ADVANCES IN IMAGING AND ELECTRON PHYSICS, VOL. 151

Color Spaces and Image Segmentation

LAURENT BUSIN1, NICOLAS VANDENBROUCKE1,2, AND LUDOVIC MACAIRE1

1Laboratoire LAGIS UMR CNRS 8146 – Bâtiment P2, Université des Sciences et Technologies de Lille, 59655 Villeneuve d’Ascq, France
2Ecole d’Ingénieurs du Pas-de-Calais (EIPC), Campus de la Malaisse – BP39, 62967 Longuenesse Cedex, France

I. Introduction ........................................ 66
II. Color Spaces ........................................ 66
   A. Digital Color Image ............................... 67
      1. Human Color Vision ............................ 67
      2. Color Image Acquisition ....................... 71
      3. Color Image Visualization ..................... 74
   B. Color Spaces ...................................... 75
      1. Primary Spaces .................................. 76
      2. Luminance-Chrominance Spaces ............... 84
      3. Perceptual Spaces ............................... 92
      4. Independent Axis Spaces ....................... 100
      5. Hybrid Color Space ............................ 102
   C. Digital Color Images and Color Spaces .......... 104
      1. Color Space Coding for Image Analysis ....... 104
      2. Application to Color Spaces ................. 106
   D. Summary .......................................... 109
III. Color Image Segmentation ......................... 110
   A. Introduction ..................................... 110
   B. Edge Detection ................................... 111
      1. Overview ...................................... 111
      2. Edge Detection by the Analysis of a Color Gradient Vector .... 114
   C. Region Construction Based on a Spatial Analysis ........ 115
      1. Segmentation by Region Growing ............... 115
      2. Segmentation by Region Merging ............... 117
   D. Region Construction Based on a Color Space Analysis .......... 119
      1. Introduction ................................... 119
      2. Analysis of One-Dimensional Histograms .......... 123
      3. Analysis of the Three-Dimensional Color Histogram ....... 129
      4. Clustering-Based Segmentation ................. 133
      5. Spatial-Color Classification .................... 136
   E. Summary .......................................... 139
IV. Relationships between Segmentation and Color Spaces .............. 140
   A. Introduction ..................................... 140
   B. Edge Detection ................................... 141
I. INTRODUCTION

One of the most important problems in color image analysis is segmentation. This chapter considers color uniformity as a relevant criterion to partition an image into significant regions. For this purpose, the color of each pixel is initially represented in the \((R,G,B)\) color space where the color component levels of the corresponding pixel are namely red \((R)\), green \((G)\), and blue \((B)\).

Our goal is to describe the relationships between the color spaces and the color image segmentation. The different color spaces used to characterize the colors in the digital images are detailed in Section I. Section II is devoted to the segmentation methods designed for color image analysis. In Section III, we present several studies about the impact of the choice of the color space on the quality of results provided by segmentation methods.

II. COLOR SPACES

This section provides the details of how the colors of the pixels of digital color images can be represented in different color spaces for color image analysis applications. The first part of this section shows how color is quantized to produce a digital color image. The basics of human color vision are first summarized to facilitate understanding of this scheme. Then the problem of color image acquisition is discussed. In the second part, the most classical color spaces are regrouped into different families. In the framework of the use of these color spaces by color image analysis applications, the last part highlights the need to efficiently code the components of these color spaces to preserve their intrinsic properties.
A digital color image acquisition device is composed of different elements (Figure 1). First, the characteristics of the light and the material that constitute the observed object determine the physical properties of its color. Then, color images are acquired by an image sensor with an optical device and finally digitized thanks to a digital processor (Sharma and Trussell, 1997; Trussell, Saber and Vrhel, 2005; Vrhel, Saber and Trussell, 2005) before being displayed on a monitor or printed by a printer.

To understand how color is represented in digital images, it is first necessary to explain how it is perceived by a human observer. Thus, the basics of the color human vision are presented first. Then, the color image acquisition device is briefly described, and finally, the color image display device is presented.

1. Human Color Vision

The human perception of color is the response of the human receptor (the eye) and the human interpretation system (the brain) to a color stimulus—the reflection or the transmission of a light source by a material. In addition to the light and the material characterized by their physical properties, the eye and the brain, which are specific to each observer, provide physiologic and psychological features to the color sensation (Wyszecki and Stiles, 1982). This section presents these different properties of the color, which are very complex notions.

a. Physical Properties. First, there is no color without any light. The light, at the beginning of the color sensation, is defined as a electromagnetic radiation—a set of electromagnetic waves produced by the propagation of luminous particles, the photons. An electromagnetic radiation is characterized...
by its wavelength $\lambda$ expressed in meters (m). Visible light is the region of the
electromagnetic radiation emitted by the sun for which our eyes are sensitive
(between $\sim$ 380 and 780 nm). This light can be scattered in a spectrum of
luminous rays thanks to a prism. Each ray is composed of radiations with
the same wavelength, the monochromatic waves. Thus, each light source is
characterized by its spectral power distribution (SPD), the quantity of energy
per unit of wavelength (usually 1 nm). Certain sources corresponding with
classical observation conditions have been normalized by the Commission Inter-
ationale de l’Éclairage (CIE) under the name of illuminant (Commission
Internationale de l’Éclairage, 1986). CIE is an international standardization
organization devoted to establishing recommendations for all matters relating
to the science and art of lighting. Another characteristic of an illuminant is
its correlated color temperature—the temperature for which it would be nec-
essary to carry a black body radiator to obtain the visual impression nearest
to that produced by the corresponding source of light. Illuminants A (light
emitted by a black body radiator at the temperature of 2856 $^\circ$K or incandes-
cence lamp), C (direct light of the sun at midday), D (daylight), F (fluorescent
lamp), and E (light of equal energy) are the most widely used illuminants.

When the light source lights a material, the luminous rays emitted by
this source are reflected or transmitted by this material. Thus, the material
transforms the properties of the light to produce a color stimulus. When the
light comes into contact with a material, two phenomena occur: a surface
reflection of the incident radiations and a penetration of the incident light into
the material. When the light penetrates into the material, it meets the pigments.
Pigments are particles that determine the color of material while absorbing,
diffusing, or transmitting the light that reaches them. The pigments modify
the SPD of the light by selectively absorbing a part of the corresponding
electromagnetic waves. The light that is not absorbed by the pigments is
diffused or transmitted outside the material and carries the color information
of this material. According to its nature, a material can be characterized by its
capacity to reflect (reflectance) or to transmit (transmittance) incident light
energy. Conversely, it also can be characterized by its capacity to absorb
(absorption) the incident light energy. Finally, the color of a material depends
not only on its characteristics but also on the conditions of lighting and
viewing (position, orientation, distance, and so on).

b. Physiologic and Psychological Properties. After crossing several ele-
ments of the eye, the color stimulus reaches a photosensitive zone at the back
of the eye, the retina, where the images of the observed scene are projected.
The retina contains two kinds of photosensitive cells: cones and rods. The
rods allow night vision (scotopic vision) while the cones allow diurnal vision
(photopic vision). There are three kinds of cones: the S cones, which are
sensitive to short wavelengths close to blue; the M cones, which are sensitive
to medium wavelengths close to green; and the L cones, which are sensitive
to long wavelengths close to red.

The eye converts a color stimulus into a color signal at the entry of the optic
nerve. The optic nerve transmits this electric signal to the external (or lateral)
geniculate body, a relay charged to establish connections with fibers going
to the brain. A first analysis of the data is achieved there. According to the
opposite colors theory of Hering (1875), it seems that the color signal is coded
in an antagonistic fashion—by an achromatic signal (black-white opposition)
and by red-green and blue-yellow opposition signals (Hering, 1875). These
signals finally are transmitted to another area of the brain, the visual cortex,
where color interpretation is performed.

c. Trichromacy. Color human perception is characterized by its three-
dimensional (3D) aspect. The works of Young at the beginning of the
nineteenth century and Helmholtz in 1866 showed that any color stimulus
can be reproduced by the mixture of three other stimuli: the red, the green
and the blue stimuli, termed primaries or reference stimuli (Helmholtz, 1866;
Young, 1807). This principle is known as trichromacy, trichromatic theory,
or color mixture. Three primaries are therefore necessary and sufficient to
match any color by mixture; colorimetry, the science of color measurement,
is based on this theory. The amounts of each of the primaries necessary
to match a color are called the tristimulus values. There are two kinds of
color mixture: the additive and subtractive color mixtures (Figure 2). The

![Color mixture](image)

**Figure 2.** Color mixture. (a) Additive color mixture. (b) Subtractive color mixture. (See Color Insert.)
additive color mixture is the result of the juxtaposition of colored lights corresponding to each of the three primaries. The additive mixture in equal amount of the three primaries creates white. Additive color mixture is used for constituting the image of a television, a monitor, or the image acquired by a color video or still camera. The subtractive color mixture is the result of the selective absorption principle of the light by a material with respect to different wavelengths. Subtractive color mixture is used in printing or painting. The primaries used differ according to the type of color mixture. The primaries used by additive color mixture are the red, the green and the blue colors, whereas the magenta, the cyan, and the yellow colors are used by subtractive color mixture (the color black is often added as a primary in printing applications). The primaries of these two types of color mixture are complementary. The additive color mixture of two complementary colors creates the color white. On the basis of the trichromatic theory established by Young, in 1853 Grassman proposed the laws consigning the fundamental properties of the color mixtures that were supplemented by Abney in 1913 (Abney, 1913; Grassman, 1853). These laws make it possible to apply the additive, associative, multiplicative, and transitive properties of the algebraic equalities to the colorimetric equalities. Today Grassman's laws are the mathematical basis of colorimetry.

d. Perceptual Attributes. The human perception of color is a subjective reaction to the stimulation of the eye, and it seems to be more adapted to characterize a color in terms of brightness, hue, and saturation.

Brightness is the attribute of a visual sensation according to which an area appears to emit more or less light. Thus, it corresponds to a feeling in terms of light (like dark or luminous) and characterizes the luminous level of a color stimulus. The terms brightness, lightness, luminance, luminosity, radiant intensity, luminous intensity, illumination, luma, and so on are often used in the literature to indicate the concept of luminosity. Poynton (1993) highlights the confusion between these terms.

Hue corresponds to the main primary colors: red, green, blue, and yellow. It is the attribute of a visual sensation according to which an area appears to be similar to one of these perceived colors. It corresponds to the dominant wavelength of a SPD—the wavelength with the highest energy. Hue can be defined by an angle termed hue angle. Since black, white, or gray are colors without any hue, they are called neutral colors or achromatic colors.

Saturation is an attribute that allows the colorfulness of an area to be estimated with respect to its brightness. The chroma is a brightness-dependent attribute that also estimates the colorfulness of the hue. Saturation represents the purity of a perceived color, such as bright, pale, dusty, and so on. It
COLOR SPACES AND IMAGE SEGMENTATION

2. Color Image Acquisition

The previous section explained that, in the framework of human vision, the eye and the brain are the receptor and the interpretation systems of the color stimuli, respectively. In the framework of computer vision, the receptor is a color camera and the interpretation system is a computer (see Figure 1). This section describes how a color image can be acquired by a color camera and be digitized by an image acquisition device.

a. Color Camera. The main element of a camera is the image sensor (Trussell, Saber and Vrhel, 2005; Vrhel, Saber and Trussell, 2005). The image sensor is composed of a set of photosensitive elements that convert the luminous flux (photon) into an electric information (electron) to provide one or several video signals to the digitization system. Each photosensitive receptor produces an increasing voltage in function of the received light intensity. They can be arranged either as a row (line scan camera) or as an array (area array camera). The obtained image is made up of a set of points called pixels,1 which correspond to the photosensitive receptor.

Two types of technology exist to make a sensor array: charged coupled device (CCD) technology and complimentary metal-oxide–semiconductor (CMOS) technology. The CCD sensors provide higher signal-to-noise ratio (SNR) levels, whereas the CMOS sensors have a lower SNR but allow the integration of other components into each photosensitive receptor of the sensor.

By analogy with the human vision system, color information is created by means of three color filters. Each is sensitive to the small (red), the medium (blue), or the large (green) wavelengths.

Two kinds of color camera exist (Figure 3):

- One-chip color cameras are designed with one single-sensor array and three interlaced color filters, termed a color filter array. A Bayer color filter array is the most widely used device. Thus, the photosensitive receptors that are laid sequentially on a same line of the sensor are fitted with red, green, and blue filters in succession. The color information is obtained by receptors located at different places. This technology generates a loss of resolution and chromatic aberration (false colors). Therefore, it requires interpolation techniques (termed demosaicking) so that the color of each pixel of the image is defined by three components (Gunturk et al., 2005). Other

1 *PICture Elemen*t = pixel.
technologies, based on only one sensor array, resolve these problems (Lyon and Hubel, 2002).

• The three-chip color cameras are fitted out with three sensor arrays associated with an optical system based on prisms. Each sensor receives the red, green, and blue stimulus, respectively, via dichroic filters fixed on prisms. The color of a pixel is provided by the response of three receptors and therefore there is not loss of resolution. Demosaicking is not necessary, and the quality of the image is better than the quality of an image acquired by a one-chip color camera. However, the three-chip technology can generate the phenomenon of shading, which creates a color scale on the image of a white background when luminous rays reaching the prisms are not perfectly parallel.
Color cameras are fitted with infrared filters because sensors are sensitive to wavelengths located over the visible domain. The spectral response of an image sensor is not the same as that of the spectral response of the human eye.

b. Color Image Digitization. A digitization system achieves the conversion (amplification, sampling, and so on) of one or more video signals derived from an image sensor into a digital signal where each element of a color image is represented by a triplet of digital values that can be stored in the computer memory for further processing and display application. This stage of acquisition is called color quantization.

Color quantization consists of associating, for each pixel of a color image, three digital values—a red level ($R$), a green level ($G$), and a blue level ($B$)—which correspond to the tristimulus values of this color. Thus, a digital color image is an array of pixels that are located by their spatial coordinates in the image plane and whose color is defined by the three components $R$, $G$, and $B$. Generally, each of these three components is coded on 8 bits and can take 256 different unsigned integer values belonging to the interval $[0, 255]$. A color is so coded with $3 \times 8 = 24$ bits and it is then possible to represent $2^{24}$ (16,777,216) colors by additive mixture, whereas the human visual system allows differentiation of approximatively 350,000 colors. However, even though this quantization represents more colors than the human observer can see, the human visual system is not uniform; for specific intervals, it is necessary to code the color component on 12 bits to discriminate all the colors that the human visual system can perceive on this interval.

Digital values, which are provided by the digitization system, depend on the characteristics of the device used, the choice of lighting system, and the setup of the acquisition system. Indeed, the SPD of a light depends on the lighting system, and the spectral responses of an image sensor depend on the chosen color camera. Thus, the $R$, $G$, and $B$ components of a given color are not the same with different equipments and parameter setups; therefore, the color components of an acquired image are device dependent (or rendered) (Trussell, Saber and Vrhel, 2005).

To ensure that the $R$, $G$, and $B$ color components are invariant to acquisition system, it is necessary to apply a calibration step. Calibration is achieved by using a chart (Macbeth Color Checker, IT8 chart, white target) with different colors whose theoretical tristimulus values are known a priori. The image of this chart is acquired by the acquisition system and the parameters of this system are adjusted so that the $R$, $G$, and $B$ components of each color on the chart in the image are the same as their theoretical values. This adjustment can be achieved via simple white balance or by more precise numerical methods (Bala and Sharma, 2005; Ramanath et al., 2005).
3. Color Image Visualization

In most applications, color images are designed to be displayed on a screen or printed on paper. For color image analysis applications, it is not always necessary to display the color when the result of the analysis is not an image but instead some data. However, many color spaces have been developed to display color images on televisions or monitors and depend on their technologies. In order to understand these color spaces, which are presented later, we describe how the colors are displayed.

a. Display. The display of acquired color image is usually achieved by a cathode ray tube (CRT) monitor or by a liquid crystal display (LCD) flat panel monitor (Vrhel, Saber and Trussell, 2005).

The screen of the CRT monitor consists of a thin layer of a mosaic of red, green, and blue phosphor cells. The CRT is to bombard the screen with electron beams that stimulate each of the three types of phosphor with varying intensities. Three colored lights (whose spectral characteristics depend on the chemical nature of the phosphor) are produced and display a color on the screen according to the additive mixture principle. For an LCD flat panel monitor, color display is produced by a mosaic of red, green, and blue color filters that are backlit by a fluorescent light source. These backlight and color filters were designed to produce chromaticities identical to those of the CRT phosphors to ensure compatibility between the two devices.

To achieve the display of a color image on a screen, the digital values of the color components of each pixel of the image are loaded into the memory of the video card. They then are converted to analog data to specify the intensity of the beam (the amount of light that reaches the screen). This intensity, denoted \( x \), is a nonlinear function, denoted \( \Gamma \), of the voltage generated by the video card. For a CRT monitor, it follows a power law defined by \( \Gamma(x) = x^\gamma \), where the value of \( \gamma \) generally ranges between 2 and 3. For an LCD monitor, the response of an LCD pixel cell is quite different from that of a CRT and tends to follow a sigmoidal law. To correct this nonlinearity, an inverse law must be applied when the image is displayed. This operation is known as gamma correction.

b. Gamma Correction. Generally, the acquisition system performs the gamma correction, which consists of compensating for the nonlinearity of display devices by transforming the acquired video signals so that the \( R \), \( G \), and \( B \) trichromatic components of the pixels of the acquired image are corrected.

For example, a television channel transmits the signals that can be displayed on a CRT TV screen and are gamma corrected. This correction depends on
the technology used, which defines the primaries necessary to achieve the additive mixture of the colors of the image. More precisely, this technology depends on the country. Thus, North American television receivers satisfy the National Television Standards Committee (NTSC) old standard, which uses the primaries defined by the Federal Communications Commission (FCC), or the Society of Motion Picture and Television Engineers (SMPTE-C) new standard. For the NTSC standard, $\gamma$ is set to 2.2. European television receivers satisfy the phase alternation by line (PAL) German standard, which uses the primaries defined by the European Broadcasting Union (EBU) or can satisfy the SÉquentiel Couleur À Mémoire (SECAM) French standard. For the PAL standard, $\gamma$ is set to 2.8.

With most color cameras, it is possible to adjust the gamma correction directly on the red, green, and blue acquired signals. This camera setup allows correction of the display problems only on monitors and so, it is not necessary to use this setup for color image analysis application. On the other hand, to render color applications or for color measure applications on a screen, it is necessary to apply gamma correction on the displayed color image. When color images whose acquisition conditions are not specified (web images) are considered, it is very difficult to know whether these images are gamma corrected.

Poynton (1996) noticed an amazing coincidence between the gamma correction law and the transfer function of the human eye. Indeed, the human eye response to a stimulus is nonlinear but corresponds to a logarithmic law very close to the gamma correction law. Thus the application of a gamma correction on the acquired color component produces a color representation close to the human color perception. According to Poynton (1993), even if the gamma correction is not necessary for technological reasons, it is useful for perceptual reasons.

B. Color Spaces

A color space is a geometrical representation of colors in a space and allows specification of colors by means of (generally) three components whose numerical values define a specific color.

The first proposed color spaces to specify colors were based on color human perception experiments designed to develop classifications of colors. The ordered color systems are collections of color samples (color charts, color atlas). These spaces are generally 3D spaces where the units of the components are merely conventional and help to specify colors within this representation. These spaces (e.g., Munsell system, Natural Color System [NCS], Optical Society of America [OSA] system, or Deutsches Institut für Normung [DIN] system) are rarely used for color image analysis applications.
The most frequently used spaces propose a metric to measure the distance between colors. The goal of this section of the text is to present the most frequently used classical color spaces in the framework of color image analysis. According to their characteristics, they are classified into four families:

1. The primary spaces are based on the trichromatic theory, assuming that it is possible to match any color by mixing appropriate amounts of three primary colors.

2. The luminance-chrominance spaces where one component represents the luminosity and the two others the chromaticity.

3. The perceptual spaces attempt to quantify the subjective human color perception by using the intensity, the hue, and the saturation components.

4. The independent axis spaces resulting from different statistical methods, which provide the least correlated components as possible.

Since many color spaces for different applications and several definitions for the same color space exist, it is difficult to select only one for a color image analysis application; therefore it is important to provide an overview of their specificities and their differences.

