A STATISTICAL APPROACH TO CULTURE COLORS DISTRIBUTION IN VIDEO SENSORS

Angela D'Angelo, Jean-Luc Dugelay

Eurecom
Multimedia Communications Department
Sophia Antipolis Cedex, FRANCE

ABSTRACT

Color is an important and powerful cue in the distinction and recognition of objects, however identifying or matching the surface color of a moving object in a video surveillance system is very critical. In this paper we describe a new approach aiming at providing a comparison of the more widely used color spaces in identifying colors in motion sensors. We employ a statistical approach by conducting extensive data mining on video clips collected under various lighting conditions and distances from several video-cameras. The goal is to learn how individual colors can drift in different illumination conditions and with different color spaces.

1. INTRODUCTION

In addition to shape, texture, and some other geometric properties, color is an important and powerful cue in the distinction and recognition of objects and for this reason it has been widely used in several applications of image and video processing. Being able to identify colors in video sequences can help, for example, tasks of object tracking and object search. One important application could be to track a given person, based on the clothing colors, across the fields-of-view (FOV) of multiple cameras. Another application could be to search a person with a specific shirt color in a camera network, such as to localize missing children in a crowded amusement park. The color cue plays an important role in enhancing the reliability of these tracking and search tasks.

Identifying or matching the surface color of a moving object in a video surveillance system, however, is very critical. Traditional color models provide only little help, since the surface of an object is usually not flat, the object's motion can alter the surface's orientation, and the lighting conditions can vary when the object moves. To tackle this research problem, many color constancy algorithms [1] have been developed to reduce color variation from fluctuation in source illumination. Despite decades of research in color constancy algorithms, these algorithms cannot be used to reliably identify colors in motion. The first issue is that most algorithms assume scene illumination to be uniform in the region of interest or changing gradually. Such an assumption almost always fails in a surveillance scenario. Second, most color-constancy algorithms depend on reliable estimates of parameters such as angles between light sources and the object, reflection angles, and surface materials. These parameters can be unknown or difficult to estimate in real-time when the object being observed is in motion.

Instead of taking the route to model variations in surface orientation, extended light, secondary reflection, and varying color sensitivity of cameras, another popular and simple approach for color identification, widely used in color based object tracking, is the introduction of color invariant models. The goal of color invariants, originally used in image retrieval, is to recognize multicolored objects invariant to substantial change in viewpoint, object geometry and illumination.

Many works have been conducted in this direction and many color spaces and distance metrics to evaluate the similarity between colors have been introduced in the scientific literature [2]. However, quite surprisingly, most of the papers does not provide strict justification of their color space choice neither a comparison of the existing color spaces. This is probably because of the possibility to obtain acceptable results on limited dataset with almost any color space.

Even if it is generally agreed that there is no single color system which is suitable for all color images, an analysis of the performance of the most popular color spaces in identifying colors in motion in video sensors would be a significant help in some tracking and search tasks. This is the goal of the proposed work.

In this paper we describe a new approach aiming at providing a comparison of the more widely used color spaces. We employ a statistical approach by conducting extensive data mining on video clips collected under various lighting conditions and distances from several video-cameras. The goal is to learn how individual colors can drift in different illumination conditions and with different color spaces. The diversified samples of pixel colors collected in real conditions will allow us to derive general conclusions on the performance of the analyzed color spaces in identifying colors in motion. This study can be considered as a preliminary study for the development of a new robust color detection framework in video.

2. RELATED WORKS

In this section a brief description of color constancy algorithms is provided[1].

Color constancy is the ability of a vision system to accurately describe the color of an object in spite of variations in illumination conditions. By ignoring atmospheric attenuation and scattering, which does not play a significant role in color appearance, the critical elements in color constancy are light sources, sensors, and how an object interacts with the incident light. This interaction is often characterized as the ratio of the reflected light and the incident light, which is commonly referred to as the bi-directional reflectance function, or the BDRF. The BDRF depends on many factors, the most important ones are the geometric configuration (i.e., the surface orientation relative to the viewer and the light source) and the wavelength. Measuring BDRF, however, is a tedious and difficult task and usually certain reasonable simplifications are made. The most common assumption is that surfaces are isotropic, or that the BDRF does not change significantly if a surface is rotated about its normal.

