

Color Matching using Time Series Searching in Color Databases

Pareo de Colores mediante Búsqueda de Series de Tiempo en Bases de Datos de Color

E. M. Felipe Riverón^a, J. G. Figueroa Nazuno^a, A. F. Gutiérrez Tornés^a,
R. Barrón Fernández^a y F. I. Cervantes Alarcón

^aCenter for Computing Research, National Polytechnic Institute, Mexico, DF
Juan de Dios Batiz s/n, esquina a Miguel Othon de Mendizabal, Col. Nueva Industrial Vallejo, P.O: 07738
edgardo@cic.ipn.mx; jfn@cic.ipn.mx; atornes@cic.ipn.mx;
rbarron@cic.ipn.mx; fabian_isr@yahoo.com.mx

Article received on December 06, 2005; accepted on April 02, 2007

Abstract

The goal of this paper is to document the use of time series in color matching. Although time series are not commonly used for this purpose, the results obtained, based on the principle of similarity in time series using the Euclidean distance, establish the validity of its use for color matching applications. The accuracy of color matching was based on the measures of reflectivity versus wavelength of samples given by a spectrophotometer. The error estimation was calculated using a database of 1001 elements. The matching module has been tested with six samples included and not included in the database. All of them gave an error lower than the estimated absolute error of 11.36.

Keywords: Time series; Color matching; Pattern recognition; Similarity; Color management systems.

Resumen

El propósito de este trabajo es documentar el uso de las series de tiempo en el pareo de colores. Aunque las series de tiempo no son usadas comúnmente con este fin, los resultados obtenidos, basados en el principio de similitud en las series de tiempo mediante la distancia euclidiana, establece la validez de su uso en las aplicaciones de pareo de colores. La exactitud del pareo de colores se basó en las medidas de la reflectividad contra la longitud de onda de las muestras, dadas por un espectrofotómetro. La estimación del error fue calculada a partir de una base de datos de 1001 elementos. El módulo de pareo ha sido probado con seis muestras incluidas y no incluidas en la base de datos. Todas ellas dieron un error menor que el error absoluto estimado de 11.36.

Palabras clave: Series de tiempo; Pareo de colores; Reconocimiento de patrones; Similitud; Sistemas de administración del color.

1 Introduction

Color mapping or color matching is a process in which a color is transformed into its visually closest color in a given target color space. A color space is a mapping of color components onto another coordinate system in three or more dimensions [1].

The principal objective of this work is to use color matching as an Internet-based service for matching paint colors [2]. Generally, factories that produce paints match colors through a color swatchbook prepared from the paints and mixes of colors produced by them. The cost in equipments of color matching in real time is relatively high, and it is done only "in situ", never remotely (e.g., via the Internet) [3], [4], [5].

It was of our interest to create a computer system for color matching. This task is not a trivial one because it depends on the preparation of color samples and their maintenance. The problem arose when it was necessary to find a procedure to achieve a fast, robust, and error-free way of color matching. We considered that the solution given by the use of the principle of similarity in time series might be viable for this purpose.

2 Some Concepts

In order to frame our contribution in the proper context we begin with a brief review of the necessary background material about Color and Time series.

2.1 Color

Color is produced when light in the visible region of the electromagnetic spectrum approximately corresponds to the wavelength from 380 to 780 nm. A color of a visible spectrum can be specified in terms of tristimulus values. The color matching properties of all spectra with respect to a given set of primaries can be specified in terms of the color matching properties of their monochromatic spectra.

The CIE 1931 recommendations defined a standard colorimetric observer by providing a set of color matching functions, the CIE RGB $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$, associated with monochromatic primaries at wavelengths of 700.0, 546.1, and 435.8 nm, respectively, with their radiant intensities adjusted so that the tristimulus values of the equi-energy spectrum are all equal (with a constant spectral power distribution). In order to avoid negative values at all wavelengths, the CIE 1931 defined also as a linear transformation the color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ in the wavelength range $\lambda = 380$ to 780 nm at wavelength intervals of $\Delta\lambda = 5$ nm. These transformations are suitable for describing color-matching when the angular subtense of the matching fields at the eye is between one and four degrees.

Because an infinite number of transformations can be defined to meet the non-negative requirement, additional criteria were used: the choice of $\bar{y}(\lambda)$ coincident with the *luminous efficiency function* (that gives the relative sensitivity of the eye to the energy at each wavelength) and the normalization of the three color matching functions so as to yield equal tristimulus values for the equi-energy spectrum [15].

CIE XYZ tristimulus space varies widely over the color space, because it is perceptually non-uniform in that equal perceptual differences between colors do not correspond to equal distances in the tristimulus space. The problem to develop a uniform color space has been decomposed into two sub-problems: one of determining a uniform lightness scale and the other of determining a uniform chromaticity diagram for equi-lightness color stimuli. Based on this, the CIE has recommended the approximately perceptually uniform CIE 1976 L*a*b* (CIELAB) space, employing a lightness scale, L*, that depends only on the luminance value Y, and two chromatic values a* and b* [15].

