

Color-Based Image Retrieval from High-Similarity Image Databases

Michael Edberg Hansen and Jens Michael Carstensen

Image Processing and Computer Vision
Informatics and Mathematical Modelling
DK-2800 Lyngby, Denmark
{meh, jmc}@imm.dtu.dk
<http://www.imm.dtu.dk/image/>

Abstract. Many image classification problems can fruitfully be thought of as image retrieval in a "high similarity image database" (HSID) characterized by being tuned towards a specific application and having a high degree of visual similarity between entries that should be distinguished. We introduce a method for HSID retrieval using a similarity measure based on a linear combination of Jeffreys-Matusita (JM) distances between distributions of color (and color derivatives) estimated from a set of automatically extracted image regions. The weight coefficients are estimated based on optimal retrieval performance. Experimental results on the difficult task of visually identifying clones of fungal colonies grown in a petri dish and categorization of pelts show a high retrieval accuracy of the method when combined with standardized sample preparation and image acquisition.

1 Introduction

Ever since the early 1970's the rapid increase in the number and size of digital collections, for describing and storing video and image data, has stressed the importance of efficient and fast methods for searching and querying. Up until today, different solutions to these problems have been proposed, directed by specific needs offered by specific applications [1, 3, 4, 9–11]. They span over a large variety of domains, from being specific to the more general (e.g fingerprints and photographs). While the solutions most often address the problem of retrieving images from high-diversity databases, this paper will draw the attention towards databases with entries having low degree of diversity. We categorize the image databases in two main categories: "Low Similarity Image Databases" (LSID) and "High Similarity Image Databases" (HSID), focusing on the latter.

The problem of retrieving images from a HSID will be illustrated by two examples: Identification and retrieval of 1) fungal colonies, and 2) Scanbrown mink pelts. In both cases acquisition procedures has been applied [8] fulfilling the requirements described in section 2.

2 Density-based image retrieval

Working successfully with HSID's means having more restrictions put on the acquisitions steps involved. Retrieving entries based on subtle differences, require the control of all influencing parameters involved in the acquisition process to reduce unnecessary variation that could affect the retrieval performance. The *sample preparation* and *presentation* procedures has to be standardized and well-defined. *Image acquisition* is standardized by color and geometric calibrations, and by using an optimal illumination geometry. By *region extraction* computations are focused to consistently defined regions within the image, from which *features* are extracted. For most HSID's *a priori* knowledge has to be used in this process. By proper choice of region extraction it is possible to describe some of the spatial relations. Finally, the *region-by-region similarity* is compared across all images. We propose a method in which we make a classification based on distributions of the observations extracted from these regions.

It is important to notice, that the standardizations applied in the acquisition procedures will affect the choice of features in such a way that a good standardization will simplify the features necessary.

2.1 Density estimation

Given N sample images containing K consistently defined regions. Each region, k , contain M_k observations based on L features, $\mathbf{x}_{m,k} = \{x_1, \dots, x_L\}$, describing a certain property on which segregation can be made. From each region, k , the output of the feature extraction is a matrix of observations, $\mathbf{X}_k \in \mathbb{R}^{M \times L}$. To illustrate the use of indices (see Figure 1), N could be the number of images in a HSID database, K a number of consistently defined regions, M_k the number of pixels within each region, k , and finally, $x_{m,k}$ would be a vector of L pixel-specific values, e.g. red, green and blue ($L = 3$).

In the simple case we have $M = 1$ observations within each region, and \mathbf{X}_k is a vector of K features. If $M > 1$ we seek a compact representation. This is done by estimating the feature distributions for all regions, either by a parametric, nonparametric or combined approach.

