A comparative study of several color models for color image invariant retrieval
Gevers, T.; Smeulders, A.W.M.

Published in:
Proceedings of the First international workshop on image databases and multimedia search, Amsterdam

Citation for published version (APA):
A Comparative Study of Several Color Models for Color Image Invariant Retrieval

T. Gevers and A.W.M. Smeulders
Department of WINS, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
e-mail: gevers@wins.uva.nl

abstract

In this paper, our goal is to analyze and evaluate various color features to be used for the purpose of image retrieval by color-metric histogram matching independent of varying imaging conditions.

In theory, we will examine the effect of a change in camera viewpoint, object surface orientation, illumination intensity, and illumination color for the various color features under a simple reflection model.

To evaluate photometric color invariance in practice, experiments have been carried out on a database consisting of 500 images taken from multicolored man-made objects in real world scenes. Images in the database show a considerable amount of noise, shadows, shading and specularities resulting in a good representation of views from everyday life as it appears in home video and consumer photography in general.

1 Introduction

A promising approach to image retrieval is image retrieval by content [2], [3], for example. The search is carried out from a pictorial specification defined by an example picture image given by the user on input. Image retrieval is then the problem of identifying the query image as (a part of) target images in the image database.

A simple and effective retrieval scheme is to represent and match images on the basis of color-metric histograms [4]. In this regard, images are retrieved that have similar color information distribution with respect to the color information distribution of the query image.

In this paper, our aim is to analyze and evaluate various color features to be used for the purpose of image retrieval by color-metric histogram matching according to the following criteria:

- Invariant to a change in viewpoint of the camera.
- Invariant to a change in position and orientation of the object in view.
- Invariant to a change in the direction and intensity of the illumination.
- Invariant to a change in the color of the illumination.
- High discrimination power to allow for color image retrieval of large image databases.
- Robust to noise in the images or model deviations in the object.
- Easy to compute to allow for fast image retrieval.

The general application is considered of matching and retrieving images taken from 2-D and 3-D multicolored objects in real-world 3-D scenes which is representative for applications such as the retrieval of postal stamps, museum objects, textile patterns, supermarket objects and 2-D and 3-D multicolored objects in general.

This paper is organized as follows. In Section 2, basic color features are defined and a simple reflection model is discussed. The reflection model is used to analyze color features with respect to the first four above mentioned criteria. In Section 3, image retrieval by color-metric histogram matching is discussed. The performance of the various color features is evaluated on the image database in Section 4.

2 Photometric Invariant Color Models

Commonly used well-known color spaces include: RGB, YIQ, XYZ, I₁I₂I₃, rgb, xyz, U*V*W*, L*a*b*, Luv and ISH. However, a number of these color features are related to intensity I (L*, W*), or they are linear combinations of RGB (YIQ, XYZ and I₁I₂I₃) or normalized with respect to intensity rgb (xyz, U*V*, a*b*, w). Therefore, in this paper, we concentrate on the following well-known standard, essentially different, color features: RGB, I, rgb, H, S. We define these color features in Section 2.1. Further, in Section 2.2, we study the effect of varying imaging conditions on the various color features for the Lambertian reflection model. The effect of highlights is studied in Section 2.3. Because all color features depend on the illumination color, a color constant color model is proposed in Section 2.4.

2.1 Basic Well-known Color Definitions

Let R, G and B, obtained from a color camera, represent the tri-stimulus components defining a mapping from image space to a 3-D sensor space:

\[ C = \int_{\lambda} E(\lambda)U_{c}(\lambda) d\lambda \] (1)

for tri-stimulus values \( C \in (R, G, B) \), where \( E(\lambda) \) is the radiance spectrum and \( U_{c} \) are the three color filter transmission functions.

