Perception Based Image Retrieval

Dirk Neumann* Karl R. Gegenfurtner†

October 30, 2002

Address for Correspondence:

Prof. Dr. Karl R. Gegenfurtner

Justus Liebig University Giessen

Cognitive Psychology Dept., FB06

Otto-Behaghel-Strasse 10F

D-35394 Giessen

Germany

*neumannd@in.tum.de

†gegenfurtner@uni-giessen.de

Abstract

We created an image indexing system based on some of the known properties of the early stages of human vision. We used a color space known to underlie the second stage of human color vision and stored chromaticity and luminance information in two logarithmic-radial histograms. A third, spatial index encodes – in analogy to the spatial frequency representation in the visual cortex – information about orientation and spatial scale.

The indices were evaluated by comparing the computed similarity values with human judgments quantitatively and objectively in a 2AFC design. For the experiments we used a heterogeneous database of 60,000 digitized photographs.

1 Introduction

The goal of image indexing systems is to find a set of images that is similar to the target image the user is looking for. Depending on the intention of the user, similarity may be defined by objects, configuration, illumination, camera position or zoom, by semantic aspects, or any combination of the above. Preferably, computer vision algorithms should extract all the relevant features from the image in the same way a human observer would do. But that is far beyond our current knowledge of human vision, cognition and emotion.

Most of the current algorithms are quite successful in using low-level features of the images, as this quite frequently results in semantically related images. Color, in particular, has proven to be very effective for the calculation of image similarity, since an object's color is independent of viewing position or viewing distance. The most prominent color statistics are histograms with equally sized bins in RGB or HSI color space. In HSI space the luminance (intensity) axis is often ignored since it is argued that the overall brightness of an image is irrelevant with respect to image similarity. Hence, luminance independent indices should be more robust with respect to different illumination conditions. Other color statistics include correlation or covariance coefficients encoding spatial information about the color distribution (e.g. Huang et al., 1999, Stricker and Dimai, 1997). Features such as texture and shape are other subjects of current research (e.g. Flickner et al., 1995).

All these approaches are based on physical, low-level features. But in the end the similarity of a given image with a target image will always be judged by a human observer. Therefore, our goal was to construct indices which are based on some of the known properties of the early levels of the human vision system. Furthermore, we used a strictly quantitative and objective approach

to evaluate the resulting image metric by comparing it with measurements of perceived image similarity.

As the first step of seeing, light is absorbed and converted into neural signals by the three different classes of cone photoreceptors in the retina. The details of the absorption spectra of the cones were well studied in the past and are known quite precisely by now. While these cones are often called red-, green- and blue-cones, they all do absorb light over a wide range of the visible spectrum. Most notably, they do not have much to do with the RGB-triplets of modern image sensors. There is no simple transformation converting between the two, and any such conversion process requires careful calibration of the image acquisition device (Wandell, 1993).

Still at the level of the retina, the signals from the cones get transformed by a complex network of retinal cells (Wässle and Boycott, 1991) into color opponent signals. Electrical recordings from single neurons in the retina and the lateral geniculate nucleus (LGN) have shown three different classes of neurons. A "luminance-type" neuron simply takes the sum of the outputs from all three cone classes. "Red-green" opponent neurons take the difference between the red- and the green-cone signal. "Blue-yellow" opponent neurons take the difference between the blue-cone signal and the sum of the red- and green-cone signal (Gegenfurtner and Kiper, in press). Incidentally or not, this basically results in a principal components analysis of the cone signals (Buchsbaum and Gottschalk, 1983, Ruderman et al., 1998) and thus removes any correlation between the signal channels, resulting in nearly optimal information transmission from the eye to the brain.

In the visual cortex, spatial information is extracted from these signals using an array of linear and non-linear filters tuned to spatial frequency and orientation (Hubel and Wiesel, 1998). The tuning characteristics of these fil-

ters seem to be highly optimized for viewing our typical visual environment (Field, 1987). The efficiency of these spatial filters was modeled by Watson (1987). He transformed the grayscale image using the discrete Fourier transformation (DFT) and applied a filter to the coefficient matrix with the same orientation and bandwidth properties found in early spatial vision. In a psychophysical experiment he found that a reconstructed image is typically indistinguishable from the original at a code size of 1 bit/pixel.

We tried to use these established coding principles of the human visual system to construct image histograms optimized to these characteristics. The color space (DKL) proposed by Derrington et al. (1984), based on recordings of single neurons in the LGN of macaques was also found to play an important role in human color vision. Krauskopf et al. (1982) showed that the increase of thresholds for detecting colors after adaptation to isoluminant color changes is independent along the color-opponent axes.

We used that color-opponent space to create two color indices. A first index encodes only chromaticity and is therefore luminance independent. The second index encodes the mean luminance of the color tones. For the color bins we did not use a equidistant spacing, the size of the bins reflects the granularity of higher order color perception e.g. for saturated colors the resolution for hue is much finer than it is for unsaturated colors while the resolution of saturation decreases, in agreement with the results of Krauskopf and Gegenfurtner (1992).

To extract the spatial information we used the luminance dimension only, which, in the DKL color space, is orthogonal to the color-opponent subspace used for the color indices. For each image the Fourier power coefficients were calculated using DFT. The 2-dimensional power spectrum was further segmented into bins representing spatial contrasts of distinct orientation and

frequency ranges. The resolution of the Fourier index is a function of frequency and orientation, representing the orientation of high frequency contrasts more precisely than that of lower frequencies.

The second major aspect of our system is its evaluation. Most of the earlier search systems were evaluated informally. The prevalent method is to have the experimenter decide which images found by the search algorithm are similar to the query image. As Cox et al. (2000) criticize, the results obtained by such methods are highly dependent on the strictness of the similarity criteria the observers use, the homogeneity of the images in the database and the number of images displayed.

We evaluated our color indexing system based on perception (CISBOP) by comparing the similarity judgments made by the algorithm to ratings made by human observers. Unlike Rogowitz et al. (1998) who calculated a similarity metric for a set of 96 images with a multidimensional scaling technique, we tried to measure similarity directly. By looking at the degree with which the algorithm correlates with the judgments of the observers, we can estimate how much different features (indices) contribute.

In three experiments we measured the relationship between the perceived similarity judgments and the computed similarity distance over a broad range of images. The similarity of images was varied from highly similar best matches to relatively distinct for images with rank 2000 in the result list.

In the first experiment, we investigated the influence of the degree of similarity between the test images on the similarity judgments. In the second experiment, the three indices were compared to each other to determine their individual contributions. In the third experiment, the indices where combined iteratively to measure the improvement in prediction when color, spatial, and luminance information are considered.