1. Primary Spaces

The primary spaces are based on the trichromatic theory, which assumes that we are able to match any color by mixing appropriate amounts of three primary colors. By analogy to physiology of the human eye, the primary colors are close to red ($R$), green ($G$), and blue ($B$). Therefore, a primary space depends on the choice of a set of primary colors denoted here $[R]$, $[G]$, and $[B]$. Ratios fixed for each primary are defined in order to reproduce a reference white (or white point), denoted $[W]$, by the additive mixing of the same amount of each primary. The choice of the primary depends on the device used, and the reference white is generally a CIE illuminant. The next sections of the text present the most useful primary spaces. They can be separated into the real primary spaces, for which the primary colors are physically realizable, and the imaginary primary spaces, whose primaries do not physically exist.

a. Real Primary Spaces. The CIE ($R$, $G$, $B$) color space is derived from color-matching experiments led in 1931 by Wright and Guild that used the primaries, denoted here $[RC]$, $[GC]$ and $[BC]$ (where $C$ indicates CIE), respectively, to match all the available colors of the visible spectrum. These primaries are red, green, and blue monochromatic color stimuli with wavelengths of 700.0, 546.1, and 435.8 nm, respectively, and the reference white
is the equal-energy illuminant E (Commission Internationale de l’Éclairage, 1986). The CIE \((R, G, B)\) color space can be considered as the reference \((R, G, B)\) color space because it defines a standard observer whose eye spectral response approximates the average eye spectral response of a representative set of human observers.

For each of the three \([R_\text{C}], [G_\text{C}], [B_\text{C}]\) primaries there correspond three normalized vectors \(\overrightarrow{R_\text{C}}, \overrightarrow{G_\text{C}}, \overrightarrow{B_\text{C}}\), respectively, that constitute the reference of a vectorial space whose original point is denoted \(O\). In this space, each color stimulus \([C]\) is represented by a point \(C\) that defines the color vector \(\overrightarrow{OC}\). The coordinates of this vector are the tristimulus values \(R_\text{C}, G_\text{C},\) and \(B_\text{C}\).

The coordinates of specific color vectors are negative because they correspond to color stimuli that are not reproducible (matched) by additive mixture. Points that correspond to color stimuli with positive tristimulus values are inside a cube, named the RGB color cube (Figure 4). The original point \(O\) corresponds to the black color \((R_\text{C} = G_\text{C} = B_\text{C} = 0)\), whereas the reference white is defined by the additive mixture of equal quantities of the three primaries \((R_\text{C} = G_\text{C} = B_\text{C} = 1)\).
The straight line joining the points Black and White in Figure 4 is called the 
gray axis, the neutral color axis, or the achromatic axis. Indeed, the points of
this line represent gray nuances from black to white.

The tristimulus values of a color stimulus depend on the color’s luminance.
Two different color stimuli can be described by the same chromatic character-
stics (here called chrominance), but their tristimulus values can be different
due to their luminance. In order to obtain color components that do not depend
on the luminance, it is necessary to normalize their values. For this purpose, it
is possible to divide each color component value by the sum of the three ones.
The three thus-obtained color components, called chromaticity coordinates
or normalized coordinates, are denoted \( r_C \), \( g_C \), and \( b_C \), respectively, and are
defined by:

\[
\begin{align*}
    r_C &= \frac{R_C}{R_C + G_C + B_C}, \\
    g_C &= \frac{G_C}{R_C + G_C + B_C}, \\
    b_C &= \frac{B_C}{R_C + G_C + B_C}.
\end{align*}
\]

The transformation defined by Eq. (1) corresponds to the projection of the \( C \)
point on the plane that is normal to the achromatic axis. This plane is defined
by the equation: \( R_C + G_C + B_C = 1 \), and the intersection between this plane
and the RGB color cube constitutes an equilateral triangle whose summits are
the three primaries \([R_C]\), \([G_C]\), and \([B_C]\). This triangle, termed the Maxwell
triangle, is represented by a dotted line in Figure 4.

The color space that can be associated with the chromaticity coordinates is
called the normalized \((R_C, G_C, B_C)\) color space and is denoted \((r_C, g_C, b_C)\).
As \( r_C + g_C + b_C = 1 \), two components are sufficient to represent the
chrominance of a color stimulus. Thus, Wright and Guild have proposed a
diagram called the chromaticity diagram \((r_C, g_C)\). Figure 5 represents this
diagram, which contains a curve (the spectrum locus), joining the points
corresponding to monochromatic color stimuli whose wavelengths range
between 380 and 780 nm. The two extremities of this curve are connected
by a straight line known as the nonspectral line of purples. All the colors of
the visible spectrum are contained in the area defined by the spectrum locus
and the purple line. In this figure, the Maxwell triangle does not contain all
the colors because several colors cannot be matched by additive mixture and
are defined by negative chromaticity coordinates.

As mentioned previously, the CIE \((R, G, B)\) color space can be consid-
ered as the reference \((R, G, B)\) color space because it defines the standard
observer. In many application fields, however, it is not possible to use this
color space and other primary colors must be used. These other \((R, G, B)\)
color spaces differ by the primaries used to match colors by additive mix-
ture and by the used reference white. Thus, they are device dependent
COLOR SPACES AND IMAGE SEGMENTATION

Color spaces (Süssstrunk, Buckley and Swen, 1999). The most widely known \( (R, G, B) \) color spaces used by the following application fields are as follows:

- **Video and image acquisition.** In color image analysis applications, the images can be acquired by analog or digital video camera, by digital still camera, by scanner, and so on. The color of each pixel of an image is defined by three numerical components that depend on the acquisition device and its setup to obtain a reference white. Thus, an image acquired with the same lighting and observation conditions by two different cameras produces different colors if the primaries’ colors associated with the two cameras are not the same. In this chapter, the \( (R, G, B) \) color space used for acquiring images is called the *image acquisition space*.

- **Television image display.** For TV analog signal transmission, it is assumed that the chromaticity coordinates of the primaries are different from those of
the CIE primaries because they correspond to phosphors or color filters with
different wavelengths. The different TV signal transmission standards use
\((R, G, B)\) color spaces that are different. For example, the NTSC standard
uses the \((R_F, G_F, B_F)\) color space based on the FCC primaries; the PAL
and SECAM standards use the \((R_E, G_E, B_E)\) color space based on the
EBU primaries, and SMPTE-C standard uses the \((R_S, G_S, B_S)\) color space
based on the SMPTE primaries. In the same way, the reference white to
reproduce depends on the display technology and its setup. For example,
we generally assume that the reference white used by the NTSC standard is
the C illuminant, whereas the D65 illuminant is used by the PAL/SECAM
or SMPTE-C standard.
• Computer graphics. Primaries and reference white used by the monitors
of computers are defined by their salesmen and differ from those used in
television or defined by CIE. In the same way, when an image is generated
by computer, the color of the pixel depends on the \((R, G, B)\) color space
that uses the software for image coding.
• Image printing. Printing systems usually use color spaces based on a
subtractive color mixture. The primaries used are cyan, magenta, and
yellow, and the primary color space is the \((C, M, Y)\) color space that can
be derived from an \((R, G, B)\) color space by the relation:

\[
\begin{bmatrix}
C \\
M \\
Y
\end{bmatrix} = 1 - \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}.
\]

The reference white used in printing systems is often the D50 illuminant.

This multitude of primary color spaces leads to many errors in color
management. It is necessary to dispose of an \((R, G, B)\) color space well suited
to CRT and LCD monitors, TVs, scanners, cameras, and printers to ensure
compatibility between these different devices and so that the colors are ren-
dered identically (Bala and Sharma, 2005; Ramanath et al., 2005). This is why
the sRGB color space has been proposed as a standard default \((R, G, B)\) color
space that characterizes an average monitor (IEC 61966-2-1/FDIS, 1999).
This color space can be considered an independent device (or unrendered)
color space for color management applications. The reference white and the
primaries are defined by the International Telecommunication Union (ITU)
standard according to the ITU-R BT.709 recommendation (ITU-R BT.709-5,
2002). Thus, the sRGB color space is denoted here as \((R_I, G_I, B_I)\).

Regardless of which primary space is used, it is possible to achieve a
primary conversion as a simple linear transformation of the component values
by means of a matrix equation. So there exists a transformation matrix \(P\) that
For each primary space, it is possible to define a RGB color cube and chromaticity coordinates.

b. Imaginary Primary Space. The \((R, G, B)\) color spaces present some major drawbacks:

- Because it is not possible to match all the colors by additive mixture with a real primary space, the tristimulus values and chromaticity coordinates can be negative.
- The tristimulus values depend on the luminance, which is a linear transformation of the \(R, G, B\) color components.
- Because \((R, G, B)\) color spaces are device dependent, there is a multitude of \((R, G, B)\) color spaces with different characteristics.

The CIE therefore defines the imaginary primary space, named the CIE \((X, Y, Z)\) color space, where the primary colors are imaginary (virtual or artificial) in order to overcome the problems of the primary spaces. In this space, the \([X]\), \([Y]\), and \([Z]\) primaries are not physically realizable, but they have been defined so that all the color stimuli are expressed by positive tristimulus values and so that one of these primaries, the \([Y]\) primary, represents the luminance component. Because all the \((R, G, B)\) color spaces can be converted to the \((X, Y, Z)\) color spaces by linear transforms, this space is an independent device color space. It defines an ideal observer, the CIE 1931 standard colorimetric observer, and all the colorimetry applications are based on it.

Chromaticity coordinates can be deduced from the \((X, Y, Z)\) color space to obtain a normalized \((X, Y, Z)\) color space denoted \((x, y, z)\). Thus, the chromaticity coordinates \(x\), \(y\), and \(z\) are derived from the \(X\), \(Y\), and \(Z\) tristimulus values by:

\[
\begin{align*}
x &= \frac{X}{X + Y + Z}, \\
y &= \frac{Y}{X + Y + Z}, \\
z &= \frac{Z}{X + Y + Z}.
\end{align*}
\]

As \(x + y + z = 1\), \(z\) can be deduced from \(x\) and \(y\). Thus, the colors can be represented in a plane called the \((x, y)\) chromaticity diagram (Figure 6). In this diagram, all the colors are inside the area delimited by the spectrum locus and the purple line. Since the chromaticity coordinates of these colors are positive, the CIE primaries \([X]\), \([Y]\), and \([Z]\) allow matching of any color by additive color mixture.
FIGURE 6. (x, y) Chromaticity diagram.

The conversion from an (R, G, B) color space to the (X, Y, Z) color space is defined by the following equation:

\[
\begin{bmatrix}
X \\
Y \\
Z \\
\end{bmatrix} = \mathbf{P} \times \begin{bmatrix}
R \\
G \\
B \\
\end{bmatrix}, \quad \text{with } \mathbf{P} = \begin{bmatrix}
X_R & X_G & X_B \\
Y_R & Y_G & Y_B \\
Z_R & Z_G & Z_B \\
\end{bmatrix}.
\]  \hspace{1cm} (4)

The coefficients of the \( \mathbf{P} \) matrix are defined with respect to the \([R], [G], \) and \([B] \) primaries and the reference white used to reproduce the equal mixing of the three \([X], [Y], [Z] \) primaries.

Generally, the primaries and the reference white are characterized by their \( x \) and \( y \) chromaticity coordinates. Table 1 lists the \( x \) and \( y \) chromaticity coordinates of the primaries used by different \((R, G, B)\) color spaces, and Table 2 lists those of the A, C, D65, D50, F2, and E illuminants.
COLOR SPACES AND IMAGE SEGMENTATION

TABLE 1

<table>
<thead>
<tr>
<th>Standard</th>
<th>Primaries</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIE</td>
<td>[RC]</td>
<td>0.735</td>
<td>0.265</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>[GC]</td>
<td>0.274</td>
<td>0.717</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>[BC]</td>
<td>0.167</td>
<td>0.009</td>
<td>0.824</td>
</tr>
<tr>
<td>FCC</td>
<td>[RF]</td>
<td>0.670</td>
<td>0.330</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>[GF]</td>
<td>0.210</td>
<td>0.710</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>[BF]</td>
<td>0.140</td>
<td>0.080</td>
<td>0.780</td>
</tr>
<tr>
<td>SMPTE</td>
<td>[RS]</td>
<td>0.630</td>
<td>0.340</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>[GS]</td>
<td>0.310</td>
<td>0.595</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>[BS]</td>
<td>0.155</td>
<td>0.070</td>
<td>0.775</td>
</tr>
<tr>
<td>EBU</td>
<td>[RE]</td>
<td>0.640</td>
<td>0.330</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>[GE]</td>
<td>0.290</td>
<td>0.600</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>[BE]</td>
<td>0.150</td>
<td>0.060</td>
<td>0.790</td>
</tr>
<tr>
<td>ITU</td>
<td>[RI]</td>
<td>0.640</td>
<td>0.330</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>[GI]</td>
<td>0.300</td>
<td>0.600</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>[BI]</td>
<td>0.150</td>
<td>0.060</td>
<td>0.790</td>
</tr>
</tbody>
</table>

TABLE 2

<table>
<thead>
<tr>
<th>Illuminant</th>
<th>Tp (°K)</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2856</td>
<td>0.448</td>
<td>0.407</td>
<td>0.145</td>
</tr>
<tr>
<td>C</td>
<td>6774</td>
<td>0.310</td>
<td>0.316</td>
<td>0.374</td>
</tr>
<tr>
<td>D50</td>
<td>5000</td>
<td>0.346</td>
<td>0.358</td>
<td>0.296</td>
</tr>
<tr>
<td>D65</td>
<td>6504</td>
<td>0.313</td>
<td>0.329</td>
<td>0.358</td>
</tr>
<tr>
<td>E</td>
<td>5400</td>
<td>0.333</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>F2</td>
<td>4200</td>
<td>0.372</td>
<td>0.375</td>
<td>0.253</td>
</tr>
</tbody>
</table>

A set of primaries defines a triangle in the (x, y) chromaticity diagram whose summits are the chromaticity coordinates of these primaries. Examples of this triangle and white points have shown in the (x, y) chromaticity diagram in Figure 7. This figure shows that none of these triangles contains all the visible colors; this means that all the colors cannot be matched by additive mixture with an (R, G, B) color space. Each (R, G, B) color space defines a set of matched colors called a gamut.
Table 3 provides examples of P transformation matrices used to convert (R, G, B) color spaces to the (X, Y, Z) color space (Pascale, 2003).

2. Luminance-Chrominance Spaces

Because a color can be expressed by its luminosity and chromaticity, several color spaces separate the luminance from the chrominance information. These color spaces can be categorized in the family of luminance-chrominance spaces.

The components of a luminance-chrominance space are derived from the component of an (R, G, B) color space by linear or nonlinear transformations. The type of transformation depends on the type of luminance-chrominance spaces that can be classified in the following spaces:
TABLE 3
(R, G, B) TO (X, Y, Z) COLOR SPACE CONVERSION (Pascale, 2003)

<table>
<thead>
<tr>
<th>Primary system</th>
<th>Color space</th>
<th>Reference white</th>
<th>P transformation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIE (R_C, G_C, B_C)</td>
<td>E</td>
<td>0.489 0.311 0.200</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.176 0.813 0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000 0.010 0.990</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.607 0.174 0.200</td>
<td></td>
</tr>
<tr>
<td>FCC (R_F, G_F, B_F)</td>
<td>C</td>
<td>0.299 0.587 0.114</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000 0.066 1.116</td>
<td></td>
</tr>
<tr>
<td>SMPT (R_S, G_S, B_S)</td>
<td>D65</td>
<td>0.394 0.365 0.192</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.212 0.701 0.087</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.019 0.112 0.958</td>
<td></td>
</tr>
<tr>
<td>EBU (R_E, G_E, B_E)</td>
<td>D65</td>
<td>0.431 0.342 0.178</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.222 0.707 0.071</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.020 0.130 0.939</td>
<td></td>
</tr>
<tr>
<td>ITU (R_I, G_I, B_I)</td>
<td>D65</td>
<td>0.412 0.358 0.180</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.213 0.715 0.072</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.019 0.119 0.950</td>
<td></td>
</tr>
</tbody>
</table>

- The perceptually uniform spaces, which propose a metric to establish a correspondence between a color difference perceived by a human observer and a distance measured in the color space.
- The television spaces, which separate luminosity signal to chromaticity signals for the television signal transmission.
- The antagonist (or opponent color) spaces, which aim to reproduce the model of the opponent color theory proposed by Hering.

Let \( L \) be the luminance component of a luminance-chrominance space. In function of the specified luminance-chrominance space, the luminance component can represent the lightness, brightness, luminous intensity, luma, or luminance. The term luminance used in this chapter is a general terminology of one of these magnitudes. Let \( Chr_1 \) and \( Chr_2 \) be the two chrominance components of a luminance-chrominance space. A luminance-chrominance space is there denoted as \((L, Chr_1, Chr_2)\).

a. Perceptually Uniform Spaces. Nonuniformity is one drawback of the \((X, Y, Z)\) color space. Indeed, in the \((x, y)\) chromaticity diagram, an equal distance between two color points does not correspond to the same color difference perceived by a human observer according to the considered area of the diagram. Thus there are areas in the \((x, y)\) chromaticity diagram for which color differences are not perceptible by a human observer. The size and orientation of these regions, known as the MacAdam ellipses, depend on their...
positions in the diagram (MacAdam, 1985). This nonuniformity is a problem for applications where it is necessary to measure color difference. Indeed, colors that are perceptually close can be separated by greater distances, whereas colors that are perceptually different can be close in the \((x, y)\) chromaticity diagram.

Because the Euclidean distances evaluated in the \((R, G, B)\) or \((X, Y, Z)\) color spaces do not correspond to the color differences that are actually perceived by a human observer, the CIE recommends two perceptually uniform spaces—the \((L^*, u^*, v^*)\) and \((L^*, a^*, b^*)\) color spaces—where \(L^*\) represents the lightness (luminance component) and where \(u^*\), \(v^*\) and \(a^*, b^*\) are chromaticity coordinates (chrominance components) (Commission Internationale de l’Éclairage, 1986).

The first perceptually uniform color space proposed by the CIE in 1960 was the \((U, V, W)\) color space, which is derived from the \((X, Y, Z)\) color space. In this color space, \(V\) is a luminance component and it is possible to define a uniform chromaticity diagram (called the 1960 Uniform Chromaticity Scale [UCS] diagram or \((u, v)\) chromaticity diagram). Because the \((U, V, W)\) color space does not satisfy the nonuniformity problem, the CIE proposed the \((U^*, V^*, W^*)\) in 1964 (Pratt, 1978, 1991). This color space finally underwent significant modification in 1976 to become the \((L^*, u^*, v^*)\) color space (also named the CIELUV color space). The luminance component of the \((L^*, u^*, v^*)\) color space is expressed as:

\[
L^* = \begin{cases} 
116 \times \sqrt[3]{\frac{Y}{Y_W}} - 16 & \text{if} \quad \frac{Y}{Y_W} > 0.008856, \\
903.3 \times \frac{Y}{Y_W} & \text{if} \quad \frac{Y}{Y_W} \leq 0.008856,
\end{cases}
\]

where \(X^W, Y^W,\) and \(Z^W\) are the tristimulus values of the reference white.

The chrominance components are

\[
u^* = 13 \times L^* \times (u' - u'W),
\]

\[
v^* = 13 \times L^* \times (v' - v'W),
\]

with

\[
u' = \frac{4X}{X + 15Y + 3Z},
\]

\[
v' = \frac{9Y}{X + 15Y + 3Z},
\]

where \(u'W\) and \(v'W\) are the chrominance components of \(u'\) and \(v'\) for the reference white, respectively.

Figure 8 shows the CIE 1976 \((u', v')\) chromaticity diagram derived from the \((L^*, u^*, v^*)\) color space.
COLOR SPACES AND IMAGE SEGMENTATION

The \((L^*, a^*, b^*)\) color space proposed by the CIE (also named the CIELAB color space) is derived from the \((X, Y, Z)\) color space by nonlinear relations. The luminance component is defined by Eq. (5) and the chrominance components are expressed as

\[
a^* = 500 \times \left( f\left(\frac{X}{X_W}\right) - f\left(\frac{Y}{Y_W}\right) \right),
\]

\[
b^* = 200 \times \left( f\left(\frac{Y}{Y_W}\right) - f\left(\frac{Z}{Z_W}\right) \right),
\]

with

\[
f(x) = \begin{cases} 
\sqrt{x} & \text{if } x > 0.008856, \\
7.787x + \frac{16}{116} & \text{if } x \leq 0.008856.
\end{cases}
\]
In these two last uniform color spaces, the luminance component $L$ corresponds to the lightness (or brightness) and represents the human eye response to a level of luminance. The CIE models this nonlinear response by a cubic root relation. The first chrominance component ($a^*$ or $u^*$) of these two color spaces corresponds to a green-red color opposition, whereas the second chrominance component ($b^*$ or $v^*$) corresponds to a blue-yellow color opposition.

