Image Indexing using Composite Color and Shape Invariant Features

T. Gevers & A. W. M. Smeulders
ISIS, University of Amsterdam, faculty WINS
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
{gevers, smeulders}@wins.uva.nl

Abstract

New sets of color models are proposed for object recognition invariant to a change in view point, object geometry and illumination. Further, computational methods are presented to combine color and shape invariants to produce a high-dimensional invariant feature set for discriminatory object recognition.

Experiments on a database of 500 images show that object recognition based on composite color and shape invariant features provides excellent recognition accuracy. Furthermore, object recognition based on color invariants provides very high recognition accuracy whereas object recognition based entirely on shape invariants yields very poor discriminative power.

The image database and the performance of the recognition scheme can be experienced within PicToSeek on-line as part of the ZOMAX system at: **http://www.wins.uva.nl/research/isis/zomax/**.

1 Introduction

Most of the work on object recognition is based on matching sets of geometric image features (e.g. edges, lines and corners) to 3D object models and significant progress has been achieved, [5], for example. As an expression of the difficulty of the general problem, most of the geometry-based schemes can handle only simple, flat and rigid man-made objects. Geometric features are rarely adequate for discriminatory object recognition of 3-D objects from arbitrary viewpoints.

Color provides powerful information for object recognition as well. A simple and effective recognition scheme is to represent and match images on the basis of color-metric histograms as proposed by Swain and Ballard [7]. This method is extended by Funt and Finlayson [2] to make the method illumination independent by indexing on an invariant set of color descriptors computed from neighboring points. However, objects should be composed of flat surfaces and the method may fail when images are contaminated by shading cues. Further, Finlayson *at al.* [1], and Healey and Slater [4] use illumination-invariant moments for object recognition. The methods fail, however, when objects are occluded as the moments are defined as an integral property on the object as one.

In this paper, our aim is to propose new image features for 3D object recognition according to the following criteria: invariance to the geometry of the object and illumination conditions, high discriminative power and low computational effort, and robustness against fragmented, occluded and overlapping objects. First, in Section 2, assuming white illumination and dichromatic reflectance, we propose new color models invariant to a change in view point, object geometry and illumination. Also a change in spectral power distribution of the illumination is considered to propose a new set of color models which is an invariant for matte objects. Simple shape invariants are discussed in Section 3. In Section 4, we propose computational methods to produce composite color and shape invariant features. Finally, in Section 5, the performance of different invariant image features is evaluated on a database of 500 images. No constraints are imposed on the images in the image database and the imaging process other than that images should be taken from multicolored objects.

2 Photometric Color Invariance

Consider an image of an infinitesimal surface patch. Using the red, green and blue sensors with spectral sensitivities given by $f_R(\lambda)$, $f_G(\lambda)$ and $f_B(\lambda)$ respectively, to obtain an image of the surface patch illuminated by a SPD of the incident light denoted by $e(\lambda)$, the measured sensor values will be given by Shafer [6]:

$$C = m_b(\vec{n}, \vec{s}) \int_\lambda f_C(\lambda) e(\lambda) c_b(\lambda) d\lambda + m_s(\vec{n}, \vec{s}, \vec{v}) \int_\lambda f_C(\lambda) e(\lambda) c_s(\lambda) d\lambda \quad (1)$$

for $C = \{R, G, B\}$ giving the Cth sensor response. Further, $c_b(\lambda)$ and $c_s(\lambda)$ are the albedo and Fresnel reflectance respectively. λ denotes the wavelength, $\vec{n}$

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 m_b and m_s denote the geometric dependencies on the body and surface reflection respectively.

