Evaluating Color and Shape Invariant Image Indexing of Consumer Photography

T. Gevers and A.W.M. Smeulders
Faculty of Mathematics & Computer Science, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
E-mail: gevers@fwi.uva.nl

Abstract
In this paper, indexing is used as a common framework to represent, index and retrieve images on the basis of color and shape invariants.
To evaluate the use of color and shape invariants for the purpose of image retrieval, experiments have been conducted on a database consisting of 500 images of multicolored objects. Images in the database show a considerable amount of noise, specularities, occlusion and fragmentation resulting in a good representation of views from everyday life as it appears in home video and consumer photography in general.
The experimental results show that image retrieval based on both color and shape invariants provides excellent retrieval accuracy. Image retrieval based on shape invariants yields poor discriminative power and worst computational performance whereas color based invariant image retrieval provides high discrimination power and best computational performance.
Furthermore, the experimental results reveal that identifying multicolored objects entirely on the basis of color invariants is to a large degree robust to partial occlusion and a change in viewing position.

1 Introduction

For the management of archived image data, an image database (IDB) system is needed which supports the analysis, storage and retrieval of images. Over the last decade, much attention has been paid to the problem of combining spatial processing operations with DBMS capabilities for the purpose of storage and retrieval of complex spatial data. Most image database systems are still based on the paradigm to store a key-word description of the image content, created by some user on input, in a database in addition to a pointer to the raw image data. Image retrieval is then based on the standard DBMS capabilities.

A different approach is required when we consider the wish to retrieve images by image example, where a query image or sketch of image segments is given by the user on input. Then, image retrieval is the problem of identifying a query image as a part of target images in the image database and when identification is successful establishing a correspondence between query and target.

The basic idea to image retrieval by image example is to extract characteristic features from target images which are matched with those of the query. These features are typically derived from shape, texture or color properties of query and target. After matching, images are ordered with respect to the query image according to their similarity measure and displayed for viewing [7], [9], [10].

The matching complexity of image retrieval by image example is similar to that of traditional object recognition schemes. In fact, image retrieval by image example shares many characteristics with model-based object recognition. The main difference is that model-based object recognition is done fully automatically whereas user intervention may be allowed for image retrieval by image example. To reduce the computational complexity of traditional matching schemes, the *indexing* or *hashing* paradigm has been proposed (for example [1], [12], [13], [15]) only recently. Because indexing avoids exhaustive searching, it is a potentially efficient matching technique. A proper indexing technique will be executed at high speed allowing for real-time image retrieval by example image. This is the case even when the image database is large as may be anticipated for multimedia and information services.

The type of indices used in existing indexing schemes are either based on geometric (shape) or on photometric (color) properties. Shape based indexing schemes, for instance [1] and [15], used indices constructed from geometric invariants which are independent of a given coordinate transformation such as similarity and affine. Because of the extensive research on shape-based indexing, we concentrate on [1], [5], [11], [12] and [15] in this paper. In general, shape invariants are computed from object features extracted from images such as intensity region outlines, edges and high curvature points. However, intensity region outlines, edges or high curvature do not necessarily correspond to material boundaries. Objects cast shadows which appear in the scene as an intensity edge without a material edge to support it. In addition, the intensity of solid objects varies due to surface orientation change yielding prominent intensity edges when the orientation of the surface changes abruptly. In this way, a large number of accidental feature points are introduced. In this paper, viewpoint independent photometric image descriptors are devised making shape based indexing more appealing.

As opposed to geometric information, other indexing schemes are entirely on the basis of colors. Swain and Ballard [13] propose a simple and effective indexing scheme using the color at pixels in the image directly as indices. If the RGB (or some linear combination of RGB) color distributions of query and target image are globally similar, the matching rate is high. Swain's color indexing method has been extended by [6] to become illumination independent by indexing on an illumination-invariant set of color descriptors. The color based schemes fail, however, when images are heavily contaminated by shadows, shading and highlights. Also, although these color based schemes are insensitive to translation and rotation, they will be affected negatively by other practical coordinate transformations such as similarity, affine and projection.

In this paper, indexing is used as a common framework to index and retrieve images on the basis of photometric and geometric invariants. The indexing scheme makes use of *local* photometric information to produce *global* geometric invariants to obtain a viewpoint invariant, highly-dimensional *similarity* descriptor to be used as an index. No constraints are imposed on the images in the image database and the camera image process other than that images should be taken from multicolored objects illuminated by a single type of source.

This paper is organized as follows. Viewpoint invariant color descriptors are proposed in Section 2. In Section 3 geometric invariants are discussed. Image indexing and retrieval is discussed in Section 4 and 5. The experimental results are given in Section 6.

2 Photometric Invariant Indexing Functions

In this paper, we concentrate on *color* images. In this context, photometric invariants are defined as functions describing the *local color* configuration of each image coordinate (x, y) by its color and colors of its neighboring pixels but discounting shading, shadows and highlights.

