Pattern Recognition 44 (2011) 1892–1902

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Learning effective color features for content based image retrieval in dermatology

Kerstin Bunte [a,*], Michael Biehl [a], Marcel F. Jonkman [b], Nicolai Petkov [a]

[a] Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands
[b] Department of Dermatology, University Medical Center Groningen, University of Groningen, The Netherlands

A R T I C L E I N F O

Available online 7 November 2010

Keywords:
Machine learning
Learning vector quantization
Adaptive distance measures
Content based image retrieval

A B S T R A C T

We investigate the extraction of effective color features for a content-based image retrieval (CBIR) application in dermatology. Effectiveness is measured by the rate of correct retrieval of images from four color classes of skin lesions. We employ and compare two different methods to learn favorable feature representations for this special application: limited rank matrix learning vector quantization (LiRaM LVQ) and a Large Margin Nearest Neighbor (LMNN) approach. Both methods use labeled training data and provide a discriminant linear transformation of the original features, potentially to a lower dimensional space. The extracted color features are used to retrieve images from a database by a k-nearest neighbor search. We perform a comparison of retrieval rates achieved with extracted and original features for eight different standard color spaces. We achieved significant improvements in every examined color space. The increase of the mean correct retrieval rate lies between 10% and 27% in the range of k=1–25 retrieved images, and the correct retrieval rate lies between 84% and 64%. We present explicit combinations of RGB and CIE-Lab color features corresponding to healthy and lesion skin. LiRaM LVQ and the computationally more expensive LMNN give comparable results for large values of the method parameter κ of LMNN ($\kappa \geq 25$) while LiRaM LVQ outperforms LMNN for smaller values of κ. We conclude that feature extraction by LiRaM LVQ leads to considerable improvement in color-based retrieval of dermatologic images.

1. Introduction

In the last decades the availability of digital images produced by scientific, educational, medical, industrial and other applications has increased dramatically. Thus, the management of the expanding visual information has become a challenging task. Since the 1990s Content Based Image Retrieval (CBIR) is a rapidly advancing research area, which uses visual content to search images from large databases according to the user's interest [36,23,21,12,22,26,41,17]. A typical CBIR system extracts visual information from an image and converts it internally to a multidimensional feature vector representation. For retrieval, the dissimilarities (distances) between the feature vector of a query image and the feature vectors of the images in the database are computed. Then, the database images most similar to the query are presented to the user. CBIR may especially be interesting in the field of computer aided diagnostics when it is partly based on images. An intelligent pre-selection of images with a trained system might help a medical doctor to efficiently search for patients, who had problems similar to the actual case.

The visual content of an image can be described by color, texture, shape or spatial relationship. A good visual content descriptor should be insensitive to the specific imaging process, e.g. invariant under changes of illumination. The prevalent visual content for image retrieval is color. Frequently used color descriptors are color moments, histograms, coherence vectors and correlograms [33,24]. Before a color descriptor can be selected, the underlying color space has to be specified.

There are many different color spaces available, which may be beneficial in different application domains. The color representations most commonly used in electronic systems are RGB and CIE-XYZ. CIE-XYZ and the related CIE-Lab and CIE-Luv are designed to match human perception. In [40] the authors argue, that normalized TSL (Tint, Saturation, Lightness) is superior to other color spaces for skin modeling with a unimodal Gaussian joint probability density function. The color space YCrCb is adjusted for efficient image compression, but the transformation simplicity and explicit separation of luminance and chrominance components appear attractive for skin color modeling [25,46,9]. Surveys on color spaces and their use can be found in [40,43]. We are not aware of a general rule for the choice of the color space and the representation might follow the users preference. So we decided to investigate eight different color spaces, which are commonly used and may be useful for the task at hand.

* Corresponding author. Tel.: +31 50 3637049; fax: +31 50 3633800.
E-mail addresses: k.bunte@rug.nl (K. Bunte), m.biehl@rug.nl (M. Biehl), m.f.jonkman@derm.azg.nl (M.F. Jonkman), n.petkov@rug.nl (N. Petkov).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.10.024

Color is an important attribute for primary skin efflorescences [3]. Color features have proven beneficial in many applications and medical sciences, especially for the recognition of skin lesions [14,40,43,38,34,19,37,25,46,18] or the classification of skin cancer [28,44,1,16,10,42]. A dermatologist might be interested in pictures of similar skin lesions in comparison to an actual case to verify the diagnosis or confer with similar symptoms. This can be interpreted as a problem of CBIR. The authors of [4] study the use of color features and the effectiveness of different color spaces in this context. They conclude that the representation of an image by the difference in the average color of healthy and lesion skin gives better results than the explicit use of the pair of colors. Fig. 1 shows two example retrievals for a CBIR system in the field of skin lesion comparison in dermatology. In [4], the best results were achieved with the CIE-Lab color representation.