To represent the RGB-color space, a cube can be defined on the R, G, and B axes. The axis connecting the black and white corners defines the intensity:

\[ I(R, G, B) = R + G + B \] (2)

The projection of RGB points on the chromaticity triangle is defined by:
\[ r(R, G, B) = \frac{R}{R + G + B}, \quad g(R, G, B) = \frac{G}{R + G + B}, \quad b(R, G, B) = \frac{B}{R + G + B} \]  

(3)
yielding the rgb color space.

The transformation from RGB used here to describe the hue \( H \) is given by:

\[ H(R, G, B) = \arctan\left( \sqrt{3(G - B)} \right) \]

(4)
and \( S \) measuring the relative white content of a color as having a particular hue by:

\[ S(R, G, B) = 1 - \frac{\min(R, G, B)}{R + G + B} \]

(5)
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.

### 2.2 Body Reflection Invariance

In this paper, body (diffuse) reflection is assumed to be Lambertian:

\[ L_C = \alpha I_C k_C (\mathbf{n} \cdot \mathbf{s}) \]

(6)
for \( C \in \{R, G, B\} \) giving the \( C \)th sensor response. \( I_C \) is the intensity of a point light source and \( k_C \) is the diffuse reflection coefficient of the surface. The surface normal is denoted by \( \mathbf{n} \) and illumination direction by \( \mathbf{s} \). Furthermore, \( \alpha = \frac{1}{a + h} \) where \( a \) is the distance from the viewpoint to the surface and \( h \) is a constant. Lambertian reflection models dull, matte surfaces. For the moment, we assume that the illumination color is white (i.e. \( I = I_R = I_G = I_B \)).

According to the body reflection model, the color depends only on the surface albedo \( k_C \) and the brightness on factor \( \alpha I(\mathbf{n} \cdot \mathbf{s}) \). As a consequence, a homogeneously painted surface (i.e. with fixed \( k_C \)) may give rise to a broad variance of RGB values due to the varying circumstances induced by the radiometric process. The same argument holds for intensity \( I \).

Opposed to RGB and \( I \), other color features are invariant under the given body reflection model.

rgb is an invariant for the set of matte, dull surfaces mathematically specified by:

\[ r(L_R, L_G, L_B) = \frac{\alpha I(\mathbf{n} \cdot \mathbf{s}) k_R}{\alpha I(\mathbf{n} \cdot \mathbf{s})(k_R + k_G + k_B)} = \frac{k_R}{k_R + k_G + k_B} \]

(7)
where \( I = I_R = I_G = I_B \). Equal argument holds for \( g \) and \( b \).

In fact, the scalar multiplier \( \alpha I(\mathbf{n} \cdot \mathbf{s}) \) is factored out resulting in an expression of (normalized) surface albedos independent of the viewpoint, surface orientation and illumination direction and intensity.
S is also an invariant for the set of matte, dull surfaces:

\[
S(L_R, L_G, L_B) = 1 - \frac{\min(\alpha I(n \cdot s)k_R, \alpha I(n \cdot s)k_G, \alpha I(n \cdot s)k_B)}{\alpha I(n \cdot s)(k_R + k_G + k_B)} = 1 - \frac{\min(k_R, k_G, k_B)}{(k_R + k_G + k_B)} \tag{8}
\]

Similarly, \(H\) is an invariant:

\[
H(L_R, L_G, L_B) = \arctan\left(\sqrt{\frac{3\alpha I(n \cdot s)(k_G - k_B)}{\alpha I(n \cdot s)((k_R - k_G) + (k_R - k_B))}}\right) = \arctan\left(\sqrt{\frac{3(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}}\right) \tag{9}
\]

The effect of surface reflection is discussed in following section.

2.3 Body and Surface Reflection Invariance

In this paper, the reflection from composite materials is approximated by the sum of the Lambertian body reflection component \(L_C\) and a surface reflection component \(S_C\). This is the Torrance-Sparrow model:

\[
M_C = L_C + S_C \tag{10}
\]

where \(C \in \{R, G, B\}\).