To classify and qualify colors according to attributes such as hue, saturation, chromaticity, luminosity or brightness, or other, color models are used. The color space may or may not be dependent of the device (printers, scanners, displays, etc.). The color model used for the samples in this work was the CMYK (cyan, magenta, yellow, black); color mixes were obtained by the principles of the subtractive method.

An illuminant D specifies the relative distribution of energy that nearly corresponds with the radiation emitted by the so-called black body. A specific illuminant is then referred to the temperature in degrees Kelvin of the black body that more nearly matches it. The illuminant D65 has a spectral energy distribution that nearly matches that of the black body at 6500 °K.

Finally, the reflectance spectrum $R(\lambda)$ is defined as the ratio of the reflected light flux to the incident light flux for a given wavelength λ [15]. Reflectance values obtained from the spectrophotometer were related to the perceptually quasi-uniform CIE L*a*b* color space.

2.2 Time Series

We begin with a definition of a time series:

Definition 1

A time series is an ordered set of m real-valued variables [7], [8]:

$$T = \{ t_1, \dots, t_m \mid t_i \in R \} \quad (1)$$

In general, in many applications the global properties of time series are not of interest, but the main interest is confined to parts of the time series called subsequences.

Definition 2

Given a time series T of length m , a subsequence C_p is a subset of n contiguous elements from T , where $m \geq n$.

$$C = \{ t_p, \dots, t_{p+n-1} \mid \text{for } 1 \leq p \leq m-n+1 \} \tag{2}$$

In this work our subsequence consist of a group of contiguous ordered elements associated with the number of drops of each primary color cyan, magenta and yellow, as well as black and distilled water, necessary to create a given sample of an unknown color.

Adapting the same approach of sliding windows (8), in our application we define it in the following way.

Definition 3

Given a time series T of length m , and a user defined subsequence of length w , we select a “window” with a known equal number of elements across T to the subsequence w . In our case, all elements selected have the same length, with their beginning always in the position $(nw+1)$, $n = 0, 1, 2, 3, \dots, (m/w)-1$. The number of possible windows (or sequences) is T/w . The above definitions are summarized in Fig. 1.

The shadowed area is the size of a matrix S of $(m-w+1)$ columns, the number of subsequences, by w rows, the number of elements (color components) of each subsequence. The information of the matrix is exactly the same information of T , thus occupying the same memory space. All subsequences are different; that is, there is not the possibility of two equal subsequences in our time series. Subsequences are contiguous to each other; that is, there are no free positions or overlapping between any of them. In a well-built database, subsequences are also contiguous. Independent elements of the database form an ordered set.

In our case, the subsequences are selected according to the minimal Euclidean distance between the pattern of the color sample and the elements of the database that constitutes our time series. Patterns are represented by the values of reflectivity for each wavelength in the range of the visible spectrum.

| | | | Time series T of length m | | | | | | | |
|-----------------------------|------------------|---------------------|---------------------------------|---------|----------|-----|--|------------|-----------|-----------|
| | Color components | Sub-sequence number | $n = 0$ | 1 | 2 | ... | | $(m/w)-2w$ | $(m/w)-w$ | $(m/w)-1$ |
| | | Offset | (1) | $(w+1)$ | $(2w+1)$ | ... | | $n - 2w$ | $n - w$ | $m-w+1$ |
| Sub-sequence length $w = 5$ | Cyan | x | | | | ... | | | | |
| | Magenta | x | | | | ... | | | | |
| | Yellow | x | Elements of the time series T | | | | | | | |
| | Black | x | | | | ... | | | | |
| | Distilled water | x | | | | ... | | | | |

Fig.1. Distribution of the subsequences of our time series database

2.3 Metrics

To find a particular subsequence in the time series different metrics can be used. The Euclidean distance in our application gave good results, although we know that it is sensitive to various “distortions” such as translation and amplitude scaling. However, this is not an issue in our application because our database has elements with fixed-length and the theoretically calculated error when they come from a near perceivable uniform color space like the $L^*a^*b^*$ one, do not cause a practical and a detectable visual color difference for an average human observer. Moreover, it can be considered that the 31 values of the reflectance spectrum from which distances are calculated have approximately a (geometric) plane distribution.

2.3.1 Euclidean Distance

Let there be two time series $X = x_1, \dots, x_n$ and $Y = y_1, \dots, y_m$ with $n = m$. The Euclidean distance between them is defined as [3]:

$$D(X, Y) \equiv \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

In color matching, we calculate the distance between the pattern of the unknown color sample and that of any non-overlapped subsequence of the entire time series. Both are of equal dimensions.

