Color Perception: A Comparison between Computer and Human Color Labelling

Lorenzo de Mari, **Alessandra Ventura, Lledó Museros** and **Ismael Sanz**[1]

Abstract. In the last few decades, the use of computer and electronic devices has increased considerably and a lot of attention has been brought to human-computer and human-robot interaction. In the latter case, the key for satisfying the user seems to be the introduction of robots able to interpret the world in the same way as humans. The perception of color fits in this perspective, as its elaboration and classification can be very important for communication. For this reason, this study tries to evaluate the differences between human and computer vision with a focus on the color naming process. A computerized determination of the name of a color from an image is developed through a simple algorithm and compared with the results obtained by interviewing a group of people from different countries. What emerged is that color perception and classification is highly subjective and it is advisable to customize those applications where the use of colors is fundamental.

1 INTRODUCTION

Computers, robots and electronic devices are nowadays part of everyday life for the majority of people throughout the world. For this reason, the attention over human-computer interaction (HCI) and human-robot interaction (HRI) is growing at a fast pace. When people have to communicate with robots, the communication is facilitated if they feel that they can interact in a spontaneous and natural way, possibly receiving a responsive feedback. This can be achieved via the implementation of programs that interpret the world in the same way as we humans do. An example is the use of robots for therapy with autistic children [1]. For them, in fact, it is fundamental that the interaction makes them feel as comfortable as possible.

One of the things that people have the ability to perceive is color, therefore the way color is represented when interacting with each other is fundamental in communication [2]. As a matter of fact, the concept of color labelling is extremely important, although it can be difficult to be reproduced in robots and computers. In fact, in order to recreate such human ability on a software, it is necessary to adopt a rigid and mathematical approach, that could lead to poor results. Therefore, it is important to evaluate the similarities and discrepancies in human and computer vision in this context.

In this approach, there are several models in the literature for color naming [2-6], and the model in [3] is selected for the comparison between computer and human color naming of a visualized HSL coordinate. This study proposes a set of labels for color naming, which are based on the responses obtained by asking a group of people to name a sample of colors. The labels are associated with an interval in the HSL representation, explained in detail in the following sections. This model has the advantage of being simple and well structured, so it is particularly suitable for an algorithmic implementation. It does not, however, include a label for the color *brown*, which is just a dark orange and not a distinct hue. Therefore, in this approach the label *orange* of the model is replaced by the label *orange/brown,* since not including such a common label as *brown* might be confusing for some users. Nevertheless, the method adopted in this approach for computer labelling is the same.

In conducting the experiment, a sample of images representing different colors has been selected and an algorithm, developed in MATLAB, has been used for associating the correct label to each image. The same set of colored images has been shown to a group of people, who were asked to label them according to the model. The experiment, moreover, has been specifically designed in order to obtain a significant comparison. In fact, all the samples were collected by using the same screen in the same environment. Also, the test can be considered guided since everyone is restricted to use the labels that the chosen model provides.

What emerged is that color perception, hence color labelling, is highly subjective for people. Nonetheless, it was always possible to identify the label that was most popular among the participants and compare it with the one obtained from the algorithm. The cases in which the two labels correspond perfectly are few, but in many cases at least one person used the same label as the computer. From the results obtained some indications on how to modify the model for HRI and HCI applications can be extracted. In addition, it seems clear that a dynamic labelling based on the user's preferences is advisable when dealing with color naming.

A small group of people took part in the experiment, due to the constraint of using the same screen, and this can be considered to be the main limitation of this study. In future works a larger dataset should be acquired and used to understand what are the factors that involve the color labelling process for people of different ages, nationalities and backgrounds. Based on the results of this study, moreover, it is possible to adjust the intervals that define the labels, focusing on the problems that emerged.