2.1 Reflectance with White Illumination

Considering the neutral interface reflection (NIR) model (assuming that $c_s(\lambda)$ has a constant value independent of the wavelength) and "white" illumination, then $e(\lambda) = e$ and $c_s(\lambda) = c_s$. Then, we propose that the measured sensor values are given by:

$$C_w = em_b(\vec{n}, \vec{s})k_C + em_s(\vec{n}, \vec{s}, \vec{v})c_s \int_\lambda f_C(\lambda)d\lambda \quad (2)$$

for $C_w \in \{R_w, G_w, B_w\}$ giving the red, green and blue sensor response under the assumption of a white light source. $k_C = \int_\lambda f_C(\lambda)c_b(\lambda)d\lambda$ is a compact formulation depending on the sensors and the surface albedo.

If the integrated white condition holds (as we assume throughout the paper): $\int_\lambda f_R(\lambda)d\lambda = \int_\lambda f_G(\lambda)d\lambda = \int_\lambda f_B(\lambda)d\lambda = f$, we propose:

$$C_w = em_b(\vec{n}, \vec{s})k_C + em_s(\vec{n}, \vec{s}, \vec{v})c_s f \quad (3)$$

2.2 Color Invariant Color Models

For a given point on a shiny surface, the contribution of the body reflection component $C_b = em_b(\vec{n}, \vec{s})k_C$ and surface reflection component $C_s = em_s(\vec{n}, \vec{s}, \vec{v})c_s f$ are added together $C_w = C_s + C_b$. Hence, the observed colors of the surface must be inside the triangular color cluster in the RGB-space formed by the two reflection components [3].

Hence, any expression defining colors on the same triangle, formed by the two reflection components, in RGB-space are photometric color invariants for the dichromatic reflection model with white illumination. To that end, we propose the following set of color models:

$$L^p = \frac{\sum_i a_i (C_i^1 - C_i^2)^p}{\sum_j a_j (C_j^3 - C_j^4)^p} \quad (4)$$

where $C_i^1 \neq C_i^2, C_j^3 \neq C_j^4 \in \{R, G, B\}, i, j, p \geq 1, a \in \mathcal{R}$.

Lemma 1 *Assuming dichromatic reflection and white illumination, L^p is independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and highlights.*

Proof: By substituting eq. (3) in eq. (4) we have:

$$\frac{\sum_i a_i((em_b(\vec{n}, \vec{s})k_{C_i^1}) - (em_b(\vec{n}, \vec{s})k_{C_i^2}))^p}{\sum_j a_j((em_b(\vec{n}, \vec{s})k_{C_j^3}) - (em_b(\vec{n}, \vec{s})k_{C_j^4}))^p} =$$

$$\frac{\sum_i a_i(em_b(\vec{n}, \vec{s}))^p(k_{C_i^1} - k_{C_i^2})^p}{\sum_j a_j(em_b(\vec{n}, \vec{s}))^p(k_{C_j^3} - k_{C_j^4})^p} =$$

$$\frac{(em_b(\vec{n}, \vec{s}))^p \sum_i a_i(k_{C_i^1} - k_{C_i^2})^p}{(em_b(\vec{n}, \vec{s}))^p \sum_j a_j(k_{C_j^3} - k_{C_j^4})^p} =$$

$$\frac{\sum_i a_i(k_{C_i^1} - k_{C_i^2})^p}{\sum_j a_j(k_{C_j^3} - k_{C_j^4})^p}$$

only dependent on the sensors and the material's albedo.

QED.

For instance, for $p = 1$, we have the set: $\{\frac{(R-G)}{(R-B)}, \frac{(B-G)}{(R-B)}, \frac{(R-G)+(B-G)}{(R-B)}, \frac{(R-G)+3(B-G)}{(R-B)+2(R-G)}, ...\}$ and for $p = 2$: $\{\frac{(R-G)^2}{(R-B)^2}, \frac{(B-G)^2}{(R-B)^2}, \frac{(R-G)^2+(B-G)^2}{(R-B)^2}, \frac{(R-G)^2+3(B-G)^2}{(R-B)^2+2(R-G)^2}, ...\}$ where all elements are photometric color invariants for objects with dichromatic reflectance under white illumination.

We can easily see that hue given by:

$$H(R, G, B) = \arctan\left(\frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}\right) \quad (5)$$

ranging from $[0, 2\pi)$ is a function of an instantiation of L^p with $p = 1$ i.e. $\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}$.

