Adaptive Metrics for Content Based Image Retrieval in Dermatology

Bunte, K.; Biehl, M.; Petkov, N.; Jonkman, M. F.

Published in:
17th European Symposium on Artificial Neural Networks (ESANN 2009)

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

Adaptive Metrics for Content Based Image Retrieval in Dermatology

Kerstin Bunte[1], Michael Biehl[1], Nicolai Petkov[1], Marcel F. Jonkman[2]
1 – Institute of Mathematics and Computing Science, University of Groningen
P. O. Box 407, 9700AK Groningen, The Netherlands
2 – Department of Dermatology
University Medical Center Groningen, University of Groningen
P. O. Box 30.001, 9700 RB Groningen, The Netherlands

Abstract. We apply distance based classifiers in the context of a content based image retrieval task in dermatology. In the present project, only RGB color information is used. We employ two different methods in order to obtain a discriminative distance measure for classification and retrieval: Generalized Matrix LVQ and Large Margin Nearest Neighbor approach. Both methods provide a linear transformation of the original features to lower dimensions. We demonstrate that both methods lead to very similar discriminative transformations and improve the classification and retrieval performances significantly.

1 Introduction

Content Based Image Retrieval (CBIR) constitutes an important tool for the handling of large amounts of visual information in medical applications [1, 2]. Three key steps can be identified in a generic CBIR system: (a) the extraction of information from images and its conversion to multi-dimensional feature vectors, (b) the computation of a suitable distance measure which quantifies the (dis-) similarity of a query image from the reference images, and (c) the identification of a set of data base images which display the smallest distances from given query.

Potential descriptors of image content include features which relate to color, texture, shape, or spatial relationship. Color has proven to be an effective descriptor regarding skin [3], especially in dermatology [4]. Here we address a specific problem of defining effective color descriptors for CBIR in dermatology. In [5] it was shown that the representation of an image in terms of color differences between lesion and healthy skin is advantageous over the use of the combined color features. Here, we extend this idea to the use of more general, discriminative linear combinations To this end we employ two learning techniques: Generalized Matrix LVQ (GMLVQ)[6] and the Large Margin Nearest Neighbor algorithm (LMNN)[7]. The distance measures are parameterized in terms of a transformation matrix which is adapted in a data driven learning process. Both methods lead to very similar features, which significantly improve the classification and retrieval performance.

After explaining the two approaches, we present first experiments with a database containing 211 images from four different classes of skin lesions in Sec. 3. We discuss the results in Sec. 4 and conclude with a summary and outlook.

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

2 Methods

We will make use of two trainable, distance based classifiers. In both approaches we employ parameterized distance measures which are adapted in a training phase, to achieve good classification and retrieval performance. In both schemes, the adaptive metrics defines a linear mapping of the original features to a lower-dimensional space, in which standard Euclidean distance can be used.

2.1 Matrix Relevance Learning Vector Quantization

Learning Vector Quantization (LVQ) and its variants belong to a popular family of prototype-based classifiers. In the following, training is based on examples of the form $\{(\boldsymbol{x}_i, y_i) \in \mathbb{R}^N\}_{i=1}^P$, where N is the dimension of feature vectors, P is the number of samples provided for training and the y_i are the corresponding discrete labels. In a C-class problem, at least C prototypes $\boldsymbol{w}_i \in \mathbb{R}^N$ are determined as typical representatives carrying labels $c(\boldsymbol{w}_i) \in \{1, \ldots, C\}$. Distances between prototypes and input vectors are determined according to a quadratic measure:

$$d^\Lambda(\boldsymbol{w}, \boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{w})^\top \Lambda (\boldsymbol{x} - \boldsymbol{w}) \in \mathbb{R}^N. \quad (1)$$

After training, LVQ realizes a "winner takes all" or "nearest prototype" classification scheme. We consider training by a stochastic gradient descent procedure, which presents a single example data at a time. It is guided by the minimization of a cost function defined by single example contributions as introduced in [8]:

$$f = \left(d_J^\Lambda - d_K^\Lambda\right) / \left(d_J^\Lambda + d_K^\Lambda\right) \text{ where } d_J^\Lambda = d^\Lambda(\boldsymbol{w}_J, \boldsymbol{x}_i) \text{ and } d_K^\Lambda = d^\Lambda(\boldsymbol{w}_K, \boldsymbol{x}_i) \quad (2)$$

correspond to the distances of feature vector $\boldsymbol{x}_i$ from the closest correct (wrong) prototype $\boldsymbol{w}_J$ ($\boldsymbol{w}_K$), respectively.