Since the difference of two color values is a special case of a linear transformation, the question arises whether better results can be achieved by more general linear transformations. Of course, it is possible that the use of a combination of a cyclic distance measure in the case of color spaces containing a "hue"-descriptor might lead to superior results. We will address this interesting questions in further studies. One well known technique to achieve a linear projection of feature vectors to a subspace which minimizes the overlap between different classes is Linear Discriminant Analysis (LDA) [13]. In this paper we employed and compared two different recent techniques, which are able to find discriminant feature transformations based on a supervised training procedure. The Large Margin Nearest Neighbor (LMNN) [45] approach has the advantage that it is based on a convex cost function, so it returns the global optimum for the current configuration of training data and parameters, based on the kNN approach. The Limited Rank Matrix Learning Vector Quantization (LiRaM LVQ) [29,30,8,32] on the other hand follows a stochastic gradient descent procedure and may get stuck in local minima. On the other hand, it has the advantage of low computational costs. It is a prototype-based method, in which the decision boundary is defined by the Voronoi cells of prototypes following the large margin principle [11]. Both algorithms are available in general form and turned out to be effective classifiers in many applications. In our real world example application of CBIR in dermatology, the LiRaM LVQ approach turned out to be quite robust concerning the initialization and parameter setting. With comparably low computational costs it leads to similar or better results than the LMNN approach with optimal parameter setting on most color spaces discovered. We improve the correct retrieval rate in CBIR of dermatological images significantly by applying adaptive linear transformations.

The main aim of this work is to demonstrate in terms of a real world example, that an adaptive, i.e. data driven transformation of original color features can improve the retrieval performance of a CBIR system significantly. We concentrate on the performance enhancement achieved by using the most basic, easy and fast acquirable set of important features for the problem at hand, i.e. color information only.

In Section 2 we explain the real world data set, the feature extraction process, we present and discuss the methods we use to determine optimal transformations of color features and their use in the CBIR system. In Section 3 we present results and conclude in Section 4.

2. Methodology

An illustration of the Methodology is shown in Fig. 2.

2.1. Data set and feature extraction

We analyze images from a database maintained at the Department of Dermatology of the University of Groningen. At the time of this study it consisted of 47,621 images from 11,361 patient sessions, the number of images grows by about 5000 per year. Clinical images are obtained under standard light conditions and do not require further calibration. A subset of 211 images was provided and manually labeled by a dermatologist, who assigned each image to one of four classes of lesions. For better readability we refer to these classes as "red", "white", "blue" and "brown" (Fig. 3). These terms correspond to the relative tint of lesions which appear reddish, blue, brownish or hypopigmented on the background of the surrounding healthy skin. We consider a data set with 82, 46, 29 and 54 samples, respectively, which amounts to a total of 211 images.

Of course there are more characteristics then just color which identify the kind of skin lesion, e.g. the shape. The consideration of other types of features will be addressed in future work, here we concentrate on the quality the most basic set of features is able to achieve. In this particular problem color seems to be a suitable indicator for the skin lesion classes. The complete data set also contains other skin lesions, but in this study we restrict ourselves to the consideration of the above mentioned classes. Here, emphasis is not on the classification performance itself. It serves as a basis for improving the retrieval system and the supervised training yields a suitable distance measure. Further studies should address additional features, more general skin lesion classes and the handling of unknown classes.

The original images were not pre-processed. For each image a region of lesion and a region of healthy skin are manually selected and for each of them the average color values are computed (see Fig. 4). Hence, the extracted data contains three color components for each of the two regions, resulting in a six-dimensional (6D) feature vector. As a normalization step we perform a z-score-transformation resulting in zero mean and unit variance features. This normalization is reasonable in the RGB color space and linear domains. In case of cyclic descriptors, like the "hue", this might not be appropriate. The combination of cyclic distances and linear dissimilarities and their normalization concerning this specific task will be addressed in future studies. Nevertheless, for the sake of comparison and completeness we show the results on different color spaces under the same conditions.

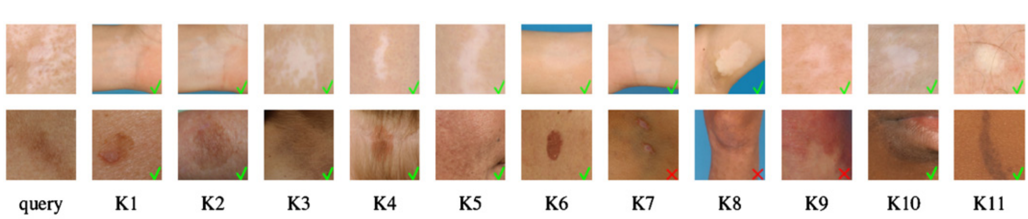

Fig. 1. Two example retrievals of the 11 most similar images for a given query image. The first image in a row is the query image, followed by the images returned from the retrieval system [4]. The green tick marks images with the same class label like the query.