A Computational Model for Color Naming and Describing Color Composition of Images

A. Mojsilović, *Member, IEEE*

***Abstract*—The extraction of high-level color descriptors is an increasingly important problem, as these descriptions often provide link to image content. When combined with image segmentation, color naming can be used to select objects by color, describe the appearance of the image and generate semantic annotations. This paper presents a computational model for color categorization, naming and extraction of color composition. In this work we start from the *National Bureau of Standards'* recommendation for color names [1], and through subjective experiments develop our color vocabulary and syntax. To assign a color name from the vocabulary to an arbitrary input color, we then design a perceptually based color naming metric. The proposed algorithm follows relevant neurophysiological findings and studies on human color categorization. Finally, we extend the algorithm and develop a scheme for extracting the color composition of a complex image. According to our results, the proposed method identifies known color regions in different color spaces accurately, the color names assigned to randomly selected colors agree with human judgments, and the description of the color composition of complex scene is consistent with human observations.**

***Index Terms*—Color naming, color composition, segmentation**

I. Introduction

Color is one of the main visual cues and has been studied extensively on many different levels, starting from the physics and psychophysics of color to the use of color principles in practical problems, such as accurate rendering, display and reproduction, segmentation, and numerous other applications in image processing, visualization and computer graphics. Although color naming represents one of the most common visual tasks, it has not received significant attention in the engineering community. Yet today, with rapidly emerging visual technologies, sophisticated user interfaces and human-machine interactions, the ability to name individual colors, point objects of certain color, and convey the impression of color composition becomes an increasingly important task. The extraction of higher-level color descriptors represents a challenging problem in image analysis and computer vision, as these descriptors often provide link to image content. When combined with image segmentation, color naming can be used to select objects by color, describe the appearance of the image and even generate semantic annotations. For example, regions labeled as *light blue* and *strong green* may represent sky and grass, vivid colors are typically found in man-made objects, while modifiers such as *brownish*, *grayish* and *dark* convey the impression of the atmosphere in the scene. All the applications mentioned so far require flexible computational model for color categorization, naming or extraction of color composition. However, modeling human behavior in color categorization involves solving, or at least providing some answers to several important problems. The first problem involves the definition of the basic color categories and "most representative examples", called prototypical colors, which play a special role in structuring these color categories. Another open issue is how to expand the notion of basic color terms into a "general" yet precise vocabulary of color names that can be used in different applications. Another problem involves the definition of category membership. Although the idea that color categories are formed around prototypical examples has received striking support in many studies, the mechanisms of color categorization and category membership are not yet fully understood. And finally, assuming that we have been able to provide some solutions to all these non-trivial problems and develop an algorithm that assigns a color name to an arbitrary color sample, we are still very far away from capturing how the color appearance of a complex scene may be described by a human observer. The objective of this paper is to provide the first steps in addressing these issues, as we make an attempt to develop a computational model for naming individual colors, as well as generating useful descriptors of color composition. To achieve these goals we need to consider relevant neuro-physiological findings and some well-known studies on human color categorization, as they set the directions for our work.

A. Color perception, categorization and naming

Color vision is initiated in retina where the three types of cones receive the light stimulus. The cone responses are then coded into one achromatic and two antagonistic chromatic signals. These signals are interpreted in the cortex, in the context of other visual information received at the same time and the previously accumulated visual experience (memory). Once the intrinsic character of colored surface has been represented internally, one may think that the color processing is complete. However, an ever-present fact about human cognition is that people go beyond the purely perceptual experience to classify things as members of categories and attach linguistic labels to them. Color is no exception. Sky and sea are classified as *blue*, despite the differences in the perceived color. That color categories are perceptually significant can be demonstrated by the "striped" appearance of the rainbow. In physical terms, the rainbow is just a light with the wavelength changing smoothly from 400-700 nm. The unmistakable stripes of color in the rainbow suggest an experimental basis for the articulation of color into at least some categories [3]. A breakthrough in the current understanding of color categorization came from a cross-cultural study conducted by Berlin and Kay [4]. They studied the color naming behavior with subjects from variety of languages. They examined 20 languages experimentally and another 78 through the literature review and discovered remarkable regularities in the shape of the basic color vocabulary. As a result of their study, Berlin and Kay introduced a concept of basic color terms, and worked on defining the color categories corresponding to these basic terms. They identified 11 basic terms in English (*black*, *white*, *red*, *green*, *yellow*, *blue*, *brown*, *pink*, *orange*, *purple* and *gray*). Berlin and Kay's experiments also demonstrated that humans perform much better in picking the "best example" for each of the color terms than in establishing boundaries between the categories. This lead to the definition of focal colors representing the centers of color categories, and the hypothesis of graded (*fuzzy*) membership. Many later studies have proven this hypothesis, indicating that prototypical colors play a crucial role in internal representation of color categories, and the membership in a color category seem to be represented relative to the prototypes [5].

Unfortunately, the mechanism of color naming is still not completely understood. The only existing theoretical models of color naming based explicitly on neurophysiology of color vision and addressing the universality of color foci and graded membership are [6] and [7]. Apart from not being developed or implemented as full-fledged computational models, both of these have important drawbacks. In Kay and McDaniel's model [6] membership in color categories is formalized in terms of fuzzy set theory, by allowing objects to be members of a given set to some degree. In terms of color categories, this means that a focal or prototypical color will be represented as having a membership degree of 1 for its category. Other, non-focal colors will have membership degrees that decrease systematically with the distance from the focal color in some color space. However, this model considers only four fuzzy sets (*red*, *green*, *yellow* and *blue*) and supporting other color terms requires the introduction of new and ad hoc fuzzy set operations. Furthermore, it is not clear how the non-spectral basic color categories, such as brown, pink and gray are to be dealt with, nor how to incorporate the learning of color names into the model. Cairo's model of color naming is based on findings in the physiology of the pre-cortical system [7]. It defines four physical parameters of the stimulus: wavelength, intensity, purity and adaptation state of the retina. According to the model, the pre-cortical visual system performs analog-to-digital conversion of these four parameters, and represents 11 basic color categories as specific combinations of the quantized values. As already observed, although interesting for its attempt to take adaptation into account, this model is clearly a gross simplification, which cannot hold in general [5].

