An Automatic Approach to Invariant Radiograph Classification

J. Dahmen, D. Keysers, M. Motter, H. Ney, T. Lehmann[1], B. Wein[2]

Lehrstuhl für Informatik VI, Computer Science Department
[1]Institute of Medical Informatics
[2]Department of Diagnostic Radiology
RWTH Aachen – University of Technology, D-52056 Aachen, Germany
Email: dahmen@informatik.rwth-aachen.de

Abstract In this paper we present an invariant statistical approach to classifying medical radiographs, being an important step in the RWTH Aachen – University of Technology IRMA system (Image Retrieval in Medical Applications). We propose a Bayesian classifier based on Gaussian kernel densities, where invariance is incorporated by using invariant distance measures. The performance of the classifier is evaluated on a dataset of 1,617 radiographs coming from daily routine. The obtained error rate of 7.6% is significantly better than the results reported in other works, using the same dataset. Furthermore, the presented probabilistic framework is also applicable to other (multi-) object recognition tasks.

1 Introduction

The importance of digital image retrieval techniques increases in the emerging fields of medical imaging and picture archiving and communication systems. Up to now, textual index entries are mandatory to retrieve medical images from a hospital archive, even if the archive is DICOM-compliant (Digital Imaging and Communications in Medicine) [1]. Furthermore, queries of diagnostic relevance include searching for organs, their relative locations and other distinct features like morphological appearances. Therefore, common retrieval systems cannot guarantee a meaningful query completion when used within medical context [2]. In contrast to this, the IRMA system - a joint project between three RWTH Aachen – University of Technology institutes - is being developed for use in daily clinical routine. This paper deals with the image classification step within the IRMA system, which is crucial, as the retrieval system needs to be familiar with the anatomical region presented in a given image in order to be able to answer complex medical queries. Detailed information on the motivation and the architecture of the IRMA system is given in [2]. We present a general probabilistic framework for object recognition and show its effectivity for the special case of radiograph classification, where invariance is incorporated by using invariant distance measures.

2 The IRMA database & feature analysis

The radiograph database used in our experiments consists of medical radiograph images taken from daily routine, which are secondary digital (that is they have

Figure1. Example radiographs taken from the IRMA database, scaled to common height. Top-left to bottom-right: abdomen, limbs, breast, skull, chest and spine.

been scanned from conventional film-based radiographs using 256 grayscales, cp. Fig. 1). The sizes of the anonymized images range from about 200×200 pixels to about $2,500 \times 2,500$ pixels and all images were labeled into six classes by an expert. The corpus consists of 110 abdomen, 706 limbs, 103 breast, 110 skull, 410 chest and 178 spine radiographs, summing up to a total of 1,617 images. Furthermore, a smaller set of 332 images exists which is used to test the generalization abilities of the classifier. To speed up the classification process, the original images are scaled down to a common height of 32 pixels (keeping the original aspect ratio). In the experiments, we make use of *appearance based* pattern recognition, that is each pixel of an image is interpreted as a feature. Furthermore, because there are only 1,617 images available, we make use of a *leaving-one-out* approach (L-1-o), that is to classify an image we use the remaining 1,616 images as references.

3 Classification

In many cases, classification of an observation $x \in \mathbb{R}^D$ is performed using the Bayesian decision rule

$$x \longmapsto r(x) = \underset{k}{\operatorname{argmax}} \left\{ p(k) p(x|k) \right\}, \tag{1}$$

where $p(k)$ is the prior probability of class k and $p(x|k)$ is the class-conditional probability for the observation x given class k [5, 3, 6]. Here, direct application of this rule is impossible, as the image sizes vary, resulting in different feature vector dimensions. Thus, we have to find the correct position of the object within the observed image, interpreting the remaining pixels as background. For this problem, we present a general, statistical multi-object recognition approach in the following, where M denotes the hypothesized number of objects present in a scene and radiograph classification will then be performed as the special case $M = 1$. We assume that the scene to be classified contains an unknown number $m = 0, ..., M$ of objects belonging to the classes $k_1, ..., k_M$, abbreviated as k_1^M in the following. Furthermore, reference models $p(x|\mu_k)$ exist for each of the known objects, μ_0 representing background. These references are subject to certain transformations (such as the position of the object in the image, its scale etc.). That is, given transformation parameters ϑ_1^M, the m-th reference is mapped to

