Color-based object recognition

Theo Gevers*, Arnold W.M. Smeulders

ISIS, Faculty of WINS, University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands

Abstract

The purpose is to arrive at recognition of multicolored objects invariant to a substantial change in viewpoint, object geometry and illumination. Assuming dichromatic reflectance and white illumination, it is shown that normalized color rgb, saturation S and hue H, and the newly proposed color models $c_1 c_2 c_3$ and $l_1 l_2 l_3$ are all invariant to a change in viewing direction, object geometry and illumination. Further, it is shown that hue H and $l_1 l_2 l_3$ are also invariant to highlights. Finally, a change in spectral power distribution of the illumination is considered to propose a new color constant color model $m_1 m_2 m_3$. To evaluate the recognition accuracy differentiated for the various color models, experiments have been carried out on a database consisting of 500 images taken from 3-D multicolored man-made objects. The experimental results show that highest object recognition accuracy is achieved by $l_1 l_2 l_3$ and hue H followed by $c_1 c_2 c_3$, normalized color rgb and $m_1 m_2 m_3$ under the constraint of white illumination. Also, it is demonstrated that recognition accuracy degrades substantially for all color features other than $m_1 m_2 m_3$ with a change in illumination color. The recognition scheme and images are available within the PicToSeek and Pic2Seek systems on-line at: http://www.wins.uva.nl/research/isis/zomax/. © 1999 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Object recognition; Multicolored objects; Color models; Dichromatic reflection; Reflectance properties; Photometric color invariants; Color constancy

1. Introduction

Color provides powerful information for object recognition. A simple and effective recognition scheme is to represent and match images on the basis of color histograms as proposed by Swain and Ballard [1]. The work makes a significant contribution in introducing color for object recognition. However, it has the drawback that when the illumination circumstances are not equal, the object recognition accuracy degrades significantly. This method is extended by Funt and Finlayson [2], based on the retinex theory of Land [3], to make the method illumination independent by indexing on illumination-invariant surface descriptors (color ratios) computed from neighboring points. However, it is assumed that neighboring points have the same surface normal. Therefore, the derived illumination-invariant surface descriptors are negatively affected by rapid changes in surface orientation of the object (i.e. the geometry of the object). Healey and Slater [4] and Finlayson et al. [5] use illumination-invariant moments of color distributions for object recognition. These methods are sensitive to object occlusion and cluttering as the moments are defined as an integral property on the object as one. In global methods, in general, occluded parts will disturb recognition. Slater and Healey [6] circumvent this problem by computing the color features from small object regions instead of the entire object.

From the above observations, the choice which color models to use does not only depend on their robustness against varying illumination across the scene (e.g. multiple light sources with different spectral power distributions), but also on their robustness against changes in surface orientation of the object (i.e. the geometry of the object), and on their robustness against object occlusion and cluttering. Furthermore, the color models should be concise, discriminatory and robust to noise. Therefore, in this paper, our aim is to analyze and evaluate various color models to be used for the purpose of recognition of multicolored objects according to the following criteria:

- Robustness to a change in viewing direction;
- robustness to a change in object geometry;
- robustness to a change in the direction of the illumination;
- robustness to a change in the intensity of the illumination;
- robustness to a change in the spectral power distribution (SPD) of the illumination.

Next to defining color models which have:

- High discriminative power;
- robustness to object occlusion and cluttering;
- robustness to noise in the images.

It can be expected that two or more of the above criteria are interrelated. For example, Funt and Finlayson [2] show that when illumination is controlled Swain's color-based recognition method performs better than object recognition based on illumination-independent image descriptors. However, Swain's method is outperformed when illumination varies across the scene. Supposedly, their is a tradeoff between the amount of invariance and expressiveness of the color models. To that end, our goal is to get more insight to decide which color models to use under which imaging parameters. This is useful for object recognition applications where no constraints on the imaging process can be imposed as well as for applications where one or more parameters of the imaging process can be controlled such as for robots and industrial inspection (e.g. controlled object positioning and lightning conditions). For such a case, color models can be used for object recognition which are less invariant (at least under the given imaging conditions), but having higher discriminative power.

The paper is organized as follows. In Section 2, basic color models are defined for completeness. In Section 3, assuming white illumination and dichromatic reflectance, we examine the effect of a change in viewpoint, surface orientation, and illumination for the various color models. From the analysis, two new invariant color models are proposed. Further, in Section 4, a change in spectral power distribution (SPD) of the illumination is considered to propose a new color constant color model.

A summary of the theoretical results is given in Section 5. In Section 6, experiments are carried out on an image database of 500 images taken from 3-D multicolored man-made objects. In Section 7, we conclude with a guideline which color models to use under which imaging conditions for both invariant and discriminatory object recognition.

2. Basic color definitions

Commonly used well-known color spaces include: (for display and printing processes) RGB, CMY; (for television and video) YIQ, YUV; (standard set of primary colors) XYZ; (uncorrelated features) $I_1 I_2 I_3$; (normalized color) rgb, xyz; (perceptual uniform spaces) $U^*V^*W^*$, $L^*a^*b^*$, Luv; and (for humans) HSI. Although, the number of existing color spaces is large, a number of these color models are correlated to intensity I: Y, L^* and W^*; are linear combinations of RGB: CMY, XYZ and $I_1 I_2 I_3$; or normalized with respect to intensity rgb: IQ, xyz, UV, U^*V^*, a^*b^*, uv. Therefore, in this paper, we concentrate on the following standard, *essentially different*, color features: intensity I, RGB, normalized color rgb, hue H and saturation S.

In the sequel, we need to be precise on the definitions of intensity I, RGB, normalized color rgb, saturation S, and hue H. To that end, in this section, we offer a quick overview of well-known facts from color theory.

Let R, G and B, obtained by a color camera, represent the 3-D sensor space

$$C = \int_\lambda p(\lambda) f_C(\lambda) \, \mathrm{d}\lambda \tag{1}$$

for $C \in (R, G, B)$, where $p(\lambda)$ is the radiance spectrum and $f_C(\lambda)$ are the three color filter transmission functions.

