

Method to determine the degrees of consistency in experimental datasets of perceptual color differences

SAMUEL MORILLAS,¹ LUIS GÓMEZ-ROBLEDO,² RAFAEL HUERTAS,² AND MANUEL MELGOSA^{2,*}

¹Instituto Universitario de Matemática Pura y Aplicada, Universidad Politécnica de Valencia, 46022 Valencia, Spain

²Departamento de Óptica, Universidad de Granada, 18071 Granada, Spain

*Corresponding author: mmelgosa@ugr.es

Received 15 August 2016; revised 10 October 2016; accepted 12 October 2016; posted 13 October 2016 (Doc. ID 273789); published 4 November 2016

We propose a fuzzy method to analyze datasets of perceptual color differences with two main objectives: to detect inconsistencies between couples of color pairs and to assign a degree of consistency to each color pair in a dataset. This method can be thought as the outcome of a previous one developed for a similar purpose [J. Mod. Opt. 56, 1447 (2009)], whose performance is compared with the proposed one. In this work, we present the results achieved using the dataset employed to develop the current CIE/ISO color-difference formula, CIEDE2000, but the method could be applied to any dataset. Specifically, in the mentioned dataset, we find that some couples of color pairs have contradictory information, which can interfere in the successful development of future color-difference formulas as well as in checking the performance of current ones. © 2016 Optical Society of America

OCIS codes: (330.1690) Color; (330.1730) Colorimetry; (330.1720) Color vision.

<http://dx.doi.org/10.1364/JOSAA.33.002289>

1. INTRODUCTION/PURPOSE

Development of accurate experimental datasets of perceptual color differences is of paramount importance since they play a crucial role in improving the correlation between visually perceived (ΔV) and instrumentally measured (ΔE) color differences [1–3]. To date, several experimental datasets with limited accuracy have been proposed for developing and testing the merits of different color-difference formulas used in industrial applications [4–9]. Undoubtedly, a key point for successful future advances in this field would be to achieve a reliable and broad set of color pairs, distributed throughout all regions of color space, visually assessed by a high number of observers with normal color vision using an appropriate methodology. This has long been sought by the International Commission on Illumination (CIE) and different researchers [10–16].

Usual datasets of perceptual color differences are sets of numerous color pairs, where for each color pair (at least) the instrumentally measured color coordinates of the two samples and the average visual difference from assessments performed by a panel of observers are reported. Because of the subjective nature of these visual experiments and the influence of different experimental conditions, differences may exist between the results reported for similar color pairs in different datasets, and even within one dataset, as the consequence of inherent observer variability or even errors [3,6,17,18]. In any case,

despite these potential differences, it is important to guarantee that each dataset has sufficient internal consistency, and also that different datasets agree relatively well with each other, at least in the cases where the viewing conditions and methodologies followed by different experiments were similar. Thus, it is advisable to have the appropriate methods to analyze the consistency of color-difference datasets. We can use such methods to test currently available and future datasets to be employed in color-difference evaluations.

In this work, we introduce a general method to analyze the consistency of a color-difference dataset, and we use it for the so-called COM dataset [3], which was employed to develop the current CIE/ISO-recommended color-difference formula, CIEDE2000 [19]. The word COM comes from “combined,” because this dataset was formed by the combination of color pairs from four different subsets [6]. After the development of CIEDE2000, a mistake in the use of one of these four subsets was detected [3] and repaired, leading to the so-called COM-corrected dataset, which is the one considered in the current work. The COM-corrected dataset has 3813 color pairs from four different subsets (provided by four different renowned laboratories), which were rescaled before combination. Specifically, these four subsets are designated as BFD-P [20], Leeds [21], RIT-DuPont [2], and Witt [22] and have 2776, 307, 312, and 418 color pairs, respectively. Despite the precautions taken,

some inconsistent couples of color pairs can unfortunately be detected in the current COM-corrected dataset. For example, in the case from the BFD-P subset shown in Table 1, a couple of color pairs have quite similar CIELAB color coordinates and CIEDE2000 color differences (columns 2–8) but with visual differences (column 9) differing by more than a factor of 6. This kind of inconsistent situation is undesirable for reliable datasets to be used by the color-difference community, and it seems necessary to have a method to detect and solve them appropriately (e.g., by removing at least one of the two inconsistent color pairs), as intended in the current paper.

We propose a new method to analyze datasets of perceptual color differences, which is composed of two different steps. The first step analyzes the whole dataset by taking couples of color pairs and determining which of them are inconsistent. The second step complements the previous one and assigns a degree of consistency to each color pair by comparison with similar color pairs in the dataset. In a previous paper [23], we proposed a one-step method to detect inconsistent color pairs in datasets of perceptual color differences and analyzed the COM-corrected dataset using such a method, assigning a degree of consistency to each of the color pairs. However, there is room for some relevant improvements in our previous method [23], considering specifically the following three points:

(1) In [23], to determine the consistency of a color pair, we compared it with other color pairs positioned in the same region of the color space, and the degree of nearness between two color pairs was measured using the CIELAB distance between the central points of the color pairs. However, now we find that this criterion of nearness is not very appropriate because nearness of central points does not necessarily imply similarity of two color pairs. Now, we have redefined the concept of nearness since, instead of depending on the distance between the central points of the two color pairs, results improve when it depends on the distance between the color samples in the pairs.

(2) Given a set of near color pairs, to determine consistencies in [23], we used the ratio between the visual color difference ΔV and the CIEDE2000 color difference (i.e., $\Delta V/\Delta E_{00}$) for all near color pairs. However, it is well known that human perception behaves nonlinearly for different magnitudes of color differences [4,6,9,10,14]. Therefore, a color pair will now be compared only with near color pairs having similar visual color differences ΔV (which implies that there will be also similarity with respect to ΔE_{00}).

(3) In [23], in addition to determining degrees of consistency, we used a statistical analysis based on the average and standard deviation of the quotient $\Delta V/\Delta E_{00}$ in each neighborhood. These two measurements (in particular the standard deviation) are particularly sensitive to the presence of noise and outliers, which may limit the accuracy of the final results.

