PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval

Theo Gevers and Arnold W. M. Smeulders, *Member, IEEE*

***Abstract*—We aim at combining color and shape invariants for indexing and retrieving images. To this end, color models are proposed independent of the object geometry, object pose, and illumination. From these color models, color invariant edges are derived from which shape invariant features are computed. Computational methods are described to combine the color and shape invariants into a unified high-dimensional invariant feature set for discriminatory object retrieval. Experiments have been conducted on a database consisting of 500 images taken from multicolored man-made objects in real world scenes. From the theoretical and experimental results it is concluded that object retrieval based on composite color and shape invariant features provides excellent retrieval accuracy. Object retrieval based on color invariants provides very high retrieval accuracy whereas object retrieval based entirely on shape invariants yields poor discriminative power. Furthermore, the image retrieval scheme is highly robust to partial occlusion, object clutter and a change in the object's pose. Finally, the image retrieval scheme is integrated into the PicToSeek system on-line at http://www.wins.uva.nl/research/isis/PicToSeek/ for searching images on the World Wide Web.**

***Index Terms*—color invariant edges, color invariants, combining color and shape information, dichromatic reflection, image retrieval, object search, query by example, reflectance properties, shape invariants.**

I. Introduction

FOR THE management of archived image data, an image database system is needed that 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 in geographic information systems. In contrast, image database systems are still based on the idea of storing a keyword description of the image content, created by a user on input, in addition to a pointer to the raw image data. Image retrieval is then shifted to standard DBMS capabilities.

A different approach is required when we consider the retrieval of images by image example, where a query image or sketch 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. In this paper, we focus on the problem of retrieving images containing instances of particular objects. Then, the query is specified by an example image taken from the object(s) at hand. In this context, image retrieval is similar to object search.

The basic idea of image retrieval by image example is to extract characteristic features from target images which are then matched with those of the query image. These features are typically derived from shape, texture, or color properties of query and target images. After matching, images are ordered with respect to the query image according to their similarity measure and displayed for viewing, see [1]–[7], for example.

The matching complexity of image retrieval by image example is similar to that of model-based 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 interaction is 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 [8]–[13]). Indexing based matching schemes have a similar underlying structure. First, a lookup table is formed by quantization of the index parameter space. Then, index vectors are generated, computing shape, color, or texture properties from target images in the image database. At run-time, these features are extracted from the query image, and indexes are computed and used to look up images in the lookup table. Because indexing based matching avoids exhaustive search, it is a potentially efficient search technique. A proper indexing technique will be executed at high speed allowing for fast image retrieval by image example. This is useful when the image database is large as may be anticipated for multimedia and information services.

Ideally, the value of the index vectors, derived from images taken from the same object, should remain the same regardless of the varying circumstances induced by the imaging process. For instance, when images are taken from the same object from different viewpoints, the shape of the recorded object will exhibit a geometric distortion. Also photometric changes may occur when the viewpoint is changed, yielding different shadowing, shading and highlighting cues for the same object. In other words, the value of index vectors should be *invariant* with respect to the varying imaging conditions.

Most of the work on shape-based object recognition rely on matching sets of local image features (e.g., edges, lines and corners) to three-dimensional (3-D) object models invariant to geometric transformations (e.g., translation, rotation, scale, and affine transformation) and significant progress has been achieved (for example [8], [9], [11], [13]). As an expression of the difficulty of the general problem, most of the geometry-based matching schemes can handle only simple, flat, and

Manuscript received December 21, 1998; revised July 7, 1999. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. B. S. Manjunath.

The authors are with the ISIS Group, Faculty of WINS, 1098 SJ Amsterdam, The Netherlands (e-mail: gevers@wins.uva.nl; smeulders@wins.uva.nl).

Publisher Item Identifier S 1057-7149(00)00235-9.

rigid man-made objects. Shape features are rarely adequate for discriminatory object recognition of 3-D objects from arbitrary viewpoints in complex scenes.

As opposed to shape information, other retrieval schemes are entirely on the basis of color. Swain and Ballard [12] made a significant contribution in introducing color for object search. Based on the opponent color model, they show that image retrieval based on histogram matching is to a large degree robust to changes in object pose and shape. The histogram based matching scheme is extended by Funt and Finlayson [14] and Nayar and Bolle [15] to make the method illumination independent by indexing on color ratio's computed from neighboring image points. However, the color ratio's are negatively affected by the geometry of the object. Further, Finlayson *et al.* [16], Healey and Slater [17], and Slater and Healey [18] introduced illumination-invariant moments of color histogram distributions.

In addition, general purpose image retrieval systems have been developed based on multiple features (e.g., color, shape, and texture) describing the image content [3], [4], [6]. We implemented the Enigma system [19], retrieving images based on query by example. QBIC [20] allows for content-based retrieval for large image and video databases. Photobook [5] reduces images to a small set of perceptually significant coefficients for the purpose of image retrieval. In [21], shape information has been used for image retrieval. In contrast to full content-based image retrieval, Chabot [22], [23] uses a combination of visual appearance and text-based cues to retrieve images. Image retrieval using combined color and shape information has been proposed by [24]. However, the retrieval scheme is suited for flat-images of trademarks. Recently, a number of image browsers are available for retrieving images from the World Wide Web, for example [1], [25]–[28]. These retrieval systems use color and shape information separately for the purpose of image retrieval. Moreover, the features used during the retrieval process depend on the shape of the object, camera viewpoint, and on the illumination. As a consequence, the performance of these systems may decrease when the query and target image taken from the same object are recorded under different imaging conditions.

In this paper, we want to arrive at *combining* color and shape invariants for the purpose of image indexing and retrieval. To that end, a retrieval scheme is proposed making use of *local* color invariant information to produce *semiglobal* shape invariants to obtain a viewpoint invariant, high-dimensional object descriptor to be used as an index for discriminatory image retrieval. To achieve this, color invariant features are proposed according to the following criteria: invariance to the viewpoint, geometry of the object, and illumination conditions. Then, from these color models, color invariant edges are derived from which the shape features are computed. Shape features are independent up to a change in viewpoint (i.e., projective transformation). Computational methods are proposed to combine color and shape invariants into a unified high-dimensional invariant feature space. The image retrieval scheme is designed according to the following criteria: high discriminative power, and robustness against fragmented, occluded and overlapping objects.

The paper is organized as follows. First, in Section II, we propose new color models invariant to a change in view point, object geometry and illumination. Color invariant edges are proposed in Section III. Shape invariants are discussed in Section IV. In Section V, we propose computational methods to produce a composite color and shape invariant indexing scheme. The matching scheme is given in Section VI. Finally, in Section VII, the performance of different invariant image features is evaluated on a dataset of 500 images.

II. Color Invariants

As discussed, attention is to be paid to the desired classes of invariance. For each image retrieval query, a proper definition of the desired invariance is essential. A concise list of the most important invariance properties is as follows.

- Is the search for objects in different orientations and scales?
- Is the search for objects in a large variety of scenes?
- Is the search for objects in other kind of light?
- Is the search for objects from different viewpoints?
- Is the search for an object irrespective occlusion?

In this section, we propose new sets of color models independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and highlights.

A. The Reflection Model

Let $E(\vec{x}, \lambda)$ be the spectral power distribution of the incident light at the object surface at $\vec{x}$, and let $L(\vec{x}, \lambda)$ be a complex function based on the geometric and spectral properties of the object surface at $\vec{x}$. The spectral sensitivity of the kth sensor is given by $F_k(\lambda)$. Then ρ_k, the sensor response of the kth channel, is given by

$$\rho_k(\vec{x}) = \int_\lambda E(\vec{x}, \lambda)L(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \tag{1}$$

where λ denotes the wavelength. The integral is taken from the visible spectrum (e.g., 380–700 nm).

Further, consider an opaque inhomogeneous dielectric object, then the geometric and surface reflection component of function $L(\vec{x}, \lambda)$ can be decomposed in a body (matte) and surface (specular) reflection component as described by Shafer [29]:

$$\begin{aligned}\phi_k(\vec{x}) = {} & G_B(\vec{x}, \vec{n}, \vec{s}) \int_\lambda E(\vec{x}, \lambda)B(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \\ & + G_S(\vec{x}, \vec{n}, \vec{s}, \vec{v}) \int_\lambda E(\vec{x}, \lambda)S(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \end{aligned} \tag{2}$$

giving the kth sensor response. Further, $B(\vec{x}, \lambda)$ and $S(\vec{x}, \lambda)$ are the surface albedo and Fresnel reflectance at $\vec{x}$, respectively. $\vec{n}$ is the surface patch normal, $\vec{s}$ is the direction of the illumination source, and $\vec{v}$ is the direction of the viewer. Geometric terms G_B and G_S denote the geometric dependencies on the body and surface reflection component, respectively.

B. Reflectance with White Illumination

Considering the neutral interface reflection (NIR) model [assuming that $S(\vec{x}, \lambda)$ has a constant value independent of the wavelength] and white illumination, then $S(\vec{x}, \lambda) = S(\vec{x})$, and $E(\vec{x}, \lambda) = E(\vec{x})$. Then, we put forward that the measured sensor values are given by [30]:

$$\begin{aligned}\omega_k(\vec{x}) = {} & G_B(\vec{x}, \vec{n}, \vec{s})E(\vec{x}) \int_\lambda B(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \\ & + G_S(\vec{x}, \vec{n}, \vec{s}, \vec{v})E(\vec{x})S(\vec{x}) \int_\lambda F_k(\lambda)\, d\lambda \quad (3)\end{aligned}$$

giving the kth sensor response of an infinitesimal surface patch under the assumption of a white light source.

If the integrated white condition holds (i.e., the area under the sensor spectral functions is approximately the same)

$$\int_\lambda F_i(\lambda)\, d\lambda = \int_\lambda F_j(\lambda)\, d\lambda. \quad (4)$$

We propose that the reflection from inhomogeneous dielectric materials under white illumination is given by:

$$\begin{aligned}\omega_k(\vec{x}) = {} & G_B(\vec{x}, \vec{n}, \vec{s})E(\vec{x}) \int_\lambda B(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \\ & + G_S(\vec{x}, \vec{n}, \vec{s}, \vec{v})E(\vec{x})S(\vec{x})F. \quad (5)\end{aligned}$$

If $\omega(\vec{x})$ is not dependent on $\vec{x}$, we obtain

$$\omega_k = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_k(\lambda)\, d\lambda + G_S(\vec{n}, \vec{s}, \vec{v})ESF. \quad (6)$$

In the next section, this reflection model is used to derive color invariants.

C. Body Reflectance Invariance

Consider the body reflection term of (5)

$$\beta_k(\vec{x}) = G_B(\vec{x}, \vec{n}, \vec{s})E(\vec{x}) \int_\lambda B(\vec{x}, \lambda)F_k(\lambda)\, d\lambda \quad (7)$$

giving the kth sensor response of an infinitesimal *matte* surface patch under the assumption of a white light source.

We now consider the shape of the color clusters which will be formed in RGB space by pixels coming from the same uniformly colored surface of matte material according to the reflectance model. In fact, the color depends on $\int_\lambda B(\vec{x}, \lambda)F_k(\lambda)\, d\lambda$ (i.e., surface albedo), and the length of the cluster depends on the illumination $E(\vec{x})$ and roughness and shape of the object $G_B(\vec{x}, \vec{n}, \vec{s})$. In other words, a uniformly colored surface which is curved (i.e., varying surface orientation) gives rise to a broad variance of sensor values. Any expression defining colors on the same elongated color cluster spanned by the body reflection vector in sensor space, originating from the origin (i.e., black point), is a color invariant for matte objects under white illumination.

To that end, we propose the following basic set of *irreducible color invariants* at a specific location $\vec{x}$:

$$\frac{\beta_i(\vec{x})}{\beta_j(\vec{x})} = \frac{\beta_i}{\beta_j} \quad (8)$$

where $\vec{x}$ is discarded as the color ratio is taken from the same surface location.

The expression is a color invariant for the dichromatic reflection model for matte objects under white illumination as follows from substituting (7) in (8):

$$\begin{aligned}\frac{\beta_i}{\beta_j} &= \frac{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_i(\lambda)\, d\lambda}{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_j(\lambda)\, d\lambda} \\ &= \frac{\int_\lambda B(\lambda)F_i(\lambda)\, d\lambda}{\int_\lambda B(\lambda)F_j(\lambda)\, d\lambda} \quad (9)\end{aligned}$$

only dependent on the surface albedo and the sensors and factoring out dependencies on the viewpoint, surface orientation, illumination direction, and illumination intensity.