To represent the RGB-sensor space, a cube can be defined on the R, G, and B axes. White is produced when all three primary colors are at M, where M is the maximum light intensity, say $M = 255$. The main diagonal-axis connecting the black and white corners defines the intensity

$$I(R, G, B) = R + G + B. \tag{2}$$

All points in a plane perpendicular to the grey axis of the color cube have the same intensity. The plane through the color cube at points $R = G = B = M$ is one such plane. This plane cuts out an equilateral triangle which is the standard rgb chromaticity triangle

$$r(R, G, B) = \frac{R}{R + G + B}, \tag{3}$$

$$g(R, G, B) = \frac{G}{R + G + B}, \tag{4}$$

$$b(R, G, B) = \frac{B}{R + G + B}. \qquad (5)$$

The transformation from RGB used here to describe the color impression hue H is given by

$$H(R, G, B) = \arctan\left(\frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}\right) \qquad (6)$$

and saturation S measuring the relative white content of a color as having a particular hue by

$$S(R, G, B) = 1 - \frac{\min(R, G, B)}{R + G + B}. \qquad (7)$$

In this way, all color features can be calculated from the original R, G, B values from the corresponding red, green, and blue images provided by the color camera.

3. Reflectance with white illumination

3.1. The reflection model

Consider an image of an infinitesimal surface patch of an inhomogeneous dielectric object. 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 is given by Shafer [7]

$$C = m_b(\mathbf{n}, \mathbf{s}) \int_\lambda f_C(\lambda) e(\lambda) c_b(\lambda) \, d\lambda$$

$$+ m_s(\mathbf{n}, \mathbf{s}, \mathbf{v}) \int_\lambda f_C(\lambda) e(\lambda) c_s(\lambda) \, d\lambda \qquad (8)$$

for $C = \{R, G, B\}$ giving the Cth sensor response. Further, $c_b(\lambda)$ and $c_s(\lambda)$ are the surface albedo and Fresnel reflectance respectively. λ denotes the wavelength, $\mathbf{n}$ is the surface patch normal, $\mathbf{s}$ is the direction of the illumination source, and $\mathbf{v}$ is the direction of the viewer. Geometric terms m_b and m_s denote the geometric dependencies on the body and surface reflection component, respectively.

Considering the neutral interface reflection (NIR) model (assuming that $c_s(\lambda)$ has a constant value independent of the wavelength) and white illumination (equal energy density for all wavelengths within the visible spectrum), then $e(\lambda) = e$ and $c_s(\lambda) = c_s$, and hence being constants. Then, we put forward that the measured sensor values are given by

$$C_w = em_b(\mathbf{n}, \mathbf{s})k_C + em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s \int_\lambda f_C(\lambda) \, d\lambda \qquad (9)$$

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. Further,

$$k_C = \int_\lambda f_C(\lambda) c_b(\lambda) \, d\lambda \qquad (10)$$

is the compact formulation depending on the sensors and the surface albedo only.

If the integrated white condition holds (as we assume throughout the paper)

$$\int_\lambda f_R(\lambda) \, d\lambda = \int_\lambda f_G(\lambda) \, d\lambda = \int_\lambda f_B(\lambda) \, d\lambda = f \qquad (11)$$

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

$$C_w = em_b(\mathbf{n}, \mathbf{s})k_C + em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s f. \qquad (12)$$

In the next section, this reflection model is used to study and analyze the RGB- subspace on which colors will be projected coming from the same uniformly colored surface.

3.2. Photometric color invariant features for Matte, Dull surfaces

Consider the body reflection term of Eq. (12)

$$C_b = em_b(\mathbf{n}, \mathbf{s})k_C \qquad (13)$$

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

According to the body reflection term, the color depends on k_C (i.e. sensors and surface albedo) and the brightness on illumination intensity e and object geometry $m_b(\mathbf{n}, \mathbf{s})$. If a matte surface region, which is homogeneously colored (i.e. with fixed albedo), contains a variety of surface normals, then the set of measured colors will generate an elongated color cluster in RGB-sensor space, where the direction of the streak is determined by k_C and its extent by the variations of surface normals $\mathbf{n}$ with respect to the illumination direction $\mathbf{s}$. As a consequence, a uniformly colored surface which is curved (i.e. varying surface orientation) gives rise to a broad variance of RGB values. The same argument holds for intensity I.

In contrast, rgb is insensitive to surface orientation, illumination direction and illumination intensity mathematically specified by substituting Eq. (13) in Eqs. (3)–(5)

$$r(R_b, G_b, B_b) = \frac{em_b(\mathbf{n}, \mathbf{s})k_R}{em_b(\mathbf{n}, \mathbf{s})(k_R + k_G + k_B)}$$

$$= \frac{k_R}{k_R + k_G + k_B}, \qquad (14)$$

$$g(R_b, G_b, B_b) = \frac{em_b(\mathbf{n}, \mathbf{s})k_G}{em_b(\mathbf{n}, \mathbf{s})(k_R + k_G + k_B)}$$

$$= \frac{k_G}{k_R + k_G + k_B}, \tag{15}$$

$$b(R_b, G_b, B_b) = \frac{em_b(\mathbf{n}, \mathbf{s})k_B}{em_b(\mathbf{n}, \mathbf{s})(k_R + k_G + k_B)}$$

$$= \frac{k_B}{k_R + k_G + k_B}, \tag{16}$$

factoring out dependencies on illumination e and object geometry $m_b(\mathbf{n}, \mathbf{s})$ and hence only dependent on the sensors and the surface albedo.

