Image Classification from Generalized Image Distance Features: Application to Detection of Interstitial Disease in Chest Radiographs

Yulia Arzhaeva and Bram van Ginneken
Image Sciences Institute,
University Medical Center Utrecht, The Netherlands
yulia@isi.uu.nl, *bram@isi.uu.nl*

David Tax
Delft University of Technology,
The Netherlands
d.m.j.tax@ewi.tudelft.nl

Abstract

One of the most important tasks in medical image analysis is to detect the absence or presence of disease in an image, without having precise delineations of pathology available for training. A novel method is proposed to solve such a classification task, based on a generalized representation of an image derived from local per-pixel features. From this representation, differences between images can be computed, and these can be used to classify the image requiring knowledge of only global image labels for training. It is shown how to construct multiple representations of one image to get multiple classification opinions and combine them to smooth over errors of individual classifiers. The performance of the method is evaluated on the detection of interstitial lung disease on standard chest radiographs. The best result is obtained for the combining classification scheme yielding an area under the ROC curve of 0.955.

1. Introduction

In this paper we discuss how to solve a medical image classification task, where depicted objects of the same kind have to be classified either normal or diseased, and there are only differences in local texture, often ill-defined, between objects from both classes. Previous pattern recognition approaches to this problem are based on pixel or region classification and subsequent fusion of local posterior probabilities to obtain an overall decision for the image under consideration. These approaches require pixel or region labels to train local classifiers. For example, to detect interstitial disease on chest radiographs pixel- and region-based classification approaches were applied in [1] and [2] respectively. In practice, a pixel or region ground truth is often unavailable, or is unreliable for ill-defined lesions. A ground truth on the image level, on the other hand, is almost always available during the collection of a data set, or is easier to obtain. Therefore we aim at a classification approach that allows us to classify an image as a whole from only overall image labels. Since the information that concerns the presence or absence of pathology is local, an image representation is introduced where global image features are derived from local per-pixel features.

The starting point of the method is the extraction of local features from spatially corresponding pixels in all images under consideration. One way to obtain corresponding pixels across images is to warp a mean image to all images. We segment an image to get a number of fixed landmarks that can be used to establish a warping function. Next, a new set of image features is derived from local features of that image and local features of another image, which we call a reference image. Features are calculated as distances between two vectors whose elements are values of a certain per-pixel feature in all corresponding points from a given image and a reference image. In this feature space supervised classification of images can be performed using only global class labels of training images. With several reference images, different image representations can be constructed and a pool of classifiers can be trained. Then, for an unseen image, it is possible to combine multiple classification opinions in order to smooth over mistakes of individual classifiers.

The experiments in this paper focus on a medical classification task, but the framework can be used for any image classification task where only overall image class labels are available. In other domains, such as object detection, this type of task is also recognized to be important (e.g. in [3]) and often referred to as weakly labelled image data.

In the next Section, the method is described in detail. Section 3 presents experimental results. Finally, Section 4 provides a discussion and conclusions.

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

2. Method

Further in this section we assume that the classification task relates to an area of interest within an image, and the word "shape" is used to denote such an area. Shapes of the same kind are aimed to be classified, e.g. lung fields in chest radiographs.

2.1. Image representation and pixel correspondence

For our method it is important to establish the correspondence between points within analogous shapes on two different images. Theoretically, various approaches to that might be considered. We describe here an approach that uses Active Shape Models (ASM) for segmentation (see [4]). After segmenting each shape X using ASM, positions of a number of corresponding landmark points on the outlines of images become available. From a large set of images, the mean position of these points can be computed. An image with these mean points is called a mean image. Next, a warping function is determined between an image I containing a shape X, and the mean image by demanding that the mean points are warped to the ASM landmark points in the image I. We use a warping algorithm that is described in [5]. A warping function finds for a point in the mean image a related point in the image I.

We consider the following representation of X. Assume, that a number of features M are extracted per pixel $i, i \in X$. A pixel i is represented by a feature vector $x_i = (x_{i1}, x_{i2}, \ldots, x_{iM})$. Thus we get a matrix representation of X:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{pmatrix} \quad (1)$$

where N is the number of pixels selected in X. To be able to represent every X by an $N \times M$ matrix, we look for correspondent N pixels in every image. We select N pixels uniformly within the shape of interest in the mean image and warp these to the image under consideration.

The representation (1) can be re-written as

$$X = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_M), \quad (2)$$

where $\bar{x}_j = (x_{1j}, x_{2j}, ..., x_{Nj})^T, j \in [1, M]$. We will use this representation of X in further calculations.

2.2. Reference image and distance features

A general pattern classification framework that uses the notion of proximity between objects is described in [6]. Our methods differs in that that the multidimensional feature space we construct combines distance measures w.r.t. one reference image.