In this section, a quantitative viewpoint-independent color feature is discussed first on which the color invariants will be based.

2.1 Viewpoint-independent Color Feature

To obtain a quantitative viewpoint-independent color descriptor, a color feature is required independent of object's surface shape and viewing geometry discounting shading, shadows and highlights. It is shown that hue is a viewpoint independent color feature for both the Phong and the Torrance-Sparrow reflection model [8]. The well-known standard color space $L^*a^*b^*$, perceptually uniform and possesses a Euclidean metric [16], is used to compute the hue. Let the color image be represented by R, G, and B images. Then the RGB values at image location (x, y) are first transformed into CIE XYZ values:

$$X(x,y) = a_1 R(x,y) + b_1 G(x,y) + c_1 B(x,y) \tag{1}$$

$$Y(x,y) = a_2 R(x,y) + b_2 G(x,y) + c_2 B(x,y) \tag{2}$$

$$Z(x,y) = a_3 R(x,y) + b_3 G(x,y) + c_3 B(x,y) \tag{3}$$

where a_i, b_i and c_i, for $i = (1,2,3)$ are camera dependent.

Then the a^*b^* values are:

$$a^*(x,y) = 500 \left[(\frac{X(x,y)}{X_0})^{\frac{1}{3}} - (\frac{Y(x,y)}{Y_0})^{\frac{1}{3}} \right] \tag{4}$$

$$b^*(x,y) = 200 \left[(\frac{Y(x,y)}{Y_0})^{\frac{1}{3}} - (\frac{Z(x,y)}{Z_0})^{\frac{1}{3}} \right] \tag{5}$$

where X_0, Y_0 and Z_0 are X, Y and Z values for the reference white, respectively.

From the a^*b^* values, we get the hue value:

$$H(x,y) = \arctan(\frac{b^*(x,y)}{a^*(x,y)}) \tag{6}$$

Hence, $H(x,y)$ denotes the hue value at image coordinates (x,y) ranging from $[0,2\pi]$.

2.2 Points

A simple invariant labeling function is defined as a function which measures the hue at coordinate (x,y):

$$l_p(x,y) = H(x,y) \tag{7}$$

In this way, each location (x,y) in a color image is given an invariant value ranging from $[0,2\pi]$.

The indexing scheme could be used directly with the hue serving as color invariant. However, geometric invariants are computed from the coordinates of color invariants. Also, all spatial color information will be lost. To add some local spatial color information, color invariant functions are proposed by considering the local topographic edge configuration of $H(x,y)$ such as hue edges and corners.

These will be described in 2.3 and 2.4 respectively.

2.3 Edges

In this section, each hue edge point is given a label based on the hue-hue transition at that location.

Unlike intensity, hue is defined on a ring ranging from $[0,2\pi]$. The standard difference operator is not suitable to compute the difference between hue values, because a low and a high hue value produce large difference although they are positioned nearby on the ring. Due to the wrap-around nature of hue, we define the angular distance between two hue values h_1 and h_2 as follows:

$$d(h_1,h_2) = \arccos(\cos h_1 \cos h_2 + \sin h_1 \sin h_2) \tag{8}$$

yielding the relative angle between h_1 and h_2.

To find hue edges in images we follow the method of Canny [2] where, instead of the standard difference function, subtractions are defined by (8). Let $G(x,y,\sigma)$ denote the Gaussian function with scale parameter σ. The partial derivatives of $H(x,y)$ are given by:

$$\nabla H(x,y) = \begin{pmatrix} H_x \\ H_y \end{pmatrix} = \begin{pmatrix} H(x,y) \otimes G_x(x,y,\sigma) \\ H(x,y) \otimes G_y(x,y,\sigma) \end{pmatrix} \tag{9}$$

where $\otimes$ denotes the angular convolution operator. The gradient magnitude is represented by:

$$||\nabla H(x,y)|| = \sqrt{{H_x}^2 + {H_y}^2} \tag{10}$$

After computing the gradient magnitude based on the angular distance, a non maximum suppression process is applied to $||\nabla H(x,y)||$ to obtain local maximum gradient values:

$$\mathcal{M}(x,y) = \begin{cases} ||\nabla H(x,y)||, & \text{if } (||\nabla H(x,y)|| > t_\sigma) \text{ is a local maximum} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where t_σ is a threshold based on the noise level.

Let the set of image coordinates of local edge maxima be denoted by $E = \{(x,y) \in H : \mathcal{M}(x,y) > 0\}$.

Then, for each local maximum, two neighboring points are computed based on the direction of the gradient to determine the hue value on both sides of the edge:

$$l_e^l(\vec{x}) = \begin{cases} H(\vec{x} - \Delta\vec{n}), & \text{if } \vec{x} \in E \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

$$l_e^r(\vec{x}) = \begin{cases} H(\vec{x} + \Delta n), & \text{if } \vec{x} \in E \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

over E where $\vec{n}$ is normal of gradient at $\vec{x}$ and Δ is a preset value.

