Intelligent Sensory Information Systems
Faculty of Mathematics and Computer Science
University of Amsterdam
The Netherlands

PicToSeek: A Content-Based Image Search System for the World Wide Web

Theo Gevers
Faculty of Mathematics, Computer Science, Physics and Astronomy,
University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
tel: +31 20 525 7516,
fax: +31 20 525 7490,
e-mail: `gevers@wins.uva.nl`

We describe a content-based image search system for exploring visual information on the World Wide Web. As the amount of visual information (e.g. images, animations, graphics and videos) being published on the Web exponentially escalate, demands are also growing for search engines which catalog and index visual information. While text-based search engines have been successful in locating textual documents, exploring visual information is beyond their capabilities since processing textual data requires completely different computational methods than visual data processing. Therefore, we propose an image search system by which visual information is automatically collected, indexed and cataloged entirely on the basis of the pictorial content. The system allows for fast on-line image search by combining: (1) visual browsing through the precomputed image catalogue; (2) query by pictorial example; (3) query by image features. The system shows high image retrieval accuracy and fast image search.
keywords: content-based image indexing and searching, image cataloging, color invariant image features, World Wide Web.

http://target.wins.uva.nl:5345/ret_user

Status: Submitted to VISUAL'97

Contents

Intelligent **S**ensory **I**nformation **S**ystems
Faculty of Mathematics, Computer Science, Physics and Astronomy
University of Amsterdam
Kruislaan 403
1098 SJ Amsterdam
The Netherlands
tel: +31 20 525 7516
fax: +31 20 525 7490
e-mail: `gevers@wins.uva.nl`

1 Introduction

With the growth and popularity of the Web, a tremendous amount of information is made accessible publicly. Although the Web allows for storage of great amounts of multimedia information, computational methods are needed to explore and locate specific information. Currently, a large number of text-based search engines are available which explore the Web for text documents. The textual content of the documents are analyzed and condense into concise searchable indexes. The problem of locating documents on the Web shifts to fast index-searching. To explore visual information, text-based search engines assume that textual descriptions of the visual data are available. However, people are reluctant in categorizing visual information on the Web. Moreover, pictorial data is very hard to be described by verbal descriptions. Often no textual description of the pictorial information is present at all. Hence, the capabilities of current text-based search engines for cataloging visual information is very limited. Therefore, we concentrate on the use of pictorial information for cataloging and exploring visual data on the Web.

In this paper, we propose an image search engine, called PicToSeek, for cataloging and searching images on the Web entirely on the basis of the pictorial content. In the first stage, PicToSeek collects images on the Web by means of autonomous Web-crawlers. Then, the collected images are automatically cataloged by image analysis methods into various image styles and types: JFIF-GIF, grey-color, photograph-synthetic, size, date of creation, and color depth. After cataloging images, invariant color image features are extracted from the images. Color invariant features are properly integrated to produce a high-dimensional image index independent of the accidental imaging conditions discounting the disturbing influences of 1: an arbitrary viewpoint from which the shape of the physical object will appear differently for each view; 2: photometric variations in the recordings due to changes in shadowing, shading, highlights and illumination; 3: object fragmentation, occlusion and clutter by the presence of other objects in the scene. When images are automatically collected, cataloged and indexed, PicToSeek allows for fast on-line image search by combining: (1) visual browsing through the precomputed image catalogue; (2) query by pictorial example; (3) query by image features.

The paper is organized as follows. In Section 2 related work is discussed. The Web-crawler and the image classification scheme are given in Section 3. In Section 4, the invariant color image features are presented. PicToSeek is outlined in Section 5.

PicToSeek is on-line at http://target.wins.uva.nl:5345/ret_user/ as part of the ZOMAX system at http://www.wins.uva.nl/research/isis/zomax/.

2 Related Work

The matching complexity of image retrieval by image example is similar to that of traditional object recognition schemes. In fact, image retrieval by image example shares many characteristics with model-based object recognition. The main difference is that model-based object recognition is done fully automatic, whereas user intervention may be allowed for image retrieval by image example.

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, [1], [9], [17], [21], 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 image retrieval due to the 3-fold higher specificity of each pixel in the image. 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 [22]. Test images are identified that have similar opponent color (some linear transformation of RGB) distributions with respect to the opponent color distribution of the reference image. The work makes a significant contribution in introducing color for object indexing. Swain and Ballard show that the use of color for object indexing is to a large degree robust to changes in position and orientation of objects as well as changes in object shapes.

Further, attempts have been made to develop general purpose image retrieval systems based on multiple features (e.g. color, shape and texture) describing the image content and significant results have been achieved [4], [5], [6], [20], [24]. IBM research at Almaden implemented the QBIC [2] system allowing for queries on large image and video databases on the basis of color, texture and shape features. Fast indexing techniques and similarity functions are proposed. The key idea behind Photobook [15] is semantics-preserving image compression, which reduces images to a small set of perceptually significant coefficients describing the shape and texture of the images. In [16] content-based image retrieval is based on texture features. In [12] image retrieval is done entirely on the basis of shape information. In contrast to full content-based image retrieval, Chabot [14] and [19] uses a combination of visual appearance and text-based cues to retrieve images.

