Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Signal Processing 85 (2005) 233–246

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Compact colour descriptors for colour-based image retrieval

Linh Viet Tran, Reiner Lenz*

Department of Science and Technology, Bredgatan, Linköping University, SE 601-74 Norrköping, Sweden

Received 1 August 2002; received in revised form 1 July 2004

Abstract

In many colour-based image retrieval systems the colour properties of an image are described by its colour histogram. Histogram-based search is, however, often inefficient for large histogram sizes. Therefore we introduce several new, Karhunen–Loève transform (KLT)-based methods that provide efficient representations of colour histograms and differences between two colour histograms. The methods are based on the following two observations; Ordinary KLT considers colour histograms as signals and uses the Euclidian distance for optimization; KLT with generalized colour distance measures that take into account both the statistical properties of the image database and the properties of the underlying colour space should improve the retrieval performance. Image retrieval applications compare similarities between different images. Relevant for the decision is only the local structure of the image space around the current query image since the task is to find those images in the database that are most similar to this given query image. Therefore only the local topology of the feature space is of interest and compression methods should preserve this local topology as much as possible. It is therefore more important to have a good representation of the differences between features of similar images than good representations of the features of the images themselves. The optimization should therefore be based on minimizing the approximation error in the space of local histogram differences instead of the space of colour histograms. In this paper we report the results of our experiments that are done on three image databases containing more than 130,000 images. Both objective and subjective ground truth queries are used in order to evaluate the proposed methods and to compare them with other existing methods. The results from our experiments show that compression methods based on a combination of the two observations described above provide new, powerful and efficient retrieval algorithms for colour-based image retrieval.

Keywords: Colour-based image retrieval; Karhunen–Loève transform; Colour distribution; Local histogram differences

1. Introduction

Colour has been widely used for content-based image retrieval, multimedia information systems and digital libraries. In many colour-based image retrieval (CBIR) applications, the colour

*Corresponding author. Tel.: +46 11 36 32 78; fax: +46 11 36 32 70.

E-mail addresses: linh.v.tran@ericsson.com (L.V. Tran), reile@itn.liu.se (R. Lenz).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2004.10.001

properties of an image are characterized by the probability distribution of the colours in the image. The colour histogram remains the most popular representation of colour distributions since it is insensitive to small object distortions and easy to compute. However, it is not very efficient due to its large memory requirement. For typical applications a colour histogram might consist of $N = 512$ bins. With such a large number of bins N (i.e. $N \geqslant 20$), the performance of current indexing techniques is reduced to a sequential scanning [15,20]. To make colour histogram-based image retrieval truly scalable to large image databases it is desirable to reduce the number of parameters needed to describe the histogram while still preserving the retrieval performance. Approaches to deal with these problems include the usage of coarser histograms [9,12], dominant colours or signature colours [2,5,14] and application of signal processing compression techniques such as the Karhunen–Loève transform, discrete cosine transform, Hadamard transform, Haar transform, and wavelets, etc. [1,3,7,8,11]. Some of them are also suggested in the context of the MPEG-7 standard [8]. The optimal way to map N-dimensional vectors to lower K-dimensional vectors ($K \ll N$) is the Karhunen–Loève transform (KLT) [6]. The KLT is optimal in the sense that it minimizes the mean squared error of the Euclidian distance between the original and the approximated vectors. However, a straightforward application of the KLT (as well as other transform-based signal processing compression techniques) to the space of colour histograms gives poor retrieval performance. In this paper we will modify the conventional KLT approach in the following two points:

(1) Traditional KLT is derived from a minimum-mean-squared-error (MMSE) criterion which requires the squared error between the original and the approximated vectors to be minimal in the mean. This assumes implicitly that the vectors to be approximated (in our case the histograms) are elements in a Euclidian vector space with the Euclidian distance measure. In the case of colour histograms this assumption is not valid since the elements in neighbouring (in colour sense) bins represent similar colours whereas elements in distant bins represent very different colours. It is thus natural to view histograms as points in a space with a non-euclidian (colour-based) distance measure. One goal of this paper is thus to investigate how colour-based distances can be incorporated into the search.
(2) The other main observation used in this paper is that KLT is mainly based on the idea to find the best representation of the vectors. In image retrieval the representation of the histograms is, however, of minor importance. The main goal in image retrieval is to find good descriptors for the *difference* of two vectors. Furthermore, only similar histograms have to be compared since distances between very different images are not of interest in retrieval. We therefore develop a method that applies the KLT to histogram differences between neighbouring images.

We investigated methods based on these two ideas and a combination of both and evaluated their performance with the help of a database of 1000 images from the Corel database, the MPEG-7 database with 5466 images and a large database consisting of 126,604 low-resolution images from a commercial database. We describe the results of extensive evaluation experiments from which we draw the conclusion that a combination of the difference-based KLT compression and the colour-based distance measure in histogram space give better retrieval performance than comparable retrieval methods. The paper is organized as follows: Our proposed methods are presented in the next section; Section 3 describes our experiments in which both objective and subjective ground truth queries are used to evaluate our methods and to compare them with other existing methods; finally, some concluding remarks are given in Section 4.

2. Compact descriptors for colour-based image retrieval

In colour-based image retrieval we want to find all images I which have similar colour properties as a given query image Q. In this paper we describe the colour properties of images by their colour histograms and we define the similarity between images as the similarity between their colour histograms. If the colour histograms of the images I and Q are given by h_I and h_Q we represent the two images I and Q by two points h_I and h_Q in the colour histogram space $\mathbb{P}$ and define the distance between the images as the distance between the two points h_I and h_Q in $\mathbb{P}$. Popular choices for computing the distances in the colour histogram space are histogram intersection [18], L_p norm, quadratic forms [7,11], the earth mover distance (EMD) [14], and other statistical distance measures [13,15,16]. The EMD and the quadratic form methods are of special interest since they take into account the properties of the colour space and the underlying colour distributions. The EMD is computational demanding. Basically it computes the minimal cost to transform one histogram into the other. An optimization problem has to be solved for each distance calculation which makes the EMD less attractive in terms of computational speed. The quadratic form distance between colour histograms is defined as