Because values of l_e^l and l_e^r interchange due to the orientation of the edge, a unique non ordered hue-hue transition function is defined by:

$$l_e(\vec{x}) = \begin{cases} l_e^r(\vec{x}) + \max(H) l_e^l(\vec{x}), & \text{if } l_e^r(\vec{x}) \leq l_e^l(\vec{x}) \\ l_e^l(\vec{x}) + \max(H) l_e^r(\vec{x}), & \text{otherwise} \end{cases} \tag{14}$$

where $\max(H) = 2\pi$ denotes the maximum hue.

The invariant is quantitative, non geometric and viewpoint-independent and can be derived from any view of a planar or 3D multicolored object. The color invariant provides powerful discrimination as for a resolution of 1 out of 100 there are 10000 combinations of hue pairs which may identify the object with the big advantage that hue is viewpoint independent.

2.4 High Curvature Points

In this paper, the measure of cornerness is defined as the change of gradient direction along an edge contour [4]:

$$\kappa(H(x,y)) = \frac{-H_y^2(x,y)H_{xx}(x,y) + 2H_x(x,y)H_y(x,y)H_{xy}(x,y) - H_x^2(x,y)H_{yy}(x,y)}{(H_x^2(x,y) + H_y^2(x,y))^{3/2}} \tag{15}$$

To isolate high curvature points, $\kappa(H(x,y))$ is multiplied with $\mathcal{M}(x,y)$:

$$\mathcal{C}(x,y)) = \kappa(H(x,y))\mathcal{M}(x,y) \tag{16}$$

Let the set of image coordinates of high curvature maxima be denoted by $C = \{(x,y) \in H : \mathcal{C}(x,y) > 0\}$.

Then, for $\vec{x} \in C$ two neighboring points are computed based on the direction of the gradient to determine the hue value on either side of the high curvature point yielding the hue-hue transition at $\vec{x}$:

$$l_c^l(\vec{x}) = \begin{cases} H(\vec{x} - \Delta\vec{n}), & \text{if } \vec{x} \in C \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

$$l_c^r(\vec{x}) = \begin{cases} H(\vec{x} + \Delta\vec{n}), & \text{if } \vec{x} \in C \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

over C where $\vec{n}$ is normal of gradient at $\vec{x}$ and Δ is a preset value.

Further,

$$l_c(\vec{x}) = \begin{cases} l_c^r(\vec{x}) + \max(H) l_c^l(\vec{x}), & \text{if } l_c^r(\vec{x}) \leq l_c^l(\vec{x}) \\ l_c^l(\vec{x}) + \max(H) l_c^r(\vec{x}), & \text{otherwise} \end{cases} \tag{19}$$

3 Geometric Invariant Indexing Functions

In this section, geometric invariants are discussed measuring geometric properties between a set of coordinates of an object in an image independent of a given coordinate transformation. These are called *algebraic invariants*. We restrict ourselves to coordinates coming from planar objects. However, many man-made objects can often be decomposed into planar sub objects.

Euclidean and projective invariants are discussed in the sequel.

3.1 Euclidean Invariant

It is known that when an object is transformed rigidly by rotation and translation, then its length is an invariant.

For image locations (x_1, y_1) and (x_2, y_2), $g_E()$ is defined as a function which is unchanged as the points undergo any two-dimensional Euclidean transformation leaving to our first geometric invariant indexing function:

$$g_E((x_1,y_1),(x_2,y_2)) = \sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)} \tag{20}$$

3.2 *Projective Invariant*

For the projective case, geometric properties of the shape of a planar object should be invariant under a change in the point of view. From the classical projective geometry we have that the cross ratio of sines between five points on a plane is a projective invariant [14].

For cases where projective invariants are of importance, the projective invariant function $g_P()$ is defined as:

$$g_P((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4),(x_5,y_5)) = \frac{\sin(\theta_1+\theta_2)\cdot\sin(\theta_2+\theta_3)}{\sin(\theta_2)\cdot\sin(\theta_1+\theta_2+\theta_3)} \tag{21}$$

where $\theta_1, \theta_2, \theta_3$ are the angles at (x_1,y_1) between $(x_1,y_1)(x_2,y_2)$ and $(x_1,y_1)(x_3,y_3)$, $(x_1,y_1)(x_3,y_3)$ and $(x_1,y_1)(x_4,y_4)$, $(x_1,y_1)(x_4,y_4)$ and $(x_1,y_1)(x_5,y_5)$ respectively.