A major drawback of the above mentioned systems is, however, that the extracted features used during the retrieval process depend on the geometry of the object, on the viewpoint of the camera and on the accidental illumination conditions. As a result, using any of these systems, major problems occur when the query and target image are recorded under different imaging conditions. The practical usefulness of such a system where only identical views can be analyzed will be an order of magnitude less than a system where objects can be searched for on the basis of a random views. Consequently, there is an inevitable need for image features which are highly robust and invariant to the image-forming process. We aim to formulate demands to computational methods on their invariance to the object's geometry and illumination parameters.

3 Image Domain

On the Web, images are published by different people with various interests. As a consequence, a collection is created of images of various types and styles.

In Section 3.1, the different image formats dominating the WWW are discussed. A Web-crawler is outlined in Section 3.2 to acquire images from a variety of sites. In Section 3.3, an image classification method is proposed to classify images into photographical and synthetical images.

3.1 Image Formats

Two formats dominate the WWW: GIF and JFIF (JPEG). JFIF is particularly suited for images that contain many colors and color variations and for which the loss of some detail does not matter. Photographs are examples for JFIF compression. GIF is suitable for images that make use of 256 colors or less. Both image formats are raster images, i.e. the images consist of pixel values. There is no generally accepted format for vector images on the WWW.

3.2 Image Collection

A Web-crawler has been implemented. The results of the Web-crawler show a broad spectrum of image addresses. The 100.000 images in the final test set originate from 1345 sites. The number of images per class format in the test collection is 20 % photographs and 80 % synthetic images. 11 % of the collection are grey value images.

3.3 Image Classification

We catalog images according to the creation method of the image into:

- Photographical images: pictures of 2D and 3D objects in real-world cluttered scenes as recorded by a sensor (camera).
- Synthetical images: computer-generated pictures of abstractions of 2D and 3D real-world objects such as icons, graphics, drawings and legend baring images.

Although, images can have both types (e.g. a photographical image containing a graphical symbol), our attention is focussed on classifying images into photographical and synthetical images.

Images are automatically cataloged on the basis of their pictorial content. The following features are computed from the color images:

- Color variation: The number of distinct hue values in an image relative to the total number of hues. Synthetic images tend to have fewer distinct colors then photographs.
- Color saturation: The accumulation of the saturation of colors in an image relative to the total number of pixels. Colors in synthetic images are likely to be more saturated.
- Color transition strength: The pronouncement of hue edges in an image. Synthetic images tend to have more abrupt color transitions than photograph images.
- Color background: Share of supposed background in the total number of pixels. The supposed background is computed as the largest accumulation in the bin in the grey-value or hue histogram. Synthetic images are likely to score higher than photographs.
- Greyness: Share of grey valued pixels in the total number of pixels. Because the background is often grey, the value is similar to the color background.

To classify images into photographical and synthetical images, we used the Fisher's linear discriminant method for classification. The Fisher's approach minimizes the classification error. We assessed the precision of the automated image type classification method, for a training and test set of 100 images from each class. The results give an overall error rate of 0.12 for synthetical images and an overall error rate of 0.18 for photographical images. From the results it is concluded that automated type classification provides satisfactory distinction.

Further, images are automatically cataloged according to the following characteristics: JFIF-GIF, grey-color, size, date of creation, and color depth.

4 Image Retrieval by Image Example

The user can visually browse through the precomputed image catalog. However, searching for images in a specific image class is still not feasible as many images are to be inspected. To proceed, we consider image retrieval by image example, where an example query image is given by the user on input, for example [13], [15]. A typical application case is the problem of retrieving images containing an instance of a particular object. Then, the query is specified by an example picture image taken from the object at hand. The basic idea to image retrieval by image example is to extract characteristic features from target images which are matched with those of the query. These features are typically derived from shape, texture or color information of query and target. In this paper, we focus on using color information for the job at hand.

A major problem of retrieval by image example occurs when the query and target image containing the same object are recorded under different imaging conditions. In general, images are taken from objects from different viewpoints and with different object orientation. When two recordings are made of the same object from different viewpoints, the perceivable shape of the object will be distorted. Also variation in viewpoint and object orientation yield different shadowing, shading and highlighting cues changing the intensity data fields considerably. Moreover, differences in the illumination change the photometric composition of images.

Therefore, image color features are required which are:

- Invariant to a change in viewpoint;
- Invariant to a change in object orientation;
- Invariant to a change in the intensity and direction of the illumination.