Because S corresponds to the radial distance from the color to the main diagonal in the RGB-color space, S is an invariant for matte, dull surfaces illuminated by white light cf. Eqs. (13) and (7)

$$S(R_b, G_b, B_b)$$

$$= 1 - \frac{\min(em_b(\mathbf{n}, \mathbf{s})k_R, em_b(\mathbf{n}, \mathbf{s})k_G, em_b(\mathbf{n}, \mathbf{s})k_B)}{em_b(\mathbf{n}, \mathbf{s})(k_R + k_G + k_B)}$$

$$= 1 - \frac{\min(k_R, k_G, k_B)}{(k_R + k_G + k_B)}, \tag{17}$$

only dependent on the sensors and the surface albedo.

Similarly, H is an invariant for matte, dull surfaces illuminated by white light cf. Eqs. (13) and (6)

$$H(R_b, G_b, B_b)$$

$$= \arctan\left(\frac{\sqrt{3}em_b(\mathbf{n}, \mathbf{s})(k_G - k_B)}{em_b(\mathbf{n}, \mathbf{s})((k_R - k_G) + (k_R - k_B))}\right)$$

$$= \arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right). \tag{18}$$

In fact, any expression defining colors on the same linear color cluster spanned by the body reflection vector in RGB-space is an invariant for the dichromatic reflection model with white illumination. To that end, we put forward the following invariant color model

$$c_1 = \arctan\left(\frac{R}{\max\{G, B\}}\right), \tag{19}$$

$$c_2 = \arctan\left(\frac{G}{\max\{R, B\}}\right), \tag{20}$$

$$c_3 = \arctan\left(\frac{B}{\max\{R, G\}}\right), \tag{21}$$

denoting the angles of the body reflection vector and consequently being invariants for matte, dull objects cf. Eqs. (13) and (19)–(21)

$$c_1(R_b, G_b, B_b) = \arctan\left(\frac{em_b(\mathbf{n}, \mathbf{s})k_R}{\max\{em_b(\mathbf{n}, \mathbf{s})k_G, em_b(\mathbf{n}, \mathbf{s})k_B\}}\right)$$

$$= \arctan\left(\frac{k_R}{\max\{k_G, k_B\}}\right), \tag{22}$$

$$c_2(R_b, G_b, B_b) = \arctan\left(\frac{em_b(\mathbf{n}, \mathbf{s})k_G}{\max\{em_b(\mathbf{n}, \mathbf{s})k_R, em_b(\mathbf{n}, \mathbf{s})k_B\}}\right)$$

$$= \arctan\left(\frac{k_G}{\max\{k_R, k_B\}}\right), \tag{23}$$

$$c_3(R_b, G_b, B_b) = \arctan\left(\frac{em_b(\mathbf{n}, \mathbf{s})k_B}{\max\{em_b(\mathbf{n}, \mathbf{s})k_R, em_b(\mathbf{n}, \mathbf{s})k_G\}}\right)$$

$$= \arctan\left(\frac{k_B}{\max\{k_R, k_G\}}\right), \tag{24}$$

only dependent on the sensors and the surface albedo.

Obviously, in practice, the assumption of objects composed of matte, dull surfaces is not always realistic. To that end, the effect of surface reflection (highlights) is discussed in the following section.

3.3. Photometric color invariant features for both Matte and Shiny surfaces

Consider the surface reflection term of Eq. (12)

$$C_s = em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s f \tag{25}$$

for $C_s \in \{R_s, G_s, B_s\}$ giving the red, green and blue sensor response for a highlighted infinitesimal surface patch with white illumination.

Note that under the given conditions, the color of highlights is not related to the color of the surface on which they appear, but only on the color of the light source. Thus, for the white light source, the set of measured colors from a highlighted surface region is on the grey axis of the RGB-color space. The extent of the streak depends on the roughness of the object surface. Very shiny object regions generate color clusters which are spread out over the entire grey axis. For rough surfaces, the extent will be small.

For a given point on a surface, the contribution of the body reflection component C_b and surface reflection component C_s are added cf. Eq. (12). Hence, the measured colors of a uniformly colored region must be on the triangular color plane in the RGB-space spanned by the two reflection components.

Because H is a function of the angle between the main diagonal and the color point in RGB-sensor space, all possible colors of the same (shiny) surface region (i.e. with

fixed albedo) have to be of the same hue as follows from substituting Eq. (12) into Eq. (6)

$$H(R_w, G_w, G_w) = \arctan\left(\frac{\sqrt{3}(G_w - B_w)}{(R_w - G_w) + (R_w - B_w)}\right)$$

$$= \arctan\left(\frac{\sqrt{3}\,em_b(\mathbf{n}, \mathbf{s})(k_G - k_B)}{em_b(\mathbf{n}, \mathbf{s})((k_R - k_G) + (k_R - k_B))}\right)$$

$$= \arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right), \quad (26)$$

factoring out dependencies on illumination e, object geometry $m_b(\mathbf{n}, \mathbf{s})$, viewpoint $m_s(\mathbf{n}, \mathbf{s}, \mathbf{v})$, and specular reflection coefficient c_s and hence only dependent on the sensors and the surface albedo. Note that $R_w = em_b(\mathbf{n}, \mathbf{s})k_R + em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s f$, $G_w = em_b(\mathbf{n}, \mathbf{s})k_G + em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s f$, and $B_w = em_b(\mathbf{n}, \mathbf{s})k_B + em_s(\mathbf{n}, \mathbf{s}, \mathbf{v})c_s f$.

Obviously, other color features depend on the contribution of the surface reflection component and hence are sensitive to highlights.

In fact, any expression defining colors on the same linear triangular color plane, spanned by the two reflection components in RGB-color space, are invariants for the dichromatic reflection model with white illumination.

To that end, a new color model $l_1 l_2 l_3$ is proposed uniquely determining the direction of the triangular color plane in RGB-space

$$l_1 = \frac{(R - G)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}, \quad (27)$$

$$l_2 = \frac{(R - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}, \quad (28)$$

$$l_3 = \frac{(G - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}, \quad (29)$$

the set of normalized color differences which is, similar to H, a photometric color invariant for matte as well as for shiny surfaces which follows from substituting Eq. (12) into Eqs. (27)–(29), which for l_1 results in

4. Reflectance with colored illumination

4.1. The reflection model

We consider the body reflection term of the dichromatic reflection model

$$C_c = m_b(\mathbf{n}, \mathbf{s}) \int_\lambda f_C(\lambda) e(\lambda) c_b(\lambda) \, d\lambda \quad (31)$$

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 of an inhomogeneous dielectric object under unknown spectral power distribution of the illumination.