4 Indexing

Let the image database consist of a set $\{I_k\}_{k=1}^{N_b}$ of color images. Histograms are created to represent the distribution of quantized invariant indexing function values in a multidimensional invariant space. Histograms are formed on the basis of:

Color Invariants:

Color invariants listed in Section 2 are computed from the images in the image database. These invariants are quantitative, non geometric and viewpoint-independent and can be derived from any view of a planar or 3D multicolored object. By using hue at image location (x,y) directly as an index, the histogram defined by:

$$\mathcal{G}_A^{I_k}(i) = \sum_{(x,y)\in H} z = \begin{cases} 1, & \text{if } l_p(x,y) = i \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

represents the distribution of hue values in an image, where $l_p()$ is given by equation (7).

Instead of considering only the hue at (x,y), the histogram representing the distribution of hue-hue transitions is given by:

$$\mathcal{G}_B^{I_k}(i) = \sum_{(x,y)\in E} z = \begin{cases} 1, & \text{if } l_e(x,y) = i \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where E is the set of coordinates of local edge maxima and $l_e()$ is defined by equation (14).

The histogram of hue-hue corners is given by:

$$\mathcal{G}_C^{I_k}(i) = \sum_{(x,y)\in C} z = \begin{cases} 1, & \text{if } l_c(x,y) = i \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

where C is the set of coordinates of corners and $l_c()$ is given by (19).

Shape Invariants:

Secondly, shape based invariant histograms are constructed. Euclidean and projective invariants, presented in Section 3, are computed. Although both geometric invariants are qualitative and geometric, the projective invariant can be derived from any image projection of a planar object whereas the Euclidean invariant requires orthogonal projection of the planar object with fixed distance to the camera. To reduce the number of coordinates, the set C of corner coordinates is taken from which geometric invariants are computed.

The histogram expressing the distribution of distances between hue corners is given by:

$$\mathcal{G}_D^{I_k}(i) = \sum_{(x_1,y_1),(x_2,y_2)\in C} z = \begin{cases} 1, & \text{if } g_E((x_1,y_1),(x_2,y_2)) = i \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

where $g_E()$ is given by equation (20).

In other words, between each pair of corner coordinates, the Euclidean distance denoted by i is computed and used as an index. In a similar way, the distribution of cross ratios between corners is given by:

$$\mathcal{G}_E^{I_k}(i) = \sum_{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4),(x_5,y_5)\in C} z = \begin{cases} 1, & \text{if } g_P((x_1,y_1),(x_2,y_2),(x_3,y_3), \\ & \qquad (x_4,y_4),(x_5,y_5)) = i \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

and $g_P()$ is defined by (21).

Color and Shape Invariants:

A 3-dimensional histogram is created counting the number of corner pairs with labels i and j which are at distance k from each other:

$$\mathcal{G}_F^{I_k}(i,j,k) = \sum_{(x_1,y_1),(x_2,y_2)\in C} z = \begin{cases} 1, & \text{if } l_c(x_1,y_1) = i, \\ & \text{if } l_c(x_2,y_2) = j, \\ & \text{if } g_E((x_1,y_1),(x_2,y_2)) = k \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

All histograms are precomputed when an image is stored into the database.

5 Retrieval

The object to be retrieved is acquired directly from a color image. The advantage of this is that the geometric and color configuration of the object including manufacturing artifacts is expressed immediately through its color image. For image retrieval based on color and/or projective invariants, the query image can be taken from an isolated multicolored object from any viewpoint. However, for the Euclidean invariant, which is viewpoint-dependent, the image is taken orthographically from a planar object with fixed distance to the camera.

Color-metric and geometric invariants are computed from query image $\mathcal{Q}$ and used to create the query histogram $\mathcal{G}^{\mathcal{Q}}$. Then, $\mathcal{G}^{\mathcal{Q}}$ is matched against the same type of histogram stored in the database. Matching is expressed by:

$$\mathcal{H}(\mathcal{G}_j^{\mathcal{Q}}, \mathcal{G}_j^{I_i}) = \frac{\sum_{k=1}^{N_{d_j}} \min\{\mathcal{G}_j^{\mathcal{Q}}(\vec{s}_k), \mathcal{G}_j^{I_i}(\vec{s}_k)\}}{N_{q_j}} \tag{28}$$

where $\mathcal{G}_j^{\mathcal{Q}}$ and $\mathcal{G}_j^{I_i}$, for $j = \{A,B,C,D,E,F\}$, are histograms of type j derived from $\mathcal{Q}$ and image I_i respectively. N_{q_j} is the number of invariant index values derived from $\mathcal{Q}$ yielding N_{d_j}, $1 \leq N_{d_j} \leq N_{q_j}$, nonzero bins in $\mathcal{G}_j^{\mathcal{Q}}$.

Histogram matching requires time proportional to $O(N_b \cdot N_d)$.

After matching, images are ranked with respect to their proximity to the query image.

6 Experiments

To evaluate color and shape based invariant indexing, the following issues will be addressed in this section:

- The discriminative power and computational complexity of color invariant image index.
- The discriminative power and computational complexity of shape invariant image index.
- The discriminative power and computational complexity of color and shape invariant index.
- The effect of occlusion and viewpoint.