$$d_M^2(h_1, h_2) = (h_1 - h_2)^{\mathrm{T}} M (h_1 - h_2), \tag{1}$$

where $M = [m_{ij}]$ is a positive semi-definite matrix defining the properties of the colour space. Each entry m_{ij} captures the perceptual similarity between colours represented by bins i and j. The naive implementation of the quadratic form-based metric is computationally demanding with a complexity of computing one distance of $\mathrm{O}(N^2)$ where N is the number of bins. Efficient implementations are, however, as fast as simple bin-by-bin distance methods such as histogram intersection or the L_p norm. It has also been reported that these metrics provide more desirable results than bin-by-bin distance methods [7], and comparable results to the more complicated EMD method [13]. The quadratic form-based distances are thus often used as distance measure in colour-based image retrieval. Using the full histogram to compute the distances in Eq. (1) is unrealistic for large image databases because of computational and storage demands. Methods for estimating the distances using fewer parameters are needed in order to speed up the search engine and to minimize storage requirements. Thus compression techniques should be used to compress the description of colour histograms. Often traditional statistical methods are applied to choose the com pression method but in the following we will see that colour-based image retrieval performance can be improved by selection of different compression strategies. In the following we consider a histogram h as a vector in N-dimensional space. Selecting N basis functions $\varphi_k, (k = 1, \ldots, N)$ we can describe h by K numbers x_k as follows:

$$h \approx \tilde{h}_K = \sum_{k=1}^{K} x_k \varphi_k. \tag{2}$$

The approximation error is given by

$$\begin{aligned} \varepsilon_K(h) = h - \tilde{h}_K &= h - \sum_{k=1}^{K} x_k \varphi_k \\ &= \sum_{k=K+1}^{N} x_k \varphi_k. \end{aligned} \tag{3}$$

Ordinary KLT in the histogram space $\mathbb{P}$ selects the basis functions φ_k such that the mean squared error in the Euclidian norm, $\bar{\varepsilon}_E^2$, is minimized:

$$\bar{\varepsilon}_E^2 = E\{||\varepsilon_K(h)^2||\} = E\{\varepsilon_K(h)^{\mathrm{T}} \varepsilon_K(h)\}. \tag{4}$$

Instead of using the Euclidian distance, a colour-based distance can be used where relations between different regions in colour space are taken into account. This results in a better correspondence to human perception. The basis functions φ_k are then selected such that the mean squared error with respect to the colour-based distances, $\bar{\varepsilon}_M^2$, is minimized:

$$\bar{\varepsilon}_M^2 = E\{||\varepsilon_K(h)||_M^2\} = E\{\varepsilon_K(h)^{\mathrm{T}} M \varepsilon_K(h)\}. \tag{5}$$

The computation of the coefficients and the basis functions in this new metric is done by using the following modified scalar product:

$$\langle h_1, h_2 \rangle_M = h_1^{\mathrm{T}} M h_2 = h_1^{\mathrm{T}} U^{\mathrm{T}} U h_2 = (U h_1)^{\mathrm{T}} (U h_2), \tag{6}$$

where U is an invertible matrix so that $M = U^{\mathrm{T}}U$. A given histogram can now be approximated by K numbers using the orthonormal basis $\{\varphi_k\}$ with respect to the new scalar product: $\langle\varphi_i, \varphi_j\rangle_M = \delta_{ij}$.

$$h \approx \tilde{h} = \sum_{k=1}^{K} \langle h, \varphi_k\rangle_M \varphi_k = \sum_{k=1}^{K} f_k \varphi_k. \tag{7}$$

Once the basis vectors φ_k are given, the coefficients f_k in the Eq. (7) are computed by:

$$f_k = \langle h, \varphi_k\rangle_M = h^{\mathrm{T}} M \varphi_k. \tag{8}$$

The new basis functions φ_k can be found by imitating the construction for the Euclidean case. The squared norm of the approximation of a histogram h is given by

$$\begin{aligned} ||\tilde{h}||_M^2 &= \langle\tilde{h}, \tilde{h}\rangle_M \\ &= \left\langle \left(\sum_{l=1}^{K} \langle h, \varphi_l\rangle_M \varphi_l\right), \left(\sum_{k=1}^{K} \langle h, \varphi_k\rangle_M \varphi_k\right) \right\rangle_M \\ &= \sum_{k=1}^{K} \langle\varphi_k, h\rangle_M \langle h, \varphi_k\rangle_M \\ &= (U\varphi_k)^{\mathrm{T}} U h h^{\mathrm{T}} U^{\mathrm{T}} (U\varphi_k). \end{aligned} \tag{9}$$

Computing the mean length and using the notation $\Sigma_M = E(Uhh^{\mathrm{T}}U^{\mathrm{T}})$ we see that the basis vectors with the smallest approximation error can be found by solving the Euclidean eigenvector problem $\Sigma_M \psi_k = c_k \psi_k$ as in the standard KLT method. From them the basis vectors are computed as $\varphi_k = U\psi_k$. Ordinary KLT technique is a special case where the relations between colour bins is ignored (M = identity). When the correlations between the input images in the database are ignored ($E\{hh^{\mathrm{T}}\}$ = identity) the solution is identical to the QBIC approach in [7]. Given two colour images I, and Q their histograms can be approximated by using only K coefficients as follows:

$$\begin{aligned} \tilde{h}_I &= \sum_{k=1}^{K} \langle h_I, \varphi_k\rangle_M \varphi_k = \sum_{k=1}^{K} f_k^I \varphi_k, \\ \tilde{h}_Q &= \sum_{k=1}^{K} \langle h_Q, \varphi_k\rangle_M \varphi_k = \sum_{k=1}^{K} f_k^Q \varphi_k. \end{aligned} \tag{10}$$

The distance between the two histograms is

$$\begin{aligned} d_M^2(I,Q) &= (h_I - h_Q)^{\mathrm{T}} M (h_I - h_Q) \\ &= ||h_I - h_Q||_M^2 \approx ||\tilde{h}_I - \tilde{h}_Q||_M^2 \\ &= \langle\tilde{h}_I - \tilde{h}_Q, \tilde{h}_I - \tilde{h}_Q\rangle_M \\ &= ||\tilde{h}_I||_M^2 + ||\tilde{h}_Q||_M^2 \\ &\quad - 2\sum_{k=1}^{K} \langle\tilde{h}_I, \varphi_k\rangle_M \langle\tilde{h}_Q, \varphi_k\rangle_M \\ &= \sum_{k=1}^{K} (f_k^I)^2 + \sum_{k=1}^{K} (f_k^Q)^2 - 2\sum_{k=1}^{K} f_k^I \cdot f_k^Q. \end{aligned} \tag{11}$$