Suppose that the sensor sensitivities of the color camera are narrow band with spectral responses approximated by delta functions $f_C(\lambda) = \delta(\lambda - \lambda_C)$, then the measured sensor values are

$$C_c = m_b(\mathbf{n}, \mathbf{s}) e(\lambda_C) c_b(\lambda_C). \quad (32)$$

By simply filling in C_c in the color model equations given in Section 2, it can be easily seen that all color model values change with a change in illumination color. To that end, a new color constant color model is proposed in the next section.

4.2. Color constant color feature for Matte, Dull surfaces

Existing color constancy methods require specific *a priori* information about the observed scene (e.g. the placement of calibration patches of known spectral reflectance in the scene) which will not be feasible in practical situations [8,9,3], for example. To circumvent these problems, Funt and Finlayson [2] propose simple and effective illumination-independent color ratios for the purpose of object recognition. However, it is assumed that the neighboring points, from which the color ratios are computed, have the same surface normal. Therefore, the method depends on varying surface orientation of the object (i.e. the geometry of the objects) affecting negatively the recognition performance. To this end, we propose

$$l_1(R_w, G_w, G_w) = \frac{(R_w - G_w)^2}{(R_w - G_w)^2 + (R_w - B_w)^2 + (G_w - B_w)^2}$$

$$= \frac{(em_b(\mathbf{n}, \mathbf{s})(k_R - k_G))^2}{(em_b(\mathbf{n}, \mathbf{s})(k_R - k_G))^2 + (em_b(\mathbf{n}, \mathbf{s})(k_R - k_B))^2 + (em_b(\mathbf{n}, \mathbf{s})(k_G - k_B))^2}$$

$$= \frac{(k_R - k_G)^2}{(k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2}, \quad (30)$$

only dependent on the sensors and the surface albedo. Equal arguments hold for l_2 and l_3.

a new color constant color ratio not only independent of the illumination color but also discounting the object's

geometry

$$m(C_1^{x_1}, C_1^{x_2}, C_2^{x_1}, C_2^{x_2}) = \frac{C_1^{x_1} C_2^{x_2}}{C_1^{x_2} C_2^{x_1}}, \quad C_1 \neq C_2, \tag{33}$$

expressing the color ratio between two neighboring image locations, for $C_1, C_2 \in \{R, G, B\}$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the image locations of the two neighboring pixels. Note that the set $\{R, G, B\}$ must be colors from narrow-band sensor filters and that they are used in defining the color ratio because they are immediately available from a color camera, but any other set of narrow-band colors derived from the visible spectrum will do as well.

If we assume that the color of the illumination is locally constant (at least over the two neighboring locations from which the ratio is computed, i.e. $e^{x_1}(\lambda) = e^{x_2}(\lambda)$) the color ratio is independent of the illumination intensity and color, and also to a change in viewpoint, object geometry, and illumination direction as shown by substituting Eq. (32) into Eq. (33)

$$m(c_1^{x_1}, c_1^{x_2}, c_2^{x_1}, c_2^{x_2})$$

$$= \frac{(m_b^{x_1}(\mathbf{n}, \mathbf{s})e^{x_1}(\lambda_{C_1})c_b^{x_1}(\lambda_{C_1}))(m_b^{x_2}(\mathbf{n}, \mathbf{s})e^{x_2}(\lambda_{C_2})c_b^{x_2}(\lambda_{C_2}))}{(m_b^{x_2}(\mathbf{n}, \mathbf{s})e^{x_2}(\lambda_{C_1})c_b^{x_2}(\lambda_{C_1}))(m_b^{x_1}(\mathbf{n}, \mathbf{s})e^{x_1}(\lambda_{C_2})c_b^{x_1}(\lambda_{C_2}))}$$

$$= \frac{c_b^{x_1}(\lambda_{C_1})c_b^{x_2}(\lambda_{C_2})}{c_b^{x_2}(\lambda_{C_1})c_b^{x_1}(\lambda_{C_2})}, \tag{34}$$

factoring out dependencies on object geometry and illumination direction $m_b^{x_1}(\mathbf{n}, \mathbf{s})$ and $m_b^{x_2}(\mathbf{n}, \mathbf{s})$, and illumination e^{x_1} and e^{x_2} as $e^{x_1}(\lambda_{C_1}) = e^{x_2}(\lambda_{C_1})$ and $e^{x_1}(\lambda_{C_2}) = e^{x_2}(\lambda_{C_2})$, and hence only dependent on the ratio of surface albedos, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two neighboring locations on the object's surface not necessarily of the same orientation.

Note that the color ratio does not require any specific *a priori* information about the observed scene, as the color model is an illumination-invariant surface descriptor based on the ratio of surface albedos rather than the recovering of the actual surface albedo itself. Also, the intensity and spectral power distribution of the illumination is allowed to vary across the scene (e.g. multiple light sources with different SPDs), and a certain amount of object occlusion and cluttering is tolerated due to the local computation of the color ratio. The color model is not restricted to Mondrian worlds where the scenes are flat, but any 3-D real-world scene is suited as the color model can cope with varying surface orientations of objects. Further note that the color ratio is insensitive to a change in surface orientation, illumination direction and intensity for matte objects under white light, but without the constraint of narrow-band filters, as follows from substituting Eq. (13) into Eq. (33):

$$\frac{(e^{x_1}m_b^{x_1}(\mathbf{n}, \mathbf{s})k_{C_1}^{x_1})(e^{x_2}m_b^{x_2}(\mathbf{n}, \mathbf{s})k_{C_2}^{x_2})}{(e^{x_2}m_b^{x_2}(\mathbf{n}, \mathbf{s})k_{C_1}^{x_2})(e^{x_1}m_b^{x_1}(\mathbf{n}, \mathbf{s})k_{C_2}^{x_1})} = \frac{k_{C_1}^{x_1}k_{C_2}^{x_2}}{k_{C_1}^{x_2}k_{C_2}^{x_1}}, \tag{35}$$

only dependent on the sensors and the surface albedo.