Although color matching must be exact in case of bad matching, it will be very convenient to consider some valid equivalence. Theoretically, any color database could have an infinite number of elements, but in practice the human vision system does not discriminate differences of more than two dozens levels in each color plane. This represents something like more than 8000 different colors. We can consider always that the database is restricted in size without being afraid of committing an unacceptable error. In fact, due to the cumbersome process of creating our color database, in our experiments it had only 1001 different patterns. However, the results demonstrated the validity of the process and of the use of the similarity of equidistant subsequences in time series for color matching.

2.3.2 Equivalent Distance

In this section we define the concept of equivalent distances for color matching. It is assumed that the elements of the database and any query subsequence have the same length. Given a real positive number r , called range, and a time series T with a subsequence C beginning in the position p , and another subsequence M beginning in the position q , then the equivalence is defined in the following way:

$$\text{If } D(C, M) \leq r, \text{ then } M \text{ is the } \textit{equivalence} \text{ of } C \text{ in } T \quad (4)$$

The value of r in our case is found empirically. In general it depends on the size of the database and on the values used to calculate the Euclidean distance (31 reflectivity values in our case). Its utility is only to give the user (or the system) a measure of the quality of a match.

3 Background

The matching of subsequences by similarity in time series databases is not new [9], [10], [11], [12], among others. On the other hand, many results about color matching have been reported, but not using the similarity in time series [13].

In the field of color matching there are a lot of work done, but mainly in the commercial environment [3], [5]. Some desktop publishing programs support the internationally recognized PANTONE® Color Matching System [4]. Nevertheless, theoretical works had been also developed. However, papers describing the use of time series for color matching were not found anywhere.

The main interest found in color matching has been its use to equate profiles of different type of equipments; namely, displays, printers, scanners, photographic cameras, and others. Also in printing industry the color matching in high quality works is fundamental [6]. However, they are not of public knowledge.

The main purpose of this work was oriented to find some solution for color matching far from that of use of swatchbooks or color cards.

4 Preparation of Color Samples

The task of preparing the color samples for color matching is of the most important. If color samples to be measured are not rigorously prepared, the color recognition is prone to errors. The color samples produced by mixing three primary colors cyan, magenta and yellow, together with black and distilled water in previously established proportions will constitute the patterns to be recognized.

Colors parameters can be measured by colorimeters, spectrophotometers, and others equipments. In our case a spectrophotometer was used. Color measurement by means of a spectrophotometer can be implemented by transmittance or by reflectance. Thus, several considerations were imposed on the color samples to be used to carry out our experiments.

Firstly, the type of color source must be selected; that is, the substances having the required primary colors to be combined later to prepare the mixes. It could be liquid, powder, pastes, and others. If transmittance is used to measure the colors, like in the case of liquids, the use of expensive quartz vessels is required, according to the spectrophotometer specifications. That is because the transparence of the vessel must be considerably high in order to guarantee that its contribution to color be practically null. Thus it was decided to measure colors by reflectance.

The use of powder was also analyzed, but it was rejected because the preparation of homogeneous mixes with them could be a very cumbersome task. Furthermore, in the market there are no a great variety of powders with an acceptable purity and color constancy. The use of paints was also rejected because the preparation of the color samples, the measurement of the appropriate proportions (doses) to create the mixes, and the cleaning of the utensils and materials once used, is a very awkward task.

For all these reasons, it was decided to use as the color source to prepare the color samples the Hewlett Packard cyan, magenta, yellow and black printer inks. Distilled water was used in order to simulate the presence of white when the mixes were prepared over a white substrate [1]. For their normal use, they were considered our best choice as primary colors to prepare the color samples used in our experiments.

As a second step, the choice of the substrate where the printer inks will be applied turned out to be of importance. The main characteristic to be fulfilled by the substrate was that its contribution in color when the inks were applied over it be a minimum, and independent on the mix prepared. A second requirement was that the application of a liquid or a paste onto it does not moisten the substrate in such a form to make it useless. Also its white color and its reflectivity must be uniform in the whole range of visible wavelengths and in its entire flat surface.

With all this in mind, three types of substrates were considered: photographic paper, white couche paper, and common white kromekote paper. The reflectance vs. wavelength curves obtained from the spectrophotometer for these three types of papers are shown in Fig. 2.

From these characteristics, it may be seen that the kromekote paper gives a better behavior than the other two, because its characteristic is flatter and more uniform. However, when an opaque mix of color inks covers uniformly the substrate, the reflectance of the paper must not intervene unfavorably on the measurements of the color reflectance of the mix.

The number of drops of ink deposited over the substrate was crucial also. The first condition to be fulfilled was that the reflectance of the paper does not affect the reflectance of the color sample. The second condition was that an excessive number of drops do not moisten the paper in a measure that the color reflectance varies with the characteristics of the paper. Some tests were carried in this regard.