In brief, we found a good agreement between the images selected with the perception-based indices and the perceived similarity. The observers' judgments can be best predicted with the chromaticity histogram. The luminance and the Fourier histogram both contribute to the similarity judgments and the percentage of agreement increases considerably if the luminance and the Fourier information is combined with the chromaticity index. We found that the percentage of agreement decreases linearly as a function of the logarithmic rank position, from the first, best matching image up to the 2000th image in the result list. The correlation was found for each of the three indices and for the index combinations.

2 Color and Spatial Indexing

The aim of an image index is to create a metric for the similarity of images. Such a metric should ideally correspond to the users' perception of similarity. The idea of our color image indexing system based on perception (CISBOP) was to use some the the established coding principles of human vision to compare images.

For each image the distribution of the features hue, luminance and spatial frequency were determined. These distributions were summarized into three histograms (feature vectors). The similarity between two images was then defined as the distance between the feature vectors. In contrast to previous approaches, we normalized the distance values with regard to the mean and standard deviation of the distribution of similarity distances. The normalization transformation is computed on a per-image basis and leads to index weights that reflect the saliency of the corresponding features.

The main focus of our research, besides evaluation, was the first step in similarity metric creation, the modeling of appropriate image features. For this purpose we used the DKL color space (Derrington et al., 1984) that was found to underlie primate color vision. The chromaticity and the luminance information was stored in two separate indices, the information about the Fourier energy distribution in a third histogram.

2.1 Color Transformation

It was mentioned in the introduction that there is no simple transformation to convert RGB-triplets into human cone photoreceptor excitations. But these receptor excitations are necessary to calculate the DKL cone-opponent coordinates for these stimuli. We circumvented this problem by using the photore-

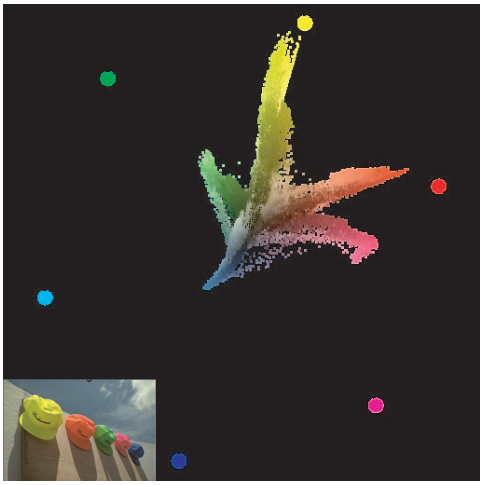
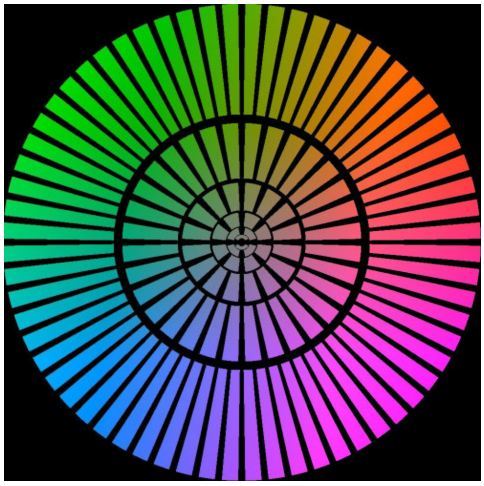
a)  b) 

Figure 1: Color distribution in the DKL space. a) The color distribution of the image with the hats in the bottom left plotted with respect to the coordinates at the color-opponent axes in DKL color space. When multiple pixels had the same position, luminance was averaged. b) Subdivision of the chromaticity subspace of DKL color space into 127 logarithmic-radial color bins used for the color indices.

ceptor excitations resulting from the display of the images on a standardized and calibrated display monitor (Sony GDM-F500). Since our observers made their judgments looking at the images displayed on that monitor, it is correct to use these excitations as the basis for the similarity metric.

The resulting color distribution of an image in the color-opponent DKL space could be visualized by plotting the pixels of the image with respect to their coordinates on the red-green and yellow-blue axis (Fig. 1a). In the chart, the pixels of the differently colored hats fall into separate radial sections.

The Cartesian color-opponent coordinates can be transformed into the polar coordinates hue and saturation. In the chart hue is the color angle ($hue = \arg(rg + i \cdot yb)$), saturation is the distance from the neutral gray ($sat = |rg + i \cdot yb|$). The cylindrical hue-saturation coordinates were then used to define the histogram bins.

2.2 Chromaticity and Luminance Index

The color histograms were created by dividing the chromaticity plane into logarithmic-radial segments (Fig. 1b). The resolution for saturation of these bins decreases with increasing saturation. Six different rings were used to discriminate saturation; the remaining unsaturated color tones in the center were averaged into a single, gray bin. The ring a color tone belongs to can be calculated using the logarithm of the saturation ($r = \lfloor -\log_2 sat \rfloor$). For hue, the histogram resolution is lowest for unsaturated, gray colors and doubles with increasing saturation ($n_r = 2^{\lfloor 6-r \rfloor}$). This yields 127 bins; 64 bins for the most saturated colors.

Furthermore, color values exceeding 95% or falling below 5% of the maximally possible luminance value of RGB space were classified as white and black. This is essential since the hue resolution for very dark and very bright colors is limited.

To calculate the two color indices the color distributions of the images were mapped into these 129 bins. For each image two vectors were stored: the frequency of the color tones f and the average luminance level of the pixels in each bin l. If a bin was empty the luminance level was set to zero.

$$
\begin{aligned}
f_{bin} &= \frac{|pixels|bin|}{|pixels|} \\
l_{bin} &= \begin{cases} \frac{\sum lum|bin}{|pixels|bin|}, & \text{if } f_{bin} \neq 0 \\ 0, & \text{if } f_{bin} = 0 \end{cases}
\end{aligned}
$$

Figure 2a shows the frequency of the color bins for the images with the hats. The luminance of the segments in the chart indicates their frequency.

The most frequent segment is white. In Figure 2b the color tones belonging to each color bin were plotted with the average luminance of that bin giving an impression of the information stored in the luminance vector.