The datasets on which the experiments will be conducted are described in Section 6.1. Error measures and performance criteria are given in 6.2 and 6.3 respectively. The discriminative power of each of the indices is evaluated with respect to the performance criteria. Finally, the performance of the different image indices are compared.

In the experiments, we set $\sigma = 1.0$ for the Gaussian derivative operator and all pixels in a color image were discarded with saturation below 15 (this number was empirically determined) before the calculation of hue because hue becomes unstable when saturation is low [3]. Consequently, grey-value parts of objects or background recorded by the color image will not be considered in the histogram matching process.

6.1 Datasets

The database consists of 500 images of domestic objects, tools, toys, food cans, art artifacts etc., all taken from two households. Objects were recorded in isolation with the aid of a low cost color camera in 3 RGB colors. The digitization was done in 8 bits per color. Objects were recorded against a white cardboard background. Two light sources of average day-light color are used to illuminate the objects in the scene. There was no intention to individually control focus or illumination. Objects were recorded at a pace of a few shots a minute. They show a considerable amount of noise, specularities, occlusion and fragmentation. As a result, recordings are best characterized as snap shot quality, a good representation of views from everyday life as it appears in home video, the news, consumer photography in general.

A second, independent set (the query set) of recordings was made of randomly chosen objects already in the database. These objects, 70 in number, were recorded again with arbitrary position and orientation with respect to camera (some upside down, some rotated) compared to previous recordings.

6.2 Error Measures

A match between an image from the query set and an arbitrary image from the database is defined by equation (28) of Section 5. For a measure of match quality, let rank r^{Q_i} denote the position of the correct match for query image Q_i, $i = 1, ..., 70$, in the ordered list of 500 match values, where $r^{Q_i} = 1$ denotes a perfect match.

Then, the average ranking percentile is defined by:

$$\overline{r} = \left(\frac{1}{70} \sum_{i=1}^{70} \frac{500 - r^{Q_i}}{500 - 1}\right) \cdot 100 \tag{29}$$

Furthermore, the number of query images yielding the same rank k is given by:

$$\eta(r^{Q_i} = k) = \sum_{i=1}^{i=70} z = \begin{cases} 1, & \text{if } (r^{Q_i} = k) \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

and the percentage of query images producing a rank smaller or equal to j is:

$$\mathcal{X}(j) = \left(\frac{1}{70} \sum_{k=1}^{j} \eta(r^{Q_i} = k)\right) \cdot 100 \tag{31}$$

Let N^{Q_i} be the number of different index values derived from query image Q_i. Then the average number of different index values $\overline{N} = \frac{1}{70} \sum_{i=1}^{70} N^{Q_i}$, determines the computational complexity during histogram matching $O(N_b \cdot \overline{N})$, where $N_b = 500$.

6.3 Performance Criteria

Good performance is achieved when the recognition rate is high and the computational complexity is low. To that end, the following criteria should be minimized:

- *the average ranking percentile:* $1 - \overline{r}$ (the discrimination power).
- *the average number of different invariant values:* $\overline{N}$ (time and storage complexity).

6.4 Histogram Binning

First, we will determine the appropriate bin size for color invariant histograms in Section 6.5. As stated, color invariants are based on hue. It is reasonable to assume that hue values appear with equal likelihood in the images. Therefore, hue is partitioned uniformly with fixed intervals. We will determine the appropriate bin size for our application empirically by varying the number of bins on the hue axis over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ and choose the q for which the performance criteria are met.

Second, optimal bin sizes will be determined for shape invariant histograms in Section 6.6. Although distances and cross ratios do not appear with equal likelihood, fixed intervals will be used for the ease of illustration. The appropriate bin size will be determined empirically by varying the number of bins over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$.

As will be seen in Section 6.5 and 6.6, the number of bins is of little influence on the retrieval accuracy when the number of bins ranges from $q = 16$ to $q = 256$.

6.5 *Color Invariant Image Index*

In this subsection, we report on the performance of the indexing scheme for the 70 query images on the database of 500 images on the basis of only color invariants alone. Attention is focussed on histogram matching based on the following histograms: $\mathcal{G}_A^{I_i}$ is the distribution of hue values in I_i and $\mathcal{G}_B^{I_i}$ and $\mathcal{G}_C^{I_i}$ give the distribution of respectively hue-hue edges and corners as defined in Section 4.

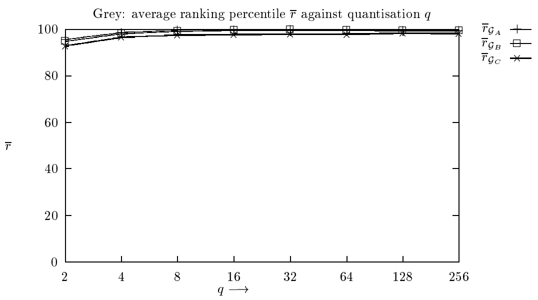

Fig. 1. Average ranking percentile of $\mathcal{G}_A$, $\mathcal{G}_B$ and $\mathcal{G}_C$.