Although any other instantiation of L^p could be taken, in this paper, we concentrate on the photometric color invariant model $l_1 l_2 l_3$ uniquely determining the direction of the linear triangular color cluster: $l_1 = \frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$, $l_2 = \frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$, $l_3 = \frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$ the set of normalized color differences which is, similar to H, an invariant for objects with dichromatic reflectance and white illumination.

2.2.1 Color Invariant Image Features

In this section, we propose different image features (i.e. edges and corners) derived from the above proposed invariant color models. Although any instantiation of L^p could be taken to produce color invariant image features, in this paper, hue is taken as an instantiation of L^p to generate color invariant image features, because hue is intuitive and well-known in the color literature.

Color Invariant Edge Pairs:

Due to the circular nature of hue, the standard difference operator is not suited for computing the difference between hue values. Therefore, we define the difference between two hue values h_1 and h_2 as follows:

$$d(h_1, h_2) = \sqrt{(\cos h_1 - \cos h_2)^2 + (\sin h_1 - \sin h_2)^2} \quad (6)$$

yielding a difference $d(h_1, h_2) \in [0, 2]$ between h_1 and h_2. Note that the difference is a distance because it satisfies the metric criteria. To find hue edges we use an edge detector, currently of the Sobel type, to suppress marginally visible edges. Then, for each local hue maximum, two opposite neighboring points are computed based on the direction of the gradient to determine the hue value on the left side of the edge: $l_e^l(\vec{x}) = H(\vec{x} - \Delta\vec{n})$ and the right side of the edge: $l_e^r(\vec{x}) = H(\vec{x} + \Delta\vec{n})$. Only computed for image location $\vec{x}$ at the two sides of a local hue maximum. Furthermore, $\vec{n}$ is the normal to the intensity gradient at image location $\vec{x}$ and Δ is a preset value (e.g. $\Delta =$ 3 pixels).

To obtain a unique characterization, we impose an order where $l_e^l(\vec{x})$ always points at the maximum hue value and $l_e^r(\vec{x})$ at the lesser hue value:

$$l_e(\vec{x}) = \begin{cases} l_e^r(\vec{x}) + 2\pi l_e^l(\vec{x}), & \text{if } l_e^r(\vec{x}) \leq l_e^l(\vec{x}) \\ l_e^l(\vec{x}) + 2\pi l_e^r(\vec{x}), & \text{otherwise} \end{cases} \quad (7)$$

The hue-hue edge pair $l_e(\vec{x})$ is quantitative, non-geometric and viewpoint independent and can be derived from any view of a 3D multicolored object.

Color Invariant Corner Pairs:

A measure of cornerness is defined as the change of gradient direction along an edge contour, which for hue H results in:

$$\kappa(H(\vec{x})) = \frac{-H_y^2 H_{xx} + 2H_x H_y H_{xy} - H_x^2 H_{yy}}{(H_x^2 + H_y^2)^{3/2}} \quad (8)$$

where the partial derivatives at image location $\vec{x}$ take into account the circular nature of H.

Then, two opposite neighboring points are computed based on the direction of the gradient to determine the hue value on either side of the corner point yielding the hue-hue corner pair at $\vec{x}$: $l_c^l(\vec{x}) = H(\vec{x} - \Delta\vec{n})$ and $l_c^r(\vec{x}) = H(\vec{x} + \Delta\vec{n})$, only computed for image location $\vec{x}$ at the two sides of a high-curvature maximum.