B. From color spaces to color naming models

Color spaces allow us to specify or describe colors in unambiguous manner, yet in everyday life we mainly identify colors by their names. Although this requires a fairly general color vocabulary and is far from being precise, identifying a color by its name is a method of communication that everyone understands. Hence, there were several attempts towards designing a standard method for choosing color names. The Munsell color order system is widely used in applications requiring precise specification of colors, such as production of paints and textiles [8], [9]. Two notable disadvantages of the Munsell system for the color-based processing are: 1) the lack of a color vocabulary and 2) the lack of exact transform from any color space to Munsell. For example, a transform proposed by Miyahara [10] is fairly complicated and sometimes inaccurate for certain regions of *CIE XYZ*. The first listing of over 3000 English words and phrases used to name colors was devised by Maerz and Paul and published in the Dictionary of colors [11]. Even more detailed was a dictionary published by The National Bureau of Standards (NBS). It contained about 7500 different names that came to general use in specific fields such as biology, textile, dyes and paint industry [1]. Both dictionaries include examples of quite esoteric words and the terms are listed in an unsystematic manner, making them unsuitable for general use. Following the recommendation of the Inter-Society Council, NBS developed the ISCC-NBS dictionary of color names for 267 regions in color space [1]. This dictionary employs English terms to describe colors along the three dimensions of the color space: hue, lightness and saturation. There are five values for lightness (*very dark*, *dark*, *medium*, *light* and *very light*), four values for saturation (*grayish*, *moderate*, *strong* and *vivid*), three terms that address both lightness and saturation (*brilliant*, *pale* and *deep*), and 28 names for hues constructed from a basic set (*red*, *orange*, *yellow*, green, *blue*, *violet*, *purple*, *pink*, *brown*, *olive*, *black*, *white* and *gray*). One problem with the ISCC-NBS model is the lack of systematic syntax. This was addressed during the design of a new Color-Naming System (CNS) [12], which was based on the ISCC-NBS model. CNS uses the same three dimensions, however the rules used to combine words from these dimensions are defined in a formal syntax. An extension of the CNS model, called the Color-Naming Method (CNM), was proposed by Tominaga in [13]. Tominaga used a predefined set of color names in the Munsell color space and developed a method for specifying color names of individual pixels or surface color samples [13]. Color names in the CNM are specified at one of four accuracy levels (fundamental, gross, medium, and minute), so that names from the higher accuracy level correspond to smaller color regions in the Munsell space. However, the method has several drawbacks. First, it uses a non-standard vocabulary of color names (e.g. *lilac*, *lavender*, *sky*, *gold*). Furthermore, the method is based on the optical measurement system, which converts the input color surface into the Munsell color space. In order to apply such a system to recorded images one needs to deal with the issues of *RGB* to Munsell conversion [10], [20], [23] -- a setback for applications that go beyond closely controlled settings such as Tominaga's (for example, diverse digital image libraries or web images, which are often not obtained with calibrated cameras). Finally, it is not obvious how to extend Tominaga's methods to automatically assign a color name to a sample image, point out examples of named colors, describe color regions and objects in the scene and communicate the color composition of the image. A computational model that provides the solution to some of these problems was proposed by Lammens, who used Berlin and Kay's color naming data and applied a variant of the Gaussian normal distribution as a category model [5]. The model was fitted to the 11 basic color names and does not account for commonly used saturation or luminance modifiers, such as *vivid orange* or *light blue*. Since the quality of color categorization depends on an intricate fitting procedure, there is no straightforward extension of the model to include these attributes. In [27], [28] Belpaeme offers another approach to the formation and computational simulation of color categorization - categorization based on the notion of color primitives surrounded by color regions with fuzzy boundaries, and modeling via adaptive radial basis function networks.

The goal of our work is to develop a broader computational color naming method, which will provide more detailed color descriptions, allow higher-level color communication, and satisfy the following properties. Color naming operation should be performed in a perceptually controlled way, so that the names attached to different colors reflect perceived color differences among them. Segmenting a color space into the color categories should produce smooth regions. The method should account for the basic color terms and use systematic syntax to combine them. It should respect the graded nature of category membership, the universality of color foci, and produce results in agreement with human judgments. The first step in our work, described in Section 2, involves the design of a balanced and well-represented set of color prototypes, *vocabulary*, and the corresponding *syntax*. In Section 3, we describe the design of a color naming metric, which for an arbitrary input color determines the category membership. In Sections 4 and 5 we extend this approach to name color regions and provide the description of the color composition for complex images. Some applications for color naming, directions for future work and concluding remarks are given in Section 6.

II. Color Naming Vocabulary and Syntax

As a starting point in our vocabulary, we adopted the ISCC-NBS dictionary [1], since it provides a model developed using controlled perceptual experiments and includes the basic color terms. Each color category is represented with its centroid color, thus preserving the notion of color foci. Yet, due to the strict naming conviction the ISCC-NBS dictionary includes several color names that are not well understood by general public (i.e. *blackish red*) and lacks systematic syntax. As the centroid colors span the color space in uniform fashion and allow grading between the categories, we decided to use these points as the prototypes in our color naming algorithm, but had to devise our own name structure that follows few simple systematic rules. To determine a reliable color vocabulary, we have performed a set of subjective experiments aimed at testing the agreement between the names from the ISCC-NBS dictionary and human judgments, adjusting the dictionary for the use in automatic color naming applications and gain better understanding of human color categorization and naming.

A. Experiments

We have conducted four experiments: Color Listing Experiment aimed at testing 11 basic color categories from Berlin and Kay study, Color Composition Experiment aimed at determining color vocabulary used in describing complex scenes, and two Color Naming Experiments aimed at understanding human behavior in color naming and adjusting the differences between the human judgments and the semantics of the ISCC-NBS vocabulary. Ten subjects participated in the experiments. All subjects had normal color vision and normal or corrected-to-normal vision.

Color Listing Experiment In addition to the 11 basic color terms in English, some studies indicated few marginal cases such as *beige/tan* [3], *olive* and *violet* [1]. To test the relevance of these terms we asked each subject to name at least twelve "most important" colors.

Color Composition Experiment In this experiment the subjects were presented with 40 photographic images in a sequence and asked to name all colors in the image. The images were selected to provide broad content, different color compositions, spatial frequencies and arrangements among the colors. Each image was displayed on a calibrated monitor against light gray background. The order of presentation was randomly generated for each subject. The subjects were advised to use common color terms and avoid rare color names. If they found a certain color difficult to name, we advised them to describe it in terms of other colors.

Color Naming Experiments In these experiments the subjects were presented with 267 centroid colors from the ISCC-NBS color dictionary and asked to name each color. The color patches were displayed on the computer monitor calibrated so that there was no difference between the colors on the monitor and corresponding chips form the Munsell Book of Colors [9] when viewed under same conditions. In the first experiment, 64×64 pixel patches were arranged into 9×6 matrix and displayed against light gray background. The names were assigned by typing into a text box below each patch. The display was then updated with the new set of patches, until all 267 colors have been named. The placement of colors within the matrix was determined randomly for each subject. In the second color naming experiments only one 200×200 pixels color patch was displayed on the screen. As in the Color Composition Experiment, in both Color Naming Experiments subjects were advised to use common color names, common modifiers for brightness or saturation, and avoid names derived from objects/materials. (Similar experiment has been recently described by Moroney in [29]).

B. Experimental results: Findings, vocabulary and syntax

Here we summarize the most important findings from the experiments and describe the resulting color naming vocabulary and syntax.

In the Color Listing Experiment 11 basic colors were found on the list of every subject. Nine subjects included beige and four included *violet*. Modifiers for hue, saturation and luminance were not used. None of the subjects listed more than 14 color names. The subjects maintained almost identical vocabulary when describing images in the Color Composition Experiment. The modifiers for hue, saturation and luminance were used only to distinguish between different types of the same hue in the single image (such as *light* blue for sky and *dark blue* for water) and were otherwise seldom included. Although most of the images had rich color histograms the subjects never listed more than ten colors.