**b. Television Spaces.** The transmission of TV signals requires the separation between the luminance and chrominance components. This separation can be achieved by a linear transformation of the component values of an $(R, G, B)$ color space. The luminance component corresponds to the $Y$ color component of the $(X, Y, Z)$ color space. The $Chr_1$ and $Chr_2$ chrominance components are as follows:

\[
\begin{align*}
Chr_1 &= a_1 (R - Y) + b_1 (B - Y), \\
Chr_2 &= a_2 (R - Y) + b_2 (B - Y).
\end{align*}
\]  

(13)

The coefficients $a_1$, $b_1$, $a_2$, and $b_2$ are specific to the norms used, standards, or commissions (e.g., NTSC, PAL, SECAM, SMPTE). Because the $Y$ color component is evaluated by means of a linear transformation of the $R$, $G$, and $B$ components, it is possible to express the conversion of an $(R, G, B)$ color space [or $(X, Y, Z)$ color space] to $(Y, Chr_1, Chr_2)$ color space by using a transformation matrix $Q$:

\[
\begin{bmatrix}
Y \\
Chr_1 \\
Chr_2
\end{bmatrix} = Q \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} = Q \times P \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.
\]  

(14)

Notice that the signal transmitted by TV channels are gamma corrected for different reasons (display, bandwidth, human eye response and so forth). Therefore, the conversion of an $(R, G, B)$ color space to a television space is applied to gamma-corrected $R$, $G$, and $B$ components, denoted $R'$, $G'$, and $B'$. The use of the prime (') notation is extended to all color components that are gamma corrected.

The main television spaces are as follows:

- $(Y', I', Q')$ color space used by the NTSC television standard, where $Y$ is the luminance component and $I$ and $Q$ are the chrominance components. The $Q$ transformation matrix (listed in Table 3) is used to define the $Y$ color component as:

\[
Y' = 0.299 \times R' + 0.587 \times G' + 0.114 \times B'.
\]  

(15)

The chrominance components are defined by Eq. (13) with $a_1 = 0.74$, $b_1 = -0.27$, $a_2 = 0.48$ and $b_2 = 0.41$. 

\[
\begin{align*}
Chr_1 &= a_1 (R - Y) + b_1 (B - Y), \\
Chr_2 &= a_2 (R - Y) + b_2 (B - Y).
\end{align*}
\]  

(13)
• \((Y', U', V')\) color space used by the EBU broadcasting standard where the chrominance components are \(U\) and \(V\). To ensure compatibility among the different TV standards, the luminance component of the EBU standard is also defined by Eq. (15). The chrominance components are defined by Eq. (13) with \(a_1 = 0, b_1 = 0.493, a_2 = 0.877\) and \(b_2 = 0\).

• \((Y', D_r', D_b')\) color space used by the SECAM color TV broadcasting standard, where the chrominance components are \(D_r\) and \(D_b\). The luminance component is also defined by Eq. (15), and the chrominance components are defined by Eq. (13) with \(a_1 = -1.9, b_1 = 0, a_2 = 0\) and \(b_2 = 1.5\).

• \((Y', C'_b, C'_r)\) color space is an international standard for digital image and video coding. ITU proposes two recommendations:
  o The ITU-R BT.601 recommendation for digital coding of standard-definition television (SDTV) signals. According to this recommendation, the \((Y', C'_b, C'_r)\) color space is independent of the primaries and the reference white (ITU-R BT.601-7, 2007) and it is also used by video and image-compression schemes such as MPEG and JPEG. The luminance component is also defined by Eq. (15), and the chrominance components are defined by Eq. (13) with \(a_1 = 0, b_1 = 0.564, a_2 = 0.713\) and \(b_2 = 0\).
  o The ITU-R BT.709 recommendation for digital coding of high-definition television (HDTV) signals (ITU-R BT.709-5, 2002) with \(a_1 = 0, b_1 = 0.534, a_2 = 0.635\) and \(b_2 = 0\).

Table 4 lists some transformation matrices to convert a primary space to one of the main TV spaces.

c. Antagonist Spaces. The antagonist spaces are based on the Hering color opponent theory in order to model the human visual system. According to this theory, the color information acquired by the eye is transmitted to the brain in three components: an achromatic component \(A\) and two chrominance components \(C_1\) and \(C_2\). The \(A\) color component integrates the signals derived from the three types of cones of the human retina and represents a black-white opposition signal, whereas the \(C_1\) and \(C_2\) components integrate only the signals provided by different types of cones and correspond to green-red and yellow-blue oppositions, respectively. The \((L^*, u^*, v^*)\) and \((L^*, a^*, b^*)\) color spaces previously presented also can be considered as antagonist spaces because they share these same properties. Different antagonist spaces have been proposed for color image analysis.

In 1976, Faugeras proposed a human visual system model in which \(A\), \(C_1\), and \(C_2\) color components are evaluated from three primaries, denoted \([L], [M],\) and \([S]\), which correspond to the three types of cones of the human retina (Faugeras, 1979). He proposed transformation matrices to convert an \((R, G, B)\) color space to the \((L, M, S)\) color space. For example,
the conversion of an \((R, G, B)\) color space used by a CRT monitor to the \((L, M, S)\) color space with the C illuminant as reference white can be achieved by the following transformation matrix:

\[
P = \begin{bmatrix}
0.3634 & 0.6102 & 0.0264 \\
0.1246 & 0.8138 & 0.0616 \\
0.0009 & 0.0602 & 0.9389
\end{bmatrix}.
\]

\[A, C_1, \text{and } C_2\] color components are defined by:

\[
A = a(\alpha \log(L) + \beta \log(M) + \gamma \log(S)),
\]

\[
C_1 = u_1(\log(L) - \log(M)),
\]

\[
C_2 = u_2(\log(L) - \log(S)).
\]

These equations indicate that the cone response to a color stimulus is nonlinear. Therefore, Faugeras proposed to model this nonlinearity by using the logarithmic function, while the CIE method uses the cube root function [see Eq.(5)].

By adjusting the \(a, \alpha, \beta, \gamma, u_1,\) and \(u_2\) parameters, Faugeras proposes different applications of the model. For example, he uses the following coefficients for color image analysis: \(a = 22.6, \alpha = 0.612, \beta = 0.369, \gamma = 0.019, u_1 = 64,\) and \(u_2 = 10.\)

In the framework of artificial vision, Garbay propose application of this space directly on the \(R, G,\) and \(B\) color components of an image acquisition...
system (Chassery and Garbay, 1984; Garbay, Brugal and Choquet, 1981). The \(a, \alpha, \beta, \gamma, u_1\), and \(u_2\) parameters are defined by:

\[
A = \frac{1}{3} \times (\log(R) + \log(G) + \log(B)),
\]

\( (20) \)

\[
C_1 = \frac{\sqrt{3}}{2} \times (\log(R) - \log(G)),
\]

\( (21) \)

\[
C_2 = \log(B) - \frac{\log(R) + \log(G)}{2}.
\]

\( (22) \)

Ballard and colleagues proposed the use of an antagonist space that does not take into account the nonlinearity of the human eye response (Ballard and Brown, 1982; Swain and Ballard, 1991). Thus, the equation of this space, denoted as \((wb, rg, by)\), can be written with a transformation matrix:

\[
\begin{bmatrix}
wb \\
rg \\
by
\end{bmatrix}
= P \times
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix},
\]

with

\[
P = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 \\
-\frac{1}{2} & -\frac{1}{2} & 1
\end{bmatrix}.
\]

\( (23) \)

Other human visual system models have been applied in the color image analysis field, such as the \((G_1, G_2, G_3)\), \((G^*_1, G^*_2, G^*_3)\) or \((H_1, H_2, H_3)\) color spaces (Braquelaire and Brun, 1997; Robinson, 1977; Pratt, 1978).

d. Other Luminance-Chrominance Spaces. Other luminance-chrominance spaces, which cannot be directly classified in the previous subfamilies of luminance-chrominance spaces, are applied in color image analysis.

By studying the properties of different luminance-chrominance spaces, Lambert and Carron (1999) proposed a transformation matrix to convert an \((R, G, B)\) color space to the luminance-chrominance space denoted \((Y, Ch_1, Ch_2)\) in which the luminance component is defined by the \(wb\) color component of Eq. \((23)\), whereas the chrominance components are defined by the following:

\[
\begin{bmatrix}
Y \\
Ch_1 \\
Ch_2
\end{bmatrix}
= P \times
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix},
\]

with

\[
P = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
1 & -\frac{1}{2} & -\frac{1}{2} \\
0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2}
\end{bmatrix}.
\]

\( (24) \)

The \(Ch_1\) and \(Ch_2\) color components correspond to a cyan-red and green-blue color oppositions, respectively.

By associating the \(r\) and \(g\) chromaticity coordinates with the luminance component defined by Eq. \((23)\) and denoted here \(I\), the \((I, r, g)\) luminance-chrominance space is defined. This space is often used in color image

CIE proposed use of the \((Y, x, y)\) color space, where \(Y\) represents the luminance component [see Eq. (15)] and \(x\) and \(y\), which are the chromaticity coordinates of the \((X, Y, Z)\) color space [see Eq. (3)], represent the chrominance.

3. Perceptual Spaces

Humans do not directly perceive color as a combination of tristimulus values related to primary colors but according to more subjective entities related to luminosity, hue, and saturation. Therefore, it is natural that many color spaces quantify the color according to these perceptual data and are grouped here into the perceptual space family. This perceptual approach allows an adequate communication between human vision and machinery for describing colors. There exist many such color spaces presented with different notations, such ISH, HSB, HSL, HSV, LCH, LSH, and so on. These different notations correspond to the same color components, but the equations to determine them are different. Two kinds of perceptual spaces can be distinguished:

- The polar (or cylindrical) coordinate spaces that correspond to expressions in polar coordinates of the luminance-chrominance components
- The perceptual coordinate spaces that are directly evaluated from primary spaces.

Perceptual spaces also can be considered as luminance-chrominance ones because they are consist of a luminance component and two chrominance components.

a. Polar Coordinates Spaces. This family of color spaces is derived from the color spaces that separate color information into a luminance axis and a chrominance plane by transposition of Cartesian coordinates to polar coordinates. Let \(P\) be a point that represents a color in an \((L, Chr_1, Chr_2)\) luminance-chrominance space. This point is defined by the coordinates of the color vector \(\overrightarrow{OP}\) in the reference \((O, \overrightarrow{L}, \overrightarrow{Chr_1}, \overrightarrow{Chr_2})\). Let \(P'\) be the projection of \(P\) on the \((O, Chr_1, Chr_2)\) plane along the \(L\) axis and let \(O'\) be the orthogonal projection of \(P\) on the \(L\) axis. By construction, the norm and the orientation of \(O'O\) and \(PP'\) vectors are equal as well as for the \(OO'\) and \(PP'\) vectors. Thus it is possible to locate the \(P\) point via the norm of the \(\overrightarrow{OP'}\) vector, the angle between the \(\overrightarrow{OP'}\) and \(\overrightarrow{Chr_1}\) vectors, and its coordinate along the \(L\) axis, which is equal to the norm of the \(\overrightarrow{OO'}\) vector. The three components so obtained are denoted \(L, C,\) and \(H\) and constitute the \((L, C, H)\) polar coordinate space (see Figure 9).

\(^2\) These color spaces are also named spherical color transform (SCT).
The first component of a polar coordinate space represents the luminance component, $L$, which is identical to the first color component of the corresponding luminance-chrominance space.

The norm of the $\mathbf{OP}'$ vector represents the chroma, $C$, which is defined by

$$C = \|\mathbf{OP}'\| = \sqrt{\text{Chr}_1^2 + \text{Chr}_2^2}.$$  \hspace{1cm} (25)

The chroma corresponds to the distance between the $P$ point and the luminance axis.

The angle of the $\mathbf{OP}'$ vector with the $\mathbf{Chr}_1$ vector represents the hue, $H$.

In order to obtain hue values ranging between 0 and $2\pi$, the evaluation of $H$ must satisfy these conditions:

$$H = \left(\mathbf{Chr}_1, \mathbf{OP}'\right) \text{ with}$$

$$\begin{align*}
\text{if } \text{Chr}_1 > 0 \text{ and } \text{Chr}_2 \geq 0 & \text{ then } 0 \leq H < \frac{\pi}{2}, \\
\text{if } \text{Chr}_1 \leq 0 \text{ and } \text{Chr}_2 > 0 & \text{ then } \frac{\pi}{2} \leq H < \pi, \\
\text{if } \text{Chr}_1 < 0 \text{ and } \text{Chr}_2 \leq 0 & \text{ then } \pi \leq H < \frac{3\pi}{2}, \\
\text{if } \text{Chr}_1 \geq 0 \text{ and } \text{Chr}_2 < 0 & \text{ then } \frac{3\pi}{2} \leq H < 2\pi.
\end{align*}$$  \hspace{1cm} (26)
The trigonometric function is used to evaluate $H$. For example, $H$ can be defined by:

$$H = \arctan \left( \frac{Chr_2}{Chr_1} \right).$$

This equation gives an angle belonging in the interval $[-\pi/2, \pi/2]$. In order to obtain an angle belonging in the interval $[0, 2\pi]$ and satisfying Eq. (26), it is necessary to first apply an operation on $Chr_2$ when $Chr_1 < 0$: if $Chr_2 \geq 0$ then $H'' = \pi + H$; else $H'' = \pi - H$. This first operation gives an angle belonging in the interval $[-\pi, \pi]$. So the following second operation is applied: $H' = H'' + 2\pi$, if $H'' < 0$.

By definition, it is possible to construct an $(L, C, H)$ color space from any of the luminance-chrominance spaces. For example, by using Eqs. (25) and (27), CIE defines the components of two $(L, C, H)$ polar coordinate spaces from the perceptually uniform spaces $(L^*, u^*, v^*)$ and $(L^*, a^*, b^*)$, respectively. These spaces are denoted $(L^*_{uv}, C^*_{uv}, h_{uv})$ and $(L^*_{ab}, C^*_{ab}, h_{ab})$, respectively (Commission Internationale de l’Éclairage, 1986). The components of these color spaces are used to evaluate color differences (Commission Internationale de l’Éclairage, 1986, 1995).

b. Perceptual Coordinate Spaces. The perceptual coordinate spaces are evaluated directly from a primary space and represent the subjective entities of the color human perception in terms of intensity ($I$), saturation ($S$), and hue ($T$). The intensity $I$ corresponds to a luminance component, and the saturation $S$ is related to a chroma component expressed by the relation $S = C / L$. The perceptual coordinate spaces are denoted here as $(I, S, T)$ to differentiate them from the polar coordinate spaces. This part presents the most widely used perceptual coordinate spaces (Shih, 1995):

- The triangle model
- The hexcone model
- The double hexcone model.

To propose a perceptual coordinate space that takes into account most of the models, Levkowitz and Herman (1993) proposed a generalized lightness, hue, and saturation (GLHS) color model.

i. Triangle Model. The triangle model corresponds to the expression of the $I$, $S$, and $T$ color components in the RGB color cube (Kender, 1976). In the representation of the RGB color cube (see Figure 4), the achromatic axis corresponds to the intensity axis in the $(I, S, T)$ color space. A point $P$, whose coordinates are the values of the $R$, $G$, and $B$ color components, is located on a plane perpendicular to the achromatic component. The intersections of this
FIGURE 10. Triangle model. (See Color Insert.)

plane with the red, green, and blue axes are the summits of a triangle \((\alpha, \beta, \gamma)\), which is proportional to the Maxwell triangle (Figure 10). On the \((\alpha, \beta, \gamma)\) triangle, it is possible to evaluate the saturation and hue components. Different methods have been proposed, and many formulations of the \((I, S, T)\) color spaces exist. Some of them are presented here.

Let \(O'\) be the orthogonal projection of \(P\) on the achromatic axis. The \(I\) color component corresponds to the norm of the \(\overrightarrow{OO'}\) vector. To maximize its value to the unity when \(R = G = B = 1\), the \(I\) color component is evaluated by the equation:

\[
I = \frac{1}{\sqrt{3}} \lVert \overrightarrow{OO'} \rVert = \frac{R + G + B}{3}. \tag{28}
\]

To simplify this calculation, a nonweighted formulation of the previous equation is often used:

\[
I = R + G + B. \tag{29}
\]

The saturation corresponds to the ratio between the norm of the \(\overrightarrow{OP}\) vector (distance between the \(P\) point and the achromatic axis) and the norm of the
vector, which represents the intensity component. The calculation of the $S$ color component is by

$$S = \frac{\|\vec{O'P}\|}{\|\vec{OO'}\|} = \sqrt{2} \times \frac{\sqrt{R^2 + G^2 + B^2 - RG - GB - RB}}{R + G + B}. \quad (30)$$

In this formulation, colors with equal saturation values are located on a circular base cone centered on the achromatic axis and with $O$ as summit. The saturation value reaches its maximal value only for the three primary colors. In order to maximize the saturation value for all points belonging to the boundaries of the $(\alpha, \beta, \gamma)$ triangle (like those corresponding to complementary colors), the saturation component is evaluated by the minimum distance between the $P$ point and one boundary of this triangle:

$$S = 1 - \frac{3 \times \min(R, G, B)}{R + G + B}. \quad (31)$$

Different formulations of this equation exist and it is important to show their relations. Thus, another similar equation can be used to define the $S$ color component:

$$S = 1 - 3 \times \min(r, g, b). \quad (32)$$

This equation uses the $r$, $g$, and $b$ chromaticity coordinates and corresponds to the evaluation of the saturation component directly in the Maxwell triangle. The saturation component can be expressed in function of the intensity component of Eq. (29):

$$S = 1 - \frac{3 \times \min(R, G, B)}{I}. \quad (33)$$

The hue component corresponds to the orientation of the $\vec{O'P}$ vector. Let $M$ be the intersection point between the achromatic axis and the Maxwell triangle, and let $P'$ be the projection of the $P$ point on this triangle along the achromatic axis. Generally, the axis defined by the $M$ point and the Red point [coordinates $(1, 0, 0)$] in the RGB color cube is the reference axis to define the orientation of the $\vec{O'P}$ vector and to specify the value of the hue component. So

$$T = (\vec{M \text{ Rouge}}, \vec{MP'})$$

and the value of the red color is set to 0. The hue component can be evaluated via trigonometric relations. For example, the following equation defines the $T$ color component:

$$T = \arctan \left( \frac{\sqrt{3}(G - B)}{2R - G - B} \right). \quad (34)$$
To obtain hue values included between 0 and $2\pi$, it is necessary to test the sign of the numerator and the denominator.

Another analog trigonometric equation can be used to define the $T$ color component:

$$T = \begin{cases} \arccos \left( \frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right) & \text{if } B \leq G, \\ 2\pi - \arccos \left( \frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right) & \text{if } B > G. \end{cases} \quad (35)$$

With this formulation, it is necessary to test if $B > G$ in order to consider only those angles included between $\pi$ and $2\pi$.

By using the chromaticity coordinates, Eq. (35) can be written as

$$T = \begin{cases} \arccos \left( \frac{2r-g-b}{\sqrt{6\times[(r-\frac{1}{3})^2+(g-\frac{1}{3})^2+(b-\frac{1}{3})^2]}} \right) & \text{if } b \leq g, \\ 2\pi - \arccos \left( \frac{2r-g-b}{\sqrt{6\times[(r-\frac{1}{3})^2+(g-\frac{1}{3})^2+(b-\frac{1}{3})^2]}} \right) & \text{if } b > g. \end{cases} \quad (36)$$

All these formulas correspond to the triangle model and can be associated to constitute the $(I, S, T)$ color space.

