A continuous and probabilistic framework for medical image representation and categorization

A. Pinhas and H. Greenspan*
Department of Biomedical Engineering, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

ABSTRACT

This work focuses on a general framework for image representation and image matching that may be appropriate for medical image archives. The proposed methodology is comprised of a continuous and probabilistic image representation scheme using Gaussian mixture modeling (GMM) along with information-theoretic image matching measures (KL). The GMM-KL framework is used for matching and categorizing x-ray images by body regions and orientation. A 4-dimensional feature space is used to represent the x-ray image input, including intensity, texture (contrast) and spatial information (x,y). Unsupervised clustering via the GMM is used to extract coherent regions in feature space, and corresponding coherent segments ("blobs") in the image content. The blobs are used in the matching process. A dominant characteristic of the radiological images is their poor contrast and large intensity variations. This presents a challenge to matching between the images and is handled via a post-processing stage that provides an invariant blob-signature per image input. In a leave-one-out procedure, each image out of 851 is used once as a test-image, and is categorized by the remaining (labeled) images. The GMM-KL classifier was tested using 851 radiological images with error-rate of 1%. The classification results compare favorably with reported global representation schemes, such as histograms.

Keywords: Statistical medical image modeling, image distance, medical image categorization, PACS, Segmentation and grouping, Content-Based Image Retrieval (CBIR)

1. INTRODUCTION

Medical image databases are a key component in future diagnosis and preventive medicine. There is an increasing trend towards the digitization of medical imagery and the formation of adequate archives. With the growing size of medical image libraries, there is a need for efficient tools that can analyze medical images content and represent it in a way that can be efficiently searched and compared. The objective of this research is to explore a statistical framework for visual information management in medical archives. The framework provides a novel representation for modeling image content, as well as novel distance measures for image comparisons and categorization. The focus of this work is matching and categorizing x-ray images by body regions and orientation. The proposed methodology is general and can be extended to additional modalities and labeled categories.

Content-based indexing and retrieval is expected to have a great impact on medical image databases. Image selection is currently based on alphanumerical information only. However, information contained in medical images differs considerably from that residing in alphanumerical format. Recent published works in medical content-based retrieval propose solutions, which are limited to images with specific organ, modality or diagnostic study, and are usually not directly transferable to other medical applications. A few examples are: lung CT scans [7], mammogram [8,9], brain CT [10], and spine radiographs [11].

This work presents the initial phase of a large-scope research, in which a statistical framework, called GMM-KL, is used as a probabilistic framework to represent and match medical images in large archives. The task of categorizing image content within a large archive is a new challenge in the community (as opposed to focusing on exact

* Corresponding author. Tel.: +972-3-6407398; Fax: +972-3-6407939.
Email address: hayit@eng.tau.ac.il (H.Greenspan)

segmentation within a particular modality). To the best of our knowledge, this is the first time a statistical framework is applied in the context of categorizing medical imagery. Additional unique challenges treated in the work include: exploring the feature-space appropriate to represent x-ray images; transitioning from global-based representations (such as histograms) to localized, region-based representations (via GMM) and illumination-invariant signatures for image matching.

2. DATA CHARACTERISTICS

The database used in this research is a subset of eight classes from the IRMA x-ray library [3]. The images were classified by medical experts according to the imaging modality, the examined region, the image orientation with respect to the body and the biological system under evaluation. This classification is used as the "ground-truth" for this research.

Classification of x-ray images (radiographs) is a non-trivial task, due to the complex nature of the information in the image. In a single chest image, there are lungs, heart, ribs cage, diaphragm, clavicle, shoulder blade, spine and blood vessels, any of which may be the region of interest for the radiologist. X-ray images of the same class share a strong visual similarity (Figure 1). However, there is a great variation within a class, caused by different doses of x-ray, varying orientation, alignment and pathology. In many images there are cloths, jewels, artificial-implants and medical instruments. In addition to content variation, the quality of the x-ray images may vary considerably. The images are characterized with contrast variation and non-uniform intensity background, weak signal-to-noise ratio, digitized x-ray projections noise, and high frequency noise.