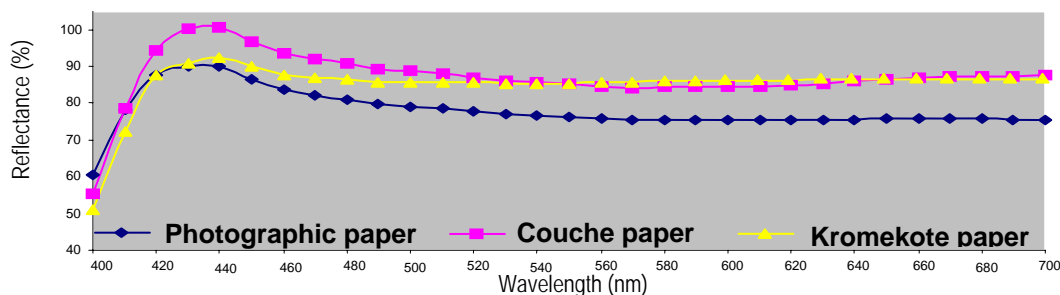


Fig. 2. Reflectance vs. wavelength characteristics for the three types of papers analyzed

Tables 1, 2 and 3 show the luminance level L^* of the three primary colors cyan, magenta and yellow, respectively (using from 1 to 10 ink drops) with respect to the standard cyan, magenta and yellow colors reflectance given by the calibrated spectrophotometer. Figures 3, 4 and 5 show the respective characteristics. It may be noted that except for

yellow, the luminance with six drops of ink on the substrate is similar to the standard one (first row in the tables), which means that the “pure” primary color covers completely the substrate, that is, the influence of the higher luminance of the white color of the substrate is minimum or nearly null. It can be noted also that yellow gives higher luminance values than the other two primary colors, shows a more uniform characteristic, and its luminance value is nearest to that of the substrate (94.29). The fact that only five drops of yellow cover the substrate (lesser luminance value) means that it is the color that better cover the substrate. The variable values of luminance obtained in all cases when more than six drops of ink were used, is an indication that the substrate has been already moistened, thus changing the properties of the color. In these figures, the letter G means drops and letter S means standard white of the equipment.

Based on the preceding results, we conclude that using only six drops of mixed inks achieved the covering of the substrate in all color samples.

Table 1. Luminance values for the cyan ink

| Cyan ink | | |
|--------------|-----------------|------------|
| No. of drops | Luminance Level | Difference |
| Standard | 59.05 | -35.90 |
| 1 | 69.08 | 10.02 |
| 2 | 66.83 | 7.78 |
| 3 | 67.80 | 8.75 |
| 4 | 66.86 | 7.81 |
| 5 | 60.59 | 1.53 |
| 6 | 58.39 | 0.66 |
| 7 | 62.34 | 3.29 |
| 8 | 60.17 | 1.12 |
| 9 | 60.69 | 1.64 |
| 10 | 61.37 | 2.32 |

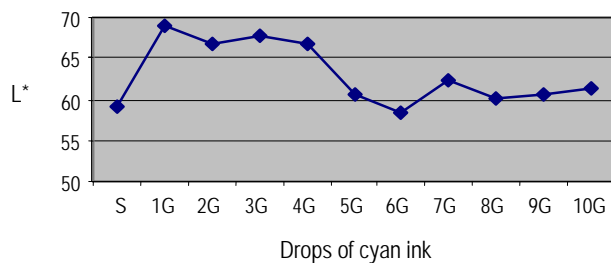


Fig. 3. Luminance values L^* for the cyan ink

Table 2. Luminance values for the magenta ink

| Magenta ink | | |
|--------------|-----------------|------------|
| No. of drops | Luminance Level | Difference |
| Standard | 49.59 | -44.28 |
| 1 | 60.94 | 11.35 |
| 2 | 57.01 | 7.42 |
| 3 | 55.07 | 5.48 |
| 4 | 53.81 | 4.22 |
| 5 | 52.96 | 3.36 |
| 6 | 50.01 | -0.42 |
| 7 | 53.43 | 3.84 |
| 8 | 50.49 | 0.90 |
| 9 | 52.98 | 3.38 |
| 10 | 48.35 | -1.25 |