Therefore, in this section, we will make a detailed account of the effect of the variability of the object's geometry and illumination parameters on the perceived color and to propose invariant measures.

In addition, a proper retrieval scheme should be robust to object fragmentation, occlusion and clutter by the presence of other objects in the view.

To that end, similarity functions are required which have:

- High discriminative power and easy to compute to allow for fast object recognition of large reference image databases;
- Robustness to object occlusion and cluttering;
- Robustness to noise in the images or model deviations in the object.

Our aim is to define invariant similarity functions satisfying the above criteria.

This section is organized as follows. First, in Section 4.1.1, the basic color features are defined and the dichromatic reflection model is discussed in 4.1.2. In Section 4.1.3, we analyze the effect of various imaging conditions on the different color features for the dichromatic reflection model. In Section 4.1.4, a guideline is given which color features to use under which imaging circumstances. Finally, in Section 4.2, various invariant similarity functions are evaluated for robust histogram matching.

4.1 Invariant Image Attributes

4.1.1 Basic Color Definitions

Commonly used well-known color spaces include: RGB, YIQ, XYZ, $I_1I_2I_3$, 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^*), linear combinations of RGB (YIQ, XYZ and $I_1I_2I_3$) or normalized with respect to intensity rgb (xyz, U^*V^*, a^*b^*, uv). Therefore, in this paper, we concentrate on the following standard, essentially different, color features: I, RGB, rgb, H and S.

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

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:

$$C = \int_\lambda p(\lambda) s_C(\lambda) d\lambda \tag{1}$$

for tri-stimulus values $C \in (R, G, B)$, where $p(\lambda)$ is the radiance spectrum and s_C are the three color filter transmission functions, see Figure 1.a.

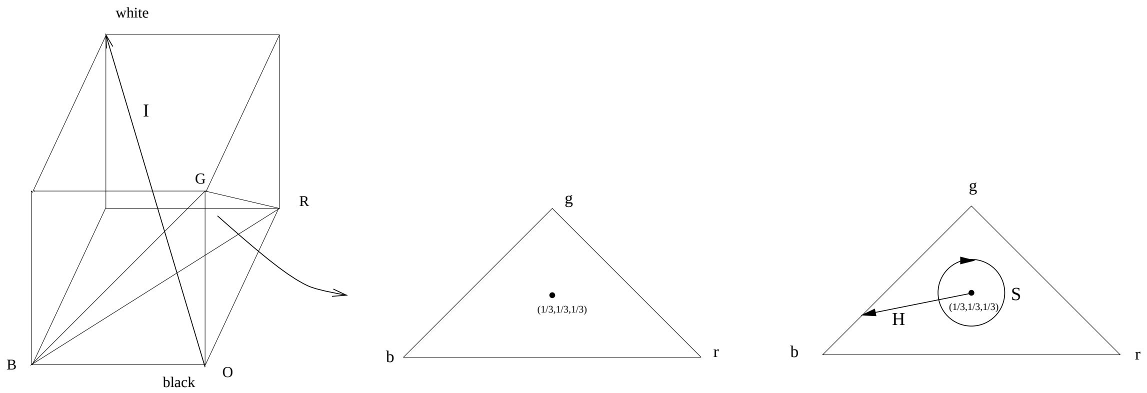

Figure 1: *a. RGB-color space b. Chromaticity plane. c. Definition of hue and saturation in the chromaticity plane.*

The 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 $R + G + B = M$, where M is the maximum light intensity (say $M = 255$), is one such plane. This plane cuts out an equilateral triangle which is the standard rgb chromaticity triangle:

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

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

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

and graphically represented by Figure 1.a and 1.b.

Hue and saturation are defined in the chromaticity triangle in the standard way as follows. If we assume a white light source (i.e. $R = G = B$) then saturation S_{rgb} is defined as the radial distance of a point from $r = g = b = 1/3$ mathematically specified as:

$$S_{rgb}(r, g, b) = \sqrt{(r - 1/3)^2 + (g - 1/3)^2 + (b - 1/3)^2} \tag{6}$$

and graphically represented by Figure 1.c.

H_{rgb} is a function of the angle between a reference line (e.g. horizontal axis) and the color point:

$$H_{rgb}(r, g, b) = \arctan\left(\frac{r - 1/3}{g - 1/3}\right) \tag{7}$$

see Figure 1.c as well.

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

$$H(R, G, B) = \arctan\left(\frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}\right) \tag{8}$$

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

$$S(R,G,B) = 1 - \frac{\min(R,G,B)}{R+G+B} \tag{9}$$

Note that rgb, H and S are undefined for $R = G = B = 0$.

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. In the next section, the dichromatic reflection model, describing the radiometric imaging process of inhomogeneous dielectric objects, is considered.

4.1.2 The Reflection Model

The reflection from inhomogeneous dielectric materials consists of a body and a surface reflection component: the body reflection component has the color of the object and the surface component has the same spectral power distribution (SPD) as the SPD of the light source.