2.2 Similarity estimation

The difference or (dis-)similarity between two images, i and j , evaluated by one of the consistently defined regions, k , can be expressed as a distance, $d_{ij,k}$, calculated between the feature distributions, $\mathbf{X}_{i,k}$ and $\mathbf{X}_{j,k}$ (see Figure 1 for illustration). The distances between all regions can then be summarized by the vector, $\mathbf{d}_{ij} = (d_{ij,1}, \dots, d_{ij,K})^t$, representing the overall similarity between all corresponding regions (low distance = high similarity). If a database contain many entries the parametric approach could serve as a rough "model", returning a list of suitable candidates, and the nonparametric approach for the final evaluation. If \mathbf{d}_{qi} is the distance vector between features extracted from a query

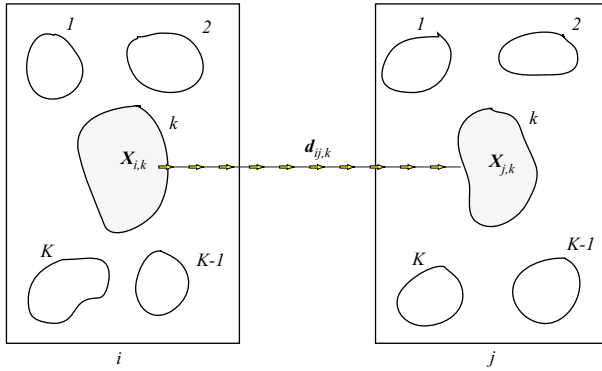


Fig. 1. Illustration of the principle of the density estimation. The similarity between two images, i and j , evaluated by one of the consistently defined regions, k , can be expressed as a distance, $d_{ij,k}$, calculated between the feature distributions, $\mathbf{X}_{i,k}$ and $\mathbf{X}_{j,k}$.

image, q , and the i 'th entry in the database, the problem of retrieving images is now reduced to compare the vectors

$$\mathbf{d}_{qi}, i = 1, \dots, N \tag{1}$$

which may be condensed to one overall similarity by

$$g : \mathbf{d}_{qi} \mapsto d_{qi} \tag{2}$$

where, g , can be any mapping suitable (e.g. convex linear combination, the p -norm, $\|\cdot\|_p$, etc.). The performance of the retrieval system depends on the feature representation and the matching scheme employed. In general, g is a function condensing significant information contained in \mathbf{d}_{qi} , and a natural choice is the convex linear combination

$$d_{qi} = g(\mathbf{d}_{qi}; \mathbf{w}) = \mathbf{w}^t \mathbf{d}_{qi} = \sum_k w_k d_{qi,k} \tag{3}$$

where $0 \leq w_k \leq 1$. \mathbf{w} may be chosen in such a way, that regions contributing with generally more information should have higher weights than those contributing with less. Equation 3 may, if necessary, be extended to a full metric.

3 Similarity evaluation

In literature different distance metrics, has been proposed [1]. The Jeffreys-Matusita distance (JM-distance) has shown to provide a reliable criteria, as measure of class separability [2]. For two feature distributions $p_i(\mathbf{x})$ and $p_j(\mathbf{x})$, the JM-distance, J_{ij} , is given by

$$J_{ij} = \left[\int_{\mathbf{x}} \left(\sqrt{p_i(\mathbf{x})} - \sqrt{p_j(\mathbf{x})} \right)^2 d\mathbf{x} \right]^{1/2} \tag{4}$$

and may be rewritten as

$$J_{ij} = \sqrt{2(1 - e^{-\alpha_{ij}})} \quad J_{ij} \in [0; \sqrt{2}] \quad (5)$$

where α_{ij} is the Bhattacharyya distance [5] computed as

$$\alpha_{ij}^2 = \frac{1}{8} (\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}_i)^t \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}_i) + \frac{1}{2} \ln \frac{|\hat{\boldsymbol{\Sigma}}|}{\sqrt{|\hat{\boldsymbol{\Sigma}}_i| |\hat{\boldsymbol{\Sigma}}_j|}}, \quad \hat{\boldsymbol{\Sigma}} = \frac{\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j}{2} \quad (6)$$

when assuming normality. Using the JM-distance in this context, the distance vector between regions is

$$\mathbf{d}_{ij} = \mathbf{J}_{ij} \quad (7)$$

To retrieve an image we need a scheme for selection, based on similarities. One option could be to use the k -nearest neighbor selection criteria [7].