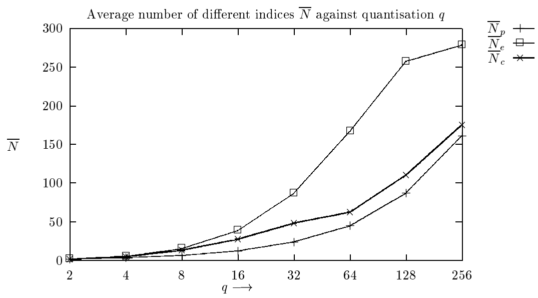

Fig. 2. Average number of different indices against number of bins

First, the average ranking percentile of hue points $\overline{r}_{\mathcal{G}_A}$, hue-hue edges $\overline{r}_{\mathcal{G}_B}$ and corners $\overline{r}_{\mathcal{G}_C}$, is tested in related to q, see Figure 1. The influence of the number of bin levels on the average ranking percentile based on color hue points, hue-hue edges and corners is negligible. Furthermore, $\overline{r}_{\mathcal{G}_A}$ gives the same results as $\overline{r}_{\mathcal{G}_B}$ which are slightly better then $\overline{r}_{\mathcal{G}_C}$. Beyond $q \geq 16$, retrieval accuracy is constant, so it is concluded that $q = 16$ bins is sufficient for proper color invariant retrieval.

Second, the average number of nonzero buckets for hue point $\overline{N}_p$, hue-hue edges $\overline{N}_e$ and corners $\overline{N}_c$ with respect to q is considered, see Figure 2. From the results we can see that the rate of incline of $\overline{N}_e$ is a constant higher (approximately 2) then for $\overline{N}_p$ and $\overline{N}_c$.

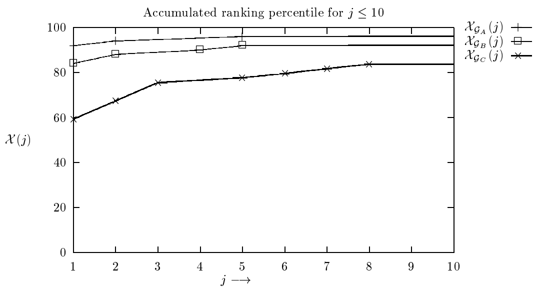

Fig. 3. Accumulated ranking percentile of $\mathcal{G}_A$, $\mathcal{G}_B$ and $\mathcal{G}_C$ for $q = 16$.

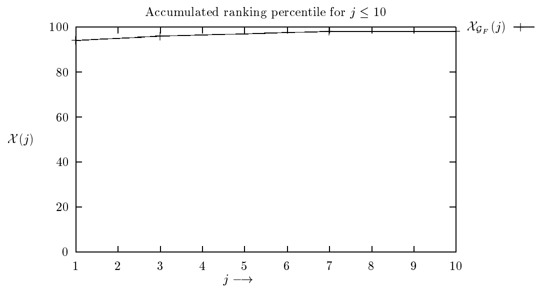

Fig. 4. Accumulated ranking percentile for $\mathcal{G}_F$.

To compromise, minimizing the two performance criteria expressing discrimination power and computational complexity, the appropriate bin number is $q = 16$ used in the sequel.

Figure 3 shows the accumulated ranking $\mathcal{X}$ for $q = 16$ of the 70 query images based on $\mathcal{G}_A$, $\mathcal{G}_B$ and $\mathcal{G}_C$ respectively. Excellent performance is shown for both $\mathcal{X}_{\mathcal{G}_A}$ and $\mathcal{X}_{\mathcal{G}_B}$, where respectively 92% and 87% of the position of the correct match in the ordered list of match values is within the first 2 and respectively 97% and 92% within the first 5 rankings. Misclassification occurs when the query image consists of very few different hue-hue edges or corners (i.e. small object).

6.6 *Shape Invariant Image Index*

In this section, the discrimination power of Euclidean and projective invariant indices are examined. The features under consideration are corner points.

Most existing shape-based matching techniques use intensity edge or high curvature points as feature points. However, intensity edge or high curvature points do not necessarily correspond to material boundaries. Shadows and abrupt surface orientation change of a solid object appear in the scene as an intensity edge without a material edge to support it introducing a large number of accidental feature points. To that end, we use hue corners as feature points. Hue corners are viewpoint-independent discounting shading, shadows and highlights.

To evaluate the discriminative power of shape invariant index, the following histograms, defined in Section 4, are considered: $\mathcal{G}_D$ and $\mathcal{G}_E$. Histogram $\mathcal{G}_D^{I_i}$ gives the distribution of Euclidean distances and $\mathcal{G}_E^{I_i}$ the distribution of cross ratios between hue corners.

