrieved images are to the initial query, we measure precision and recall [24], defined as

$$\text{precision} = \frac{\text{No. relevant images retrieved}}{\text{Total No. images retrieved}}$$

$$\text{recall} = \frac{\text{No. relevant images retrieved}}{\text{Total No. relevant images in collection}}$$

where high precision indicates that from all the images returned to a query, a large proportion of the images are relevant to the search (purity of retrieval). A high recall indicates that from all the images in the repository that are relevant to the query, a large number of these images are indeed returned (completeness of retrieval). In our case, the number of relevant images is the number of images that are of the same class as the initial query image, i.e., images by expert consensus classified as vertebra with lower anterior osteophyte.

We also address the question of how well the retrieval method searches the space of solutions. We express the trade-off be-

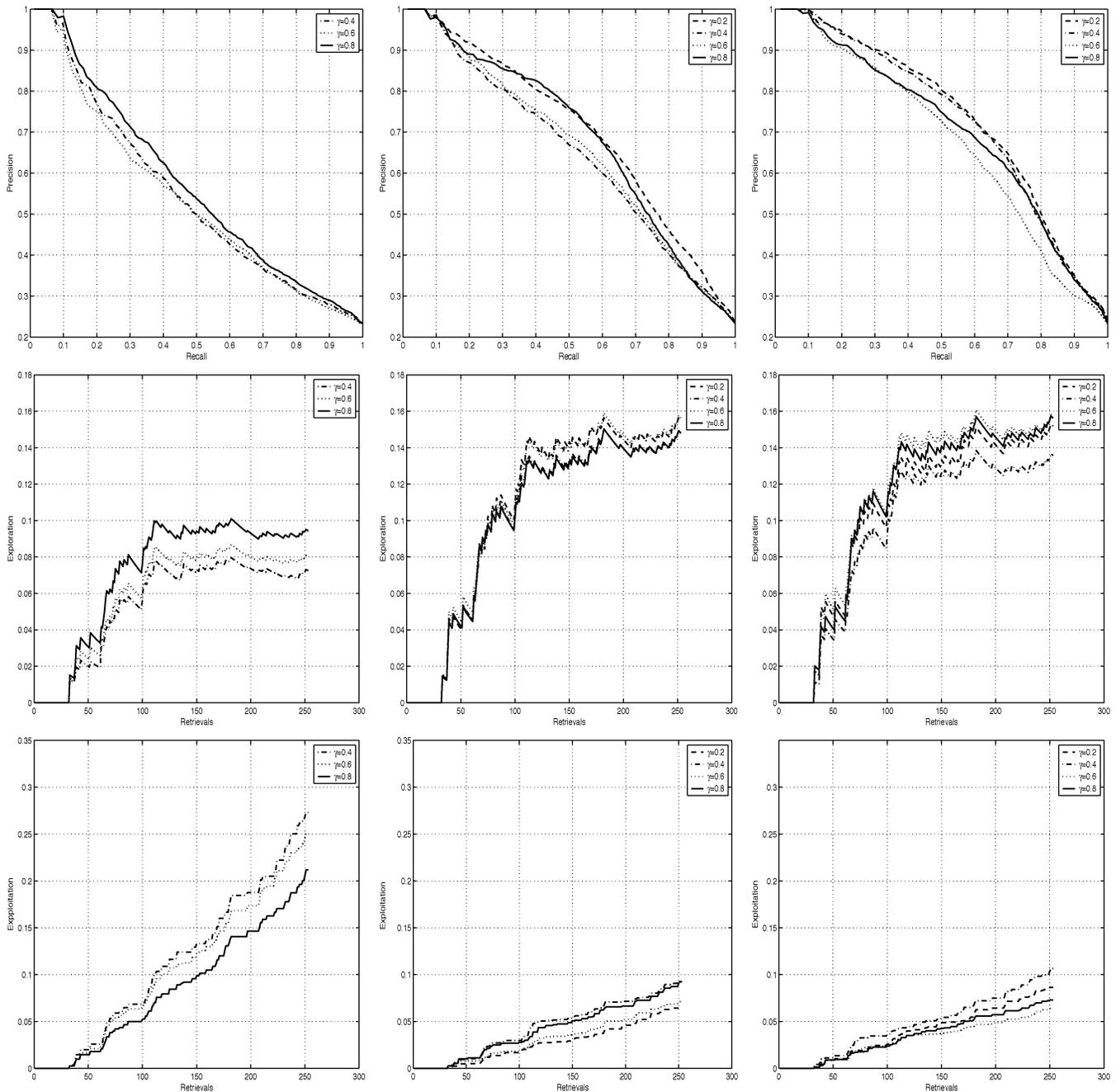


Fig. 9. Precision versus recall, exploration and exploiting results for the three different starting points of the initial object concept. From left to right are shown results when starting from: a low-level, intermediate-level and high-level concept. Each figures show the average of a total of 58 retrieval sessions using  $\gamma$  values: 0.2, 0.4, 0.6, 0.8. It can be shown that when using a high-level object concept higher precision versus recall values are reached. Also more effective exploration is done. More effective exploitation is done when starting from a low-level concept.

tween exploration of the search space and exploitation of previous results by

$$\text{exploration} = \frac{\text{No. previously unretrieved relevant images}}{\text{Total No. of retrieved images}}$$

$$\text{exploitation} = \frac{\text{No. previously retrieved relevant images}}{\text{Total No. relevant images in collection}}$$

where previously unretrieved images are images that where no part of the top 10 best results in the history of the browsing session. In this context, exploration is the ability of the retrieval

method to investigate the search space thoroughly, while exploitation refers to the method's ability to use the information about the search space it has gained to narrow its future search.

## B. Results

Fig. 9 shows precision versus recall, exploration and exploiting results for the three different starting points of the initial object concept. From left to right are shown results when starting from: a low-level, intermediate-level and high-level concept. Each figures show the average of a total of 58 retrieval