**ii. Hexcone Model.** Each point $P$, whose coordinates are the values of the $R$, $G$, and $B$ color components, belongs to a face of a color cube whose summit $O'$ corresponds to the maximum level of the $R$, $G$, and $B$ components. The projection of the points of this color cube along the achromatic axis on the plane perpendicular to this axis and joining the $O'$ point constitutes a hexagonal closed area whose summits are the projections of the summits of the so defined color cube and whose center is the $O'$ point. In this model, which is represented by Figure 11, it is thus possible to define the $I$, $S$, and $T$ color components.

The intensity component, known by the term *value* and denoted $V$, is represented by this achromatic point and is expressed as

$$I = V = \frac{1}{\sqrt{3}} \| \overrightarrow{OO'} \| = \max(R, G, B). \quad (37)$$

The projection on the plane perpendicular to the achromatic axis and joining the summit $O'$ of the so constructed color cube defines a hexagon in which is located the $P'$ point, projection of $P$. The saturation component corresponds to the length of the $O'P'$ segment divided by the maximal length for a same hue. It is expressed by Eq. (33):

$$S = \frac{V - \min(R, G, B)}{V}. \quad (38)$$
In the case when \( V = 0 \), the saturation components cannot be defined (hence \( S = 0 \)).

By construction, the White point is also the projection of the Black point. The projection of the \( \overrightarrow{Black \ Red} \) vector is thus the \( \overrightarrow{White \ Red} \) vector. Let \( M \) be the projection of \( P' \) on the plane perpendicular to the achromatic axis and joining the White point. The hue component is defined as the angle between the \( \overrightarrow{White \ M} \) vector and the \( \overrightarrow{White \ Red} \) vector. Therefore,

\[
T = \left( \overrightarrow{White \ Red}', \overrightarrow{White \ M} \right)
\]

and the \( T \) color component is evaluated, for all \( S \neq 0 \), by

\[
T = \begin{cases} 
\frac{G - B}{V - \min(R,G,B)} & \text{if } V = R, \\
\frac{B - R}{V - \min(R,G,B)} & \text{if } V = G, \\
\frac{R - G}{V - \min(R,G,B)} & \text{if } V = B.
\end{cases}
\]  

(39)

If \( S = 0 \) (when \( R = G = B \)), then the hue component is not defined. Moreover, if \( V = R \) and \( \min(R,G,B) = G \), then \( T \) is negative.

The hexcone model is defined by Eqs. (37), (38), and (39) (Foley, Dam and Feiner, 1990; Marcu and Abe, 1995; Shih, 1995).
iii. Double Hexcone Model. This model is based on the same principle as the previous model, except that the projections of the subcube are achieved on each side of the plane perpendicular to achromatic axis and joining the middle of this axis (Foley, Dam and Feiner, 1990; Marcu and Abe, 1995; Shih, 1995).

Figure 12 shows a representation of the double hexcone model. If \( \min = \min(R, G, B) \) and \( \max = \max(R, G, B) \), the intensity component is expressed as

\[
I = \frac{\max + \min}{2}. \tag{40}
\]

Let \( I_{\text{max}} \) be the maximum value of the \( I \) color component. The saturation component is defined, for all \( I \neq 0 \), by

\[
S = \begin{cases} 
\frac{\max - \min}{\max + \min} & \text{if } I \leq \frac{I_{\text{max}}}{2}, \\
\frac{\max - \min}{2 \times I_{\text{max}} - \max - \min} & \text{if } I > \frac{I_{\text{max}}}{2}.
\end{cases} \tag{41}
\]
The saturation component is equal to 0 \( (S = 0) \) if \( I = 0 \) and the hue component is not defined if \( S = 0 \) (when \( R = G = B \)). Otherwise, the relations that allow processing the \( T \) color component are the same as those of the above hexcone color model [see Eq. (39)].

iv. \((L^*_uv, S^*_uv, h_{uv})\) Color Space. This space was defined by the CIE at the same time that the \((L^*_ab, C^*_ab, h_{ab})\) and \((L^*_uv, C^*_uv, h_{uv})\) color spaces were defined. In the \((L^*, u^*, v^*)\) color space, CIE defines the saturation component as the ratio

\[
S^*_uv = \frac{C^*_uv}{L^*_uv} \tag{42}
\]

The components \(L^*_uv\) and \(h_{uv}\) are defined by Eqs. (5) and (25), respectively (Commission Internationale de l’Éclairage, 1986).

4. Independent Axis Spaces

Because the color components of an \((R, G, B)\) color space are strongly correlated, they share common luminance information (Lee, Chang and Kim, 1994; Ohta, Kanade and Sakai, 1980). Therefore, many authors attempt to determine color spaces whose components are independent (i.e., components that share different noncorrelated and nonredundant informations). A general solution consists in applying a Karhunen–Loeve transformation or a principal component analysis (PCA) to the components of a color space.

a. Principal Component Analysis. Principal component analysis is a data analysis method. Its goal is to analyze a set of quantitative data represented in a multidimensional space to obtain a representation space with (eventually) a reduced dimension whose components, named principal components, are uncorrelated and share different informations.

When the data are the values of the \(R\), \(G\), and \(B\) color components of pixels of an image, the principal components analysis provides a color space whose color components are as uncorrelated as possible and can be independently considered.

For this purpose, the set of color vectors within a color image is characterized by its diagonalized covariance matrix. The eigenvectors (denoted \(w_i\) of this matrix) are determined and the principal components (denoted \(X_i\)) are processed by the relation: \(X_i = w_i [R \ G \ B]^T\). The Karhunen–Loeve transformation is used to apply this relation to each of the \(X_i\) components. Thus, this linear transformation consists of projecting the data into another space. The maximal eigenvalue \(\lambda_i\) corresponds to the first principal component \(X_1\)—the component that mainly represents the image. The components
are thus ordered with respect to their decreasing discriminant powers (their eigenvalues \( \lambda_i \)).

This transformation, which also allows reduction of the dimension of the color space, is principally applied from an \((R, G, B)\) color space but can be derived from any of color spaces (Savoji and Burge, 1985; Tominaga, 1992).

The drawback of the PCA-based methods is that it depends on the statistical properties of a data set. In color image analysis, the application of PCA on each analyzed color image is time consuming. To reduce computation time, several authors determined independent axis spaces that approximate the Karhunen–Loeve transformation by applying PCA on several different sets of images, such as \((I_1, I_2, I_3)\), \((P_1, P_2, P_3)\), \((I, J, K)\), or \((i_{1\text{new}}, i_{2\text{new}}, i_{3\text{new}})\) color spaces (Ohta, Kanade and Sakai, 1980; Philipp and Rath, 2002; Robinson, 1977; Verikas, Malmqvist and Bergman, 1997).

In the framework of natural color image analysis, an experiment led by Ohta et al. in 1980 on eight different images allows determination of a color space based on the Karhunen–Loeve transformation (Ohta, Kanade and Sakai, 1980). These authors proposed segmenting color images into regions by a recursive thresholding method and applying a PCA at each iteration step of the algorithm. They showed that there exists a single transformation that allows efficient conversion of the \((R, G, B)\) color space to a space denoted \((I_1, I_2, I_3)\). This transformation can be defined by this transformation matrix:

\[
\begin{bmatrix}
I_1 \\
I_2 \\
I_3
\end{bmatrix} = P \times \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}, \quad \text{with} \quad P = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{2} & 0 & -\frac{1}{2} \\
-\frac{1}{4} & \frac{1}{2} & -\frac{1}{4}
\end{bmatrix}.
\]

The first component of the \((I_1, I_2, I_3)\) color space is the most discriminating one and represents a luminance component because it verifies Eq. (28). The two other color components represent the blue-red and magenta-green color oppositions, respectively. These two color components are less discriminating than the first one because their corresponding eigenvalues are low and the third component is not always used by color image analysis. Thus, the \((I_1, I_2, I_3)\) color space can also be considered a luminance-chrominance space.

*b. Independent Component Analysis.* Another solution to obtain an independent axis space consists of transforming the \(R\), \(G\), and \(B\) color components of the pixels to independent components via an independent component analysis method. This method allows determination of color spaces whose components are statistically independent but where information is dispersed on all components, whereas the first component deduced by PCA contains the maximum information (Lee and Lewicki, 2002).
5. Hybrid Color Space

The analysis of the colors of the pixels in a color space is not restricted to the \((R, G, B)\) color space. Indeed, the previous text shows that there exist a large number of color spaces that can be used to represent the colors of the pixels (Sharma and Trussell, 1997). They are supported by specific physical, physiologic, and psychovisual properties (Sangwine and Horne, 1998). The multitude and the diversity of available color spaces requires classifying them into categories according to their definitions (Poynton, 1995; Tkalcic and Tasic, 2003). We propose grouping the most classical color spaces into four principal families that are further divided into subfamilies. Figure 13 distinguishes these four main families by four main rectangles:

- Primary spaces
- Luminance-chrominance spaces
- Perceptual space
- Independent axis spaces.

The dotted rectangles within the main rectangles correspond to subfamilies of color spaces. In general, color images are acquired through the \((R, G, B)\) color image acquisition space. Therefore, all color spaces are defined by an equation whose inputs are the \(R\), \(G\), and \(B\) color components. Figure 13 shows how a color space is determined by following the arrows starting from the \((R, G, B)\) color space.

We have compiled a nonexhaustive list of the most widely used color spaces by image analysis applications. Authors sometimes propose color spaces that are specific to their applications. For example, CIE presently proposes a color appearance model (CIECAM) that attempts to consider all the visual phenomena interfering on the color human perception (Commission Internationale de l’Éclairage, 2004). This empirical model defines the color of a surface by taking into account its neighborhood. In the color image analysis framework, the use of such a color model leads to considering the neighborhood of the pixels to evaluate color components.

On the one hand, when considering the multitude of available color spaces, it is difficult to select a specific color space adapted to all the color image analysis applications. On the other hand, the performance of an image-processing procedure depends on the choice of the color space. Many authors have tried to determine the color spaces that are well suited for their specific color image segmentation problems. Authors provide contradictory conclusions about the pertinence of the available color spaces in the context of color image analysis; it is thus easy to conclude that, there currently does not exist any classical color space that provides satisfying results for the analysis of all types of color images.
FIGURE 13. Color space families. (See Color Insert.)
Instead of searching for the best classical color space for color image analysis, Vandenbroucke, Macaire, and Postaire proposed an original approach to improve the results of image processing. They defined a new kind of color space by selecting a set of color components that can belong to any of the different classical color spaces listed in Figure 13 (Vandenbroucke, Macaire and Postaire, 1998). These spaces, which have neither psychovisual nor physical color significance, are called hybrid color spaces and can be automatically selected by means of an iterative feature selection procedure operating with supervised learning scheme (Vandenbroucke, Macaire and Postaire, 2003).

C. Digital Color Images and Color Spaces

In the previous section, the most classical color spaces were presented and classified into different families. However, the use of these color spaces in color image analysis requires some precautions. Indeed, when a color image is converted to an image where colors are represented in another color space, their values belong to different value domains in function of the applied transformation.

In analyzing the color distribution of an image in a color space (e.g., Euclidean distance comparison, histogram evaluation, memory storage, image coding, color display), the values of the transformed color components often must belong to the same value domain as that in the original color space. To satisfy this constraint, it is necessary to code each color space without changing its intrinsic properties (Vandenbroucke, Macaire and Postaire, 2000a). This last condition proposes a color space-coding scheme applied to the different families of color spaces as presented by Figure 13.

1. Color Space Coding for Image Analysis

The coding of the color spaces consists of rounding, scaling, and normalizing the values of their components in order to process values that range between the unsigned integer values 0 and 255. We propose a color space-coding scheme that preserves the properties of each of the color spaces.

a. Independent and Dependent Coding. The components of all color spaces are defined by one or several successive color transformations of the \( R \), \( G \), and \( B \) components. Generally, each of these three components is coded on 8 bits and can take 256 different unsigned integer values belonging to the interval [0, 255].

Let us denote \( T_1 \), \( T_2 \), and \( T_3 \), the transformed components of any color spaces obtained from the \( R \), \( G \), and \( B \) components by a set of color
transformation $T$. In most cases, transformed color components assume signed floating values that do not range between 0 and 255.

For several color image analysis applications, the transformed values of the color components must be comparable to those represented by the original color space. That means that a specific coding scheme is necessary. Let $T'_1$, $T'_2$, and $T'_3$, be the coded components provided by a transformation, denoted $C$ (Figure 14).

The space-coding scheme of a transformed color space $(T_1, T_2, T_3)$ is divided into different successive steps as follows:

- Shifting the color component values so that they are unsigned. This operation requires knowing the minimal value of each component of the considered color space.
- Normalizing the shifted component values so that they range between 0 and 255 (for 8 bit coding). This operation requires knowing the range of each component of the considered color space (i.e., their minimal and maximal values).
- Rounding the normalized shifted component values to obtain integer values. These values are rounded to the nearest integer.

In order to achieve the coding of the transformed color components $T_1$, $T_2$, and $T_3$, it is necessary to determine the extreme values of their ranges. Let us denote $m_k$ and $M_k$ ($k \in \{1, 2, 3\}$), the minimum and maximum values of the transformed color component $T_k$, respectively, so that $\Delta_k = M_k - m_k$ represents the range of the transformed color component $T_k$. In order to adjust the width of this range to 255, the transformed color component $T_k$ can be coded independently of the two other components as

$$T'_k = \frac{255}{\Delta_k} \times (T_k - m_k).$$

According to the properties of a color space, it is sometimes necessary to use a dependent coding scheme so that the range of at least one color component is equal to 255. Let $\Delta_{max}$ denote the larger range of the three components of a color space, defined as

$$\Delta_{max} = \max_{k} (\Delta_k).$$
The dependent coding of the components of a color space is expressed as

\[ T'_k = \frac{255}{\Delta_{\text{max}}} \times (T_k - m_k). \]  

The dependent coding scheme achieves an equal scaling process for each component of a color space. Therefore, the relative position of colors in the space defined by this coding scheme is not modified. Furthermore, the Euclidean distances between colors are preserved by such a coding scheme.

2. Application to Color Spaces

In order to preserve their properties, we apply one of the two above-defined coding schemes to the color spaces of the four families described by Figure 13.

a. Primary Spaces. The transformation of an \((R, G, B)\) color space to any other primary space is a linear transformation that depends on the choice of the primaries and the reference white. For instance, the transformation from the NTSC \((R_F, G_F, B_F)\) color space to the CIE \((X, Y, Z)\) color space with the C illuminant is defined via a transformation matrix \([X \ Y \ Z]^T = T \ [R_F \ G_F \ B_F]^T\) (see Table 3), where

\[ T = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix}. \]  

The chromaticity coordinates of the C illuminant are \(x_n = 0.310\) and \(y_n = 0.316\), and those of the primaries \(R_F, G_F, B_F\) are \(x_r = 0.670\), \(y_r = 0.330\), \(x_g = 0.210\), \(y_g = 0.710\), and \(x_b = 0.140\) and \(y_b = 0.080\), respectively. Table 5 lists the coding parameters of the \((X, Y, Z)\) color space.

Ohta, Kanade and Sakai (1980) proposed coding this color space by means of an independent coding scheme associated with a matrix \(C_{\text{ind}}\) so that \([T'_1 \ T'_2 \ T'_3]^T = C_{\text{ind}} \ T \ [R_F \ G_F \ B_F]^T\) with

\[ C_{\text{ind}} \ T = \begin{bmatrix} 0.618 & 0.177 & 0.205 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.056 & 0.944 \end{bmatrix}. \]  

<table>
<thead>
<tr>
<th>(X)</th>
<th>(Y)</th>
<th>(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_k)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(M_k)</td>
<td>250.16</td>
<td>255.00</td>
</tr>
</tbody>
</table>
With such a coding scheme, the evaluated chromaticity coordinates are \( x_n = 0.333, y_n = 0.333, x_r = 0.674, y_r = 0.326, x_g = 0.224, y_g = 0.743, \) and \( x_b = 0.162 \) and \( y_b = 0.090 \). They do not correspond to the coordinates that are computed by the matrix \( T \). Since the position of the illuminant \( C \) in the chromaticity diagram has changed with the independent coding scheme, we can conclude that the definition of the \((X, Y, Z)\) color space from the primary space NTSC is modified by such a coding scheme. By extending the case of this illuminant to all the available colors, we conclude that the spectrum locus contained by the \((x, y)\) chromaticity diagram is distorted by the independent coding scheme.

Conversely, the dependent coding scheme provides a matrix \( C_{dep} \) so that:

\[
C_{dep} T = \begin{bmatrix}
0.514 & 0.147 & 0.169 \\
0.253 & 0.497 & 0.096 \\
0.000 & 0.056 & 0.944
\end{bmatrix}. \tag{49}
\]

The dependent coding scheme does not modify the actual chromaticity coordinates. This example can be generalized to the other primary spaces. Thus, in order to preserve the colorimetric properties of a primary space, we propose to achieve a dependent coding for each of its components.

**b. Luminance-Chrominance Spaces.** The color spaces of this family are defined by linear or nonlinear transformations. The color spaces obtained through a linear transformation can be considered as primary spaces and must be coded with a dependent coding scheme for the same reasons as the primary spaces.

The color components of the \((L^*, u^*, v^*)\) color space are defined by nonlinear transformations [see Eqs. (5)–(7)]. By applying an independent coding to the components of this space, the ellipses of MacAdam are distorted so that the color distances that are evaluated in this color space do not correspond to the color differences perceived by a human observer. To preserve the shape of MacAdam ellipses and to preserve an adequacy between the Euclidean distance and the visual perception, we propose application of a dependent coding scheme to the perceptually uniform spaces. To illustrate this phenomenon, an ellipse is represented in the \((u^*, v^*)\) chromaticity diagram of Figure 15a before applying a coding scheme. In this figure, the two distances denoted \( D_1 \) and \( D_2 \) are equal. MacAdam rules assume that there are no perceptible differences between the corresponding colors for the human eye. By applying an independent coding scheme to the \((L^*, u^*, v^*)\) color space (Figure 15b), these properties are modified because \( D_1' \neq D_2' \), whereas with a dependent coding scheme, these properties are always satisfied because \( D_1' = D_2' \) (Figure 15c).

We extend this dependent coding scheme to all the luminance-chrominance spaces obtained through nonlinear transformations.
FIGURE 15. Application of the independent and dependent codings to an ellipse represented in the \((u^*, v^*)\) chromaticity diagram. (a) Ellipse in the \((u^*, v^*)\) chromaticity diagram. (b) Independent coding. (c) Dependent coding.
c. Perceptual Spaces. Since the components of the perceptual spaces represent three subjective entities to qualify a color, they can be considered independently. Furthermore, in order to compare two colors, the Euclidean distance is meaningless since the hue component is periodical. Therefore, we propose applying an independent coding scheme to the perceptual spaces.

d. Independent Axis Spaces. Because the components of the independent axis spaces are determined by linear transformations, they can be considered primary spaces so they must be coded with a dependent coding scheme.

e. Hybrid Color Spaces. The components of a hybrid color space are selected among all the components of several color spaces. Thus, this color space has neither psychovisual nor physical color significance and dependent color coding is not justified. That means that the hybrid color spaces must be coded by means of an independent color-coding scheme.

D. Summary

This first text section which dealt with the representation of the colors characterizing the pixels in a color space, underscoring two important points:

- In order to use a color space, it is important to know the acquisition conditions (i.e., to specify the reference white, the lighting system, and the camera or scanner parameters, such as gain, offset, and gamma corrections).
- A color calibration of the color image acquisition device is recommended so that the colors of the objects observed by an image sensor are represented correctly in the acquired color images.

The second part of this section presented the most commonly used color spaces for color image analysis. Each color space is characterized by its physical, physiologic or psychological properties, but most of these color spaces are not initially developed for color image analysis applications. Therefore, their exploitation by a digital color image acquisition device require that some conditions be satisfied.

The last part of this section shows that a specific coding scheme must be applied on these color spaces to correctly use them for color image analysis without modifying their intrinsic properties.

The next section of this chapter shows how the color spaces can be exploited by color image segmentation algorithms, and the last section studies the impact of a choice of a color space on the results provided by these algorithms.
III. Color Image Segmentation

A. Introduction

Generally, we assume that the different colors that are present in an image correspond mainly to different properties of the surfaces of the observed objects. The segmentation procedures analyze the colors of pixels in order to distinguish the different objects that constitute the scene observed by a color sensor or camera. It is a process of partitioning an image into disjoint regions (i.e., into subsets of connected pixels that share similar color properties).