We call an image I_r containing a shape of interest R, a reference image when it is used to derive a new set of image features for a given image I. The same representation as in (2) is computed for R, $R = (\bar{r}_1, \bar{r}_2, ..., \bar{r}_M)$. We introduce a distance vector $d(X,R)$ between shapes X and R:

$$\begin{aligned} d(X,R) &= (\| \bar{r}_1, \bar{x}_1 \|, \| \bar{r}_2, \bar{x}_2 \|, ..., \| \bar{r}_M, \bar{x}_M \|) = \\ &= (d_1, d_2, ..., d_M), \end{aligned} \quad (3)$$

where $\| \cdot \|$ denotes any metric, e.g. a Euclidean distance. We consider such a vector $d(X,R)$ a feature vector for X in an M-dimensional feature space. Note that different metrics can be used to compute different features $d_j, j \in [1, M]$. Given a certain I_r and a set of training images with known class labels, it is possible to describe each training shape X_{tr} by $d(X_{tr}, R)$ and train a classifier on a set of obtained distance vectors. In extreme case a distance vector for X might be calculated with respect to a so-called "zero image", i.e. when a representation (1) is a zero matrix $\mathbf{0}$, and $d(X, \mathbf{0})$ is a vector of norms of vectors $\bar{x}_j, j \in [1, M]$.

2.3. Image classification

We denote a classifier that is used in a feature space built with respect to (w.r.t.) a reference image I_r as f_r. For an unseen shape X_0 a classifier f_r yields either posterior probabilities or a class label. From different reference images, different distance vectors for a certain shape X can be calculated and used to train multiple classifiers. For an unseen shape X_0 and s reference images $R_1, ..., R_s$, s descriptions of X_0 are possible, and subsequently s classification opinions $f_{r1}(d(X_0, R_1)), ..., f_{rs}(d(X_0, R_s))$. A performance of f_r is dependent on how separable data is in the feature space constructed w.r.t. I_r. Hence one way to solve a classification task could be to find the best performing I_r. Since it is not trivial to do that, or there simply might be no excellently performing reference images among the available ones, it seems advantageous to combine outputs of several classifiers in order to compensate for possible mistakes of individual classifiers. In our experiments we applied static fusion schemes such as the mean, minimum, maximum, product, vote and percentile rules.

3. Experimental results

A medical image analysis classification task is considered in this work. The goal of the task is to classify

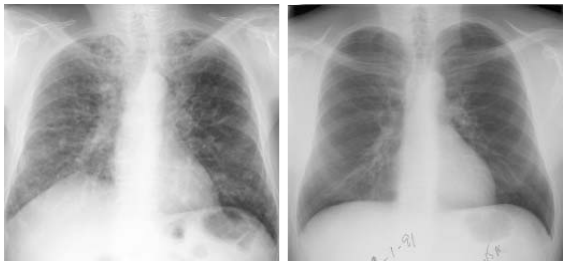

Figure 1. A radiograph of a patient with interstitial lung disease (left) and a radiograph of a healthy person (right).

chest radiographs on presence or absence of interstitial lung disease. Interstitial disease is a general term for a number of lung diseases characterized on radiographs by moderate shadowing with underlying abnormal texture patterns. Figure 1 illustrates how normal and diseased lungs might look on a radiograph. Note that the abnormal case has ill-defined, widespread subtle differences in texture compared to the normal case. Since normal anatomical structures are superimposed on these subtle textures, reliable detection of the presence of interstitial disease from these radiographs is difficult. This task was also studied in [2] and [1].

3.1. Data features and classifiers

A database of 200 posterior-anterior chest radiographs is used that is obtained from University of Chicago hospitals [7]. Both normal and abnormal (diseased) classes contain 100 images. Shapes we are interested in are the lung fields. They are segmented using the ASM algorithm, description of which can be found elsewhere (e.g. in [4]). The segmentation is performed with the same settings of parameters as in [8]. The resolution of images is sub-sampled to 700×700. For a more detailed description of the data see [2].

The mean image is calculated from available images with the previously segmented lung fields, and every $5th$ pixel in X and Y directions is selected within the mean lung fields, in total 7103 pixels. Correspondent pixels are subsequently obtained in every image using a warping algorithm as described in [5]. For every selected pixel i the local features are extracted that include the pixel intensity, the outcomes of Gaussian derivatives multi-scale filter bank in the pixel (scales 1, 2, 4, 8, and 16 from 1 to 16, derivatives of order $0, 1$and 2), and local statistics (the mean, standard deviation, skew and kurtosis) in a circular area centered in the pixel (radius 32 pixels). This feature space dimensionality equals 157.