Finally, to obtain a unique characterization we define:

$$l_c(\vec{x}) = \begin{cases} l_c^r(\vec{x}) + 2\pi l_c^l(\vec{x}), & \text{if } l_c^r(\vec{x}) \leq l_c^l(\vec{x}) \\ l_c^l(\vec{x}) + 2\pi l_c^r(\vec{x}), & \text{otherwise} \end{cases} \quad (9)$$

2.3 Reflection with Colored Illumination

Consider the body reflection term of the dichromatic reflection model defined by eq. (1):

$$C_c = m_b(\vec{n}, \vec{s}) \int_\lambda f_C(\lambda) e(\lambda) c_b(\lambda) d\lambda \quad (10)$$

for $C = \{R, G, B\}$ where $C_c = \{R_c, G_c, B_c\}$ gives the red, green and blue sensor response of a *matte* infinitesimal surface patch under unknown spectral power distribution of the illumination.

Suppose that the sensor sensitivities of the color camera are narrow-band with spectral response be approximated by delta functions $f_C(\lambda) = \delta(\lambda - \lambda_C)$, then we derive:

$$C_c = m_b(\vec{n}, \vec{s}) e(\lambda_C) c_b(\lambda_C) \quad (11)$$

2.3.1 Color Constant Color Model for Matte Surfaces

In this section, we propose a set of new color constant color models not only independent of the illumination color but also independent of the object's geometry.

The set of color constant color models is defined by:

$$M^p = \frac{(C_1^{\vec{x}_1} C_2^{\vec{x}_2})^p}{(C_1^{\vec{x}_2} C_2^{\vec{x}_1})^p}, C_1 \neq C_2 \in \{R, G, B\}, p \geq 1 \quad (12)$$

expressing the color ratio between two neighboring image locations, where $\vec{x}_1$ and $\vec{x}_2$ denote the image locations of the two neighboring pixels.

Lemma 2 *Assuming body reflection, M^p is independent of the viewpoint, surface orientation, illumination direction, illumination intensity, and illumination color.*

Proof: If we assume that the SPD of the illumination is locally constant (at least over the two neighboring locations from which ratio is computed i.e. $e^{\vec{y}_1}(\lambda) = e^{\vec{y}_2}(\lambda)$), then cf. (11) in eq. (12) we have:

$$\frac{(m_b^{\vec{y}_1}(\vec{n}, \vec{s}) c_{bC_1}^{\vec{y}_1} \int_\lambda f_{C_1}(\lambda) e^{\vec{y}_1}(\lambda) d\lambda)^p}{(m_b^{\vec{y}_2}(\vec{n}, \vec{s}) c_{bC_1}^{\vec{y}_2} \int_\lambda f_{C_1}(\lambda) e^{\vec{y}_2}(\lambda) d\lambda)^p} \cdot$$

$$\frac{(m_b^{\vec{y}_2}(\vec{n}, \vec{s}) c_{bC_2}^{\vec{y}_2} \int_\lambda f_{C_2}(\lambda) e^{\vec{y}_2}(\lambda) d\lambda)^p}{(m_b^{\vec{y}_1}(\vec{n}, \vec{s}) c_{bC_2}^{\vec{y}_1} \int_\lambda f_{C_2}(\lambda) e^{\vec{y}_1}(\lambda) d\lambda)^p} = \frac{(c_{bC_1}^{\vec{y}_1})^p (c_{bC_2}^{\vec{y}_2})^p}{(c_{bC_1}^{\vec{y}_2})^p (c_{bC_2}^{\vec{y}_1})^p} \quad (13)$$

only dependent on the surface albedo, where $\vec{y}_1$ and $\vec{y}_2$ are two neighboring locations on the object's surface not necessarily of the same orientation. *QED.*

In theory, when $\vec{y}_1$ and $\vec{y}_2$ are neighboring locations on the same uniformly painted surface, the color ratio M^p will be 1. Except along color edges, assuming that the neighboring locations are at either side of the color edge, the value of the color ratio will deviate from 1.