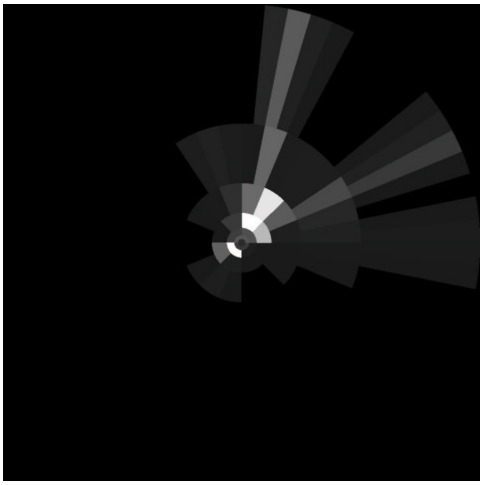
a)  b) 

Figure 2: Color histograms. a) Frequency of the color bins for the image with the hats (Fig. 1a). The brightness of the bins is proportional to their frequency in the image. The most frequent bin is white. b) Mean luminance in the color bins for the image with the hats. The color tones belonging to each color bin are plotted with the average luminance level of that bin.

2.3 Fourier Index

To determine the distribution of Fourier energy across different orientations, we used the 2-dimensional discrete Fourier transformation (DFT). The decomposition was performed in the luminance dimension of the DKL space only.

In contrast to the local Gabor transformation, the 2-dimensional discrete Fourier transformation is a global decomposition determining the amplitude for a set of 2-dimensional sine waves. These basis functions can be characterized by a single complex parameter. The polar coordinates of this Fourier parameter are the orientation and spatial frequency of the 2-dimensional sine waves. The DFT calculates for each parameter a complex coefficient storing the amplitude and the phase of the corresponding 2-dimensional sine waves. The image then can be reconstructed as a linear combination of this basis function.

The squared coefficient – the energy – respectively its distribution, the Fourier spectrum, was used to construct the spatial index. Because the grayscale values of the images are real (not complex) numbers, the resulting Fourier spectrum is symmetric and for the index creation only the upper half of the spectrum was used.

Since even the fast Fourier transformation (FFT) is computationally intensive, the 768x512 sized images were rescaled to 96x64 thumbnails using bilinear approximation. The thumbnails were then projected onto the luminance dimension of the DKL space. To correct for artifacts that arise from the rectangular form of the images, the grayscale images were multiplied with a circular mask prior to Fourier transformation. This leads to a more even distribution of Fourier energy across orientations because it removes the artifacts resulting from the image edges. Orthogonal components are still most frequent because many images contain objects with vertical or horizontal ori-

entations.

The resulting Fourier spectrum was divided into radial-logarithmic bins analogous to the chromaticity segments. The index contains 126 bins. Each segment represents contrasts of distinct orientation and frequency ranges. The bins at the origin correspond to contrasts with very low or zero frequency and store the mean luminance of the image. Because we wanted to construct a luminance independent spatial index we did not use these bins for searching; they were filtered out prior to searching.

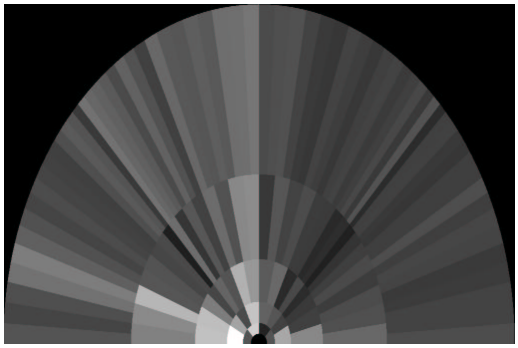


Figure 3: Fourier index for the image with the hats (Fig. 1a). Spatial frequency increases from the center to the edge of the chart. The central bins containing average luminance information were removed. The brightness of a bin is proportional to the logarithm of the average energy within the bin. The segment with the highest energy is white.

2.4 Distance metrics

To find similar images for a given query image it is necessary to compare the index vectors. Most indexing systems interpret the indices as points in an Euclidian space and use the Euclidian norm to define the distance between two images. We used it for comparing the luminance and the Fourier index.

For the chromaticity frequency index the more intuitive intersection norm was used. It defines the similarity between two images by the sum of the minimum of the corresponding bin frequencies ($s_{is} = \sum_{i=1}^{n} min\,(x_i, y_i)$). The value can be interpreted as the proportion the two color distributions share. For example, a value of 0.75 means that for 75% of the pixels in one image there exist pixels in the other image which fall into corresponding color bins. Figure 4 shows the percentage of color concordance for a sample query using the chromaticity index.

For very similar images the intersection norm results in a value near one because the sum over the bins of the chromaticity histogram is one. Images that do not share a single color bin are considered completely dissimilar. The similarity value would be zero. To use the intersection norm as a distance measure the similarity value must be negated. The transformation $d_{is} = 1 - s_{is}$ was used.