$$\mu_{k_m} \rightarrow \tilde{\mu}(\mu_{k_m}, \vartheta_m). \tag{2}$$

Furthermore, the original scene is implicitly partitioned into $M+1$ regions I_0^M, where I_m is assumed to contain the m-th object and I_0 represents the background. The idea is now to hypothesize all unknown parameters, i.e. M, k_1^M, ϑ_1^M and I_0^M and to look for the hypothesis which best explains the given scene. Note that this means that any pixel in the scene has to be assigned either to an object or to the background class. Formally, the approach can be written as

$$r(\{x_{ij}\}) = \underset{M,k_1^M,\vartheta_1^M,I_0^M}{\operatorname{argmax}} \left\{ p(k_1^M) \prod_{(i,j)\in I_0} p_0(x_{ij}|\mu_0) \prod_{m=1}^{M} p_{k_m}(x_{I_m}|\tilde{\mu}(\mu_{k_m},\vartheta_m)) \right\} \quad (3)$$

where $\{x_{ij}\}$ denotes the scene to get classified and x_{I_m} is the feature vector extracted from I_m. For radiograph classification, the 'scene' equals the radiograph to be classified and we assume $M = 1$. Furthermore, the only transformation regarded for the reference images in our experiments is horizontal shift (vertical shifts do not occur as all images are scaled to the same height). A very simple background model is used, assuming a constant background of grayvalue zero. Furthermore, a penalty term is introduced, based on the different sizes of observation and reference image (preferring images of roughly the same size). To model the references $p_{k_m}(x_{I_m}|\tilde{\mu}(\mu_{k_m},\vartheta_m))$, kernel densities (with class specific standard deviations) are used and the the prior probabilities $p(k)$ are modeled via relative frequencies. More information on these models can be found in [3]. Note that in this work, all images were scaled down to a common size. Thus, Bayes rule was applied unchanged as given in Eq. (1). The statistical framework presented above is entirely new and also suited for other (multi-) object recognition tasks.

4 Invariant distance measures

In our experiments, the Mahalanobis distance present in the Gaussian kernel density approach is replaced by invariant distance measures. To compensate for global image transformations, single sided tangent distance is used as proposed by SIMARD in 1993. Due to space limitations, we cannot go into details of tangent distance here, more information about it can be found in [4] or [5]. In our experiments, we used a total of seven tangents (six for affine transformations and one for additive illumination variations [3]). To compensate for local image transformations, such as varying scribor positions or the presence/ absence of pathologies, the following image distortion model was used:

When calculating the distance between two images x and μ, small local deformations are allowed. That is, the image distortion model does not compute the squared error between a pixel (i,j) in x and its counterpart in μ, but it looks for the 'best-fitting' pixel in μ within a certain neighbourhood R_{ij} around the corresponding pixel:

$$D_{dist}(x,\mu) = \sum_{i=1}^{I} \sum_{j=1}^{J} \min_{(i',j')\in R_{ij}} \{\|x_{ij} - \mu_{i'j'}\|^2 + C(i,i',j,j')\} \quad (4)$$

for images with dimension $I \times J$. The cost function $C(i, i', j, j')$ models the costs for deforming a source pixel (i, j) in the input image to a target pixel (i', j') in the reference image. In the experiments, a weighted Euclidean distance between source and target pixel was used. Thus, short-ranged transformations are preferred to (most probably unwanted) long-range transformations. Furthermore, a region size $R_{ij} = (2r + 1) \times (2r + 1)$ was used with $r = 1$.

Note that tangent distance and the proposed distortion model can be easily combined to *distorted tangent distance*. In that case, tangent distance is used to register the (sub)images and the distortion distance ist then computed between the registered images.

5 Results

The experiments were started by using Mahalanobis distance within the classifier presented above, resulting in an error rate of 14.0%. Using single-sided tangent distance for recognition, this error rate could be reduced to 13.3%. Interestingly, using the image distortion model with a region size $r = 1$ significantly outperformed tangent distance on this particular dataset, yielding an error rate of 12.1%. In another experiment, it was investigated on the question whether the improvements of tangent distance and the image distortion model are additive. This sounds reasonable, as tangent distance compensates for global image transformations, whereas the image distortion model deals with local image perturbations. Indeed, using *distorted tangent distance* as proposed in Section 4, the error rate could be further reduced to 10.4%.