4 Region-weight estimation

From Equation 3, the essential question is how to chose the weights, \mathbf{w} . We propose a method to estimate $\hat{\mathbf{w}}$, based on optimal retrieval performance. The optimization is done based on a training set (of M images) with *a priori* knowledge about common relations. First, we define the *class correspondence matrix* as $\mathbf{C} \in \mathbb{R}^{M \times M}$, explaining the correspondence between each combination of pairs of images (i, j) in the database, where

$$c_{ij} = \begin{cases} 1 & \text{if image } i = \text{image } j \\ -1 & \text{if image } i \neq \text{image } j \end{cases} \quad (8)$$

In the case of optimal retrieval performance, validation would give us a classification scheme with relations given by \mathbf{C} , and this matrix describe the *ground truth*. Next, the distance between each of the entries in the database is estimated, and the resulting matrix is then given by $\mathbf{D} = \{\mathbf{d}_{ij}\}$. We assume, that \mathbf{C} and \mathbf{D} are symmetric. Finally, the objective is to find the set of weights, $\hat{\mathbf{w}}$, that optimizes

$$\hat{\mathbf{w}} = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_k)^t = \arg \max_{\mathbf{w}} [\Phi(\mathbf{D}, \mathbf{C}; \mathbf{w})] \quad (9)$$

where Φ is the performance evaluation function. This function could be anything from retrieval percentage to an overall difference in distances between equal images. Common for all is that the estimation of $\hat{\mathbf{w}}$ is based on the ground truth \mathbf{C} and the weighted similarity between entries, \mathbf{D} . A proper choice of Φ is the function

$$\Phi(\mathbf{D}, \mathbf{C}; \mathbf{w}) = \frac{-2}{M(M-1)} \sum_{i=1}^M \sum_{j<i}^M c_{ij} \mathbf{w}^t \mathbf{d}_{ij} \quad (10)$$

that maximizes the average distance between groups of similar and non-similar entries.

5 Experimental section

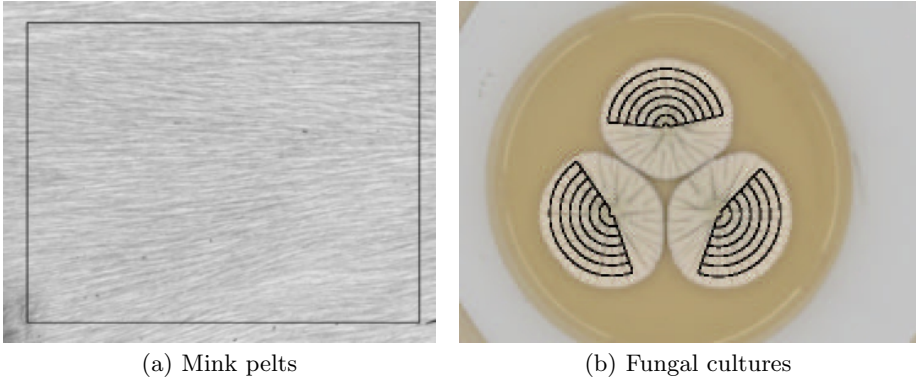


Fig. 2. Domains of images to be indexed. 2(a): Only one region, $K = 1$, is used. 2(b): For $K = 1$ the whole area is used ($\hat{w} = w_1 = 1.0$), and for $K > 1$ the distance from the center to the area is increasing for higher indices.

5.1 Query of mink pelt images

700 mink pelts of the type Scanbrown were scanned [8]. Seven classes are given, characterized by seven different nuances of brown: xxdk, xdk, dk, medium, pale, xpale, and xxpale - 100 pelts from each class. The pelts were selected to capture the full variation within each class. The goal is to measure the "pelts color" defined by the industry and maintained by a set of highly trained and calibrated human sorters. The procedure of educating human sorters may take up to three years, so having an automated and objective system for scoring pelts is appreciable. "Pelts color" should ideally be uncorrelated with hair length, hair density, hair direction, gloss and other disturbances.