Any (linear) combination of the basic set of irreducible color invariants will result in a new color invariant. For the ease of illustration, we now focus on the 3-D RGB-space given by

$$R_b = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_R(\lambda)\, d\lambda \quad (10)$$

$$G_b = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_G(\lambda)\, d\lambda \quad (11)$$

$$B_b = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_B(\lambda)\, d\lambda \quad (12)$$

where $C \in \{R_b, G_b, B_b\}$ giving the red, green, and blue sensor response of an infinitesimal *matte* surface patch under the assumption of a white light source.

Then, having red, green, and blue as primary colors yielding the basic set of irreducible color invariants [cf. (8)]:

$$\begin{aligned}\frac{R_b}{G_b} &= \frac{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_R(\lambda)\, d\lambda}{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_G(\lambda)\, d\lambda} \\ &= \frac{\int_\lambda B(\lambda)F_R(\lambda)\, d\lambda}{\int_\lambda B(\lambda)F_G(\lambda)\, d\lambda} \quad (13)\end{aligned}$$

$$\frac{B_b}{R_b} = \frac{G_B(\vec{n}, \vec{s})E\int_\lambda B(\lambda)F_B(\lambda)\,d\lambda}{G_B(\vec{n}, \vec{s})E\int_\lambda B(\lambda)F_R(\lambda)\,d\lambda} = \frac{\int_\lambda B(\lambda)F_B(\lambda)\,d\lambda}{\int_\lambda B(\lambda)F_R(\lambda)\,d\lambda} \tag{14}$$

$$\frac{G_b}{B_b} = \frac{G_B(\vec{n}, \vec{s})E\int_\lambda B(\lambda)F_G(\lambda)\,d\lambda}{G_B(\vec{n}, \vec{s})E\int_\lambda B(\lambda)F_B(\lambda)\,d\lambda} = \frac{\int_\lambda B(\lambda)F_G(\lambda)\,d\lambda}{\int_\lambda B(\lambda)F_B(\lambda)\,d\lambda}. \tag{15}$$

Then, other color invariants can be computed in a systematic manner in terms of R_b, G_b, and B_b:

$$C(R_b, G_b, B_b) = \frac{\sum_i a_i(R_b)_i^p(G_b)_i^q(B_b)_i^r}{\sum_j b_j(R_b)_j^s(G_b)_j^t(B_b)_j^u} \tag{16}$$

where $p+q+r = s+t+u$, and $p, q, r, s, t, u \in \mathcal{R}$. Further, $i, j \geq 1$ and $a_i, b_j \in \mathcal{R}$.

Lemma 1: Assuming dichromatic reflection and white illumination, C is independent of the viewpoint, surface orientation, illumination direction, and illumination intensity.

Proof: By substituting (10)–(12) in (16) we have (16a), shown at the bottom of the page, factoring out dependencies on the viewpoint, surface orientation, illumination direction, and illumination intensity. $K_C(\lambda) = \int_\lambda B(\lambda)F_C(\lambda)\,d\lambda$ for $C \in \{R, G, B\}$ is the compact formulation depending on the sensors and surface albedo only. Further, $p+q+r = s+t+u$, and $p, q, r, s, t, u \in \mathcal{R}$. Finally, $i, j \geq 1$ and $a_i, b_j \in \mathcal{R}$. ■

For instance, for the first order color invariants (i.e., $p+q+r = s+t+u = 1$), we have the set

$$\left\{\frac{R}{B}, \frac{-B}{G}, \frac{R+G+B}{3R+B}, \frac{R}{R+G+B}, \frac{3(B-G)}{2R+G}, \cdots, \right\} \tag{17}$$

and for the second order color invariants (i.e., $p+q+r = s+t+u = 2$):

$$\left\{\frac{RB}{B^2}, \frac{4BR}{5B^2}, \frac{R^2+G^2}{B^2}, \frac{B^2+3R^2}{R^2}, \cdots, \right\} \tag{18}$$

and for the third order color invariants:

$$\left\{\frac{G^3}{R^3+5B^3}, \frac{RGB}{R^3}, \frac{RG^2+B^3}{B^3}, \frac{BR^2+G^3}{R^3+G^3}, \cdots, \right\} \tag{19}$$

etc., where each expression is a color invariant for the dichromatic reflectance under white illumination.

We can easily see that normalized color given by [31]

$$r(R, G, B) = \frac{R}{R+G+B}, \tag{20}$$

$$g(R, G, B) = \frac{G}{R+G+B}, \tag{21}$$

$$b(R, G, B) = \frac{B}{R+G+B} \tag{22}$$

$$\begin{aligned} C(R_b, G_b, B_b) &= \frac{\sum_i a_i(R_b)_i^p(G_b)_i^q(B_b)_i^r}{\sum_j b_j(R_b)_j^s(G_b)_j^t(B_b)_j^u} \\ &= \frac{\sum_i a_i(G_B(\vec{n}, \vec{s})EK_R(\lambda))_i^p(G_B(\vec{n}, \vec{s})EK_G(\lambda))_i^q(G_B(\vec{n}, \vec{s})EK_B(\lambda))_i^r}{\sum_j a_j(G_B(\vec{n}, \vec{s})EK_R(\lambda))_j^s(G_B(\vec{n}, \vec{s})EK_G(\lambda))_j^t(G_B(\vec{n}, \vec{s})EK_B(\lambda))_j^u} \\ &= \frac{\sum_i a_i(G_B(\vec{n}, \vec{s})E)^{p+q+r}(K_R(\lambda))_i^p(K_G(\lambda))_i^q(K_B(\lambda))_i^r}{\sum_j b_j(G_B(\vec{n}, \vec{s})E)^{s+t+u}(K_R(\lambda))_j^s(K_G(\lambda))_j^t(K_B(\lambda))_j^u} \\ &= \frac{\sum_i a_i(K_R(\lambda))_i^p(K_G(\lambda))_i^q(K_B(\lambda))_i^r}{\sum_j b_j(K_R(\lambda))_j^s(K_G(\lambda))_j^t(K_B(\lambda))_i^u} \end{aligned} \tag{16a}$$

is an instantiation of the first order color invariant of (16) and hence being independent of the viewpoint, surface orientation, illumination direction, and illumination intensity as shown in (23), shown at the bottom of the page, where again

$$K_C(\lambda) = \int_\lambda B(\lambda) F_C(\lambda) d\lambda \qquad \text{for } C \in \{R, G, B\} \tag{24}$$

is the compact formulation depending on the sensors and surface albedo only. Equal arguments hold for g and b.

Although any instantiation of C can be taken for the purpose of viewpoint independent image retrieval, in this paper, normalized color rgb is considered as an instantiation of C because normalized color is intuitive and well-known in the color literature. In addition to rgb, the following first-order color invariant has been selected as an instantiation of C for viewpoint-invariant object search:

$$c_4(R, G, B) = \frac{R - G}{R + G} \tag{25}$$

$$c_5(R, G, B) = \frac{R - B}{R + B} \tag{26}$$

$$c_6(R, G, B) = \frac{G - B}{G + B} \tag{27}$$

being invariants for matte, dull objects [cf. (10)–(12) and (25)–(27)]:

$$\begin{aligned} c_4(R_b, G_b, B_b) &= \frac{G_B(\vec{n}, \vec{s}) E K_R(\lambda) - G_B(\vec{n}, \vec{s}) E K_G(\lambda)}{G_B(\vec{n}, \vec{s}) E K_R(\lambda) + G_B(\vec{n}, \vec{s}) E K_G(\lambda)} \\ &= \frac{K_R(\lambda) - K_G(\lambda)}{K_R(\lambda) + K_G(\lambda)} \end{aligned} \tag{28}$$

$$\begin{aligned} c_5(R_b, G_b, B_b) &= \frac{G_B(\vec{n}, \vec{s}) E K_R(\lambda) - G_B(\vec{n}, \vec{s}) E K_B(\lambda)}{G_B(\vec{n}, \vec{s}) E K_R(\lambda) + G_B(\vec{n}, \vec{s}) E K_B(\lambda)} \\ &= \frac{K_R(\lambda) - K_B(\lambda)}{K_R(\lambda) + K_B(\lambda)} \end{aligned} \tag{29}$$

$$\begin{aligned} c_6(R_b, G_b, B_b) &= \frac{G_B(\vec{n}, \vec{s}) E K_G(\lambda) - G_B(\vec{n}, \vec{s}) E K_B(\lambda)}{G_B(\vec{n}, \vec{s}) E K_G(\lambda) + G_B(\vec{n}, \vec{s}) E K_B(\lambda)} \\ &= \frac{K_G(\lambda) - K_B(\lambda)}{K_G(\lambda) + K_B(\lambda)} \end{aligned} \tag{30}$$

only dependent on the sensors and the surface albedo.

The effect of surface reflection (highlights) is discussed in the following section.

D. Body and Surface Reflectance Invariance

Consider the surface reflection term of (5)

$$\gamma_k(\vec{x}) = G_S(\vec{x}, \vec{n}, \vec{s}, \vec{v}) E(\vec{x}) S(\vec{x}) F \tag{31}$$

giving the kth sensor response for an infinitesimal *shiny* surface patch under white illumination.

For a given point on a shiny surface, the contribution of the body reflection component β and surface reflection component γ are added cf. (5). As a consequence, in RGB-color space, the observed colors of a uniformly colored (shiny) surface will be formed on the dichromatic plane spanned by the body and surface reflection components.

Under the condition of the NIR model and white light, this dichromatic plane originates from the main diagonal axis. Therefore, any expression defining colors on this dichromatic plane is a color invariant for the dichromatic reflection model. To that end, we propose the following basic set of *irreducible color invariants* at location $\vec{x}$:

$$\frac{\omega_i(\vec{x}) - \omega_j(\vec{x})}{\omega_k(\vec{x}) - \omega_l(\vec{x})} = \frac{\omega_i - \omega_j}{\omega_k - \omega_l} \tag{32}$$

where $\omega_k \neq \omega_l$, and $\vec{x}$ is omitted as the color ratio is taken from the same surface location.

$$K_k(\lambda) = \int_\lambda B(\lambda) F_k(\lambda)\, d\lambda \tag{33}$$

This expression is a color invariant for the dichromatic reflection model under white illumination as follows from substituting (5) in (32) as in (32a), shown at the bottom of the next page, only dependent on the sensors and the surface albedo, where is the compact formulation for the kth the channel.

Any (linear) combination of the basic set of irreducible color invariants will result in a new color invariant. For the ease of illustration, we again focus on the 3-D RGB-space given by

$$\begin{aligned} R_w = {} & G_B(\vec{n}, \vec{s}) E \int_\lambda B(\lambda) F_R(\lambda)\, d\lambda \\ & + G_S(\vec{n}, \vec{s}, \vec{v}) E S F \end{aligned} \tag{34}$$

$$\begin{aligned} G_w = {} & G_B(\vec{n}, \vec{s}) E \int_\lambda B(\lambda) F_G(\lambda)\, d\lambda \\ & + G_S(\vec{n}, \vec{s}, \vec{v}) E S F \end{aligned} \tag{35}$$

$$\begin{aligned} B_w = {} & G_B(\vec{n}, \vec{s}) E \int_\lambda B(\lambda) F_B(\lambda)\, d\lambda \\ & + G_S(\vec{n}, \vec{s}, \vec{v}) E S F \end{aligned} \tag{36}$$

$$\begin{aligned} r(R_b, G_b, B_b) &= \frac{G_B(\vec{n}, \vec{s}) E K_R}{G_B(\vec{n}, \vec{s}) E K_R(\lambda) + G_B(\vec{n}, \vec{s}) E K_G(\lambda) + G_B(\vec{n}, \vec{s}) E K_B(\lambda)} \\ &= \frac{K_R(\lambda)}{K_R(\lambda) + K_G(\lambda) + K_B(\lambda)} \end{aligned} \tag{23}$$

giving the red, green, and blue sensor response of an infinitesimal surface patch under the assumption of a white light source. Then, having red, green and blue as primary colors yielding the following basic set of irreducible color invariants:

$$\frac{(R_w - G_w)}{(B_w - R_w)} \tag{37}$$

$$\frac{(R_w - G_w)}{(G_w - B_w)} \tag{38}$$

$$\frac{(G_w - B_w)}{(B_w - R_w)} \tag{39}$$

color invariants can be computed in a systematic manner:

$$\begin{aligned} L(R_w, G_w, B_w) \\ = \frac{\sum_i a_i (R_w - G_w)_i^p (B_w - R_w)_i^q (G_w - B_w)_i^r}{\sum_j b_j (R_w - G_w)_j^s (B_w - R_w)_j^t (G_w - B_w)_j^u} \end{aligned} \tag{40}$$

where $p + q + r = s + t + u$, and $p, q, r, s, t, u \in \mathcal{R}$. Further, $i, j \geq 1$ and $a_i, b_j \in \mathcal{R}$.