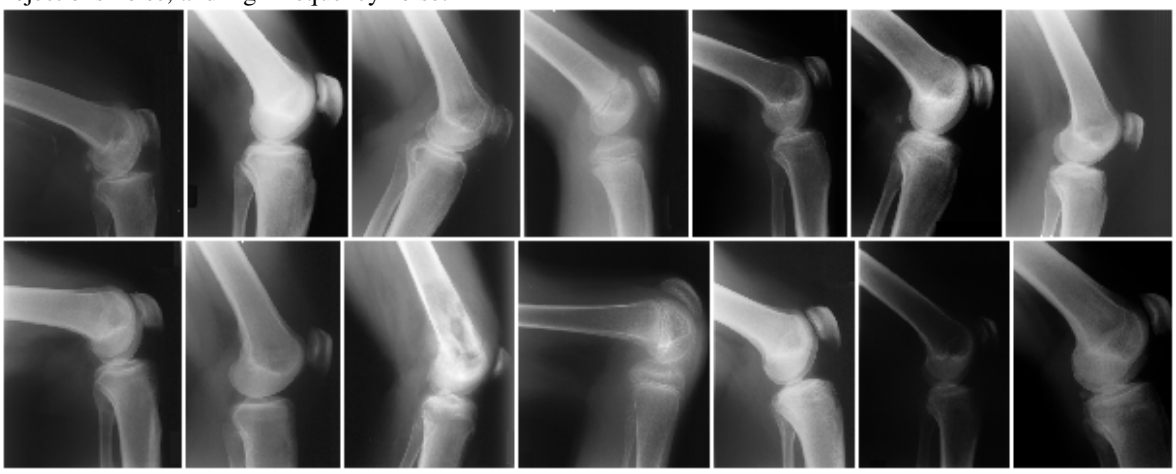

Figure 1: Sample x-ray images from the knee class.

3. FEATURE SPACE SELECTION

In the current work a 4-dimensional feature space is used to represent the x-ray image input, including intensity, spatial information and texture. An initial transition is made from pixels to the selected 4-D feature space. Each pixel is represented with a feature vector and the image as a whole is represented by a collection of feature vectors. In the following, we briefly review the texture feature used.

3.1 Texture feature extraction

The texture descriptor used in this work is based on the windowed second moment matrix, as defined in [2]. The gradient of the image intensity, ∇I, is computed using the first difference approximation along each dimension. Scale is defined to be the width of the Gaussian window within which the gradient vectors of the image are pooled. The second moment matrix for the vectors within this window, computed about each pixel in the image, can be approximated using:

$$M_{\sigma}(x, y) = G_{\sigma}(x, y) * (\nabla I)(\nabla I)^{T} \quad (1)$$

where $G_\sigma(x,y)$ is a separable binomial approximation to a Gaussian smoothing kernel with variance σ^2. At each pixel location, $M_\sigma(x,y)$ is a 2x2 symmetric positive semidefinite matrix. Rather than work with the raw entries in M_σ, it is more common to deal with its eigenstructure. Consider a fixed scale and pixel location. Let λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) denote the eigenvalues of $M\sigma$ at that location, and let φ denote the argument of the principal eigenvector of $M\sigma$. When λ_1 is large compared to λ_2, the local neighborhood possesses a dominant orientation, as specified by φ. When the eigenvalues are comparable, there is no preferred orientation, and when both eigenvalues are negligible, the local neighborhood is approximately constant.

We may think of σ as controlling the size of the integration window around each pixel within which the outer product of the gradient vectors is averaged. Note that $\sigma=\sigma(x,y)$; the scale varies across the image. In order to select the scale at which $M\sigma$ is computed, i.e. to determine the function $\sigma(x,y)$, we make use of a local image property known as polarity. The polarity is a measure of the extent to which the gradient vectors in a certain neighborhood point in the same direction. The polarity at a given pixel is computed with respect to the dominant orientation φ in the neighborhood of that pixel. For each pixel, we select the scale as the first value of σ for which the difference between successive values of polarity $P(\sigma_k)$-$P(\sigma_{k-1})$ is less than 2%. Once a scale σ is selected for each pixel, that pixel is assigned three texture descriptors. The first two, which are taken from M_σ, are the anisotropy, defined as $A=1-\lambda_2/\lambda_1$, and the normalized contrast, defined as $C=(\lambda_1+\lambda_2)^{1/2}$. The third is the polarity, P, which is defined as:

