Visual Perception with Color for Architectural Aesthetics

Michael S. Bittermann
Department of Architecture
Maltepe University
Maltepe-Istanbul, Turkey
michaelsbittermann@maltepe.edu.tr

Özer Ciftcioğlu, Senior Member, IEEE
Department of Architecture
Delft University of Technology, The Netherlands
Maltepe University, Maltepe - Istanbul, Turkey
o.ciftcioglu@tudelft.nl  ozerciftcioglu@maltepe.edu.tr

Abstract—Studies on computer-based visual perception and aesthetical judgment for architectural design are presented. In the model, both color and the geometric aspects of human vision are jointly taken into account, quantifying the perception of an individual object, as well as a scene consisting of several objects. This is accomplished by fuzzy neural tree processing. Based on the perception model, aesthetical color compositions are identified for a scene using multi-objective evolutionary algorithm. The methodology is described together with associated computer experiments verifying the theoretical considerations. Modeling of aesthetical judgment is a significant step for applications, where human-like visual perception and cognition are of concern. Examples of such applications are architectural design, product design, and urbanism.

Keywords—visual perception; color difference; fuzzy neural tree; architectural design; genetic algorithm; Pareto front

I. INTRODUCTION

Visual perception is human’s main source of information. Therefore understanding and modeling human interaction with environment, inevitably involves this subject. Developing models of visual perception is relevant to the diverse fields where human interaction with environment is concerned, such as cybernetics, robotics, medicine, architecture, and industrial design, making the subject an important one. Perception has been extensively treated in the literature, for instance in philosophy and psychology. Descriptions of the phenomenon in these fields are generally qualitative or statistical in nature. Despite their validity, descriptions of the basic nature of perception are lacking in precision. The same issue also applies to other areas dealing with perception, such as psychophysics [1, 2], and image processing [3-8], where the perception concept referred to is generally expressed not mathematically but linguistically. For instance, in the psychophysics and cognition works, brain processing in human visual system is explained via neurobiological terms rather than mathematical ones. Yet, establishing a model of human perception implies that the phenomenon should be treated in computational form for minimal ambiguity. Although image processing studies are a matter of computation, and they traditionally do make reference to biological vision in order to justify or have inspiration in the development of machine vision algorithms, the algorithms resemble to human vision only in a restricted sense. Generally an image processing algorithm singles-out a component of perception process occurring in human visual system. Examples are the ample edge detection studies in the literature, e.g. [9, 10], and works on recovering three-dimensional object information from two-dimensional image data, e.g. [11, 12]. However, due to the specific nature of the image processing applications, there is no need that the computations reflect some general characteristics of human vision that are due to the totality of interrelated brain processes. One of the most observable of such characteristics is the uncertainty of remembering visual information.

When an observer pays visual attention to multiple objects existing in his visual scope, there is likelihood the observer does not notice the presence of some objects, i.e. he is unable remember them. The cause of the phenomenon is the complexity of human visual system. Dealing with applications that need not closely resemble human activities, image processing works generally can afford to ignore such complexity induced properties. Hence, the works generally do not refer to human vision as the object of the modeling effort. The complexity is due to the multiple, interrelated, and merely partially understood brain processes that are involved in perception. Therefore probabilistic treatment of the phenomenon is most convenient, as the imprecisions in description of the concept are subject to absorption in a probabilistic model. In order to delineate the probabilistic treatment in this work from existing probabilistic treatment in vision related research, one notes that in object detection studies perception is considered to be the engagement of pattern recognition. In this case Bayesian methods are appropriate [13]. However, for human perception modeling Bayesian approach turns out to be trivial [14]. This is because for human the probability of a retinal image given a certain scene is almost certain, which implies that the recognition of the scene given a retinal image is also almost certain. Therefore, as the subject matter of this work is human perception, notwithstanding the validity of the Bayesian approaches for computer vision, and the same approaches are of minor importance here.