Lemma 2: Assuming dichromatic reflection and white illumination, L is independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and highlights.

Proof: By substituting (37)–(39) in (40) we have as shown in (40a), shown at the bottom of the page, independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and highlights. Further, $p + q + r = s + t + u$, and $p, q, r, s, t, u \in \mathcal{R}$. $i, j \geq 1$ and $a_i, b_j \in \mathcal{R}$. Furthermore, $C_w = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_C(\lambda)\, d\lambda + G_S(\vec{n}, \vec{s}, \vec{v})ESF$, and $C_b = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_C(\lambda)\, d\lambda$, and $K_C(\lambda) = \int_\lambda B(\lambda)F_C(\lambda)\, d\lambda$ for $C \in \{R, G, B\}$. ■

For instance, for the first-order color invariants (i.e., $p + q + r = s + t + u = 1$), we have the set

$$\begin{aligned} \left\{ \frac{(R-G)}{(R-B)}, \frac{(B-G)}{(R-B)}, \frac{(R-G)+(B-G)}{(R-B)}, \right. \\ \left. \frac{(R-G)+3(B-G)}{(R-B)+2(R-G)}, \cdots \right\} \end{aligned} \tag{41}$$

and for the second order color invariants (i.e., $p + q + r = s + t + u = 2$)

$$\begin{aligned} \left\{ \frac{(R-G)(R-B)}{(R-B)^2}, \frac{(B-G)(R-B)}{(R-B)^2}, \right. \\ \left. \frac{(R-G)^2+(B-G)^2}{(R-B)^2}, \frac{(R-G)^2+3(B-G)^2}{(R-B)^2+2(R-G)^2}, \cdots \right\} \end{aligned} \tag{42}$$

$$\begin{aligned} \frac{\omega_i - \omega_j}{\omega_k - \omega_l} &= \frac{(G_B(\vec{n}, \vec{s})EK_i(\lambda) + G_S(\vec{n}, \vec{s}, \vec{v})ESF) - (G_B(\vec{n}, \vec{s})EK_j(\lambda) + G_S(\vec{n}, \vec{s}, \vec{v})ESF)}{(G_B(\vec{n}, \vec{s})EK_k(\lambda) + G_S(\vec{n}, \vec{s}, \vec{v})ESF) - (G_B(\vec{n}, \vec{s})EK_l(\lambda) + G_S(\vec{n}, \vec{s}, \vec{v})ESF)} \\ &= \frac{(G_B(\vec{n}, \vec{s})EK_i(\lambda)) - (G_B(\vec{n}, \vec{s})EK_j(\lambda))}{(G_B(\vec{n}, \vec{s})EK_k(\lambda)) - (G_B(\vec{n}, \vec{s})EK_l(\lambda))} = \frac{G_B(\vec{n}, \vec{s})E(K_i(\lambda) - K_j(\lambda))}{G_B(\vec{n}, \vec{s})E(K_k(\lambda) - K_l(\lambda))} \\ &= \frac{K_i(\lambda) - K_j(\lambda)}{K_k(\lambda) - K_l(\lambda)} \end{aligned} \tag{32a}$$

$$\begin{aligned} L(R_w, G_w, B_w) &= \frac{\sum_i a_i (R_w - G_w)_i^p (B_w - R_w)_i^q (G_w - B_w)_i^r}{\sum_j b_j (R_w - G_w)_j^s (B_w - R_w)_j^t (G_w - B_w)_j^u} \\ &= \frac{\sum_i a_i (R_b - G_b)_i^p (B_b - R_b)_i^q (G_b - B_b)_i^r}{\sum_j b_j (R_b - G_b)_j^s (B_b - R_b)_j^t (G_b - B_b)_j^u} \\ &= \frac{\sum_i a_i (G_B(\vec{n}, \vec{s})E)^{p+q+r} (K_R(\lambda) - K_G(\lambda))_i^p (K_R(\lambda) - K_B(\lambda))_i^q (K_G(\lambda) - K_B(\lambda))_i^r}{\sum_j b_j (G_B(\vec{n}, \vec{s})E)^{s+t+u} (K_R(\lambda) - K_G(\lambda))_j^s (K_R(\lambda) - K_B(\lambda))_j^t (K_G(\lambda) - K_B(\lambda))_j^u} \\ &= \frac{\sum_i a_i (K_R(\lambda) - K_G(\lambda))_i^p (K_R(\lambda) - K_B(\lambda))_i^q (K_G(\lambda) - K_B(\lambda))_i^r}{\sum_j b_j (K_R(\lambda) - K_G(\lambda))_j^s (K_R(\lambda) - K_B(\lambda))_j^t (K_G(\lambda) - K_B(\lambda))_j^u} \end{aligned} \tag{40a}$$

and for the third order color invariants

$$\left\{ \frac{(R-G)^3}{(R-B)^3}, \frac{(B-G)^3}{(R-B)^3}, \frac{(R-G)^3+(B-G)^3}{(R-B)^3}, \right. \\ \left. \frac{(R-G)(R-B)(G-B)+3(B-G)^3}{(R-B)^3+2(R-G)^3}, \cdots \right\} \quad (43)$$

etc., where each expression is a color invariant for the dichromatic reflectance under white illumination.

We can easily see that hue given by [31]:

$$H(R, G, B) = \arctan\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right) \quad (44)$$

ranging from [0, 2π) is an instantiation of the first order color invariant of (40), as a function of $\arctan()$, with $a_1 = \sqrt{3}$, $a_2 = 0$, $b_1 = 1$, $b_2 = 1$.

Although any instantiation of L can be taken for the purpose of viewpoint independent image retrieval, in this paper, the following first-order color invariant has been selected as an instantiation of L for viewpoint-invariant image retrieval:

$$l_4(R, G, B) = \frac{|R-G|}{|R-G|+|B-R|+|G-B|}, \quad (45)$$

$$l_5(R, G, B) = \frac{|R-B|}{|R-G|+|B-R|+|G-B|}, \quad (46)$$

$$l_6(R, G, B) = \frac{|G-B|}{|R-G|+|B-R|+|G-B|} \quad (47)$$

which is the set of normalized (absolute) color differences (ncd), where $0 \leq l_i \leq 1$ and $l_4 + l_5 + l_6 = 1$.

E. Noise Analysis of Color Invariants

In this section, the aim is to study the robustness of the different color invariants with respect to sensing and measurement errors. For example, it is known that normalized color become more sensitive to noise when $R+G+B$ is near zero [32]. To get more insight in the noise stability of the newly proposed color invariants, we analyze and compare the noise sensitivity of the color invariants rgb, $c_4c_5c_6$, hue and $l_4l_5l_6$.

It is known that noise sensitivity of a function can be derived from the stability of its variables. The idea is that the uncertainty in a function is stretched with the value of the derivative at that point. Then the sensitivity of a function $f(x, y, \cdots,)$ with variables $x, y, \cdots$, having values $x_0, y_0, \cdots$ is given by

$$\Delta f(x_0, y_0, \cdots,) = \left|\frac{\delta f}{\delta x}\right| x = x_0 \Delta x + \left|\frac{\delta f}{\delta y}\right| y = y_0 \Delta y + \cdots. \quad (48)$$

We have computed Δf for the different color invariants. It can be concluded that normalized color becomes unstable when intensity is small as reported by Kender [32]. Same arguments hold for $c_4c_5c_6$, where rgb is slightly more robust than $c_4c_5c_6$. For hue and $l_4l_5l_6$ it is concluded that they become unstable when intensity and saturation (i.e., near $R = G = B$) is small, where $l_4l_5l_6$ is slightly more robust than hue. Note that $l_4l_5l_6$ has a singular point at $R = G = B$.

III. Color Invariant Gradients

In the previous section, we discussed color models that are invariant under varying imaging conditions. In this section, we propose color invariant edges derived from the newly proposed color models. The color invariant edges will be used to compute the shape-based invariant features.

A. Gradients in Multivalued Images

In contrast to gradient methods that combine individual components of a multivalued image in an *ad hoc* manner without any theoretical basis (e.g., taking the sum or RMS of the component gradient magnitudes as the magnitude of the resultant gradient), we follow the principled way to compute gradients in vector images as described by Silvano di Zenzo [33] and further used in [34], which is summarized as follows.

Let $\Theta(x_1, x_2)$: $\Re^2 \rightarrow \Re^m$ be a m-band image with components $\Theta_i(x_1, x_2)$: $\Re^2 \rightarrow \Re$ for $i = 1, 2, \cdots, m$. For color images we have $m = 3$. Hence, at a given image location the image value is a vector in $\Re^m$. The difference at two nearby points $P = (x_1^0, x_2^0)$ and $Q = (x_1^1, x_2^1)$ is given by $\triangle\Theta = \Theta(P) - \Theta(Q)$. Considering an infinitesmall displacement, the difference becomes the differential $d\Theta = \sum_{i=1}^2 (\partial\Theta/\partial x_i)\, dx_i$ and its squared norm is given by

$$\begin{aligned} d\Theta^2 &= \sum_{i=1}^{2}\sum_{k=1}^{2} \frac{\partial\Theta}{\partial x_i}\frac{\partial\Theta}{\partial x_k}\, dx_i\, dx_k \\ &= \sum_{i=1}^{2}\sum_{k=1}^{2} g_{ik}\, dx_i\, dx_k \\ &= \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \end{aligned} \quad (49)$$

where $g_{ik} := (\partial\Theta/\partial x_i) \cdot (\partial\Theta/\partial x_k)$ and the extrema of the quadratic form are obtained in the direction of the eigenvectors of the matrix $[g_{ik}]$ and the values at these locations correspond with the eigenvalues given by

$$\lambda_\pm = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \quad (50)$$

with corresponding eigenvectors given by $(\cos\theta_\pm, \sin\theta_\pm)$, where $\theta_+ = (1/2)\arctan(2g_{12}/g_{11} - g_{22})$ and $\theta_- = \theta_+ + \pi/2$. Hence, the direction of the minimal and maximal changes at a given image location is expressed by the eigenvectors θ_- and θ_+, respectively, and the corresponding magnitude is given by the eigenvalues λ_- and λ_+, respectively. Note that λ_- may be different than zero and that the strength of an multivalued edge should be expressed by how λ_+ compares to λ_-, for example by subtraction $\lambda_+ - \lambda_-$ as proposed by [34], which will be used to define gradients in multivalued color *invariant* images in the next section.

B. Gradients in Multivalued Color Invariant Images

In this section, we propose color invariant gradients based on the multiband approach as described in the previous section.