The first term can be pre-computed for all images in the database, the second term is computed only once for the query image. The distance computation in the retrieval phase involves therefore only K multiplications and additions. Quadratic form-based distances have been used in colour-based image retrieval for a long time and we mention here the following selections of the matrices M: In [7] the matrix M has entries defined as

$$m_{ij} = 1 - d_{ij}/d_{\max}. \tag{12}$$

Here d_{ij} is the Euclidean distance between colour i and j in the CIELAB colour space and $d_{\max} = \max\{d_{ij}\}$. The CIELAB colour space is used since its metrical properties are well adapted to human colour difference judgments. The quadratic form distance using metric M as in Eq. (12) tends to overestimate the mutual similarity of colour distributions [14,16,17]. Several suggestions have been made to reduce the mutual similarity of dissimilar colours. One example is

$$m_{ij} = \exp(-\sigma(d_{ij}/d_{\max})^k) \tag{13}$$

described in [7]. It enforces a faster roll-off as a function of d_{ij}, the distance between colour bins. Another method uses a threshold for similar colours so that only colours which are similar will be considered in contributing to the distance. For example, m_{ij} in Eq. (12) can be redefined as follows [8]:

$$m_{ij} = \begin{cases} 1 - d_{ij}/d_{\max} & \text{if } d_{ij} \leqslant T_d, \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where T_d is the maximum distance for two colours to be considered similar. The value of $d_{\max}$ has to be redefined as αT_d where α is a constant between 1.0 and 1.5. These matrices will later be used in some of the experiments. The second problem we address in this paper is the question if a compression method based on a minimization of the reconstruction error is appropriate for retrieval. Against this choice one can argue that the ultimate aim of compressing histograms in image retrieval applications is not to reconstruct the histograms but to estimate distances to histograms similar to the histogram of the query image. In that sense, image retrieval is concerned with the (dis)similarity or the differences between histograms. In Eq. (1) the distance was defined as

$$d_M^2(h_1, h_2) = (h_1 - h_2)^{\mathrm{T}} M (h_1 - h_2).$$

It seems reasonable to expect that a KLT designed to provide the best reconstruction of the differences between colour histograms may lead to a better retrieval performance. Since we care only about similar images, only pairs of similar colour histograms are taken into account in the compression. We therefore define for a (small) constant δ the space $\mathbb{D}_\delta$ of local histogram differences as:

$$\mathbb{D}_\delta = \{\Delta h = h_1 - h_2 : h_1, h_2 \in \mathbb{P}, \\ d_M(h_1, h_2) \leqslant \delta\}. \tag{15}$$

Another way to define the space of local histogram differences is based on the set of nearest neighbours. For each colour histogram h_1, we define the local differences space at every $h_1 \in \mathbb{P}$ as

$$\mathbb{D}_n^{h_1} = \{\Delta h = h_1 - h_2 : h_2 \in \mathbb{P}, \\ d(h_1, h_2) \text{ are the } n \text{ smallest distances}\}. \tag{16}$$

The space of local histogram differences is then defined as the union of all such $\mathbb{D}_n^{h_1}$ at every $h_1 \in \mathbb{P}$

$$\mathbb{D}_n = \bigcup_{h_1 \in \mathbb{P}} \mathbb{D}_n^{h_1}. \tag{17}$$

After the construction of the spaces of local histogram differences, KLT-techniques are used as before with the only difference that now they operate on the space $\mathbb{D}_\delta$ given in Eq. (15) or the space $\mathbb{D}_n$ given in Eq. (17) instead of the histogram space $\mathbb{P}$. The basis obtained from applying KLT on $\mathbb{D}_\delta$ and $\mathbb{D}_n$ are then used for compressing the features in the space of colour histograms $\mathbb{P}$. The following remark may help to understand why this strategy gives a good estimation of the histogram distance even though it (probably) gives a poor approximation of the histograms involved:

- Denote the basis functions computed from the local histogram differences by φ_k.
- Take two neighbouring histograms h_Q, h_I and approximate them, and their difference in this basis: $h_Q \approx \sum_{k=1}^{K} \alpha_k \varphi_k, h_I \approx \sum_{k=1}^{K} \beta_k \varphi_k, h_Q - h_I \approx \sum_{k=1}^{K} \gamma_k \varphi_k$ with $\gamma_k = \alpha_k - \beta_k$.
- Now the coefficients γ_k which are optimal (since they are the coefficients in the expansion of the difference) can be computed from the coefficients of the original expansions (which are non-optimal) and we have $|\gamma_k| = |\alpha_k - \beta_k|$.

Summarizing we can say that the KLT-based methods proposed here are designed to meet the following two requirements:

- Statistical properties of the image database and properties of the underlying colour space should be incorporated into the distance measure and into the compression.
- The compression should minimize the approximation error in the space of local histogram differences instead of the space of colour histograms.

3. Experiments

We implemented the methods described above and compared their retrieval performance with some traditional colour-based methods. In the following we use the following methods:

H_K	Full colour histogram with K bins
D_K	Dominant colour-based method [5,8,14]
K_K^{QB}	KLT-based method from QBIC [7]
K_K	Ordinary KLT in the space of histograms $\mathbb{P}$
$K_K^{\mathbb{D}}$	KLT in the space of differences of neighbouring histograms $\mathbb{D}_n$
$K_K^{\mathbb{M}}$	KLT in $\mathbb{P}$ with colour metric M
$K_K^{\mathbb{D}M}$	KLT in $\mathbb{D}_n$ with colour metric M

The approximation order (or the dimension of the compressed feature space) used in the experiments is given by the subscript K and this notation will be used in the rest of this section. The following image databases of totally more than 130,000 images are used in our experiments:

Corel database: 1000 randomly chosen colour images from the Corel Gallery

MPEG-7 database: 5466 colour images and 50 standard queries [21] designed to be used in the MPEG-7 colour core experiments

Matton database: 126,604 colour images. These images are low-resolution images of the commercial image database maintained by Matton AB in Stockholm (the average size is 108×120 pixels)

In all our experiments, the retrieval performance is measured based on the average normalized modified retrieval rank (ANMRR) [8,10]. The detailed description of ANMRR is complicated but lower values indicate high retrieval rate with relevant items ranked at the top. Zero means that all the ground truth images have been retrieved, 1 means that none of the ground truth images have been retrieved (a detailed description can be found in the appendix). A colour-based search engine for image databases (CSE) was developed in order to compare the retrieval performance of the different methods described in this article. A demo of the CSE system with the Matton database of 126,604 images is currently available under http://www.ep.liu.se/databases/cse-imgdb.