In this work two causes of the uncertainty in visual perception are considered. The first cause is that in a scene with multiple objects, visual attention is only partly devoted to each object, so that the amount of attention paid for an object has a likelihood of being insufficient for yielding the remembrance...
of the object. An object subverting a larger portion of our visual field implies a greater likelihood to be seen, and hence remembered. This common phenomenon has been treated in the literature [15]. The second cause underlies the following common experience. An object having very similar color with objects behind, next to, and in front of it may not stand-out enough from its surrounding to be noticed. Conversely, it is likewise common experience that a greater color difference between an object and its background implies a greater likelihood we see and remember the object. Modeling of visual perception including both, the geometric and the color aspects, is missing in the literature. This work tackles this issue by fusing geometric and color perceptual information. The fusion is possible, since both perceptual properties are treated as likelihoods in this work, while they have alternative interpretations as fuzzy memberships. Therefore, the fusion is executed by means of a likelihood-based fuzzy computation, quantifying the intensity of a perception in the form of likelihood. This is the first item addressed in this study. Based on the first one, the second item is pinpointing the role of the intensity of perception in aesthetic judgment of the color composition of a scene. Based on these considerations, the reason why certain color compositions of a scene strike an unbiased observer as aesthetical is investigated. The organization of the paper is as follows. In section II, computation of the likelihood of visual perception is presented. Based on the perception computations, the multi-dimensional nature of color aesthetics is exposed in section III. The validity of the model is verified by means of computer experiments in section IV. This is followed by conclusions.

II. LIKELIHOOD OF VISUAL PERCEPTION

One source of the uncertainty characterizing visual perception is due to the geometry of an object in view in relation to visual scope of observation. This has been treated in the literature [15], where visual attention is modeled as a uniform probability density (pdf) with respect to solid vision angle \( \Omega \) defining the visual scope. The pdf is given by

\[
f_{\Omega}(\Omega) = \frac{1}{\Omega} \Omega.
\]  

(1)

Based on visual attention, the perception of an object due to its geometry, i.e. its occupation of visual scope, is defined as the integral of the attention over the domain subtended by the object. The domain is expressed by the solid angle \( \Omega \). The perception is quantified by the likelihood \( \mathbb{L}_v \) and obtained by

\[
\mathbb{L}_v = \frac{\int_{\Omega} \frac{1}{\Omega} \Omega \cdot \Omega \cdot \Omega}{\Omega}.
\]  

(2)

Next to geometry, the difference in color between an object and the objects surrounding it is a second important condition to be fulfilled for perceiving an object. For computing color difference, first it is imperative to represent a color numerically. This is the subject matter of colorimetry. In extensive color matching experiments, the Commission Internationale de l’Eclairage (C.I.E.) established the means to represent by vectors of three numbers the set of colors a standard human observer is able to perceive [16, 17]. The underlying theory is due to Grassmann’s laws of trichromatic generalization [18]. These state (i) stimuli with same specifications look alike to an observer with normal colour vision under the same observation conditions, (ii) stimuli that look alike have the same specification, and (iii) the numbers comprising the specification belong to continuous functions [19]. In such matching experiments a large group of observers is asked to produce a certain colour by adjusting separately the intensity of three monochromatic primary colours red, green and blue with known wavelengths. The observers should combine the three colours in such a way that the combination matches as close as possible to a given monochromatic test sample. The colours are typically presented in the two halves of a bipartite visual field. This way any influence of geometry in the colour matching is taken care of. The three numbers resulting from the conversion of a light stimulus using C.I.E.’s standard observer model [20] are termed as tristimulus values. In their normalized from they are referred to as chromaticity coordinates [21].

Estimating the difference between two colors in the context of perception modeling requires that the difference quantity obtained matches to the difference a standard human observer would attribute to the color pair. This is conveniently accomplished when the color space, in which the pair is represented, is perceptually uniform. This property stipulates that the Euclidean distance between the chromaticity coordinates of two colors quantifies the color difference attributed by human for these colors. The first color spaces introduced by C.I.E., namely the 1931 C.I.E. RGB space and a transformed version of it named 1976 C.I.E. XYZ space, have both been shown to lack in perceptual uniformity [22]. To alleviate this drawback, C.I.E. introduced two approximately uniform spaces by two different transformations of the XYZ space, known respectively as CIE 1976 \( L_u \) \( v^* \) and CIE 1976 \( L' \) \( a' \) \( b' \) spaces [20, 23]. In either space the \( L' \) dimension expresses the lightness of a color; \( u^*, v^* \) and \( a', b' \) are chromaticity coordinates, which in respective combination specify the saturation and hue of a color. Detailed definitions of these quantities can be found in [19]. It is emphasized that due to the approximate perceptual uniformity of both spaces, color differences in either space are obtained by the Euclidean distance among the chromaticity coordinates. Explicitly, the color difference \( \Delta E_{ab} \) between two colors in \( L' a' b' \) space is obtained by

\[
\Delta E_{ab} = \sqrt{(L'_1 - L'_2)^2 + (a'_1 - a'_2)^2 + (b'_1 - b'_2)^2}.
\]  