In fact, the inconsistent color pairs shown in Table 1 were not clearly detected by the method in [23], and these color pairs were assigned similar middle degrees of consistency of 0.41 and 0.66, respectively (column 10 in Table 1), but in fact they are highly inconsistent, as detected by the procedure described in the current paper (column 13 in Table 1). This was most probably because in the neighborhood of these pairs the standard deviation of the quotient $\Delta V/\Delta E_{00}$ was very large, and then the accuracy dropped. Thus, we now replace the previous quotient $\Delta V/\Delta E_{00}$ by the absolute value of the difference between ΔV and ΔE_{00} , $D = |\Delta V - \Delta E_{00}|$, which has several advantages concerning stability and invariance to translation, assuming that ΔV and ΔE_{00} are in a common scale, as has been done in the current paper (see Section 4).

The standardized residual sum of squares (STRESS) index [24] was used in [23] to test the performance of color-difference formulas for the COM-corrected dataset before and after the removal of any inconsistent color pairs detected. The use of a weighted STRESS [25,14] also proved useful to assess the degree of consistency assigned to each color pair using the new method proposed in the current paper. However, bearing in mind that the STRESS index is not very sensitive to the removal of a small number of color pairs from a dataset, in addition to STRESS we have also used the mean square error (MSE) to analyze our current results.

This paper is structured as follows: Section 2 introduces the new method for detecting inconsistencies between couples of color pairs, which is based on fuzzy rules (given that the ideas behind the proposed new method are explained by linguistic terms, we will use fuzzy logic to formulate them); Section 3 describes the proposed method for determining the degree of consistency of each color pair; Section 4 provides the results found for the COM-corrected dataset [3]; finally, conclusions are drawn in Section 5.

2. STEP 1: DETECTION OF INCONSISTENT COUPLES OF COLOR PAIRS

In the experimental color-difference dataset to be analyzed, each color pair is denoted as $S_i = \{A_i, B_i, \Delta V_i, \Delta E_{00,i}\}$, where A_i and B_i denote the CIELAB color coordinates of the two color samples in the pair, given by $A_i = \{L_i^{*,A}, a_i^{*,A}, b_i^{*,A}\}$ and $B_i = \{L_i^{*,B}, a_i^{*,B}, b_i^{*,B}\}$, respectively; ΔV_i represents the average visual color difference reported by observers; and $\Delta E_{00,i}$ is the computed color difference from the CIELAB color coordinates of the two samples in the pair using the CIEDE2000 color-difference formula [6,19]. Among many color-difference formulas currently available [1,26–30], here we use just the CIEDE2000 formula, because it is the current CIE/ISO recommendation to achieve the best agreement with

Table 1. Example of Two Inconsistent Color Pairs in the BFD-P Subset^a

Color Pair	$L^{*,A}$	$a^{*,A}$	$b^{*,A}$	$L^{*,B}$	$a^{*,B}$	$b^{*,B}$	ΔE_{00}	ΔV	FM	$\Delta M_{ij,1}$	$\Delta M_{ij,2}$	I_{ij}
<i>I</i>	40.730	−44.325	36.707	40.083	−43.851	36.144	0.611	2.257	0.41	0.00	0.50	0.86
<i>J</i>	40.730	−44.325	36.707	40.307	−43.700	36.568	0.429	0.359	0.66			

^aNote the very similar values of color coordinates of the samples and ΔE_{00} color differences, while the two visual differences ΔV are very different (more than a factor of 6). FM is the degree of consistency of each of these pairs from the method described in [23]; $\Delta M_{ij,1}$ and $\Delta M_{ij,2}$ are distances between couples of pairs; and I_{ij} is the degree of inconsistency (see text).

visually perceived color differences [14,19]. Nonetheless, the method proposed here can be used with any other color-difference formula.

Let us consider two color pairs $\mathcal{S}_i$ and $\mathcal{S}_j$, which are close in the color space and have similar ΔE_{00} values. In such a case, it is expected that these two color pairs will also have similar ΔV values. When this fails, it is considered an inconsistency. Since all this reasoning is formulated by means of vague linguistic terms, we can use fuzzy logic for its numerical representation [31]. Following the above notation, let us denote by $\Delta\Delta V_{ij} = |\Delta V_i - \Delta V_j|$ and $\Delta\Delta E_{ij} = |\Delta E_{00,i} - \Delta E_{00,j}|$ the absolute value of the differences between ΔV and ΔE_{00} for the pairs $\mathcal{S}_i$ and $\mathcal{S}_j$, respectively. To represent the nearness in the color space of the two color pairs $\mathcal{S}_i$ and $\mathcal{S}_j$, we defined the pair of distances $(\Delta M_{ij,1}, \Delta M_{ij,2})$:

$$\begin{aligned} (\Delta M_{ij,1}, \Delta M_{ij,2}) &= (\|A_i - A_j\|, \|B_i - B_j\|) \\ \text{if } \|A_i - A_j\| + \|B_i - B_j\| &\leq \|A_i - B_j\| + \|B_i - A_j\|, \\ \text{otherwise } (\Delta M_{ij,1}, \Delta M_{ij,2}) &= (\|A_i - B_j\|, \|B_i - A_j\|), \end{aligned} \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean distance in CIELAB space between the two points separated by the symbol $-$. A graphic example of the meaning of Eq. (1) is given in Fig. 1, where we can see meaningful nonnull distances in the specific case where the centers C_i and C_j of two color pairs are at the same point.

Then the following fuzzy rule is used to detect inconsistent couples of color pairs:

Fuzzy rule 1.1:

IF $\Delta M_{ij,1}$ is small, AND $\Delta M_{ij,2}$ is small, AND $\Delta\Delta E_{ij}$ is very small, AND $\Delta\Delta V_{ij}$ is not small,

THEN $\mathcal{S}_i$ and $\mathcal{S}_j$ are inconsistent.