The segmentation of the image denoted $I$ regroups connected pixels with similar colors into $N_R$ regions $R_i$, $i = 1, \ldots, N_R$ (Zucker, 1976). The pixels of each region must respect homogeneity and connectedness conditions. The homogeneity of a region $R_i$ is defined by a uniformity-based predicate, denoted $\text{Pred}(R_i)$, which is true if the colors of $R_i$ are homogeneous and false on the opposite.

The regions must respect the following conditions:

- $I = \bigcup_{i=1, \ldots, N_R} R_i$: each pixel must be assigned to one single region and the set of all the regions must correspond to the image.
- $R_i$ contains only connected pixels $\forall i = 1, \ldots, N_R$: a region is defined as a subset of connected pixels.
- $\text{Pred}(R_i)$ is true $\forall i = 1, \ldots, N_R$: each region must respect the uniformity-based predicate.
- $\text{Pred}(R_i \cup R_j) = \text{false}$ $\forall i \neq j$, $R_i$ and $R_j$ being adjacent in $I$: two adjacent regions do not respect the predicate.

The result of the segmentation is an image in which each pixel is associated with a label corresponding to a region.

Segmentation schemes can be divided into two primary approaches with respect to the used predicate (Cheng et al., 2001). The first approach assumes that adjacent regions representing different objects present local discontinuities of colors at their boundaries. (Section III.B describes the edge detection methods deduced from this assumption.)

The second approach assumes that a region is a subset of connected pixels that share similar color properties. The methods associated with this assumption are called region construction methods and look for subsets of connected pixels whose colors are homogeneous. These techniques can be categorized into two main classes—whether the distribution of the pixel colors is analyzed in the image plane or in the color space.

The spatial analysis described in Section III.C is based on a region growing or merging process.
The analysis in the color space takes advantage of the characterization of each pixel \( P \) of an image \( I \) by its color point \( I(P) \) in this space. Since the pixels are represented by points in a 3D color space, this approach assumes that homogeneous regions in the image plane give rise to clusters of color points in the color space. Each cluster corresponds to a class of pixels that share similar color properties. These clusters generally are identified by means of an analysis of the color histogram or a cluster analysis procedure and are mapped back to the original image plane to produce the segmentation (see Section III.D).

In this chapter, we do not consider the physically based segmentation techniques that use the explicit assumptions about the physics that create the images. Klinker, Shafer and Kanade (1990) use a model of the formation of the color, called the dichromatic reflection model, to segment a color image. This model contains several rigid assumptions, such as the illumination conditions and the type of observed materials. For most real scenes, these assumptions cannot always be justified. Therefore, the proposed model can be used only for a restricted number of images, such as the scenes observed within a controlled environment.

For illustration purposes, we apply segmentation schemes to the synthetic image of Figure 16a, which contains six different regions with different colors and shapes (a brown background, a small yellow square, an orange large square, a purple patch, and two concentric green disks). This image, with color coded in the \((R, G, B)\) color space, is corrupted by an uncorrelated Gaussian noise with a standard deviation \(\sigma = 5\), which is independently added to each of the three color components. To show the influence of the chosen color space on the segmentation results, the colors of the pixels are represented by some of the most widely known color spaces: the \((R, G, B)\) color space (Figure 16a), the \((Y', U', V')\) color space defined by Section II.B.2.b (Figure 16b), and the triangular \((I, S, T)\) color space defined by Section II.B.3.b (Figure 16c).

### B. Edge Detection

#### 1. Overview

Edge detection consists of detecting local discontinuities of colors. The result of the edge pixel detection procedure is a binary image composed of edge and nonedge pixels. Cumani proposed detecting the zero-crossings of second-order derivative filters (Cumani, 1991). However, this approach is very sensitive to the presence of noise in the images, why is the reason several authors apply first-order derivative filters (gradient) to detect edges. Locations of gradient maxima that are higher than a threshold generate edge pixels.
The color edge detection methods based on gradient analysis can be divided into three families according to the analyses of the color component images denoted \( I^k, k = R, G, B \) (i.e., the monochromatic images where pixels are characterized by the levels of one single color component):

1. The methods that fuse edge binary images obtained by the analyses of different component images (Rosenfeld and Kak, 1981) (Figure 17a).

2. The methods that independently analyze each color component image in order to process three marginal gradient vectors. These three vectors are combined to provide the norm of a gradient (Lambert and Carron, 1999) (Figure 17b).

3. The methods that process a color gradient vector from three gradient vectors computed in each of the three color component images. The edge pixels are detected by an analysis of a color vector gradient (Di Zenzo, 1986; Lee and Cok, 1991) (Figure 17c). Since these methods provide the
FIGURE 17. Different approaches for color edge detection. (a) Analysis of edge binary images. (b) Analysis of the norm of a color gradient. (c) Analysis of a color gradient vector. (See Color Insert.)
best results in terms of edge detection (Cheng et al., 2001), we describe 
this method, which is the most widely used method.

2. Edge Detection by the Analysis of a Color Gradient Vector

In order to detect color edge pixels, the color gradient magnitude and direction 
can be determined by Di Zenzo’s algorithm, which computes the first-order 
differential operators (Di Zenzo, 1986). The color gradient detection consists 
in determining, at each pixel, the spatial direction denoted $\theta_{\text{max}}$ along which 
the local variations of the color components are the highest. The absolute 
value of this maximum variation corresponds to the gradient module, which 
is evaluated from the horizontal and vertical first derivatives of each color 
component. Let us denote $I^R(x, y)$, $I^G(x, y)$, and $I^B(x, y)$ as the three 
($R$, $G$, $B$) color components of a pixel $P(x, y)$ with spatial coordinates $(x, y)$ 
in the color image $I$ and $\theta$ the gradient direction. Let $F$ be a variational 
function expressed as

$$F(\theta) = p \cos^2 \theta + q \sin^2 \theta + 2t \sin \theta \cos \theta,$$  
(50)

where $p$ is the squared module of the horizontal first-order partial derivatives, $q$ corresponds to the squared module of the vertical first-order partial derivatives, and $t$ is the mixed-squared module of the horizontal and vertical 
first-order partial derivatives:

$$p = \left( \frac{\partial I^R(x, y)}{\partial x} \right)^2 + \left( \frac{\partial I^G(x, y)}{\partial x} \right)^2 + \left( \frac{\partial I^B(x, y)}{\partial x} \right)^2,$$  
(51)

$$q = \left( \frac{\partial I^R(x, y)}{\partial y} \right)^2 + \left( \frac{\partial I^G(x, y)}{\partial y} \right)^2 + \left( \frac{\partial I^B(x, y)}{\partial y} \right)^2,$$  
(52)

and

$$t = \frac{\partial I^R(x, y)}{\partial x} \frac{\partial I^R(x, y)}{\partial y} + \frac{\partial I^G(x, y)}{\partial x} \frac{\partial I^G(x, y)}{\partial y} + \frac{\partial I^B(x, y)}{\partial x} \frac{\partial I^B(x, y)}{\partial y}.$$  
(53)

The direction of the color gradient that maximizes the function $F(\theta)$ is given 
by:

$$\theta_{\text{max}} = \frac{1}{2} \arctan \frac{2t}{p - q},$$  
(54)

and the gradient magnitude is therefore equal to $\sqrt{F(\theta_{\text{max}})}$.

This maximum variation is evaluated from the first horizontal and vertical 
derivatives of each color component [see Eq. (50)]. For this purpose, Deriche 
proposes to apply a recursive filter as a differential operator (Deriche, 1990).
This optimal filter is well suited for edge detection in noisy images (Stoclin, Duvieubourg and Cabestaing, 1997). The behavior of this filter is governed by a parameter $\alpha$. It is adjusted according to the expected filter performance, which is evaluated in terms of detection and localization. In order to extract well-connected edge pixels, detection of all the spatial color variations is more important than the accuracy of their localizations. The parameter $\alpha$ is then adjusted by the analyst so that the trade-off between detection and localization favors detection.

To extract thin edges from the gradient image obtained by this process, the edge pixels whose magnitude are local maxima along the gradient direction are retained. Then, an hysteresis thresholding scheme, with a low threshold $Th_l$ and a high threshold $Th_h$, provides a binary edge image. These parameters are adjusted by the analyst so that a maximum of pertinent edge pixels are detected.

To illustrate the behavior of Di Zenzo’s method, we propose its applications to the synthetic images in Figure 16. In the images of Figure 18, the detected edge pixels are represented by a white overlay. Different tests show that this approach is not very sensitive to the adjustment of the thresholds used by the hysteresis scheme on these images. The three images in Figure 18 show that the edge detection using the same parameter values provides different results with respect to the color space used.

Furthermore, one limitation of such an approach is the connectivity between the detected edge pixels (Ultré, Macaire and Postaire, 1996). Indeed, the region extraction is performed by a chaining scheme of edge pixels. This scheme assumes that the detected edge pixels that represent the boundary of a region are well connected (Rosenfeld and Kak, 1981). A postprocessing edge closing step has to be achieved when the edge pixels are not connected, which is why many authors prefer approaches based on region construction in order to segment color images.

C. Region Construction Based on a Spatial Analysis

These methods, which analyze the image plane to construct the regions in the image, can be divided into two approaches: the region growing method and the region merging method.

1. Segmentation by Region Growing

A region growing method consists of sequentially merging neighboring pixels of similar colors starting from initial seeds (Trémeau and Borel, 1997). Regions are expanded as far as possible with respective aggregating conditions.
FIGURE 18. Edge pixels of the images in Figure 16 detected by Di Zenzo’s method. The parameters $\alpha$, $Th_h$, and $Th_l$ are set to 1.0, 10, and 5, respectively. (a) $(R, G, B)$ color space. (b) $(Y', U', V')$ color space. (c) Triangular $(I, S, T)$ color space. (See Color Insert.)

The region growing method scans the image in some predetermined manner such as left-right top-bottom. The first analyzed pixel constitutes an initial region, named seed, and a next-neighboring pixel is examined. The color of this pixel is compared with the color of the already (but not necessarily) completed neighboring regions. If its color and the color of a neighboring region in progress are close enough (i.e., the Euclidean distance separating the colors is lower than a threshold), then the pixel is added to this region and the color of this region is updated. If there are several regions in progress whose colors are close enough, then the pixel is added to the region in progress whose color is the closest. However, when the colors of the two adjacent regions in progress are close enough, the two regions are merged and the pixel is added to the merged region. When the color of the examined pixel and that of any neighboring region are not close enough, a new region in progress is created (Moghaddamzadeh and Bourbakis, 1997).
Trémeau and Borel proposed three criteria to compare the color of a pixel with that of the region in progress (Trémeau and Borel, 1997). Let us denote \( R_i \), the subset of pixels that constitutes the region in progress, \( P \) the analyzed pixel, \( N(P) \) the set of neighboring pixels of \( P \), and \( I(P) \) the color of the pixel \( P \). We denote with the same manner the color of a pixel \( P \) with \((x, y)\) spatial coordinates by \( I(P) \) and \( I(x, y) \). The pixel \( P \) merges with one of its neighboring pixels \( Q \) that also belongs to \( R_i \): \((Q \in N(P) \text{ and } Q \in R_i)\), if the three following aggregating conditions are respected:

\[
\begin{align*}
&\|I(P) - I(Q)\| < T_1 \\
&\|I(P) - \mu_{N(P)}\| < T_2 \\
&\|I(P) - \mu_{R_i}\| < T_3,
\end{align*}
\]

where the thresholds \( T_j \) are adjusted by the analyst, \( \| \| \) is the Euclidean distance, \( \mu_{N(P)} \) represents the mean color of the pixels \( Q \) that belong to \( N(P) \), and \( R_i \), and \( \mu_{R_i} \) represents the mean color of the region \( R_i \) in progress.

This scheme grows the region in progress until the maximum of the differences between one color component of the region and that of the examined pixel is higher than 10 (Figure 19). The first pixel to be analyzed is located in the left-top side of the image, and the image is scanned in the left-right top-bottom order. In the images in Figure 19, the constructed regions are labeled with false colors. To show the influence of the color space on the quality of the region construction, the colors of the pixels are coded in the \((R, G, B)\) color space (see Figure 19a), in the \((Y', U', V')\) color space (see Figure 19b) and in the triangular \((I, S, T)\) color space (see Figure 19c).

These three images show that the growing scheme using the same parameter values provides different results with respect to the used color space. This scheme suffers from the difficulty of adjusting relevant thresholds for all the color spaces. Furthermore, one of the main drawbacks of these neighborhood-based segmentation methods is their sequential nature: the resulting regions depend on the order in which pixels are merged and on the selection of the initial seeds.

2. Segmentation by Region Merging

In order to avoid \textit{a posteriori} the oversegmentation problem, Trémeau and Colantoni proposed that the segmentation is achieved with two successive steps: a low-level step provides an oversegmented image by means of a region growing scheme (see Section III.C.1) and a high-level step merges adjacent regions with similar colors (Trémeau and Colantoni, 2000). He proposes that the high-level scheme is applied to a region adjacency graph that models a segmented image. A node of this graph represents a region, and an edge between two nodes corresponds to a pair of adjacent regions in the image.
Figure 19. Segmentation of the images in Figure 16 by region growing. The region grows until the max of the differences between one color component of the region and that of the examined pixel is higher than 10. (a) \((R, G, B)\) color space. (b) \((Y', U', V')\) color space. (c) Triangular \((I, S, T)\) color space. (See Color Insert.)

Figure 20b shows the adjacency graph obtained from the presegmented image of Figure 20a. The iterative analysis of the graph consists of merging two nodes relied by an edge when they respect a fusion criterion at each iteration step. More precisely, this procedure merges adjacent regions whose colors are close enough (Lozano, Colantoni and Laget, 1996).

In addition to the color homogeneity, Schettini proposes another criterion that considers the perimeters of two adjacent regions that are candidate for merging (Schettini, 1993). These criteria are adapted in case of uniform color regions but not in case of textured regions. This led Panjwani and Healey to model the regions by color Markov fields, which are designed for characterizing color textures (Panjwani and Healey, 1995).
These methods favor the spatial interactions between pixels and analyze the colors of the pixels only with respect to grow or merge regions. Since these methods require an a priori knowledge of the images in order to adjust the used parameters, many authors prefer to achieve a global analysis of the color distribution in the color space.

**D. Region Construction Based on a Color Space Analysis**

1. **Introduction**

Because the color of each pixel can be represented in a color space, it is also possible to analyze the distribution of pixel colors rather than examining the image plane. In the \((R, G, B)\) color space, a color point is defined by the color component levels of the corresponding pixel, namely, red \((R)\), green \((G)\), and blue \((B)\). It is generally assumed that homogeneous regions in the image plane create clusters of color points in the color space, each cluster corresponding to a class of pixels that share similar color properties.

Let us consider the synthetic image of Figure 21a which is composed of 6 regions, denoted \(R_i, i = 1, \ldots, 6\), and is identical to that of Figure 16a. The color points representing the pixels in the \((R, G, B)\) color space are shown in Figure 21b. To estimate the separability of the clusters, let us also examine Figures 21c–e, which show the clusters of color points of the image of Figure 21a projected onto the different chromatic planes \((R, G)\), \((G, B)\), and \((R, B)\), respectively. The regions \(R_1, R_2, R_3,\) and \(R_4\) give rise to well-separated clusters of color points, whereas those of the regions \(R_5\) and \(R_6\) form two overlapping clusters. It is difficult to identify these two clusters by the analysis of the color point distribution in the \((R, G, B)\) color space.
FIGURE 21. Clusters of color points of the image of figure (a) in the \((R, G, B)\) color space. (a) Color synthetic image. (b) Clusters of color points in the \((R, G, B)\) color space. (c) Clusters of color points in the \((R, G)\) chromatic plane. (d) Clusters of color points in the \((G, B)\) chromatic plane. (e) Clusters of color points in the \((R, B)\) chromatic plane. (See Color Insert.)
The classes of pixels are constructed by means of a cluster identification scheme that is performed either by an analysis of the color histogram or by a cluster analysis procedure. When the classes are constructed, the pixels are assigned to one of them by means of a decision rule and are mapped back to the original image plane to produce the segmentation. The regions of the segmented image are composed of connected pixels, which are assigned to the same classes. When the distribution of color points is analyzed in the color space, the procedures generally lead to a noisy segmentation with small regions scattered through the image. Usually, a spatial-based postprocessing is performed to reconstruct the actual regions in the image (Cheng, Jiang and Wang, 2002; Nikolaev and Nikolayev, 2004).

Several color spaces can be used to represent the colors of the pixels. Figure 22 shows the color points of the image in Figure 21a represented in the \((Y', U', V')\) and \((I, S, T)\) color spaces. By visually comparing this figure with Figure 21e, we see that the clusters representing the colors of the regions \(R_5\) and \(R_6\) form two overlapping clusters in the \((R, G, B)\) color space—one single cluster in the \((Y', U', V')\) color space and well-separated clusters in the \((I, S, T)\) color space. These two figures show that the color distribution and the cluster separability depend on the choice of the color space.

To illustrate this problem with a real example (soccer image segmentation), let us examine Figure 23a extracted from Vandenbroucke, Macaire and Postaire (2000b) where each of the images contains one soccer player. The players are regrouped into four classes \(\omega_j, j = 1, \ldots, 4\). Figure 23b shows
FIGURE 23. Clusters corresponding to different classes projected onto different color spaces. (a) Classes of pixels to be considered. (b) Color points in the $(R, G, B)$ color space. (c) Color points in Carron’s $(Y, Ch_1, Ch_2)$ color space. (d) Color points in the hybrid $(x, Ch_2, I_3)$ color space. (See Color Insert.)
COLOR SPACES AND IMAGE SEGMENTATION

the colors of pixels of each class projected onto the \((R, G, B)\) color space. Clusters corresponding to the classes \(\omega_2, \omega_3,\) and \(\omega_4\) are not well separated in the \((R, G, B)\) color space while they are more compact in the \((Y, Ch_1, Ch_2)\) Carron’s color space (defined in Section II.B.2.d; see Figure 23c) and well separated in the \((x, Ch_2, I3)\) hybrid color space (defined in Section II.B.2.d; see Figure 23d). This example shows that the selection of the color space is crucial for the construction of pixel classes.

The key of the segmentation problem based on color-spaced analysis consists of the construction of the pixel classes. The segmentation of color image based on pixel classification can be divided into four groups: (1) the analysis of histograms (see Section III.D.2), (2) the analysis of the 3D color histogram (see Section III.D.3), (3) the clustering approaches, which take into account only color distribution (see Section III.D.4), and (4) the spatial-color classification methods, which simultaneously consider the spatial interactions between the pixels and the color distribution (see Section III.D.5).

2. Analysis of One-Dimensional Histograms

Many statistical procedures seek to accomplish class construction by detecting modes of the color space. The modes are characterized by high local concentration of color points in the color space separated by valleys with a low local concentration of color points. For this purpose, they estimate the probability density function (pdf) underlying the distribution of the color points (Devijver and Kittler, 1982). Assuming that each mode of the color space corresponds to the definition domain of a pixel class, the color image segmentation can be viewed as a pdf mode detection problem.

The most widely used tools for the approximation of the pdf are the 1D histograms denoted \(H^k[I]\), of the color component images \(I^k, k = R, G, B,\)
FIGURE 24. One-dimensional histograms of the image in Figure 16a. (a) One-dimensional histogram of the color component image $I^R$. (b) One-dimensional histogram of the color component image $I^G$. (c) One-dimensional histogram of the color component image $I^B$. (See Color Insert.)
FIGURE 25. Mode detection in the chromatic plane \((G, B)\) by the analysis of the 1D histograms \(H_G[I]\) and \(H_B[I]\) of the image in Figure 21a. (See Color Insert.)

when the \((R, G, B)\) color space is used. The bin \(H^k[I](n)\) contains the number of pixels \(P\) whose color component level \(I^k(P)\) is equal to \(n\).

By examining Figure 24 (which contains three 1D histograms \(H^k[I], k = R, G, B, \) in the image of Figure 21a), we see that the color component distributions of the different regions create peaks of the 1D histograms. Furthermore, the histograms show that the pixel classes may not be equiprobable.

The peaks of the corresponding 1D histogram are determined by means of searching thresholds that delimit them (Schettini, 1993). When the thresholds are determined, the color space is partitioned into parallelepipedic boxes by means of the Cartesian product of the intervals delimited by these thresholds. An analysis of the populations of pixels whose color points fall into these boxes allows us to identify the modes of the color space (Figure 25).
A peak of a 1D histogram may contain the color component levels of pixels that constitute different regions of the image (see Figure 24a). Pixels of each region must be regrouped into a specific class of pixels defined by a mode of the color space. Therefore, the methods of mode detection analyzes the case when one peak of the 1D histogram corresponds to several modes of the color space (Ohlander, Price and Reddy, 1978).