The subjects showed the highest level of precision in the Color Naming Experiments. Most of them (8/10) frequently used modifiers for hue, saturation or brightness. The modifiers for hue were designed either by joining two generic hues with a hyphen, or by attaching the suffix *–ish* to the farther hue. Typically, only two adjacent hues (e.g. *purple* and *blue*) were combined. Seven subjects used *olive*, although they had not used this term in the previous experiments. On the other hand, although it had been listed in the Color Listing Experiment, *violet* was seldom used and was most of the time described as *bluish purple*. Modifiers *brilliant* and *deep*, as in the ISCC-NBS vocabulary, were not used. There was a very good degree of concordance between the subjects; In the First Color Naming Experiment, out of 267 color samples, 223 of them were assigned the same hue by all subjects (the variations were in the use of modifiers), 15 were assigned into one of two related hue categories (such as *yellowish green* and *green*), 19 were assigned into one of three related hue categories (such as *greenish yellow*, *yellowish green* and *green*). The remaining 10 color samples were not reliably assigned into any category. Out of 223 hues that were assigned into the same category by all subjects, 195 were the same as in the ISCC-NBS vocabulary, 22 were assigned to a related hue, and 8 hues were assigned entirely different color name. Similar results were obtained in the Second Color Naming Experiment. The most notable difference between subjective judgments and ISCC-NBS vocabulary involved the use of saturation modifiers. Colors appeared less saturated to our subjects and they generally applied higher "thresholds" when attaching modifiers like *vivid*, *strong* or *grayish*. These observations are in agreement with the results of Moroney's experiments [29].

To analyze the agreement between the two color naming experiments, for each experiment we have devised a list of corrected color names, i.e. the names from the ISCC-NBS vocabulary were changed to reflect the opinion of the majority of subjects. By comparing the two lists, we have observed a very good agreement between the experiments - the only difference between the two experiments was in the use of luminance modifiers. The same color was often perceived lighter when displayed in the small patch (Experiment 1) than in the large window (Experiment 2). Also, very pale and unsaturated (grayish) colors appeared more chromatic when displayed in the smaller window. Hence, colors that were perceived as grayish in the first experiment (*grayish blue* for example) were named gray (*bluish gray*) in the second.

For the final vocabulary we have adopted the list from the first color naming experiment. These names were generated in the interaction with other colors and we felt that this choice is a better representative of the real-world applications. We have generalized our findings in the following syntax (the symbol : denotes "is defined as" and symbol | denotes meta-or):

<color name> : <chromatic name> | <achromatic name>

<chromatic name> : <lightness> <saturation> <hue> | <saturation> <lightness> <hue>

<achromatic name> : <lightness> <achromatic term>

<lightness> : blackish | very dark | dark | medium | light | very light | whitish

<saturation> : grayish | moderate | medium | strong | vivid

<hue> : <generic hue> | <halfway hue> | <quarterway hue>

<generic hue> : red | orange | brown | yellow | green | blue | purple | pink | beige | olive

<halfway hue> : <generic hue> - <genereic hue>

<quarterway hue> : <ish form> <generic hue>

<ish form> : reddish | brownish | yellowish | greenish | bluish | purplish | pinkish

<achromatic term> : <generic achromatic term>| <ish form> <generic achromatic term>

<generic achromatic term>: gray | black | white

We also assume that: 1. If <lightness> is omitted, medium is assumed. 2. If <saturation> is omitted, medium is assumed. 3. Only adjacent hues may be combined to form <halfway hue> and <quarterway hue>.

Our experiments have confirmed that ISCC-NBS dictionary includes several color names/terms that are not well understood by general public. It is important to emphasize that the primary goal of our experiments was to "correct" only the syntax of these names, not the color values of corresponding prototypes. Consequently, our vocabulary can be viewed as a "renamed ISCC-NBS", as it operates on the same set of prototypes as the ISCC-NBS model. The difference between them is due to the fact that: 1) color prototypes that have not been consistently perceived by our subjects were removed from the model, and 2) some of the ISCC-NBS names were changed to reflect the majority of subjective decisions.

III. Color Naming Metric

Having established the vocabulary of color names, the next step is developing an algorithm to assign a color name to an arbitrary input color. The color naming process should address the graded nature of category membership and take into account the universality of color foci. Therefore, we will perform color categorization through the color naming metric. Assuming a well-represented set of prototypes (foci), the metric computes the distance between the input color and all prototypes, thus providing a membership value for each categorical judgment.

Although commonly used as measure of color similarity, Euclidean distance in the *CIE Lab* color space has several drawbacks for the use in color naming applications. The first problem is related to the sparse sampling of the color space. It is well known that the uniformity of the *Lab* suffers from defects, so that "nice" perceptual properties remain in effect only within a radius of few just-noticeable differences [2][14]. Since there are only 267 points in our vocabulary, the distances between the colors may be large and the metric only partially reflects the degree of color similarity. For example, when the vocabulary was used with the *Lab* distance to name regions along the gray line in the *Lab* color space ($0<L<100$, $a = 0$, $b = 0$), some regions were named *pinkish white*, *light bluish gray* and *dark greenish gray*, instead of *white*, *light gray* and *dark gray*. The other, more serious problem is related to our perception of color names and their similarity. Let us assume an arbitrary color represented by a point in the *Lab* space, and a set of neighboring colors, $\{c_{ni}\}$, on a circle with the radius L in that space. Although all pairs (c_p, c_{ni}) are equally distant, we do not perceive them as equally similar. This is illustrated in Fig. 1 (within the limit of printer gamut of course) where color c_p is compared to the colors c_{x1} - c_{x5}, all satisfying $D_{Lab}(c_p, c_{xi}) = 10$. The color coordinates and related data are given in Table I. Although all pairs share the same distance in the color space, the perceptual differences between them are not equal. The data in Table I indicate a correlation between perceptual similarity, distances $D_{HLS}(c_p, c_{xi})$, and spatial angles $\theta_{HLS}(c_p, c_{xi})$ in the *HSL* space. (Throughout this work we will be using *HSL* defined as the double-cone subset of a cylindrical space. The conversion algorithm is given in [19]. For the *Lab* conversion we used standard algorithm with *D*65 white point [24]).