We use the modified Torrance-Sparrow specular reflection term:

\[
S_C = \left(\frac{FPG}{n \cdot v}\right)L_C \tag{11}
\]

where \(v\) is a unit vector in the direction of the viewer and \(F\) is the Fresnel coefficient describing the percentage of the light reflected at the interface. Because the Fresnel coefficient is weakly dependent on incident angle and wavelength we will assume that it is a constant for a given material. \(P\) models the surface roughness and \(G\) is the geometric attenuation factor. 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.

By filling in \(L_C + S_C\) in the hue equation, we can see that all possible colors of the same shiny homogeneously colored surface, illuminated by a white light source, have to be of the same hue mathematically specified as:

\[
H(M_R, M_G, M_B) = \arctan\left(\sqrt{\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}}\right) \tag{12}
\]

where illumination and geometric components are factored out resulting in an expression of surface albedos.
Obviously other color features depend on the contribution of the surface reflection component \( S_C \) and hence are sensitive to highlights. To that end, we propose a new color model \( l \):

\[
l_1(R,G,B) = \frac{(R - G)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}, \tag{13}
\]

\[
l_2(R,G,B) = \frac{(R - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}, \tag{14}
\]

\[
l_3(R,G,B) = \frac{(G - B)^2}{(R - G)^2 + (R - B)^2 + (G - B)^2}. \tag{15}
\]

the set of normalized color differences which is, similar to \( H \), an invariant for the set of matte and shiny surfaces as follows from:

\[
l_1(M_R, M_G, M_B) = \frac{(k_R - k_G)^2}{(k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2}. \tag{16}
\]

Equal argument holds for \( l_2 \) and \( l_3 \).

Unfortunately, all above mentioned color models depend on the illumination color [2]. Therefore, in the next section, a color constant color model is proposed.

### 2.4 Color Constant Color Model

[1] proposes a simple and effective illumination-independent color feature for the purpose of image retrieval. The method fails, however, when images are contaminated by shadows, shading and highlights.

To that end, we propose a color constant feature not only independent of the illumination color but also discounting shadowing and shading cues.

The color feature is defined by:

\[
m(C_1^{x_1}, C_1^{x_2}, C_2^{x_1}, C_2^{x_2}) = \frac{C_1^{x_1} C_2^{x_2}}{C_1^{x_2} C_2^{x_1}}, \tag{17}
\]

expressing the color ratio between two neighboring image locations, for \( C_1, C_2 \in \{R, G, B\} \) and \( C_1 \neq C_2 \) where \( x_1 \) and \( x_2 \) denote the image locations of the two neighboring pixels.

Having three color components of two locations, color ratios obtained from a \( RGB \)-color image are:

\[
m_1(R^{x_1}, R^{x_2}, G^{x_1}, G^{x_2}) = \frac{R^{x_1} G^{x_2}}{R^{x_2} G^{x_1}}, \tag{18}
\]

\[
m_2(R^{x_1}, R^{x_2}, B^{x_1}, B^{x_2}) = \frac{R^{x_1} B^{x_2}}{R^{x_2} B^{x_1}}, \tag{19}
\]

\[
m_3(G^{x_1}, G^{x_2}, B^{x_1}, B^{x_2}) = \frac{G^{x_1} B^{x_2}}{G^{x_2} B^{x_1}}, \tag{20}
\]
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 \( m_3 \).

Assuming Lambertian surface reflection, the color ratio is independent of a change in surface orientation, viewpoint and direction of the illumination as follows from:

\[
m_1(L_R^y, L_G^y, L_G^x) = \frac{\alpha_x^1 I_R^y k_R^y (\mathbf{n}^y \cdot \mathbf{s}) \alpha_x^2 I_G^y k_G^y (\mathbf{n}^y \cdot \mathbf{s})}{\alpha_y^1 I_R^x k_R^x (\mathbf{n}^x \cdot \mathbf{s}) \alpha_y^2 I_G^x k_G^x (\mathbf{n}^x \cdot \mathbf{s})} = \frac{k_R^y k_G^y}{k_R^x k_G^x} \tag{21}
\]

where \( y_1 \) and \( y_2 \) are two neighboring locations on the object’s surface not necessarily of the same orientation. Further, \( \alpha^y = \frac{1}{\rho(y)} \) depends on \( \rho(y) \) denoting the relative distance of the perspective viewpoint to object’s surface location \( y \).