2.3.2 Color Constant Image Features

In this paper, the set of color ratio's is considered for $p = 1$. Then, having three color components of two locations, color ratios obtained from a RGB-color image are: $m_1(R^{\vec{x}_1}, R^{\vec{x}_2}, G^{\vec{x}_1}, G^{\vec{x}_2}) = \frac{R^{\vec{x}_1} G^{\vec{x}_2}}{R^{\vec{x}_2} G^{\vec{x}_1}}$, $m_2(R^{\vec{x}_1}, R^{\vec{x}_2}, B^{\vec{x}_1}, B^{\vec{x}_2}) = \frac{R^{\vec{x}_1} B^{\vec{x}_2}}{R^{\vec{x}_2} B^{\vec{x}_1}}$, $m_3(G^{\vec{x}_1}, G^{\vec{x}_2}, B^{\vec{x}_1}, B^{\vec{x}_2}) = \frac{G^{\vec{x}_1} B^{\vec{x}_2}}{G^{\vec{x}_2} B^{\vec{x}_1}}$, where $m_1, m_2, m_3 \in M^p$ with $p = 1$.

For the ease of exposition, we concentrate on m_1 based on RG in the following discussion. Without loss of generality, all results derived for m_1 will also hold for m_2 and m_3.

Taking logarithms of both sides of eq. (12) results for m_1 in:

$$\ln m_1(R^{\vec{x}_1}, R^{\vec{x}_2}, G^{\vec{x}_1}, G^{\vec{x}_2}) = \ln R^{\vec{x}_1} + \ln G^{\vec{x}_2} - \ln R^{\vec{x}_2} - \ln G^{\vec{x}_1} \tag{14}$$

When these differences are taken between neighboring pixels in a particular direction, they correspond to finite-difference differentiation. To find color ratio edges in images we use the edge detection proposed in [3] which is currently of the Sobel type.

The results obtained so far for m_1 hold also for m_2 and m_3, yielding a 3-tuple $(\mathcal{G}_{m_1}(\vec{x}), \mathcal{G}_{m_2}(\vec{x}), \mathcal{G}_{m_3}(\vec{x}))$ denoting the gradient magnitude for every neighborhood centered at $\vec{x}$ in the image.

3 Shape Invariants

3.1 Similarity Invariant

For image locations $\vec{x_1}$, $\vec{x_2}$ and $\vec{x_3}$, $g_E()$ is defined as the well-known similarity invariant:

$$g_E(\vec{x_1}, \vec{x_2}, \vec{x_3}) = \theta \tag{15}$$

where θ is the angle at image coordinate $\vec{x_1}$ between line $\vec{x_1}\vec{x_2}$ and $\vec{x_1}\vec{x_3}$.

3.2 Projective Invariant

From the classical projective geometry we know that the so called cross-ratio is independent of the projection viewpoint:

$$g_P(\vec{x_1}, \vec{x_2}, \vec{x_3}, \vec{x_4}, \vec{x_5})) = \frac{\sin(\theta_1 + \theta_2)\sin(\theta_2 + \theta_3)}{\sin(\theta_2)\sin(\theta_1 + \theta_2 + \theta_3)} \tag{16}$$

where $\theta_1, \theta_2, \theta_3$ are the angles at image coordinate $\vec{x_1}$ between $\vec{x_1}\vec{x_2}$ and $\vec{x_1}\vec{x_3}$, $\vec{x_1}\vec{x_3}$ and $\vec{x_1}\vec{x_4}$, $\vec{x_1}\vec{x_4}$ and $\vec{x_1}\vec{x_5}$ respectively.

4 Object Recognition: Histogram Formation and Matching

Histograms are created on the basis of the different features defined in Section 2 and 3 for each reference image in the image database by counting the number of times a discrete color feature occurs in the image. The histogram from the test image is created in a similar way. Then, the object recognition process is reduced to the problem to what extent histogram $\mathcal{H}^{\mathcal{Q}}$ derived from the test image $\mathcal{Q}$ is similar to a histogram $\mathcal{H}^{I_k}$ constructed for each reference image I_k in the image database. For comparison reasons in the literature, similarity between histograms is expressed by histogram intersection [7].