A. Testing the hypothesis: Color similarity experiment

To test the relationship between perceptual similarity, color distances and angles in the *Lab* and *HSL* color spaces we have conducted a small-scale subjective experiment. Four subjects participated in the experiment. The subjects were given 10 sets of color samples. Each set consisted of a "prototype" color c_p, and five colors, $\{c_{xi}\}_{i=1,..5}$ with $D_{Lab}(c_p, c_{xi}) = const$. The distances between the prototype and the other colors ranged from 6 to 30. For each set the subjects were asked to order the samples according to the perceived similarity to the prototype. The sets were displayed in sequence on a computer monitor with light gray background under the daylight illumination. Each color sample was displayed in the 100×100 pixels window and could be moved on the screen to allow for the comparison with the prototype c_p. By analyzing the scores, the first thing we observed is that for $D_{Lab} < 7$ all colors were perceived as equally similar to the prototype. In other cases subjects identified the best and worst match unanimously, frequently leaving other samples unranked. Typically, the colors our subjects failed to rank were close in all three values. For the colors that were ranked by the subjects, the correlation between the subjects' rankings and rankings determined based on θ_{HSL} was 0.96. The correlation between the subjects' rankings and rankings determined based on D_{HSL} was 0.85, and the correlation with the rankings determined based on θ_{Lab} was 0.70. The slope of the least square regression line for the subjects' rankings and the rankings assigned according to θ_{HSL}, D_{HSL} and θ_{Lab} was 0.97, 0.84, was 0.87, respectively. These results indicate that θ_{HSL} and D_{HSL} (alone or combined) are better predictors of perceptual similarity between equidistant colors than θ_{Lab}, although alone neither represents an accurate color naming metric.

B. Designing the color naming metric

The metric we designed models the findings from the experiment. Let us assume a prototype c_p and arbitrary input color c_x. As discussed previously, for a given $D_{Lab}(c_p, c_x)$ a combination between $\theta_{HSL}(c_p, c_x)$ and $D_{HSL}(c_p, c_x)$ reflects the "reliability" of the *Lab* distance as a measure of similarity <u>in the color name domain</u>. Thus, we will use this relationship to modify D_{Lab} in the following manner. We first compute the distances between c_p and c_x in the *Lab* and *HSL* spaces:

$$D_{Lab}(c_p, c_x) = L = \sqrt{(l_p - l_x)^2 + (a_p - a_x)^2 + (b_p - b_x)^2}$$

$$D_{HLS}(c_p, c_x) = R = \sqrt{s_p^2 + s_x^2 - 2 s_p s_x \cos(h_p - h_x) + (l_p - l_x)^2} .$$

Given R, we then find a color $c_o : (h_o, s_o, l_o)$ with

$D_{HLS}(c_p, c_o) = R$, and

$$\theta_{HSL}(c_p, c_o) = (s_p s_o + l_p l_o) / \sqrt{(s_p^2 + l_p^2)(s_o^2 + l_o^2)} = 0 . \qquad (1)$$

Solving (1) results <u>in $h_o = h_p$</u>, $s_{o1,2} = s_p (1 \pm R / \sqrt{(s_p^2 + l_p^2)})$, $l_{o1,2} = l_p (1 \pm R / \sqrt{(s_p^2 + l_p^2)})$ and we chose a point that satisfies $\theta_{HSL}(c_x, c_o) < \pi$. This is illustrated in Fig. 2. According to our hypothesis, given the distance L, the optimal perceptual match is found along $\theta_{HSL}(c_p, c_o) = 0$. Assuming a small increment ΔR, we update the initial solution c_o as:

$$R_o = D_{HSL}(c_p, c_o) ,\ s_o = s_o (1 \pm \Delta R / R_o) ,\ l_o = l_o (1 \pm \Delta R / R_o)$$

until $D_{Lab}(c_p, c_o) \approx D$. For the given R, c_o is an "optimal" perceptual match to c_p. We denote this solution c_{opt}. As an estimate of perceptual dissimilarity between c_x and c_{opt}, we use the relative difference in the *HSL* space between c_{opt} and the projection $c_x \perp c_{opt}$:

$$\Delta d(c_p, c_x) = \frac{d(c_p, c_{opt}) - d(c_p, c_{ox})}{d(c_p, c_{opt})} = \frac{R_o - R\cos\beta}{R_o} = \\ = 1 - \frac{s_p s_x \cos(h_p - h_x) + l_p l_x - s_p^2 - l_p^2}{R_o \sqrt{s_p^2 + l_p{}^2}} \qquad (2)$$

As required by our model, in predicting the amount of perceptual similarity this formula takes into account both the distance and the angle in the *HSL* space. Therefore, we use this value to increase the *Lab* distance proportionally to the amount of dissimilarity Δd :

$$D(c_p, c_x) = D_{Lab}(c_p, c_x)(1 + k(c_p, c_x)\Delta d(c_p, c_x)) \qquad (3)$$

The factor $k(L)$ is introduced to avoid modifying distances between very close points and limit the amount of increase for large L ($k(L) = 0$, if $L < 7$ and $k(L) = const$, if $L > 30$).

C. Testing the metric

To test the stability of the method we have applied the metric to name different color regions in the *RGB* and *HSV* color spaces. Fig. 3 shows the transition of color names along the "color circle" in the *HSL* space and along the "red-yellow" line in the *RGB* space. As it can be seen in both color spaces color names change smoothly and the known color regions are identified accurately. To test the agreement with human observers we asked four subjects to review the color names assigned by our method to 100 randomly selected colors. Each subject received a different set of colors. The experimental setup was the same as in the first color naming experiment. The subjects agreed with the assigned color name in 91% of cases (362/400).

IV. Extracting the Color Composition of an Image

Human observation of a scene is typically very different from the recorded image. The method we have presented so far allows us only to name isolated samples of colors or assign color names to individual image pixels - the method does not account for color constancy, spatial relationships and interactions among colors. Therefore, the histogram of color names computed from the recorded image directly does not provide an accurate description of color content. To address the issue of color composition we need to address, at least to a certain extent, the issues of color constancy and chromatic adaptation, image segmentation and scene understanding. In this section we present an algorithm that takes into account the issues listed above and provides a description of the scene consistent with human observation. The algorithm has two parts. The first deals with the problem of color constancy, while the second provides image smoothing and segmentation suitable for the extraction of perceived colors.

A. Color constancy issues

The approach we adopt here is similar to the one taken by Lammens [5], as it seems to be fairly robust with respect to different lightning conditions, and to some extent even with respect to different sensing devices. We first gamma correct the image to make the intensities perceptually linear (we use the gamma correction factor of 2.2). Next we need to compensate for the differences in illumination conditions, with respect to both intensity and spectral characteristics. Here we rely on the most accepted hypothesis, the Von Kries law of coefficients, which assumes that different adaptations of a particular retinal area modify the overall sensitivities of the three fundamental color-response mechanisms, without affecting their relative spectral sensitivities [2]. Although the spectrum of the light source cannot be completely recovered from the image, as long as the spectrum of the light source is not too distorted, the Von Kries model provides reasonable results [2], [5]. In our scheme we take a variant of the "white world approach" to estimate the scene illuminant. This process is based on the assumption that the whitest point in the image comes from a surface that reflects light equally in all directions – thus by finding the "whitest" point for the given image we have an indication on what the illumination of the scene was. We therefore search the image for the "best representatives" of white, w, and black, b, and then use these values to apply a simple chromatic adaptation transform:

$$c_i'(x, y) = \frac{c_i(x, y) - b_i}{w_i - b_i} \qquad (4)$$

where: $c(x,y) = [c_1(x,y)\ c_2(x,y)\ c_3(x,y)]$ is the original color in the linear *RGB* color space, and $c'(x, y)$ is the transformed value. The best representatives for white and black are found as follows. The original image is first median filtered to refine the well-defined color regions and remove "noisy" pixels that do not contribute to the perceived colors. Next, each pixel is represented as: $(x, y) : (d_b(x, y), d_w(x, y))$, where $d_b(x, y)$ and $d_w(x, y)$ are the color name distances (3) between the given pixel and the black and white prototypes from the vocabulary, respectively. The black and white prototypes are then chosen as:

$$b = c(x_b, y_b),\ (x_b, y_b) = \arg\min(d_b(x, y)), \text{ and}$$
$$w = c(x_w, y_w),\ (x_w, y_w) = \arg\min(d_w(x, y)). \qquad (5)$$

This procedure can be understood as stretching of the gray axis of the original image and realigning it with the theoretical gray axis for perfectly homogeneous flat-spectrum illumination.