Another complication is that there exist two major reflection mechanisms: interface reflection and body reflection. Interface reflection occurs at the junction between an object and the surrounding medium. In contrast, body reflection is usually considered Lambertian and wavelength dependent. Most color constancy algorithms concentrate on analyzing and modeling the body reflection component as it carries the most discriminative information for inferring an object true color.

Many algorithms for color constancy have been introduced in the scientific literature in the last twenty years. The simplest model that accounts for illumination variation is to compute a single statistic, such as a mean, to estimate scene illumination, which is assumed to be uniform in the region of interest. This leads to the so-called greyworld algorithms [4].

Gamut mapping colour constancy [5] attempts to determine the set of diagonal matrices taking the gamut of image colours under an unknown illuminant into the gamut of colours observed under a standard illuminant. While the algorithm performs well on images of flat, matte, uniformly illuminated scenes, its performance on images of more realistic scenes can be poor (it has a particular problem with specularities) and furthermore the algorithm is computationally intensive.

Linear decomposition methods [6] model illumination change using a linear transformation. This model is justified if illumination and surface reflectance can be expressed as linear combinations of a small number of basis functions. In particular, the diagonal linear model, which maps the image taken under one illumination to another by simply scaling each color channel independently, has been shown to be effective in some application scenarios.

In summary, color perception and understanding is an extremely complicated and nonlinear science. To simplify the analysis, many color constancy models assume a single camera, a frontal surface orientation or a spatially-invariant illumination. In the reality, unluckily, we must account for spatially-distributed surveillance cameras operating under different lighting conditions and with varying color sensitivity. The complexity of such modeling for identifying color in motion in video sensors, makes the task difficult, if not impossible.

3. CULTURE COLORS DISTRIBUTION

As explained in the introduction, the goal of the proposed work is to learn how colors can drift in different illumination conditions and with different color spaces. In order to provide a very general approach, we have to collect pixels describing different colors. To exploit the colors choice we need to refer to the branch of color categorization. Color categorization is intrinsically related to color naming, which lies at the boundary between different fields of cognitive sciences: visual perception and linguistics. Color naming is about the labeling of a given set of color stimuli according to their appearance in a given observation condition. Pioneering this field, the work of Berlin & Kay [7] traces back to early 1970s, and have settled the ground for the proliferation of the next wave of cognitive studies. Based on this work, the set of color terms that can be considered as universal constants (among the languages that have at least the necessary number of color terms) are the following: black, white, red, yellow, green, blue, brown, purple, pink, orange, and grey.

Based on the above considerations we quantize the entire color space into the above eleven bin. These colors are usually referred to as *culture colors*, which have been used in literature of different cultures in the past two thousand years to refer to colors [7]. Moreover most of the video processing applications are based on color quantization using the previous categories. It is worth pointing out that in the scientific literature many other color quantization schemes have been proposed but the approach of the proposed algorithm is equally applicable to them.

One might argue that having a finer quantization may

[1] We do not aim to draw up a complete and exhaustive overview of all color constancy methods but only to give the basic concepts. A comprehensive survey and comparison of some popular ones is provided in [3].

better discern different objects. Unfortunately, finer quantization leads to less reliable color prediction, and can be counter-productive in improving prediction accuracy.

3.1. Mining testbed

Instead of taking a generative approach as discussed in Section 2, we employ a discriminant approach to address the complexity of the color identification. The idea is to collect a big number of pixels whose colors correspond to culture colors, under various lighting conditions, from different cameras and from several distances from the cameras. Such a task could be not trivial, if not impossible, since it is very time-consuming to carry-out the experiments and, above all, it is almost impossible to test all the possible real conditions. Some authors treat this problem by recording videos of people wearing certain colored shirts and walking around in the field-of-view of a camera network. In this case, however, the results of the color detection system are limited to few sensors and few illumination conditions.