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

4.1 Color Invariant Histogram Formation

Histograms are constructed on the basis of different color features representing the distribution of discrete color feature values in a n-dimensional color feature space, where $n = 3$ for $l_1 l_2 l_3$ and $m_1 m_2 m_3$, and $n = 1$ for H. For comparison reasons in the literature, we have also constructed color feature spaces for RGB and the following standard, color features derived from RGB: intensity $I(R, G, B) = R + B + G$, normalized colors (invariant for matte objects [3]): $r(R, G, B) = \frac{R}{R+G+B}$, $g(R, G, B) = \frac{G}{R+G+B}$, $b(R, G, B) = \frac{B}{R+G+B}$ and saturation (invariant for matte objects [3]): $S(R, G, B) = 1 - \frac{\min(R,G,B)}{R+G+B}$.

We have determined the appropriate bin size for our application empirically. The histogram bin size used during histogram formation is $q = 32$.

Hence, for each test and reference image, 3-dimensional histograms are created for the RGB, $l_1 l_2 l_3$, rgb and $m_1 m_2 m_3$ color space denoted by $\mathcal{H}_{RGB}$, $\mathcal{H}_{l_1 l_2 l_3}$, $\mathcal{H}_{rgb}$ and $\mathcal{H}_{m_1 m_2 m_3}$ respectively. Furthermore, 1-dimensional histograms are created for I, S and H denoted by $\mathcal{H}_I$, $\mathcal{H}_S$, and $\mathcal{H}_H$.

4.1.1 Hue-based Histogram Formation

The histogram representing the distribution of hue-hue edge pairs is given by:

$$\mathcal{H}_B^{I_k}(i) \hat{=} \eta(l_e(\vec{x}) == i) \tag{17}$$

only computed for the set of hue edge maxima computed from I_k. η indicates the number of times $l_e(\vec{x})$ equals the value of the histogram index denoted by i with $l_e()$ given by eq. (7).

The histogram of hue-hue corners is given by:

$$\mathcal{H}_C^{I_k}(i) \hat{=} \eta(l_c(\vec{x}) == i) \tag{18}$$

only computed for the set of hue corners computed from I_k and $l_c()$ is given by eq. (9).

4.2 Shape Invariant Histogram Formation

A 1-dimensional histogram is constructed in a standard way on the angle axis expressing the distribution of angles between hue corner triplets mathematically specified by:

$$\mathcal{H}_D^{I_k}(i) \hat{=} \eta(g_E(\vec{x}_1, \vec{x}_2, \vec{x}_3) == i) \tag{19}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \in C^{I_k}$, where C^{I_k} is the set of hue corners computed from I_k and $g_E()$ is given by eq. (15).

In a similar way, a 1-dimensional histogram is defined on the cross ratio axis expressing the distribution of cross ratios between hue corner quintets:

$$\mathcal{H}_E^{I_k}(i) \hat{=} \eta(g_P(\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5) == i) \tag{20}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \neq \vec{x}_4 \neq \vec{x}_5 \in C^{I_k}$ and $g_P()$ is defined by eq. (16).

4.3 Composite Color and Shape Invariant Histogram Formation

A 4-dimensional histogram is created counting the number of corner triples with hue-hue values i, j and k generating angle l (similarity invariant):

$$\mathcal{H}_F^{I_k}(i,j,k,l) \hat{=} \eta(l_c(\vec{x}_1) == i \wedge l_c(\vec{x}_2) == j \wedge$$
$$l_c(\vec{x}_3) == k \wedge g_E(\vec{x}_1, \vec{x}_2, \vec{x}_3) == l) \tag{21}$$

only computed for $\vec{x}_1 \neq \vec{x}_2 \neq \vec{x}_3 \in C^{I_k}$, where $g_E()$ is given by eq. (15) and $\wedge$ is the the logical AND. Each histogram bin measures the number of hue-hue corner triplets generating a certain angle.