B. Spatial averaging and segmentation

An important process in the early stage of human vision is spatial averaging, which significantly accounts for the way we interpret color information. The smoothing process is partly due to the nature of the channel between the retina and visual cortex, where the neighboring receptors converge into one ganglion, while the groups of ganglions converge to single neurons in the visual cortex [15]. The amount of averaging depends on the spatial frequencies, spatial relationships among the colors, size of the observed objects and the global context. For example, the capability of human visual system to distinguish different colors drops rapidly for high spatial frequencies, consequently we describe texture areas with a single color, since only spatial averages of the microvariations are perceived. On the other hand we do not average isolated edges, as they represent object and region boundaries.

Based on these observations we model human perception as an adaptive low-pass filter operation, i.e. convolution of the input image with a localized kernel. In the proposed method we start by reducing the number of colors in the image to 128. We use the *LBG* vector quantization algorithm [16] to obtain a set of colors that optimally represent image colors in the *Lab* space (with respect to the mean square error). For each pixel (x, y), we then compute the local color contrast, $con(x,y)$, as:

$$con(x, y) = \frac{\| c(x, y) - \bar{c}(x, y) \|}{\| \bar{c}(x, y) \|} \quad (6)$$

where $\bar{c}(x, y)$ is the average color in a small neighborhood around $c(x, y)$ and $\| \cdot \|$ is the norm of the vector. The pixel (x, y) is considered an edge if its contrast exceeds a predefined threshold con_{min}. In the next step, to distinguish between the uniform regions, texture areas, and contour points, we use a sliding window to estimate the mean m, and variance v, of edge density for each pixel. Depending on these estimates we label pixels as: Type 1) uniform, $m = 0$, Type 2) noise, $m < t_{m1}$, Type 3) color edge, i.e. edge between two uniform regions, $t_{m1} < m < t_{m2}$, Type 4) texture edge, i.e. transition between uniform and textured region (or between two textured regions), $t_{m2} < m < t_{m3}$, Type 5) coarse texture, $m > t_{m3}$, $v > t_v$, or Type 6) fine texture, $m > t_{m3}$, $v < t_v$. Figs. 4 and 5 illustrate edge detection and pixel labeling processes. The labeling operation produces pixel maps, which control the smoothing process and determine the computation of dominant colors in the following way. Pixels labeled as noise are first removed and their color is changed to the neighboring uniform color. Since human eye creates a perception of a single dominant color within uniform regions, the amount of smoothing is largest for the uniform pixels. To allow for the highest amount of smoothing, the radius of the smoothing kernel is chosen adaptively for each uniform pixel, depending on the distance to the closest edge (color or texture). Pixels labeled as color edge and texture edge are not filtered. Also, since edges do not contribute to the way humans describe color content, these pixels are not used in computation of color composition. Finally, the amount of averaging performed in the textured areas is chosen based on the edge density, so that amount of averaging is higher for fine textures and lower for coarse textures. Thus, the perceived color at the location (x, y), $pc(x, y)$ is computed as $pc(x, y) = (c * g_{N(x,y)})(x, y)$, where $*$ is the convolution operator and $g_{N(x,y)}$ is the Gaussian kernel defined as:

$$g_{N(x_c,y_c)}(x, y) = k \exp(-\frac{x^2 + y^2}{\sigma^2}), \ \sum g_{N(x_c,y_c)}(x, y) = 1 \quad (7)$$

and $N(x,y)$ is the radius of the kernel. Note that $g_{N(x,y)}$ depends on the type of pixel in the center of the kernel, (x_c, y_c), as:

$$N(x, y) = \begin{cases} \| (x, y) - (x_e, y_e) \|, & (x_c, y_c) \text{ is Type } 1 \\ D, & (x_c, y_c) \text{ is Type } 5 \\ 2D, & (x_c, y_c) \text{ is Type } 6 \end{cases} \quad (8)$$

and (x_e, y_e) is the edge pixel closest to (x, y). Hence, the smoothing algorithm averages only within the uniform and textured regions, thus simulating the behavior of human visual system. Yet, due to imperfections in computing edge maps some boundary pixels are included in the smoothing operation, which produces slight "graying out" of the resulting image and causes some regions to be named differently. Therefore, we will again apply the modified Von Kries adaptation in the linear color space (the color-restoration problem can be viewed as the color-constancy problem, since our task is to preserve the same color appearance to the human observer). In computing (4), we use the same spatial location for the black and white representatives but "read" their color values from the smoothed image. The smoothed and "color-restored" image is then subjected to the mean-shift color segmentation [17] and used as an input to the color naming procedure. Prior to the segmentation the color value for each pixel labeled as color or texture edge is replaced with the color value of the closest uniform or texture region. The complete algorithm is illustrated in Figs. 4 and 5.

V. Color Naming Results

To extract the color composition of a scene, we start from the color-segmented image, and via (3) attach the color name to all pixels labeled as uniform or texture. In the next step, we compute the histogram of color names and use it to generate the description of color composition.

The structure and syntax of our color vocabulary allow us to describe color composition at different accuracy levels. According to the findings from our experiments, at the *fundamental* level, the color names are expressed as <generic hue> or <generic achromatic term> from the syntax. At the *coarse* level, color names are expressed as <luminance> <generic hue>, or <luminance> <generic achromatic term>. At the *medium* level, color names are obtained by adding the <saturation> to the coarse description. Finally, at the *minute* level, the complete <color name> as specified in the syntax is used. These different precision levels correspond to different color naming patterns in human language. For example, we use fundamental level when referring to well-know objects or when color information is not considered important (Color Listing Experiment). According to our experiments, the description of photographic images are mainly formulated with coarse or medium precision (Color Composition Experiment), while the color names at the minute level are typically used when describing isolated samples of colors (Color Naming Experiments) or specific objects and regions (Color Composition Experiment). We have applied the method to 40 images used in Color Composition Experiment. Tables II and III show color compositions extracted from images in Figs. 4 and 5, and a comparison to subjective descriptions from Color Composition Experiment. In general, the "computed" descriptions were in agreement with human judgements.