These characteristics are described by the dichromatic reflection model [18]:

$$L_b(\vec{n},\vec{s},\lambda) + L_s(\vec{n},\vec{s},\vec{v},\lambda) = m_b(\vec{n},\vec{s})c_b(\lambda) + m_s(\vec{n},\vec{s},\vec{v})c_s(\lambda) \tag{10}$$

where L_b is the body-reflectance term and L_s the surface reflectance term. Further, 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 unit vector in the direction of the viewer. Geometric terms m_b and m_s denote the geometric dependencies on the body and surface reflection respectively.

In [10] the dichromatic reflection model is evaluated on different dielectric objects and it is concluded that the model applies well to plastics, paints, ceramics, vinyls, tiles, fruits, leaves and woods whereas the model fails on metals, cloth and papers. Furthermore they concluded that the SPD of the surface reflection is the same as the SPD of the incident light. Recall that the reflection model makes no assumptions about the geometric terms and that Lambertian [8] and the Torrance-Sparrow [23] reflection models are instantiations of the dichromatic reflection model as shown by [18].

Now consider an image of the surface patch. Using the red, green and blue sensors with spectral sensitivities given by $s_R(\lambda)$, $s_G(\lambda)$ and $s_B(\lambda)$ respectively, to obtain an image of the surface patch illuminated by a SPD of the incident light denoted by $e(\lambda)$, the measured sensor value will be given by [18]:

$$C = m_b(\vec{n},\vec{s}) \int_\lambda s_C(\lambda)e(\lambda)c_b(\lambda)d\lambda + m_s(\vec{n},\vec{s},\vec{v}) \int_\lambda s_C(\lambda)e(\lambda)c_s(\lambda)d\lambda \tag{11}$$

for $C = \{R, G, B\}$ giving the Cth sensor response.

We now present the dichromatic reflection model assuming a white light spectral distribution of the incident light. Then $e(\lambda) = e$ and $c_s(\lambda) = 1$ and we have:

$$C_w = em_b(\vec{n},\vec{s})k_C \int_\lambda s_C(\lambda)d\lambda + em_s(\vec{n},\vec{s},\vec{v}) \int_\lambda s_C(\lambda)d\lambda \tag{12}$$

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, $c_b(\lambda) = k_C$ for $C = \{R, G, B\}$ is the surface color albedo.

Consider, as is usually the case, the sensitivities of the camera $s_R(\lambda)$, $s_G(\lambda)$ and $s_B(\lambda)$ to be relatively narrow-band filters. Then their spectral response can be approximated by delta functions $s_R(\lambda) = \delta(\lambda - \lambda_R)$, $s_G(\lambda) = \delta(\lambda - \lambda_G)$ and $s_B(\lambda) = \delta(\lambda - \lambda_B)$. Then we obtain:

$$C_w = C_b + C_s = em_b(\vec{n},\vec{s})k_C + em_s(\vec{n},\vec{s},\vec{v}) \tag{13}$$

If the filters are not narrow-band, then the integrated white conditions must hold:

$$\int_\lambda s_R(\lambda) = \int_\lambda s_G(\lambda) = \int_\lambda s_B(\lambda) \tag{14}$$

This can be achieved by carefully selecting spectral filters. Throughout this presentation, we assume that the sensitivities of the camera are relatively narrow-band.

4.1.3 Photometric Color Invariant Features for Matte and Shiny Chromatic Surfaces

Consider the body reflection term of equation 13:

$$C_b = em_b(\vec{n}, \vec{s})k_C \tag{15}$$

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

According to the body reflection model, the color depends only on the surface albedo k_C and the brightness on factor $em_b(\vec{n}, \vec{s})$. If a matte surface contains a variety of surface normals, then the set of C_b colors will generate a line-like color cluster in RGB-sensor space, see Figure 2.a, where the direction of the line-like color cluster is determined by k_C and its extent by the variations of surface normals $\vec{n}$ with respect to the illumination direction $\vec{s}$. A line-like color cluster with gaps will be produced by a matte, dull dielectric object composed of a few flat surfaces. In fact, all observed colors of a surface with fixed k_C are represented by the color cluster vector $\vec{B}$, see Figure 2.a, where the direction of $\vec{B}$ is based on k_C.

As a consequence, a uniformly 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 image-forming process such as a change in object orientation, illumination intensity and position. The same argument holds for intensity I which is on the diagonal axis (i.e. grey-axis) connecting the black and white corners in the RGB-space.