After acquisition, each of the pelt images were analyzed based on the RGB channels. One single region ($K=1$) was used (see Figure 2(a)), and from this the feature distribution were estimated as described in Section 3.

In Figure 3 the retrieval performance based on a leave-one-out experiment using the k -nearest neighbor as selection criteria is plotted. Here we see, that the retrieval performance of the HSID is above 96% in the best cases (xxpale) and approximately 75% in the most difficult ones (xxdk and xdk). If the number of neighbors used in the query is increased we observe an increase in overall performance (see Figure 3(b)).

5.2 Query of fungal images

A last example will be taken from the world of microbes. Here *Penicillium commune* [13, 12] is the most frequently occurring spoilage fungi on cheese. In work-

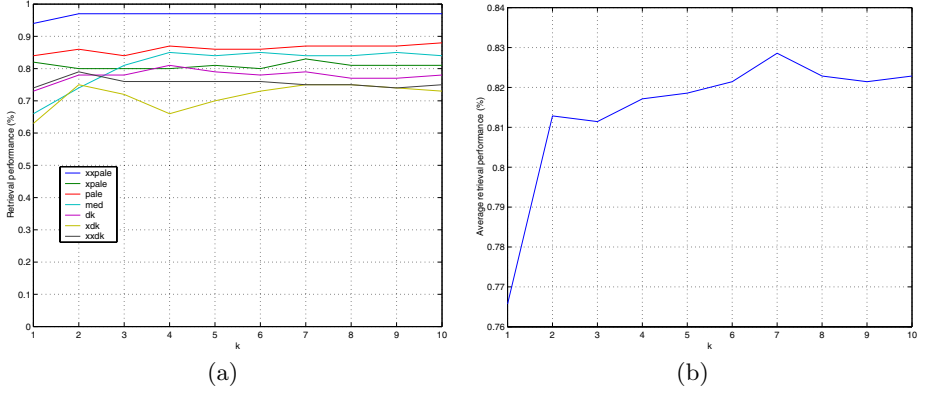


Fig. 3. Plots of the retrieval performance as function of the k -nearest neighbour, for each of the different nuances 3(a) and the average 3(b).

ing with these isolates, it was noticed that different isolates on identification media had slightly different appearance. It was the aim to design a content based HSID retrieval (fungal identification) system, based on visual features. Figure 2(b) show the result after pre-processing of the acquired images of the cultures. The images were segmented into a number of concentric regions¹, and features were extracted describing both color and texture within these regions [6]. In this context, texture is defined as the radial and angular "rippled-ness" according to a center at the inoculation point. Based on the gradients

$$\nabla I(\mathbf{x})|_{circular} = \mathbf{\Gamma} \cdot \nabla I(\mathbf{x})|_{cartesian} \quad (11)$$

where

$$\nabla I(\mathbf{x})|_{cartesian} = \begin{bmatrix} \frac{\delta I}{\delta x_1} \\ \frac{\delta I}{\delta x_2} \end{bmatrix} \left(= \frac{\delta I}{\delta x_1} \hat{x}_1 + \frac{\delta I}{\delta x_2} \hat{x}_2 \right) \quad (12)$$

is the ordinary gradient operator in rectangular coordinates and $\mathbf{\Gamma}$ is the *rotation matrix*

$$\mathbf{\Gamma} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (13)$$

in these directions, we now have a set of additional features for each pixel giving us a distribution of features for each region.

From each of the regions, the features were extracted, the mean and dispersion were estimated and the distances between the distributions estimated.