$$P = \frac{|E_+ - E_-|}{E_+ + E_-} \tag{2}$$

where the definition of E_+ and E_- are:

$$E_+ = \sum_{(x,y)\in\Omega} G_\sigma(x,y)[\nabla I \cdot \hat{n}]_+$$

$$E_- = \sum_{(x,y)\in\Omega} G_\sigma(x,y)[\nabla I \cdot \hat{n}]_-$$

In this work, we use the contrast as the texture feature. In [2], a pixel is referred to as uniform (non texture) if its mean contrast across scale is less than 0.1, where the contrast ranges from 0 to 1. Due to the varying quality of the x-ray images and the range of possible textures in the images, a constant threshold cannot be used to differentiate between the real texture and the noise. We wish to determine the threshold adaptively per image. For this task, a histogram of the texture contrast is used. This histogram is typically a mixture of two distributions; the lower-mean distribution is generated by the background (non texture), while the higher mean distribution is generated by textured objects. The optimal threshold for each image is taken as the minimum point between the two distributions, defined as the minimum point between two local maxima points, according to highest maxima to minima ratio.

4. IMAGE REPRESENTATION & MATCHING

The proposed methodology is comprised of a continuous and probabilistic image representation scheme along with information-theoretic image matching measures. Unsupervised clustering via Gaussian mixture modeling (GMM) is used to extract coherent regions in feature space, and corresponding coherent segments ("blobs") in the image content. The parameters of the GMM are determined via the maximum likelihood principle and the Expectation-Maximization (EM) algorithm. Following the image modeling stage, the image-matching problem is treated as a distribution-matching problem, and the information theoretic Kullback-Leibler (KL) distance is used as a distance measure between image GMMs. The combined GMM-KL framework has recently been proposed [1] as an extension to the content-based retrieval system of Blobworld [2].

4.1 Grouping pixels into regions

Once the feature space is selected, grouping the pixels is the next required stage. The pixels are grouped into homogeneous regions by grouping the feature vectors in the selected feature space. The feature space is searched for dominant clusters and the image samples in the feature space are then represented via the modeled clusters. The underlying assumption is that the image intensities and their spatial distribution in the image plane are generated by a

mixture of Gaussians. Each homogeneous region in the image plane is thus represented by a Gaussian distribution, and the set of regions in the image is represented by a Gaussian mixture model. Learning a Gaussian mixture model is in essence an unsupervised clustering task.

The Expectation Maximization (EM) algorithm [1] is used to determine the maximum likelihood parameters of a mixture of *k* Gaussians in the feature space. The distribution of a random variable $X \in R^d$ is a mixture of *k* Gaussians if its density function is:

$$f(X \mid \theta) = \sum_{j-1}^{k} a_j \frac{1}{\sqrt{(2\pi)^d \left|\Sigma_j\right|}} \exp\left\{-\frac{1}{2}(X-\mu_j)^T \Sigma_j^{-1}(X-\mu_j)\right\} \tag{3}$$

such that the parameter set $\theta = \left\{a_j, \mu_j, \Sigma_j\right\}_{j-1}^{k}$ consists of:

- $\alpha_j > 0 \quad , \quad \sum_{j-1}^{K} \alpha_j = 1$
- $\mu_j \in R^d \quad and \quad \Sigma_j$ is a *dxd* positive definite matrix.

where α_j is the prior probability for Gaussian *k,* and μ_k, Σ_k are the mean vector and covariance matrix of Gaussian *k*, respectively.

Given a set of feature vectors $x_1,\ldots,x_n$, the maximum likelihood estimation of θ is:

$$\theta_{ML} = \arg\max_{\theta} f(x_{1,\ldots,} x_n \mid \theta) \tag{4}$$

The EM algorithm is an iterative method to obtain θ_{ML}. The first step in applying the EM algorithm to the problem at hand is to initialize the mixture model parameters. The K-means algorithm [2] is utilized to extract the data-driven initialization. The updating process is repeated until the log-likelihood is increased by less than a predefined threshold. In this work we use a threshold value of 1%.

The number of mixture components (or number of means), *k*, is of great importance in the accurate representation of a given image. Ideally, *k* is to represent the value that best suits the natural number of groups present in the image. Note that each of these feature groups may include several spatially disjoint regions in the image. It is often accepted that the Minimum Description Length (MDL) principle [4, 5] may serve to select among values of *k*. In our experiments, the MDL criterion indicates a monotonic improvement of less than 0.5% above 7 or 8 blobs. We therefore select *k*=8 for all images.