In a similar way, a 6-dimensional invariant histogram can be constructed considering the cross-ratio between hue-hue corners:

$$\mathcal{H}_G^{I_k}(i,j,k,l,m,n) \hat{=} \eta(l_c(\vec{x}_1) == i \wedge l_c(\vec{x}_2) == j \wedge l_c(\vec{x}_3) == k$$
$$l_c(\vec{x}_4) == l \wedge l_c(\vec{x}_5) == m \wedge g_P(\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5) == n) \tag{22}$$

5 Experiments

In this section, the different invariant image features are evaluated on a database of 500 images. The data sets on which the experiments will be conducted are described in Section 5.1. Error measures are given in 5.2.

5.1 Datasets

The database consists of $N_1 = 500$ reference color images of domestic objects, tools, toys, food cans, art artifacts etc., all taken from two households. Objects were recorded in isolation (one per image) with the aid of the SONY XC-003P CCD color camera (3 chips) and the Matrox Magic Color frame grabber. 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.

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 the experiments, 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.

The image database and the performance of the recognition scheme can be experienced within the ZOMAX system on-line at:
http://www.wins.uva.nl/research/isis/zomax/.

5.2 Error Measures

For a measure of recognition quality, let rank r^{Q_i} denote the position of the correct match for test image Q_i, $i = 1, ..., N_2$, in the ordered list of N_1 match values. The 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: $\overline{r} = (\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r^{Q_i}}{N_1 - 1})100\%$. The cumulative percentile of test images producing a rank smaller or equal to j is defined as: $\mathcal{X}(j) = (\frac{1}{N_2} \sum_{k=1}^{j} \eta(r^{Q_i} == k))100\%$ where η reads as the number of test images having rank k.

5.3 Results with White Illumination

In this subsection, we report on the recognition accuracy of the matching process for $N_2 = 70$ test images and $N_1 = 500$ reference images for the various invariant image features. As stated, white lighting is used during the recording of the reference images in the image database and the independent test set. However, the objects were recorded with a new, arbitrary position and orientation with respect to camera. In Fig. 1 accumulated ranking percentile is shown for the various invariant features.

From the results of Fig. 1 we can observe that the discriminative power of $l_1 l_2 l_3$, H and hue-hue edges

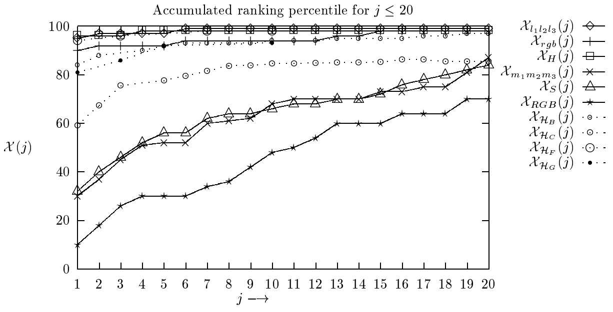

Figure 1: *The discriminative power: the cumulative percentile $\mathcal{X}$ for H, $l_1l_2l_3$, rgb, S, $m_1m_2m_3$, RGB, $\mathcal{H}_B$, $\mathcal{H}_C$, $\mathcal{H}_F$ and $\mathcal{H}_G$ is given by $\mathcal{X}_H$, $\mathcal{X}_{l_1l_2l_3}$, $\mathcal{X}_{c_1c_2c_3}$, $\mathcal{X}_{rgb}$, $\mathcal{X}_S$, $\mathcal{X}_{m_1m_2m_3}$, $\mathcal{X}_{RGB}$, $\mathcal{X}_{\mathcal{H}_B}$, $\mathcal{X}_{\mathcal{H}_C}$, $\mathcal{X}_{\mathcal{H}_F}$ and $\mathcal{X}_{\mathcal{H}_G}$ respectively.*

$\mathcal{X}_{\mathcal{H}_B}$ followed by rgb is higher then the other invariants. Furthermore, very high discriminative accuracy is shown for composite color and shape features $\mathcal{H}_F$ and $\mathcal{H}_G$ as 96% of the images are within the first 2 rankings, and 98% within the first 9 rankings.