Table 1 (left) show the retrieval performance based on a leave-one-out experiment using the 1-nearest neighbor as selection criteria. First the average distance between regions was considered, according to Equation 3. Second, weights were

¹ The regions near the center of the colony is older than those closer to the edge.

optimized using Equation 10. We see, that the method has high performance, increasing with the number of regions. Though it seems, that we have a drop in performance, when using the average distances of 6 regions, we notice that the tendency gets to be the same when optimized weights. These weights are listed

Regions, K	Mean		Optimized		Weights						
	Φ	RR	Φ	RR	w_1	w_2	w_3	w_4	w_5	w_6	
1	1.12	94.5%	1.13	94.5%	1.00	—	—	—	—	—	
2	1.17	98.2%	1.18	98.2%	0.71	0.29	—	—	—	—	
3	1.19	99.1%	1.20	99.1%	0.28	0.54	0.18	—	—	—	
6	1.22	98.2%	1.23	99.1%	0.00	0.38	0.38	0.00	0.13	0.13	

Table 1. Left: Table showing the difference between the JM-distance for identical and different images using and optimized weights on a different number of regions, and the retrieval rate. Right: Table showing optimized weights achieved by the optimization procedure.

in Table 1 (right). We conclude, that some of the regions has a larger influence on the result than others.

6 Discussion

We have introduced the term High (and Low) Similarity Image Databases (LSIDs and HSIDs) as a basic concept for handling different image classification problems. We have presented a method for retrieving similar images from HSID’s using a similarity metric based on a linear combination of Jeffreys-Matusita (JM) distances, computed on a set of features extracted from consistently defined image regions. Weight coefficients can be estimated based on optimizing the retrieval performance.

The method has proven to have a high performance on both sets of images acquired of fungal colonies and Scanbrown pelts. These results are important, since the retrieval rates are significantly higher in the case of fungal identification, and almost as high in the case of scoring pelts, compared with the performance of human observers. Hereby it is made possible to accumulate and retrieve important information about fungal and mink pelt color expression in a HSID.

The JM distance between a set of distributions, is based on the assumption normal distributed observations. An improvement of the method could be to calculate the JM-distance, based on a non-parametric density estimation and use the normal assumption as a first guess to pick out relevant candidates.

7 Acknowledgments

This project was supported by the Danish Technical Research Council under the project ”Programme for predictive biotechnology: Functional biodiversity

in *Penicillium* and *Aspergillus*' (grant no. 9901295). The authors thank Dr. Flemming Lund for kindly providing fungal cultures and Videometer A/S and Copenhagen Fur Center for providing pelt images.

References

1. D. Androustos, K.N. Plataniotis, and A.N. Venetsanopoulos. Distance measures for color image retrieval. *IEEE*, 1998.
2. C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press - Oxford, 1995. ISBN: 0-19-853864-2.
3. L. Cinque, S. Levialdi, K.A. Olsen, and A. Pellicanó. Color-based image retrieval using spatial-chromatic histograms. *IEEE*, 1999.
4. Y. Deng, B.S. Manjunath, C. Kenney, M. S. Moore, and H. Shin. An efficient color representation for image retrieval. *IEEE Transactions on Image Processing*, 10(1):140–147, 2001.
5. K. Fukunaga. *Introduction to Pattern Recognition*. Academic Press, Inc., 2nd edition, 1990.
6. M. E. Hansen, F. Lund, and J. M. Carstensen. Visual clone identification of *penicillium commune* isolates. *Journal of Microbiological Methods*, 52(2):221–229, february 2003.
7. T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning; datamining, inference and prediction*. Springer, 2002.
8. <http://www.videometer.com>.
9. A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.
10. M. S. Kankanhalli, B. M. Mehtre, and H. Y. Huang. Color and spatial feature for content-based image retrieval. *Pattern Recognition Letters*, (20):109–118, 1999.
11. M. Mirmehdi and M. Petrou. Segmentation of color textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):142–159, February 2000.
12. J.I. Pitt. *The genus penicillium and its teleomorphic states eupenicillium and talaromyces*. Academic Press, London, 1979.
13. K.B. Raper and C. Thom. *Manual of the penicillia*. Williams and Wilkins, Baltimore, 1949.