4.2 Image segmentation

An immediate transition is possible between the image representation using a Gaussian mixture model, and probabilistic image segmentation. A direct correspondence can be made between the mixture representation and the image plane. Each pixel of the original image is now affiliated with the most probable Gaussian cluster. The labeling of each pixel is done in the following manner. Suppose that the parameter set that was trained for the image is $\theta = \left\{a_j, \mu_j, \Sigma_j\right\}_{j=1}^{k}$

Denote:

$$f_j(x \mid \alpha_j, \mu_j, \Sigma_j) = \alpha_j \frac{1}{\sqrt{(2\pi)^d \mid \Sigma_j \mid}} \exp\left\{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\right\} \tag{5}$$

Then the labeling of the pixel related to the feature vector x is chosen as follows:

$$Label(x) = \arg\max_j f_j(x \mid \alpha_j, \mu_j \Sigma_j) \tag{6}$$

4.3 Image similarity and matching

The localized Gaussian mixture provides for a compact representation of the image in the feature space. This representation is used for images similarity comparisons, using the Kullback Leibler (KL) distance [6]. The KL distance (or relative entropy) is a measure of the distance between two distributions based on information theory. It is consistent with the probabilistic modeling technique and can be efficiently evaluated through Monte Carlo procedures.

Once we associate a Gaussian mixture model with an image, the image can be viewed as a set of independently identically distributed (IID) samples from the Gaussian mixture distribution. Hence, a reasonable distance measure between two images is a distance measure between the two Gaussian mixture distributions obtained from the images. Denote the Gaussian mixture models computed from the two images by f_1 and f_2. Given the two distributions: f_1 and f_2, the non-symmetric version KL distance is:

$$D(f_1 \parallel f_2) = E_{f_1} \log \frac{f_1(x)}{f_2(x)} \tag{7}$$

where E is the expected value function. Since the KL distance between two Gaussian mixture distributions cannot be analytically computed, we can instead apply the image data to approximate it. One possible approximation is to use synthetic samples, denote by $x_1 \ldots x_n$, produced from the Gaussian mixture distribution, f_1. This enables us to compute the KL distance without referring to the images from which the models were built:

$$D(f_1 \parallel f_2) \cong \frac{1}{n} \sum_{t-1}^{n} \log \frac{f_1(x_t)}{f_2(x_t)} \tag{8}$$

such that $x_{i1} \ldots x_{ini}$ is the feature set extracted from image i, (i=1,2), and n_i is the size of this set.

4.4 Normalized blobs

As discussed earlier, dominant characteristics of the radiological images are their poor contrast and large intensity variations. This presents a challenge in matching between the images (or Gaussian mixtures), as similar images may share only a limited space in the feature space. For example, dark and bright images of similar visual content may result in a no-match decision due to the large intensity variations. A post-processing stage is suggested to overcome such difficulties. In the post-processing, blob representation per image is normalized to the range of [0-1] such that the blob that represents the least dense organ (the background in most of the cases) is assigned an intensity value of 0, and the blob that represents the densest organ is assigned an intensity value of 1. The rest of the blobs' normalized-values are distributed between 0 and 1, in correspondence to the original blob-intensity value. After the normalization of each image representation, all the Gaussian mixtures are distributed in the same range. Furthermore, the corresponding blobs between any two similar images, which represent the same organ, are likely to have closer intensity values and result in a lower KL distance between the two blob representations.

5. EXPERIMENTS AND RESULTS

A set of 851 images was selected from eight different classes, where each class contains 50-200 images (Table 1). Figures 4 and 5 (top) show sample images from the eight classes. The described algorithms used 8-bit gray level uncompressed images in reduced resolution of typically 300x500 pixels (images in the database have up to 2000x3000 pixels). In the following experiments, we start by showing results related to the texture feature, we then proceed to image representation and matching results.

5.1 The texture descriptor

Figure 2 left, demonstrates that there is no clear border in the intensity image between the organ and background. Using the texture contrast feature (section 3.1), a distinct border between the organs and the background or between bones and soft tissues can be seen, as shown in Figure 2 right.