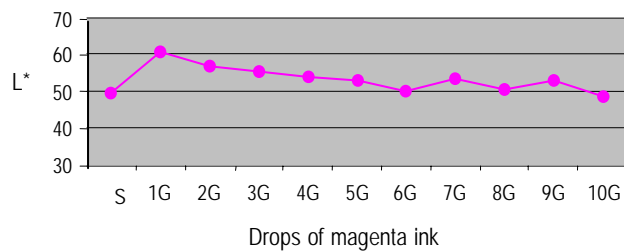
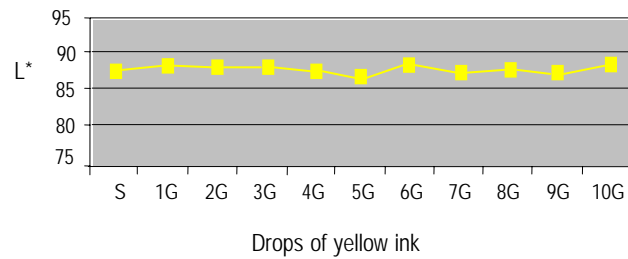


Fig. 4. Luminance values L^* for the magenta ink

Table 3. Luminance values for the yellow ink

| Yellow ink | | |
|--------------|-----------------|------------|
| No. of drops | Luminance Level | Difference |
| Standard | 87.39 | -6.65 |
| 1 | 87.69 | 0.30 |
| 2 | 87.44 | 0.05 |
| 3 | 87.42 | 0.03 |
| 4 | 87.37 | -0.01 |
| 5 | 87.12 | -0.27 |
| 6 | 87.65 | -0.26 |
| 7 | 85.18 | -2.21 |
| 8 | 85.53 | -1.86 |
| 9 | 84.92 | -2.47 |
| 10 | 87.78 | 0.39 |

**Fig. 5.** Luminance values L^* for the yellow ink

To achieve different shadows and tints of pure primary colors and their mixes, it was necessary to mix pure primary color inks with black ink and distilled water in the corresponding proportions, respectively. The white color of the substrate contributed to give tint when distilled water was used in the mixes. Afterward, other particular mixes were prepared using the corresponding shadowed and tinted primaries.

The total number of drops used in the preparation of every color mix was limited always to one hundred drops. Each color mix prepared was kept in a sterilized glass dropper, sealed with a cap, and protected from light in order to guarantee that its quality did not deteriorate.

The color samples were prepared carefully on a substrate of 5 x 7.5 cm of white kromekote paper, with a useful area of 5 x 5 cm, which guaranteed a good reading by the spectrophotometer. To apply the color mix over the substrate a printer roller was used in order to guarantee good uniformity of the mixed inks. On each substrate, for each sample and after shaking the container, were poured only six drops of the mix prepared with the proportions previously established. The roller was applied six times over the paper substrate to get uniformity in the color. Finally, the roller was cleaned with water before the preparation of the next sample.

Therefore, we produced samples from five components, that is, three primary colors cyan, magenta and yellow, together with black and distilled water to add shadow or tint, respectively. The mixes prepared were in ten-percent increments of each component, that is, the number of drops of each component varied in steps of ten drops.

To calculate exactly the number of possible color patterns that must be prepared according to the conditions stated, the generating function models were used [14].

Suppose that a_r is the number of ways to select r objects in a certain procedure. Then $g(x)$ is a generating function for a_r if $g(x)$ has the polynomial expansion:

$$g(x) = a_0 + a_1x + a_2x^2 + \cdots + a_r x^r + \cdots + a_n x^n$$

For example as:

$$(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{r}x^r + \cdots + \binom{n}{n}x^n$$

Then $g(x) = (1+x)^n$ is the generating function for $a_r = C(n, r)$, the number of ways to select an r -subset from an n -set.

The problem of determining the coefficient of x^r when we multiply several such polynomial factors together can be restated in terms of exponents.

As an example, consider the expansion of $(1 + x + x^2)^4$. The number of formal products $x^{e_1} x^{e_2} x^{e_3} x^{e_4}$, $0 \leq e_i \leq 2$ equaling x^r in the expansion, will be the number of integer solutions to:

$$e_1 + e_2 + e_3 + e_4 = r, 0 \leq e_i \leq 2$$

In our case, the total number of drops of the five possible components to create each color mix was limited to 100 drops for convenience. Additionally, the concentration of samples created for each one out of five components was established in steps of ten drops, that is, $r = 100/10 = 10$ steps. This means that:

$$e_1 + e_2 + \dots + e_5 = 10, \text{ where } 0 \leq e_i \leq 10$$

This is equivalent to solving the polynomial:

$$P(x) = (x^0 + x^1 + x^2 + x^3 + \dots + x^8 + x^9 + x^{10})^5$$

That is, $P(x)$ will be:

$$\begin{aligned} &1 + 5x + 15x^2 + 35x^3 + 70x^4 + 126x^5 + 210x^6 + 330x^7 + 495x^8 + 715x^9 + 1001x^{10} \\ &+ 1360x^{11} + 1795x^{12} + 2305x^{13} + 2885x^{14} + 3526x^{15} + 4215x^{16} + 4935x^{17} \\ &+ 5665x^{18} + 6380x^{19} + 7051x^{20} + 7645x^{21} + 8135x^{22} + 8500x^{23} + 8725x^{24} \\ &+ 8801x^{25} + 8135x^{28} + 7645x^{29} + 8725x^{26} + 8500x^{27} + 7051x^{30} + 6380x^{31} \\ &+ 5665x^{32} + 715x^{41} + 495x^{42} + 330x^{43} + 4935x^{33} + 4215x^{34} + 3526x^{35} + 2885x^{36} \\ &+ 2305x^{37} + 1795x^{38} + 1360x^{39} + 1001x^{40} + x^{50} + 210x^{44} + 126x^{45} + 70x^{46} + 35x^{47} \\ &+ 15x^{48} + 5x^{49} \end{aligned}$$