For this purpose, an iterative analysis of the 1D histograms is achieved (Busin et al., 2005). At each iteration step, this procedure constructs one class of pixels. The procedure looks for the most significant mode to construct one class, according to different criteria, such as the population size. The pixels assigned to this class and which are connected in the image constitute one of the reconstructed regions in the segmented image.

The pixels assigned to the so-constructed class are extracted from the color image so that they are not taken into account at the next iteration steps of the procedure. The iterative procedure stops when only a few pixels that are not assigned to any constructed classes remain. When the iterative procedure stops, the pixels have not been assigned to any class, that can be assigned to one of the constructed classes by means of a specific decision rule.

Figure 26 illustrates the iterative segmentation of the Hand image in Figure 26a (Busin et al., 2004). The image in Figure 26b shows pixels assigned at the first class built by an analysis of the 1D color histograms. White-labeled pixels in the image of Figure 26c are not assigned to this class and thus are considered by the second iteration step in order to build the second class. Images of Figures 26d, f, h, and j show the successively built classes, whereas images of Figures 26e, g, i, and k contain the pixels to be analyzed by the successive iteration steps. The final segmented image of Figure 26l with pixels labeled to the five built classes shows that this scheme provides satisfying results.

Actually, the performance of the mode extraction scheme is strongly affected by the choice of color space, implicitly indicating that the performance depends on the color distribution of the analyzed image (Lim and Lee, 1990). That finding led Tominaga to perform the Karhunen–Loeve transformation of the $(R,G,B)$ color components (i.e., the PCA) in order to construct the classes (Tominaga, 1992). This procedure transforms the $(R,G,B)$ color space into the $(X_1, X_2, X_3)$ principal component space. The histogram $H[I]^{X_1}$ of the most discriminating component $X_1$ is first analyzed. If this histogram is a multimodal one, the most significant peak is detected. Then a new Karhunen–Loeve transformation is achieved with the pixels that do not belong to the built class. If the histogram of the first component contains only one single peak, then Tominaga proposes analysis of the second component $X_2$. This iterative scheme stops when the histograms
COLOR SPACES AND IMAGE SEGMENTATION

FIGURE 26. Iterative segmentation of the Hand image. (a) Hand image. (b) Pixels assigned to the class built at the first iteration. (c) Pixels to be analyzed after the first iteration. (d) Pixels assigned to the class built at the second iteration. (e) Pixels to be analyzed after the second iteration. (f) Pixels assigned to the class built at the third iteration. (g) Pixels to be analyzed after the third iteration. (h) Pixels assigned to the class built at the fourth iteration. (i) Pixels to be analyzed after the fourth iteration. (j) Pixels assigned to the class built at the fifth iteration. (k) Pixels to be analyzed after the fifth iteration. (l) Segmented image. (See Color Insert.)
Figure 27. Color component images resulting from the Karhunen–Loève transformation of the image in Figure 21a. (a) Color component image $I^{X_1}$. (b) Color component image $I^{X_2}$. (c) Color component image $I^{X_3}$.

are empty or when the histogram of the third component $X_3$ contains one single peak.

Figure 27 shows the three color component images $I^{X_1}$, $I^{X_2}$, and $I^{X_3}$ resulting from the Karhunen–Loève transformation of the image in Figure 21a. The $I^{X_1}$ color component image contains the greatest amount of information. Tominaga proposes to first extract the most significant peak from the histogram $H^{X_1}[I]$ (Figure 28a). Pixels of the image $I^{X_1}$ whose $X_1$ levels fall into the interval delimiting this peak are shown with a false color in the image of Figure 29a, where unassigned pixels are labeled as black. By examining this image, we see that the pixels of the background constitute the first built class. The scheme is iterated to yield the final satisfying segmentation shown by Figure 29b.
3. Analysis of the Three-Dimensional Color Histogram

Figure 24 shows that a peak of each 1D histogram of a color image can reflect several classes. Even when the mode detection is separately performed on the three 1D histograms, the appropriate number of classes may not be detectable because of this superposition. Moreover, since the thresholds obtained from the analysis of the 1D histograms are used to partition the color space into parallelepipedic boxes, the boundaries between classes cannot be adequately determined, sometimes yielding false segmentation results.

Because color is 3D information, several authors use the 3D color histogram for approximating the pdf. The color histogram is composed of bins whose coordinates are the three color components. Each bin $H[I](C)$ indicates the number of pixels $P$ whose color $I(P)$ is equal to the color $C$. 

Figure 28. One-dimensional histograms of the $I^{X_1}$, $I^{X_2}$, and $I^{X_3}$ color component images in Figure 27. (a) One-dimensional histogram of the image in Figure 27a. (b) One-dimensional histogram of the image in Figure 27b. (c) One-dimensional histogram of the image in Figure 27c. (See Color Insert.)
Since the colors of pixels are usually quantified with 256 levels, a 3D histogram requires a considerable amount of computer memory. That explains why only a few authors have analyzed the 3D histogram for color image segmentation (Lambert and Macaire, 2000). A simple solution consists of color quantization by considering only the highest bits for each component in order to reduce the computational complexity. This solution yields poor results, because this quantization phase leads to a very crude segmentation. Another approach consists of mode detection by thresholding the bins. Nevertheless, the determination of the threshold is very delicate, which is why several mode detection methods use specific morphological transformations of the color histogram to enhance the modes (Park, Yun and Lee, 1998; Shaffarenko, Petrou and Kittler, 1998).

Let us consider the image of Figure 30a, which is made of two concentric disks lying on a distinct background. This image is corrupted by a noncorrelated Gaussian noise, with a standard deviation equal to 9, that is independently added to the color components. We can visually distinguish three regions, each being characterized by a specific shade of green. Therefore, pixels of this image must be grouped into three classes. In order to simplify the illustration, the blue component has been set to 0 throughout the image. The color histogram of this image is plotted in Figure 30b over a restricted domain of the chromatic plane \((R, G)\) where its values are non-zero. Since the distributions of those colors overlap, the three color modes are barely detected by an automatic processing of this histogram.

Gillet et al. define a mode as a parallelepipedic box of the color space where the histogram function is locally concave while reaching values high
FIGURE 30. Mode detection by convexity analysis of the color histogram. (a) Image. (b) Histogram. (c) Modes of the histogram detected by the convexity test; their identification is illustrated by false colors. (d) Prototypes of the three built pixel classes. (e) Pixels assigned to the three classes. (See Color Insert.)
enough to be relevant (Gillet et al., 2002). In order to assign a “concave” or “convex” label to each color \( C \) of the color space, a local convexity test is performed on the histogram. Let \( D(C, l) \) be a cubic box, centered at the color \( C = [C^R, C^G, C^B]^T \) in the \((R, G, B)\) color space, whose edges of length \( l \) (an odd integer) are parallel to the color component axes. The box \( D(C, l) \) therefore includes all the color points \( c = [c^R, c^G, c^B]^T \) such that 
\[
C^k - \frac{l-1}{2} \leq c^k \leq C^k + \frac{l-1}{2}, \quad k = R, G, B.
\]
The mode detection is based on the property stating that, at each color point \( C \) where a trivariate function \( f \) is locally concave, the average value of \( f \) processed over a box \( D(C, l) \) is a decreasing function of \( l \) (Postaire and Vasseur, 1980). The histogram local convexity is therefore evaluated at each color point \( C \) by examining the variation of the average histogram, denoted \( \mu \), according to the size of the cubic box centered at \( C \). First, we average the histogram over the box \( D(C, l_1) \):
\[
\mu_1(C) = \frac{\sum_{c \in D(C, l_1)} H(c)}{(l_1)^3}.
\]
We then compute the second average value \( \mu_2 \) using a slightly larger box \( D(C, l_2) \):
\[
\mu_2(C) = \frac{\sum_{c \in D(C, l_2)} H(c, l)}{(l_2)^3}, \quad \text{with } l_2 > l_1.
\]
Empirically, length \( l_1 \) is set to 5, while length \( l_2 \) is set to \( l_1 + 2 \). In these conditions, if \( \mu_1(C) > \mu_2(C) \), the histogram is considered as locally concave. If this condition is satisfied and if \( H[I](C) \) is high enough (i.e., higher than a threshold adjusted by the analyst), we decide that the color point \( C \) belongs to a mode.

This mode detection algorithm is applied to the 2D histogram of the synthetic image (see Figure 30b). Because small modes may be detected by convexity analysis, a binary morphological opening is applied to the modes to preserve only the significant ones—modes whose volume is larger than that of the structuring element that is used. Figure 30c shows the three detected modes in the chromatic plane \((R, G)\).

The detected modes then are identified using a connected component analysis procedure in the color space. Each mode identified in this manner is the definition domain of a pixel class. All the pixels whose colors are located in an identified mode define the prototypes of the corresponding class and are assigned to it. Figure 30d shows the prototypes of the three pixel classes labeled as false colors, constructed by the analysis of the histogram in Figure 30b. The black-marked pixels are the ones that are not yet assigned. This image shows that the prototypes correctly represent the background and
the two concentric shapes. This example illustrates that the convexity analysis of the histogram efficiently constructs the pixel classes even when the color distributions of the different regions strongly overlap. The next classification procedure step consists of assigning all the pixels that are not yet labeled by use of a specific decision rule in order to produce the final segmented image in Figure 30e (Vannoorenberghe and Flouzat, 2006).

4. Clustering-Based Segmentation

The multithresholding methods based on histogram analysis require the computation of the histogram, which is both time and memory space consuming. For this reason, many authors have proposed a looking for clusters of color points in the color space.

Among the clustering techniques based on the least sum of squares criterion, the c-means algorithm is one of the most widely used methods for clustering multidimensional data (Ismail and Kamel, 1989). This iterative approach requires the analyst to adjust the desired number $N_\omega$ of pixel classes.

Let $C_j$ denote the gravity center of the cluster that is associated with the class $\omega_j$. This approach tends to minimize the global within-class dispersion, which is defined by

$$S_W = \sum_{j=1}^{N_\omega} \sum_{P \in \omega_j} \| I(P) - C_j \|.$$  \hspace{1cm} (57)

At the initial iteration step, the locations of the $N_\omega$ gravity centers are randomly fixed in the color space. At each iteration step denoted $t$ of the process, the $N_\omega$ gravity centers $C_j(t)$ are updated via this construction scheme:

- Each pixel $P$ is assigned to the class $\omega_{j0}$ whose gravity center $C_{j0}(t)$ is the nearest one to the color $I(P)$ in the color space:

$$\| C_{j0}(t) - I(P) \| = \min_{j=1, \ldots, N_\omega} (\| C_j(t) - I(P) \|).$$

- The gravity center $C_j(t)$ of each class $\omega_j$ is updated by taking into account the pixels assigned to $\omega_j$:

$$C_j(t) = \sum_{P \in \omega_j} I(P).$$

- The variation denoted $\varepsilon$ of the gravity centers between the preceding $(t-1)$ and current $(t)$ steps is determined by:

$$\varepsilon = \sum_{j=1, \ldots, N_\omega} \| C_j(t) - C_j(t - 1) \|.$$
• If $\varepsilon$ is higher than a threshold, then the process achieves a new iteration step.

Whatever the initialization of the gravity centers, the scheme converges to a local minimum of $S_W$. In this way, the algorithm stops at a local optimum, which can be far away from the real global optimum, especially when large numbers of pixels and clusters are involved. Since the convergence and the classification results depend on the initialization step, different approaches are proposed as follows.

• The initial locations of the gravity centers are randomly selected (see Figure 31a)
During a supervised learning, the initial locations of the gravity centers are interactively selected so that they are uniformly scattered in the color space (see Figure 31b).

The isodata method adds two rules to determine the number $N_\omega$ of classes: if the within-class dispersion of a class is greater than a threshold, then the class is split into two different classes. If the distance between two gravity centers is too low, then two corresponding classes are merged into one class (Takahashi, Nakatani and Abe, 1995).

In order to minimize the within-class dispersion $S_W$, Uchiyama and Arbib constructed the $N_\omega$ classes by means of an iterative process based on competitive learning (Uchiyama and Arbib, 1994). This process, which does not require an initialization step, converges to approximate the optimum solution. We apply the competitive learning scheme to the image in Figure 21a (see Figure 31c). By examining the segmented image, we notice that the regions $R_5$ and $R_6$ are merged into one class, because the color component distributions overlap. The regions $R_4$ and $R_2$ are represented by one class, because the regions with the lowest population (in our case $R_4$) are not taken in account by the competitive learning scheme.

Liew et al. apply the fuzzy c-means, a fuzzy classification scheme that considers that each pixel $P$ belongs to each class $\omega_j$ according to a membership degree denoted $U_j(P)$ ranging between 0 and 1 (Liew, Leung and Lau, 2000). This method constructs the classes by minimizing the $m$-fuzzy within-class dispersion $J(m)$:

$$J(m) = \sum_P \sum_{j=1}^{N_\omega} (U_j(P))^m \left\| C_j - I(P) \right\|, \quad (58)$$

where

$$(U_j(P))^m = \left( \sum_{i=1}^{N_\omega} \left( \frac{\|I(P) - C_i\|}{\|I(P) - C_j\|} \right)^{2/(m-1)} \right)^{-1}. \quad (59)$$

The color components of the gravity center $C_j$ associated with the class $\omega_j$ are defined by:

$$C_j^k = \frac{\sum_P (U_j(P))^m I^k(P)}{\sum_P (U_j(P))^m}, \quad \text{with } k = R, G, B. \quad (60)$$

When the gravity centers of the classes are estimated, the membership degrees of each pixel to the classes are determined. The pixel $P$ is assigned to the class for which the membership degree is the highest (Scheunders, 1997).
5. Spatial-Color Classification

Let us consider the synthetic image of Figure 32a, which is composed of six regions, denoted $R_i$, $i = 1, \ldots, 6$, with different sizes and shapes. The color distribution in this synthetic image is different from that of Figure 16a. For the sake of illustration simplicity, the blue level of the pixels is set to 0. The projections of the color points representing the regions onto the $(R, G)$ chromatic plane are displayed in Figure 32b. The regions $R_1$ and $R_2$ give rise to well-separated clusters of color points, whereas those of regions $R_3$ and $R_4$ form two overlapping clusters. Note that as the color points of regions $R_5$ and $R_6$ constitute a single cluster, they cannot be discriminated by the analysis of the color point distribution in the $(R, G)$ chromatic plane. This image shows that there is not always a one-to-one correspondence between the regions in the image and the clusters of color points in the color space. Classical classification schemes cannot identify clusters of color points in the color space that correspond to several different regions in the image. A class construction procedure that would consider simultaneously the color properties of pixels as well as their spatial arrangement in the image could be appealing to identify pixel classes corresponding to the actual regions (Balasubramanian, Allebach and Bouman, 1994; Orchard and Bouman, 1991).

The JSEG algorithm proposed by Deng and Manjunath (2001) separates the segmentation procedure into two successive steps. In the first step, pixels are classified in the color space without considering their spatial distribution. Then the colors of the pixels are replaced by their corresponding color class
COLOR SPACES AND IMAGE SEGMENTATION

labels, thus forming a class-map of the image. At each pixel the denoted JSEG criterion, which depends on the dispersion of the labels associated with its neighboring pixels, is measured. In the second step, the application of this criterion in the class-map yields an image in which high and low values correspond to possible boundaries and interiors of regions, respectively. A region growing method (see Section III.C.1) using this criterion provides the final segmentation.

Cheng, Jiang and Wang (2002) proposes a fuzzy homogeneity approach to take into account simultaneously the color and spatial properties of the pixels. Their segmentation scheme also is divided into two steps. In the first step, each of the three color component images $I^k$, $k = R, G, B$, is analyzed to aggregate its pixels into classes. For this purpose, the authors introduce the homogram that is defined for each color component level as the mean measure of homogeneity degrees described in Cheng et al. (1998) between pixels with this level and their neighbors. A fuzzy analysis detects the peaks of each homogram in order to identify the prominent classes of pixels for each color component image. The prominent classes built by the analyses of the three color component images are combined to form the classes of pixels of the color image. When some of these classes contain too few pixels, the procedure may lead to oversegmentation. This problem is partially addressed in the second step by merging neighboring classes. This interesting approach is very sensitive to the adjustment of the parameters required for determining the peaks of each homogram in the first step and for merging classes in the second step.

Another solution consists of estimating the gradient of the probability density of color occurring jointly in spatial and color domains of the color image (Comaniciu and Meer, 2002). Each pixel is associated with the closest local mode of the density distribution determined by means of the nonparametric “mean-shift” estimation. The quality of the segmentation depends on the precise adjustment of two parameters that control the resolutions in the spatial and in the color domains.

We propose to segment the image of Figure 32a by use of the EDISON software developed by Comaniciu and Meer (available at the web address: http://www.caip.rutgers.edu/riul/research/code.html).

For this purpose, the pair of parameters $h_s$ and $h_r$, which are the side lengths of the windows used to compute the gradient in the spatial and color domains, are set to the values suggested by the authors for segmenting most color images. The parameter $M$, which corresponds to the minimal population sizes of the classes, is set to 1 so that all the classes, even those with small population sizes, are selected.

Figures 33a and b yield the best results obtained by the “mean-shift” analysis. Figure 33a shows that when the side length of the color domains is
(a) \( h_s = 8, h_r = 4, M = 1 \). (b) \( h_s = 8, h_r = 8, M = 1 \). (See Color Insert.)

low (\( h_r \) is set to 4), the procedure based on the “mean-shift” method succeeds in separating the two clusters of color points representing the two green concentric disks. However, it also splits the cluster of color points representing the background of the image, which is segmented into several different regions. When the side length of the color domains increases (\( h_r \) is set to 8), the procedure fails to separate the two clusters of color points corresponding to the two green concentric disks (see Figure 33b). It is difficult, or even impossible, to adjust the parameters required by this procedure in order to identify the two classes of pixels representing the two concentric green disks without splitting the other clusters corresponding to the other regions of the image. However, the mean-shift method does not require adjustment of the desired number of pixel classes. When the number of pixel classes is given, a postprocessing step to the mean-shift method would lead to good segmentation results.

Macaire, Vandenbroucke and Postaire (2006) assume that each region can be considered as a subset of strongly connected pixels with homogeneous colors. Hence, they propose selecting parallelepipedic boxes in the color space that define subsets (named color subsets) of strongly connected pixels in the image with as homogeneous colors as possible. In other words, pixels in the image whose color points fall into such boxes (named color domains) of the color space constitute color subsets, so that each color domain is associated with a color subset. The color and connectedness properties of the color subsets associated with the considered color domains are simultaneously taken into account by the pixel class construction scheme. To measure these properties, they introduced the concept of spatial-color compactness degree (SCD) of a color subset, which measures the spatial arrangement of its pixels.
in the image plane (Fontaine, Macaire and Postaire, 2000) and the dispersion of the color points representing its pixels in the color space. In order to select the color domains that define color subsets corresponding to the actual regions in the image, the pixel class construction procedure looks for maxima of the SCDS in the color space.

The labeled image in Figure 34b shows how the regions of the image in Figure 32a are reconstructed from the identified class color domains displayed in Figure 34a according to specific rules (Macaire, Vandenbroucke and Postaire, 2006). In Figure 34b the boundaries between the reconstructed regions are overlaid in black. This image shows that the six regions are well reconstructed by the SCD analysis. The overlapping of the color point distributions of the regions $R_3$ and $R_4$ explains why a few pixels of the region $R_3$ are misclassified. Note that even with the strong overlapping of the color point distributions of the two regions $R_5$ and $R_6$, most pixels of the regions corresponding to the two green concentric disks are correctly classified. This result shows that this method is able to handle unequiprobable and overlapping classes of pixels.

**E. Summary**

In this section, we have attempted to present classical color segmentation methods. Each method is adapted to a type of color image, so a method that provides a satisfying segmentation of all images does not exist. Several
IV. Relationships between Segmentation and Color Spaces

A. Introduction

The first part of this chapter has shown that the information of color can be coded by several color spaces that respect their own specific properties. The second part of the chapter has shown that color image segmentation could be divided into two main approaches according to whether the pixels are analyzed in the image plane or in a color space. If the pixels are analyzed in the image plane, then two dual type of methods are available. On the one hand, the segmentation can be produced by an edge detection scheme. On the other hand, the segmentation can be produced via region construction analysis.