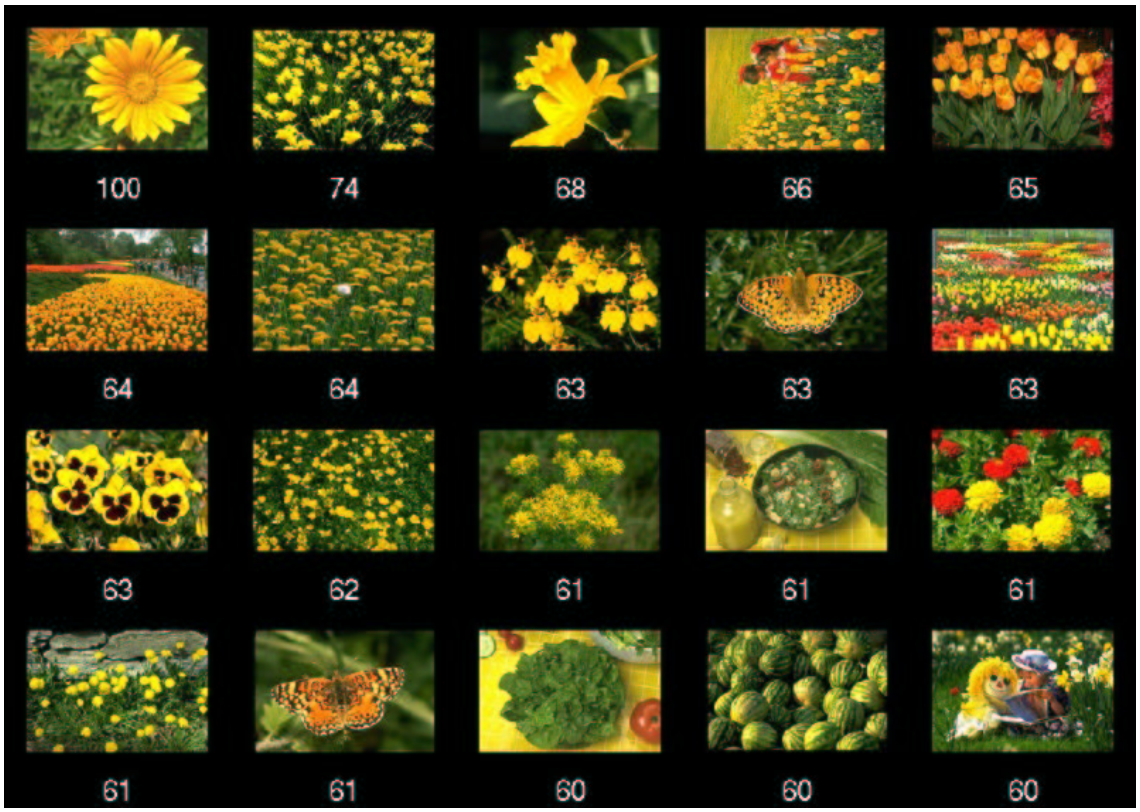


Figure 4: Sample query results comparing images using the chromaticity histogram and the intersection norm. The number below the images shows the percentage of pixels the images shares with the query image at the top left. The query images shares with itself 100% of the color distribution.

2.5 Cue Combination

It is necessary to normalize the different distance functions before combining the different indices. This is not only a consequence of using two different distance functions. Even if different indices are combined using the same distance function normalization is necessary because the variances of the indices can be quite different.

We used the z-transformation to normalize the distance values. The use of the z-transformation seems reasonable inasmuch as the distance distributions we inspected manually show the form of a slightly skewed Gaussian density function (Fig. 5). The parameters of the z-transformation are estimated before each query. For this purpose a set of 1000 images R is randomly selected. The mean and the standard deviation of the distribution of distances between the query image and the 1000 random images is then estimated ($z_d = \frac{d-\bar{d}}{\hat{\sigma}_d}$).

It is necessary to estimate the parameters for each query image again because the (query) images differ in their mean distances to all images and in the variance of the distance distribution (Fig. 5a and b). These values can be used to characterize the image. If $\bar{d}_s(q, I)$ is great and $\sigma_s(q, I)$ is small then the image q would be quite distinct regarding the similarity metric s used for comparisons. If an image is compared using multiple indices then the means and standard deviations characterize the importance of the different features for the comparison within the image database. The distance of an image to itself is zero per definition. Thus the z-value of an image to itself is the (negative) distance to the mean of the database in terms of the standard deviation. If the image is quite distinct from the images in the database then the absolute z-distance would be great.

The distance values of the different indices could then be combined by

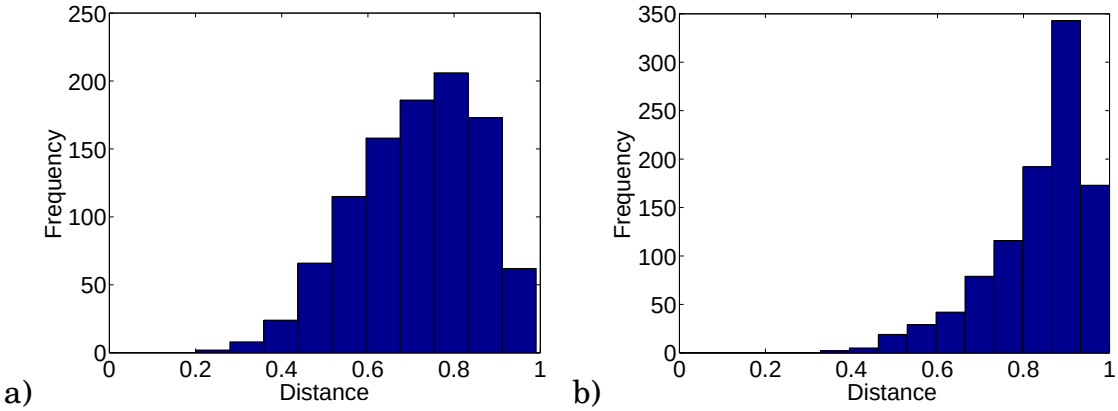


Figure 5: Histograms of distance values from two randomly choosen image to a set of 1000 randomly selected images using chromaticity histogram and intersection norm. The distance distributions have different means that will result in different z-values (see text).

averaging the z-values. The z-transformation – calculated for each index separately – therefore ranks the indices in terms of their salience. The indices that are more characteristic for an image are automatically weighted higher than the other indices.

$$z_S(q,i) = \frac{1}{|S|} \sum_{s \in S} \frac{d_s(q,i) - \bar{d}_s(q,I)}{\sigma_s(q,I)}$$

Figure 6: Sample query results for a combined search using z-transformation. Below each image the average of the z-values of the chromaticity, the luminance and the Fourier histogram distances is shown.

3 Implementation

The algorithms were implemented into a Java framework. It consists of five modules: database, features, feature combination, filters, a graphical user interface (GUI) and communication protocols.

The image database module is URL based and can be used to address either image files or images on http or ftp servers. For a query the different indices can be selected and combined with filters (e.g. removal of DC components in the Fourier spectrum) and one of the two distance metrics (Fig. 7).