Average ranking percentile for $\mathcal{G}_D$ and $\mathcal{G}_E$ is denoted by $\overline{r}_{\mathcal{G}_D}$ and $\overline{r}_{\mathcal{G}_E}$ respectively and shown for different $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ is shown in Figure 5. The average number of different distance values $\overline{N}_D$ and cross ratios $\overline{N}_E$ is shown is Figure 6.

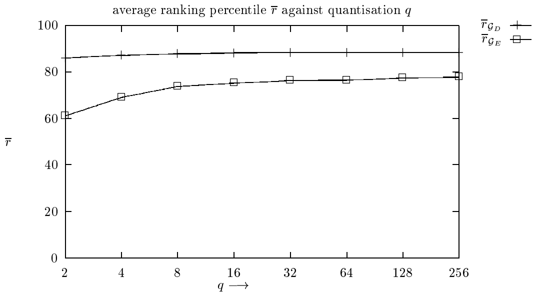

Fig. 5. Average ranking percentile for $\mathcal{G}_D$ and $\mathcal{G}_E$.

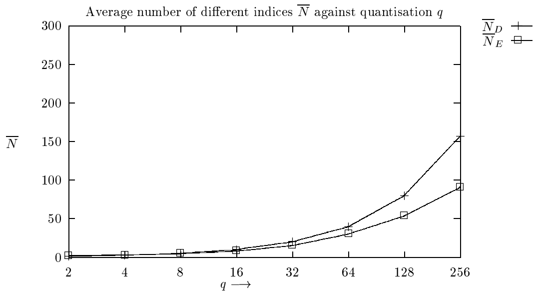

Fig. 6. Average number of different indices.

As expected, projective invariant values are less constrained (i.e. more coordinate combinations produce the same invariant value) and hence the discrimination performance expressed by $\overline{r}_{\mathcal{G}_E}$ is significantly worse than that of $\overline{r}_{\mathcal{G}_D}$. For proper retrieval accuracy, the number of bins is $q = 16$.

The average number of different distance values $\overline{N}_D$ and cross ratios $\overline{N}_E$ is shown is Figure 6.

To minimize the two performance criteria, $q = 16$ is taken for $\mathcal{G}_D$ and $\mathcal{G}_E$ in the sequel.

Note that the discriminative power of color invariant image index is significantly better than shape invariant matching. Shape can only serve as additional information.

6.7 *Color and Shape Invariant Image Index*

In this section, the discriminative power of shape and color invariant histogram matching is examined by considering $\mathcal{G}_F$ as defined in Section 4. Features used are hue-hue corners. There is no need for tuning parameter q because $\mathcal{G}_F$ can be seen as the aggregation of $\mathcal{G}_C$ and $\mathcal{G}_D$ both with $q = 16$. The accumulated ranking $\mathcal{X}$ is shown in Figure 4.

Excellent discrimination performance is shown, where 96% of the correct images are within the first 2 and 98% within the first 7 rankings. However, because geometric invariants are computed from the coordinates of color invariants, the average number of different indices is $\overline{N}_{\mathcal{G}_F} = 898$ which is quite large compared with histogram matching based entirely on color invariants.

When the performance of different invariant image indices is compared by evaluating the results with respect to the performance criteria given in 6.3., histogram matching based on both shape and color invariants produces the highest discriminative power and worst computational complexity. Invariant shape based matching yields poor discriminative power with bad computational complexity. However, color invariant based histogram matching results in very good discrimination performance and best computational complexity. Color based invariant indexing can be used as a filter to reject a large number of possible images from the image database yielding a short list of candidate solutions. Images in this list are then verified to be an instance of the query image by histogram matching based on both shape and color invariants or another type of verification scheme.

Therefore, in the next section, we test the effect occlusion and change in viewpoint has on color based invariant histogram matching.

6.8 *Stability to Occlusion and Viewpoint*

To test the effect of occlusion on the color based color invariant histogram matching process, 10 objects, already in the database of 500 recordings, were randomly selected and in total 40 images were taken blanking out $o \in \{50, 65, 80, 90\}$ percent of the total object area. The ranking percentile $\overline{r}_{\mathcal{G}_A}$, $\overline{r}_{\mathcal{G}_B}$ and $\overline{r}_{\mathcal{G}_B}$ averaged over the 10 histogram matching values is shown in Figure 7.

From the results we see that, in general, the shape and decline of the the curves for different color invariant functions do not differ significantly, except their attitude. This means that the effect of occlusion is fairly the same for all color invariant functions: constant for $0 < o < 50$, then a linear decline for $50 < o < 80$, proceeding in a rapid decrease for $o > 80$.