This fuzzy rule provides a number in the interval [0,1] representing the degree of inconsistency of the two color pairs. This degree, associated with the degree of certainty of the consequent, is identified with the degree of certainty of the antecedent. In turn, as is common practice in fuzzy-logic systems [31], the certainty of the antecedent is computed using a continuous t-norm to perform the conjunction of the certainties of the vague terms involved. In particular, we have used the classical product t-norm, which means that the certainty of the antecedent is given by the product of the certainties of the four vague terms in fuzzy rule 1.1. In our case, using an S-type fuzzy membership function, the degree to which $\Delta M_{ij,1}$ is *small* is computed as

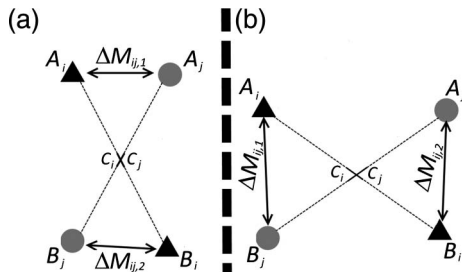


Fig. 1. Example of the two options to compute $\Delta M_{ij,1}$ and $\Delta M_{ij,2}$ from Eq. (1), in two specific cases where the distances between the centers C_i and C_j of the two color pairs (with samples indicated by triangles and circles, respectively) were null.

$$\Delta M_{ij,1}^{\text{small}} = 1 - \mu(\Delta M_{ij,1}, \alpha_1, \gamma_1), \quad (2)$$

where

$$\mu(x, \alpha, \gamma) = \begin{cases} 0 & \text{if } x \leq \alpha \\ 1 - 2\left(\frac{x-\gamma}{\gamma-\alpha}\right)^2 & \text{if } \alpha < x \leq \frac{\alpha+\gamma}{2} \\ 2\left(\frac{x-\alpha}{\gamma-\alpha}\right)^2 & \text{if } \frac{\alpha+\gamma}{2} < x \leq \gamma \\ 1 & \text{if } x > \gamma \end{cases} \quad (3)$$

Analogously, $\Delta M_{ij,2}^{\text{small}}$ is also defined as in Eqs. (2) and (3).

Similarly, the degree in which $\Delta\Delta E_{ij}$ is *very small* is defined as

$$\Delta\Delta E_{ij}^{\text{very small}} = 1 - \mu(\Delta\Delta E_{ij}, \alpha_2, \gamma_2), \quad (4)$$

where the function μ has been defined in Eq. (3).

Finally, the degree to which $\Delta\Delta V_{ij}$ is *not small* is defined as

$$\Delta\Delta V_{ij}^{\text{not small}} = \mu(\Delta\Delta V_{ij}, \alpha_3, \gamma_3), \quad (5)$$

where once again the function μ is the one defined in Eq. (3).

The conjunction of the certainties of the four statements in the antecedent of fuzzy rule 1.1 (i.e., $\Delta M_{ij,1}$ is small, AND $\Delta M_{ij,2}$ is small, AND $\Delta\Delta E_{ij}$ is very small, AND $\Delta\Delta V_{ij}$ is not small) is computed by the product of these four variables. Therefore, if we identify the certainty of the consequent with that of the antecedent, the degree of inconsistency between the pairs $\mathcal{S}_i$ and $\mathcal{S}_j$ is given by the parameter

$$I_{ij} = \Delta M_{ij,1}^{\text{small}} \cdot \Delta M_{ij,2}^{\text{small}} \cdot \Delta\Delta E_{ij}^{\text{very small}} \cdot \Delta\Delta V_{ij}^{\text{not small}}, \quad (6)$$

in such a way that higher I_{ij} values (its maximum value being 1) indicate greater inconsistency of the couple of color pairs $\mathcal{S}_i$ and $\mathcal{S}_j$.

Moreover, if in the foregoing fuzzy rule 1.1, we exchange the roles of ΔE and ΔV , we can set the next fuzzy rule, which is also able to identify inconsistencies:

Fuzzy rule 1.2:

IF $\Delta M_{ij,1}$ is small, AND $\Delta M_{ij,2}$ is small, AND $\Delta\Delta V_{ij}$ is very small, AND $\Delta\Delta E_{ij}$ is not small,

THEN $\mathcal{S}_i$ and $\mathcal{S}_j$ are inconsistent.

Following an analogous procedure to the one leading to I_{ij} in Eq. (6), from fuzzy rule 1.2 we can find another fuzzy degree of inconsistency between the color pairs $\mathcal{S}_i$ and $\mathcal{S}_j$, which we denote by I_{ij}^* and which is defined by

$$I_{ij}^* = \Delta M_{ij,1}^{\text{small}} \cdot \Delta M_{ij,2}^{\text{small}} \cdot \Delta\Delta V_{ij}^{\text{very small}} \cdot \Delta\Delta E_{ij}^{\text{not small}}, \quad (7)$$

where $\Delta\Delta V_{ij}^{\text{very small}}$ and $\Delta\Delta E_{ij}^{\text{not small}}$ are defined analogously as in previous Eqs. (4) and (5), respectively.

Finally, the couple of color pairs $\mathcal{S}_i$ and $\mathcal{S}_j$ are considered inconsistent if either I_{ij} [Eq. (6)] or I_{ij}^* [Eq. (7)] exceed a certain fixed value or threshold. Note that from previous fuzzy rules [Eqs. (6) and (7)] we can detect couples of inconsistent color pairs but no information concerning which of the two color pairs may be considered wrong. The adopted threshold values as well as the color pairs to be removed will be studied in Section 4, using the COM-corrected dataset.