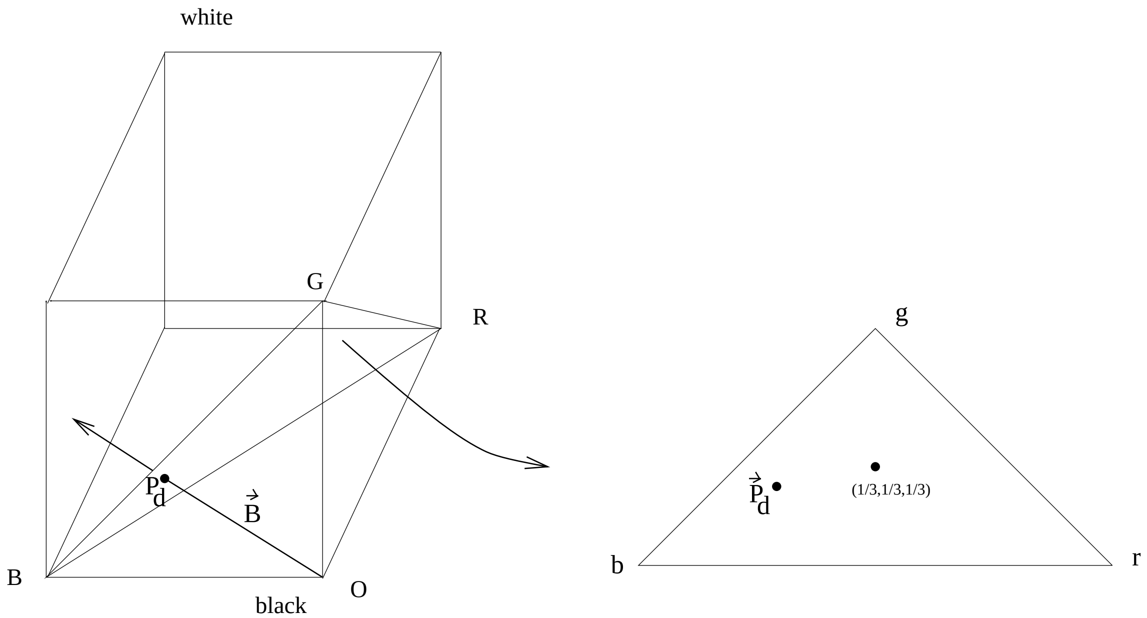

Figure 2: *a. Color cluster vector $\vec{B}$ of a matte surface in RGB. b. Projection of $\vec{B}$ onto its chroma in the chromaticity plane.*

In contrast, assuming the dichromatic reflection model with white illumination, rgb is insensitive to surface orientation, illumination direction and intensity as can be seen from:

$$r(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_R}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_R}{k_R + k_G + k_B} \tag{16}$$

$$g(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_G}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_G}{k_R + k_G + k_B} \tag{17}$$

$$b(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_B}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_B}{k_R + k_G + k_B} \tag{18}$$

The scalar multiplier $em_b(\vec{n}, \vec{s})$ is factored out resulting in an expression of (normalized) surface albedos independent of the viewpoint, surface orientation and illumination direction and intensity.

The projection of the color cluster for a matte surface onto its chroma value is graphically illustrated in Figure 2.a and 2.b, where P_d is the intersection of the color cluster vector with the unit plane in RGB-space and $\vec{p}_d$ its corresponding chroma.

Because S corresponds to the radial distance from the reference white-light point, S is an invariant for the set of matte, dull surfaces illuminated by white SPD mathematically specified by:

$$S(R_b, G_b, B_b) = 1 - \frac{\min(em_b(\vec{n}, \vec{s})k_R, em_b(\vec{n}, \vec{s})k_G, em_b(\vec{n}, \vec{s})k_B)}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} =$$

$$1 - \frac{\min(k_R, k_G, k_B)}{(k_R + k_G + k_B)} \tag{19}$$

Similarly, H is an invariant for matte, dull surfaces:

$$H(R_b, G_b, B_b) = \arctan\left(\frac{\sqrt{3}em_b(\vec{n}, \vec{s})(k_G - k_B)}{em_b(\vec{n}, \vec{s})((k_R - k_G) + (k_R - k_B))}\right) =$$

$$\arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right) \tag{20}$$

Note that surfaces of distinct surface albedo k_C (i.e. distinct uniformly painted surfaces) produce a unique rgb value. However, surfaces of distinct surface albedos may generate similar saturation value S (if their chroma values have the same radial distance from $(1/3, 1/3, 1/3)$) or similar hue value H (if their chroma values differ only in saturation).

Obviously, in practice, the assumption of objects composed of matte, dull surfaces is not always realistic.

To that end, consider the surface reflection term of equation 13:

$$C_s = em_s(\vec{n}, \vec{s}, \vec{v}) \tag{21}$$

for $C_s \in \{R_s, G_s, B_s\}$ giving the red, green and blue sensor response for a shiny 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 surface reflection color cluster $\vec{S}$ is on the diagonal grey axis of the basic RGB-color space corresponding to intensity, see Figure 3.a. The extent of the highlight color cluster depends on the roughness of the object surface. Very shiny object surfaces generate color clusters which are spread out over the entire diagonal grey axis. For rough surfaces, the extent of the observed color cluster will be small.