Several authors have studied the influence of color spaces for color image segmentation with the objective to determine if there exists a color space that improves the quality reached by color image segmentation methods. To determine whether this “best” color space exists, they estimated the results produced by color image segmentation methods via evaluation methods.

First, Ohlander, Price and Reddy (1978) and Ohta, Kanade and Sakai (1980) visually estimated the quality of color segmentation. Because it is very difficult to evaluate the segmentation of a natural scene, several authors adopted the qualitative evaluation method as the most reliable one.

Consequently, this part of text is divided as follow. Section IV.B presents the primary evaluation methods used for color image segmentation by edge detection, and Section IV.C presents the main evaluation methods used for color image segmentation by region construction. For both sections, quantitative evaluation methods with ground truth, which correspond to unsupervised evaluations, and quantitative evaluation methods without ground truth, which correspond to supervised evaluations, are presented in the subsections.

To illustrate the relationships between the color spaces and segmentation, Section IV.D details two pixel classification methods that determine the most discriminating color space (hybrid color space or classic color spaces).
B. Edge Detection

In order to evaluate the result provided by an edge detection segmentation, the most common method used in the literature consists of comparing the result of the segmentation $I$ with the ground truth $I_{\text{ref}}$ (also called the reference segmentation or gold standard). These comparison methods are presented in Section IV.B.1. Since very few methods based on an evaluation method without ground truth are available and used in the literature for color edge detection segmentation methods, none of them are presented here. Finally, a table summarizing the selected color spaces by means of these criteria that provide the best segmentation results is presented in Section IV.B.2.

1. Quantitative Evaluation Methods with Ground Truth

The most common way to evaluate the segmented image computed via an edge detection segmentation algorithm consists of evaluating the extracted edges of the segmented image $I$ with the edges of the ground truth image $I_{\text{ref}}$. For this type of evaluation, two primary methods are available: a probabilistic error approach and a spatial approach.

For both approaches, the discrepancy between a binary image computed by means of an edge detection segmentation algorithm consists of evaluating the extracted edges of the segmented image with the edges of the ground truth (Román-Roldán et al., 2001). To detail these both approaches, we define the following:

• **Mistake** — The discrepancy between the detected edge and the real edge is due to individual discrepancies arising from pixel to pixel, which are referred to as mistakes. These may be of two kinds:
  1. **Bit or overdetected edge pixel** — A mistake due to excess, when a pixel is erroneously defined as an edge pixel.
  2. **Hole or subdetected edge pixel** — A mistake due to failure of the method to identify an edge pixel as such.

• $N_b$ — The number of bits in the segmented image.
• $N_h$ — The number of holes in the segmented image.
• $N_e$ — The number of edge pixels in the ground truth.
• $N$ — The number of pixels in the segmented image.

a. **Probabilistic Error Approach.** The first discrepancy measure established by Peli and Malah (1982) is called error probability and is denoted $P_e(I, I_{\text{ref}})$ hereafter. This rate is based on a statistical comparison between the number $N_b$ of bits and the number $N_e$ of edge pixels in the ground truth image $I_{\text{ref}}$.
and can be computed as:

$$P_e(I, I_{\text{ref}}) = \frac{N_b}{N_e}.$$ (61)

Another probabilistic discrepancy measure introduced by Lee, Chung and Park (1990)—denoted $P_E(I, I_{\text{ref}})$ hereafter—is based on a two-class problem (denoted $\omega_1$ and $\omega_2$). In the edge detection segmentation methods, the class of pixels $\omega_1$ corresponds to those that are assigned as edge pixels, whereas the class of pixel $\omega_2$ corresponds to those that are assigned as nonedge pixels. This discrepancy measure is better than the measure proposed by Peli and Malah (1982) because both types of errors (holes and bits) are taken into account. The $P_E(I, I_{\text{ref}})$ probabilistic discrepancy measure based on the classical classification error measure can be expressed as:

$$P_E(I, I_{\text{ref}}) = P(\omega_1) \times P(\omega_2|\omega_1) + P(\omega_2) \times P(\omega_1|\omega_2) = \frac{N_h + N_b}{N},$$ (62)

where:

- $P(\omega_1) = \frac{N_e}{N}$
- $P(\omega_2) = \frac{N - N_e}{N}$
- $P(\omega_2|\omega_1) = \frac{N_h}{N_e}$
- $P(\omega_1|\omega_2) = \frac{N_b}{N - N_e}$.

b. Spatial Approach. Because the Peli and Lee discrepancy measures do not consider the spatial distance between the edges detected in the segmented image and the edges from the ground truth, most authors prefer Pratt’s discrepancy measure based on the mean-square distance called figure of merit (FOM) (Pratt, 1978) and defined as:

$$\text{FOM}(I, I_{\text{ref}}) = \frac{1}{M} \sum_{i=1}^{N_e-N_h+N_b} \frac{1}{1 + \alpha d(i)^2},$$ (63)

where $\alpha$ is a scaling parameter, $M = \max(N_e, N_e - N_h + N_b)$, and $d(i)$ is the distance from the $i$th edge pixel in the segmented image to its exact location in the ground truth. The measure is a normalized one so that $\text{FOM} = 1$ indicates a perfect segmentation. Pratt’s measure is the most frequently used discrepancy measure, even though it has been modified by Strastersa and Gerbrands (1991) to improve its accuracy with respect to bits and holes. Another interesting discrepancy measure proposed by Román-Roldán et al. (2001) requires a training procedure to fix the parameters used.
COLOR SPACES AND IMAGE SEGMENTATION

TABLE 6

<table>
<thead>
<tr>
<th>Reference</th>
<th>Evaluation method</th>
<th>Candidate color spaces</th>
<th>Selected color spaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rakotomalala et al. (1998)</td>
<td>Pratt’s discrepancy measure and</td>
<td>(R, G, B), (X, Y, Z),</td>
<td>(L*, a*, b*)</td>
</tr>
<tr>
<td></td>
<td>probability error</td>
<td>(I1, I2, I3), (A, C1,</td>
<td>(L*, u*, v*)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2), (I, S, T), (L*, a*, b*),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(L*, u*, v*)</td>
<td></td>
</tr>
<tr>
<td>Wesolkowski, Jernigan and</td>
<td>Pratt’s discrepancy measure</td>
<td>(R, G, B), (X, Y, Z),</td>
<td>(h1, h2, h3)</td>
</tr>
<tr>
<td>Dony (2000)</td>
<td></td>
<td>(L*, a*, b*),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(L*, u*, v*),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(r, g, b),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(l1, l2, l3), (h1, h2, h3)</td>
<td></td>
</tr>
</tbody>
</table>

2. Evaluation Methods for Color Image Segmentation by Edge Detection

Several authors apply these methods to evaluate the accuracy of the color image segmentation by edge detection for several color spaces. For the sake of simplicity, the results of these works are presented in Table 6.

Table 6 indicates that no color space is actually adapted to edge detection. Moreover, the conclusions of the authors cited are contradictory—Rakotomalala et al. (1998) select (L*, a*, b*) or (L*, u*, v*) spaces, whereas Wesolkowski, Jernigan and Dony (2000) do not select this as a color space in which the edge detection is efficient. The next text section details the primary evaluation methods for region construction segmentation methods to determine whether an efficient color space exists for such methods.

C. Region Construction

Evaluation methods for region construction segmentation methods can be divided into two main approaches according to whether a ground truth is available or not. The primary evaluation methods based on discrepancy measure with a ground truth are presented in Section IV.C.1. Without ground truth, it is impossible to compare the segmented images. Therefore, the unsupervised evaluation methods that attempt to mimic human sensitivity are presented in Section IV.C.2. Finally, the selected color spaces by means of these criteria that provide the best segmentation results are summarized in a table in Section IV.C.3.

1. Quantitative Evaluation Methods with Ground Truth

Supervised evaluation methods of color image segmentation by region construction can be divided into two families. The first family is based on the
discrepancy area of the regions between the segmented image $I$ and the ground truth $I_{\text{ref}}$. For this method, correct classification rate is the most often used (Sonka, Hlavac and Boyle, 1994). The second family consists of evaluating the color discrepancy between the pixels of the segmented image $I$ and the ground truth $I_{\text{ref}}$. For this method, two approaches are available. The first consists of analyzing color discrepancies between the segmented image $I$ and the ground truth $I_{\text{ref}}$. For this evaluation method, the color of each pixel in both the segmented image $I$ and the ground truth $I_{\text{ref}}$ must be the mean color of the pixels that belong to the same region. The most current evaluation methods for this approach are the mean square error (MSE) and the peak SNR. The second approach consists of evaluating the probability error of mislabeling pixels between the segmented image $I$ and the ground truth $I_{\text{ref}}$. Receiver operator characteristic (ROC) curves and probabilistic error approach often are used for a such analysis.

a. Correct Classification Rate. Generally, a similarity measure is used based on the correct classification rate between a segmented image $I$ and a ground truth $I_{\text{ref}}$ (Sonka, Hlavac and Boyle, 1994). According to the notation by Chabrier et al. (2006), the first step of the method is computing the superposition table $T(I, I_{\text{ref}})$ as follows:

$$T(I, I_{\text{ref}}) = \left[ \text{card}\{R_i \cap R_{\text{ref}}^j\}, i = 1, \ldots, N_R, j = 1, \ldots, N_{\text{ref}} \right],$$  \hspace{1cm} (64)

where $\text{card}\{R_i \cap R_{\text{ref}}^j\}$ is the number of pixels that belong to both the region $R_i$ in the segmented image $I$ and to the region $R_{\text{ref}}^j$ in the ground truth $I_{\text{ref}}$ while $N_R$ and $N_{\text{ref}}$ are the number of regions in the image $I$ and $I_{\text{ref}}$, respectively. From this superposition table, only the $C'$ couples that maximize $\text{card}(R_i \cap R_{\text{ref}}^j)$ are selected to compute correct classification rate. Thus, the dissimilarity measure that is a normalized one is computed, so that a measure close to 0 means that the images $I$ and $I_{\text{ref}}$ strongly share similar regions, whereas a measure close to 1 means that the images $I$ and $I_{\text{ref}}$ are very dissimilar. This measure, denoted $\text{CCR}(I, I_{\text{ref}})$, is computed as follows:

$$\text{CCR}(I, I_{\text{ref}}) = \frac{\text{card}(I) - \sum_{C'} \text{card}(R_i \cap R_{\text{ref}}^j)}{\text{card}(I)}. \hspace{1cm} (65)$$

b. Mean Square Error. The mean square error, denoted as $\text{MSE}(I, I_{\text{ref}})$, expresses the delineation accuracy and region homogeneity of the final partitioning between the segmented image $I$ and the ground truth $I_{\text{ref}}$. The lower the values of the $\text{MSE}(I, I_{\text{ref}})$, the better the segmentation results.
The $MSE(I, I_{ref})$ can be computed as

$$MSE(I, I_{ref}) = \sum_{P \in I} \| I_{ref}(P) - I(P) \|,$$  \hspace{1cm} (66)

where $I_{ref}(P)$ is the color mean of the region to which the pixel $P$ belong.

c. Peak SNR. Peak signal to noise ratio, denoted as $PSNR(I, I_{ref})$, is close to the $MSE(I, I_{ref})$ evaluation method and expresses the same concept. The main difference lies in the interpretation of the result provided by the $PSNR(I, I_{ref})$. Indeed, the higher the $PSNR(I, I_{ref})$, the better the segmentation result. The PSNR of an $X \times Y$ image (measured in decibels) can be computed as

$$PSNR(I, I_{ref}) = 10 \times \log \left( \frac{255^2 \times X \times Y \times \text{card}(k)}{\sum_{P \in I} \| I_{ref}(P) - I(P) \|} \right).$$  \hspace{1cm} (67)

d. Probabilistic Error Approach. The discrepancy measure introduced in Section IV.B.1 has been extended for the $\omega_n$ classes problem. In the case of an $\omega_n$ class problem, each $\omega_n$ class corresponds to a specific region, usually determined by a pixel classification, as in Lim and Lee (1990) and Park, Yun and Lee (1998), for example. The $P_E(I, I_{ref})$ probabilistic discrepancy measure can be calculated for an $\omega_n$ class problem as follows:

$$P_E(I, I_{ref}) = \sum_{j=1}^{\mathcal{R}_n} \sum_{\substack{i=1 \atop i \neq j}}^{\mathcal{R}_n} P(\mathcal{R}_i | \mathcal{R}_j) \times P(\mathcal{R}_j).$$  \hspace{1cm} (68)

e. ROC Curves. ROC curves are commonly used to present results for binary decision problems. By comparison of the labeled pixels provided by a classification method, four cases are possible as shown in the confusion matrix shown in Table 7.

The ROC curves can be represented by plotting in a 2D space where the $x$-axis corresponds to the fraction of false positives (false-positive rate, or FPR) and the $y$-axis corresponds to the fraction of true positives (true positive rate,

<table>
<thead>
<tr>
<th>Pixels assigned to the class $\omega_i$ ($i = 1, 2$) in $I_{ref}$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>True positive (TP)</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>False negative (FN)</td>
<td>True negative (TN)</td>
</tr>
</tbody>
</table>
TPR) (Liu and Shriberg, 2007). The TPR and FPR are expressed as:

\[ TPR = \frac{TP}{TP + FN}, \quad (69) \]

\[ FPR = \frac{FP}{FP + TN}. \quad (70) \]

Analog curves of ROC curves, denoted precision recall (PR) curves, exist. Davis and Goadrich (2006) have shown that a deep relationship exists between ROC space and PR space, such that a curve dominates in ROC space if and only if it dominates in PR space. Precision and Recall are expressed as

\[ \text{Precision} = \frac{TP}{TP + FP}, \quad (71) \]

\[ \text{Recall} = \frac{TP}{TP + FN}. \quad (72) \]

Martin, Fowlkes and Malik (2004) emphasize that PR curves are better adapted for color image segmentation by edge detection than ROC curves.

2. Quantitative Evaluation Methods without Ground Truth

Even though the mislabeling rate often is used to evaluate the quality of the segmentation. A segmented image \( I_a \) with a mislabeling rate lower than a segmented image \( I_b \) may not correspond to a better segmentation with a visual estimation, as illustrated by Liu and Yang (1994). They explain this phenomenon by the miscellaneous sensitivity of human perception according to the inspected scene. Moreover, it is demanding to construct ground truth segmentation of real scene images. Martin et al. (2001) have built a benchmark image database that contains human segmentation of real scene images. Global measures of segmentation quality were proposed first by Liu and Yang (1994) and improved by Borsotti, Campadelli and Schettini (1998). Both criteria do not require evaluation parameters.

a. Criterion of Liu and Yang. To evaluate segmentation results both locally and globally, the evaluation function \( F(I) \) is defined by Liu and Yang (1994) as

\[ F(I) = \sqrt{N_R} \times \sum_{i=1}^{N_R} \frac{e_i^2}{\sqrt{A_i}}, \quad (73) \]

where \( N_R \) is the number of regions in the segmented image, \( A_i \) is the area of the \( i \)th region, and \( e_i \) its mean error color. In this equation, the term \( \sqrt{N_R} \) penalizes an oversegmentation while the local measure is estimated by the term \( e_i^2/\sqrt{A_i} \) and penalizes small regions or regions with a large color error. The lower the value of the criterion \( F(I) \), the better the segmentation result.
b. Criterion of Borsotti, Campadelli and Schettir. Liu and Yang’s criterion and Borsotti et al.’s criterion are empirically defined as Zhang (1996) suggests. Borsotti et al. improve Liu and Yang’s criterion by a heavier penalization of the segmented image with too many small regions and inhomogeneous color regions. For this task, the sum of Eq. (73) is split into two terms rather than only one. The criterion of Borsotti et al. $Q(I)$ of an $X \times Y$ segmented image $I$ can be computed as

$$Q(I) = \frac{\sqrt{N_R}}{X \times Y} \sum_{i=1}^{N_R} \frac{e_i^2}{1 + \log A_i} + \frac{N_R(A_i)^2}{A_i^2},$$

(74)

where $N_R(A_i)$ is the number of regions having an area equal to $A_i$.

3. Evaluation Methods for Color Image Segmentation by Region Construction

This section provides an overview of several works that used the previous evaluation methods to determine whether a single color space that provides best results for color image segmentation by region construction exists. As in Section IV.B.2, these studies are summarized in a table (see Table 8).

As shown in Tables 6 and 8 there is not single color space that is recommended by all studies. Thus, we can conclude that there does not exist a single color space that provides the best results for color image segmentation. In fact, the choice of the color space varies according to the type of image to segment and the segmentation method used.

D. Selection of the Most Discriminating Color Space

Sections IV.B and IV.C have shown contradictory conclusions about the pertinence of the available color spaces in the context of image segmentation. Instead of searching for the best classical color space to improve the results of image segmentation by pixel classification, Vandenbroucke, Macaire and Postaire (2003) defined a new type of color space by selecting a set of color components that can belong to any of several color spaces. Such spaces, which have neither psychovisual nor physical color significance, are called hybrid color spaces (HCS) by the authors and have been used to extract meaningful regions representing the soccer players and to recognize their teams. More details about this original approach are presented in Section IV.D.1. Another interesting approach proposed by Busin et al. (2005) consists of selecting a specific color space for the construction of each extracted class of pixels by using an iterative procedure. We explain this method in Section IV.D.2.
### TABLE 8

**Methods Used to Evaluate Color Image Segmentation by Region Construction**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Evaluation method</th>
<th>Candidate color spaces</th>
<th>Selected color spaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas-Yedid et al. (2004)</td>
<td>Liu’s criterion</td>
<td>((R, G, B), (r, g, b), (H_1, H_2, H_3))</td>
<td>((I_1, I_2, I_3))</td>
</tr>
<tr>
<td></td>
<td>Borsotti’s criterion specific criterion</td>
<td>((1, 12, 13), (X, Y, Z), (Y, I, Q) (L^<em>, a^</em>, b^*))</td>
<td>((Y', I', Q'))</td>
</tr>
<tr>
<td>Makrogiannis, Economou and Fotopoulos (2005)</td>
<td>PSNR</td>
<td>((R, G, B), (Y, C_b, C_r), (I_1, I_2, I_3), (L^<em>, a^</em>, b^*))</td>
<td>((L^<em>, a^</em>, b^*))</td>
</tr>
<tr>
<td>Liu and Yang (1994)</td>
<td>Liu’s criterion</td>
<td>((R, G, B), (I, S, T), (I_1, I_2, I_3))</td>
<td>Varies according to the image type</td>
</tr>
<tr>
<td>Phung, Bouzerdoum and Chai (2005)</td>
<td>ROC curves</td>
<td>((R, G, B), (H, S, V), (r, g), (H, S), (C_b, C_r), (a^<em>, b^</em>))</td>
<td>((C_b, C_r))</td>
</tr>
<tr>
<td>Lezoray (2003)</td>
<td>MSE</td>
<td>((R, G, B), (X, Y, Z), (Y', I', Q'), (Y', U', V'), (Y, C_b, C_r), (L^<em>, a^</em>, b^*))</td>
<td>Color spaces ordered</td>
</tr>
<tr>
<td>Lim and Lee (1990)</td>
<td>Probability error</td>
<td>((R, G, B), (X, Y, Z), (Y', I', Q'), (U^<em>, V^</em>, W^*))</td>
<td>((I_1, I_2, I_3))</td>
</tr>
<tr>
<td>Park, Yun and Lee (1998)</td>
<td>Probability error</td>
<td>((R, G, B), (X, Y, Z), (Y', I', Q'), (I_1, I_2, I_3))</td>
<td>Segmented images are unaffected by the choice of the color space</td>
</tr>
</tbody>
</table>
1. **Pixel Classification in an Adapted Hybrid Color Space**

Instead of searching for the best classical color space to improve the results of image segmentation by pixel classification, Vandenbroucke, Macaire and Postaire (2003) defined an HCS by selecting a set of color components that can belong to any of the different color spaces listed in Table ???. The HCSs have been used to extract meaningful regions representing the soccer players and to recognize their teams. The goal of the analysis of these images is to extract the players and the referee (Figure 35). In this context, there are three pixel classes of interest—the pixels representing the players of the two opposing teams and the referee—which are identified by the colors of their uniforms.