The color gradient for RGB is as follows:

$$\nabla\mathcal{C}_{RGB} = \sqrt{\lambda_+^{RGB} - \lambda_-^{RGB}} \quad (51)$$

for

$$\lambda_{\pm} = \frac{g_{11}^{RGB} + g_{22}^{RGB} \pm \sqrt{(g_{11}^{RGB} - g_{22}^{RGB})^2 + 4(g_{12}^{RGB})^2}}{2} \tag{52}$$

where

$$\begin{aligned} g_{11}^{RGB} &= \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2, \\ g_{22}^{RGB} &= \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2, \\ g_{12}^{RGB} &= \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y}. \end{aligned}$$

Further, we propose that the color invariant gradient (based on $c_4c_5c_6$) for matte objects is given by

$$\nabla\mathcal{C}_{c_4c_5c_6} = \sqrt{\lambda_{+}^{c_4c_5c_6} - \lambda_{-}^{c_4c_5c_6}} \tag{53}$$

for

$$\lambda_{\pm} = \frac{g_{11}^{c_4c_5c_6} + g_{22}^{c_4c_5c_6} \pm \sqrt{(g_{11}^{c_4c_5c_6} - g_{22}^{c_4c_5c_6})^2 + 4(g_{12}^{c_4c_5c_6})^2}}{2} \tag{54}$$

where

$$\begin{aligned} g_{11}^{c_4c_5c_6} &= \left|\frac{\partial c_4}{\partial x}\right|^2 + \left|\frac{\partial c_5}{\partial x}\right|^2 + \left|\frac{\partial c_6}{\partial x}\right|^2, \\ g_{22}^{c_4c_5c_6} &= \left|\frac{\partial c_4}{\partial y}\right|^2 + \left|\frac{\partial c_5}{\partial y}\right|^2 + \left|\frac{\partial c_6}{\partial y}\right|^2, \\ g_{12}^{c_4c_5c_6} &= \frac{\partial c_4}{\partial x}\frac{\partial c_4}{\partial y} + \frac{\partial c_5}{\partial x}\frac{\partial c_5}{\partial y} + \frac{\partial c_6}{\partial x}\frac{\partial c_6}{\partial y}. \end{aligned}$$

Similarly, we propose that the color invariant gradient (based on $l_4l_5l_6$) for shiny objects is given by

$$\nabla\mathcal{C}_{l_4l_5l_6} = \sqrt{\lambda_{+}^{l_4l_5l_6} - \lambda_{-}^{l_4l_5l_6}} \tag{55}$$

for

$$\lambda_{\pm} = \frac{g_{11}^{l_4l_5l_6} + g_{22}^{l_4l_5l_6} \pm \sqrt{(g_{11}^{l_4l_5l_6} - g_{22}^{l_4l_5l_6})^2 + 4(g_{12}^{l_4l_5l_6})^2}}{2} \tag{56}$$

where

$$\begin{aligned} g_{11}^{l_4l_5l_6} &= \left|\frac{\partial l_4}{\partial x}\right|^2 + \left|\frac{\partial l_5}{\partial x}\right|^2 + \left|\frac{\partial l_6}{\partial x}\right|^2, \\ g_{22}^{l_4l_5l_6} &= \left|\frac{\partial l_4}{\partial y}\right|^2 + \left|\frac{\partial l_5}{\partial y}\right|^2 + \left|\frac{\partial l_6}{\partial y}\right|^2, \\ g_{12}^{l_4l_5l_6} &= \frac{\partial l_4}{\partial x}\frac{\partial l_4}{\partial y} + \frac{\partial l_5}{\partial x}\frac{\partial l_5}{\partial y} + \frac{\partial l_6}{\partial x}\frac{\partial l_6}{\partial y}. \end{aligned}$$

Note that $l_4l_5l_6$ varies with a change in material only, $c_4c_5c_6$ with a change in material and highlights, and RGB vary with a change in material, highlights, and geometry of an object. Based on these observation, we may conclude that $\nabla\mathcal{C}_{RGB}$ measures the presence of 1) shadow or geometry edges, 2) highlight edges, and 3) material edges. Further, $\nabla\mathcal{C}_{l_4l_5l_6}$ measures the presence of 1) highlight edges, 3) material edges. And $\nabla\mathcal{C}_{l_4l_5l_6}$ measures the presence of only 3) material edges.

Note that $l_4l_5l_6$ varies with a change in material only, $c_4c_5c_6$ with a change in material and highlights, and RGB vary with a change in material, highlights, and geometry of an object. Based on these observation, we may conclude that $\nabla\mathcal{C}_{RGB}$ measures the presence of 1) shadow or geometry edges, 2) highlight edges, and 3) material edges. Further, $\nabla\mathcal{C}_{l_4l_5l_6}$ measures the presence of 1) highlight edges, 3) material edges. And $\nabla\mathcal{C}_{l_4l_5l_6}$ measures the presence of only 3) material edges.

IV. Shape Invariants

In this section, shape invariants are discussed measuring geometric properties of a set of coordinates of an image object independent of a coordinate transformation. We discuss similarity and projective invariants.

A. *Similarity Invariant*

For image locations (x_1, y_1), (x_2, y_2), and (x_3, y_3), $g_E(\,)$ is defined as a function which is unchanged as the points undergo any two-dimensional (2–D) translation, rotation and scaling transformation, yielding the well-known similarity invariant:

$$g_E((x_1, y_1), (x_2, y_2), (x_3, y_3)) = \theta \tag{57}$$

where θ is the angle at image coordinate (x_1, y_1) between line $(x_1, y_1)(x_2, y_2)$ and $(x_1, y_1)(x_3, y_3)$.

B. *Projective Invariant*

For the projective case, geometric properties of the shape of an object should be invariant under a change in the point of view. From the classical projective geometry we know that the so called cross-ratio is independent of the projection viewpoint.

From [35], we derive the projective invariant $g_P(\,)$ defined as

$$\begin{aligned} &g_P((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)) \\ &\quad = \frac{\sin(\theta_1 + \theta_2)\,\sin(\theta_2 + \theta_3)}{\sin(\theta_2)\,\sin(\theta_1 + \theta_2 + \theta_3)} \end{aligned} \tag{58}$$

where $\theta_1, \theta_2, \theta_3$ are the angles at image coordinate (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.

Noise sensitivity and probabilistic analysis of using the cross ratio for model-based object recognition is discussed in [36].

V. Invariant Image Indexing

Let the reference image database consist of a set $\{I_k\}_{k=1}^{N_b}$ of color images. Invariant feature spaces are created for each image I_k to represent the distribution of quantized invariant values in a high-dimensional invariant space. In this section, invariant feature spaces are formed on the basis of photometric color invariants, geometric invariants and combination of both.

A. Color Invariant Histogram Formation

By using $c_4c_5c_6$ at a pixel as a direct index, a 3-D histogram is constructed in a standard way on the c_4, c_5, and c_6 axes as shown in (59), at the bottom of the page, where η indicates the number of times c_4, c_5, and c_6 equals the value of index $(i,\, j,\, k)$. N is the total number of image locations. $\wedge$ denotes the logical AND.

$$\mathcal{H}_A(i,\, j,\, k) \hat{=} \frac{\eta((c_4(R,\, G,\, B) = i) \wedge (c_5(R,\, G,\, B) = j) \wedge (c_6(R,\, G,\, B) = k))}{N} \qquad \text{for } \forall\, \vec{x} \in I \tag{59}$$

The total accumulation for a particular histogram bin represents a measure of the area of a uniformly colored surface patch being imaged. Because each nonzero bin indicates the presence of a distinctively colored patch, the histogram is indicative for the color variety of the object in view independent of object geometry, shadows, and camera viewpoint.

The 3-D histogram of $l_4l_5l_6$ is defined as in (60), shown at the bottom of the page, where η indicates the number of times l_4, l_5, and l_6 equals the value of index $(i,\, j,\, k)$.

$$\mathcal{H}_B(i,\, j,\, k) \hat{=} \frac{\eta((l_4(R,\, G,\, B) = i) \wedge (l_5(R,\, G,\, B) = j) \wedge (l_6(R,\, G,\, B) = k))}{N} \qquad \text{for } \forall\, \vec{x} \in I \tag{60}$$

The histogram representing the distribution of $l_4l_5l_6$ edges is given by

$$\mathcal{H}_C(i) \hat{=} \eta(\nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}) = i) \tag{61}$$

only computed for locations $\vec{x} \in E^{I_k}$, where E^{I_k} is the set of $l_4l_5l_6$ edge maxima computed from image I_k. Edge maxima are obtained by applying nonmaximum suppression on the gradient to obtain local maxima in the gradient values [37].

The total accumulation for a particular bin represents a measure of the length of a certain color edge. For example, accumulation in a particular bin may represent the length of a yellow-green edge in the image. In this way, the measure of color area expressed by $\mathcal{H}_A$ and $\mathcal{H}_B$ is replaced with a measure of edge length.

B. Shape Invariant Histogram Formation

In this section, shape invariant histograms are constructed. We use $l_4l_5l_6$-based color invariant edges as feature points. These edges are viewpoint-independent, discounting shading, illumination intensity and direction, shadows and highlights.

A one-dimensional (1–D) histogram is constructed in a standard way on the angle axis expressing the distribution of angles between color invariant edge triplets mathematically specified by

$$\mathcal{H}_D(i) \hat{=} \eta(g_E(\vec{x}_1,\, \vec{x}_2,\, \vec{x}_3) = i) \tag{62}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \in E^{I_k}$, where E^{I_k} is the set of edge maxima computed from I_k and $g_E(\,)$ is given by (57).

Thus, between each triplet of color edge maxima, the angle denoted by i is computed and used as an index. Hence, each particular bin sum can be seen as the number of color edge triplets generating the same angle.

In a similar way, a 1-D histogram is defined on the cross ratio axis expressing the distribution of cross ratios between color edge quintets

$$\mathcal{H}_E(i) \hat{=} \eta(g_P(\vec{x}_1,\, \vec{x}_2,\, \vec{x}_3,\, \vec{x}_4,\, \vec{x}_5) = i) \tag{63}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \neq \vec{x}_4 \neq \vec{x}_5 \in E^{I_k}$ and $g_P(\,)$ is defined by (58).

C. Composite Color and Shape Invariant Histogram Formation

In this section, photometric color and geometric invariants are combined to construct a high-dimensional invariant histogram.

A four-dimensional (4–D) histogram is created counting the number of color invariant edge triples with values i, j, and k generating angle l (similarity invariant):

$$\begin{aligned}\mathcal{H}_F(i,\, j,\, k,\, l) \hat{=} \eta(\nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_1) = {}& i \wedge \\ \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_2) = {}& j \wedge \\ \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_3) = {}& k \wedge \\ g_E(\vec{x}_1, \vec{x}_2, \vec{x}_3) = {}& l)\end{aligned} \tag{64}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \in E^{I_k}$, where E^{I_k} is the set of (color invariant) edge maxima computed from I_k and $\nabla \mathcal{C}_{l_4l_5l_6}(\vec{x})$ the value of the color edge at $(\vec{x})$.

Each histogram bin measures the number of color edge triplets generating a certain angle. For example, a particular bin accumulation may represent the number of red-blue, orange-blue, and yellow-green edges in an image generating the angle $\theta = 1/4\pi$. In this way, both color and shape invariants are used during histogram formation. As a consequence, each object in view should generate a highly object-specific histogram.

In a similar way, a six-dimensional (6–D) invariant histogram can be constructed considering the cross-ratio between color edges as follows:

$$\begin{aligned}\mathcal{H}_G(i,\, j,\, k,\, l,\, m,\, n) \hat{=} {}& \eta(\nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_1) = i \wedge \\ & \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_2) = j \wedge \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_3) = k \wedge \\ & \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_4) = l \wedge \nabla \mathcal{C}_{l_4l_5l_6}(\vec{x}_5) = m \wedge \\ & g_P(\vec{x}_1,\, \vec{x}_2,\, \vec{x}_3,\, \vec{x}_4,\, \vec{x}_5) = n).\end{aligned} \tag{65}$$

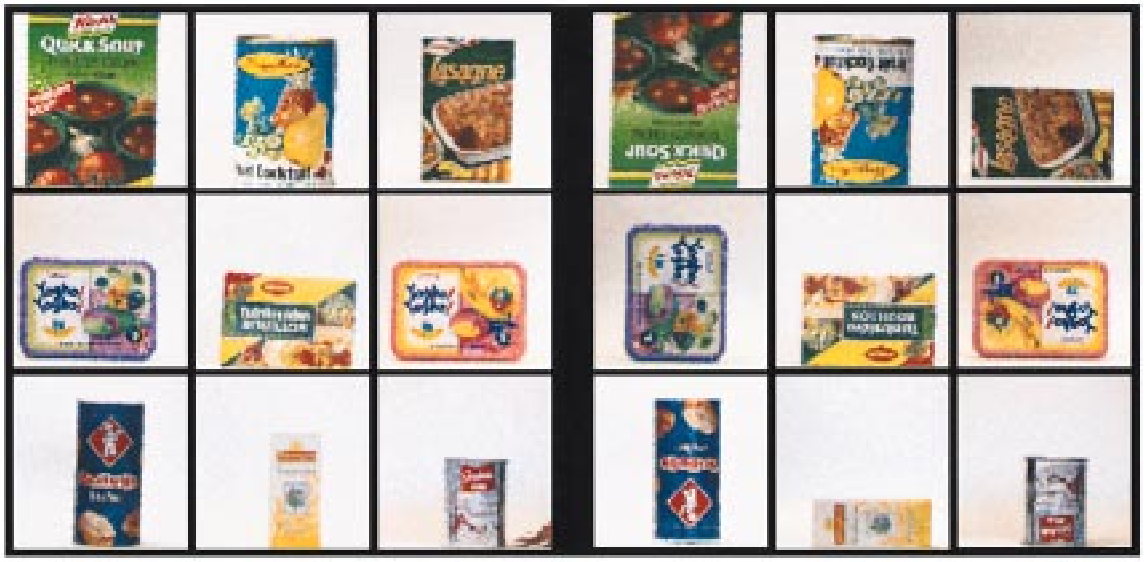


Fig. 1. Left: Various images which are included in the image database of 500 images. The images are representative for the images in the database. Right: Corresponding images from the query set.