sessions using  $\gamma$  values: 0.2, 0.4, 0.6, 0.8. Precision and recall, exploration and exploitation rates are computed at the end of each retrieval session. A retrieval session ends when the stopping criteria  $\lambda = 5$  is met, which indicates the minimal number of relevant images that need to be in the top of the ranked retrieved images. Consequently, the precision value up to 5 retrievals is 100%. This is also reflected in the first part of the precision versus recall graph. Note also that the first part of the exploration and exploitation graphs indicates 0 values. This is because the retrievals along the  $x$  axis, are not ordered by the degree of concept matching but by database index position; the removal of relevant images from the first positions in the database for inductive concept learning leads to 0 exploration and exploitation values.

The first precision versus recall graph in the first row of Fig. 9 shows how precision decreases as increasingly large fractions of the images are retrieved. It shows, for example, that to retrieve 50% of the relevant images, about 45% of the retrieved images will not be relevant. The graphs show that for  $\gamma = 0.8$  the best performance is obtained, whereas for  $\gamma = 0.2$  the stopping criteria was never met within the maximum number of 50 relevance feedback step. This is natural as with  $\gamma = 0.2$  only a small area of the search space is investigated and, thus, a large number of steps is required to get at the relevant part of the search space. The values  $\gamma = 0.4$ ,  $\gamma = 0.6$ ,  $\gamma = 0.8$  allow investigation of a larger area of the search space and, thus, guarantee convergence within 50 relevance feedback step. It can be seen that higher precision versus recall values are obtained when departing from an intermediate-level concept. For example, for  $\gamma = 0.8$  to retrieve 50% of the relevant images, now only about 25% of the retrieved images will not be relevant. The high-level concept produces the best precision versus recall values because it immediately brings the user to a position in the search space where vertebra with similar abnormalities as the query vertebra are found. This can also be deduced from the fact that the best precision versus recall values are reached for  $\gamma = 0.2$ . From the above, we conclude that our method integrally handles a variety of search strategies and performs well in terms of precision and accuracy of retrieval, considering the complexity of the visual appearance of the vertebral structures in our database.