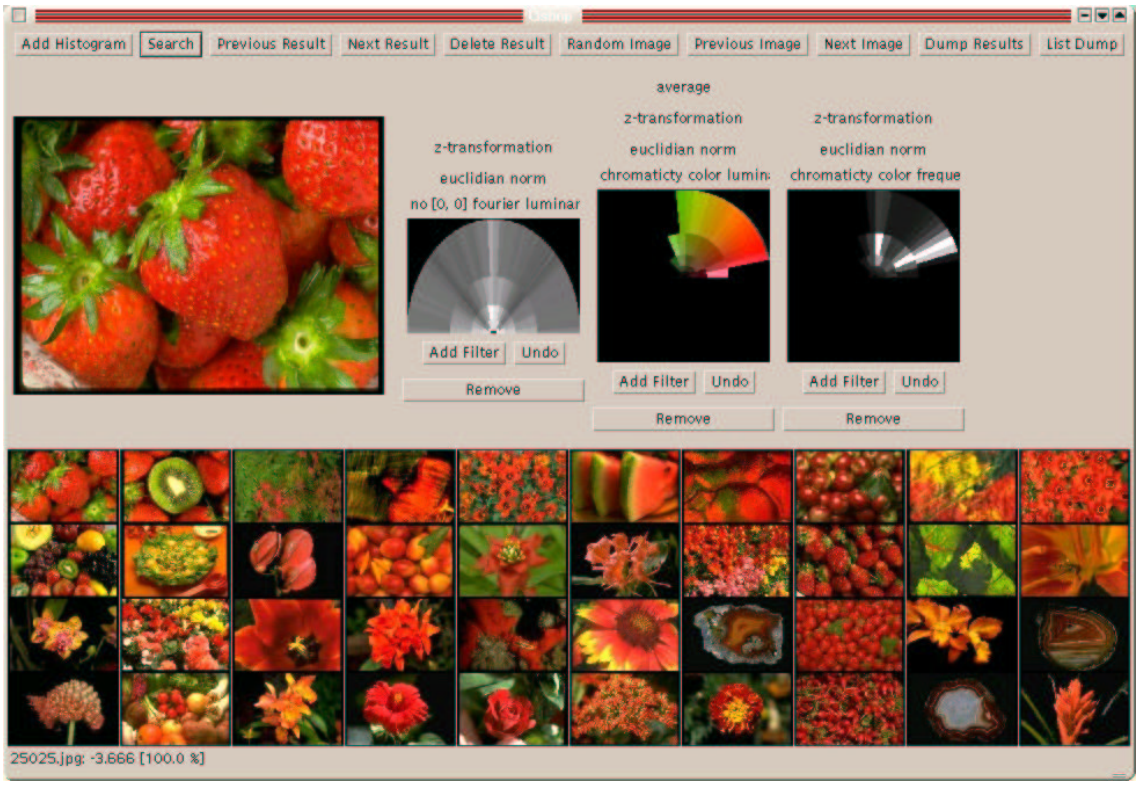


Figure 7: Screenshot of the client applet.

The program can be run either as a standalone program or as a client-server combination. The clients can be run as applets in internet browsers which handle the local query definitions and store the query results. The Java remote method invocation protocol (RMI) is used for client-server com-

munication. The server receives the hierarchical query definitions, searches the feature database for images and returns the results and the image URLs to the clients. The images are retrieved by the client applets from the server or the web.

The process of feature creation is script-based. For the 2-dimensional discrete Fourier transformation and the bilinear rescaling we used the 'hips' image software package (Landy et al., 1984).

The performance of the program depends on the speed of the computer and the implementation of the Java virtual machine. For the database of 60,000 images used in the experiments it took between one and two seconds to find the best 100 matches for a given image on a 750 MHz Pentium-III based computer running the standard Java implementation of Linux. When all three indices were combined a query was processed in less than five seconds. It should be noted that we only used a linear search strategy not taking advantage of possible bounding assumptions and tree-based data structures. The search could be further speeded up by reducing the number of bins using principal component analysis (PCA).

4 Psychophysical Evaluation

The purpose of the psychophysical experiments was to measure the relationship between the similarity norms created by the indices and the perceived similarity. The similarity between two images was interpreted as a probabilistic function and was measured in a two-alternative-forced-choice (2AFC) design. The correspondence between the similarity computations and the judged similarity was measured for the three indices and three index combinations. In contrast to previous evaluation strategies histogram-based similarity was not only evaluated for a few best matches but for relatively distinct images too.

4.1 Measuring Similarity

For precision/recall evaluation methods similarity is implicitly defined by the judgments (a) of the experimenter. The precision (A) of the result set can be written as a function defining which images are relevant or similar for each query image:

$$a : I \times I \rightarrow \{0, 1\}$$

The precision A of a result list D of the m best matching images (images having a rank below than or equal to m) can then be simply defined:

$$A = \frac{1}{m} \sum_{i \in D} a\,(q, i)$$

In contrast, we here define the similarity between two images as a probabilistic function:

$$p : I \times I \rightarrow [0, 1]$$

Figure 8: 2AFC display used in the experiments. At the top the query image is show, below are two test images (target and distractor positions were randomized).

The perceived similarity can be measured using either an absolute or relative scale. Both methods are highly correlated; the relationship between both measures follows the form of a psychometric function (Papathomas et al., 1998). In the relative case the user indicates the degree of similarity between the query image and the test image on a rating scale. This requires the users to adjust their scale to the content of the database prior to the experiment, e.g. by seeing a list of random images. This may increase the standard error of evaluation measurements. A more objective method is the two-alternative forced-choice configuration (2AFC).

We used a configuration with three images: the query image at the top together with two test images below (Fig. 8). The task of the user was to compare the similarity of the test images with the query image and to select the test image that is more similar. We defined the (probabilistic) similarity between the query image and a test image $p\left(q,i\right)$ by the probability of preferring the image $p\left(i\right)$.

We were interested in comparing the judged image similarity $p(q, i)$ with the similarity order of the indices. In the forced choice configuration the rate of preferring an image is dependent on the degree of similarity between the two alternatives. With the exception of experiment one we kept the similarity of one image, the distractor image, constant. The similarity of the other image, the target image, was varied.

For a given query image q an image i can be chosen that is either relatively similar or quite dissimilar to q, using a similarity metric s. If the similarity metric s corresponds to the observers' varying the computed degree of similarity $s(q, i)$ should have an influence on the selection rate. For weak indexing approaches, however, the observed correlation would be low. Therefore, the relationship between the computed similarity $s(q, i)$ and the rate of preferring the image $p(q, i)$ can be used to evaluated the similarity metric s.

4.2 Objectives

Currently, similarity measures were only evaluated for the first best matching images using precision information. But little is known if histogram-based indexing can be used to compare the similarity of relatively distinct images, too. In the experiments we varied the similarity between the target and the query images over a broad range, from best matches to the 2000th image in the result list to test whether the information stored in the indices can predict the similarity judgments of the observers for the different levels of similarity.

The advantage of the 2AFC design over a rating scale is its simplicity. The observers do not need to maintain an internal scale but can directly compare the images. The disadvantage is that the selection probability $p(i)$ is not only dependent on q but also on the alternative d (distractor). If d for instance is a random image the selection rate should be always greater than 0.5. If d

is a quite similar to q than achieving a selection rate greater than 0.5 would be a challenge. Therefore, the aim of the first experiment was the determine how the similiarity judgments $p(i)$ are influenced by either very similar, less similar, or completely random distractor images d.

In a second experiment the three indices were compared. The indices stored distinct information: the chromaticity histogram encodes the frequency of the color bins, the luminance histograms stores average luminance levels but no frequency information, and both color histograms contain no spatial information. From the Fourier index the bins containing average luminance information were removed so that it only stores the orientation and spatial frequency information. We wanted to known which index would be best suitable for image indexing and to what extent the information sources (chromaticity, luminance, spatial) contribute to the perceived similarity.

In the third experiment the indices' information were combined. We wanted to know if the chromaticity histogram can be improved by using the spatial information and whether the luminance index would further improve, or worsen the concordance with the observers' judgments.