VI. Invariant Image Retrieval

Color and shape invariants are computed from query image $\mathcal{Q}$ and used to create the query histogram $\mathcal{H}^{\mathcal{Q}}$. Then, $\mathcal{H}^{\mathcal{Q}}$ is matched against the same type of histogram precomputed and stored for each reference image in the database. For comparison reasons in the literature, matching is expressed by normalized histogram intersection as defined by

$$\mathcal{D}\left(\mathcal{H}_j^{\mathcal{Q}}, \mathcal{H}_j^{I_i}\right) = \frac{\sum_{k=1}^{N_{d_j}} \min\left\{\mathcal{H}_j^{\mathcal{Q}}(\vec{s}_k), \mathcal{H}_j^{I_i}(\vec{s}_k)\right\}}{N_{q_j}} \tag{66}$$

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

Note that normalized histogram intersection is robust to substantial object occlusion and cluttering [12]. In contrast, similarity functions based on eigenvalues or moments may run short in case of object occlusion and cluttering, as they are defined as an integral property on the invariant feature distributions.

VII. Experiments

To evaluate color and shape invariant indexing and retrieval, the following issues will be addressed in this section: 1) the discriminative power of color invariant object indexes, shape invariant object indexes, and of combined color and shape invariant indexes; and 2) robustness of the image retrieval scheme to occlusion, clutter and a change in viewpoint.

The data sets on which the experiments will be conducted are described in Section VII-A. The same dataset has been used to compare different color models for object recognition [30], [38]. Error measures and performance criteria are given in Section VII-B and VII-C, respectively.

A. Datasets

The dataset consists of $N_1 = 500$ color images taken from multicolored man-made objects composed of a large variety of materials including plastic, textile, paper, wood, rubber, painted metal, and ceramic. The SONY XC-003P CCD color camera (3 chips) and the Matrox Magic Color frame grabber were used to record the objects. The objects were recorded in isolation (one per image) against a white cardboard background. The digitization was done in 8 b per color. Two light sources of average day-light color were used to illuminate the objects in the scene. There was no attempt to individually control the focus of the camera or the illumination. Objects were recorded at a pace of a few shots a minute. They show a considerable amount of noise, shadows, shading, specularities, and self occlusion resulting in a good representation of views from everyday life.

A second, independent set (the test set) of recordings was made of randomly chosen objects already in the database. These objects, $N_2 = 70$ in number, were recorded again (one per image) with a new, arbitrary position and orientation with respect to the camera [some recorded upside down, some rotated, some at different distances (different scale)].

In Fig. 1, various images from the image database of 500 images are shown on the left, whereas various images coming from the query set are shown on the right.

In the experiments, all pixels in a color image are discarded having intensity and saturation smaller then 5% of the total range otherwise calculation of rgb, $c_4c_5c_6$, hue, and $l_4l_5l_6$ become unstable, see Section II-E. Consequently, the white cardboard background as well as the grey, white, dark or nearly colorless parts of objects as recorded in the color image will not be considered in the matching process.

B. Error Measures

For a measure of recognition quality, let rank r^{Q_i} denote the position of the correct match for query image Q_i, $i = 1, \cdots, N_2$, in the ordered list of N_1 match values. The

Fig. 2. One of the ten images generating four images by blanking out $o \in \{50, 65, 80, 90\}$ percent of the total object area.

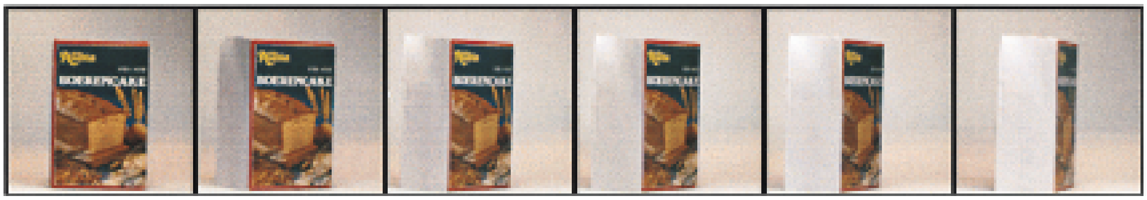


Fig. 3. One of the ten images generating four images by varying the angle between the camera for $s = \{45, 60, 75, 80\}$ degrees with respect to the object's surface normal (see the color plate for the color figures).

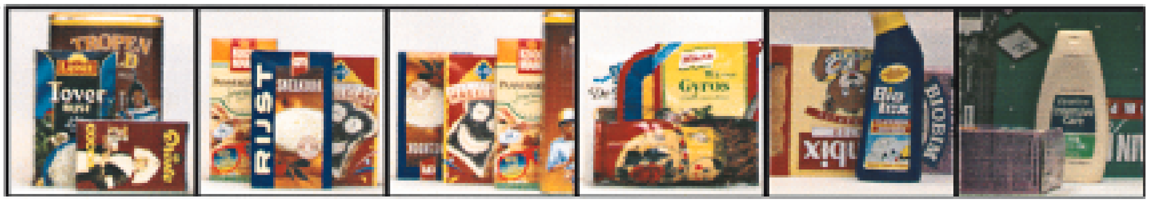


Fig. 4. Six of the 30 images taken from cluttered scenes.

rank r^{Q_i} ranges from $r = 1$ from a perfect match to $r = N_1$ for the worst possible match.

Then, for one experiment, the average ranking percentile is defined by

$$\overline{r} = \left(\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r^{Q_i}}{N_1 - 1} \right) 100\%. \tag{67}$$

The cumulative percentile of test images producing a rank smaller or equal to j is defined as

$$\mathcal{X}(j) = \left(\frac{1}{N_2} \sum_{k=1}^{j} \eta(r^{Q_i} == k) \right) 100\% \tag{68}$$

where η reads as the number of test images having rank k.

Further, let N^{Q_i} be the number of nonzero bins in the test histogram H^{Q_i}. Then the average number of nonzero bins $\overline{N} = (1/N_2) \sum_{i=1}^{N_2} N^{Q_i}$ determines the average run time complexity of the histogram matching process

$$O(N_1 \overline{N}) \tag{69}$$

where N_1 is the number of reference images in the image database.

C. Performance Criteria

Good performance is achieved when the recognition rate is high and the average run time complexity is low. To that end, the following criterion should be maximized: *the average ranking percentile* $\overline{r}$ (the discriminative power) resulting from matching the test set on the reference database; and the following criterion should be minimized: *the average number of nonzero bins* $\overline{N}$ (average run time complexity) to be used during histogram matching to compute the number of common hits between $\mathcal{H}^Q$ and $\mathcal{H}^{I_k}$.

D. Image Retrieval by Photometric Color Invariant Image Indexing

In this section, we report on the performance of the indexing and retrieval scheme for the $N_2 = 70$ test images on the database of $N_1 = 500$ reference images on the basis of photometric color invariants. To that end, attention is focussed on retrieval by histogram matching based on the following color-based histograms: $\mathcal{H}_A^{I_i}$, $\mathcal{H}_B^{I_i}$ and $\mathcal{H}_C^{I_i}$ as defined in Section V.

First, we will determine the appropriate bin size. We determine the appropriate bin size for our application empirically by varying the number of bins on the color invariant axes over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ and choose the smallest q for which the performance criteria, given in Section VII-C, are met. To that end, the average ranking percentile of $c_4c_5c_6$ denoted by $\overline{r}_{\mathcal{H}_A}$, $l_4l_5l_6$ denoted by $\overline{r}_{\mathcal{H}_B}$ and color edges denoted by $\overline{r}_{\mathcal{H}_C}$, is tested in relation to q (see Fig. 5). The influence of the number of bins on the average ranking percentile based on the different color invariants is the same: $\overline{r}_{\mathcal{H}_A}$ gives the same results as $\overline{r}_{\mathcal{H}_B}$ which are slightly better then $\overline{r}_{\mathcal{H}_C}$. Beyond $q \geq 16$, retrieval accuracy is constant, so it is concluded that $q = 16$ bins are sufficient for proper photometric color invariant object retrieval.

Second, the average number of nonzero bins determining the computational complexity for $c_4c_5c_6$ denoted by $\overline{N}_p$, $l_4l_5l_6$ given by $\overline{N}_e$ and color edges by $\overline{N}_c$ with respect to q is

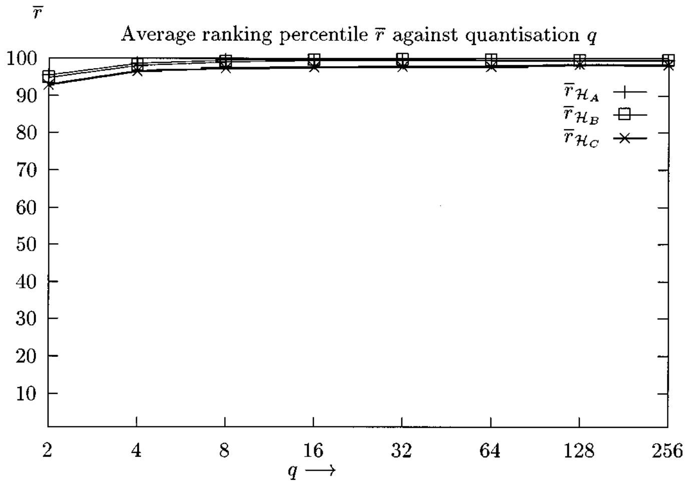


Fig. 5. Average ranking percentile of $c_4c_5c_6$ denoted by $\overline{r}_{\mathcal{H}_A}$, $l_4l_5l_6$ given by $\overline{r}_{\mathcal{H}_B}$ and color invariant edge maxima denoted by $\overline{r}_{\mathcal{H}_C}$, plotted against quantization q.

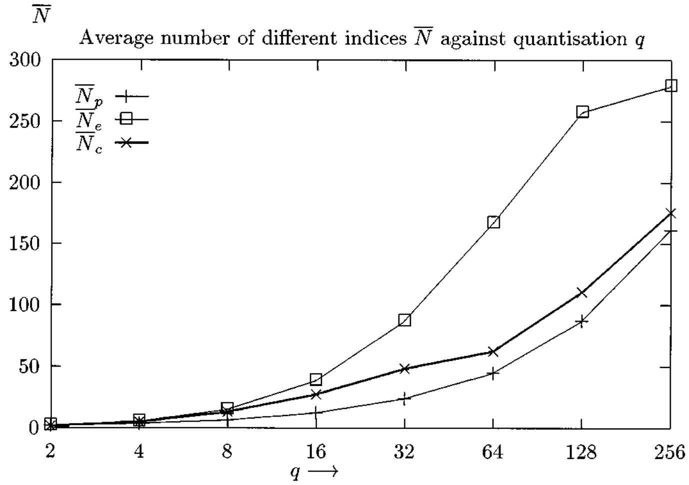


Fig. 6. Average number of nonzero bins for $c_4c_5c_6$ given by $\overline{N}_p$, $l_4l_5l_6$ denoted by $\overline{N}_e$ and color edges given by $\overline{N}_c$, plotted against quantization q.