3. STEP 2: CONSISTENCY OF EACH COLOR PAIR

In this section, we describe the second step of our proposed method for consistency analysis, improving the one previously proposed by us in [23]. The aim of the current procedure is to determine the degree to which each color pair in a dataset can be considered *consistent*, by comparison with *similar* color pairs in the same dataset. Since both *consistent* and *similar* are linguistic terms, they can be modeled by means of fuzzy sets [32]. The degree of membership to the fuzzy set *consistent* represents the consistency of each color pair. Specifically, as mentioned before, the improvement of the current method with respect to the one in [23] is based on the following three points: First, we redefine the concept of nearness considering the positions of the samples of the two color pairs in the CIELAB color space instead of the distance between their corresponding centers (denoted as C_i and C_j in Fig. 1); second, we consider that a color pair S_i should be compared only with those color pairs having similar ΔV_i values (which implies them also to have similar $\Delta E_{00,i}$ values; otherwise the pair would have been detected and removed in step 1 of our method); third, we no longer need to use the quotient $\Delta V/\Delta E_{00}$, which was justified in [23] because experimental data with very different ΔV_i values could be involved in the computations of the degree of consistency using our previous method. Now, assuming that ΔV_i and $\Delta E_{00,i}$ values are in a common scale, we can replace the above quotient by the difference $D_i = |\Delta V_i - \Delta E_{00,i}|$, noting that D_i has several advantages concerning stability and invariance to translation. Below, we will detail the procedure of incorporating the aforementioned improvements.

First, we build the fuzzy set of color pairs *similar* to the pair S_i , denoted as S^{S_i} . We consider two color pairs *similar* when they are *not far* apart in the color space and their ΔV values are *similar*. To represent nearness in color space we will use the pair of distances $(\Delta M_{ij,1}, \Delta M_{ij,2})$ defined in Eq. (1), and for the similarity between ΔV values, we will use again $\Delta \Delta V_{ij} = |\Delta V_i - \Delta V_j|$. Then, the reasoning behind the concept of similarity that we employ follows the next fuzzy rule:

Fuzzy rule 2.1:

IF $\Delta M_{ij,1}$ is not large, AND $\Delta M_{ij,2}$ is not large, AND $\Delta \Delta V_{ij}$ is not large,

THEN S_j is similar to S_i .

Fuzzy inference [31] can be used to determine a degree of certainty in the interval [0,1] for the expression in the consequent: " S_j is similar to S_i ." This degree will be identified with the membership of S_j to the fuzzy set of neighbors similar to S_i , denoted by $S^{S_i}(S_j)$.

Once again, the certainty of the consequent will be identified with the certainty of the antecedent. In the antecedent of fuzzy rule 2.1, we need to model three vague terms, for which we will use the fuzzy sets described in Eqs. (8) and (9).

The degree of certainty of " $\Delta M_{ij,1}$ is not large," denoted by $\Delta M_{ij,1}^{\text{not large}}$, is computed using an S-type function as

$$\Delta M_{ij,1}^{\text{not large}} = 1 - \mu(\Delta M_{ij,1}, \alpha_4, \gamma_4), \quad (8)$$

where the function μ was defined in Eq. (3).

Analogously, the certainty of " $\Delta M_{ij,2}$ is not large," denoted as $\Delta M_{ij,2}^{\text{not large}}$, is computed in the same way indicated in Eq. (8).

For the certainty of " $\Delta \Delta V_{ij}$ is not large," denoted as $\Delta \Delta V_{ij}^{\text{not large}}$, we perform in a similar way, using the following expression:

$$\Delta \Delta V_{ij}^{\text{not large}} = 1 - \mu(\Delta \Delta V_{ij}, \alpha_5, \gamma_5). \quad (9)$$

Finally, to model the conjunction operation AND in fuzzy rule 2.1, we make use of the product t-norm. Therefore, we compute $S^{S_i}(S_j)$ as

$$S^{S_i}(S_j) = \Delta M_{ij,1}^{\text{not large}} \cdot \Delta M_{ij,2}^{\text{not large}} \cdot \Delta \Delta V_{ij}^{\text{not large}}. \quad (10)$$

Given the fuzzy set of color pairs *similar* to S_i , denoted as S^{S_i} , the degree of consistency of the color pair S_i , denoted by $C(S_i)$, is computed by comparing the value of the difference $D_i = |\Delta V_i - \Delta E_{00,i}|$ with an average of the D_j differences observed for the *similar* color pairs in S^{S_i} . To do so, we used the fuzzy metric in Eq. (11) [32], because it was successfully employed in previous works [33,34] and was able to compare those values, taking into account the weighted average and standard deviation of the D_j values [Eqs. (12) and (13), respectively]:

$$C(S_i) = \text{FM}_S(D_i, \tilde{D}_i, \tilde{\sigma}_i) = \frac{\tilde{\sigma}_i}{\tilde{\sigma}_i + |D_i - \tilde{D}_i|}, \quad (11)$$

where $\tilde{D}_i$ and $\tilde{\sigma}_i$ are the *fuzzy mean* and *fuzzy standard deviation* of the D_j values of the color pairs in S^{S_i} , computed as

$$\tilde{D}_i = \frac{\sum_{S_j \in S, S_j \neq S_i} S^{S_i}(S_j) D_j}{\sum_{S_j \in S, S_j \neq S_i} S^{S_i}(S_j)} \quad (12)$$

and

$$\tilde{\sigma}_i = \sqrt{\frac{\sum_{S_j \in S, S_j \neq S_i} S^{S_i}(S_j) (D_j - \tilde{D}_i)^2}{\sum_{S_j \in S, S_j \neq S_i} S^{S_i}(S_j)}}, \quad (13)$$

respectively. The degree of consistency $C(S_i)$ [Eq. (11)] represents the degree to which D_i is near to $\tilde{D}_i$, taking into account the *fuzzy standard deviation* $\tilde{\sigma}_i$. A high value (close to 1) of $C(S_i)$ indicates that the agreement is good and the value of D_i is close to its neighbors. On the other hand, if $C(S_i)$ is low (close to 0), the agreement is small and the color pair S_i is noisy or has low consistency. Again, for practical purposes, we must fix a value or threshold to consider a pair consistent or inconsistent, as will be studied for a specific dataset in the next section.

4. RESULTS OF THE FUZZY CONSISTENCY ANALYSIS FOR THE COM-CORRECTED DATASET

We have used the method proposed in the two previous sections to analyze the consistency of the 3813 color pairs in the COM-corrected dataset [3].