2 COMPUTER PERCEPTION: COLOR LABELLING

The color naming model used in this paper for human color naming comparison is the one presented in [3,7] which is a Qualitative Color Description (QCD) model. This QCD model defines a reference

[1] Computer Science and Engineering department, Universitat Jaume I, Spain, email: isanz@uji.es

system in the HSL colour space for qualitative colour description, which is built according to Figure 1 and 2, and defined as:

$$QC_{RS} = \{uH, uS, uL, QCNAME1..5, QCINT1..5\}$$

where uH is the unit of Hue; uS is the unit of Saturation; uL is the unit of Lightness; QCNAME1..5 refers to the colour names; and QCINT1..5 refers to the intervals of HSL coordinates associated with each colour. The chosen QCNAME and QCINT are:

QCNAME1 = {*black, dark grey, grey, light grey, white*}
QCINT1 = {[0, 20), [20, 30), [30, 50), [50, 75), [75, 100) ∈ uL | ∀ uH ∧ uS ∈ [0, 20] }
QCNAME2 = {*red, orange, yellow, green, turquoise, blue, purple, pink*}
QCINT2 = {(335, 360] ∧ [0, 20], (20,50], (50, 80], (80, 160], (160, 200], (200, 260], (260, 300], (300, 335] ∈ uH | uS ∈ (50, 100] ∧ uL ∈ (40, 55]}
QCNAME3 = {pale-red, pale-orange, pale-yellow, ..., pale-blue, pale-purple, pale-pink}
QCINT3 = {∀ QCINT2 | uS ∈ (20, 50] ∧ uL ∈ (40, 55]}
QCNAME4 = {light-red, light-orange, light yellow, ..., light blue, light purple, light pink}
QCINT4 = {∀ QCINT2 | uS ∈ (50, 100] ∧ uL ∈ (55, 100]}
QCNAME5 = {dark red, dark orange, dark yellow, ..., dark blue, dark purple, dark pink}
QCINT5 = {∀ QCINT2 | uS ∈ (50, 100] ∧ uL ∈ (20, 40]}

As a baseline, the Qualitative Color Reference System (QCRS) was calibrated according to the vision system used.

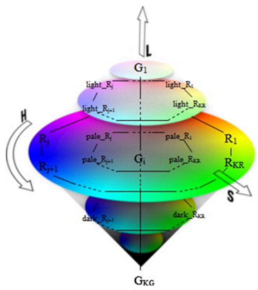

Figure 1. Describing the QCD: discretization of HSL

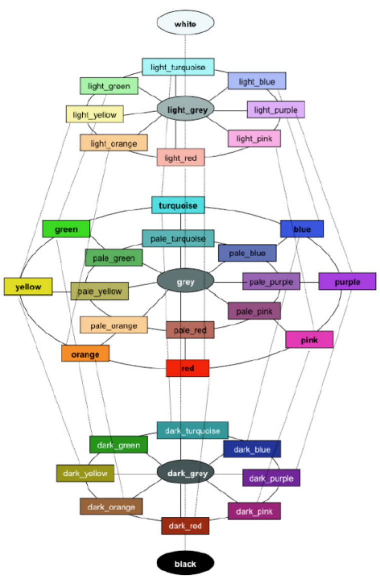

Figure 2. Labels and CND of the QCD

The QCD model has a relational structure and it can be organized in a conceptual neighborhood diagram (CND) [8] according to how a colour can be transformed into another by changing its luminosity, saturation or hue. For example, the colours red and orange are conceptual neighbors since a continuous change in hue causes a direct transition from red to orange. However, blue and red are not conceptual neighbors, since a continuous transformation of hue from blue to red finds other colours in between. A CND for the computational QCD is built and shown in Figure 2.

In this approach the *orange* label has been replaced by the *orange/brown*. We consider that *brown* is a label vey commonly used by humans, and therefore it is important to include it in the study. The *brown* HSL coordinates seem to be localized in the range corresponding to the hue value of *orange*, but not in a continuous way with respect to saturation and lightness, and thus a simple discrimination is not possible. Therefore, the choice adopted here is to merge the two labels. This is, of course, a flaw in the model, as in human perception the two colours are quite different.

A simple algorithm based on this QCD model has been implemented with MATLAB, in order to perform color labeling (Figure 3). The script receives as input an image of a color in the RGB domain and, first of all, performs a transformation in order to obtain the HSL representation of the same image. Then, the mean value of hue, saturation and lightness is evaluated and finally the mean color of the image is named according to the model.