The second row of Fig. 9 shows the percentage of the retrieved relevant images that are not ranked in the top 10 during the entire retrieval session as a function of the number of retrieved images. In other words, at the end of the retrieval session we look at the top 10 images and count the number of relevant images that never have been ranked in the top 10 during the same retrieval session. In the left graph, it can be seen that if 150 images are retrieved, about 10% of the retrieved relevant images have never been ranked in the top 10. The graph also shows that the innovation parameter has a considerable effect on the exploration rate. As expected the exploration rate is higher for  $\gamma = 0.8$ . When the intermediate-level concept is used, the innovation parameter has a minor effect on the exploration rate, as shown in the middle figure. In fact the graph shows that the highest exploration rate often occurs for  $\gamma = 0.2$ . This is because the intermediate-level concept instantly brings the user to the relevant search space and, hence, a low innovation parameter suffices to successfully find new but previously unretrieved

relevant images. This also explains why the exploration rates are generally higher when starting from the intermediate-level concept. The same also holds for the high-level concept (see right figure) with the difference that higher values for the innovation parameter now produce better exploration rates because the high-level concept may deviate considerably from the typical class characteristics and, thus, may require more adaptation to successfully investigate the search space. In any case, the effect of the innovation parameter diminishes when performing a direct search. We can state that our method truly assists in quick and effective searching when doing an open-ended as well as a direct search.

The third row of Fig. 9 shows the percentage of the retrieved images that have been in the top 10 at least once during the retrieval session as a function of the number of retrieved images. That is, at the end of retrieval session we look at the 10 best images and count the number of relevant images that already have been ranked in the top 10 at least once during the same retrieval session. The graph shows for  $\gamma = 0.8$  that if 150 images are retrieved, only 10% of the retrieved relevant images have already been ranked in the top 10. As expected, the lower the innovation parameter the higher the exploitation rate. The figure in the middle shows that the exploitation rate is generally lower when starting from an intermediate-level concept. This is because the intermediate-level concept brings the user immediately in the relevant search space so that the user has a larger variety of relevant images to choose from, whereas the low-level concept, once it reaches a part of the search space that is relevant, it is more likely to exhaust that part until the convergence criteria is met. This also holds for the high-level concept, which produces exploitation rates similar to that of the intermediate-level concept. We conclude that our method facilitates learning from the user's search history to focus in on promising parts of the search space.

We also investigated the convergence of retrieval for various values of the stopping criteria  $\lambda$  and innovation parameter  $\gamma$  when starting from a high-level concept. As mentioned earlier, the stopping criteria  $\lambda$  simply indicates the number of relevant images in the first 10 of the retrieved images. The value of  $\lambda$  is chosen such that it reflects the user search intention as well as the certainty the user seeks about the relatedness of the retrieved images. In general, when browsing a low-value suffices, when doing a direct search higher values are required. Fig. 10 shows that for  $\lambda = 1$  on average a little over 1 iteration step is required to convergence. For  $\lambda = 3$  this is approximately 2 and for  $\lambda = 5$  approximately 3.5. Only in open-ended searches or only in the extreme case where one aims at a direct search, starting from an unrelated object and using parameters tuned for exploitation, one may require number of steps that exceed 20. A single iteration approximately takes 1 s. Hence, we conclude that our method converges within an acceptable number of concept customization steps and time span, particularly considering the low number of features used and high values of the stopping criteria.

### C. Examples

We visually illustrate population-based incremental interactive concept learning for image retrieval. We concentrate on