Having three color components of two locations, color ratios obtained from a *RGB*-color image are

$$m_1 = \frac{R^{x_1} G^{x_2}}{R^{x_2} G^{x_1}}, \tag{36}$$

$$m_2 = \frac{R^{x_1} B^{x_2}}{R^{x_2} B^{x_1}}, \tag{37}$$

$$m_3 = \frac{G^{x_1} B^{x_2}}{G^{x_2} B^{x_1}}. \tag{38}$$

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

Taking the natural logarithm of both sides of Eq. (33) results for m_1 in

$$\ln m_1(R^{x_1}, R^{x_2}, G^{x_1}, G^{x_2}) = \ln\left(\frac{R^{x_1} G^{x_2}}{R^{x_2} G^{x_1}}\right)$$

$$= \ln R^{x_1} + \ln G^{x_2} - \ln R^{x_2} - \ln G^{x_1}$$

$$= \ln\left(\frac{R^{x_1}}{G^{x_1}}\right) - \ln\left(\frac{R^{x_2}}{G^{x_2}}\right). \tag{39}$$

Hence, the color ratios can be seen as differences at two neighboring locations $\mathbf{x}_1$ and $\mathbf{x}_2$ in the image domain of the logarithm of R/G

$$d_{m_1}(\mathbf{x}_1, \mathbf{x}_2) = \left(\ln\left(\frac{R}{G}\right)\right)^{x_1} - \left(\ln\left(\frac{R}{G}\right)\right)^{x_2}. \tag{40}$$

By taking these differences in a particular direction between neighboring pixels, the finite-difference differentiation is obtained of the logarithm of image R/G which is independent of the illumination color, and also a change in viewpoint, the object geometry, and illumination intensity. We have taken the gradient magnitude by applying Canny's edge detector (derivative of the Gaussian with $\sigma = 1.0$) on image $\ln(R/G)$ with non-maximum suppression in a standard way to obtain gradient magnitudes at local edge maxima denoted by $\mathscr{G}_{m_1}(\mathbf{x})$, where the Gaussian smoothing suppresses the sensitivity of the color ratios to noise. The results obtained so far for m_1 hold also for m_2 and m_3, yielding a 3-tuple ($\mathscr{G}_{m_1}(\mathbf{x})$, $\mathscr{G}_{m_2}(\mathbf{x})$, $\mathscr{G}_{m_3}(\mathbf{x})$) denoting gradient magnitude at local edge maxima in images $\ln(R/G)$, $\ln(R/B)$ and $\ln(G/B)$, respectively. For pixels on a uniformly colored region (i.e. with fixed surface albedo), in theory, all three components will be zero whereas at least one the three components will be non-zero for pixels on locations where two regions of distinct surface albedo meet.

	viewing direction	surface orientation	highlights	illumination direction	illumination intensity	illumination color	inter reflection
I	-	-	-	-	-	-	-
RGB	-	-	-	-	-	-	-
rgb	+	+	-	+	+	-	-
S	+	+	-	+	+	-	-
c1c2c3	+	+	-	+	+	-	-
H	+	+	+	+	+	-	-
l1l2l3	+	+	+	+	+	-	-
m1m2m3	+	+	-	+	+	+	+

Fig. 1. Overview of the various color models and their invariance to various imaging conditions. $+$ denotes invariant and $-$ denotes sensitivity of the color model to the imaging condition.

Fig. 2. Left: 16 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.

5. Summary of the theoretical results

In conclusion, assuming dichromatic reflection and white illumination, normalized color rgb, saturation S and hue H, and the newly proposed color models $c_1c_2c_3$, $l_1l_2l_3$ and $m_1m_2m_3$ are all invariant to the viewing direction, object geometry and illumination. Further, hue H and $l_1l_2l_3$ are also invariant to highlights. $m_1m_2m_3$ is independent of the illumination color and inter-reflections (i.e. objects receiving reflected light from other objects) under the assumption of narrow-band filters. These results are summarized in Fig. 1.

To evaluate photometric color invariant object recognition, in practice, in the next section, the various color models are evaluated and compared on an image database of 500 images taken from 3-D multicolored man-made objects.

6. Color-based object recognition: experiments

In the experiments, we focus on object recognition by histogram matching for comparison reasons in the literature. Obviously, transforming RGB to one of the

invariant color models can be performed as a preprocessing step by other matching techniques.

This section is organized as follows. First, in Section 6.1, the experimental setup is given. The experimental results are given in Section 6.2.

6.1. Experimental setup

The following section is outlined as follows. First, the data sets on which the experiments will be conducted are described in Section 6.1.1. Error measures are given in Section 6.1.2. Histogram formation and similarity measure are given in Section 6.1.3.

6.1.1. Datasets

The database consists of $N_1 = 500$ reference images of multicolored 3-D domestic objects, tools, toys, etc. 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. 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.

In Fig. 2, 16 images from the image database of 500 images are shown on the left. Corresponding images coming from the query set are shown on the right.

More information about color-based object recognition can be found in [10]. The image database and the performance of the recognition scheme can be experienced within the PicToSeek and Pic2Seek systems on-line at http: //www.wins.uva.nl/research/isis/zomax/.