All the possible color labels are stored in a file with the range of HSL values that correspond to each label. Therefore, the MATLAB script reads the file and performs a comparison between the HSL values of the given image and the values assigned to each label. The given model avoids any overlap in HSL intervals for the labelling, hence the algorithm cannot give more than one label to the same HSL value.

Algorithm:

$labels \leftarrow labeling.xls$
for each image **do**:
  $img_{RGB} \leftarrow image$
  $img_{HSL} \leftarrow rgb2hsl(img_{RGB})$
  $mean_H \leftarrow mean(img_{HSL}(1))$
  $mean_S \leftarrow mean(img_{HSL}(2))$
  $mean_L \leftarrow mean(img_{HSL}(3))$
  for each *line* in *labels* **do**:
    $min_H \leftarrow line.Hue_{min}$
    $max_H \leftarrow line.Hue_{max}$
    $min_S \leftarrow line.Saturation_{min}$
    $max_S \leftarrow line.Saturation_{max}$
    $min_L \leftarrow line.Lightness_{min}$
    $max_L \leftarrow line.Lightness_{max}$
    if $(min_i < mean_i < max_i)$for i=H,S,L **then**:
      $color\_label \leftarrow line.Label$
  return *color_label*

Figure 3: The algorithm returns the label of the image given as input.

3 HUMAN PERCEPTION: QUESTIONNAIRE

For evaluating human perception of color, a Google Form questionnaire has been designed. An example of it can be seen in

Figure 4. The questionnaire was developed in English, as this does not compromise the validity of the results even though the test was performed with people of different nationalities [6].

In this study we wanted to take into account that the screen and the relative setting of the computer and its environment can affect color perception. For this reason, all the people that were interrogated were asked to do the questionnaire in the same conditions and on the same computer, in order to limit the discrepancies due to external causes.

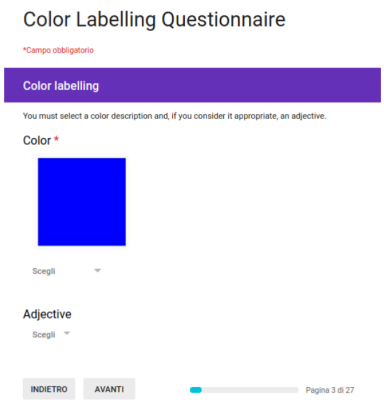

Figure 4: Example of question: the color is shown and the person can choose among the available colors (black, red, orange/brown, …) and, if they believe it is necessary, an adjective (dark, pale or light).

A set of 24 colors shown in Figure 5 has been used for the test. In addition, one color (corresponding to the red in the RGB definition) was shown twice, to detect whether people gave the same label to it or not.

Figure 5: Set of colors selected for the experiment.

Because of that, the questionnaire is composed by 25 questions, that are considered to be sufficient for the study and not overwhelming for the participants. If the questionnaire were too long, it is likely that people would lose attention before the end.

The order in which the colors were shown was the same for all participants. What we tried to avoid is to have similar colors appearing near one another, as this could influence the answers.

Also, color perception and categorization can be subject to cultural influences and other factors, so before performing the color labeling process all subjects were asked to provide some personal information. All information was acquired anonymously, for statistical purposes only.

The required information was: age, gender, nationality, country where the person lives and for how long they have been living there, level of English, experience with color (work, hobby, none).

The expectation is that people that work with colors or have a better level of English will use the labels in a broader way. Male and females, moreover, are known to perceive color differently [9], so the gender is taken into consideration. Finally, questions about age, nationality and country of living were asked for taking into account whether culture can have an impact in color naming.

4 RESULTS

As mentioned above, the test was based on a 24 colour model (Figure 5). Based on the selected labels and their relative intervals in the HSL pyramid, the computer was able to identify all 24 colors. As for the human color labelling, a total of 11 people took part in the experiment. The limited number of participants is due to the desire of executing the test with the same screen and on the same conditions. This guarantees that there is no unknown discrepancy between the image shown by the computer, as different screens with different settings can influence the answer.