VI. Results, Discussion and Conclusions

We have presented a new framework for color naming, which follows relevant studies on human categorization, and, as our results demonstrate, captures human behavior in describing individual colors and color composition of complex images. It is important to note that although the color-naming metric (3) has been designed for the use in color naming applications (and with the given set of prototypes), the framework and algorithms can be easily modified to use any other distance function that satisfies the requirements for the color-naming metric. Candidate approaches include CIE94 [21] and CIECAM97s [22], or the recent color difference metrics (such as CMC or dE2000) addressing the non-uniformity of the *CIE-Lab*. Regardless of the metric used, there are numerous interesting applications for color naming in image processing and analysis. Using color names to label regions can often improve segmentation result, since neighboring regions that share the same color name can be merged. As illustrated in Fig. 5f, this also reduced false contours introduced by the color quantization process. In many cases color names only, or in combination with other features (spatial attributes, boundary, size features, etc.) can help enhance information about analyzed images and reveal their semantics. Therefore, our color-naming scheme might be seen as a "distant relative" of the spatially driven color retrieval algorithms, see for example [30]. Color name representation is especially suited for image/video retrieval tasks, where depending on the application domain and users needs color similarity criteria might include colors and relationships between the most important regions, color similarity of main objects, similarity in terms of color composition, or higher-level criteria, such as "images are pale", "bright", or "grayish". The descriptions derived from the color name histogram can be used to address the latter queries, as we can search for "dark" images, images with "impression of single color", or "monochromatic" images (Fig. 6a). Overall number of colors in the image also caries a degree of semantics. This query is used in Fig. 6b to identify and filter-out banners and graphics commonly used on Internet sites (images with a limited number of colors and color composition with predominantly vivid colors). By adding the semantics of other pictorial features (lines, energy, texture, shape) we can continue to build our knowledge of the image world. Figs. 6c,d show results of modeling categories "Outdoor Architecture" (based on the presence of "sky" region and long straight lines) and "Flowers" (based on the spatial relationships between vivid colored curved regions and green regions with irregular boundaries). (For a detailed overview of the image retrieval system used to provide these examples, underlying algorithms and image features see [31], [32]). These are only few interesting demonstrations of what can be accomplished with perceptually based color naming models and the whole area is certainly open for improvement. In generating the color description of a scene our model takes into account direct color measurements (hue, saturation and luminance), image elements (regions and edges), and their spatial properties (uniformity and "texturenes"). However, these three aspects are only elementary building blocks, which contribute to our color perception in an exceptionally complex way. As pointed by Jacobson and Bender in [18] the interactions between these building blocks produce many dimension of color experience. Hence, we need to go further and develop better models to predict how the appearance of color changes depending upon its neighboring colors and overall color content. Yet, even though we are still beginning to tackle this problem, our results demonstrate potential value of the color naming algorithms in many areas of image processing, visualization, computer graphics and human-machine interaction.

References

[1] K. Kelly, and D. Judd, "The ISCC-NBS color names dictionary and the universal color language (The ISCC-NBS method of designating colors and a dictionary of color names)", NBS Circular 553, November 1, 1955.

[2] G. Wyszecki, W. S. Stiles, Color science: Concepts and methods, quantitative data and formulae, John Wiley&Sons, New York, 1982.

[3] S. E. Palmer, Vision Science: Photons to phenomenology, MIT Press, 1999.

[4] B. Berlin and P. Kay, Basic Color Terms: Their Universality and Evolution, Berkeley: University of California, 1969.

[5] J. M. Lammens, "A computational model of color perception and color naming", Ph.D. Thesis, Univ. of Buffalo, June 1994.

[6] P. Kay and C. K. McDaniel, "The linguistic significance of the meaning of basic color terms", Language, vol. 54, no. 3, pp. 610-646, 1978.

[7] J. E. Cairo, "The neurophysiological basis of basic color terms", PhD Thesis, State University of New York at Binghamton, 1977.

[8] A. Munsell, A Color Notation, Munsell Color Company, Baltimore, 1946.

[9] Munsell Color, Munsell Book of Color, Macbeth, a division of Kollmorgen Corp., Baltimore MD, 1979.

[10] M. Miyahara, and Y. Yoshida, "Mathematical transform of (r,g,b) colour data to Munsell (h,v,c) colour data", Proc. SPIE Visual Communications and Image Processing '88, vol. 1001, pp. 650-657, 1988.

[11] A. Maerz, and M. R. Paul, A dictionary of color, McGraw-Hill Book Co., Inc., New York, NY 1930.

[12] T. Berk, L. Brownston, and A. Kaufman, "A new color-naming system for graphic languages", IEEE Computer Graphics and Applications, vol. 2, no. 3, pp. 37-44, May 1982.

[13] S. Tominaga, "A colour-naming method for computer color vision", Proc. of the 1985 IEEE Int. Conf. on Cybernetics and Society, pp. 573-577, Tucson, Arizona.

[14] D. H. Alman, "Industrial color difference evaluation", Color Research and Applications, vol. 18, pp. 137-139, 1993.

[15] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, Principles of neural science, Appleton and Lange, New York, 1991.

[16] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantization design", IEEE Trans. Communications, vol. 28, pp. 84-95, January 1980.

[17] D.Comaniciu, P.Meer, "Mean shift analysis and applications", IEEE Int. Conf. Comp. Vis., pp.1197-1203, Greece, 1999.

[18] N. Jacobson, and W. Bender, "Color as determined communication", IBM Systems Journal, vol. 35, no. 3&4, pp. 526-538, 1996.

[19] J. D. Foley, Fundamentals of Interactive Computer Graphics, Pearson Addison Wesley, 1982.

[20] S. Tominaga, "Color conversion using neural networks", SPIE Proceedings Vol. 3300 , pp.66-75, 1998.

[21] CIE Publication 116-1995, 'Industrial Colour-Difference Evaluation' (Vienna: CIE 1995)

[22] R.W.G. Hunt and M.R. Luo, The structure of the CIE 1997 Colour Appearance Model (CIECAM97S), Proceedings of the CIE Expert Symposium '97 Colour Standards for Image Technology, CIE Publ. No. x000, Central Bureau of the CIE, Vienna, 000-000 (1997).

[23] John J. McCann, "Uniform color spaces: 3D LUTs vs. algorithms". Proc, IS&T's 1999 PICS Conference, 204-208, 1999.

[24] C. Poynton, A technical introduction to digital video, John Wiley, 1996.

[25] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC 79 colour-difference formula," J. Soc. Dyers Colour. 100(4),128-132 (1984).

[26] R. McDonald, "Acceptability and Perceptibility decisions using the CMC color difference formula," Textiles chemist and colorist. 20 (6), 31-36 (1988).

[27] T. Belpaeme, "Simulating the Formation of Color Categories". In Nebel, B. (editor), Proceedings of the International Joint Conference on Artificial Intelligence 2001 (IJCAI'01), pp. 393-398.

[28] T. Belpaeme, "Reaching coherent color categories through communication". In Kröse, B. et al. (eds.), Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC'01), pp. 41-48, Amsterdam, The Netherlands.

[29] N. Moroney, “Unconstrained web-based color naming experiment”, In Color imaging: Device-dependent color, color hardcopy and graphic arts VIII, Reiner Eschbach, Gabriel Marcu, Editors, Proc. of the SPIE, 2003.