3.1. Properties of colour histogram space vs. retrieval performance

The retrieval performance of histogram-based methods using quadratic form distances depends on the construction of the colour histogram and the metric M defining the properties of the histogram space. In the first set of experiments, the following four different methods of defining the metric M are evaluated in order to find a good matrix M for the next sets of experiments:

M_1	standard method as described in Eq. (12)
M_2	exponential function as in Eq. (13)
M_3	colour threshold T_d as in Eq. (14)
M_4	combination of colour threshold and exponential roll-off

There are several parameters in the construction of each method used to define M. Changing these parameters affects the distance measure between colour histograms and consequently the retrieval performance of the colour-based image retrieval. Increasing σ in Eq. (13), for example, will reduce the influence of neighbouring colour bins and vice versa. Fig. 1 shows the ANMRR of the 50 standard queries for the MPEG-7 database when the metric is defined as M_4 and σ (or ρ, the normalized version[1] of σ for the case $k=2$), is varying. The experiment is repeated for other methods defining M. Table 1 summaries the best retrieval performance of each method for different colour spaces. The results show that the distance measure in Eq. (12) overestimates the mutual similarity of dissimilar colours. The retrieval performance is improved using the distance measures in Eqs. (13) and (14). However, when ρ in Eq. (13) increases too much and/or the value T_d in Eq. (14) decreases too much, the retrieval performance is getting worse. The experimental results show also that the optimum retrieval performance of methods M_2, M_3, and M_4 (which

[1]For the sake of simplicity in parameterizing M, parameter ρ was introduced as a simple normalized version of σ for the case $k=2$ as

$$\rho = \frac{\sigma}{d_{\max}^2 * \text{ standard deviation of all histograms}}.$$

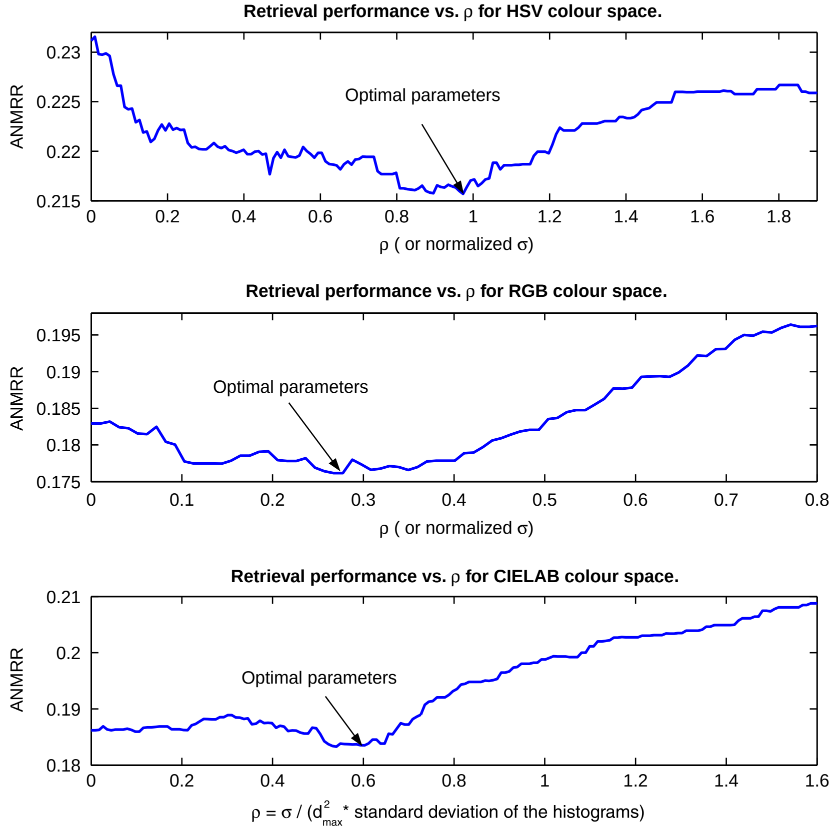

Fig. 1. Properties of metric M_4 in Eq. (13): ANMRR of 50 standard queries from the MPEG-7 database for different colour spaces when constants σ and ρ are varying. $T_d = 30, \alpha = 1.2, d_{\max} = 36$.

Table 1
Best retrieval performance (measured by ANMRR of 50 standard queries in (the MPEG-7 database) of different methods of defining the metric M for the colour histogram space in HSV $16 \times 4 \times 4$ bins, RGB $8 \times 8 \times 8$ bins, and CIELAB $8 \times 8 \times 8$ bins

M	HSV 256 bins	RGB 512 bins	Lab 512 bins
M_1	0.237	0.229	0.226
$M_2, k = 2$	0.214	0.174	0.188
M_3	0.215	0.174	0.198
M_4	0.216	0.176	0.183

is a combination of both) are comparable. The optimal parameters depend on both the colour perception of the observer and the application at hand. Finding such an optimal metric M can be done experimentally and its estimation is not discussed here. Instead we used the experiments summarized in Fig. 1 and Table 1 to determine a set of reasonable parameters for the remaining experiments.

3.2. Experiments with the Corel database

In the second set of experiments, we estimate the influence of the different approximation methods including the usage of coarser histograms [9,12], dominant colours or signature colours [2,5,14], the standard KLT, the method used in [7,11] and the proposed KLT-based methods as presented in the previous section. We compare the retrieval results of the approximation-based methods to the retrieval result achieved when the full histogram is used. One thousand images (randomly chosen) from the Corel image database were used in the experiments. In the first processing step we compute different descriptions of the colour distribution of an image. The CIELAB colour space and the distance measure using the metric M_2 as in Eq. (13) were chosen for these experiments. In the second step we use these descriptions to approximate the quadratic form-based distance measure from Eq. (1). In the retrieval simulation we use every image in the database as a query image and search the whole image database. The result is then compared to the standard method based on the full histogram. This allows us to evaluate the approximation performance of different methods in the context of colour-based image retrieval. Again ANMRR is used in the evaluation. In the dominant colour-based method, images are segmented into several homogenous regions. The clustering uses the mean shift algorithm [4]. Three different parameter settings were used to cluster each image in the database. The resulting clustered images consisted on average of 8, 25.5 and 44.5 segmented regions. The dominant colour of each region is then quantized to one of 512 CIELAB values in the original method in order to speed up the search algorithm. Each region is then described by two parameters: the probability of a pixel lying in this region and the index of the dominant colour of the region. An image which is segmented into n dominant colour regions is then described by $2 \times n$ parameters. For KLT-based methods operating on space $\mathbb{D}$, we used for every image its 40 nearest neighbours to estimate the space of local histogram differences. Figs. 2 and 3 show results with different lengths of query windows for the case where the metric M_2 is defined as in Eq. (13) using $\rho = 0.3$. Results with other choices of ρ are collected in

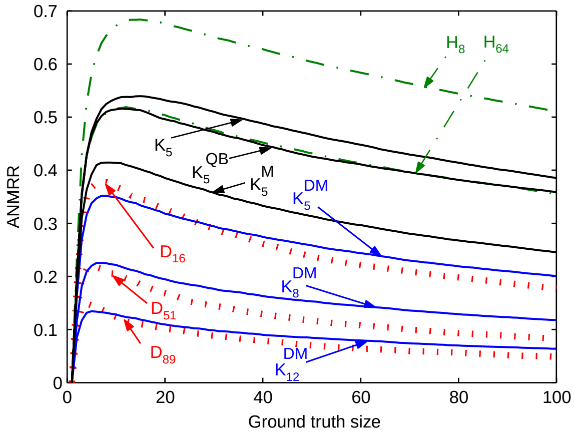

Fig. 2. ANMRR of 1000 queries in the Corel database using different histogram compression methods compared to the full histogram-based method.