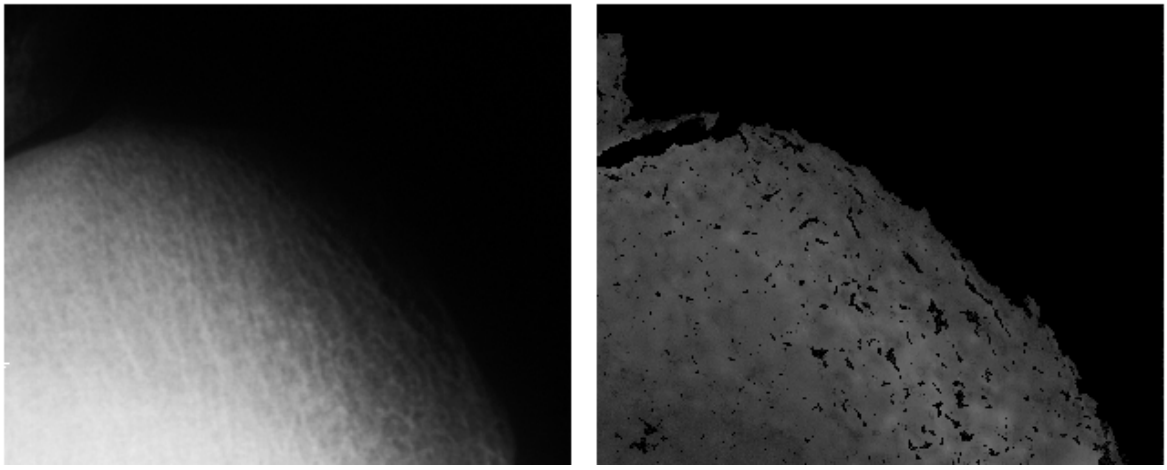

Figure 2- Left: Original x-ray image. Right: Corresponding texture contrast image, ranging from 0 (dark) no texture, to 1 (bright) strong texture.

Image segmentation results, with and without texture, are shown in Figure 3 (equations 5 and 6). The segmentation results provide a visualization tool for better understanding the image model. Uniformly colored regions (with arbitrary gray-levels) represent homogeneous regions in the feature space. The associated pixels are all linked (unsupervised) to the corresponding Gaussian characteristics. When the texture is included in the feature space (Figure 3b), a distinction is found between the soft-tissue, segment 1, and the background, segment 2. Excluding the texture from the feature space (Figure 3c) results in merging of the regions into segment 5. In a similar way, the bone, segment 3, and the background, segment 4 (figure-3b) were merged into segment 6 (figure-3c).

.

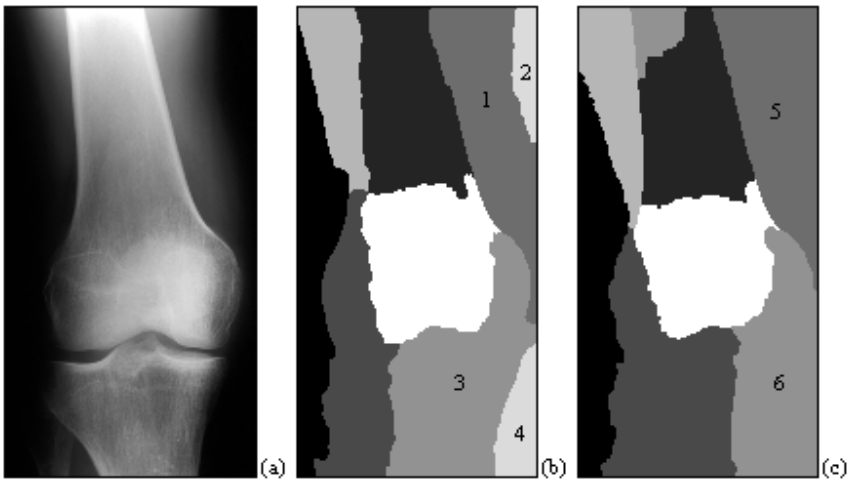

Figure 3- (a) Original x-ray image. (b) Segmentation results for a model learned using intensity, position and texture features. (c) Segmentation results for a model learned using intensity and position features only.