In the expansion of this polynomial $P(x)$ in its formal products the coefficient a_{10} of x^{10} gives the total combination number under our conditions. That coefficient is 1001. Then, the number of combinations (to be) prepared was 1001 combinations.

5 Measurements

Color-measurement devices, such as spectrophotometers, measure the light reflected from an object and gives the color a position into a set of coordinates of a three-dimensional color space. In taking measurements, the device was calibrated to measure the samples under a known specific set of conditions. These conditions involve the illuminant (light source) type, the observer's viewing angle, and the spectral inclusion (taking gloss into consideration), or the spectral exclusion (not taking gloss into consideration). We did not consider the presence of gloss.

The spectrophotometer used was a Milton Roy COLOR MATE™ HDS Color Analyzer. Before its use and after warmed, it was always meticulously calibrated with a white tile accompanying the equipment. The parameters established in the equipment to measure the color samples were:

- Observation Angle: 10°
- Sight Window (Area of View): small
- Color Space: CIE L*a*b*
- Illuminants: D65 and D50
- Average number of lectures: 5

Once the parameters were established in the spectrophotometer and after its calibration, a labeled color sample was placed manually in the sight window. Five readings were taken in different zones of the covered substrate. The equipment then averaged the results of the five readings and finally gave us, among others, thirty one values of the reflectance of the sample for each one of the thirty one wavelengths ranging from 400 to 700 nanometers of the visible spectrum, in steps of ten, of the standard and the color sample, as well as the difference between both values. The standard color could be selected by the user, between the standard white, standard black, one of the primary color selected from cyan, magenta and yellow, or another one of our convenience.

From the two illuminants selected, that is D50 and D65, the spectrophotometer gave us the values of the components L^* , a^* , b^* of the color model CIELAB selected and the values of the three primary three-stimulus values X, Y and Z, for both the standard and the sample, together with the respective difference. In order to carry out the color matching, the value of L^* related to the illuminant D65 was used as reference for daylight. Nevertheless, the color matching method presented in this paper was achieved only on the base of the Euclidean distance calculated from the difference of the 31 reflectance values given by the spectrophotometer (with L^* constant) in the range of wavelengths between 400 and 700 nm, for the parameters established and detailed above.

Then, after reading the 1001 color samples, prepared with colors cyan, magenta, yellow, black and distilled water in steps of ten drops, and taking into account that the total number of drops be 100 drops for each mix prepared and used 6 drops to cover the substrate, a database was created with the thirty one values of the reflectance of each sample.

6 Error Estimation

The difference in color between two similar samples of the database, not necessarily adjacent, depends on the number of drops of the five primary components. The effect of the number of drops used to prepare the samples is always related to the 31-reflectance values obtained for each wavelength in the range 400-700 nm.

To calculate the maximum permissible error E of our color matcher, consider that each one of the 31 components of the Euclidean distance has the same error ε . Then, the maximum error E reflected in the Euclidean distance calculated is: $E = \sqrt{31\varepsilon^2} \cong 5.568\varepsilon$. The value of ε is unknown but can be estimated. Among other factors, it depends on the maximal error of the spectrophotometer, on the quality of the samples prepared, the number of measurements averaged, and on the care with which the experiments were carried out.

According to the operator's manual of the spectrophotometer, the repeatability is 0.2% maximum peak-to-peak deviation in percent reflectance ($\pm 0.1\%$) at any wavelength for 20 readings of stable white reference material within one hour. On the other hand, the short term color space repeatability is 0.02 maximum standard deviation of CIE $L^*a^*b^*$ ΔE from the mean of a set of measurements made over one hour without instrument re-calibration. The reflectivity values obtained from the spectrophotometer were approximated by the equipment until the hundredth, between 0 (ideal black) and 100.00 (ideal white) for each wavelength. The worst case gives an $\varepsilon = 2.0$ (0.02×100.00) and the maximum absolute error will be $E = 11.136$. This error is influenced in by the five components that, combined, produce the color mix.

7 Results

Two types of experiments were carried out to test the color matching module by the similarity in time series: (1) using known samples from the database, keeping it into the database; (2) using known samples from the database, but excluding it from the database.