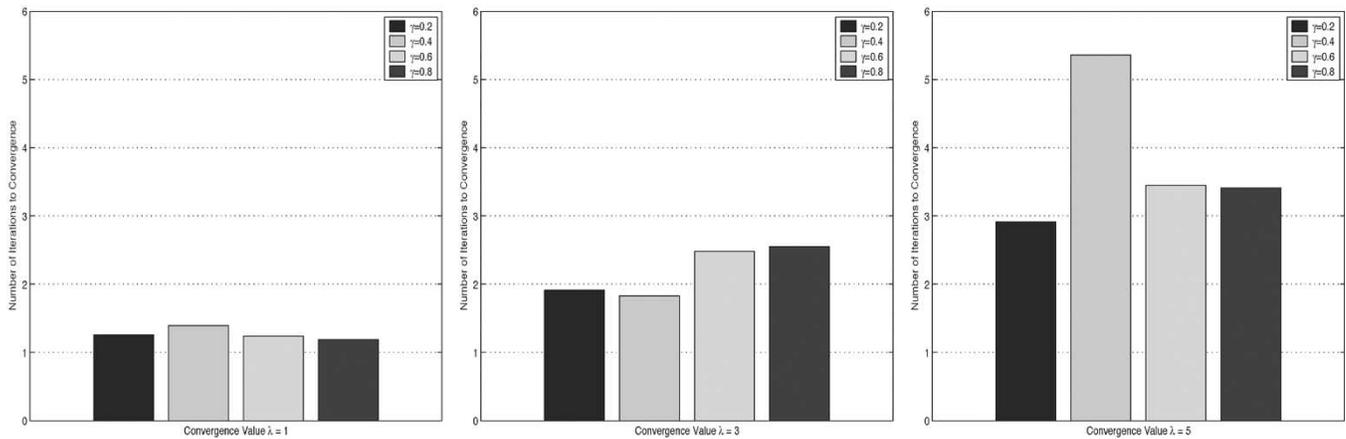


Fig. 10. Convergence performance for various values of the stopping criteria  $\lambda$  and innovation parameter  $\gamma$  when starting from a high-level concept. Left:  $\lambda = 1$ ; middle:  $\lambda = 3$ ; and right:  $\lambda = 5$ . The graph shows that on average convergence is almost always retrieved within three steps, even for a high stopping criteria.

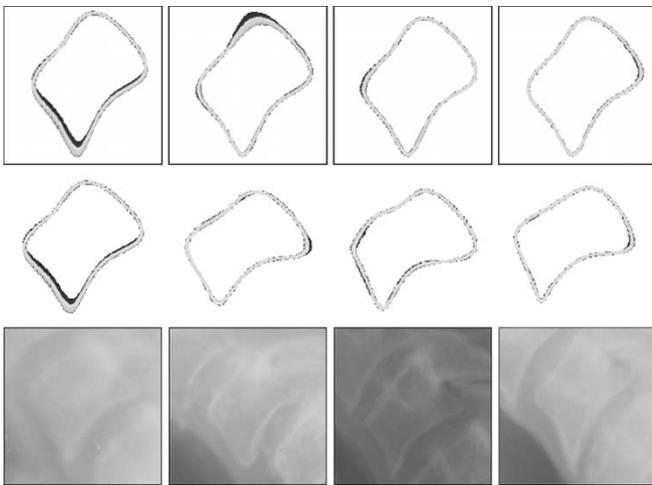


Fig. 11. The top row is a visualization of shape and image characteristics of a population of normal cervical vertebrae. Shown are the average shape plus (light) and minus (dark) up to three standard deviation in the (a) first, (b) second (c) third, and (d) fourth principal components direction. Middle row shows the mean shape of a cervical vertebra plus (light) and minus (dark) up to three standard deviations away in the direction of the first principal component. The four pictures represent the condition at the  $s = 4$  steps that were required to fetch the query image. Note the difference with the first row in that the four figures correspond with the four browsing steps, not the variations in the four principal component directions. Bottom row: illustration of a query image with retrieval results ranked by the degree of content matching. From left to right: initial query image is one of vertebra with lower anterior osteophyte, best matching image in repository after browsing, second best, third best. Note that all the retrieved images are of vertebrae with lower anterior osteophyte.

$K = 2$  shape features, capturing the projectional alignment of vertebrae in addition to local shape. These features are straightforward to visualize and can be related to the vertebral shape at a glance. The first dimension  $f^1(v)$  is the distance between sampled  $x$  coordinate values along the vertebra contour and the  $x$  coordinate value of a single reference point on it. The second dimension  $f^2(v)$  is the difference between sampled  $y$  coordinate values and the  $y$  coordinate value of the same reference point. Hence, we have feature functions  $\mathbf{f}(v) = [f^1(v), f^2(v)]^T$  in a two-dimensional (2-D) search space. We start from a concept of the “normal” cervical vertebra and gradually customize this concept to one of a vertebrae with lower anterior osteophyte.