[30] L. Cinque, G. Ciocca, S. Levialdi, A. Pellicanò, R. Schettini, “Color-based image retrieval using spatial-chromatic histograms,” Image and Vision Computing, Vol. 19, 2001. (SCI 3225), pp. 879-986.

[31] A. Mojsilovic, J. Gomes, and B. Rogowitz, “Semantic-friendly Indexing and Querying of Images Based on the Extraction of the Objective Semantic Cues”, International Journal of Computer Vision, vol. 56, pp. 79-107, 2004.

[32] A. Mojsilovic, and B. Rogowitz, "Capturing image semantics with low-level descriptors", Proc. Int. Conference on Image Processing, ICIP 2001, Thessaloniki, Greece, September 2001.

Aleksandra (Saška) Mojsilović (M’92) was born in Belgrade, Yugoslavia, in 1968. She received her Ph.D. in Electrical Engineering, from University of Belgrade, Yugoslavia, in 1997, and since then has worked at IBM Research (2000-present), Bell Laboratories (1998-2000) and as a Faculty Member University of Belgrade (1997-1998). Her main research interests include multidimensional signal processing, pattern recognition, modeling, image analysis and human perception. In 2001 Dr. Mojsilović received the Young Author Best Paper Award from the IEEE Signal Processing Society. Dr. Mojsilović is a member of the IEEE Multidimensional Signal Processing Technical Committee and Associate Editor for the IEEE Transactions on Image Processing.

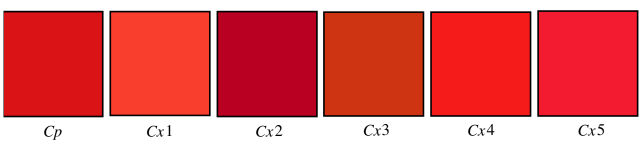


Fig. 1: Example of perceptual differences among equidistant pairs of colors. The color data is given in Table I.

TABLE I: COLOR DATA FOR FIG. 1: COLOR DISTANCES AND ANGLES BETWEEN c_p AND c_{xi} IN THE *LAB* AND *HSL* COLOR SPACES

c_p: (L,a,b) = (47,70,55) (h,s,l) = (0,90,85)						
	(L,a,b)	(h,s,l)	D_{Lab}	D_{HLS}	θ_{Lab}	θ_{HLS}
c_{x1}	(57,70,55)	(5,82,98)	10.0	17.0	4.8	7.5
c_{x2}	(37,70,55)	(349,100,73)	10.0	24.0	5.3	11.1
c_{x3}	(47,60,55)	(11,91,81)	10.0	17.8	4.4	8.3
c_{x4}	(53,76,60)	(0,89,96)	9.9	11.0	0.9	3.8
c_{x5}	(53,76,50)	(354,89,95)	9.9	13.7	4.8	5.5

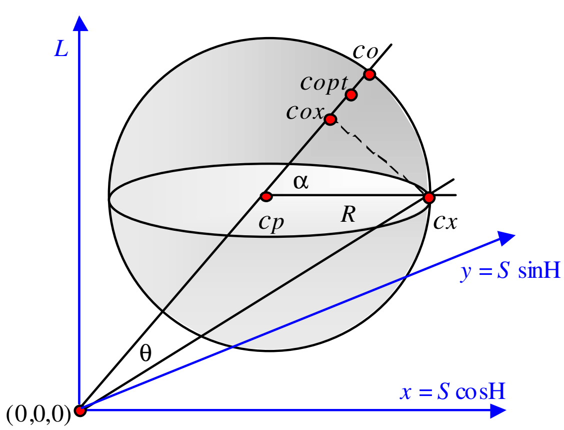


Fig. 2: Deriving the metric.

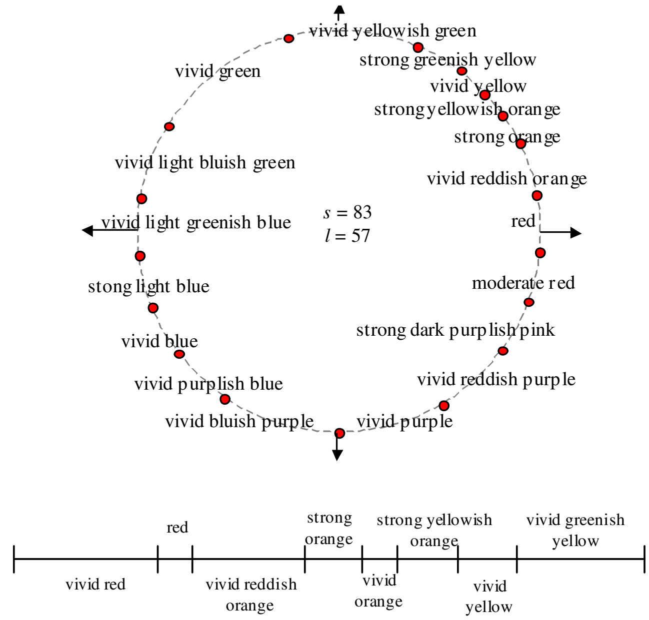


Fig. 3: Color names assigned to a “color circle” in the *HSL* space defined with s = 83 and l = 135, and along a red-yellow line (r = 235, 0 < g < 255, b = 20) in the *RGB* space. To compute the distribution of color names, each line has been quantized into 200 points and the color-naming algorithm has been applied to each point. The crosses shown on each line indicate the boundary points between two different color names.

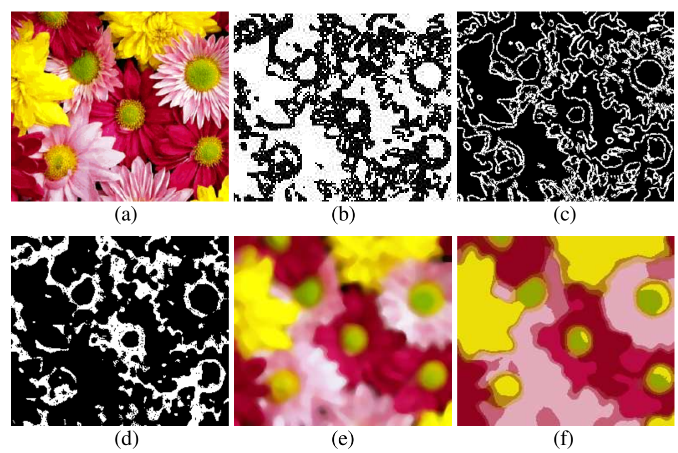
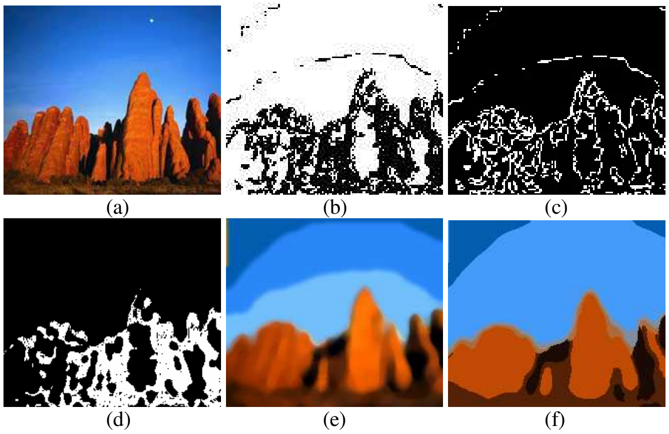


Fig. 4: (a) Original image, (b) uniform pixels (white), (c) color and texture edges (white pixels), (d) textured regions (white pixels), (e) smoothed and color-restored image, and (f) color segmented image

TABLE II: COLOR COMPOSITION EXTRACTED FROM THE IMAGE IN FIG. 4 AND THE COMPARISON TO THREE TYPICAL DESCRIPTIONS OUR SUBJECTS ENTERED IN THE COLOR COMPOSITION EXPERIMENT.