Our idea to obtain a so diversified dataset is to associate to each of the culture color a sport team (football teams, cyclist teams, rugby teams, etc.) with the color of the uniform corresponding to that color, and to randomly collect from the web video clips of the selected teams. Specifically, we collected 4 or 5 video clips for each team and around 120 pixels for each color taken in different frames of the clips. This procedure allows us to obtain 1355 pixel samples collected in real illumination conditions, different position of the objects (the players) with respect to illumunitaion and, with very high probability, taken from different cameras.

3.2. Color spaces

Colorimetry, computer graphics and video signal transmission standards have given birth to many colorspaces with different properties. In this paper we take into account five of the most popular and widely used color spaces: RGB, normalize RGB, HSV, Lab, YUV. In the following, we briefly review these color spaces[2].

3.2.1. RGB

RGB is a colorspace originated from CRT display applications, when it was convenient to describe color as a combination of three colored rays (red, green and blue). It is one of the most widely used colorspaces for processing and storing of digital image data.

RGB is a device-dependent color space. This fact, together with the high correlation between channels, the significant perceptual non-uniformity, the mixing of chrominance and luminance data, make RGB not a very favorable choice for color analysis and color based recognition algorithms.

[2] Please refer to [2] for details.

3.2.2. Normalized RGB

Normalized RGB is a representation that is easily obtained from the RGB values by a simple normalization procedure:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \tag{1}$$

As the sum of the three normalized components is known ($r+g+b=1$), the third component does not hold any significant information and can be omitted, reducing the space dimensionality.

A remarkable property of this representation is that for matte surfaces, while ignoring ambient light, normalized RGB is invariant, under certain assumptions, to changes of surface orientation relatively to the light source.

3.2.3. HSV

Hue-saturation based colorspaces (HSI, HSV, HSL) were introduced when there was a need for the user to specify color properties numerically. They describe color with intuitive values. Hue (H) defines the dominant color of an area, saturation (S) is a combination of light intensity and how much it is distributed across the spectrum of different wavelengths and value (V) is related to the color luminance, as described in the following equation:

$$H = \arccos \frac{\frac{1}{2}\left((R-G)+(R-B)\right)}{\sqrt{\left((R-G)^2+(R-B)(G-B)\right)}} \tag{2}$$

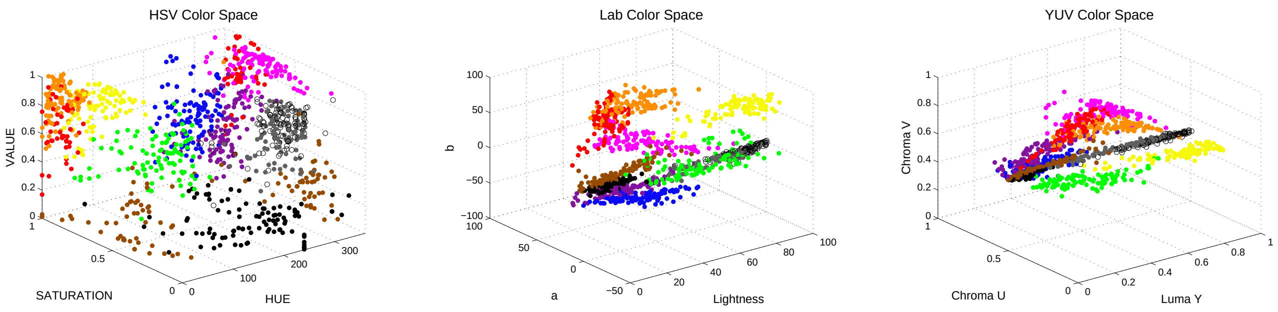
$$S = 1 - 3\frac{\min(R,G,B)}{R+G+B} \tag{3}$$

$$V = \frac{1}{3}(R+G+B) \tag{4}$$

The intuitiveness of the colorspace components, explicit discrimination between luminance and chrominance properties and its invariant to highlights at white light sources, and also, for matte surfaces, to ambient light and surface orientation relative to the light source, made this colorspaces widely used in image processing applications, such as skin color segmentation.