A distance feature vector computed for a pair of images has the same dimensionality, as well as a vector of norms for an image. A Euclidean metric is used to compute distance vectors. In further experiments the features are normalized to have total zero mean and unit variance, and subsequently principal component analysis retaining 99% of variance is applied to reduce feature space dimensionality and prevent instability of numerical computations. In all conducted experiments, posterior probabilities for an image are obtained using a linear discriminant classifier (LDC). The choice of this classifier is based on a pilot experiment, where LDC, a quadratic discriminant classifier, and a k-nearest neighbor classifier have been compared, and LDC performed notably better. The area under the receiver operating characteristic (ROC) curve, A_z, is used as performance measure. In the experiments where several classification outputs for an image are fused, mean and vote combining schemes appeared to perform better than other standard fusion schemes, and therefore only results from these schemes are given.

3.2. Results

Three types of classification experiments were conducted: one in a feature space of norms (w.r.t. a "zero image"), another experiment in a feature space constructed w.r.t. a single randomly chosen reference image, and a third experiment where outcomes of several classifiers constructed w.r.t. randomly chosen reference images were combined.

A leave-one-out technique was used to carry out the first experiment. This technique implies that each image from a data set is classified provided that a classifier is retrained each time with the remaining images. A classification result in terms of A_z is presented in Table 1 in the "zero image" row.

Table 1. Area under the ROC, A_z, performances, for varying number of reference images. When applicable, the average performance is presented, with the standard deviation in parenthesis.

Reference image(s)	A_z
"zero image"	0.926
one, normal	0.907 (0.016)
one, abnormal	0.914 (0.030)
10 normal, 10 abnormal, "mean"	0.958 (0.007)
10 normal, 10 abnormal, "vote"	0.950 (0.011)

In the second experiment a normal or abnormal image is randomly picked to be a reference image. Distance vectors are calculated for the remaining 199 images w.r.t. a reference image. Here again a leave-one-out scheme is applied to classify every image but a reference one, which is used only as a base to calculate feature vectors. The experiment is repeated ten times with a random selection of a reference image among normal images, and ten times with a random reference image selected among abnormal images. The mean value of A_z is shown in Table 1. Noteworthy is that it makes no difference for classification performance if a normal or abnormal image is used for reference: only the standard deviation appears to higher when classification is done in feature spaces constructed w.r.t. abnormal reference images.

In the third experiment ten normal and ten abnormal images are randomly selected to serve as reference images, which allows us to construct 20 different feature spaces and classify each of the remaining images 20 times using the leave-one-out technique. The classification opinions for each image are combined to obtain the final posterior probabilities for that image. The A_z performance is averaged over three runs. Results presented in Table 1 show that combining outcomes of independent classifiers improves the overall classification performance.

4. Discussion and conclusions

A new method of image representation was introduced where global image features are derived from local pixel features. This approach enables the construction of multiple representations of one image, which can be used to obtain multiple classification opinions about the image from training data that is labelled with only a single global class label for the complete image. It was shown that it is advantageous to combine those opinions to smooth over errors of individual classifiers.

The method was tested on the data set of chest radiographs representing normal lungs and lungs affected by interstitial disease. The classification performance is high when multiple image representations are used in combining classification schemes. We believe that the performance of individual classifiers might gain from more sophisticated feature selection than the simple principal component analysis used here. Nevertheless, performance of our method is already comparable with the results obtained in [1] on the same database. In that work, a combination of local per-pixel posterior probabilities was used to classify complete images.

It should also be noted that the image representation by means of a vector of norms yields equal or better classification performance than a representation w.r.t. one reference image. Another interesting observation is that A_z results are almost indiscernible for normal and abnormal reference images. We may conclude that the idea of a reference image gives us an advantage of multiple image representations and subsequent application of combining classifiers.

In future, development of a reference image selection procedure, which may be comparable to feature selection procedures, or application of more sophisticated combining rules than static fusion schemes are possible new research directions. Another interesting research question to focus on is how to exploit the notion of differences between images by not comparing corresponding pixels, but looking at the distributions of the feature values over the complete image. In this way, it is possible to obtain small distances between images with similar abnormalities but located at different positions.

References

[1] M. Loog and B. van Ginneken, "Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs," in *International Conference on Pattern Recognition*, J. Kittler, M. Petrou, and M. Nixon, Eds., 2004.

[2] B. van Ginneken, S. Katsuragawa, B. ter Haar Romeny, K. Doi, and M. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 139–149, 2002.

[3] C. M. Bishop and I. Ulusoy, "Object recognition via local patch labelling," in *Workshop on Machine Learning*, J. Winkler, N. Lawrence, and M. Niranjan, Eds., 2005.

[4] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[5] D. Ruprecht and H. Müller, "Image warping with scattered data interpolation," *IEEE Comput. Graph. Appl.*, vol. 15, no. 2, pp. 37–43, 1995.

[6] E. Pekalska, "Dissimilarity representations in pattern recognition. Concepts, theory and applications," Ph.D. dissertation, Delft University of Technology, the Netherlands, January 2005.

[7] S. Katsuragawa, K. Doi, and H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography: detection and characterization of interstitial lung disease in digital chest radiographs," *Med Phys*, vol. 15, no. 3, pp. 311–319, 1988.

[8] B. van Ginneken, M. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.