In step 1, for practical purposes, the values $\alpha_1 = 1$ and $\gamma_1 = 5$ have been reasonably well adopted [4,6,9] in Eq. (2), meaning that, in terms of nearness in color space, distances lower than or equal to 1 CIELAB unit are *small* (degree 1), and those higher than 5 CIELAB units are not

at all considered *small* (degree 0). Concerning the criteria for color differences, we adopted the values $\alpha_2 = 0.2$ and $\gamma_2 = 1$ in Eq. (4), signifying that when the CIEDE2000 color difference between two color pairs is below 0.2 units it is considered *very small* (degree 1), and when it exceeds 1 unit it is not at all considered *very small* (degree 0). In the case of Eq. (5) involving the perceived difference ΔV , it is important to note that we have rescaled the values of ΔV to CIEDE2000 color-difference units [24]. In this case, we adopted $\alpha_3 = 1$ and $\gamma_3 = 2$, which means that when the ΔV difference between two color pairs is below 1 unit it is considered *small* (degree 0), and when it exceeds 2 units it is considered *not small* (degree 1). With respect to fuzzy rule 1.2, where the roles of $\Delta E_{00,i}$ and ΔV_i are exchanged, we can use the same parameter setting as that in fuzzy rule 1.1, given that both magnitudes are in a common scale.

In step 2 of our current proposed method, we have adopted the values $\alpha_4 = 1$ and $\gamma_4 = 10$ in Eq. (8) because they indicate distances between neighboring color pairs. On the one hand, this choice is based on the fact that in most industrial applications and color atlases, distances above 10 CIELAB units are usually considered *large* [4,5], whereas shorter distances may be reasonably considered to be *not large* to some degree. On the other hand, distances smaller than 1 CIELAB unit are definitely *not large* with maximum certainty [2,4,8,9]. Regarding the criteria adopted for differences in ΔV , we considered in Eq. (9) $\alpha_5 = 0.5$ and $\gamma_5 = 2$, meaning that when $\Delta\Delta V_{ij}$ is larger or equal than 2 CIEDE2000 units, the color pairs S_i and S_j are not at all considered *similar*, but when this difference is below 0.5 units, they are considered completely *similar*. It is important to note that in step 2 we are modeling the concept of being *not large*, which is different from the concept of being *small*, modeled in the previous paragraph when we discussed the values of parameters for step 1. In step 1, we look for inconsistencies between almost identical couples of color pairs, but in step 2 we are interested in comparing a particular color pair with other similar color pairs in the same region of color

space. Therefore, in step 2 we need to be less restrictive than in step 1 in terms of both nearness of color samples and differences in ΔV values, since we do not look for almost the same values but just similar ones.

Figure 2 is a plot of the μ function in Eq. (3), considering the different couples of values of parameters α and γ mentioned in the two previous paragraphs. Note that values of parameters α and γ must be selected carefully because the wrong use of the concepts *small* and *not large* can lead to contradictory results, as we confirm after different trials.

A. Results of Step 1

As an example, the degree of inconsistency for the couple of color pairs shown in Table 1 is as high as $I_{ij} = 0.86$ from fuzzy rule 1.1, with $\Delta M_{ij,1}^{\text{small}} = \Delta M_{ij,2}^{\text{small}} = \Delta\Delta E_{ij}^{\text{small}} = 1$ and $\Delta\Delta V_{ij}^{\text{not small}} = 0.86$. This means that the two color pairs in Table 1 are very inconsistent, a fact which was not detected by our previous method in [23].

Of course, different results are found depending on the threshold fixed to consider a couple of color pairs as inconsistent. Figure 3 shows the number of inconsistent color pairs from the fuzzy rules 1.1 and 1.2 for different values of threshold (i.e., I_{ij} and I_{ij}^* values). As expected, we see that the lower the threshold, the higher the number of couple of color pairs with higher inconsistency than the threshold (it should be remembered that these degrees of inconsistency are in the range [0,1], where higher values indicate greater inconsistency). Finally, a threshold of inconsistency equal to 0.5 has been assumed henceforth, bearing in mind that a couple of color pairs with a degree of consistency higher than 0.5 can be considered more inconsistent than consistent. It was found that, assuming this 0.5 threshold, there were 117 inconsistent couples of color pairs using the fuzzy rule 1.1, and 113 inconsistent couples of color pairs using the fuzzy rule 1.2. Therefore, we detected a total of 230 inconsistencies between couples of color pairs. Bearing in mind that for this dataset the total number of comparisons

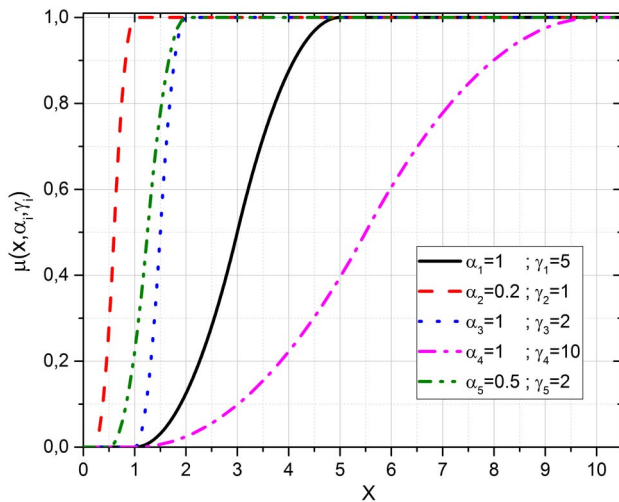


Fig. 2. Plot of μ function in Eq. (3), for the different values of parameters α_i and γ_i considered in the text. Below the value α_i the membership of the neighborhoods is equal to 0; then the membership increases smoothly until value γ_i ; and finally, above value γ_i , the membership is equal to 1.