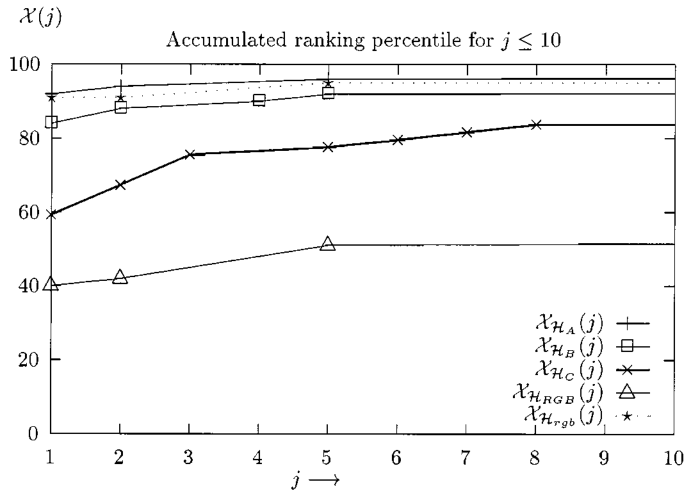


Fig. 7. Accumulated ranking $\mathcal{X}$ plotted against ranking j with $q = 16$ for $c_4c_5c_6$ denoted by $\mathcal{X}_{\mathcal{H}_A}$, $l_4l_5l_6$ denoted by $\mathcal{X}_{\mathcal{H}_B}$, color edges given by $\mathcal{X}_{\mathcal{H}_C}$, RGB given by $\mathcal{X}_{\mathcal{H}_{RGB}}$, and normalized color rgb denoted by $\mathcal{X}_{\mathcal{H}_{rgb}}$.

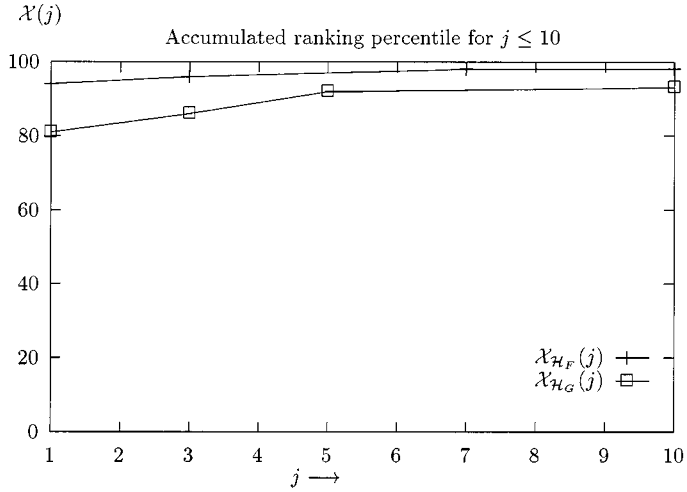


Fig. 8. Accumulated ranking plotted against ranking j with $q = 16$ for combined color-shape invariants $\mathcal{X}_{\mathcal{H}_F}$ and $\mathcal{X}_{\mathcal{H}_G}$.

considered, see Fig. 6. From the results we can see that the rate of increase of $\overline{N}_e$ is twice as much as the one for $\overline{N}_p$ and $\overline{N}_c$.

To compromise between discriminative power and average run time complexity, $q = 16$ is used in the following.

Fig. 7 shows the accumulated ranking $\mathcal{X}$ for $q = 16$, averaged over all the test images differentiated for the various photometric color invariants. Excellent performance is shown for both $\mathcal{X}_{\mathcal{H}_A}$ and $\mathcal{X}_{\mathcal{H}_B}$, where, respectively, 92% and 87% of the position of the correct match in the ordered list of match values is within the first two and, respectively, 97% and 92% within the first five rankings. Misclassification occurs when the test image consists of very few (two or three) distinct color patches mostly arising from small objects. Hence, from the results it is shown that $c_4c_5c_6$ and $l_4l_5l_6$ perform more or less the same. Color invariant edges give slightly worse retrieval accuracy.

For comparison reasons, the accumulated ranking $\mathcal{X}$ has also been computed for RGB and normalized color rgb (see Fig. 7). From the results we can observe that the discriminative power of rgb and $c_4c_5c_6$ are similar. As expected, the discrimination power of RGB has the worst performance due to its sensitivity to varying imaging conditions, see also [30].

For $q = 16$, according to (69), the average run time complexity is $O(N_1\overline{N}_p)$, $O(N_1\overline{N}_e)$ and $O(N_1\overline{N}_c)$ for $\overline{N_p} = 18$, $\overline{N_e} = 38$ and $\overline{N_c} = 27$, respectively, see Fig. 6. $c_4c_5c_6$ give slightly better run time complexity then $l_4l_5l_6$.

E. Image Retrieval by Geometric Invariant Image Indexes

In this section, the discriminative power of similarity and projective invariant indices are examined.

To evaluate the discriminative power of the geometric invariant index, the following histograms, defined in Section V, are considered: $\mathcal{H}_D$ and $\mathcal{H}_E$. Histogram $\mathcal{H}_D$ gives the distribution of angles and $\mathcal{H}_E$ the distribution of cross ratios between color edges.

Average ranking percentile for $\mathcal{H}_D$ and $\mathcal{H}_E$, denoted by $\overline{r}_{\mathcal{H}_D}$ and $\overline{r}_{\mathcal{H}_E}$, respectively, is shown for different $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ in Fig. 9. The average number of nonzero bins $\overline{N}_D$ (similarity) and $\overline{N}_E$ (cross ratio) is shown is Fig. 10.

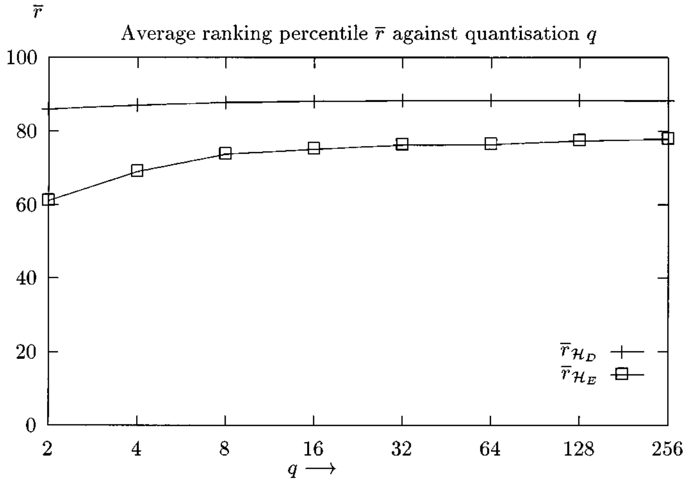


Fig. 9. Average ranking percentile for similarity $\mathcal{H}_D$ and cross-ratio $\mathcal{H}_E$ plotted against quantization q.

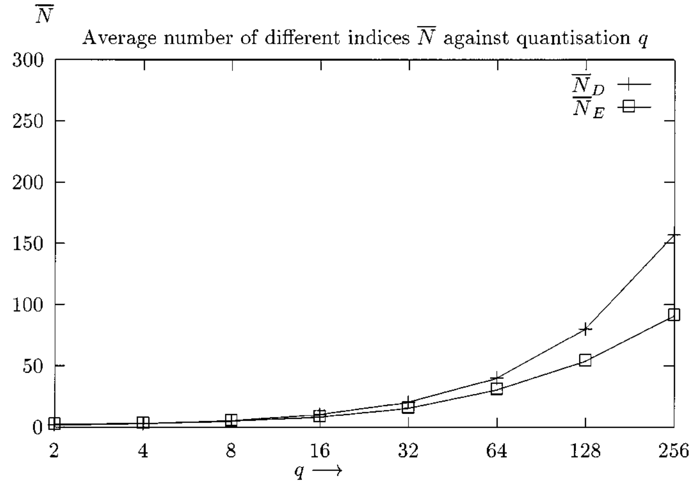


Fig. 10. Average number of nonzero bins for the similarity and cross ratio invariants plotted against quantization q.

Projective invariant values are noise sensitive [39] and less constrained (i.e., more coordinate combinations produce the same invariant value) and hence the discriminative performance expressed by $\bar{r}_{\mathcal{H}_E}$ is significantly worse than that of $\bar{r}_{\mathcal{H}_D}$. Note that the discriminative power of photometric color invariant image indices from the previous section is significantly better than shape based matching. Where average ranking percentile for $c_4c_5c_6$ and $l_4l_5l_6$ is approximately 94% for $q = 16$ within the first ten rankings, see Fig. 7, the average ranking percentile of the similarity invariant is 84% and 72% for cross ratios.

To compromise between the two performance criteria, $q = 16$ is taken for $\mathcal{H}_D$ and $\mathcal{H}_E$ in the following.

F. Image Retrieval by Composite Color and Shape Invariant Image Indexes

In this section, the discriminative power of the combination of shape and color invariant histogram matching is examined by considering $\mathcal{H}_F$ and $\mathcal{H}_G$ as defined in Section V during the histogram matching process. Note that there is no need for tuning parameter q, because $\mathcal{H}_F$ can be seen as the aggregation of $\mathcal{H}_C$ and $\mathcal{H}_D$, and $\mathcal{H}_G$ can be seen as the aggregation of $\mathcal{H}_C$ and $\mathcal{H}_E$ all with $q = 16$. The accumulated ranking $\mathcal{X}$ is shown in Fig. 8.

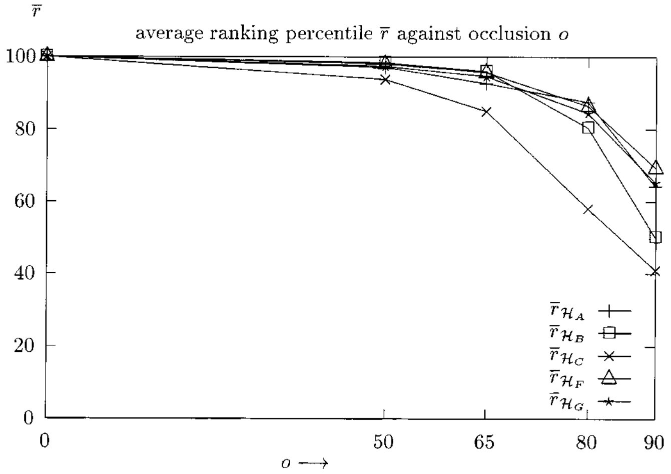


Fig. 11. Ranking percentile plotted against the percentage object area blanked out o denoted by $\bar{r}_{\mathcal{H}_A}$, $\bar{r}_{\mathcal{H}_B}$, $\bar{r}_{\mathcal{H}_C}$, $\bar{r}_{\mathcal{H}_F}$, and $\bar{r}_{\mathcal{H}_G}$.

Excellent discriminative accuracy is shown for $\mathcal{H}_F$ as 96% of the images are within the first two rankings, and 98% within the first nine rankings. $\mathcal{H}_G$ gives very good retrieval accuracy as 92% of the images are within the first five rankings.

G. Stability to Occlusion and a Change in Viewpoint

To test the effect of occlusion on the retrieval process, ten objects, already in the database of 500 recordings, were randomly selected and in total 40 images were generated by blanking out $o \in \{50, 65, 80, 90\}$ percent of the total object are (see Fig. 2). Note that white as recorded in the color image will not be considered in the matching process.

The ranking percentile $\bar{r}_{\mathcal{H}_A}$, $\bar{r}_{\mathcal{H}_B}$, $\bar{r}_{\mathcal{H}_C}$, $\bar{r}_{\mathcal{H}_F}$, and $\bar{r}_{\mathcal{H}_G}$,. averaged over the ten histogram matching values, is shown in Fig. 11.

From the results, we see that the shape and decrease of the curves for $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$, $\mathcal{H}_F$, and $\mathcal{H}_G$ do not differ significantly: namely a gradual decrease in retrieval accuracy beyond 50% blanking.

To test the effect of a change in viewpoint, the ten flat objects were put perpendicularly in front of the camera and in total 40 recordings were generated by varying the angle between the camera for $s = \{45, 60, 75, 80\}$ degrees with respect to the object's surface normal (see Fig. 3). Average ranking percentile is shown in Fig. 12.

Looking at the results, the rate of decrease is almost negligible for viewing angles up to 75°. Even when the object-side is nearly vanishing from sight, retrieval is still acceptable.

H. Discriminative Power in the Presence of Object Clutter

Another important claim is that the proposed method for object retrieval is fairly insensitive to object clutter. To test the effect of object cluttering, 30 images have been recorded from cluttered scenes. Each cluttered scene contained different multicolored objects (see Fig. 4).