In Generalized Matrix Relevance LVQ, as introduced in [6], gradient based updates concern, both, the prototypes and the distance measure. The positive (semi-) definite matrix Λ in Eq. (1) is written as

$$\Lambda = \Omega^\top \Omega \quad \text{with } \Omega \in \mathbb{R}^{M \times N} \quad \text{and, hence, } \; d^\Lambda(\boldsymbol{w}, \boldsymbol{x}) = \left[\Omega\ (\boldsymbol{x} - \boldsymbol{w})\right]^2. \quad (3)$$

For $M < N$, the matrix Ω defines a linear transformation to a lower-dimensional space in which the (squared) Euclidean distance is evaluated, see [9] for details.

For the closest correct prototype $\boldsymbol{w}_J$ and closest wrong prototype $\boldsymbol{w}_K$ one obtains an update of the form

$$\boldsymbol{w}_J^{\text{new}} = \boldsymbol{w}_J + \alpha_1 \cdot \gamma^+ \cdot 2\Lambda(\boldsymbol{x} - \boldsymbol{w}_J) \quad \text{and} \quad \boldsymbol{w}_K^{\text{new}} = \boldsymbol{w}_K + \alpha_1 \cdot \gamma^- \cdot 2\Lambda(\boldsymbol{x} - \boldsymbol{w}_K) \quad (4)$$

$$\text{with } \; \gamma^+ = 2 \cdot d_K^\Lambda / \left(d_J^\Lambda + d_K^\Lambda\right)^2 \quad \text{and} \quad \gamma^- = -2 \cdot d_J^\Lambda / \left(d_J^\Lambda + d_K^\Lambda\right)^2. \quad (5)$$

Here, the index J (K) refers to the closest correct (wrong) prototype $\boldsymbol{w}_J$ ($\boldsymbol{w}_K$). As in many LVQ variants, prototypes are moved towards (away from) the current training data, if the labels of prototype and example agree (disagree), respectively. The corresponding matrix update reads

$$\Delta\Omega_{mn} = -\alpha_2 \cdot \left(\gamma^+ \cdot \frac{\partial d_J^\Lambda}{\partial \Omega_{mn}} + \gamma^- \cdot \frac{\partial d_K^\Lambda}{\partial \Omega_{mn}}\right). \quad (6)$$

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

In addition, the transformation matrix Ω is normalized after each learning step such that $\sum_i \Lambda_{ii} = \sum_{mn} \Omega_{mn}^2 = 1$. Note that the learning rates α_1 and α_2 can be chosen independently. In general, we set $\alpha_1 \gg \alpha_2$ which implies that changes of the metric occur on a much slower time scale than those of the prototypes, for further information see [10].

2.2 Large Margin Nearest Neighbor Algorithm

The k-Nearest Neighbor (kNN) rule is one of the most basic and simplest methods for classification. It labels a novel feature vector by a majority vote among its k nearest neighbors in the training set. Thus, its performance depends crucially on the metric used for the identification of neighbors. In [7], the Large Margin Nearest Neighbor (LMNN) algorithm is introduced. The aim of the training process is that the κ nearest neighbors of an example data belong to the same class with high probability. At the same time, different classes should be separated by a large margin. The corresponding optimization problem is convex and the global optimum can be found by means of semi-definite programming [7]. Note that the computational effort of the algorithm grows with the parameter κ.

The LMNN algorithm provides a discriminative distance measure for the k-NN classifier which can be written as $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left[\Psi(\boldsymbol{x}_i - \boldsymbol{x}_j)\right]^2$. Here, Ψ denotes an $M \times N$ matrix which is the counterpart of the matrix Ω obtained in GMLVQ, cf. Eq. (3). All results presented in the following were produced with the code made available at `www.weinbergerweb.net` [7] using default parameters.

2.3 Canonical Representation of the Transformations

The matrices Ω learned by GMLVQ and Ψ obtained by the LMNN algorithm are not uniquely determined: The distance measures are, e.g., invariant under rotations in feature space. We identify unique $\widehat{\Omega}$ and $\widehat{\Psi}$ by decomposing $\Lambda = \Omega^\top \Omega$ and $\Upsilon = \Psi^\top \Psi$ in a canonical way: We determine the normalized eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M$ corresponding to the M ordered non-zero eigenvalues of Λ or Υ, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$ and define $\widehat{\Omega}$ or $\widehat{\Psi}$ as:

$$\{\widehat{\Omega}|\widehat{\Psi}\} = \left(\left[\sqrt{\lambda_1}\boldsymbol{v}_1, \sqrt{\lambda_2}\boldsymbol{v}_2, \ldots, \sqrt{\lambda_M}\boldsymbol{v}_M\right]\right)^\top \in \mathbb{R}^{M \times N}. \tag{7}$$

While this does not alter the classifier and retrieval system, it allows for direct comparison of $\widehat{\Omega}$ and $\widehat{\Psi}$.