It is important to remark that the majority of the individuals are male between 18 and 35 years of age, the main nationalities are Italian and Spanish, and all participants have a good level of English. The participants were also asked to define their experience with color, that is whether they work with colors (color scientists, painters, designers, …), if they use the color for hobby (paint in their free time, decorate, …) or have no particular experience related to color classification.

The color names obtained by people are not homogeneous, as appears from the representative example in **Figure 6.** This is due to the fact that, even when put in the same conditions, color naming is very subjective.

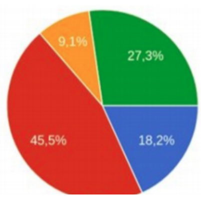

Figure 6: Example of answers: one of the results obtained for one color when people were asked to assign a label to it. A variety of answers was given to the same question; however, it is possible to identify the most popular one.

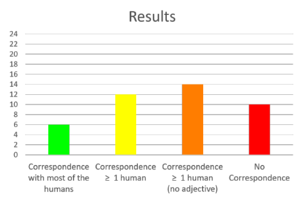

Figure 7: Results obtained: comparison between the labels obtained from the test performed by a human sample group and the ones obtained by the computerized color labelling.

Surprisingly, in one case out of 11, the color red that was repeated in the questionnaire was categorized by the same person differently in the two occurrences. In the first the color was identified as light red, while in the second it was identified as dark red. Despite this discrepancy, it was decided not to discard the sample, as this phenomenon is most likely due to the effect of the tuning process of the human categorization ability. In this case, in fact, the colors seen before and the experience on labelling acquired during the test itself can be considered the cause of the incoherence.

A plot of the results obtained can be seen in Figure 7. As it appears from the graph, the correspondence of the label obtained by the computer and the label obtained by the majority of people is perfect only 6 out of 24 times. This rate doubles if we count the times that at least one person gave the same answer as the algorithm.

It is also important to notice that the adjective assigned to the color is relevant, but it is not discriminant for the results obtained. In fact, when the adjective is disregarded the number of matches grows just from 12 to 14. Thus, it can be said that, in the majority of the cases, the labels were not different because of the adjective, but because of the identification of the hue value itself.

One of the main problems that emerged by looking at the results of the experiment is that the lack of a differentiated label between orange and brown causes a big discrepancy in the answers. What appeared is that the concept of light brown and the concept of light orange are very different in the human mind, while in the test performed by the computer they were put in the same category.

Similarly, the algorithm cannot assign an adjective to the colors *black* and *white*, while some people did.

Another discrepancy is related to *pink* and *red*. It emerged that 100% of the participants labeled the color in Figure 8a) as *pink* (pale in most cases), while for the computer the color is categorized as *light red*. This must be taken into consideration when evaluating and using the model.

a) b)

Figure 8. Example of ambiguous colors. Colors in a) and b) were subject to discrepancies between the answers received from people and the answers received from the algorithm. The color in a) was labelled as pale pink by humans and as light red by the computer. The color in b) was labelled as dark yellow by people and as orange/brown by the computer.

Finally, it is important to note that in those cases where the label given by the computer and the one obtained by the group test do not correspond, the color given by the computer is not absurd in most cases. For instance, the color shown in Figure 8b) was identified by people as *dark yellow* and by the computer as *orange/brown*. The answers can be considered close in representation, however a test on the acceptance of the labels given by the computer from people could give a less qualitative idea of the results obtained.

5 DISCUSSION AND CONCLUSIONS

It can be argued that the major limitation of this study is the small sample of participants that took part in the experiment. Nevertheless, it is possible to draw some relevant conclusions from what emerged.

It is important to remark that color perception appears to be highly subjective, hence it is very difficult to reproduce it in a software. This is part of the reason why the results obtained show a relatively low correspondence between the model and human perception.

The problem here seems to be partly due to the fact that a mathematical categorization of color, necessary for a software in order to perform color labelling, is not entirely suitable for the human mind. As we saw, in one case out of 11, the same person labelled the same color differently in two moments of the test and this is something that cannot be solved by changing the intervals in the HSL domain.