4.3 Method

For all experiments a large commercially database (Corel Corp., 1990) of 60,000 digitized photographs was used. It contains a wide range of themes. Each theme consists of 100 images. The images show, for example, natural and man-made objects, landscapes and close-ups, and were photographed under natural illumination conditions or under artificial lighting.

From the database 900 query images were randomly selected for each experiment. For each of these images the 2000 best matching images for each relevant index or index combination were retrieved. The images with the rank

numbers 2, 20, 200 and 2000 were selected for the experiments. In addition, the target image could be a random image. Using the rank numbers rather than the raw distances allows comparing the differently distributed distance functions.

To achieve comparability the distractor was always determined by the color histogram and the intersection metric. With the exception of experiment one where the influence of the distractor similarity was investigated, the rank of the distractor image was always 200. In the first experiment the similarity of distractor images was varied. The similarity rank was 1 (best matches), 200 or random.

The position of the target/distractor (left or right), the order of query images and of the similarity conditions were randomized per subject (mixed design).

The images were displayed on a 21¨ computer monitor (Sony GDM-F500) with a resolution of 1280x1024 pixels on a 50% gray background. The experiments were self-paced without decision time limits and lasted between 45 and 60 minutes. The subjects were undergraduates of psychology who were obliged to participate in experiments for their curriculum and were naive with respect to the experiments. They were instructed to compare the similarity of the two test images with the image at the top and to decide which image is more similar. Subjects were told that there is no “true or false” and to judge intuitively if unsure. The answers were giving by clicking the left or right mouse button.

4.4 Experiment 1: Influence of distractor similarity

In the first experiment the relationship between the similarity computed with the chromaticity histogram and the human judgments was measured. As a second factor, the similarity of the distractor image was varied to determine whether the degree of similarity between the two test images would influence the rate of preference for the target image.

The image similarities were calculated using the chromaticity histogram and the intersection norm. The rank of the distractor image was either 1 (best match), 200, or random. The rank of the target image was 1, 2, 20, 200, 2000, or random. This results in 3×5 conditions because the points of equal target and the distractor ranks were not measured.

The results show a strong correlation between the logarithmic rank of the test image and the probability of preference. In figure 9 the similarity rank of the target image is plotted at the x-axis. For each of the three distractor ranks, the rate of preferring the target is shown on the y-axis. The hypothetical points of equal similarity between the target and the distractor image are marked by filled circles. The correlation coefficient between the logarithmic rank and the percentage of concordance is very high ($r > .95$) and significant ($p < .05$). Table 1 shows the regression and correlation coefficients.

Distractor Rank	b	a	r
1	-.022	.49	.985
200	-.032	.68	.999
Random	-.027	.80	.965

Table 1: Regression and correlation coefficients for the relationship between logarithmic target rank and concordance for different distractor ranks.

Changing the distractor similarity does not alter the gradient of the linear regression and only shifts the functions by a constant amount. Therefore, the alternative image in the 2AFC design is not of critical importance for the

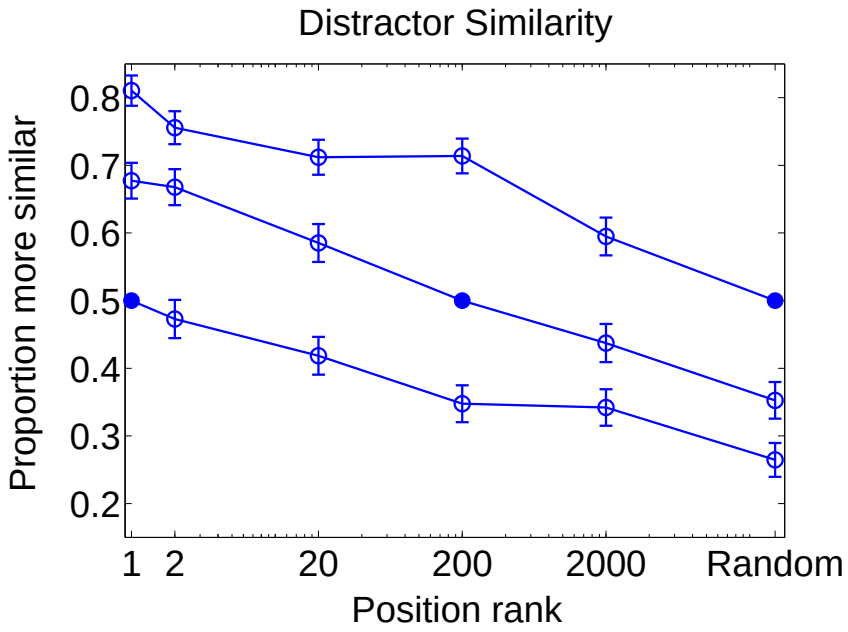


Figure 9: Probability of preferring the target image against the distractor image as a function of similarity between the target and the query image. The similarity distance was calculated using the chromaticity index and the intersection metric. The filled circles indicate hypothetic points of equal similarity of distractor and target image with expected probability .5 and therefore define the rank of the distractor image. The standard error is shown for each condition.

evaluation of the indices. For the following experiments we kept the distractor similarity constant. We decided to use the medium distractor similarity (rank 200) to allow an optimal use of the range of the scale.

4.5 Experiment 2: Index comparison

Experiment 1 showed that the computed chromaticity-based similarity correlates with the observers' judgments. In experiment 2 we wanted to know, to what extend the other information sources (luminance, orientation and spation frequency) contribute to the similarity judgments.

The distractor image was determined by the chromaticity histogram and had rank 200. The rank of the target image was again 1, 2, 20, 2000, or random. For the chromaticity histogram the minimum norm was used. The distances between the other vectors were calculated with the Euclidian norm.

The results show a strong advantage for the chromaticity histogram. For relatively distinct images the luminance and spatial index show similar performance (rank 20, 2000). For highly similar images (rank 1,2), the luminance index provided better matches.