The initial concept is defined in terms of the stochastic characteristics of a population of normal cervical vertebrae. The images in the top row of Fig. 11 illustrate the average shape of a normal cervical vertebra plus (light gray values) and minus (dark gray values) up to three standard deviations away from the average shape. The four columns correspond to the shape variation in the first four principal component directions, together capturing 93% of the total variability in shape. The number of principal components has been set to  $Q = 4$  because we expect for our vertebra application there are 4 corners and, hence, 4 places, where the data in the learning phase might exhibit independent shape variation. It can be seen in the first figure that the main variation in the shape of the normal cervical vertebrae occurs at the lower anterior corner. The variation extends to the entire anterior and lower part. The second figure shows that the second locus of variation is at the right upper corner. The third mode of variation concentrates at the upper anterior corner as illustrated in the third figure, exhibiting minor shape variation. The right figure indicates little variation at the right lower corner in the fourth principal component direction. As expected in this application, shape variations concentrate at the four corners of the normal cervical vertebrae.

We demonstrate how the normal vertebra shape evolves during concept customization to one of lower anterior osteophyte. We visualize the stochastic characteristics of the vertebra shapes emanating from the virtual feature functions that are sampled during concept customization.

The shapes in the second row of Fig. 11 indicate the evolution trajectory from normal cervical vertebra to one which possesses a lower anterior osteophyte, requiring only four concept customization steps. From the figures it can be seen which shapes are gradually explored to arrive at lower anterior osteophyte, with the modes of variation showing the parts of the vertebra contour that are adapted during concept customization. It can be seen that there is almost no variation at the end of the retrieval session as, in this case, the same abnormal vertebra has been used for exploration. When investigating the database on the basis of a population of abnormal, the average indicates the common characteristics of the browsed abnormal, whereas the modes of variation indicate parts of the abnormal vertebrae where the characteristics are equivocal. It might be postulated

that the evolution from normal to abnormal vertebra perhaps corresponds to some pathological process and may permit the generation of a model of the disease process.

Images of the abnormal vertebrae ranked highest after each of the 4 required concept customization steps are illustrated in the bottom row of Fig. 11. The initial image is the query image of a vertebra with lower anterior osteophyte. The best three matching images after browsing the retrieved images are shown. All three belong to the class of lower anterior osteophyte.

## V. DISCUSSION AND CONCLUSION

Our concept-based image retrieval method contributes to the evolution of content-based image retrieval in the *content understanding-query completion-user interaction space* [1] in a number of ways. At data entry the method guarantees content understanding at the semantical level thanks to expert delineation of the object of interest in the image to be stored. During image retrieval content-understanding is elevated to the conceptual level by consideration of the user's subjectivity and specific preferences. Content-understanding at the conceptual level is achieved by means of relevance feedback, which is a powerful mechanism for a high degree of query completion. As relevance feedback for query completion is inherently interactive, our method requires an active and integral role of the user. However, as noted in [1], if an image database can provide content understanding at an organ level and can guarantee a high level of query completion, the user may be willing to invest moderate effort in the entry interaction process at the time each image is added to the collection or at retrieval time.

The main difference with other, more general, concept-based image retrieval methods such as [3] is that our method offers the user the capability to incrementally and interactively zoom in on or out from subtle visual object patterns, obtained by mapping objects from a functional feature space to concept space based on their deviations from a given concept. This capability is compelling especially where: 1) the user knows the object features he/she is interested in, while the system is not immediately capable of recognizing the subtle difference in visual features between object or object classes; 2) the user is not capable of expressing the visual features that are relevant to him/her, while the system has no difficulty in discriminating between the objects or object classes on the basis of multiple features and instances. The difference with other methods manifests itself, among others, in the ability to deal with multiple types of search behavior. Methods such as those proposed in [6] and [5] allow a single search strategy, e.g., exact match, proximity match, loose match. Our method, in contrast to the major dichotomy usually found in discussions about retrieval strategies, that of passive versus an active user role and that of directed versus undirected search [25], offers a continuum of control with respect to the search strategy and, thus, can be considered a superset of common methods.

We draw the following conclusions from our experiments and results. When a high-level concept is used without concept customization, our method reduces to a "black box" where a single query/response pair is supposed to satisfy the needs of the user. In this case, the method will work well only if the visual char-

acteristics of and dissimilarities between objects are clear and unambiguous. A high-level concept in combination with concept customization particularly works well if visual characteristics are complex and differences between objects diffuse. Then, some optimal, usually low, value for the innovation parameter will bring the user to the desired search subspace in a minimal number of steps. A low-level concept requires concept customization for a direct search with a specified or an unspecified end, usually with high values for the innovation parameter. The user needs to consistently indicate what he/she finds interesting or uninteresting during the entire image retrieval process. Without consistency in relevance feedback the method will reduce to one for undirected search with an unspecified end. Note that in our experiments we use a single value for the innovation parameter within one retrieval session. We expect that adapting the innovation parameter during a retrieval session will lead to performance improvement. In any case, our method works well for supporting the user's search behavior and for meeting the user's search intention.