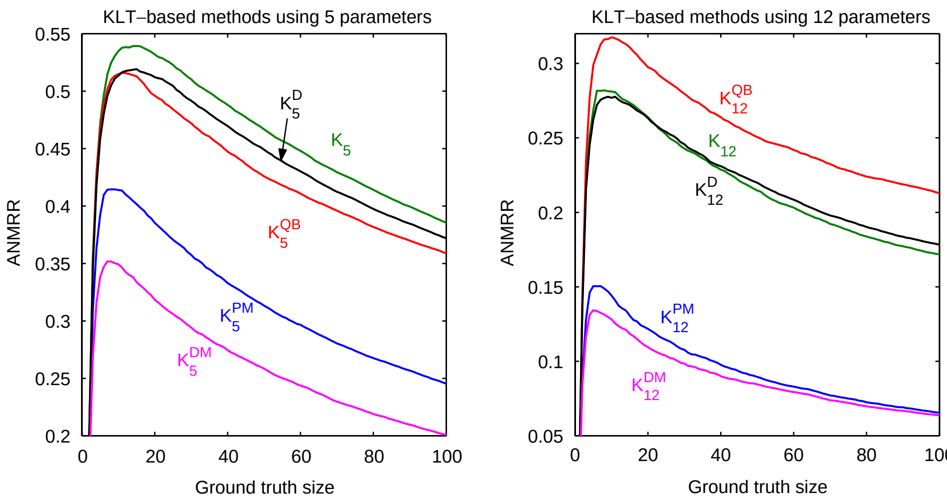

Fig. 3. ANMRR of 1000 queries in the Corel database using different KLT-based histogram compression methods compared to the full histogram-based method.

Table 2
Mean values of ANMRR of 1000 queries in the Corel database when the ground truth size varies from 10 to 40 for different histogram compression methods compared to the full histogram-based method. Different metrics M were used

ρ (normalized σ)	K_5^{QB}	K_5	$K_5^{\mathbb{D}}$	$K_5^{\mathbb{M}}$	$K_5^{\mathbb{D}M}$	D_{16}	H_8
0.08	0.418	0.575	0.561	0.154	0.116	0.259	0.640
0.15	0.441	0.542	0.526	0.237	0.204	0.275	0.643
0.3	0.484	0.519	0.500	0.373	0.308	0.310	0.661
0.7	0.545	0.513	0.482	0.441	0.409	0.374	0.693
ρ (normalized σ)	K_{12}^{QB}	K_{12}	$K_{12}^{\mathbb{D}}$	$K_{12}^{\mathbb{M}}$	$K_{12}^{\mathbb{D}M}$	D_{51}	H_{64}
0.08	0.131	0.303	0.336	0.027	0.021	0.123	0.466
0.15	0.203	0.269	0.275	0.055	0.051	0.135	0.471
0.3	0.290	0.254	0.254	0.116	0.106	0.159	0.489
0.7	0.257	0.533	0.248	0.189	0.183	0.208	0.524

Table 2. The results from these experiments show that:

- Incorporating information from the structure of the colour space and applying KLT in the space of differences between neighbouring histograms make the search results in the approximated feature space more correlated to the original full histogram method. The proposed method $K^{\mathbb{D}M}$, which combines the two ideas described above, gives the best performance compared to the other methods in all experiments. For example in Fig. 2, $K_5^{\mathbb{D}M}$, using only 5 parameters, gives the same retrieval performance as the dominant colour-based method using 16 parameters. It is superior to the full histogram-based method using 64 parameters. $K_{12}^{\mathbb{D}M}$ using only 12 parameters gives about the same retrieval performance as the dominant colour-based method using 89 parameters.
- The retrieval performance of these methods, however, depends on the matrix M, or how the information of the colour space is integrated into the standard KLT on image data. All experiments (see Fig. 1 and Table 2) show that there is an optimal matrix M for each method which is a balance between two extreme cases: ignore the colour information (take the colour histogram as ordinary signals and do not weight the mutual information between any pair of difference colours), and equal weighting of the mutual information between any two colours.
- Particularly, when σ is small, the K^{QB} method described in QBIC [7] is comparable to other KLT-based methods. This is, however, the case when the mutual similarity between dissimilar colours is overestimated. When σ is increased, or the metric M becomes more diagonally dominant, the retrieval performance of the K^{QB} method decreases, compared to other KLT-based methods which are not solely based on the matrix M.
- For large values of $K (K \geqslant 15)$, results of K^{DM} methods which incorporate both the colour metric M and image data converged to the standard method much faster than K^{QB}.
- The dominant colour-based method is fairly good while simple KLT and coarse histogram-based methods show poor results. Performance of the coarse histogram with 64 parameters is inferior than using only 4 parameters in our K_4^{DM} method.

In order to confirm these conclusions, large-scale experiments with the bigger databases were carried out.

3.3. Experiments with the MPEG-7 database

In the third set of experiments, KLT-based methods are investigated further with the MPEG-7 databases of 5466 colour images. Both objective and subjective queries are used. First, the same experiments as in the previous section are done