3 Experiments

Results presented here are based on a dataset of images provided by the Department of Dermatology at the University of Groningen. Currently, this data base contains 47621 images from 11361 patient sessions. A subset of 211 images was manually labeled by dermatologists, who assigned each image to one of four classes of lesions. We will refer to these as *brown (class 1), white (2), blue (3)* and *red (4)* with 54, 46, 29 and 82 samples respectively, see [5] for details.

The original features were manually extracted by taking the average color of a region of lesion and a region of healthy skin [5]. Three color components for each of the two regions result in six-dimensional original feature vectors which are z-transformed resulting in zero mean and unit variance features.

We represent the data by three-dimensional vectors which are obtained from the original data by means of a linear mapping. To this end, we first apply the GMLVQ(3×6) algorithm[9], as given by Eqs. (4, 6). In the training process we start matrix learning after $t_m = 50$ of, in total, 500 epochs and apply a learning rate schedule of the form $\alpha_{1,2}(t) = \alpha_{1,2}^{\text{start}} \big/ [1 + (t-1)\Delta\alpha_{1,2}]$. Here, t counts sweeps through the dataset, and α_1^{start} and α_2^{start} denote the initial rates for prototype and matrix updates. In all runs we set $\alpha_1^{\text{start}} = 10^{-2}$, $\Delta\alpha_1 = \Delta\alpha_2 = 10^{-4}$ and $\alpha_2^{\text{start}} = 10^{-3}$. Initial positions $\boldsymbol{w}_i(0)$ of the prototypes were determined by randomly selecting 1/3 of the feature vectors in class $c(\boldsymbol{w}_i)$ and taking the respective mean. Relevance initialization was done by generating independent random Ω_{ij} uniformly in $[-1, 1]$ and subsequent normalization.

The dataset D is divided in 10 disjoint subsets D_s, $s = 1 \ldots 10$, of approximately equal size which yields 10 training datasets $D_s^t = D/D_s$. For each set we compute the canonical representation $\widehat{\Omega}$ and average it over 10 initializations and subsequently over the 10 training sets. LMNN obtains a unique, global optimum of the cost function and when using the original features there is no training process involved. Thus, in both cases, the outcome is not influenced by initialization or the randomized training procedure.

4 Results

Figure 1 shows the retrieval rates vs. the number k of retrieved images. The accuracies quantify the percentage of images from the same class among the k retrieved data base images. Results are displayed for the GMLVQ(3×6) algorithm and for two versions of LMNN. They differ with respect to the number κ of neighbors taken into account in the training phase: In the first, κ equals the number k of retrieved images in the working phase, in the second $\kappa = 25$ for all k. In the latter case the retrieval performances of LMNN and GMLVQ are comparable, which is also reflected in the fact that the obtained matrices $\widehat{\Omega}$ and $\widehat{\Psi}$ are very similar, cf. Fig. 2. Using smaller κ reduces the computational cost of LMNN, but at the same time its performance deteriorates.

In general, the adaptation of the distance measure improves the retrieval performance significantly over the use of the original features or the difference features suggested in [5]. As can be seen in Fig. 2, the canonical transformations $\widehat{\Omega}$ obtained by GMLVQ and $\widehat{\Psi}$ from LMNN with $\kappa = 25$, on average over the ten fold training processes, are almost identical.

5 Conclusion and Outlook

Our results show that the performance of the retrieval system can be improved significantly by choosing an appropriate distance measure. The application of

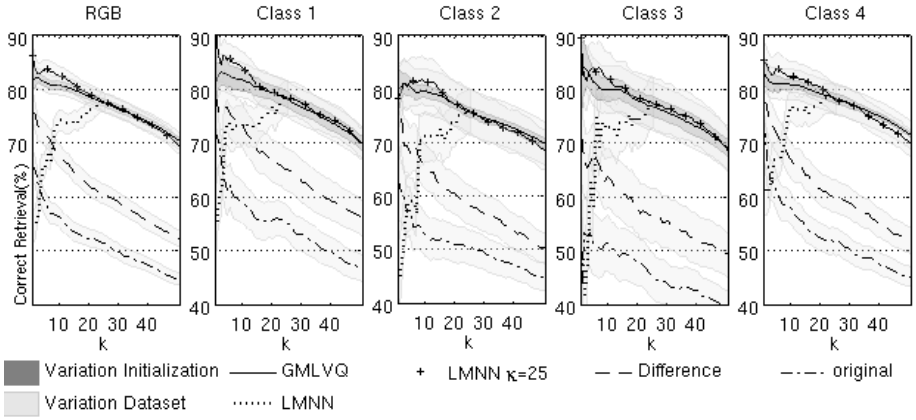

Fig. 1: The retrieval rates for colorspace RGB vs. the number k of retrieved images. The four rightmost figures show the class-specific retrieval rates and their variation with random initialization σ_{init} and with the dataset ϵ_{data}. LMNN is trained with, either, $\kappa = k$ (dotted lines) or with $\kappa = 25$ (dashed lines).