Then, ten objects were randomly selected which participated in exactly one of the cluttered scenes. These objects were

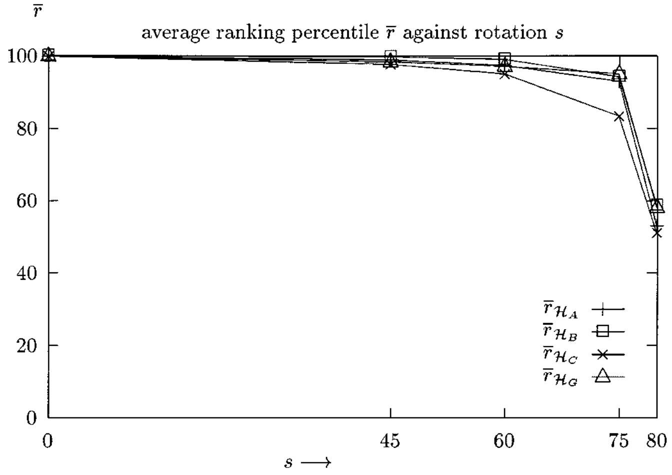


Fig. 12. Ranking percentile plotted against the angle of rotation s denoted by $\bar{r}_{\mathcal{H}_A}$, $\bar{r}_{\mathcal{H}_B}$, $\bar{r}_{\mathcal{H}_C}$, and $\bar{r}_{\mathcal{H}_G}$.

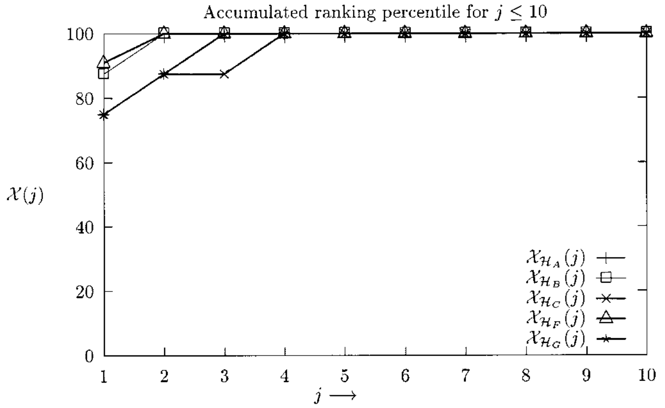


Fig. 13. Discriminative power plotted against the ranking j for $\bar{r}_{\mathcal{H}_A}$, $\bar{r}_{\mathcal{H}_B}$, $\bar{r}_{\mathcal{H}_C}$, $\bar{r}_{\mathcal{H}_F}$, and $\bar{r}_{\mathcal{H}_G}$.

recorded in isolation against a white background yielding the test set. The test set has been matched against the database of 30 images. Fig. 13 shows the accumulated average ranking percentile for different invariant indexes.

From the results it can be observed that the invariant indexes are fairly insensitive to object clutter.

VIII. Discussion

When the performance of different invariant indices is compared, histogram matching based on both shape and color invariants produces the highest discriminative power: 96% of the images are within the first two rankings, and 98% within the first nine rankings. Image retrieval based entirely on shape invariants yields poor discriminative power. As opposed to shape invariant matching, color invariant based histogram matching results in very high discriminative performance. While the average ranking percentile for $c_4 c_5 c_6$ and $l_4 l_5 l_6$ is 94% , the average ranking percentile of the similarity invariant is 84% and 72% for cross ratios.

Furthermore, the experimental results reveal that identifying multicolored objects on the basis of only photometric color invariants, and the combination of shape and color invariants, is to a very large degree robust to partial occlusion, object clutter and a change in viewing position.

In the next section, the image retrieval scheme is integrated into the PicToSeek system for searching images on the World Wide Web.

IX. PicToSeek: A Content-Based Image Search System

We have implemented a content-based image search system, called PicToSeek, for exploring visual information on the World Wide Web. In the first stage, PicToSeek collects images on the World Wide Web by means of autonomous Web-crawlers. Then, the collected images are automatically cataloged into various image styles and types: JFIF-GIF, grey-color, size, date of creation, and color depth. Further, the system automatically classifies (by supervised learning) images into the following classes: photograph-synthetic, (photographs) indoor-outdoor, (photographs) portraits, and (synthetics) buttons. After cataloging images, the proposed invariant image features are extracted from the images to produce a high-dimensional image index independent of the accidental imaging conditions. When images are automatically collected, cataloged and indexed, PicToSeek allows for fast on-line image search by combining: 1) visual browsing through the precomputed image catalogue, 2) query by pictorial example, and 3) query by image features. The content-based image retrieval process is conducted in an interactive, iterative manner guided by the user by relevance feedback.

In Section IX-A, an overview of the system is given. In Section IX-B, the implementation of PicToSeek is discussed. Finally, the query capability of the system is outlined in Section IX-C. PicToSeek is on-line at http://www.wins.uva.nl/research/isis/zomax/. A more detailed report on PicToSeek appeared in [1].

A. System Overview

The major components of the PicToSeek system are described in detail below.

1) Interactive Query Formulation: An image is sketched, recorded or selected from a repository. This is the query definition with the aim to find a similar image in the database. Note that “similar image” may imply a partially identical image (as in the case of finding stamps), or a partially identical object in the image (as in the case of a stolen goods database), or a similar styled image (as in the case of a fashion design support system).

PicToSeek offers snakes for interactive image segmentation, described in [40], for the purpose of content-based image retrieval by query-by-example. We proposed the use of color *invariant* gradient information to guide the deformation process to obtain snake boundaries which correspond to material boundaries in images discounting the disturbing influences of surface orientation, illumination, shadows, and highlights. The key idea is to allow the user to specify in an interactive way salient subimages of objects on which the image object search will be based. In this way, confounding and misleading image information is discarded. In conclusion, PicToSeek offers interactive query formulation either by query (sub)image(s) or by offering a pattern of feature values and weights.

2) Image Features: PicToSeek allows the user to choose the desired classes of invariance. For each image retrieval query a proper definition of the desired invariance is in order. Does the applicant wish search for the object in rotation and scale invariance? Illumination invariance? Viewpoint invariance? Occlusion invariance? In the current state of the art of query engines, invariance receives little attention. But for large databases, the availability at the time of query definition is essential. The shape and color invariants proposed in this paper are the core of the PicToSeek system.

3) Feature Representation and Weighting: The image feature sets are represented by n-dimensional feature space. In this way, the domain dependent part of the whole image retrieval system is reduced to a minimum.

To be precise, let an image I be represented by its *image feature vectors* of the form $I = (f_0, w_{I0}; f_1, w_{I1}; \cdots; f_t, w_{It})$ and a typical query Q by $Q = (f_0, w_{Q0}; f_1, w_{Q1}; \cdots; f_t, w_{Qt})$, where w_{Ik} (or w_{Qk}) represent the weight of image feature f_k in image I (or query Q), and t image features are used for image object search. The weights are assumed to be between zero and one.

Weights can be assigned corresponding to the feature frequency ff as defined by

$$w_i = \mathrm{ff}_i \tag{70}$$

giving the well-known histogram form where ff_i (feature frequency) is the frequency of occurrences of the image feature values i in the image or query. However, for accurate image object search, it is desirable to assign weights in accordance to the importance of the image features. To that end, the image feature weights used for both images and queries are computed as the product of the features frequency multiplied by the inverse collection frequency factor, defined by [41]

$$w_i = \left(0.5 + \frac{0.5\mathrm{ff}_i}{\max\{\mathrm{ff}\}_{i=1}^{t}}\right)\log\left(\frac{N}{n}\right) \tag{71}$$

where N is the number of images in the database and n denotes the number of images to which a feature value is assigned. In this way, features are emphasized having high feature frequencies but low overall collection frequencies.

4) Searching: In the field of pattern recognition, several methods have been proposed that improve classification automatically through experience such as artificial neural networks, decision tree learning, Bayesian learning, and k-nearest neighbor classifiers. Except for the k-nearest neighbor classifier, the other methods construct a general, explicit description of the target function when training examples are provided. In contrast, k-nearest neighbor classification consist of finding the relationship to the previously stored images each time a new query image is given. When a new query is given by the user, a set of similar related images is retrieved from the image database and used to classify the new query image. The advantage of k-nearest neighbor classification is that the technique construct a local approximation to the target function that applies in the neighborhood of the new image query images, and never construct an approximation designed to perform well over the entire instance space. To that end, PicToSeek uses the k-nearest neighbor classifier for image search.

5) Relevance Feedback: Relevance feedback is an automatic process designed to produce improved query formulations following an initial retrieval operation. Relevance feedback is needed for image retrieval where the users find it difficult to formulate pictorial queries which are well designed for accurate retrieval purposes. For example, without any specific query image example, the user might find it difficult to formulate a query (e.g., to retrieve an image of a car) by an image sketch or by offering a pattern of feature values and weights. This suggests that the first search operation should be conducted with a tentative, initial query formulation, and should be processed as a trial search. These initially retrieved images should then be examined for relevance, and a (new) improved query formulation should be constructed with the purpose to retrieve more relevant images in subsequent search operations. The system use the feature weighting given by the user to find the images in the image database which are most similar with respect to the feature weighting.

B. Implementation

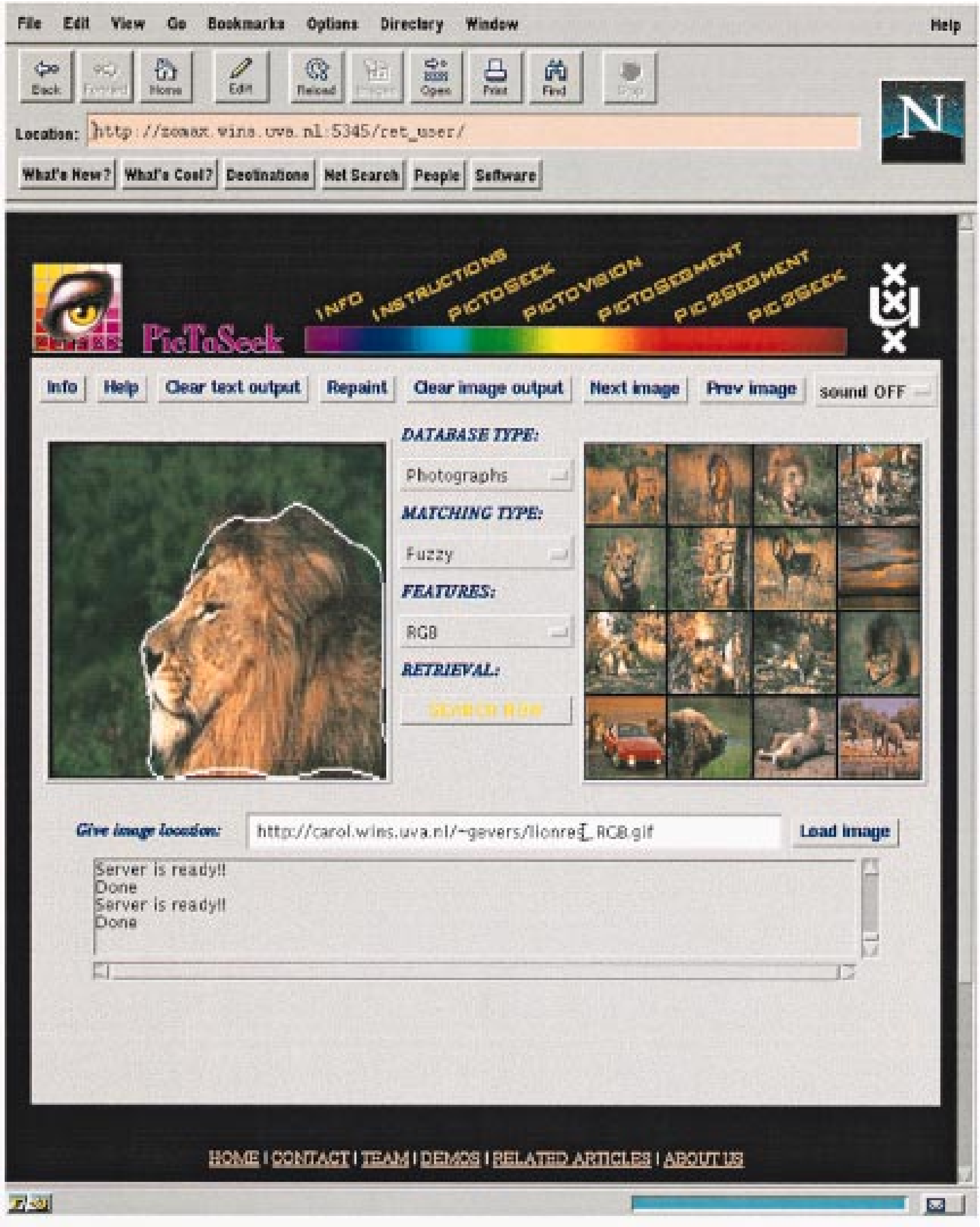
The PicToSeek system is based on a client-server paradigm. The client part is a Java Applet and correspond to the graphical user interface. The client part takes care of interactive query formulation, the display of the results, and the relevance feedback specification given by the user. The server part of PicToSeek takes care of the image feature extraction, feature weighting from relevance feedback, k-nearest neighbor feature classification, and image sorting. The server is implemented in C. The interface between client (Java) and server (C) is written in Java. The Web-crawler, image analysis and feature extraction methods have been implemented in C.