A number of issues remain to be investigated. We expect Gaussian distributions in our search space and, therefore, use an adaptive Gaussian probabilistic model. We have not explored the benefits of using a parameter-free density functions as proposed by Vasconcelos *et al.* [6]. Also, it is expected that negative feedback improves the discrimination ability of the retrieval method in addition to dealing with local minima of the search space. The consequence of negative feedback to the exploration and exploitation of the search space needs further study. Furthermore, in this application we have preselected a number of features for the definition of the object boundaries. When the different object classes are known, as in our clinical example application, features can be studied separately for each class in order to build a different type of concept for each different object class. This is expected to improve the discrimination ability of our method. Finally, we note that, although in this paper we applied the method for retrieval of vertebrae images, the combination of multifeature object description with concepts from population-based incremental learning techniques [21] naturally deals with challenges and opportunities provided by other (medical) images as well.

## APPENDIX

- *Feature function definition for object description.* We compute a continuous approximation of an object boundary by interpolation of a curve  $\mathbf{s}(v)$  through discrete boundary points.  $\mathbf{s}(v)$  and its corresponding  $K$ -dimensional feature function  $\mathbf{f}^{**}(v)$  are tensor product B-Splines

$$\mathbf{s}(v; \mathbf{p}) = \sum_{i=1}^I B_i(v) \mathbf{p}_i \quad (17)$$

$$\mathbf{f}^{**}(v; \mathbf{q}) = \sum_{j=1}^J B_j(v) \mathbf{q}_j \quad (18)$$

where  $\mathbf{p}_i$  denotes discrete control points of  $\mathbf{s}(v)$  and  $B_i(v)$  the corresponding basis functions defined on an uniform knot vector. Similarly,  $\mathbf{f}^{**}(v; \mathbf{q}_j)$  is defined by

basis functions  $B_j(v)$ , corresponding to control points  $\mathbf{q}_j$ , i.e., the sampled feature values. We impose regularity on  $\mathbf{f}^{**}(v; \mathbf{q}_j)$  by using basis expansions with a relatively small number of basis functions [20].

- *Feature function alignment for concept formalization.* Alignment of  $\mathbf{f}_1^{**}(v), \dots, \mathbf{f}_M^{**}(v)$  is done by the iterative Procrustes method [26] using a global alignment criteria that computes the least squares distance to the estimated overall average  $\hat{\mathbf{f}}(v)$ . This reduces to finding the nonlinear strictly monotonic warping function  $\omega_m(v)$  such that

$$\omega_m(v) = \arg \min_{\omega_m^*(v)} \sum_{m=1}^M \int_v \left\| \mathbf{f}_m^{**}(\omega_m^*(v)) - \hat{\mathbf{f}}(v) \right\|^2 dv. \quad (19)$$

The function  $\omega_m(v)$ ,  $v = 0, \dots, \mathcal{V}$ , is differentiable up to a certain order and has properties  $\omega_m(0) = 0$  and  $\omega_m(\mathcal{V}) = \mathcal{V}$ . It takes care of a shift and a nonlinear transformation by the roughness penalty approach. We penalize by the size of the third derivative of  $\omega_m(v)$ . Corresponding starting points for and direction of alignment need to be known.

- *Feature function normalization for concept formalization.* The average feature function  $\bar{\mathbf{f}}(v)$  is subtracted from each feature function to normalize the range of feature values, reducing the influence of variational differences due to measurements in different units. This yields

$$\mathbf{f}_m(v) = \frac{\mathbf{f}_m^*(v) - \bar{\mathbf{f}}(v)}{\sigma_{\mathbf{f}}(v)} \quad (20)$$

with units of variance due to normalization by the variance vector of functions

$$\sigma_{\mathbf{f}}(v) = \left( \frac{1}{M} \sum_{m=1}^M \left\| \mathbf{f}_m^*(v) - \bar{\mathbf{f}}(v) \right\|^2 \right)^{\frac{1}{2}}. \quad (21)$$