In another experiment, a *thresholding* approach was applied using $S = 5000$ (that is, the maximum local distance between two pixels was restricted to a maximum value S), in combination with the different distance measures discussed above. Doing so, the best error rate could be significantly reduced from 10.4% to 8.2%. Astonishingly, the result of tangent distance in that case is only slightly better than that of Mahalanobis distance (11.1% vs. 11.2%). One thing to be learned from this is that using the thresholding approach possibly mimics the behaviour of tangent distance in this particular application. It should also be noted that in previous experiments all IRMA images were scaled down to a common size of 32×32 pixels prior to classification (more information on that approach is given in [3]). In these experiments, tangent distance significantly outperformed Mahalanobis distance (with and without the thresholding approach). Thus, it seems possible that the main effect of tangent distance is the compensation of image shifts (which is now inherent to the classification approach by optimizing over all possible image positions). Surprisingly, not regarding background pixels at all improved the error rate to 7.6%. In that case, only the penalty term based on different image sizes between observation and reference was used. An overview of the results obtained on the radiograph database is given in Table 1. Note that no other group reports error rates of below 29% on the same dataset. More information on this topic can be found in [7, 3]. To make sure that no overfitting occurred in the experiments, 332 previously unseen radiographs were used as test images and the 1,617 images of the IRMA database as references,

Distance Measure	Thresholding	
	no	yes
Mahalanobis Distance	14.0	11.2
Tangent Distance	13.3	11.1
Image Distortion Model	12.1	9.0
Distorted Tangent Distance	10.4	8.2

Table1. L-1-0 IRMA error rates [%] for kernel densities and background model with respect to varying distance measures (with and without thresholding for S=5000).

using the optimal parameter set determined on the IRMA images. The obtained error rate of 9.0% shows, that the classifier proposed here generalizes very well.

6 Conclusion and outlook

In this paper, we presented a probabilistic framework for (multi-) object recognition and proved its effectivity by applying it to radiograph classification (being a single-object recognition task), obtaining an excellent result of 7.6%. Invariance was incorporated into the appearance based approach by using invariant distance measures. The proposed *distorted tangent distance*, being an extension of SIMARD's tangent distance, proved to be especially effective here. The presented approach also obtained promising results in multi-object digit recognition. These results will be published elsewhere.

References

1. M. Kohnen, H. Schubert, B. Wein, R. Günther, J. Bredno, T. Lehmann, J. Dahmen, "Qualität von DICOM Informationen in Bilddaten aus der klinischen Routine", to appear in *Bildverarbeitung für die Medizin*, Lübeck, Germany, 2001.
2. T. Lehmann, B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, M. Kohnen, "Content-based Image Retrieval in Medical Applications: A Novel Multi-step Approach", *Proceedings of the SPIE*, Vol. 3972(32), pp. 312–320, 2000.
3. J. Dahmen, T. Theiner, D. Keysers, H. Ney, T. Lehmann, B. Wein, "Classification of Radiographs in the 'Image Retrieval in Medical Applications' System (IRMA)", *6th International RIAO Conference on Content-Based Multimedia Information Access*, Paris, France, pp. 551–566, April 2000.
4. P. Simard, Y. Le Cun, J. Denker, "Efficient Pattern Recognition Using a New Transformation Distance," S.J. Hanson, J.D. Cowan, C.L. Giles (eds.): *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, San Mateo CA, pp. 50–58, 1993.
5. J. Dahmen, D. Keysers, H. Ney, M. Güld, "Statistical Image Object Recognition using Mixture Densities", to appear in *Journal of Mathematical Imaging and Vision*, Kluwer Academic Publishers, 2001.
6. D. Keysers, J. Dahmen, T. Theiner, H. Ney, "Experiments with an Extended Tangent Distance", *15th International Conference on Pattern Recognition*, Barcelona, Spain, pp. 38–42, September 2000.
7. J. Bredno, S. Brandt, J. Dahmen, B. Wein, T. Lehmann, "Kategorisierung von Röntgenbildern mit aktiven Konturmodellen", *Bildverarbeitung für die Medizin 2000*, München, Germany, pp. 356–360, März 2000.