6.1.2. Error measures

For a measure of match quality, let rank r^{Q_i} denote the position of the correct match for test image Q_i, $i = 1, \dots, N_2$, 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

$$\bar{r} = \left(\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r^{Q_i}}{N_1 - 1} \right) 100\%. \qquad (41)$$

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

$$\mathcal{X}(j) = \left(\frac{1}{N_2} \sum_{k=1}^{j} \eta(r^{Q_i} = k) \right) 100\%, \qquad (42)$$

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

6.1.3. Similarity measure and histogram formation

Histograms are constructed on the basis of different color features representing the distribution of discrete color feature values in an n-dimensional color feature space, where $n = 3$ for RGB, rgb, $l_1 l_2 l_3$, $c_1 c_2 c_3$ and $m_1 m_2 m_3$, and $n = 1$ for I, S and H. During histogram construction, all pixels in a color image are discarded with a local saturation and intensity smaller than 5% of the total range. 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
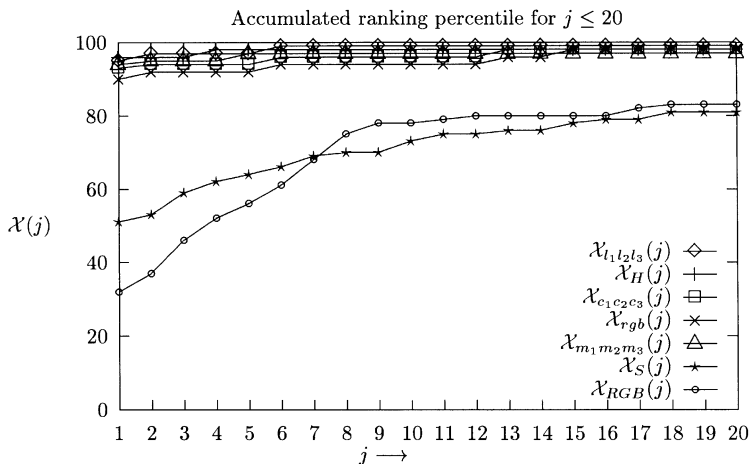

Fig. 3. The discriminative power of the histogram matching process differentiated for the various color features plotted against the ranking j. The cumulative percentile $\mathcal{X}$ for $\mathcal{H}_{l_1 l_2 l_3}$, $\mathcal{H}_H$, $\mathcal{H}_{c_1 c_2 c_3}$, $\mathcal{H}_{rgb}$, $\mathcal{H}_{m_1 m_2 m_3}$, $\mathcal{H}_S$, and $\mathcal{H}_{RGB}$ is given by $\mathcal{X}_{l_1 l_2 l_3}$, $\mathcal{X}_H$, $\mathcal{X}_{c_1 c_2 c_3}$, $\mathcal{X}_{rgb}$, $\mathcal{X}_{m_1 m_2 m_3}$, $\mathcal{X}_S$ and $\mathcal{X}_{RGB}$, respectively.
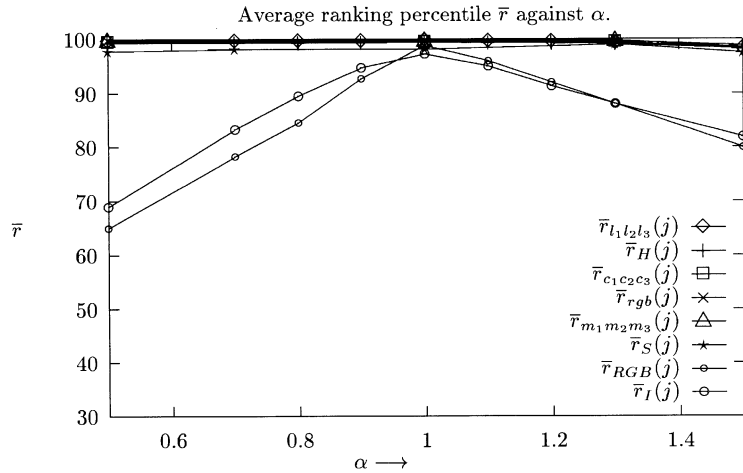
Fig. 4. The discriminative power of the histogram matching process differentiated for the various color features plotted against the illumination intensity represented by variation as expressed by the factor α. The average percentile $\bar{r}$ for $\mathcal{H}_{l_1 l_2 l_3}$, $\mathcal{H}_H$, $\mathcal{H}_{c_1 c_2 c_3}$, $\mathcal{H}_{rgb}$, $\mathcal{H}_{m_1 m_2 m_3}$, $\mathcal{H}_S$, $\mathcal{H}_{RGB}$ and $\mathcal{H}_I$ is given by $\bar{r}_{l_1 l_2 l_3}$, $\bar{r}_H$, $\bar{r}_{c_1 c_2 c_3}$, $\bar{r}_{rgb}$, $\bar{r}_{m_1 m_2 m_3}$, $\bar{r}_S$, $\bar{r}_{RGB}$ and $\bar{r}_I$, respectively.

Fig. 5. Four of the 10 objects with spatially varying illumination.

considered in the matching process. For comparison reasons in the literature, in this paper, the histogram similarity function is expressed by histogram intersection [1].

Histogram axes are partitioned uniformly with fixed intervals. The resolution on the axes follows from the amount of noise and computational efficiency considerations. We determined the appropriate bin size for our application empirically. This has been achieved by varying the same number of bins on the axes over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ and chose the smallest q for which the number of bins is kept small for computational efficiency and large for recognition accuracy. The results show (not presented here) that the number of bins was of little influence on the recognition accuracy when the number of bins ranges from $q = 32$ to 256 for all color spaces. Therefore, the histogram bin size used during histogram formation is $q = 32$ in the following. For each test and reference image, 3-D histograms are created for the RGB, $l_1 l_2 l_3$, rgb and $c_1 c_2 c_3$ color space denoted by $\mathcal{H}_{RGB}$, $\mathcal{H}_{l_1 l_2 l_3}$, $\mathcal{H}_{rgb}$ and $\mathcal{H}_{c_1 c_2 c_3}$, respectively. Furthermore, 1-D histograms are created for I, S and H denoted by $\mathcal{H}_I$, $\mathcal{H}_S$, and $\mathcal{H}_H$.