5.2 Image representation

Image representation using GMM modeling can be seen in Figures 4 and 5. In this visualization, each localized Gaussian mixture is shown as a set of ellipsoids. Each ellipsoid represents the support, mean intensity and spatial layout, of a particular Gaussian in the image plane. The blob representation demonstrates that even in the simplified blob-visualization one can identify the image type, since generally, the regions correspond to objects or parts of objects. For example, in the 'Chest' image, the two lungs are represented by the two dark ellipsoids, while in 'Chest side view' image, a single dark ellipsoid represents a side view of the lungs, and in the 'Hand' image the thumb-blob is separated from the fingers-blobs.

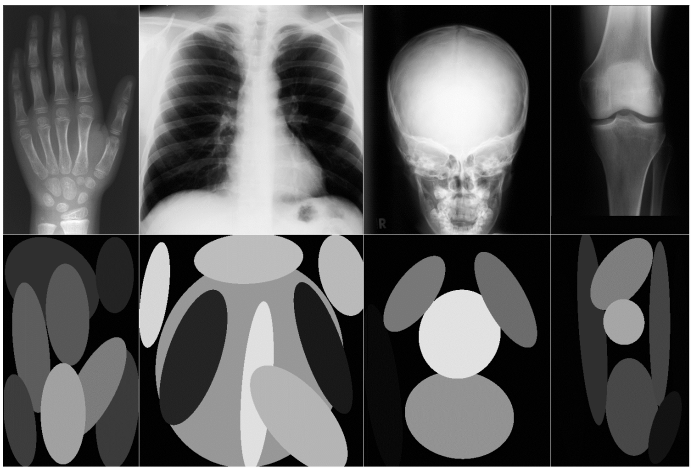

Figure 4-

Top: Example images. From left to right: hand, chest, skull, and knee classes.

Bottom: Corresponding GMM (blobs) representation.

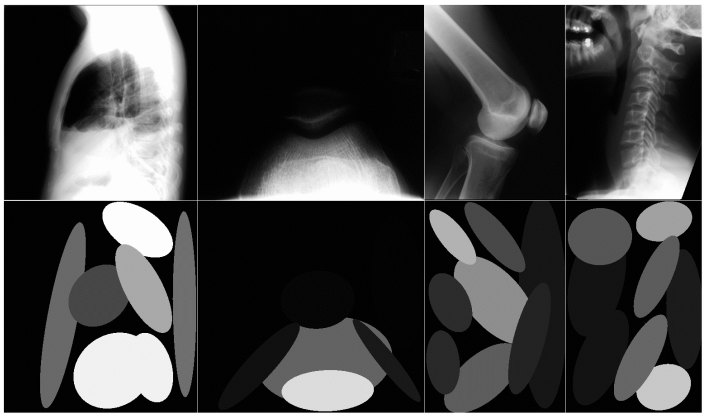

Figure 5-

Top: Example images. From left to right: chest side view, patella, knee side view, and Neck classes.

Bottom: Corresponding GMM (blobs) representation.

In order to evaluate the GMM blob representation for the x-ray images, we use an intra-inter class statistical evaluation methodology. The intra-class set consists of image pairs from the same class. The inter-class set corresponds to pairing of images from different classes. A histogram of the intra-class and inter-class distances is given in Figure 6. The x-axis is the KL distance and the y-axis is the frequency of occurrence of the respective distance in each of the two distance sets. The two distinct peaks in the graph demonstrate the separation between the sets. An overlap exists; this demonstrates the complexity of the task. As expected, the intra-class distances have low values, while the inter-class set is more widely spread over larger distance values. We next wish to investigate if the separation between the sets is sufficient for the image classification task.

5.3 Image classification

The final set of experiments consist of classification experiments, to evaluate image matching via the GMM-KL framework. A leave-one-out procedure is used. Each image, out of 851, is used once as a test-image, and is categorized by the remaining (labeled) images. Table 1 presents the categorization results per image category. Results are an

average of several K nearest-neighbor voting cycles (varying K). A 2% error rate is achieved with the initial blobs model. Using the normalized blob models the error rate is less than 1%.