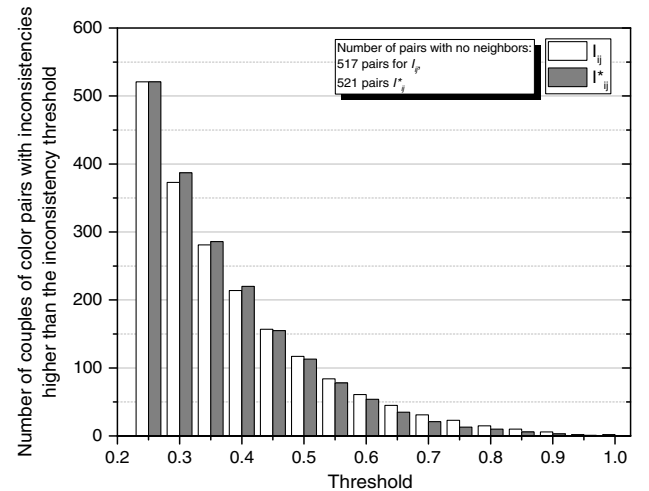


Fig. 3. Number of couples of color pairs with inconsistencies higher than the inconsistency thresholds indicated in the x-axis, using the I_{ij} (inconsistencies when $\Delta\Delta E_{00}$ is very small and $\Delta\Delta V$ is not small) and I_{ij}^* (inconsistencies when $\Delta\Delta E_{00}$ is not small and $\Delta\Delta V$ is very small) indices.

between neighboring color pairs at least partially fulfilling the closeness conditions given by the μ function with $\alpha_1 = 1$ and $\gamma_1 = 5$ and $\alpha_2 = 0.2$ and $\gamma_2 = 1$, using fuzzy rules 1.1 and 1.2, was 118,700, the number of 230 inconsistencies represents only 0.19%. Thus, the COM-corrected dataset can be considered overall to be considerably consistent under step 1 of our method. However, as stated above, step 1 informs us about inconsistencies only between couples of color pairs, but no information is provided about which is the less inconsistent color pair in a given couple of inconsistent color pairs nor about the degree of consistency of a given color pair. Finally, it is also worth pointing that there were 517 and 521 pairs for which we found no neighbors to compare with for fuzzy rule 1.1 and fuzzy rule 1.2, respectively. Therefore we have been unable to compare these color pairs with any others.

Roughly we may assume that, in a couple of inconsistent color pairs, the color pair with higher $|\Delta E - \Delta V|$ is the most inconsistent, and therefore the one we must select to be removed. This procedure gives us 67 inconsistent pairs, from the 230 couples of inconsistent pairs, because using this criterion some pairs were inconsistent in more than one couple. To analyze the influence of these pairs, we can add that, after removing these 67 pairs, the STRESS value for the CIEDE2000 color-difference formula reduced from 29.20 to 28.05, which represents more than a 1-point decrease. While this small improvement in STRESS values is not statistically significant at a confidence level of 95% ($F = 0.923$, $F_C = 0.938$, $N = 3746$) [23,3], it can be considered as an indicator of both the validity of our proposed method to remove inconsistent couples of color pairs and the high consistency or reliability of the COM-corrected dataset. Using a more outlier- and error-sensitive index, the mean square error, we found that, for the CIEDE2000 formula, by removing these 67 color pairs, the MSE decreased from 0.36 to 0.33.

B. Results of Step 2

We applied the procedure proposed in Section 3 after removing the 67 inconsistent pairs mentioned above. Using this procedure, each pair is compared, through its $D_i = |\Delta V_i - \Delta E_{00,i}|$ value, with its fuzzy neighborhoods. As we can see in Fig. 4, the color

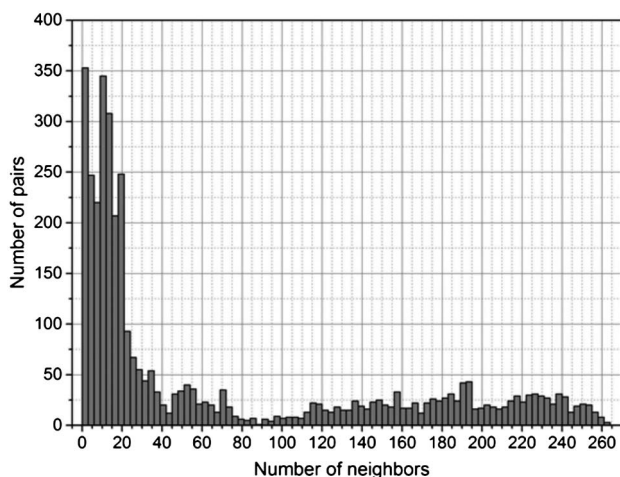


Fig. 4. Number of color pairs (y-axis) for different numbers of fuzzy neighbors (x-axis, where the bar width is equal to 3). It should be remembered that $\alpha_4 = 1$ and $\gamma_4 = 10$ fixed the adopted neighborhood criterion.

pairs in the COM-corrected dataset are not uniformly distributed in color space. Note that the histogram in Fig. 4 is not accumulative but just shows the number of color pairs with different numbers of fuzzy neighbors indicated on the abscissa axis. There are many pairs with a low number of neighbors. If we consider the addition of the neighborhood degrees of all pairs S_j that have nonnull values in $S^{S_i}(S_j)$, we find that, for example, 668 pairs (17.5%) have less than 6 fuzzy neighbors and 2294 pairs (60.2%) have less than 20 fuzzy neighbors. It should be noted that step 2 works by comparison with similar color pairs, and therefore it is meaningless to consider step 2 for color pairs with only a very few neighbors. In this context, from trial and error, we think that, using this dataset, the results of step 2 can be considered reliable if either S^{S_i} has a value of more than six units or there are at least three neighbors with membership $S^{S_i}(S_j)$ higher than 0.35 (i.e., there are three pairs with very few, but quite similar, neighbors). Thus, we found that 176 color pairs did not fulfill these requirements, and they were discarded in such a way that only the remaining 3570 color pairs were considered for the next analyses (it should be remembered that 67 color pairs were already removed in step 1).