Hue-hue corners $\mathcal{X}_{\mathcal{H}_C}$, saturation S and color ratio $m_1m_2m_3$ provides slightly worse recognition accuracy. As expected, the discrimination power of RGB has the worst performance due to its sensitivity to varying imaging conditions. Shape-based object recognition (not within the first 20 rankings and hence not shown here) yields very poor discriminative power.

5.4 Results with Colored Illumination

Based on the coefficient rule or von Kries model, in this paper, the change in the illumination color is approximated by a 3x3 diagonal matrix among the sensor bands and is equal to the multiplication of each RGB-color band by an independent scalar factor [1]. Note that the diagonal model of illumination change holds in the case of narrow-band sensors. To measure the sensitivity of the various image invariant feature in practice with respect to a change in the color of the illumination, the R, G and B-color bands of each image of the test set are multiplied by a factor $\beta_1 = \beta$, $\beta_2 = 1$ and $\beta_3 = 2 - \beta$ respectively (i.e. $\beta_1 R$, $\beta_2 G$ and $\beta_3 B$) by varying β over $\{0.5, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.5\}$. The discrimination power of the histogram matching process differentiated for the various invariant features plotted against the illumination color is shown in Fig. 2. For $\beta < 1$ the color is reddish whereas bluish for $\beta > 1$.

As expected, only the color ratio $m_1m_2m_3$ is insensitive to a change in illumination color. From Fig. 2 we can observe that invariant features H, hue-based

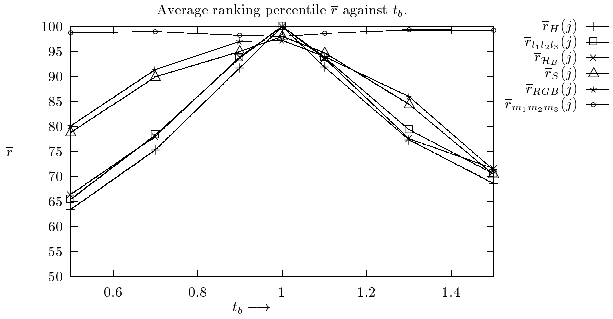

Figure 2: *The discriminative power plotted against the change β in the color composition of the illumination spectrum.*

composite invariant features, and $l_1l_2l_3$ which achieved best recognition accuracy under white illumination, see Fig. 1, are highly sensitive to a change in illumination color followed by S and RGB. Even for a slight change in the illumination color, their recognition potential degrades drastically.

6 Discussion and Conclusion

On the basis of the above reported theory and experiments, it is concluded that the proposed invariant $l_1l_2l_3$ and (hue-based) composite shape and color invariant features, followed by H and hue-hue edges are most appropriate to be used for invariant object recognition under the constraint of a white illumination source. When no constraints are imposed on the imaging conditions (i.e. the most general case), the newly proposed color ratio $m_1m_2m_3$ is most appropriate.

References

[1] Finlayson, G. D., Chatterjee S. S., and Funt B. V., *Color Angular Indexing*, ECCV96, II, pp. 16-27, 1996.

[2] Funt, B. V. and Finlayson, G. D., *Color Constant Color Indexing*, IEEE PAMI, 17(5), pp. 522-529, 1995.

[3] Gevers, T., *Color Image Invariant Segmentation and Retrieval*, PhD Thesis, ISBN 90-74795-51-X, University of Amsterdam, The Netherlands, 1996.

[4] Healey, G. and Slater D, *Global Color Constancy: Recognition of Objects by Use of Illumination Invariant Properties of Color Distributions*, J. Opt. Soc. Am. A, Vol. 11, No. 11, pp. 3003-3010, Nov 1995.

[5] *Geometric Invariance in Computer Vision*, Mundy, J. and Zisserman, A. (eds.), MIT Press, Cambridge, Massachusetts, 1992.

[6] Shafer, S. A., *Using Color to Separate Reflection Components*, COLOR Res. Appl., 10(4), pp 210-218, 1985.

[7] Swain, M. J. and Ballard, D. H., *Color Indexing*, International Journal of Computer Vision, Vol. 7, No. 1, pp. 11-32, 1991.