Assuming a uniform distribution of the RGB colors implies, however, a non-uniform distribution of color ratios m_1, m_2 and m_3 and corresponding $\mathcal{G}_{m_1}(\mathbf{x})$, $\mathcal{G}_{m_2}(\mathbf{x})$,

	$l_1 l_2 l_3$	H	$c_1 c_2 c_3$	rgb	$m_1 m_2 m_3$	S	RGB	I
rank 1	8	7	9	9	7	0	0	0
rank 2	2	1	1	0	1	0	0	0
rank 3	0	0	0	1	1	0	0	0
rank >3	0	2	0	0	1	10	10	10

Fig. 6. Ranking statistics of matching the 10 images with spatially varying illumination against the database of 500 images.

and $\mathcal{G}_{m_3}(\mathbf{x})$ denoting the gradient magnitude at local edge maxima in images $\ln(R/G)$, $\ln(R/B)$ and $\ln(G/B)$, respectively. Unfortunately, we observed from the reference images in the datasets that RGB colors are non-uniformly distributed and hence a theoretical model of the probability distribution of ratios is not feasible. To that end, an experimental probability distribution is generated by computing $\mathcal{G}_{m_1}(\mathbf{x})$, $\mathcal{G}_{m_2}(\mathbf{x})$, and $\mathcal{G}_{m_3}(\mathbf{x})$ for the 500 images in the image database. According to the experimentally determined probability distribution (not shown here), we partition the gradient magnitude axes finely near 0 and sparsely when reaching maximum by their projection onto the log axis. In this way, a 3-dimensional histogram is created for $\mathcal{G}_{m_1}(\mathbf{x})$, $\mathcal{G}_{m_2}(\mathbf{x})$, and $\mathcal{G}_{m_3}(\mathbf{x})$ denoted by $\mathcal{H}_{m_1 m_2 m_3}$.
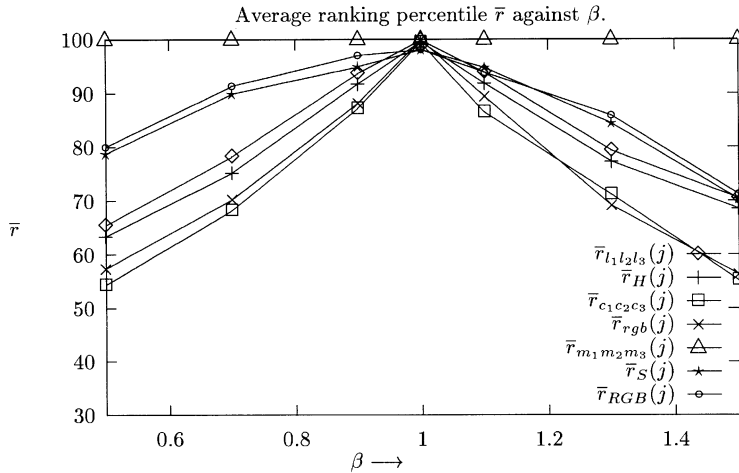
Fig. 7. The discriminative power of the histogram matching process differentiated for the various color features plotted against the change β in the color composition of the illumination spectrum. The average percentile $\bar{r}$ for $\mathcal{H}_{l_1l_2l_3}$, $\mathcal{H}_H$, $\mathcal{H}_{c_1c_2c_3}$, $\mathcal{H}_{m_1m_2m_3}$, $\mathcal{H}_S$ and $\mathcal{H}_{RGB}$ is given by $\bar{r}_{l_1l_2l_3}$, $\bar{r}_H$, $\bar{r}_{c_1c_2c_3}$, $\bar{r}_{m_1m_2m_3}$, $\bar{r}_S$ and $\bar{r}_{RGB}$, respectively.

	RGB	$c_1c_2c_3$ rgb	$l_1l_2l_3$ H	$c_1c_2c_3$ rgb	$m_1m_2m_3$	$l_1l_2l_3$ H	$c_1c_2c_3$ rgb	$m_1m_2m_3$	$l_1l_2l_3$ H	$m_1m_2m_3$	$m_1m_2m_3$	$m_1m_2m_3$	$m_1m_2m_3$	$l_1l_2l_3$ H	$m_1m_2m_3$
viewpoint & geometry	+	-	+	+	+	-	-	-	+	+	+	+	-	-	-
highlights	+	+	-	+	+	-	+	+	-	-	+	-	+	-	-
illumination intensity	+	+	+	-	+	+	-	+	-	+	-	-	-	+	-
illumination color	+	+	+	+	-	+	+	-	+	-	-	-	-	+	-

Fig. 8. Overview of which color models to use under which imaging conditions + denotes controlled and − denotes uncontrolled imaging condition.

6.2. Experimental results

6.2.1. Results with white illumination

In this section, 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 color 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. 3 accumulated ranking percentile is shown for the various color features.

From the results of Fig. 3 we can observe that the discriminative power of $l_1l_2l_3$, H followed by $c_1c_2c_3$, rgb and $m_1m_2m_3$ is higher than the other color models achieving a probability of, respectively, 97, 96, 94, 92 and 89 perfect matches out of 100. Saturation S provides significantly worse recognition accuracy. As expected, the discriminative power of RGB has the worst performance due to its sensitivity to varying viewing directions and object positionings.

6.2.2. The effect of a change in the illumination intensity

The effect of a change in the illumination intensity is equal to the multiplication of each RGB-color by a uniform scalar factor α. In theory, we have shown that only RGB and I-color features are sensitive to changes in the illumination intensity. To measure the sensitivity of different color features, in practice, RGB-images of the test set are multiplied by a constant factor varying over $\alpha \in \{0.5, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.5\}$. The discriminative power of the histogram matching process differentiated for the various color features plotted against illumination intensity is shown in Fig. 4.

As expected, RGB and I-color features depend on the illumination intensity. The further illumination intensity deviates from the original value (i.e. $\alpha = 1$), the worse discriminative power is achieved. Note that objects are recognized randomly for $\bar{r} = 50$. Furthermore, all other color feature are fairly independent under varying intensity of the illumination.