- *Feature space reduction for concept formalization.* The central concept of functional principal component analysis is that of taking the linear combination

$$g_{mq} = \sum_{k=1}^K \int_v f_{mk}(v) \alpha_{qk}(v) dv \quad (22)$$

To obtain the value of the principal component score  $g_{mq}$  for all  $q = 1, \dots, Q$  the corresponding vectors of weighting functions  $\alpha_q(v) = [\alpha_{q1}(v), \dots, \alpha_{qK}(v)]$  are sought for one-by-one

$$\begin{aligned} \alpha_q(v) &= \arg \min_{\alpha_q^*(v)} \frac{1}{M} \sum_{m=1}^M g_{mq}^2 \\ &= \arg \min_{\alpha_q^*(v)} \frac{1}{M} \sum_{m=1}^M \left( \sum_{k=1}^K \int_v f_{mk}(v) \alpha_{qk}(v) dv \right)^2 \end{aligned} \quad (23)$$

where  $\alpha_l(v)$ , for each iteration  $l$ , is subject to the following orthonormal constraints:

$$\sum_{k=1}^K \int_v \alpha_{qk}(v)^2 dv = 1 \quad (24)$$

$$\sum_{k=1}^K \int_v \alpha_{lk}(v) \alpha_{qk}(v) dv = 0, \quad l \leq q. \quad (25)$$

- *Principal components regression for concept formalization* We define  $\mathbf{F}^T(v) = [\mathbf{f}_1(v), \dots, \mathbf{f}_M(v)]^T$  and the matrix  $\mathbf{G}$  by

$$\mathbf{G} = \mathbf{F}(v) \mathbf{A}(v) \quad (26)$$

with scalar elements according to the dot product defined in (22). To find the values of  $\mathbf{B}(v) = [\beta_1(v), \dots, \beta_Q(v)]^T$ , with elements of the same  $K$ -dimensional functional form as the elements of  $\mathbf{F}(v)$ ,  $\mathbf{F}(v)$  is expressed as

$$\mathbf{F}(v) = \mathbf{G} \mathbf{B}(v) + \mathbf{E}(v) \quad (27)$$

with  $\mathbf{E}(v) = [\epsilon_1(v), \dots, \epsilon_M(v)]^T$  being the matrix of residual functions. Since there are no particular restrictions on the way in which the matrix of functions  $\mathbf{B}(v)$  varies as a function of  $v$ , the solution can be obtained by minimizing the least squares difference for each  $v$  separately. After least squares minimization we have

$$\mathbf{E}(v) = \left\| \mathbf{F}(v) - \mathbf{G} \mathbf{B}(v) \right\|^2. \quad (28)$$

- *Score estimation for object organization.* Once  $\mathbf{B}(v)$  is known the organization of database objects reduces to estimate the scores corresponding to their feature functions  $\mathbf{f}_1(v), \dots, \mathbf{f}_N(v)$  using the principal component regression model and least squares minimization, such that

$$\mathbf{g}_n = \arg \min_{\mathbf{g}_n^*} \int_v \left\| \mathbf{g}_n^* \mathbf{B}(v)^T - \mathbf{f}_n(v) \right\|^2 dv. \quad (29)$$

#### ACKNOWLEDGMENT

The authors are thankful to the Lister Hill National Center for Biomedical Communications, NLM/NIH for providing NHANES II images from the National Center for Health Statistics.