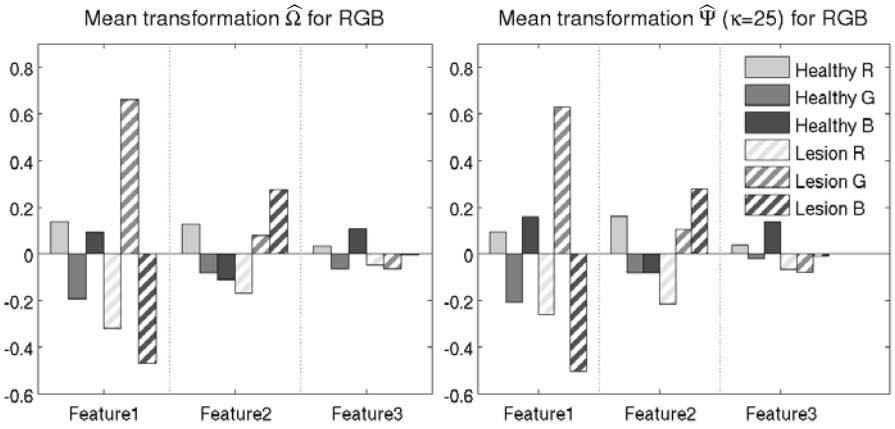

Fig. 2: Canonical transformations of RGB features on average over data sets and initialization. Left panel: Multipliers $\widehat{\Omega}$ from GMLVQ which define the new features as linear combinations of the six original features. Right panel: the same for $\widehat{\Psi}$ from LMNN.

LMNN seems natural, since the retrieval is also based on a kNN approach. However, our investigation shows that the GMLVQ approach outperforms LMNN if the latter takes only a relatively small number κ of neighbors into account in the training process. For larger κ the obtained metric becomes very similar to that of GMVLQ and, consequently, the retrieval performances are comparable. The computational effort for GMLVQ training is typically lower than that of the

LMNN optimization where it grows with κ. The use of large κ decreases the influence of single examples. The same effect is achieved in GMLVQ through the use of prototypes. An important advantage of the GMLVQ approach is its greater potential with respect to extensions. We will address, e.g., the use of local metrics defined in different areas of the features space.

We have concentrated here on the RGB color space as one example. Similar improvements of retrieval by metric adaptation are observed for representations like LCH, HSV, or XYZ, details will be published elsewhere. The use of color features only constitutes, of course, a significant restriction of this initial study. Obviously, other features like shape and texture should play an important role in lesion classification and retrieval. Therefore, we intend to apply the metrics adaptation schemes in CBIR using more general sets of features.

Acknowledgment This work was supported by the "Nederlandse organisatie voor Wetenschappelijke Onderzoek (NWO)" under project code 612.066.620.

References

[1] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. In *International Journal of Medical Informatics*, volume 73, pages 1–23, 2004.

[2] T. Lehmann, M. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. Wein. Content-based image retrieval in medical applications. In *Methods of Information in Medicine*, volume 43, pages 354–361, 2004.

[3] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40:1106–1122, 2007.

[4] Y. Cheng, R. Swamisai, S. E. Umbaugh, R. H. Moss, W.V. W. V. Stoecker, S. Teegala, and S. K. Srinivasan. Skin lesion classication using relative color features. In *Skin Research and Technology*, volume 14, pages 53–64, 2008.

[5] H. H. W. J. Bosman, N. Petkov, and M. Jonkman. Comparison of color representations for content based image retrieval of dermatological images. submitted, 2008.

[6] P. Schneider, M. Biehl, and B. Hammer. Relevance matrices in LVQ. In M. Verleysen, editor, *Proc. of European Symposium on Artifiacal Neural Networks (ESANN)*, pages 37–42, Bruges, Belgium, April 2007.

[7] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems (NIPS)*, volume 18, 2006.

[8] A. S. Sato and K. Yamada. Generalized learning vector quantization. In *Advances in Neural Information Processing Systems*, volume 8, pages 423–429, 1996.

[9] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. Discriminative visualization by limited rank matrix learning. Technical Report MLR-03-2008, 2008. http://www.uni-leipzig.de/∼compint/mlr/mlr_03_2008.pdf.

[10] B. Hammer and T. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.