The client and server components are described here more in detail.

1) Client Site: Using a standard web-browser, the PicToSeek Applet is sent to the client. After the Applet has started, the user can load any image available at the WWW by giving the URL address. After the user has loaded an image, the user is allowed to specify (sub)images by the interactive snake segmentation method. After interactive query formulation, the user specifies the preferred invariance, and the similarity measure. Then, the image query formulation is send to the server. In conclusion, the client-part is a Java Applet and can be started by a standard web browser. The Java Applet allows the user to

1) select/load an (external) image;
2) select appropriate subimages of objects (instead of the entire image) on which the image object search will be conducted;
3) select color features (invariants) and similarity measure;
4) send the query formulation to the server.

2) Server Site: The server receives the query image formulation send by the client. After receiving the query image, the server convert the image to the desired format, enabling the image processing routines, implemented in C, to extract the required invariant image features. Query image features are weighted. In this way, features are emphasized having



Location: http://zomax.wins.uva.nl:5345/ret_user/

DATABASE TYPE: Photographs

MATCHING TYPE: Fuzzy

FEATURES: RGB

RETRIEVAL:

Give image location: http://carol.wins.uva.nl/~gevers/lionre_RGB.gif

```
Server is ready!!
Done
Server is ready!!
Done
```

Fig. 14. Content-based image retrieval by query-by-example based on the region denoting the lion (without the background) as specified by the user.

high feature frequencies but low overall collection frequencies. K-nearest neighbors are found in this weighted vector representation. The k-nearest neighbors are sorted with respect to their similarity and send back to the client for display. In conclusion, the server receives the image query formulation from the client. Then, the following operations are performed:

1) image feature extraction;
2) image feature weighting;
3) k-nearest neighbors are found and sorted;
4) results are send back to the client for display.

C. Query Scenario

All queries follow the same scenario, listed here.

Step 1) *Image domain selection:* Visual browsing through the precomputed image catalogue;
Step 2) *Image selection:* select an image from the catalogue or capture the query image from an object by giving a URL address.
Step 3) *Query image:* the query image is defined as an user-specified interesting part of the selected image.

Step 4) *Invariance selection:* the required invariance is selected from the list of available invariant indices.

Step 5) *Search:* the same invariant indices are computed from the query and matched with those stored in the database.

Step 6) *Display:* an ordered list of most similar images is shown.

Step 7) *Image Selection:* if the right image is found, the image can be displayed at full resolution.

Step 8) *Rerun:* if the right image is not found the query image is adjusted (go to Step 1) or the most similar image is used to refine query definition (go to Step 3).

To illustrate the query capability of the system, typical applications are considered of retrieving images containing an instance of a given object. To that end, the query is specified by an example image taken from the object at hand. Typical query specifications are shown in Fig. 14. The images come from Corel © Stock Photo Libraries.

Consider Fig. 14, where the user has specified the region showing a lion. The region is used as the query. Images in the image database are compared to the lion query based on their color invariant information. After image matching, images are shown in order of resemblance to the user. Note that within the first 16 images, 12 images contain a lion.

X. Conclusion

In this paper, new sets of color models have been proposed invariant to the viewpoint, geometry of the object and illumination conditions. Color invariant edges have been proposed from which shape invariant features are computed. Computational methods are given to combine color and shape invariants into a unified high-dimensional invariant feature set for discriminatory object search.

From the theoretical and experimental results, it is concluded that object search based on composite color and shape invariant features provides excellent recognition accuracy. Object search based on color invariants provides very high retrieval accuracy whereas object search based entirely on shape invariants yields poor discriminative power. Furthermore, the image retrieval scheme is highly robust to partial occlusion, object clutter and a change in viewing position.

Finally, the image retrieval scheme is integrated into the PicToSeek system on-line at http://www.wins.uva.nl/research/isis/PicToSeek/ for searching images on the World Wide Web.

References

[1] T. Gevers and A. W. M. Smeulders, “PicToSeek: A content-based image search engine for the World Wide Web,” in *Proc. Visual Information Systems*, San Diego, CA, 1997, pp. 93–100.

[2] W. Grosky and R. Mehrotra, “Special issue on image database management,” *Computer*, vol. 22, no. 12, Sept. 1989.

[3] IFIP, *Visual Database Systems I and II*, Amsterdam, The Netherlands: Elsevier, 1989 and 1992.

[4] R. Jain, “NSF workshop on visual information management systems,” *SIGmod Record*, vol. 22, pp. 57–75, 1993.

[5] A. Pentland, R. W. Picard, and S. Sclaroff, “Photobook: Tools for content-based manipulation of image databases,” in *Proc. Storage and Retrieval for Image and Video Databases II*. Bellingham, WA: SPIE, 1994, vol. 2, pp. 34–47.

[6] W. Niblack and R. Jain, Eds., *Proc. Storage and Retrieval for Image and Video Databases I, II, and III*. Bellingham, WA: SPIE, 1993, 1994 and 1995, vol. 1,908; 2,185; and 2,420.

[7] *Proc.Visual Information Systems: The 1st Int. Conf.Visual Information Systems*, Melbourne, Vic., Australia, 1996.

[8] A. Califano and R. Mohan, “Multidimensional indexing for recognizing visual shapes,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 373–392, Apr. 1994.

[9] Y. Lamdan and H. J. Wolfson, “Geometric hashing: A general and efficient model-based recognition scheme,” in *Proc. 2nd ICCV*, 1988, pp. 238–249.

[10] I. Rigoutsos and R. Hummel, “On a scalable parallel implementation of geometric hashing on the connection machine,” Courant Inst. Math. Science, New York Univ., New York, Tech. Rep. 554, 1991.

[11] F. Stein and G. Medioni, “Structural indexing: Efficient 2-D object recognition,” *IEEE Trans.Pattern Anal. Machine Intell.*, vol. 14, pp. 1198–1204, Dec. 1992.

[12] M. J. Swain and D. H. Ballard, “Color indexing,” *Int. J. Comput. Vis.*, vol. 7, pp. 11–32, Nov. 1991.

[13] H. J. Wolfson, “Object recognition by transformation invariant indexing,” in *Proc. Invariance Workshop, ECCV*, 1992.

[14] B. V. Funt and G. D. Finlayson, “Color constant color indexing,” *IEEE Trans.Pattern Anal. Machine Intell.*, vol. 17, pp. 522–529, May 1995.

[15] S. K. Nayar and R. M. Bolle, “Reflectance based object recognition,” *Int. J. Comput. Vis.*, vol. 17, pp. 219–240, Mar. 1996.

[16] G. D. Finlayson, S. S. Chatterjee, and B. V. Funt, “Color angular indexing,” in *ECCV’96*, 1996, pp. 16–27.

[17] G. Healey and D. Slater, “Global color constancy: Recognition of objects by use of illumination invariant properties of color distributions,” *J. Opt. Soc. Amer.*, vol. 11, pp. 3003–3010, Nov. 1995.

[18] D. Slater and G. Healey, “The illumination-invariant recognition of 3D objects using local color invariants,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 206–210, Feb. 1996.

[19] T. Gevers and A. W. M. Smeulders, “Enigma: An image retrieval system,” in *Proc.11th ICPR*, 1992, pp. 697–700.

[20] M. Flickner *et al.*, “Query by image and video content: The QBIC system,” *Computer*, vol. 28, pp. 23–33, Sept. 1995.

[21] R. Mehrotra and J. E. Gary, “Similar-shape retrieval in shape data management,” *Computer*, vol. 28, pp. 7–14, Sept. 1995.

[22] V. E. Ogle and M. Stonebraker, “Chabot: Retrieval from a relational database of images,” *Computer*, vol. 28, pp. 40–49, Sept. 1995.

[23] R. H. Srihari, “Automatic indexing of content-based retrieval of captioned images,” *Computer*, vol. 28, pp. 49–56, 1995.

[24] A. K. Jain and A. Vailaya, “Image retrieval using color and shape,” *Pattern Recognit.*, vol. 29, pp. 1233–1244, 1996.

[25] S. Sclaroff, L. Taycher, and M. La Cascia, “ImageRover: A content-based image browser for the World Wide Web,” in *Proc. IEEE Workshop on Content-based Access and Video Libraries, CVPR*, 1997.

[26] C. Frankel, M. Swain, and A. Webseer, “An image search engine for the World Wide Web,” Univ. Chicago, Chicago, IL, Tech. Rep. TR-96-14, 1996.

[27] J. R. Smith and S.-F. Chang, “VisualSEEK: A fully automated content-based image query system,” in *Proc. ACM Multimedia*, 1996.

[28] A. Gupta, “Visual information retrieval technology: A Virage perspective,” Virage Inc., TR 3A, 1996.

[29] S. A. Shafer, “Using color to separate reflection components,” *Color Res. Appl.*, vol. 10, pp. 210–218, 1985.

[30] T. Gevers and A. W. M. Smeulders, “Color based object recognition,” *Pattern Recognit.*, vol. 32, pp. 453–465, Mar. 1999.

[31] H. Levkowitz and G. T. Herman, “GLHS: A generalized lightness, hue, and saturation color model,” *Comput. Vis. Graph. Image Process.: Graph. Models Image Process.*, vol. 55, pp. 271–285, 1993.

[32] “Saturation, hue, and normalized colors: Calculation, digitization effects, and use, Tech Rep.,” Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, 1976.

[33] S. di Zenzo, “Gradient of a multi-images,” *Comput. Vis. Graph. Image Process.*, vol. 33, pp. 116–125, 1986.

[34] G. Sapiro and D. L. Ringach, “Anisotropic diffusion of multi-valued images with applications to color filtering,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 1582–1586, Nov. 1996.

[35] O. Veiblen and J. W. Young, *Projective Geometry*. Boston, MA: Ginn., 1910.

[36] S. J. Maybank, “Probabilistic analysis of the application of the cross ratio to model based vision,” *Int. J. Comput. Vis.* , vol. 16, pp. 5–33, Sept. 1995.

[37] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, Nov. 1986.

[38] T. Gevers, “Color image invariant segmentation and retrieval,” Ph.D. dissertation, Univ. Amsterdam, The Netherlands, May 1996.

[39] C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy, “Planar object recognition using projective shape representation,” *Int. J. Comput. Vis.*, vol. 16, pp. 57–99, 1995.

[40] T. Gevers and A. W. M. Smeulders, “Interactive query formulation for object search,” in *Proc. Visual Information Systems*, Amsterdam, The Netherlands, 1999.

[41] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Process. Manage.*, 1988.

Theo Gevers is an Assistant Professor of Computer Science at the University of Amsterdam, The Netherlands. His main research interests are in the fundamentals of image database system design, image retrieval by content, theoretical foundation of geometric and photometric invariants and color image processing. He has led several national and international projects and acts as a reviewer. He has published more than 40 papers on color image processing, physics-based vision, content-based image retrieval and image database design.

Dr. Gevers is co-organizer of the First International Workshop on Image Databases and Multimedia Search and the Third International Conference on Visual Information Systems.

Arnold W. M. Smeulders (S’80–M’82) is a Professor of computer science in multimedia information analysis at the University of Amsterdam, The Netherlands, where he also heads the Intelligent Sensory Information Systems Group. He has been in computer vision since 1975. He has published more than 200 papers and 200 conference contributions, mostly on vision and recognition, with a new emphasis on multimedia analysis. His current research interests are in industrial vision from specification, color vision, image search by pictorial example and image databases, intelligent interactive analysis, and system design aspects of multimedia systems. He is particularly interested in the correspondence between language and picture.

He is co-chair of IAPR’s TC12 on Multimedia, Associate Editor for IEEE TRANSACTIONS ON PAMI , and the journal *Cytometry*, and a member of the Visual Information Systems steering committee. He is also director of the Research Institute of Computer Science and Department Head of the University of Amsterdam, and Director of the Intelligent Systems Lab at Amsterdam.