For a given point on a shiny surface, the contribution of the body reflection component C_b and surface reflection component C_s are added together $C_w = C_s + C_b$. Hence, the observed

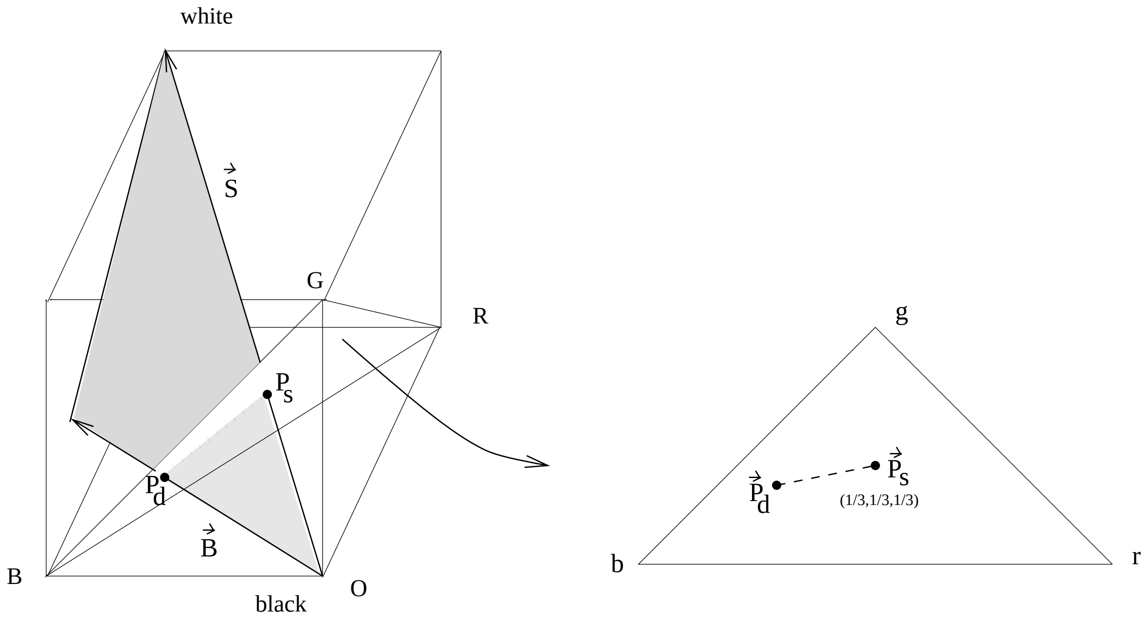

Figure 3: *a. Triangular color cluster $\vec{S}$ from a shiny uniformly colored surface in RGB. b. Projection of the triangular color cluster onto a line in the chromaticity plane originating at* $(1/3, 1/3, 1/3)$.

colors of the surface must be inside the triangular color cluster in the RGB-space formed by the two reflection components, see Fig 3.a.

The triangular color cluster in the RGB-space forms a line-cluster in rgb-space, see Fig 3.a and 3.b, as can be seen as follows. Let P_d be the intersection of body color cluster vector $\vec{B}$ in RGB-sensor space. Then the chroma values of P_d are defined by:

$$\vec{p}_d = \left(\frac{k_R}{k_R + k_G + k_B}, \frac{k_R}{k_R + k_G + k_B}, \frac{k_R}{k_R + k_G + k_B}\right) \tag{22}$$

Let P_s be the intersection of surface color cluster vector $\vec{S}$ with the unit plane. Then for the dichromatic reflection off an object illuminated with white light, we have:

$$\vec{p}_s = \left(\left(\frac{em_s(\vec{n}, \vec{s}, \vec{v})}{3em_s(\vec{n}, \vec{s}, \vec{v})}\right), \left(\frac{em_s(\vec{n}, \vec{s}, \vec{v})}{3em_s(\vec{n}, \vec{s}, \vec{v})}\right), \left(\frac{em_s(\vec{n}, \vec{s}, \vec{v})}{3em_s(\vec{n}, \vec{s}, \vec{v})}\right)\right) = (1/3, 1/3, 1/3) \tag{23}$$

That is, the reflection is also white.

Hence, $C_b + C_s$ is projected onto a line in the chromaticity originating at $\vec{p}_s$ and ending in $\vec{p}_d$, see Fig 3.b.

Because H is a function of the angle between the reference line and color point $\vec{p}_d$, all possible colors of the same shiny uniformly colored surface have to be of the same hue mathematically specified as:

$$H(R_w, G_w, G_w) =$$

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

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

$$\arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right) \tag{24}$$

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

4.1.4 Photometric Color Invariant Image Retrieval

In [3] the different color features were evaluated for the purpose of image retrieval by histogram matching.