Six out of 1001 independent samples were selected at random from the database for testing and excluded from it each time. Then, the database had only 1000 elements (color combinations) in each case. The characteristic reflectivity versus wavelength was measured by using the same spectrophotometer. The 31 values obtained were introduced to the matching module created by using the similarity of time series and then we got the number of drops to be compared. Table 4 shows the values of the number of drops of the six samples, the number of drops given by the matching module, and the smallest Euclidean distance between them.

The correct matches in the number of drops of each component for each mix are shown in Table 4 in white cells. We show three gray levels for failures, noting that the darker the cell the more distant is the result obtained from that expected. Note that in all cases the Euclidean distance is below the maximum absolute error E calculated in section 6.

It is important to observe that the number of matches in drops and how far the components are from those of the sample has no relation with to Euclidean distance calculated. However, the influence on the final color depends on the particular characteristics and properties of the color components.

Table 4. Number of drops of six matching examples

| Number of drops of the color sample to match | | | | | Number of drops of the matching with the color sample | | | | | Euclidean Distance |
|--|---------|--------|-------|-------|---|---------|--------|-------|-------|--------------------|
| Cyan | Magenta | Yellow | Black | Water | Cyan | Magenta | Yellow | Black | Water | |
| 0 | 10 | 0 | 20 | 70 | 0 | 10 | 10 | 40 | 40 | 1.92 |
| 40 | 0 | 10 | 10 | 40 | 40 | 0 | 0 | 10 | 50 | 2.99 |
| 0 | 40 | 30 | 20 | 10 | 0 | 30 | 30 | 30 | 10 | 6.13 |
| 20 | 40 | 0 | 20 | 20 | 20 | 60 | 0 | 10 | 10 | 5.23 |
| 30 | 10 | 20 | 20 | 20 | 30 | 10 | 30 | 10 | 20 | 2.44 |
| 10 | 0 | 0 | 70 | 20 | 10 | 0 | 0 | 60 | 30 | 5.81 |

The characteristics reflectivity vs. wavelength of six samples and its corresponding graphical matching are shown in Figures 6, 7, 8, 9, 10 and 11. Note the small difference between both characteristics, in spite of the few elements in the database, which means that the principle of similarity in time series can be used reliably in this application. Undoubtedly, with a larger database the coincidence between the samples and its color matching would be higher and the error would be lower.

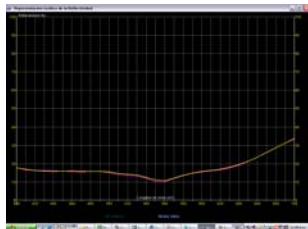


Fig. 6. Sample 1

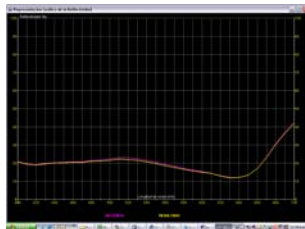


Fig. 7. Sample 2

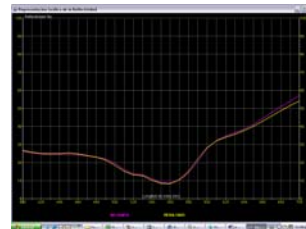


Fig. 8. Sample 3

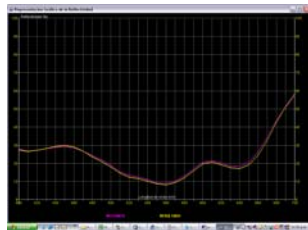


Fig. 9. Sample 4

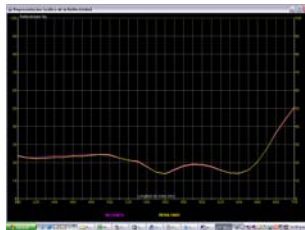


Fig. 10. Sample 5

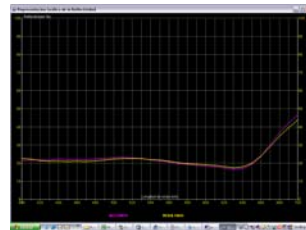


Fig. 11. Sample 6

8 Conclusions

This work has been oriented toward the use of the principle of similarity in time series for color matching. The similarity was calculated based upon the Euclidean distance. The task of creating color samples with an acceptable quality resulted cumbersome. For that reason, the universe of samples is a relatively small database with 1001 samples, but prepared carefully. Two types of experiments were done to test the color matching module by the similarity in time series: (1) using known samples from the database, keeping it into the database; (2) using known samples from the database, but excluding it from the database. The quality of color matching was based on the measures of reflectivity versus wavelength of samples given by a spectrophotometer. The results demonstrate the validity of the use of time series in color matching.

Acknowledgments

The authors of this paper wish to thank the Center for Computing Research (CIC); Research and Postgraduate Secretary (SIP), and National Polytechnic Institute (IPN), Mexico, for their support.