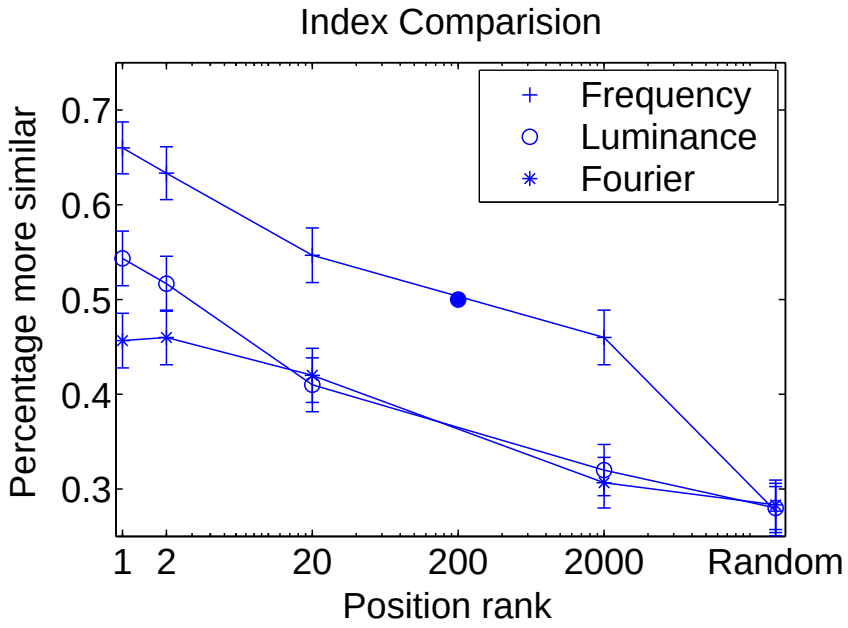


Figure 10: Relationship between the probability of preferring the target image against the distractor image as a function the similarity between query and target image. The similarity distances were calculated using the chromaticity, luminance and the spatial index. The filled circle indicates the theoretic point where the target and the distractor image are equally similar.

Index	b	a	r
Color Frequency	-.034	.66	.976
Luminance	-.025	.52	.978
Fourier spectrum	-.019	.47	.989

Table 2: Regression and correlation coefficients for the relationship between logarithmic target rank and concordance for the different indices.

4.6 Experiment 3: Index combination

The performance of the luminance and spatial index is clearly worse than that of the chromaticity histogram. However, we wanted to know if the correspondence with the observers' judgments could be improved if the spatial or the spatial and the luminance information were used in addition to the color frequencies.

We used the average of the z-transformed distance values as our similarity function. All other conditions were identical to the ones in experiment two.

Figure 11 shows a general improvement of the selection rates of the image retrieved when either spatial or spatial and luminance information is combined with the chromaticity histogram. The correspondence with the judgments was improved by 5% if the spatial information is used (Table 3). The luminance information further enhances the concordance by 4%. Therefore, the luminance and the spatial index contribute to image similarity independent of the chromaticity histogram.

The overall improvement of 9% clearly indicates that the z-transformation is a good choice for combining index information.

Index Combination	b	a	r
Freq.	-.021	.60	.967
Freq. + Fourier	-.028	.65	.959
Freq. + Fourier + Lum.	-.031	.69	.998

Table 3: Regression and correlation coefficients for the relationship between logarithmic target rank and concordance for combinations of the indices.

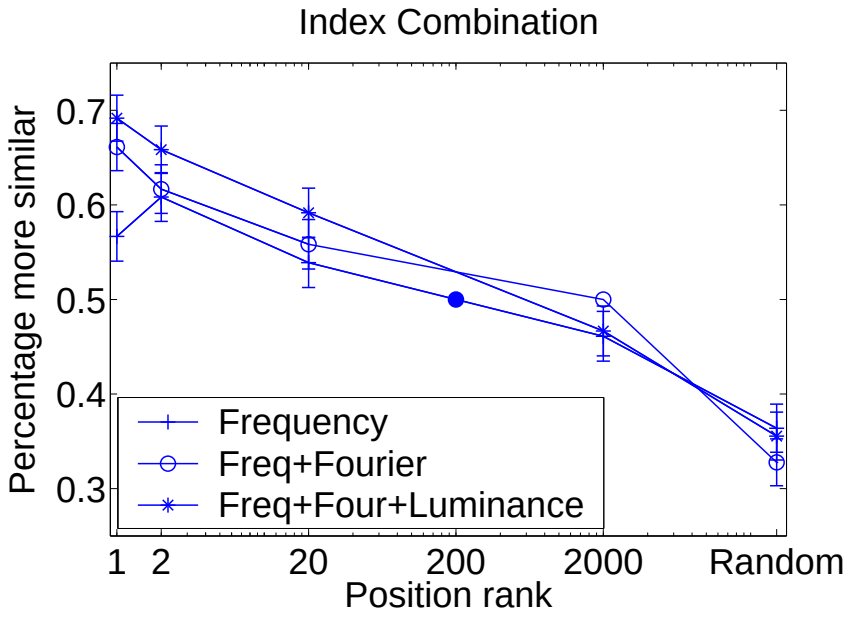


Figure 11: Relationship between the probability of preferring the target image against the distractor image as a function of the similarity between the query and the target image. The similarity distances were calculated using the chromaticity, the chromaticity and the spatial and all three indices. The filled circle indicates the theoretic point where the target and distractor are equally similar. The standard error is shown for each data point.

5 Discussion

5.1 Perception Based Image Indexing

Color histograms and their variations are used with great success in many current CBIR systems. The histograms are typically built in RGB or HSI space. For objects that are photographed under different illuminations, color constancy algorithms can improve the retrieval quality (Funt and Finlayson (1995), Gevers and Smeulders (1996), Alferez and Wang (1999)). The spatial information of images was encoded a variety of ways: by multiple histograms (Stricker and Dimai, 1997), correlograms (Huang et al., 1999), explicit texture statistics (Liu and Picard, 1996) and by wavelet transformations (Manjunath and Ma (1996), Wang et al. (1997), Liang and Kuo (1999)).

Most of these features were constructed to optimally represent some, in most cases statistical properties of images. We created a chromaticity histogram in analogy to processing of color in the human vision system. The color-opponent code in the optical nerve is a nearly optimal representation of color because it removes the correlation between the channels (Buchsbaum and Gottschalk, 1983). One could argue that optimal representations for image colors would be better constructed by principal components or cluster analyses. However, in the end the similarity of the images is always judged by human observers. Therefore, using a psychophysical color space should result in an optimized metric which better corresponds to human similarity judgments.

In addition to the frequency encoding chromaticity histogram, we stored the average luminance levels of the color bins for each image in a second vector. The separate processing of chromaticity and luminance (contrast) information is part of various models of visual processing. We used the separa-

tion of chromaticity and luminance information to compare the contribution of both information sources to the similarity judgments. We found that the luminance index is correlated with the observers' judgments, too. The prediction of the perceived similarity is not as good as with the chromaticity histogram. But luminance appears to be considered when images are compared, even if the luminance of the color tones seems not to be as important as their frequency. Illumination invariant indices may therefore only be superior if the database contains many similar objects photographed under various illumination conditions.