Histograms were 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 and rgb, and $n = 1$ for I, S and H.

To evaluate expressiveness and robustness of photometric color invariant image retrieval by histogram matching, experiments have been conducted 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.

On the basis of the reported theory outlined above and experimental results given in [3], it is concluded that RGB and I are sensitive to shadows, surface orientation and illumination intensity. RGB, I, rgb and S are all sensitive to highlights. The color model m is insensitive for smoothly sloping highlights. H is independent of surface specularities. m is the only color model invariant to illumination color and hence inter reflections. Inter reflections occur when an object receives the light reflected from another object. These results are summarized in Figure 4.

	shadows	surface orientation	highlights	illumination direction	illumination intensity	illumination color	inter reflection
I	-	-	-	-	-	-	-
RGB	-	-	-	-	-	-	-
rgb	+	+	-	+	+	-	-
S	+	+	-	+	+	-	-
H	+	+	+	+	+	-	-
hue-hue edge	+	+	+	+	+	-	-
hue-hue corner	+	+	+	+	+	-	-
hue-hue Y-fork	+	+	+	+	+	-	-
color ratio m	+	+	- /+	+	+	+	+

Figure 4: *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.*

The difference between the hue H and hue-hue salient points is a larger specificity in the selection of the targeted object.

4.2 Invariant Similarity Functions

A proper retrieval scheme should be robust to object fragmentation, occlusion and clutter by the presence of other objects in the view.

To that end, similarity functions are required which have high discriminative power and easy to compute to allow for fast object recognition of large reference image databases; robustness to object occlusion and cluttering; robustness to noise in the images or model deviations in the object.

In [3] various similarity functions were presented for histogram matching.

From the results it is concluded that retrieval accuracy of similarity functions depend on the presence of object clutter in the scene. The histogram cross correlation (e.g. Chisquare) provide best retrieval accuracy without any object clutter. This is due to the fact that this similarity functions is symmetric and can be interpreted as the number of pixels with the same values in the query image which can be found present in the retrieved image and vice versa. This is a desirable property when one object per image is recorded without any object clutter. In the presence of object clutter, highest image retrieval accuracy is provided by the quadratic similarity function (e.g. histogram intersection). This is because this similarity measure count the number of similar hits and hence are insensitive to false positives.

5 PicToSeek

In the first stage, PicToSeek collects images on the Web by means of autonomous Web-crawlers. Then, the collected images are automatically cataloged by image analysis methods into various image styles and types: JFIF-GIF, grey-color, photograph-synthetic, size, date of creation, and color depth. After cataloging images, invariant color image features are extracted from the images. When images are automatically collected, indexed and cataloged, PicToSeek allows for fast on-line image search.

There are many ways in which images can be retrieved. To that end, in Section 5.1, the retrieval methods used by PicToSeek are discussed. In Section 5.2, the query capability of the system is illustrated.

5.1 Image Retrieval Methods

PicToSeek allows the following retrieval methods:

Visual browsing through the precomputed image catalogue: In the first stage, the specific image class is selected. For example, the user can select the image class of small, photographical color images. Then, PicToSeek allows visual browsing through the collection of images of the same image class.

query by pictorial example: Query by pictorial example is the retrieval method in which a query for images consists of one or more images. PicToSeek allows users to bring in their own image by means of a URL-address.

query by image features: Query by image features is the retrieval method in which a query for images consists of values of user-specified features only. PicToSeek allows the user to specify image features. For example, a user can specify "dark and reddish" images by setting the image feature parameters.

5.2 Query Scenario

All queries follow the same scenario:

Step 0 **Image Domain Selection:** Visual browsing through the precomputed image catalogue;

Step 1 **Image selection:** select an image from the catalogue or capture the query image from an object by giving an URL-address.

Step 2 **Query image:** the query image is defined as an user-specified interesting part of the selected image.

Step 3 **Invariance selection:** the required invariance is selected from the list of available invariant indices. They have been computed and stored in separate hash tables for each image when it entered the database.

Step 4 **Search:** the same invariant indices are computed from the query and matched with those stored in the database hash tables.

Step 5 **Display:** an ordered list of most similar images is shown.

Step 6 **Image selection:** if the right image is found, the image can be displayed at full resolution.

Step 7 **Rerun:** if the right image is not found the query image is adjusted (go to step 1) or the most similar image is used to refine query definition (go to step 3).

To illustrate the query capability of the system, the typical application is considered of retrieving images containing an instance of a given object. To that end, the query is specified by an example image taken from the object at hand. A typical query specification is shown in Figure 5.

PicToSeek enables the user to select and display the image of the object at hand by a Url-address. At run time, the user specifies the preferred invariance. Then, the required color invariants are extracted from the query and matched with those of the target images in the database.