In theory, when \( y_1 \) and \( y_2 \) are neighboring locations on the same homogeneously painted surface, the color ratio 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. Thus, in theory, the number of distinct color ratio values is a measure for the amount of distinct color edges.

If we assume that the color of the illumination is locally constant (at least over the two neighboring locations from which ratio is computed), the color ratio is also independent of the illumination color:

\[
m_1(L_R^y, L_G^y, L_G^x) = m_1(\beta_1 L_R^y, \beta_1 L_R^x, \beta_2 L_G^y, \beta_2 L_G^x) = \frac{\alpha_y^1 \beta_1 I_R^y k_R^y (\mathbf{n}^y \cdot \mathbf{s}) \alpha_y^2 \beta_2 I_G^y k_G^y (\mathbf{n}^y \cdot \mathbf{s})}{\alpha_y^1 \beta_1 I_R^x k_R^x (\mathbf{n}^x \cdot \mathbf{s}) \alpha_y^2 \beta_2 I_G^x k_G^x (\mathbf{n}^x \cdot \mathbf{s})} = \frac{k_R^y k_G^y}{k_R^x k_G^x} \tag{22}
\]

where we assume that the change in illumination color is equal to the multiplication of each \( I_R, I_G \) and \( I_B \) by an independent scalar factor \( \beta_1, \beta_2, \beta_3 \in [0, 1] \). Hence, color ratio \( m_1 \) is independent of a change in viewpoint, surface orientation, illumination intensity, and illumination spectral color composition.

Taking logarithms of both sides of equation 17 results for \( m_1 \) in:

\[
\ln m_1(R^x, R^y, G^x, G^y) = \ln R^y + \ln G^y - \ln R^x - \ln G^x \tag{23}
\]

The color ratios can be seen as differences at two neighboring locations \( x_1 \) and \( x_2 \) in the image domain:

\[
d_{m_1}(x_1, x_2) = \ln R^y + \ln G^y - \ln R^x - \ln G^x \tag{24}
\]

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 [2].

The results obtained so far for \( m_1 \) hold also for \( m_2 \) and \( m_3 \), yielding a 3-tuple \((\mathcal{R}_{m_1}(x), \mathcal{R}_{m_2}(x), \mathcal{R}_{m_3}(x))\) for every neighborhood centered at \( x \in T \).

For pixels on a homogeneously painted region, 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 color meet.
3 Image Retrieval

Color-metric histograms are created on the basis of each color feature defined in Section 2 for each image in the image database by counting the number of times a discrete color feature occurs in the image. The color-metric histogram from the query image is created in a similar way. Then, image retrieval is reduced to the problem of what extent histogram $H^Q$ derived from the query image $Q$ is similar to a histogram $H^{I_k}$ constructed for each image $I_k$ in the image database. A similarity function $D(H^Q, H^{I_k})$ is required returning a numerical measure of similarity between $H^Q$ and $H^{I_k}$.

For comparison reasons in the literature, in this paper, $D$ is expressed by histogram intersection [4]:

$$D_a(H^Q, H^{I_k}) = \frac{\sum_{k=1}^{N_d} \min \{H^Q(k), H^{I_k}(k)\}}{N_d}$$

(25)

where $k$ denote the bin index and $N_d$ the number of non zero bins.

4 Experiments

To evaluate the various color models, the criteria 1-7 of Section 1 will be addressed.

The data sets on which the experiments will be conducted are described in Section 4.1. The error measures and histogram formation are given in 4.2 and 4.3 respectively.