As in step 1, different results were found depending on the fixed threshold to consider a color pair as consistent from its $C(S_i)$ value. The higher the value adopted as the threshold, the greater the number of color pairs with a consistency below the fixed threshold. Figure 5 shows, for the pairs considered from the COM-corrected dataset, the number of color pairs with a degree of consistency below different threshold values. STRESS values found for the CIEDE2000 color-difference formula after removing color pairs with a degree of consistency equal to or lower than the values on the abscissa axis are also shown by the continuous line in Fig. 5. After removing the 67 pairs detected in step 1 plus the 176 pairs with no neighbors, there were really few color pairs with a low degree of consistency. For example, only a few color pairs can be noted in Fig. 5 below a threshold value of 0.2. In any case, removing

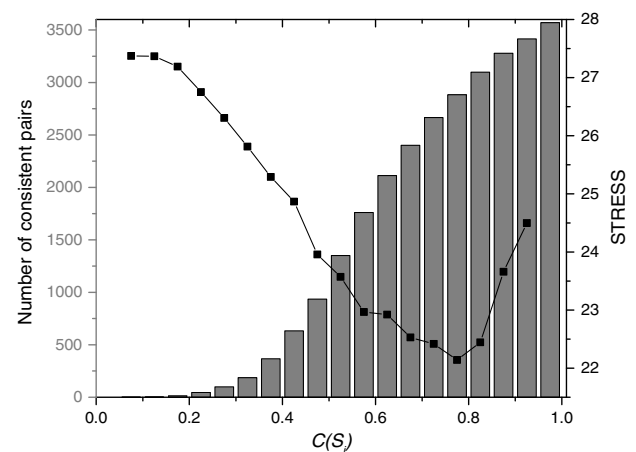


Fig. 5. Histogram with number of color pairs with a degree of consistency lower than $C(S_i)$ values shown in the x-axis. The continuous line in the plot (associated to the right y-axis units) corresponds to STRESS values calculated for the CIEDE2000 color-difference formula after removing color pairs with degrees of consistency equal to or lower than specific $C(S_i)$ values shown on the x-axis.

the pairs with a degree of consistency lower than 0.50 (935 pairs), STRESS reduces up to 23.96. If we are more restrictive, STRESS continues to decrease until we achieve a degree of consistency approaching 0.8. At this point, if we continue to remove color pairs STRESS rises, because in fact these last color pairs removed are quite consistent and there is no strong reason to remove them. As desired, removal of highly inconsistent color pairs improves the performance of the CIEDE2000 color-difference formula measured using the STRESS index.

In summary, the proposed method provides information for removing inconsistent color pairs from a dataset. Specifically, step 2 gives a degree of consistency for each color pair, which is highly useful information for any dataset. It should be noted that, with a few exceptions [2,25], the reliability of individual color pairs is not provided in available color-difference datasets, where only average ΔE and ΔV values are provided. The reliability of ΔV may be the standard deviation of the answers reported by the observers or something similar. When this information is not available, at least our method is able to provide information on the reliability of an individual color pair, based on its agreement or consistency with neighboring color pairs (assuming that there are a reasonable minimum number of such neighbor color pairs). This information on the specific reliability of each color pair can also be considered to compute a weighted STRESS [8,14,24,25] for any color-difference formula, which may be a good option for a better assessment of the merit of such a formula. For example, using CIEDE2000 as the color-difference formula, and the weights provided by the consistency index $C(S_i)$ described in step 2, we find that the weighted STRESS value (WSTRESS) for the COM-corrected dataset was 26.14 while the STRESS value was 27.53. In this case the difference between STRESS and WSTRESS was small because all 3570 selected color pairs are quite reliable. When removal of many or all inconsistent color pairs is not advisable, the use of WSTRESS may be preferable to the use of STRESS.

Finally, Fig. 6 shows that results from step 2 of current method are quite different from those of our previous method in [23]. Pearson's linear correlation coefficient for data shown in

Fig. 6 is quite low ($r = 0.039$), and it bears noting that the degrees of consistency provided by our current method are in general slightly higher than those provided by our previous method.

5. CONCLUSIONS

In this paper, we have introduced a two-step method to analyze datasets of perceptual color differences, improving the accuracy achieved by the method in [23]. The first step analyzes the dataset to determine inconsistencies between couples of pairs. Thus, it provides couples of pairs in the dataset that are not consistent with each other. The second step assigns each color pair a degree of consistency by comparing each color pair with similar color pairs in the dataset. The best results are found using steps 1 and 2 sequentially. This procedure can be used to remove inconsistent data in a perceived color-difference database, enabling STRESS values provided by a good color-difference formula to be reduced simply by controlling a threshold consistency-degree parameter.

For the COM-corrected dataset, we conclude that there are 696 (18%) color pairs which cannot be properly analyzed because they lack adequate similar color pairs in the same region of the color space: Specifically, 520 color pairs had no other color pair to compare with in step 1, and 176 color pairs did not have enough neighbors to make the results of step 2 meaningful. The results for the remaining color pairs of the COM-corrected dataset can be considered quite satisfactory since only 67 inconsistent color pairs were found in step 1, and there were only 37 color pairs that reached a degree of consistency lower than 0.25 in step 2.

Funding. Ministerio de Economía y Competitividad (MINECO) del Gobierno de España (FIS2013-40661-P, FIS2013-45952-P, MTM2015-64373-P); European Regional Development Fund (ERDF).

Acknowledgment. Funding has been provided by the Ministerio de Economía y Competitividad (MINECO) del Gobierno de España, research projects FIS2013-40661-P, FIS2013-45952-P, and MTM2015-64373-P, with support from the European Regional Development Fund (ERDF).

REFERENCES

1. M. Melgosa, A. Trémeau, and G. Cui, "Colour difference evaluation," in *Advanced Color Imaging Processing and Analysis*, C. Fernandez-Maloigne, ed. (Springer, 2013), pp. 59–79.
2. R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen, "Visual determination of suprathreshold color-difference tolerances using probit analysis," *Color Res. Appl.* **16**, 297–316 (1991).
3. M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008).
4. M. Melgosa, E. Hita, J. Romero, and L. Jiménez del Barco, "Some classical color differences calculated with new formulas," *J. Opt. Soc. Am. A* **9**, 1247–1254 (1992).
5. M. Melgosa, J. J. Quesada, and E. Hita, "Uniformity of some recent color metrics tested with an accurate color-difference tolerance dataset," *Appl. Opt.* **33**, 8069–8077 (1994).