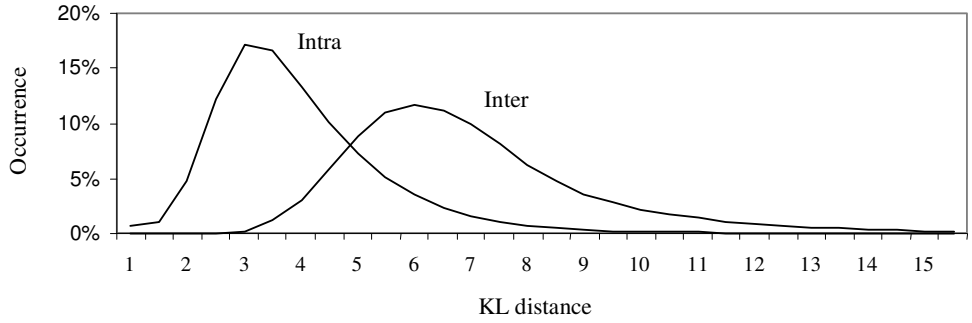

Figure 6: Inter-intra graph

Table 1: (a) Image Classes; (b) Number of images per class; (c) Number of misclassified images per class.

Image Classes (a)	Number of Images in a class (b)	Number of misclassified images (c)	
		Initial blobs	Normalized blobs
Hand	120	2	2
Chest	168	0	0
Chest side view	208	6	0
Skull	121	2	2
Knee	77	4	2
Patella	48	1	0
Knee side view	53	5	3
Neck	56	3	0
Total:	851	18 (2.11%)	8 (0.94%)

6. CONCLUSIONS

The GMM-KL framework was shown to provide a high precision of 99.06% in classifying x-ray images. Results presented compare favorably with global and local based representations schemes as reported in the literature [3]. Several research questions are still open with respect to the GMM representation, which is a localized representation. An important issue is the ability to match similar images with variations in the alignment or zoom. We are currently investigating invariant image descriptors, such as blobs with relative layout features, as a solution to this challenge. Another open question is how to retrieve images from a sub-class; for example, retrieving the images in which there is only a single healthy lung, from the chest-class. Future work entails expanding the image dataset with additional classes, and performing large-scope validation and comparative studies.

ACKNOWLEDGEMENT

The image data used in this study is courtesy of the Image Retrieval in Medical Application (IRMA) group, Aachen, Germany, http://irma-project.org.

REFERENCES

1. H. Greenspan, J. Goldberger, L. Ridel, A Continuous probabilistic framework for image matching, Journal of Computer Vision and Image Understanding, Vol. 84, No. 3, pp.384-406, 2001.
2. C. Carson, S. Belongie, H. Greenspan, J. Malik, Recognition of images in large databases using color and texture, IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE-PAMI, 24(8): 1026-1038, 2002.
3. T. Lehmann, B. Wein, D. Keysers, M. Kohnen, H. Schubert, *A mono-hierarchical multi-axial classification code for medical images in content-based retrieval*, Proceedings 1st IEEE International Symposium on Biomedical Imaging, pp. 313-316, 2002.
4. J. Rissanen, *Modeling by shortest data description*, Automatica 14, 465-471, 1978.
5. J. Rissanen, *Stochastic complexity in statistical inquiry*, World scientific, 1989.
6. S. Kullback, *Learning Textures*, Dover, 1968.
7. A.C. Kak, C. Pavlopoulou, *Computer Vision Techniques for Content-Based Image Retrieval from Large Medical Databases*, 7th Workshop on Machine Vision Applications, IAPR, Tokyo, Japan, 2000.
8. A. Maria-Luiza, O.R. Zaïane, A. Coman, *Application of Data Mining Techniques for Medical Image Classification*, in Proc. of Second Intl. Workshop on Multimedia Data Mining (MDM/KDD'2001) in conjunction with Seventh ACM SIGKDD, pp. 94-101, San Francisco, CA, 2001.
9. P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, Z. Protopapas, *Fast and effective retrieval of medical tumor shapes*, IEEE Transactions on Knowledge and Data Engineering, 10(6): 889–904, 1998.
10. Y. Liu, F. Dellaert, *Classification Driven Medical Image Retrieval,* Proc. of the Image Understanding Workshop, 1998.
11. L.R. Long, S. Antania, D.J. Leeb, D.M. Krainakc, G.R. Thomaa, *Biomedical information from a national collection of spine x-rays – Film to content-based retrieval*, Proceedings SPIE 2003.