After matching, images are ordered with respect to the query according to their similarity measure and displayed in the retrieval unit one by one through image browsing or as an as an ordered set according to the user preferences.

6 Implementation

The graphical-user interface of PicToSeek has been implemented in Java. The Web-crawler, image analysis and feature extraction methods have been implemented in C++. The monet database is used to store the images and indexes. The server runs on a SPARCstation 5 with 110 Mhz.

7 Summary

We described a content-based image search system for exploring visual information on the World Wide Web. PicToSeek collects images on the Web by means of autonomous Web-crawlers. Collected images are automatically cataloged by image analysis methods into various image styles and types: JFIF-GIF, grey-color, photograph-synthetic, size, date of creation, and color depth. After cataloging images, invariant color image features are extracted from the images to produce a high-dimensional image index independent of the accidental imaging conditions. PicToSeek allows for fast on-line image search by combining: (1) visual browsing through the precomputed image catalogue; (2) query by pictorial example; (3) query by image features.

References

[1] Califano, A. and Mohan, R., *Multidimensional Indexing for Recognizing Visual Shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16m No. 4, pp. 373-392, 1994.

[2] Flickner, M. et al, *Query by Image and Video Content: the QBIC system*, IEEE Computer, Vol. 28, No. 9, 1995.

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

[4] Grosky W. and Mehrotra R., Special Issue on Image Database Management, Computer, Vol. 22, No. 12, 1989.

[5] IFIP, Visual Database Systems I and II, Elsevier Science Publishers, North-Holland, 1989 and 1992.

[6] Jain, R., NSF Workshop on Visual Information Management Systems, SIGmod Record, Vol. 22, No. 3, pp. 57-75, 1993.

[7] Kender, J. R., *Saturation, Hue, and Normalized Colors: Calculation, Digitization Effects, and Use*, Technical Report, Department of Computer Science, Carnegie-Mellon University, 1976.

[8] Lambert J. H., Photometrica Sive de Mensura de Gratibus Luminis, Colorum et Umbrae, Eberhard Klett, Augsberg, Germany, 1760.

[9] Lamdan, Y. and Wolfson, H. J., *Geometric Hashing: A General and Efficient Model-Based Recognition Scheme*, In Proceedings of the Second International Conference on Computer Vision, IEEE CS Press, Order no. 883-17, pp. 238-249, 1988.

[10] Lee H.-C., Breneman E, J. and Schulte C. P., *Modeling Light Reflection for Computer Color Vision* IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 12, No. 3, pp. 402-409, 1990.

[11] Levkowitz, H. and Herman G. T., *GLHS: A Generalized Lightness, Hue, and Saturation Color Model*, CVGIP: Graphical Models and Image Processing, Vol. 55, No. 4, pp. 271-285, 1993.

[12] Mehrotra, R. and Gary, J. E., *Similar-shape Retrieval in Shape Data Management*, IEEE Computer, Vol. 28, No. 9, 1995.

[13] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D. and Yanker, P., *The QBIC Project: Querying Images by Content Using Color, Texture, and Shape*, In Proceedings of Storage and Retrieval for Image and Video Databases, SPIE, 1993.

[14] Ogle, V. E. and Stonebraker, M., *Chabot: Retrieval from a Relational Database of Images*, IEEE Computer, Vol. 28, No. 9, 1995.

[15] Pentland, A., Picard, R. W. and Sclaroff, S., *Photobook: Tools for Content-based Manipulation of Image Databases*, In Proceedings of Storage and Retrieval for Image and Video Databases II, 2, 185, SPIE, Bellingham, Wash. pp. 34-47, 1994

[16] Picard, R. W. and Minka, T. P., *Vision Texture for Annotation.*, Journal of Multimedia Systems, Vol. 3, pp. 3-14, 1995.

[17] Rothwell, C. A., Zisserman, A., Forsyth, D. A. and Mundy, J. L., *Planar Object Recognition Using Projective Shape Representation*, International Journal of Computer Vision, Vol. 16, pp. 57-99, 1995.

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

[19] Srihari, R. H., *Automatic Indexing of Content-based Retrieval of Captioned Images*, IEEE Computer, Vol. 28, No. 9, 1995.

[20] In proc. Storage and Retrieval for Image and Video Databases I, II, and III, Vol. 1,908; 2.185; and 2,420; W. Niblack and R. Jain, (eds.), SPIE, Bellingham, 1993, 1994 and 1995.

[21] Stein, F. and Medioni, G., *Structural Indexing: Efficient 2-D Object Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 1198-1204, 1992.

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

[23] Torrance, K. and Sparrow, E., *Theory for Off-specular Reflection from Roughened Surfaces*, J. Opt. Soc. Am., Vol. 57, 1967.

[24] Proceedings of Visual96 Information Systems: The First International Conference on Visual Information Systems, Melbourne, Victoria, Australia, 1996.

Figure 5: *Overview of the system.*