This section is divided as follow. First, we detail the determination of an HCS. Second, we present the color pixels classification algorithm. Finally, we show several results obtained with this method.

*a. Adapted HCS Construction.* For this study, the color components of the HCS are chosen among a set of color components \( \prod^K = (R, G, B, r, g, b, I, H, S, X, Y, Z, x, y, z, L^*, a^*, b^*, u^*, v^*, I_1, I_2, I_3, A, C_1, C_2, Y', I', Q', U', V', C_{uv}, h_{uv}^o, S_{uv}^o, C_{ab}^o, h_{ab}^o) \), where \( K \) is the number of available color components. Each player pixel \( P(x, y) \) in Figure 35 is represented by a point in the space \( \prod^K \) whose \( k \)th coordinate is denoted \( \pi^k(x, y) \). \( \omega_j \) is a class of player pixels (\( j = 1, \ldots, N_\omega \) where \( N_\omega \) is the number of classes). They also denote \( N_{\omega j} \), the number of player pixels in the class of player pixels \( \omega_j \). For each color feature of this space, they compute \( m^k_j \) the average of the player pixels values in class of player pixels \( \omega_j \):

\[
m^k_j = \frac{1}{N_{\omega j}} \times \sum_{P(x,y) \in \omega_j} \pi^k(x, y). \tag{75}
\]

![Figure 35](image.png)  
*FIGURE 35. Example of three classes of pixels representing the players of the two opposing teams and the referee. (See Color Insert.)*
Thus, they define a $K$-dimensional color feature vector

$$M_j = [m_j^1, \ldots, m_j^k, \ldots, m_j^K]^T$$

for each class of player pixels $\omega_j$. In order to select the set of most discriminating color features among $K$ available color features, they propose use of a specific informational criterion. They assume that the better the classes are separated and compact in the HCS, the higher the discriminating power of the selected feature. That assumption leads to their choice of the measures of separability and compactness. The measure of compactness of each class $\omega_j$ ($j = 1, \ldots, N_{\omega}$) is defined by the within-class dispersion matrix $\Sigma_j$

$$\Sigma_j = \frac{1}{N_{\omega_j}} \times \sum_{P(x,y) \in \omega_j} (X_j - M_j)(X_j - M_j)^T,$$

(76)

where $X_j = [\pi^1(x, y), \ldots, \pi^k(x, y), \ldots, \pi^K(x, y)]^T$ is the color point of the pixel $P(x, y)$ that belongs to the class $\omega_j$. They define the total within-class dispersion matrix $\Sigma_C$ as

$$\Sigma_C = \frac{1}{N_{\omega}} \times \sum_{j=1}^{N_{\omega}} \Sigma_j.$$

(77)

The measure of the class separability is defined by the between-class dispersion matrix

$$\Sigma_S = \frac{1}{N_{\omega}} \times \sum_{j=1}^{N_{\omega}} (M_j - M)(M_j - M)^T,$$

(78)

where $M = [m^1, \ldots, m^k, \ldots, m^K]^T$ is the mean vector of all the classes

$$M = \frac{1}{N_{\omega}} \times \sum_{j=1}^{N_{\omega}} M_j.$$

(79)

The most discriminating set of color features maximizes the criterion $J = \text{trace}(\Sigma_C^{-1} \Sigma_S)$. The HCS is constructed by means of the “knock-out” algorithm (Firmin et al., 1996), and its dimension $d$ is based on a correlation measure. The dimensionality of the HCS increases while this correlation measure is lower than a given threshold. With this procedure, they select the most discriminating color features among the $K$ available ones. The player pixels are classified in this most discriminating HCS.
b. Color Pixels Classification Algorithm. Before classifying the player pixels of a current color image, the player pixels are extracted. Then the $R$, $G$, and $B$ features of each player pixel $P(x, y)$ are transformed into HCS features. To classify a player pixel $P(x, y)$ of a color image, they consider the set of player pixels falling into a neighborhood of $P(x, y)$. The size of this neighborhood depends on the mean player size in the image. Then for each player pixel, they evaluate a mean vector $M_P = [m^1_p, \ldots, m^d_p]^T$ of the HCS features of the player pixels belonging to neighborhood. For each class $\omega_j$, they evaluate the Euclidean distance $D_j(x, y)$ between the mean vector $M_j$ of the class $\omega_j$ and the mean vector $M_P$ in the HCS

$$D_j(x, y) = \|M_j - M_P\| = \sqrt{\sum_{k=1}^{d} (m^k_j - m^k_P)^2}.$$  

Finally, a minimum decision rule is used to assign $P(x, y)$ to the class $\omega_j$ ($j = 1, \ldots, N_\omega$) for which $D_j(x, y)$ is minimum.

c. Results. Vandenbroucke, Macaire and Postaire (2003) illustrate the effectiveness of the HCS for color image segmentation by pixels classification to extract meaningful regions representing the soccer players and to recognize their teams. The images in Figure 36 constitute a test sample extracted from the same sequence. These images are not easy to segment because each of them contains at least two adjacent players. The player pixels extracted from the images of Figure 36 by means of the multi-thresholding scheme of Ohlander, Price and Reddy (1978) are shown in Figure 37.

The images in Figure 38 show how the extracted player pixels have been classified in the $(a^*, C_U V)$ adapted HCS. The player pixels assigned to the same class are labeled with the same color. The connected player pixels with the same label constitute regions that correspond to the different soccer players.

Vandenbroucke, Macaire and Postaire (2003) compare the results (see Figure 38) with the ground truth of Figure 39 by means of the error classification rate (see Table 9). Since the mean classification error rate associated with the adapted hybrid color space $(a^*, C_U V)$ is the lowest one, they conclude that the classification in the HCS provides better results than the classification in the classical color spaces.

2. Unsupervised Selection of the Most Discriminating Color Space

The selection of the HCS is achieved by a supervised learning scheme. For several applications such as image retrieval, color space must be selected...
FIGURE 36. Color soccer images (125 × 125 pixels). (See Color Insert.)

FIGURE 37. Player pixels extracted from the images in Figure 36. (See Color Insert.)

FIGURE 38. Player pixels of Figure 37 classified in the hybrid color space. (See Color Insert.)

FIGURE 39. Ground truth of the player pixels of Figure 37. (See Color Insert.)
COLOR SPACES AND IMAGE SEGMENTATION

TABLE 9
CLASSIFICATION ERROR RATES OF PLAYER PIXELS IN FIGURE 37

<table>
<thead>
<tr>
<th>Color space</th>
<th>Classification error rates</th>
<th>Mean error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>image (a)</td>
<td>image (b)</td>
</tr>
<tr>
<td>(r, g, b)</td>
<td>32.63</td>
<td>10.01</td>
</tr>
<tr>
<td>(H, S, I)</td>
<td>8.63</td>
<td>27.77</td>
</tr>
<tr>
<td>(Y', CUV, rUV)</td>
<td>12.28</td>
<td>17.15</td>
</tr>
<tr>
<td>(a*, CUV)</td>
<td>11.5</td>
<td>7.35</td>
</tr>
<tr>
<td>(R, G, B)</td>
<td>24.12</td>
<td>29.25</td>
</tr>
</tbody>
</table>

without any learning. Busin et al. (2005) proposed a selection scheme based on an unsupervised learning analysis of color histogram.

Because the classes constructed by a classification procedure depend on the color space used, it would be prudent to select the color space that is the most relevant for detecting the modes that correspond to regions. For this purpose, Busin et al. (2005) assumed that the higher the discriminating power of the 1D histogram, the more probable the detected modes correspond to regions with the same colors. Thus from among several color spaces, the proposed procedure selects those for which the discriminating powers of the 1D histograms are the highest.

The discriminating power of a 1D histogram depends on the number of detected modes and on the connectedness properties of pixels whose color component levels fall into these detected modes. They consider that the higher the number of detected modes and the greater the connectivity in the image plane of the pixels belonging to the detected modes, the more discriminating the considered 1D histogram. The main originality of the proposed unsupervised procedure is the selection of the most relevant color space for constructing each class of pixels at each iteration step. The selection simultaneously takes into account the color and spatial connectedness properties of the pixels in the image. Each stage of the proposed iterative method is shown by Figure 40 and is detailed in the following subsections.

a. One-Dimensional Histograms Determination. At each iteration step, the color vectors of the pixels submitted to the analysis are represented into the \( N_S = 11 \) color spaces \(((R, G, B), (r, g, b), (X, Y, Z), (x, y, z), (Y', I'), Q), (Y', U', V'), (bw, rg, by), (Y, Ch_1, Ch_2), (L*, a*, b*), (L*, u*, v*), and (I_1, I_2, I_3))\). In the \( i \)th color space, each of the three 1D histograms \( H^{i,j}(x) \) of each color component numbered \( j (j = 1, 2, 3) \), where \( x \) is the color component level, is determined. The 1D histograms of the red, green, and blue
components of the House image in Figure 41a are represented in Figures 41b, c, and d, respectively.

**b. One-Dimensional Histograms Smoothing.** Because the 1D histograms are corrupted by noise, it is difficult to detect their modes. Thus, they propose to smooth them by means of adaptive filtering. A smoothed histogram $H_{i,j}^{i,j}(x)$ is computed by the convolution between the 1D histogram $H_{i,j}^{i,j}(x)$ and a Gaussian kernel $g_\sigma(x)$, where $\sigma$ is the standard deviation

$$h_{i,j}^{i,j}(x) = h_{i,j}^{i,j}(x) * g_\sigma(x),$$

where “*” denotes the convolution operator and

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

The effect of the smoothing depends on the standard deviation $\sigma$ used to define the Gaussian kernel. For each 1D histogram, $\sigma$ is automatically determined by means of the procedure proposed by Lin, Wang and Yang (1996), so that the smoothed 1D histogram reveals its modes. Figure 42 shows the smoothed 1D histograms of the House image of Figure 42a. These smoothed 1D histograms can be easily analyzed for mode detection.

**c. Most Relevant Color Space Selection.** The thresholds that delimit the modes of each smoothed histogram $H_{i,j}^{i,j}(x)$ are determined by the analysis of the zero-crossing of its first-derivative function. A threshold is detected by a zero-crossing of the first-derivative function of $H_{i,j}^{i,j}(x)$ whose sign changes.

![Color image segmentation flowchart.](image-url)
COLOR SPACES AND IMAGE SEGMENTATION

from minus to plus (local minimum). A mode is detected by a zero-crossing of the first-derivative function of $H_{\sigma}^{i,j}(x)$ whose sign changes from plus to minus (local maximum). The number of these detected modes is denoted $N_{i,j}$.

Three features characterize the $k$th ($k = 1, \ldots, N_{i,j}$) detected mode of the 1D histogram $H_{\sigma}^{i,j}(x)$:

- The left and right detected thresholds $T_{\text{left}}^{i,j,k}$ and $T_{\text{right}}^{i,j,k}$
- The amplitude

$$A_{i,j,k}^{i,j,k} = \max_{l=T_{\text{left}}^{i,j,k}}^{T_{\text{right}}^{i,j,k}} h_{\sigma}^{i,j}(l).$$
For each 1D histogram $H^{i,j}_{\sigma}(x)$, $K(i, j)$ denotes the rank order of the detected mode with the highest amplitude.

To illustrate the procedure, Table 10 shows the features of the detected modes in the smoothed histograms of Figures 42b–d with only the $(R, G, B)$ color space ($i = 1$). Figure 43 shows the features of the detected modes from the smoothed histogram $H^{1,3}_{\sigma}(x)$ in Figure 42c.

The main problem of analysis for schemes 1D histograms is that they analyze only the similarities between the colors of the pixels and ignore their spatial arrangement in the image. However, a region is a subset of pixels that shares similar color properties and is strongly connected in the image plane.
To select the most relevant color space, they measure the connectedness degrees of subsets of pixels $S$ by means of connected degrees $CD(S)$ introduced by Macaire, Vandenbroucke and Postaire (2006). The connectedness degree $CD(S)$ is a normalized measure, so that a connectedness degree close to 1 means that the pixels belonging to the subset $S$ are strongly connected in the image, whereas a connectedness degree close to 0 means that the pixels are sparsely scattered throughout the image.

Table 10 shows the connectedness degrees $CD(S)$ of the detected modes of the smoothed histograms shown in Figures 42b–d. Among the three smoothed 1D histograms of the $i$th color space, $i = 1, \ldots, N_S$, the most discriminating
1D histogram is determined. For this purpose the discriminating power, denoted \( R_{i,j} \), is evaluated for each smoothed 1D histogram \( H_i^{\gamma,j}(x) \). \( R_{i,j} \) is the sum of the connectedness degrees of the subsets \( S[T_{i,j,k}^{\text{left}}, T_{i,j,k}^{\text{right}}] \) associated with the detected modes of the histogram

\[
R_{i,j} = \sum_{k=1}^{N_{i,j}} CD(S[T_{i,j,k}^{\text{left}}, T_{i,j,k}^{\text{right}}]).
\]  

(82)

Because the Gaussian smoothing eliminates the nonsignificant modes, the 1D histogram allows discrimination of the classes if the number of the detected modes is high and if the sum of their connectedness degrees is high. Thus, the higher the discriminating power \( R_{i,j} \), the more probable the modes correspond to regions with the same colors in the image. Therefore, the most discriminating 1D histogram of the \( i \)th color space is the 1D histogram with the highest discriminating power \( R_{i,j} \). They denote \( J(i) \) as the rank order of the color component that corresponds to the most discriminating 1D histogram of the \( i \)th color space—that is, the 1D histogram with the highest discriminating power \( R_{i,j} \). Table 10 shows that, in our example, among the 3D histograms of Figure 42, the most discriminating histogram of the \((R, G, B)\) color space \((i = 1)\) is the histogram of the component \( G \), so \( J(1) \) is set to 2.

The most relevant color space is selected among the \( N_S \) color spaces as that with the highest discriminating power. If the discriminating powers of several color spaces are equal, the most relevant color space among those spaces is selected as that which contains the 1D histogram with the second highest value of discriminating power \( R_{i,j} \). They denote \( I \) the rank order of the color space that is selected as the most relevant one. For the House image in Figure 41a, the most relevant color space that is selected at the first step is the \((Y', U', V')\) color space \((I = 6)\). The most discriminating 1D histogram of the selected color space is the \( V' \) component \((J(6) = 3)\).

d. One Class Construction. One class of pixels is constructed by analyzing the most relevant color space with rank order \( I \), which has been determined at the current iteration step of the algorithm. This class of pixels is defined by a parallelepipedic box in the most relevant color space. Only one parallelepipedic box is selected to build the class of pixels. This box is delimited by two thresholds defined along each color component of the most relevant color space.

Along the color component with rank order \( J(I) \) which corresponds to the most discriminating 1D histogram, the two thresholds \( T_{i,J(I),K(I,J(I))}^{\text{left}} \) and \( T_{i,J(I),K(I,J(I))}^{\text{right}} \) are those that delimit the mode with the highest amplitude. The thresholds along the two other color components are selected among the
COLOR SPACES AND IMAGE SEGMENTATION

thresholds $T_{left}^{I,j,k}$ and $T_{right}^{I,j,k}$, $j \neq J(I)$, determined by the mode detection stage. The selected thresholds delimit the box into which fall the color vectors of the highest population of pixels. The pixels whose color vectors fall into this box constitute the class of pixels constructed at the current iteration step.

e. **Pixels of the Class Extraction.** The pixels assigned to the constructed class are extracted from the color image so that they are not taken into account at the next iteration steps of the procedure. The pixels that are assigned to this class and which are connected in the image constitute one of the reconstructed regions in the segmented image.

f. **Stopping Criterion.** The iterative procedure stops when a percentage $p$ of pixels of the image have not been assigned to any of the previously constructed classes. The parameter $p$, adjusted by the analyst, allows the desired coarseness of the segmentation be tuned. When the iterative procedure stops, the pixels that have not been assigned to any class could be assigned to one of the constructed classes by means of a specific decision rule.

g. **Results.** This class construction scheme is based on the analysis of both the connectedness and the color properties of the subsets of pixels. To demonstrate the interest of this approach, they propose segmenting the benchmark image named House (see Figure 44a) by means of the presented procedure (see Figure 44b).

The extracted pixels of a constructed class at each iteration step of the algorithm are labeled with a false color in the segmented image of Figure 44b. These labels are represented in the first column of Table 11. As these false colors are ordered, the iteration steps at which the classes of pixels are constructed by examining the proposed segmented images results can be determined.

Table 11 indicates the discriminating powers of the color spaces selected at each step of the procedure applied to the image in Figure 44a. It shows that there does not exist one single color space that is the most relevant at all iteration steps of the procedure.

Figure 44b shows that this procedure provides satisfying segmentation results in terms of pixel classification. Indeed, the regions representing the different objects in the images are well reconstructed.

The result obtained with the proposed method is compared with the one obtained when only the $(R, G, B)$ color space is taken into account by each iteration step (see Figure 44c), and with the one obtained when only the most relevant color space selected at the first iteration step is taken into account by the other iteration steps (see Figure 44d). By examining the segmented images, they conclude that the selection of the most relevant color space
Figure 44. Segmentation of the House image (255 × 255 pixels) by the proposed method. (a) Original image House. (b) Segmented image House by the proposed method. (c) Segmented image House when the (R, G, B) color space is selected at each iteration. (d) Segmented image House when the \( (Y', U', V') \) color space is selected at each iteration. (See Color Insert.)

Table 11

<table>
<thead>
<tr>
<th>Iteration step</th>
<th>Color space</th>
<th>( N_{I, J(I)} )</th>
<th>( R_{I, J(I)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Y', U', V')</td>
<td>3</td>
<td>4.42</td>
</tr>
<tr>
<td>2</td>
<td>(Y', U', V')</td>
<td>3</td>
<td>5.34</td>
</tr>
<tr>
<td>3</td>
<td>(R, G, B)</td>
<td>2</td>
<td>4.38</td>
</tr>
<tr>
<td>4</td>
<td>(R, G, B)</td>
<td>2</td>
<td>5.25</td>
</tr>
<tr>
<td>5</td>
<td>(bw, rg, by)</td>
<td>1</td>
<td>1.75</td>
</tr>
<tr>
<td>6</td>
<td>(I1, I2, I3)</td>
<td>1</td>
<td>1.94</td>
</tr>
<tr>
<td>7</td>
<td>(r, g, b)</td>
<td>1</td>
<td>1.20</td>
</tr>
<tr>
<td>8</td>
<td>(Y', U', V')</td>
<td>1</td>
<td>1.49</td>
</tr>
</tbody>
</table>

(See Color Insert.)
provides results that are more acceptable in terms of segmentation quality than the class construction achieved in one single color space. These results show that the selection of different color spaces at the iteration steps of the procedure, which is designed to discriminate the considered pixel classes, is a relevant solution for the construction of the regions.

E. Conclusion

This section has presented several evaluation methods of color image segmentation to highlight the relationships between segmentation and color spaces. These evaluation methods have been used to compare the segmentation results according to the chosen color space used to represent the colors of the pixels. We have shown that the choice of the color space depends on the kind of image to be segmented and the segmentation method.

Because no single “best” color space exists for color image segmentation, several studies select them by means of decision rules. The results provided by these color image segmentation methods allow improved image segmentation results.

V. CONCLUSION

In this chapter we have first described the most used color spaces for digital color image analysis. We have pointed out that most transformations from \((R, G, B)\) to another color space require the prior knowledge of the acquisition conditions (reference white, illuminating properties). Moreover, the coding of the color components must be adapted to the used color space.

In the second section we presented classical segmentation schemes designed for exploiting colors, which can be divided in two families. Since edge detection requires postprocessing steps to yield closed boundaries, most schemes construct the regions in the image via either image plane analysis or color space analysis. Recent approaches tend to combine the spatial and the colorimetric analysis to improve the quality of the segmentation.

The third section deals with the relationships between color spaces and segmentation. Regardless of the criteria used to evaluate the quality of a segmentation, no color spaces effect in which segmentation schemes provide efficient results of \(all\) images, which prompted our development of schemes to determine the color space that is the most adapted to a specific set of images. These schemes are based on multidimensional statistical criteria but not on psychovisual criteria. One improvement in our approaches could be the integration of these criteria to the selection of the color space for image segmentation.
REFERENCES


COLOR SPACES AND IMAGE SEGMENTATION


COLOR SPACES AND IMAGE SEGMENTATION