References

1. **Kaverman, J. (2002)**. Color matching for pad printing, Screen Printing magazine. <http://www.screenweb.com/index.php/channel/7/id/1604/>
2. **Aquino, D., Cervantes, F., Linares, B., Felipe, E., De la Cruz, C. (2004)**. Telecolorimetría basada en el reconocimiento de patrones de color, Thesis, Interdisciplinary Professional School in Engineering and Advanced Technologies, National Polytechnic Institute, Mexico.
3. **Butts, K. (2003)**. Database Preparation for Successful Computer Color Matching. Techexchange.com.
4. **White, J.W. (1997)**. <http://www.trumatch.com/articles/truwhite.htm>
5. **Coudray, M.A. (2004)**. The Realities of Color Matching, the dynamics of color matching and color perceptions. Dynamesh.
6. **Kaverman, J. (2004)**. Diversified Printing Technique, Screen Printing magazine. <http://www.screenweb.com/pad/cont/colormatch.html>
7. **Keogh, E., Lin, J., Truppel, W. (2003)**. Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research, ICDM.
8. **Lin, J., Keogh, E., Lonardi, S. and Chui, B. (2003)**. A Novel Symbolic Representation of Time Series, with Applications to Classification, Clustering, Query by Content and Anomaly Detection, ACM SIGKDD'03, August 24-27, Washington, D.C..
9. **Orenstein, J. (1986)**. Spatial query processing in an object-oriented database system. Proc. ACM SIGMOD, pages 326-336.
10. **Faloutsos, C., Ranganathan, M. and Manolopoulos, Y. (1994)**. Fast Subsequence Matching in Time-Series Databases, SIGMOD 94-5/94, ACM.
11. **Agrawal, R., Faloutsos, C. & Swami, A. (1993)**. Efficient similarity search in sequence databases. In proceedings of the 4th Int'l Conference on Foundations of Data Organizations and Algorithms. Chicago, IL, Oct. 13-15. 69-84.
12. **Faloutsos, C., Equitz, M., Flickner, M., Niblack, W., Petkovic, D., and Barber, R. (1993)**. Efficient and effective querying by image content. Journal of Intelligent Information System.
13. **Zhang, H., Dong Bi, J., Back, B. (2002)**. The Design and Implementation of Color Matching System Based on Back Propagation, Lecture Notes in Computer Science, Vol. 2358, Springer, Proc. of the 15th Intl. Conf. on Industrial and Engineering, Applications of Artificial Intelligence and Expert Systems: Developments in Applied Artificial Intelligence, pp. 557-566.
14. **Tucker A. (2002)**. *Applied Combinatorics*, John Wiley & Sons, 4th Edition, Chapter 6.
15. **Sharma G. (2003)**. *Digital Color Imaging Handbook*, CRC Press, 797 p.



Edgardo Manuel Felipe Riverón He received the B.Sc. degree in Electrical Engineering from the Higher Polytechnic Institute Jose Antonio Echeverria, in Havana, Cuba, in 1967. He received the Ph.D. degree in Technical Sciences from the Computer and Automation Research Institute, in Budapest, Hungary, in 1977. He is currently Full Professor and Senior Researcher at the Center for Computing Research of the National Polytechnic Institute of Mexico. His research interests include Image Processing and Analysis, Computer Vision and Pattern Recognition, in particular color quantization, retina analysis, biometric solutions, document analysis and mathematical morphology applications.



Jesús Guillermo Figueroa Nazuno was born in Mexico, 1945. He received the B.Sc. degree in Physics from the National Autonomous University of Mexico in 1970, the Ph.D. degree in Artificial Intelligence in 1979 and the Sc.D. in Theoretical Physics in 1980, both from Edinburgh University, U.K. Currently, he is Professor at the Center for Computing Research of the National Polytechnic Institute of Mexico. His research interests include neural networks, mathematical modeling, machine learning and prediction models.



Agustín Francisco Gutiérrez Tornés He received the B.Sc. degree in Economics from the Institute of Economics, University of Havana, Cuba, in 1970, and his Ph.D. degree from the Enterprise Improvement Research Center, Agricultural Academy (SGGW-AR), Warsaw, Poland, in 1984. Actually he works as a Systems Coordinator at Banamex, S.A. (CITIGROUP) and as an Adjoin Professor at the Instituto Tecnológico y de Estudios Superiores de Monterrey. His research interests are on different topics of Software Engineering.



Ricardo Barrón Fernández was born in México City. He received the Mathematical degree from the National Autonomous University of Mexico (UNAM) in 1985, the M. Sc. degree in Computer Engineering in 1998 and his Ph.D. degree in Computer Science in 2006, both from the National Polytechnic Institute, in Mexico. His research interest includes Artificial Intelligence, Computational Vision and Images.