4.1 Datasets

The database consists of $N_1 = 500$ images of 2-D and 3-D 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. Objects were recorded at a pace of a few shots a minute. There was no attempt to individually control focus or illumination. Images show a considerable amount of noise, shadows, shading and highlights. As a result, recordings are best characterized as snap shot quality, a good representation of views from everyday life as it appears in home video, the news, and consumer photography in general.

A second, independent set (the query set) of recordings was made of randomly chosen objects already in the database. These objects, $N_2 = 70$ 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 scale).
In the experiments, all pixels in a color image are discarded with a local saturation smaller than 5 percent of the total range (this number was empirically determined by visual inspection); otherwise calculation of $H$, $S$, rgb, $m$ and $l$ becomes unstable. 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.

### 4.2 Error measures

For a measure of match quality, let rank $r_{Q_i}$ denote the position of the correct match for query image $Q_i$, $i = 1, ..., N_2$, 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:

$$
\bar{r} = \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r_{Q_i}}{N_1 - 1} \times 100\%
$$

(26)

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

$$
\mathcal{N}(j) = \left( \frac{1}{N_2} \sum_{k=1}^{j} \eta(r_{Q_i} = k) \right) \times 100\%
$$

(27)

where $\eta$ reads as the number of query images having rank $k$.

### 4.3 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 RGB, rgb, $l$ and $m$, and $n = 1$ for $I$, $S$ and $H$.

Except for $m$, see [2], 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 retrieval accuracy. The results show (not presented here) that the number of bins was of little influence on the retrieval accuracy when the number of bins ranges from $q = 32$ to $q = 256$ for all color spaces. Therefore, the histogram bin size used during histogram formation is $q = 32$ in the following.
4.4 Discriminative Power

In this subsection, we report on the image retrieval accuracy of the matching process for \( N_2 = 70 \) query images and \( N_1 = 500 \) target images for the various color features. As stated, white lighting is used during the recording of the images in the image database and the independent query set. However, the objects were recorded with a new, arbitrary position and orientation with respect to camera. In Fig. 1.a accumulated ranking is shown for similarity function based on histogram intersection.

![Fig. 1. a. The accumulated ranking percentile. b. The discriminative power plotted against the change \( \beta \) in the color composition of the illumination spectrum.](image)

From the results of Fig. 1.a we can observe that the discrimination power of \( l, rgb \) and \( H \) is higher than the other color features. Saturation \( S \) and color ratio \( m \) provide slightly worse image retrieval accuracy. As expected, the discrimination power of \( RGB \) has the worst performance due to its sensitivity to varying imaging conditions.

Hence, for image retrieval purposes under white lighting, color features \( l, rgb \) and \( H \) are most appropriate achieving a probability of respectively 99, 98 and 92 perfect matches out of 100.

4.5 The Effect of a Change in the Illumination Color

The effect of multiple light sources of different color distributions and a change in the illumination color is equal to the multiplication of each \( RGB \) image by an independent scalar factor. 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 query 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 \( \beta \) 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 color features plotted against the illumination color is shown in Fig. 1.b.
As expected, only the color ratio $m$ is insensitive to a change in illumination color. From Fig. 1.b we can observe that color features $H$, $l$ and $rgb$, which achieved best retrieval accuracy under white illumination, see Fig. 1.a, are highly sensitive to a change in illumination color followed by $S$. Even for a slight change in the illumination color, their retrieval potential degrades drastically.

5 Conclusion

In this chapter, various color features have been analyzed and evaluated for the purpose of image retrieval by color-metric histogram matching under varying illumination circumstances.

On the basis of the above reported theory and experiments, it is concluded that $l$ (invariant for both matte and shiny surfaces) followed by $H$ and $rgb$ are most appropriate to be used for image retrieval by color-metric histogram matching under the constraint of a white illumination source. When no constraints are imposed on the imaging conditions (i.e. the most general case), the proposed color ratio $m$ is most appropriate discounting the disturbing influences of shadows, shading, illumination intensity and illumination color.

References


This article was processed using the \LaTeX macro package with LLNCS style