Since the tests were performed in the same conditions and on the same screen, it can be said that the discrepancies that resulted between different people and, as mentioned, between the same person, are due to other factors. It is plausible that the order in which the colors were shown to the participants influenced some of the answers.

Moreover, what probably happened is that the use of the adjectives was tuned after a few color samples, and people learnt to use the labels in a coherent way while the test was going on, reducing what can be considered an error of the first few samples. This would also explain why one person categorized the RGB (255,0,0) first as light red and later as dark red.

In order to avoid this error, the test could be performed with more color images and the first answers of the test could be discarded. The reasons this was not done here are several. First, the number of participants in the test is small, so it was considered advisable to take as many answers as possible. Second, the test is rather monotonous and, if too many color images were presented to the participants, they could have been subject to losing their focus and they would have given less precise answers. Last, it is very complex to determine how many answers should be discarded. This could depend on the person themselves and a deeper analysis should be performed.

As mentioned before, further analysis should be done regarding the label *brown*. The model should be coherent and provide the definition for the different labels *light brown*, *dark brown*, *pale brown*, and *brown*. However, the interval in which the brown can be identified is not mathematically continuous and should be studied in depth.

A similar problem emerged with the discrimination of the color red and the color pink, as for most humans the color labelled as light red corresponds to pink. Further studies should focus on that as well.

Future works could be focused on finding which factors (gender, age, nationality, …) have an impact in color labelling, but first the dataset has to be extended with more diverse samples.

Besides, a second test could be done for testing people's acceptance of the labels assigned by the computer for a complete evaluation of the model.

In conclusion, what can be said is that color perception can be very useful in HCI and HRI, as the more the external world is perceived in a similar way, the easier the communication is. This work does not provide a robust solution for the problem of discrepancies in color perception between human and computer, but it can be considered a starting point for future work. What comes clear from this study is that, if a static and a priori color labelling is required, the users should be able to customize their choices and the model should adapt to their color interpretation.

The study highlights some of the problems in the interpretation of color not only from person to computer, but also from person to person. For this reason, designing a robot that is good in color recognition is difficult, but necessary for many applications.

ACKNOWLEDGEMENTS

The funding and support by the Spanish Ministry of Economy and Competitiveness (project TIN2017-88805-R), and Universitat Jaume I (project UJI-B2017-73) are gratefully acknowledged.

REFERENCES

[1] Hoang Long Cao et al. Robot-enhanced therapy: development and validation of a supervised autonomous robotic system for autism spectrum disorders therapy. In: *IEEE Robotics and Automation Magazine* (2019).

[2] Luc Steels, Tony Belpaeme, et al. Coordinating perceptually grounded categories through language: A case study for colour. In: *Behavioral and brain sciences 28.4* (2005), pp. 469-488.

[3] Ismael Sanz et al. Customising a qualitative colour description for adaptability and usability. In: *Pattern Recognition Letters 67* (2015), pp. 2-10.

[4] Lledó Museros et al. Creating, Interpreting and Rating Harmonic Colour Palettes Using a Cognitively Inspired Model. In: *Cognitive Computation* (2018), pp. 1-18.

[5] Rashad J Rasras, Ibrahiem MM El Emary, and Dimitri E Skopin. Developing a new color model for image analysis and processing. In: *Comput. Sci. Inf. Syst.* 4.1 (2007), pp. 43-55.

[6] Paul Kay and Terry Regier. Resolving the question of color naming universals. In: *Proceedings of the National Academy of Sciences 100.15* (2003), pp. 9085-9089.

[7] Z. Falomir, L. Museros, L. Gonzalez-Abril, A model for colour naming and comparing based on conceptual neighbourhood. An application for comparing art compositions, *Knowledge-Based Systems* 81: 1–21, 2015.

[8] Freksa C. Spatial computing. Cognitive and Linguistic Aspects of Geographic Space: New Perspectives on Geographic Information Research. In: *Raubal M, Mark DM, and Frank AU, editors*. Berlin: Springer Berlin Heidelberg. p. 23–42, 2013.

[9] Israel Abramov et al. Sex and vision II: color appearance of monochromatic lights. In: *Biology of sex differences 3.1* (2012), p. 21.