#### REFERENCES

- [1] H. D. Tagare, C. C. Jaffe, and J. Duncan, "Medical image databases: a content-based retrieval approach," *Amer. Med. Informatics Assoc.*, vol. 4, pp. 184–198, 1997.
- [2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. C. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1349–1380, Dec 2000.
- [3] B. Bhanu and A. Dong, "Concepts learning with fuzzy clustering and relevance feedback," *Eng. Applicat. Artif. Intell.*, vol. 15, pp. 123–138, 2002.
- [4] A. L. Ratan, O. Maron, W. E. L. Grimson, and T. Lozano-Prez, "A framework for learning query concepts in image classification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1999, pp. 1423–1443.
- [5] C. Meilhac and C. Nastar, "Relevance feedback and category search in image databases," in *IEEE Int. Conf. Multimedia Computing and Systems*, vol. 1, 1999, pp. 512–517.
- [6] N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval," in *Adv. Neural Inform. Processing Syst.*, S. A. Sara A. Solla, T. K. Todd K. Leen, and K.-R. Klaus-Robert Müller, Eds., 1999, pp. 977–986.
- [7] X. S. Zhou and T. S. Huang, "Relevance feedback for image retrieval: a comprehensive review," *Multimedia Syst.*, vol. 8, no. 6, pp. 536–544, 2003.
- [8] T.-S. Chua, H.-K. Pung, G.-J. Lu, and H.-S. Jong, "A concept-based image retrieval system," in *Proc. 20th Annu. Hawaii Int. Conf. System Sciences*, 1994, pp. 590–598.
- [9] G. R. Thoma, L. R. Long, and S. Antani, "Content Based Image Retrieval of Biomedical Images, Report to the NLM/LHC Board of Scientific Counselors," Washington, DC, 2002.

- [10] E. G. M. Petrakis, "Content-based retrieval of medical images," *Int. J. Comput.*, vol. 11, no. 2, pp. 171–182, 2002.
- [11] Y. Liu and F. Dellaert, "Classification-driven medical image retrieval," in *Proc. IUW 98*, 1998, pp. 1261–1267.
- [12] A. Mojsilovic and J. Gomes, "Semantic based categorization, browsing and retrieval in medical image databases," in *Proc. Int. Conf. Image Processing*, 2002, pp. 145–148.
- [13] D. Keysers, J. Dahmen, H. Ney, B. Wein, and T. M. Lehmann, "A statistical framework for model-based image retrieval in medical applications," *J. Electron. Imag.*, vol. 12, no. 1, pp. 59–68, 2003.
- [14] H. J. Lowe, I. Antipov, W. R. Hersh, and C. A. Smith, "Toward knowledge-based retrieval of medical images: the role of semantic indexing, image content representation, and knowledge-based retrieval," in *Proc. 1998 AMIA Annu. Fall Symp.*, 1998, pp. 882–886.
- [15] W. W. Chu, C.-C. Hsu, A. F. Cardenas, and R. K. Taira, "Knowledge-based image retrieval with spatial and temporal constructs," *IEEE Trans. Knowledge Data Eng.*, vol. 10, no. 6, pp. 872–888, 1998.
- [16] R. Chbeir, A. Youssef, and A. Flory, "A four-dimensional approach to medical image retrieval," *Meth. Inform. Med.*, vol. 40, no. 3, pp. 178–184, 2001.
- [17] D. Comaniciu, P. Meer, and D. J. Foran, "Image-guided decision support system for pathology," *Machine Vis. Applicat.*, vol. 11, no. 4, pp. 213–224, 1999.
- [18] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases," *Comput. Vis. Image Understanding*, vol. 75, no. 1–2, pp. 111–132, 1999.
- [19] S. Ghebreab and A. W. M. Smeulders, "Strings: variational deformable models of multivariate ordered features," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, pp. 1399–1410, Nov. 2003.
- [20] J. Ramsay and B. W. Silverman, *Functional Data Analysis*. Berlin, Germany: Springer-Verlag, 1997.
- [21] S. Baluja, "Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning," Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-94-163, 1994.
- [22] R. R. Long and G. R. Thoma, "Image query and indexing for digital X-rays," *Proc. SPIE (Storage and Retrieval for Image and Video Databases VII)*, pp. 12–21, 1999.
- [23] National Center for Health Statistics. National Health and Nutrition Examination Survey (1976–1980). National Library of Medicine. [Online]. Available: <http://archive.nlm.nih.gov/>
- [24] D. M. Squire, W. Muller, H. Muller, and J. Raki, "Content-based query of image databases, inspirations from text retrieval," *Pattern Recogn. Lett.*, vol. 21, no. 11, pp. 1193–1198, 2000.
- [25] M. J. Bates, "An exploratory paradigm for online information retrieval," in *Intell. Inform. Syst. Inform. Soc.*, 1986.
- [26] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Roy. Statist. Soc.. Ser. B. Methodological*, vol. 53, no. 2, pp. 285–339, 1991.