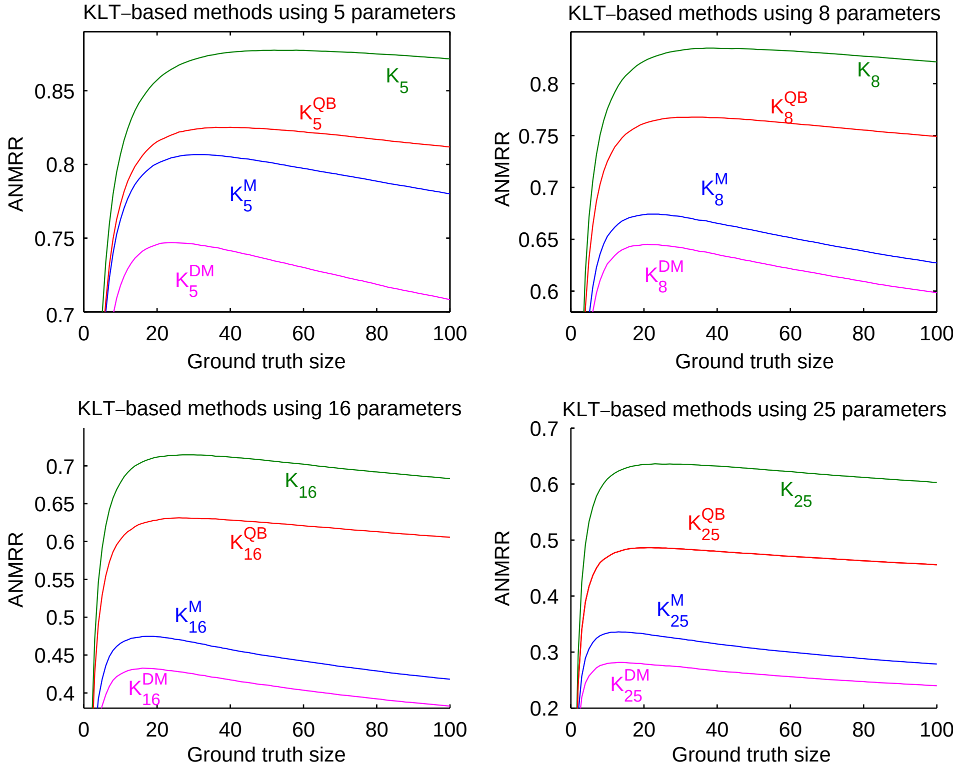

Fig. 4. ANMRR of 5466 queries in the MPEG-7 database using different KLT-based histogram compression methods compared to the full histogram-based method.

Table 3
Mean values of ANMRR of 5466 queries in the MPEG-7 image database when the ground truth size varies from 10 to 40 for different methods compared to the full histogram method

Colour space and Desc. of the method	K^{QB}	K	K^{M}	K^{DM}
HSV $16 \times 4 \times 4$, # of parameters $K = 5$	0.673	0.628	0.491	0.490
HSV $16 \times 4 \times 4, K = 8$	0.544	0.544	0.386	0.365
HSV $16 \times 4 \times 4, K = 16$	0.377	0.414	0.197	0.182
HSV $16 \times 4 \times 4, K = 25$	0.266	0.314	0.114	0.107
RGB $8 \times 8 \times 8, K = 5$	0.775	0.576	0.436	0.419
RGB $8 \times 8 \times 8, K = 8$	0.729	0.405	0.268	0.243
RGB $8 \times 8 \times 8, K = 16$	0.546	0.227	0.102	0.091
RGB $8 \times 8 \times 8, K = 25$	0.450	0.153	0.044	0.041
CIELAB $8 \times 8 \times 8, K = 5$	0.558	0.579	0.475	0.455
CIELAB $8 \times 8 \times 8, K = 8$	0.505	0.453	0.319	0.292
CIELAB $8 \times 8 \times 8, K = 16$	0.425	0.251	0.151	0.137
CIELAB $8 \times 8 \times 8, K = 25$	0.345	0.165	0.075	0.072

with the MPEG-7 database. The only different setting was that the number of neighbouring images of each image used when constructing the space of local histogram differences is 100 images. Several colour spaces, including HSV, RGB and CIELAB, are used in these experiments. Fig. 4 and Table 3 show the results for different colour spaces. We also used 50 standard queries as subjective search criteria to compare the retrieval performance of these KLT-based methods. The results are shown in Table 4. In another experiment, we select a set of 20 images, where 10 of them are from standard queries, and the other 10 are well-known images such as Lena, Peppers, Mandrill, Parrots, etc. From each of these 20 images a new set of 20 images is generated by adding noise and sub-sampling the images. There are totally 420 images. The parameters that control the generated images are: P_s = percentage

Table 4
ANMRR of 50 standard queries in the MPEG-7 image database

Colour space and Desc. of the method	K^{QB}	K	K^M	K^{DM}
HSV $16 \times 4 \times 4$, # of parameters = 8	0.422	0.337	0.337	0.333
HSV $16 \times 4 \times 4, K = 16$	0.352	0.247	0.257	0.263
HSV $16 \times 4 \times 4, K = 25$	0.297	0.238	0.248	0.247
RGB $8 \times 8 \times 8, K = 8$	0.487	0.381	0.311	0.316
RGB $8 \times 8 \times 8, K = 16$	0.347	0.283	0.232	0.229
RGB $8 \times 8 \times 8, K = 25$	0.288	0.275	0.200	0.200
CIELAB $8 \times 8 \times 8, K = 8$	0.336	0.383	0.322	0.301
CIELAB $8 \times 8 \times 8, K = 16$	0.287	0.298	0.251	0.233
CIELAB $8 \times 8 \times 8, K = 25$	0.266	0.256	0.224	0.222

Table 5
ANMRR of 20 generated queries for the MPEG-7 image database

P_s	P_n	R_n	# of Dim.	K^{QB}	K	K^M	K^{DM}
20	20	20	5	0.0181	0.0119	0.0111	0.0060
20	20	20	8	0.0098	0.0084	0.0059	0.0049
20	20	20	16	0.0111	0.0051	0.0042	0.0035
20	20	20	25	0.0046	0.0033	0.0032	0.0031
20	20	40	5	0.1225	0.0429	0.0403	0.0346
20	20	40	8	0.0458	0.0200	0.0235	0.0206
20	20	40	16	0.0215	0.0142	0.0181	0.0172
20	20	40	25	0.0139	0.0134	0.0173	0.0172
40	20	20	5	0.0181	0.0116	0.0121	0.0063
40	20	20	8	0.0098	0.0084	0.0060	0.0051
40	20	20	16	0.0111	0.0048	0.0043	0.0035
40	20	20	25	0.0041	0.0031	0.0030	0.0029
60	10	50	5	0.0302	0.0110	0.0144	0.0111
60	10	50	8	0.0192	0.0090	0.0071	0.0068
60	10	50	16	0.0115	0.0045	0.0053	0.0040
60	10	50	25	0.0038	0.0030	0.0029	0.0028

of sampled pixels, P_n = percentage of pixels with added noise, and R_n = the range of the noise magnitudes. Noise is uniformly distributed. Only the RGB colour space is used in this experiment. Each set of 20 generated images is supposed to have similar colour distributions as the original image. We then take these 20 images as the ground truth when retrieving the original image. The average results of 20 different queries are collected in Table 5. The results from the simulation of the search process on both objective and subjective queries of the MPEG-7 database all agreed with the results obtained from the Corel database in Section 3.1.

3.4. Experiments with the Matton database of 126,604 images

Finally we extend the comparison to the large Matton image database containing 126,604 images. The experiment set-up is as in the second set of experiments described in Section 3.2. The colour histograms were computed in the HSV colour space using $16 \times 4 \times 4$ bins. A set of 5000 images was selected randomly, the basis of different KLT-based methods are then computed from this set. For KLT-based methods operating on the space $\mathbb{D}$, we used for every image its 100 nearest neighbours to represent the local histogram differences. Fig. 5 shows the average results when all 5000 images in the training set were used as query images. We also selected another 5000 images, not in the training set, as query images in the image retrieval simulation, the average results for this set are collected in Fig. 6. Twenty queries from the set of 420 generated images as described in Section 4.3 are also used to evaluate KLT-based methods in the Matton database. The results are shown in Table 6. As expected, the results obtained from the large database also agreed with earlier results of the small-scale experiments on the Corel database of 1000 images.