3.2.4. Lab

The Lab color space is a color-opponent space with dimension *L* for lightness and *a* and *b* for the color-opponent dimensions. Lab is now more often used as an informal abbreviation for the CIE (*International Commission on Illumination*) 1976 color space (also called CIE LAB). CIE LAB is



(a) (b) (c)

Fig. 1. Distribution of the culture colors: (a) HSV color space; (b) Lab color space; (c) YUV color space.

a perceptually uniform colorspace. Perceptual uniformity means that a small perturbation to a component value is approximately equally perceptible across the range of that value. The price for better perceptual uniformity is complex transformation functions from and to RGB space (for sake of brevity we do not report these functions), demanding far more computation than most other colorspaces.

3.2.5. YUV

In the YUV color space, the Y component determines the brightness of the color (referred to as luminance or luma), while the U and V components determine the color itself (the chroma). It is defined, with respect to the RGB color space, by the following equation:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.288 & 0.436 \\ 0.615 & -0.515 & -0.1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5}$$

One aspect of YUV is that it is possible throw out the U and V components and get a grey-scale image. Since human eye is more responsive to brightness than it is to color, many image compression formats throw away half or more of the samples in the chroma channels to reduce the amount of data to deal with, without severely destroying image quality.

3.3. Visual distribution

We want now to analyze how individual colors can drift from the real color in different conditions and in the described color spaces. In Fig.1 the plots of the colors distributions for each culture color in different color spaces are shown[3]. Ideally, each color should appear as one point in the color space. However, as expected, with the interferences of various enviromental factors, the pixel colors of a culture color spread out like a cloud. While this effect is present in all the analyzed color spaces, it changes drastically from one color space to another and for the different colors, as we can notice by observing the plots. Obviously, the more the cloud of a color spreads in the space, the more difficult will be the color identification. This means, in fact, that the color changes too much its value in different illumination conditions. In the same way, overlapping distributions shows a difficulty of the color space in discriminating different colors. Based on these two main observations, some preliminary considerations can be done by observing the plots.

From a general point of view, the more dispersive color space seems to be the HSV color space. In this case, in fact, the clouds of the distribution of almost all the colors are quite spread in the space, as we can see looking at the plot in Fig.1.(a). On the other side, we expect that the Lab or the YUV color space provide a good tool to identify the eleven culture colors because the distribution, for almost all the colors, is usually restricted on a specific area of the space, as we can see looking at the plot in Fig.1.(b) and Fig.1.(c). The HSV color space is widely used in different applications of image and video processing (color segmentation, skin detection, shadow removing, etc.) thanks to its property to be robust against change in illumination and shadow. Thus the colors distribution we obtained for the HSV space is quite surprising. On the other hand, it is well known that the Lab color space provides good results in color identification since it is perceptually uniform.

Analyzing more in details the plots in Fig.1, we can also observe that the most difficult task for all the color spaces is discriminating grey and white. These two colors, in fact, are always overlapped thus no distance metric[4] will be able to robustly identify the two colors. Some problems appear also in discriminating red and orange in the HSV color space, since these two colors can drift each other changing the il-

[3]For sake of brevity only three of the five plots are reported, the ones showing more interesting results.

[4]Some popular distance metrics usually used in color detection algorithms are the euclidean, cylindric and angular metrics.

lumination condition.

The above considerations are confirmed by evaluating the first and second order statistics of the distributions, which provide, respectively, an indication of the overlap of the colors and their spread in a specific color space. For sake of brevity, the plots of the statistics are not reported.

4. FUZZY CLUSTERING

In the previous section we have seen how the culture colors spread in different color spaces and drift from one color to other colors changing sensors and illumination conditions. While the spreading factor is reasonable and intuitive to understand the tricky problem of color identification, the most important aspect is the possibility to design a detector able to partition the colors, thus we need to refer to the problem of data clustering.

Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. While in hard clustering data is divided into distinct clusters where each data element belongs to exactly one cluster, in fuzzy clustering data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These values indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is the process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

Fuzzy clustering is particular suited to color quantization since color boundaries are not well defined (as we can noticed by the plots in Fig.1). In the proposed system we adopt the **fuzzy k-nearest neighbors algorithm** (KNN) introduced by Keller & al. [8], which works as follows. Let us assume that a training set of *m* samples vectors $Z_1, Z_2, ..., Z_m$ is available. Let X be a new vector considered as the input to be classified. Fixed a value of k, the first step consists in identifying, among these sample vectors, the k nearest neighbors $Y_1, Y_2, ..., Y_k$ of the input X. Then the membership vectors of the selected labeled samples Y are combined to find the membership vector of the input X, where the membership vector describes the probabilities of the membership to the possible C classes. Let $u_i(X)$ be the membership of the input X to the i^{th} class (with $i \leq C$), and w_{ij} the membership of its j^{th} neighbor Y_j to the same class ($w_{ij} = u_i(Y_j)$), then (with $m > 1$):

$$u_i(X) = \frac{\sum_{j=1}^{k} w_{ij} \left(\frac{1}{\|X - Y_j\|} \right)^{\frac{2}{(m-1)}}}{\sum_{j=1}^{k} \left(\frac{1}{\|X - Y_j\|} \right)^{\frac{2}{(m-1)}}} \tag{6}$$

In the above formula, the inverse distance is used to weight the memberships degrees of the samples by assigning a higher weight to closest vector.

	RGB	norm RGB	HSV	Lab	YUV
Accuracy	91.3	89.9	86.8	94.1	91.4

Table 1. Accuracy of the proposed color space.

OUTPUT CLASS \ TARGET CLASS	Red	Orange	Black	Pink	White	Gray	Purple	Brown	Blu	Yellow	Green
Red	97	10	0	0	0	0	0	0	0	0	0
Orange	1	90	0	0	0	0	0	0	0	0	0
Black	0	0	96	0	0	1	1	7	1	0	1
Pink	1	0	0	99	0	1	0	0	0	0	0
White	0	0	0	0	88	19	0	0	0	0	0
Gray	0	0	0	1	12	79	0	0	0	0	0
Purple	0	0	0	0	0	0	102	0	2	0	0
Brown	1	0	4	0	0	0	0	96	0	0	0
Blu	0	0	0	0	0	0	1	0	101	0	0
Yellow	0	0	0	0	0	0	0	0	0	96	0
Green	0	0	0	0	0	0	0	1	0	0	95

Table 2. Confusion matrix of the Lab color space.

We tested the above classifier on a new dataset of 1104 samples collected as described in Sec.3.1. It is quite usual to describe the performance of a classifier in terms of *true positive rate* (TPR) and *false positive rate* (FPR), where TPR is the proportion of positive instances that were correctly reported as being positive and FPR is the proportion of negative instances that were erroneously reported as being positive. These values are commonly depicted in the form of a receiver operating characteristic (ROC) curve, as reported in Fig.2 for all the culture colors (for a better comparison the curves are zoomed in the upper left part). By observing the plots we can notice how the HSV color space provides the worst results, while the Lab color spaces show good performances in color identification task. These consideration are confirmed by the values of the accuracy reported in table 1. To investigate more in details the performance of the Lab color space for all the culture colors, we can have a look at the confusion matrix reported in table 2. The system is able to perfectly recognize yellow, pink, green or purple (which have a correct classification rate of respectively 100%, 99%, 99% and 98.1%), while it has some problems, as expected, in identifying gray and white (correct classification rate of 88% and 79% respectively).

5. CONCLUSIONS

In this paper we describe a new statistical approach to provide a comparison of the color spaces in describing and identifying colors in motion in video sensors. The proposed approach allow us to test the system for different colors in real illumination conditions in order to draw very general considerations on the performances of different color spaces. Even if we believe that it is almost impossible to define a single color system suitable for all color images and for all images and video processing applications, the proposed analysis is a contribution for the comprehension of