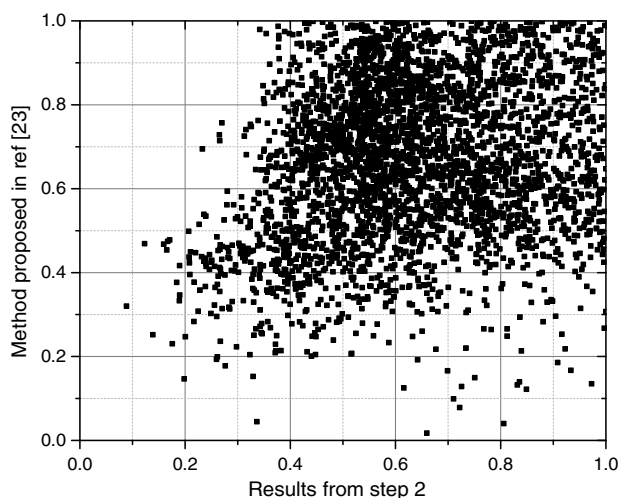


Fig. 6. Degrees of consistency for color pairs in the COM-corrected dataset from the method in [23] against those taken from step 2 of our current method.

6. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.* **26**, 340–350 (2001).
7. I. Farup, "Hyperbolic geometry for colour metrics," *Opt. Express* **22**, 12369–12378 (2014).
8. M. Melgosa, J. Martínez-García, L. Gómez-Robledo, E. Perales, F. M. Martínez-Verdú, and T. Dauser, "Measuring color differences in automotive samples with lightness flop: a test of the AUDI2000 color-difference formula," *Opt. Express* **22**, 3458–3467 (2014).
9. M. Huang, G. Cui, M. Melgosa, M. Sánchez-Marañón, C. Li, M. R. Luo, and H. Liu, "Power functions improving the performance of color-difference formulas," *Opt. Express* **23**, 597–610 (2015).
10. International Commission on Illumination (CIE), *Parametric Effects in Colour Difference Evaluation* (CIE Central Bureau, 1993).
11. A. R. Robertson, "CIE guidelines for coordinated research on color-difference evaluation," *Color Res. Appl.* **3**, 149–151 (1978).
12. K. Witt, "CIE guidelines for coordinated future work on industrial colour-difference evaluation," *Color Res. Appl.* **20**, 399–403 (1995).
13. M. Melgosa, "Request for existing experimental datasets on color differences," *Color Res. Appl.* **32**, 159 (2007).
14. CIE 217:2016, *Recommended Method for Evaluating the Performance of Colour-Difference Formulae* (CIE Central Bureau, 2016).
15. E. D. Montag and D. C. Wilber, "A comparison of constant stimuli and gray-scale methods of color difference scaling," *Col. Res. Appl.* **28**, 36–44 (2003).
16. E. Kirchner, N. Dekker, M. Lucassen, L. Njo, I. van der Lans, P. Urban, and R. Huertas, "How psychophysical methods influence optimizations of color difference formulas," *J. Opt. Soc. Am. A* **32**, 357–366 (2015).
17. M. Melgosa, P. A. García, L. Gómez-Robledo, R. Shamey, D. Hinks, G. Cui, and M. R. Luo, "Notes on the application of the standardized residual sum of squares index for the assessment of intra- and inter-observer variability," *J. Opt. Soc. Am. A* **28**, 949–953 (2011).
18. R. Shamey, J. Lin, W. Sawatwarakul, and R. Cao, "Evaluation of performance of various color-difference formulae using an experimental black dataset," *Color Res. Appl.* **39**, 589–598 (2014).
19. International Commission on Illumination (CIE) and International Organization for Standardization (ISO), "Colorimetry—Part 6: CIEDE2000 Colour-Difference Formula," ISO/CIE 11664-6:2014 (CIE Central Bureau, 2014).
20. M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colors," *Color Res. Appl.* **11**, 25–42 (1986).
21. D. H. Kim and J. Nobbs, "New weighting functions for the weighted CIELAB color difference formula," in *Proceedings of AIC Colour 97* (Color Science Association of Japan, 1997), Vol. **1**, pp. 446–449.
22. K. Witt, "Geometrical relations between scales of small colour differences," *Color Res. Appl.* **24**, 78–92 (1999).
23. S. Morillas, L. Gómez-Robledo, R. Huertas, and M. Melgosa, "Fuzzy analysis for detection of inconsistent data in experimental datasets employed at the development of the CIEDE2000 colour-difference formula," *J. Mod. Opt.* **56**, 1447–1456 (2009).
24. P. A. García, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," *J. Opt. Soc. Am. A* **24**, 1823–1829 (2007).
25. R. S. Berns and B. Hou, "RIT-DuPont supra-threshold color-tolerances individual color-difference pair dataset," *Color Res. Appl.* **35**, 274–283 (2010).
26. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," *Color Res. Appl.* **31**, 320–330 (2006).
27. G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," *Color Res. Appl.* **27**, 282–290 (2002).
28. C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space," *J. Opt. Soc. Am. A* **26**, 121–134 (2009).
29. D. H. Kim, "The ULAB colour space," *Color Res. Appl.* **40**, 17–29 (2015).
30. M. W. Derhak and R. S. Berns, "Introducing WLab—Going from Wpt (Waypoint) to a uniform material color equivalence space," *Color Res. Appl.* **40**, 550–563 (2015).
31. E. E. Kerre, *Fuzzy Sets and Approximate Reasoning* (Xian Jiaotong University, 1998).
32. A. George and P. Veeramani, "On some results in fuzzy metrics spaces," *Fuzzy Sets Syst.* **64**, 395–399 (1994).
33. S. Morillas, V. Gregori, G. Peris-Fajarnes, and A. Sapena, "New adaptive vector filter using fuzzy metrics," *J. Electron. Imaging* **16**, 33007 (2007).
34. S. Morillas, "Fuzzy metrics and fuzzy logic for colour image filtering," Ph.D. thesis (Universidad Politécnica de Valencia, 2007).