4. Conclusions

We applied KLT-based approximation methods to colour-based image retrieval. We presented different strategies combining two ideas: Incorporating information from the structure of the colour space with information from images; and using projection methods in the space of colour histograms and the space of differences between neighbouring histograms. The experiments with three databases of totally more than 130,000 images using different sets of parameters such as

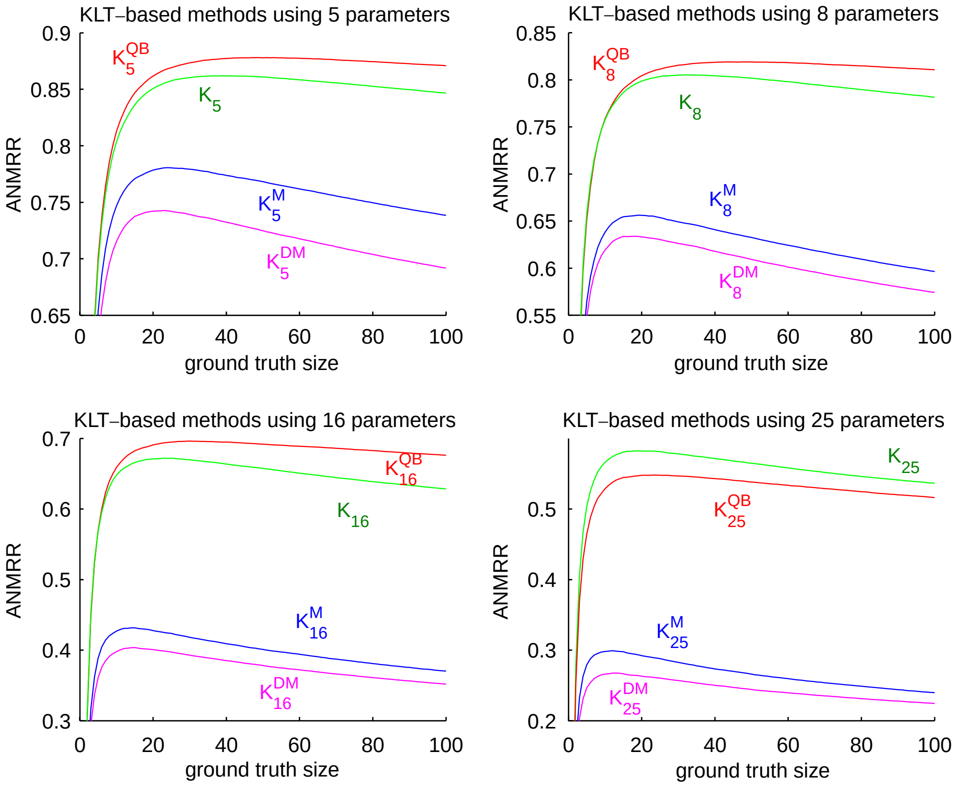

Fig. 5. ANMRR of 5000 queries (selected from the training set) in the Matton database using different KLT-based histogram compression methods compared to the full histogram-based method.

colour spaces, bin size and bin-locations of histograms, etc. show that the method which combines both the colour metric and the difference of histograms space gives very good results compared to other existing methods. Investigating which set of parameters one should use for a given CBIR application was not our primary interest in this paper. However, some comments which might be of interest are as follows: Regarding the selection of colour space, our experiments show that there is a difference in the retrieval performance for different colour representations. They indicate that the common assumption that CIELAB descriptions are better for colour-based image retrieval since they are better adapted to human colour vision have to be investigated further. This argument ignores that the CIELAB system is strictly valid only in the framework of colour matching, i.e. in comparing single colours under controlled conditions. This is very different from the colour-based image retrieval situation in which statistical distributions of colours have to be compared. A comprehensive investigation of this problem has, to our knowledge, not been done. The problem of choosing the parameters used in the computation of the histogram is not investigated here but some relevant results (and a comparison with kernel-based density estimators for image retrieval) is reported in [19]. We also want to mention that the general strategy of using problem-based distance measures and differences of histograms outlined above is quite general and can be applied for other features used in content-based image retrieval applications.

Appendix: Definition of the ANMRR

Given a query q with NG_q ground truth images, suppose that the kth ground truth image I_k is

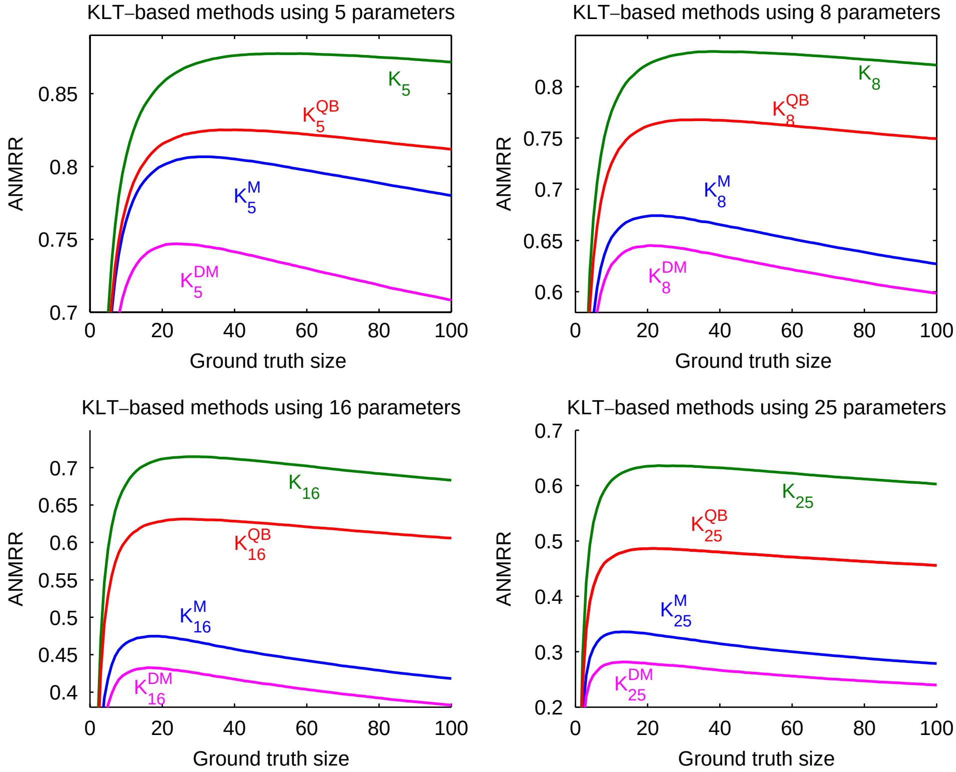

Fig. 6. ANMRR of 5000 queries (not selected from the training set) in the Matton database using different KLT-based histogram compression methods compared to the full histogram-based method.