<table>
<tr><td>Minute</td><td>Medium</td><td colspan="2">Coarse</td></tr>
<tr><td>strong purplish red (29%)
vivid yellow (21%)
moderate pink (22%)
moderate dark pink (9%)
strong light yellowish green (4%)</td><td>strong red (29%)
vivid yellow (21%)
moderate pink (22%)
moderate dark pink (9%)
light green (4%)</td><td colspan="2">red (29%)
yellow (21%)
pink (22%)
dark pink (9%)
light green (4%)</td></tr>
<tr><td>Fundamental</td><td colspan="3">Subjects’ descriptions</td></tr>
<tr><td>red (29%)
yellow (21%)
pink (31%)
green (4%)</td><td>purple
yellow
light green
pink
dark pink</td><td>red
yellow
pink</td><td>yellow
pink
green
purple</td></tr>
</table>

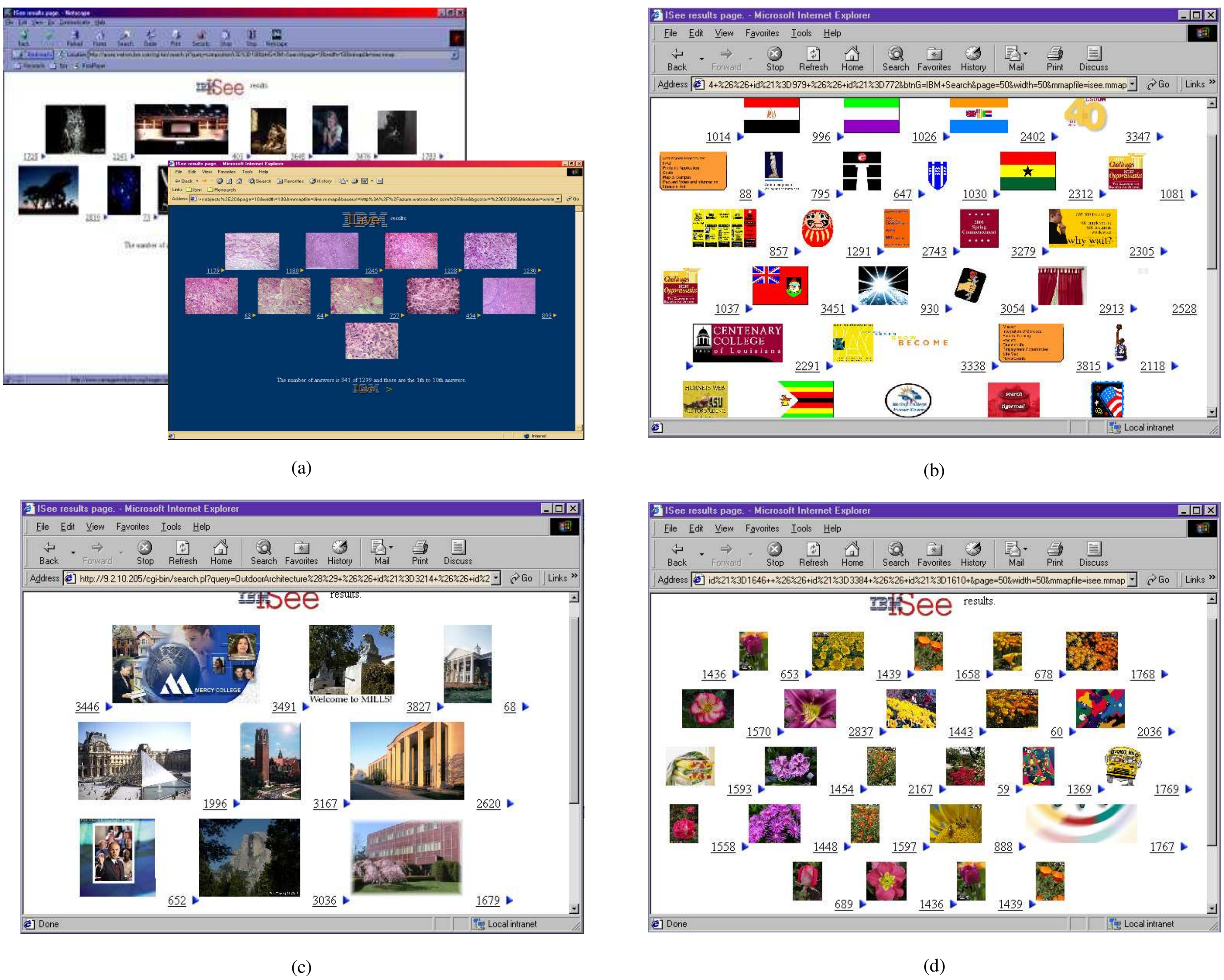


(a) (b) (c)
(d) (e) (f)

Fig. 5: (a) Original image, (b) uniform pixels (white), (c) color and texture edges, (d) textured regions (white pixels), (e) smoothed and color-restored image, and (f) color segmented image. The mean-shift color segmentation is further improved by merging regions that share the same color name.

TABLE III: COLOR COMPOSITION EXTRACTED FROM THE IMAGE IN FIG. 5 AND THE COMPARISON TO THE TYPICAL DESCRIPTIONS OUR SUBJECTS ENTERED IN THE COLOR COMPOSITION EXPERIMENT.

Minute	Medium	Coarse	
vivid blue (49%) *strong purplish blue* (7%) *vivid reddish orange* (23%) *deep brown* (12%) *brownish black* (4%)	*vivid blue* (49%) *strong blue* (7%) *vivid orange* (23%) *brown* (12%) *black* (4%)	*blue* (56%) *orange* (23%) *brown* (12%) *black* (4%)	
Fundamental	Subjects' descriptions		
blue (56%) *orange* (23%) *brown* (12%) *black* (4%)	*blue* *light blue* *dark orange* *brown* *black*	*orange* *brown* *very light blue* *black*	*blue* *brown* *black*



(a) (b)
(c) (d)

Fig. 6: Semantic cues derived from color composition: (a) "dark" images and "monochromatic" images, (b) Web banners: Images with a limited number of colors and a color composition with predominantly vivid colors, (c) "Outdoor architecture" and (d) "Flowers."