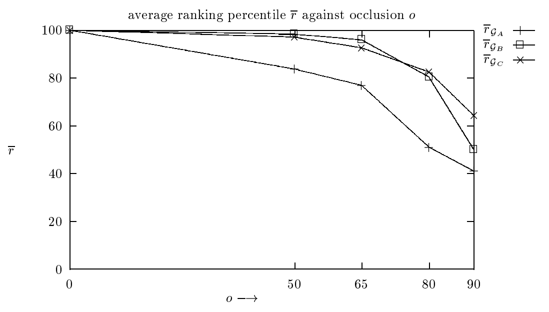

Fig. 7. Average ranking against occlusion $o \in \{50, 65, 80, 90\}$.

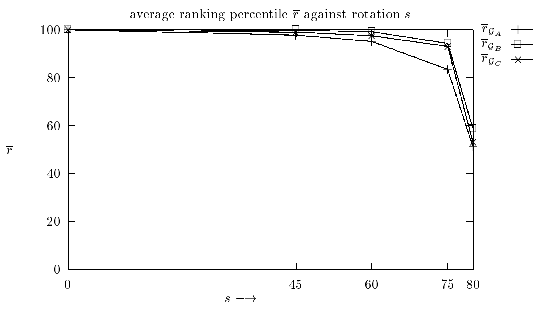

Fig. 8. Average ranking against rotation $s \in \{0, 45, 60, 75, 80\}$.

To test the effect of change in viewpoint, the 10 flat objects were put perpendicularly in front of the camera and in total 50 recordings were made by varying the angle between the camera for $s = \{0, 45, 60, 75, 80\}$ with respect to the object. Average ranking percentile is shown in Figure 8.

Looking at the results, the rate of decline is almost negligible for $0 < s < 60$, resulting in rapid decline for $s > 60$. This includes that the color invariant is highly robust to a change in viewpoint up to 60^o between the object and the the camera.

7 Summary

In this paper, indexing is used as a common framework to represent, index and retrieve images on the basis of color and shape invariants.

Experimental results showed that image retrieval based on both color and shape invariants provides excellent retrieval accuracy. Shape based invariant image retrieval yields poor discriminative power and worst computational performance whereas color based invariant image retrieval provides high discrimination power and best computational performance. Hence, shape can only serve as additional information for the purpose of invariant image retrieval. Another drawback of shape based invariant image retrieval is that it is restricted to planar objects from which geometrical properties are derived whereas color invariants can be derived from any view of a planar or 3D object.

The experimental results further showed that identifying multicolored objects entirely on the basis of color invariants is to a large degree robust to partial occlusion and a change in viewing position.

References

[1] A. Califano and R. Mohan, Multidimensional indexing for recognizing visual shapes, *IEEE PAMI*, 16(4), pp. 373-392, 1994

[2] J. Canny, A computational approach to edge detection, *IEEE Transactions on PAMI*, Vol. 8, No. 6, pp. 679-698, 1986.

[3] J. Kender, Saturation, hue, and normalized color: calculation digitization, and use, *Computer science technical report*, Carnegie-Mellon University, 1976.

[4] L. Kitchen and A. Rosenfeld, Gray-level corner detection, *Patt. Rec. Lett.*, 1, pp. 95-102, 1982.

[5] D. T. Clemens and D. W. Jacobs, Model group indexing for recognition, *Proc. conf. CVPR*, IEEE CS Press, Los Alamitos, Calif., pp 4-9, 1991

[6] B. V. Funt and G. D. Finlayson, Color constant color indexing, *IEEE PAMI*, 17(5), pp. 522-529, 1995.

[7] T. Gevers and A. W. M. Smeulders, Enigma: an image retrieval system, *Proc. ICPR*, The Hague, The Netherlands, vol. II, pp. 697-700, 1992.

[8] T. Gevers and A. W. M. Smeulders, Color and shape invariant image indexing, submitted for publication.

[9] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic and P. Yanker, The QBIC project: querying images by content using color, texture, and shape, Proc. storage and retrieval for image and video databases, SPIE, 1993

[10] A. Pentland, R. W. Picard and S. Sclaroff, Photobook: tools for content-based manipulation of image databases, *Proc. storage and retrieval for image and video databases II*, 2, 185, SPIE, Bellingham, Wash. pp. 34-47, 1994

[11] C. A. Rothwell, A. Zisserman, D. A. Forsyth and J. L. Mundy, Planar object recognition using projective shape representation, *Int'l, J. Comput. Vision*, 16, pp. 57-99, 1995.

[12] F. Stein and G. Medioni, Structural indexing: efficient 2-D object recognition, *IEEE PAMI*, 14, pp. 1198-1204, 1992

[13] M. J. Swain and D. H. Ballard, Color indexing, *Int'l, J. Comput. Vision*, 7(1), pp. 11-32, 1991.

[14] O. Veiblen and J. W. Young, Projective Geometry, Ginn. Boston, 1910.

[15] H. J. Wolfson, Object recognition by transformation invariant indexing, *Proc. Invariance Workshop*, ECCV, 1992.

[16] G. Wyszecki and W. S. Stiles, Color science: concepts and methods, quantitative data and formulae, Wiley: New York, 1982.