Table 6
ANMRR of 20 generated queries for the Matton database

P_s	P_n	R_n	# of Dim.	K^{QB}	K	K^M	K^{DM}
40	30	60	5	0.317	0.520	0.050	0
40	30	60	8	0.336	0.083	0.014	0.001
40	30	60	16	0.507	0.007	0	0
40	30	60	25	0.174	0.001	0	0
40	30	50	5	0.312	0.445	0.045	0
40	30	50	8	0.305	0.068	0.007	0.001
40	30	50	16	0.442	0.005	0	0
40	30	50	25	0.135	0.001	0	0
40	25	50	5	0.240	0.353	0.032	0
40	25	50	8	0.232	0.054	0.002	0
40	25	50	16	0.332	0.003	0	0
40	25	50	25	0.093	0.0030	0	0

found at specific position R_k of the retrieval result. The retrieval rank of I_k is defined specifically in our experiments as

$$\text{Rank}(I_k) = \begin{cases} R_k & \text{if } R_k \leqslant 2 \cdot NG_q, \\ 2.5 \cdot NG_q & \text{otherwise} \end{cases}$$

and the average rank (AVR) for query q is given by

$$\text{AVR}_q = \frac{1}{NG_q} \sum_{k=1}^{NG_q} \text{Rank}(I_k)$$

ANMRR of a set of queries is then defined as the average for all queries of the normalized version of AVR to a value between 0 and 1. 0 means that all the ground truth images have been retrieved, 1 means that none of the ground truth images have been retrieved. Mathematically, the definition of ANMRR is given as

$$\mathrm{ANMRR} = \frac{1}{NQ}\sum_{q=1}^{NQ} \frac{\mathrm{AVR}_q - 0.5 \cdot (1 + NG_q)}{2 \cdot NG_q - 0.5}.$$

As examples, suppose that we have a query with 30 ground truth images, if there only one ground truth image is missed in the retrieval result, the ANMRR = 0.042 if the missing image is in the 1st rank, and ANMRR = 0.025 if it is in the last rank. If we missed the first five images, ANMRR = 0.202, and if we missed the last 5 images, ANMRR = 0.132. If we missed the first six images, ANMRR = 0.240, and if we missed the first five images and the last image, ANMRR = 0.227.

Acknowledgements

This work was supported by the VISIT (VISual Information Technology) program of the Swedish Foundation for Strategic Research (SSF). Reiner Lenz was supported by Center for Industrial Information Technology, Linköping University (CENIIT) and the Swedish Research Council (VR).

References

[1] E. Albuz, E. Kocalar, A.A. Khokhar, Scalable color image indexing and retrieval using vector wavelets, IEEE Trans. Knowl. Data Eng. 13 (5) (2001) 851–861.
[2] D. Androutsos, K.N. Plataniotis, A.N. Venetsanopoulos, A novel vector-based approach to color image retrieval using a vector angular-based distance measure, Computer Vision Image Under. 75 (1/2) (1999) 46–58.
[3] J. Berens, G.D. Finlayson, G. Gu, Image indexing using compressed colour histogram, IEE Proc. Vision Image Signal Process. 147 (4) (2000) 349–353.
[4] D. Comaniciu, P. Meer, Distribution free decomposition of multivariate data, Pattern Anal. Appl. 2 (1) (1999) 22–30.
[5] Y. Deng, B.S. Manjunath, C. Kenney, M.S. Moore, H. Shin, An efficient color representation for image retrieval, IEEE Trans. Image Process. 10 (1) (2001) 140–147.
[6] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1990.
[7] J. Hafner, H.S. Sawhney, W. Equitz, M. Flickner, W. Niblack, Efficient color histogram indexing for quadratic form distance functions, IEEE Trans. Pattern Anal. Mach. Intell. 17 (7) (1995) 729–736.
[8] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Systems Video Technol. 11 (6) (June 2001) 703–715.
[9] M. Mitra, J. Huang, S.R. Kumar, Combining supervised learning with color correlograms for content-based image retrieval, in: Proceedings of Fifth ACM Multimedia Conference, 1997.
[10] P. Ndjiki-Nya, J. Restat, T. Meiers, J.-R. Ohm, A. Seyferth, R. Sniehotta, Subjective evaluation of the MPEG-7 retrieval accuracy measure (ANMRR), Technical Report, ISO/IEC JTC1/SC29/WG11, Doc. M6029, May 2000.
[11] R.T. Ng, D. Tam, Multilevel filtering for high-dimensional image data: Why and how, IEEE Trans. Knowl. Data Eng. 11 (6) (1999) 916–928.
[12] G. Pass, R. Zabih, Comparing images using joint histograms, Multimedia Systems 7 (3) (1999) 234–240.
[13] J. Puzixha, J.M. Buhmann, Y. Rubner, C. Tomasi, Empirical evaluation of dissimilarity measures for color and texture, in: Proceedings of IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1165–1172.
[14] Y. Rubner, Perceptual metrics for image database navigation, Ph.D. Thesis, Stanford University, May 1999.
[15] Y. Rui, T.S. Huang, S.-F. Chang, Image retrieval: current techniques, promising directions, and open issues, J. Vis. Commun. Image Represent. 10 (1) (1999) 39–62.
[16] R. Schettini, G. Ciocca, S. Zuffi, A survey on methods for colour image indexing and retrieval in image database, in: R. Luo, L. MacDonald (Eds.), Color Imaging Science: Exploiting Digital Media, Wiley, New York, 2001.
[17] M.A. Stricker, M. Orengo, Similarity of color images, in: W. Niblack, R. Jain (Eds.), Storage and Retrieval for Image and Video Databases, SPIE, 1995, pp. 381–392.
[18] M.J. Swain, D.H. Ballard, Color indexing, Int. J. Comput. Vision 7 (1) (1991) 11–32.
[19] L.V. Tran, R. Lenz, Characterization of color distributions with histograms and kernel density estimators, in: Proceedings of SPIE-2003, Internet Imaging VIII, 2003.
[20] R. Weber, H. Schek, S. Blott, A quantitative analysis and performance study for similarity search methods in high-dimensional spaces, in: Proceedings of 24th VLDB, 1998, pp. 194–205.
[21] D. Zier, J.R. Ohm, Common datasets and queries in MPEG-7 color core experiments, Technical Report, ISO/IEC JTC1/SC29/WG11, Doc. M5060, October 1999.