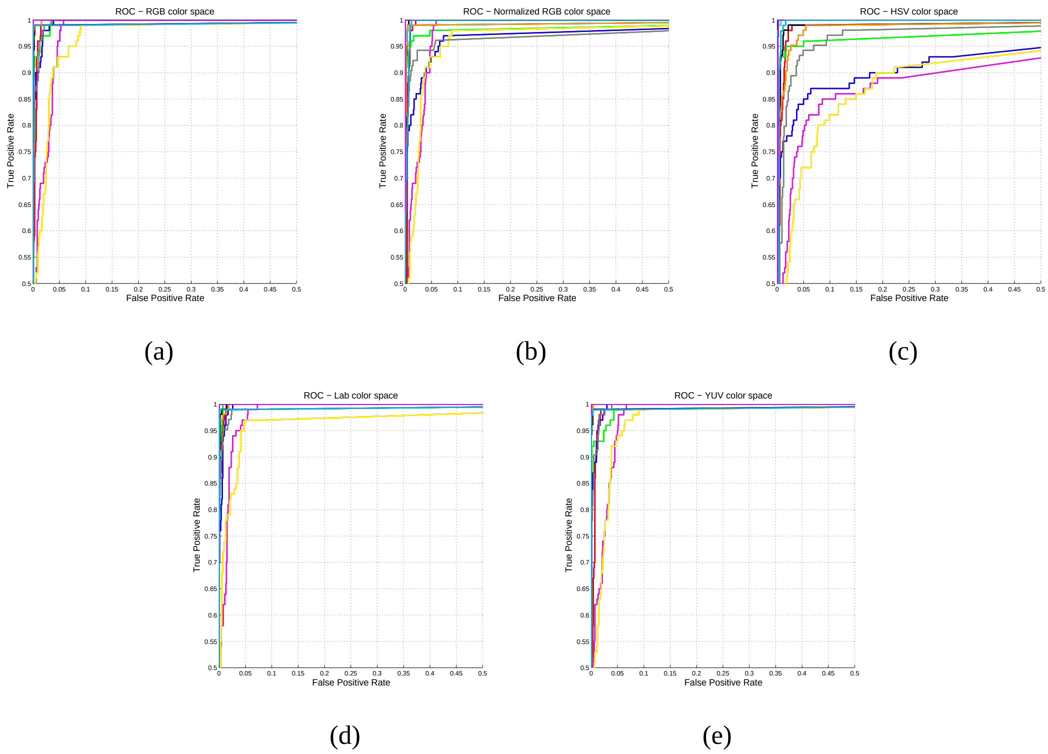


Fig. 2. ROC curves: (a) RGB; (b) Normalized RGB; (c) HSV; (d) Lab; (e) YUV.

the performance of some popular color spaces in representing the surface color of a moving object in video. The proposed work could be an useful tool for the development of a new robust and general purpose color detection framework that can find application in searching and tracking tasks in video surveillance system.

Acknowledgment

The research described in this paper has been partially funded by the FR OSEO BIORAFALE project. Information regarding projects activities can be found at the Multimedia Image processing Group (MIG) website: http://image.eurecom.fr/.

6. REFERENCES

[1] D. Hilbert, "Color constancy and the complexity of color," *Philosophical Topics*, vol. 33, no. 1, pp. 141, 2005.

[2] G. Wyszecki and WS Stiles, "Color science: concepts and methods, quantitative data and formulae," 2000.

[3] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms-part i: Methodology and experiments with synthesized data," *IEEE transactions on Image Processing*, vol. 11, no. 9, pp. 972–984, 2002.

[4] R. Gershon, A.D. Jepson, and J.K. Tsotsos, "From [R, G, B] to surface reflectance: computing color constant descriptors in images," *Perception*, vol. 17, pp. 755–758, 1988.

[5] G. Finlayson and S. Hordley, "A theory of selection for gamut mapping color constancy," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. Institue of Electrical Engineers Inc. (IEEE), 1998, pp. 60–65.

[6] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Spectral sharpening: sensor transformations for improved color constancy," *Journal of the Optical Society of America A*, vol. 11, no. 5, pp. 1553–1563, 1994.

[7] B. Berlin, F. Birren, J. Gage, J.A. Gatto, T. Hoban, and H. Kuppers, "Basic color terms; their universality and evolution [by] Brent Berlin and Paul Kay," 1993.

[8] J.M. Keller, M.R. GRAY, and JR. GIVENS, "A Fuzzy K-Nearest Neighbor Algorithm ," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 581, 1985.