The spatial information was extracted using the 2-dimensional DFT. The index was constructed to represent the distribution of orientation and spatial frequency in the image in analogy to the processing of contrast information in the visual cortex. Although many texture feature sets describe similar information, Fourier analysis offers a mathematically profound way to extract the information. The explicit representation of orientation and spatial frequency in the Fourier spectrum allows the simple filtering of these dimensions. In our program the user could for example select a radial filter to use the Fourier index as an orientation index only. By averaging the bins of equal frequency ranges a rotation invariant search could be conducted.

5.2 Evaluation

Currently, CBIR systems are evaluated by precision and sometimes recall measures. Relevant images are – in the majority of the publications – defined by similarity judgments of a single person deciding whether an image belongs to the same category as the query image. The image databases used for evaluation typically contain between hundred and some thousand images.

A very well evaluated system is the PicHunter system (Cox et al. (2000),

Papathomas et al. (1998)). The system includes simple features like the image height and width, color histograms, color autocorrelogram, a color-coherence vector and – for a subset of the images – semantic annotations. The features were tested in a forced choice design similar to one we used and then later compared within the target testing paradigm (Cox et al., 2000).

However, the target testing evaluation procedure used in Papathomas et al. (1998) is limited to relevance feedback systems. It is ideal to evaluate a complete system but it is less ideal to measure the relationship between single features and human perception. With target testing the functional relationship between an index similarity space and the similarity space of human perception cannot determined.

Cox et al. (2000) used a sigmoid function to transform the raw distances to a probabilistic scale. Our results suggest that the relationship could be modeled by a logarithmic function between the similarity rank and the selection probability for 2AFC designs. That relationship holds for all indices and is not impacted by the similarity of the distractor image.

For relevance feedback systems it is highly important to measure the relationship between the similarity metric and the perceived similarity over the whole range of similarity, from the most similar image to random images.

5.3 Conclusion

In summary, we have shown that the psychophysically motivated indices are very effective in finding similar images. The indices were constructed in accordance with some of the known properties of the early stages of human vision. The color codes in the “red-green” and “blue-yellow” channels were modeled using the color-opponent axes of the DKL color space and a logarithmic-radial scaling for the histogram bins. The luminance information was stored in a

separate index. The 2-dimensional Discrete Fourier Transformation was used to create an orientation and spatial frequency histogram in analogy to similar representations in the visual cortex. Finally, we evaluated the Color Indexing System Based On Perception (CISBOP) with a strictly qualitative and objective approach using a large, heterogeneous database of 60,000 digitized photographs.

Affiliation

Justus Liebig University Giessen

Psychology Deptartment

Otto-Behaghel-Strasse 10F

D-35394 Giessen

GERMANY

References

Ronald Alferez and Yuan-Fang Wang. Geometric and illumination invariants for object recognition. *IEEE Trans. PAMI*, 21(6):505–536, June 1999.

G. Buchsbaum and A. Gottschalk. Trichromacy opponent color coding and color transmission in the retina. In *Proc. Roy. Soc. Lond. (B)*, volume 220, pages 89–113, 1983.

Corel Corp. Corel stock photographic library, 1990.

I. Cox, M. Miller, T. Minka, T. Papathomas, and P. Yianilos. The bayesian image retrieval system, pichunter: theory, implementation and psychological experiments. *IEEE Trans. Image Processing*, 9:20–37, 2000.

A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macque. *J. Physiol.*, 357:241–265, 1984.

D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Amer.*, 4 (12):2379–2394, 1987.

M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query

by image and video content: The qbic system. *IEEE Computer Magazine*, 28:23 – 32, 1995.

Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *IEEE Trans. PAMI*, 15(5):522–529, May 1995.

K.R. Gegenfurtner and D.C. Kiper. Color vision. *Annual Review of Neuroscience*, in press.

T. Gevers and A. Smeulders. A comparative study of several color models for color image invariant retreival. In *Proc. 1st Int. Workshop on Image Databases and Multimedia Search*, pages 17–, Amsterdam, Netherlands, 1996.

J. Huang, SR Kumar, M Mitra, WJ Zhu, and R Zabih. Spatial color indexing and applications. *International Journal of Computer Vision*, 35 (3):245–268, 1999.

D.H. Hubel and T.N. Wiesel. Early exploration of the visual cortex. *Neuron*, 20(3):401–412, 1998.

J. Krauskopf, D. R. Williams, and D. W. Heeley. Cardinal directions of color space. *Vision Res.*, 22:1123–1131, 1982.

John Krauskopf and Karl R Gegenfurtner. Color discrimination and adaptation. *Vision Res.*, 32 (11):2165–2175, 1992.

Michael S. Landy, Yoav Cohen, and George Sperling. Hips: a Unix-based image processing system. *Computer Vision, Graphics, and Image Processing*, 25(3):331–347, March 1984. ISSN 0734-189X.

Kai-Chieh Liang and C.-C. Jay Kuo. Waveguide: A joint wavelet-based image

representation and description system. *IEEE Trans. Image Processing*, 8 (11), 1999.

F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. PAMI*, 18(7):517–549, 1996.

B. S. Manjunath and M. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. PAMI*, 18(8):837–842, August 1996.

T. Papathomas, T. Conway, I. Cox, J. Ghosn, M. Miller, T. Minka, and P. Yianilos. Psychophysical studies of the performance of an image database retrieval system, 1998.

Bernice E. Rogowitz, Thomas Frese, John R. Smith, Charles A. Bouman, and Edward Kalin. Perceptual image similarity experiments. *Human Vision and Electronic Imaging III*, 1998.

Daniel L. Ruderman, Thomas W. Cronin, and Chuan-Chin Chiao. Statistics of cone responses to natural images: implications for visual coding. *J Opt Soc America A*, 15:2036–2045, 1998.

M. Stricker and A. Dimai. Spectral covariance and fuzzy regions for image indexing. *Machine Vision and Applications*, 10 (2):66–73, 1997.

B.A. Wandell. Color appearance: The effects of illumination and spatial resolution. In *Proc. Nat. Acad. Sci.*, volume 90, pages 1494–1501, 1993.

James Ze Wang, Gio Wiederhold, Oscar Firschein, and Sha Xin Wei. Content-based image indexing and searching using daubechies' wavelets. *Int. J. Digit. Libr.*, 1:311–328, 1997.

Andrew B. Watson. Efficiency of a model human image code. *J. Opt. Soc. Am.*, 4(12):2401–2417, 1987.

H. Wässle and B. B. Boycott. Functional architecture of the mammalian retina. *Physiol Rev*, 71:447–480, 1991.