To test recognition accuracy for real images under varying illumination intensity, an independent test set of

recordings was made by randomly chosen 10 objects already in the database of 500 images. These objects were recorded again with the same pose but with spatially varying illumination intensity, see Fig. 5.

Then these 10 images were matched against the database of 500 images. From Fig. 6 is can be observed that the discriminative power of $c_1c_2c_3$ and rgb (with 9 perfect matches out of 10) with respect to $l_1l_2l_3$ and H is similar or even better due to minor amount of highlights in the test set. Further, $m_1m_2m_3$ shows very high matching accuracy, whereas S, RGB and I provide very poor matching accuracy under spatially varying illumination.

6.2.3. The effect of a change in the illumination color

Based on the coefficient rule or von Kries model, the change in the illumination color is approximated by a 3×3 diagonal matrix among the sensor bands and is equal to the multiplication of each RGB-color band by an independent scalar factor [3,11]. Note that the diagonal model of illumination change holds exactly in the case of narrow-band sensors. In theory, all color features except color ratio $m_1m_2m_3$ are sensitive to changes in the illumination color. To measure the sensitivity of the various color feature, in practice, with respect to a change in the color of the illumination, the R, G and B-images 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 discriminative power of the histogram matching process differentiated for the various color features plotted against the illumination color is shown in Fig. 7. For $\beta < 1$ the color is bluish whereas reddish for $\beta > 1$.

As expected, only the color ratio $m_1m_2m_3$ is insensitive to a change in illumination color. From Fig. 7 we can observe that color features $l_1l_2l_3$, H, $c_1c_2c_3$ and rgb which achieved highest recognition accuracy under white illumination, see Figs. 3, 4 and 6, 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.

7. Discussion

From the experimental results it is concluded that, under the assumption of a white light source, the discriminative power of $l_1l_2l_3$, H followed by $c_1c_2c_3$, rgb and $m_1m_2m_3$ is approximately the same. Saturation S provides significantly worse recognition accuracy. The discriminative power of RGB has the worst performance due to its sensitivity to varying imaging conditions. When no constraints are imposed on the illumination, the proposed color ratio $m_1m_2m_3$ is most appropriate. Based on both the reported theory and the experimental results, we now present a schema which color models to use under which imaging conditions to achieve both

invariant and discriminatory object recognition, see Fig. 8.

The schema is useful for object recognition applications where no constraints on the imaging process can be imposed as well as for applications where one or more parameters of the imaging process can be controlled such as for robots and industrial inspection (e.g. controlled object positioning and lightning).

For such a case, color models can be used for object recognition which are less invariant (at least under the given imaging parameters), but having higher discriminative power. For example, an inspection task for which lighting is controlled, but not the exact position of the object (on the conveyer belt), color model $l_1l_2l_3$ is most appropriate for the inspection task at hand. In addition, when the object does not produce a significant amount of highlights, then $c_1c_2c_3$ or rgb should be taken.

8. Conclusion

In this paper, new color models have been proposed which are analyzed in theory and evaluated in practice for the purpose of recognition of multicolored objects invariant to a substantial change in viewpoint, object geometry and illumination.

In conclusion, RGB is most appropriate for multicolored object recognition when all imaging conditions are controlled. Without the presence of highlights and under the constraint of white illumination, $c_1c_2c_3$ and normalized color rgb are most appropriate. When images are also contaminated by highlights, $l_1l_2l_3$ or H should be taken for the job at hand. When no constraints are imposed on the SPD of the illumination, $m_1m_2m_3$ is most appropriate.

We concluded by presenting a schema on which color models to use under which imaging conditions to achieve on both invariant and discriminatory recognition of multicolored objects.

References

[1] M.J. Swain, D.H. Ballard, Color indexing, Int. J. Computer. Vision 7(1) (1991) 11–32.

[2] B.V. Funt, G.D. Finlayson, Color constant color indexing, IEEE Trans. PAMI 17(5) (1995) 522–529.

[3] E.H. Land, J.J. McCann, Lightness and retinex theory, J. Opt. Soc. Am. 61 (1971) 1–11.

[4] G. Healey, D. Slater, Global color constancy: recognition of objects by use of illumination invariant properties of color distributions, J. Opt. Soc. Am. 11(11) (1995) 3003–3010.

[5] G.D. Finlayson, S.S. Chatterjee, B.V. Funt, Color angular indexing, ECCV96, Vol. II, 1996, pp. 16–27.

[6] D. Slater, G. Healey, The illumination-invariant recognition of 3-D objects using local color invariants, IEEE Trans. PAMI 18(2) (1996) 206–211.

[7] S.A. Shafer, Using color to separate reflection components, COLOR Res. Appl. 10(4) (1985) 210–218.

[8] D. Forsyth, A novel algorithm for color constancy, Int. J. Comput. Vision 5 (1990) 5–36.

[9] B.V. Funt, M.S. Drew, Color constancy computation in near-mondrian scenes, pp. 544–549. CVPR, IEEE Computer Society Press, Silver Spring, MD, 1988.

[10] T. Gevers, Color image invariant segmentation and retrieval, Ph.D. Thesis, ISBN 90-74795-51-X, University of Amsterdam, The Netherlands, 1996.

[11] G.D. Finlayson, M.S. Drew, B.V. Funt, Spectral sharpening: sensor transformations for improved color constancy, J. Opt. Soc. Am. 11(5) (1994) 1553–1563.

About the Author—THEO GEVERS received his Ph.D. degree in Computer Science from the University of Amsterdam in 1996 for a thesis on color image segmentation and retrieval. 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.

About the Author—ARNOLD W.M. SMEULDERS is professor of Computer Science on Multi Media Information Systems. He has been in image processing since 1977 when he completed his M.Sc. in physics from Delft University of Technology. Initially, he was interested in accurate and precise measurement from digital images. His current interest is in image databases and intelligent interactive image analysis systems, as well as method- and system engineering aspects of image processing and image processing for documents and geographical information.