(3)

One notes that the uniformity in the distance computation for small color differences, i.e. \( \Delta E_{ab} < 10 \) has been further improved by the C.I.E. due to [24]. The difference \( \Delta E_i' \) between two colors that both have chromaticity coordinates \( a' \neq 0 \land b' \neq 0 \), is generally larger compared to distance between their achromatic, i.e. grey, counterparts, denoted b \( \Delta E_i \), where \( a' = 0 \land b' = 0 \). Converting a chromatic color to its achromatic counterpart means that both chromaticity coordinates \( a^* \) and \( b^* \) are set to zero, so that the color is devoid of chroma. Geometrically, an achromatic counterpart of a color is obtained by projecting the color parallel to the \( a' b' \)-plane onto the \( L' \) axis. Referring to (3), with the exception of the trivial case the two colors at hand are equal, since always

\[
\sqrt{(L'_1 - L'_2)^2 + (a'_1 - a'_2)^2 + (b'_1 - b'_2)^2} < L'_2 - L'_1
\]  

(4)

therefore

\[
\Delta E_i' > \Delta E_i.
\]  

(5)
This explains why chromatic images are generally more memorable compared to their achromatic counterparts.

Investigating the role of color difference in perception, as an initial simple case we consider the situation of a single occlusion, namely a single object in front of a background. The likelihood the object was perceived due to color is dependent on the difference between the object’s color and the color of the background measured in a perceptually uniform color space. It is to note that, although the relative color difference due to color difference vis-à-vis every partly occluding and occluded object needs to be taken into account in this computation. We denote the number of such occlusions by \( n \), and the likelihood of perception of an object due to color difference by \( L_C \). The likelihoods of perception due to geometry \( L_G \) given by (2) of every occlusion region belonging to the object of concern determine the weights in the computation of \( L_C \) as given by

\[
L_c = \frac{\Delta E_{ab}'}{\Delta E_{ab,\text{max}}} = \frac{\sqrt{(L_c' - L_c) + (a'_c - a_c) + (b'_c - b_c)}}{\sqrt{(L_{ab,\text{max}} - L_{ab,\text{max}}) + (a'_{ab,\text{max}} - a_{ab,\text{max}}) + (b'_{ab,\text{max}} - b_{ab,\text{max}})}}
\]

where \( \Delta E_{ab,\text{max}} \) is the maximal color difference between two colors in the uniform space. In case an object has exactly the same color as the background, then the likelihood the object has been perceived due to color difference vanishes. Conversely, if an object has the maximally possible color difference in perceptually uniform color space, then the likelihood, the object has been perceived due to its color is maximal, namely unity. It is noted that in (6) \( \Delta E_{ab,\text{max}} = 375.6 \) when the entire visible spectrum of colors is considered. Color reproducing devices, such as computer monitors, however are not capable of displaying the entire visible spectrum, but merely a restricted portion of it. The portion is referred to as the device’s gamut. Therefore, in computer-based applications generally \( \Delta E_{ab,\text{max}} < 375.6 \). We consider another scene, where there is a second object located behind our first one, and in front of the background. In case object nr. 2 is larger than object nr. 1 and located in such a way that nr. 1 no color difference vis-à-vis the background anymore, but only vis-à-vis object nr 2, then \( L_C \) of object nr. 1 is given as before by (6). The only difference is, this time \( \Delta E_{ab}' \) is computed between the colors of the first and second object.

In (7) \( \Delta E_i' \) denotes the color difference given by (3) between the \( i \)-th object and the object of concern; \( L_{Gi} \) and \( L_{Gj} \) are the likelihoods of perception given by (2) of the \( i \)-th, respectively \( j \)-th, occlusion region. Comparing (6) and (7) one notes that (6) is a special case of (7), namely for \( n = 1 \). Considering the other extreme, when a certain partial occlusion is hardly noticeable due to geometry, i.e. \( L_{Gi} \ll \sum_{j=1}^{n} L_{Gj} \) it is clear that the color difference belonging to this particular occlusion hardly influences the likelihood of perception due to color. To illustrate (7) we consider the perception of object nr. 1 in the figure. The significance of the color difference the object has with the background is comparable with the difference between object nr. 1 and object nr. 2. The two occlusion regions, the one between object nr. 1 and the background and the one between objects nr. 1 and nr. 2, occupy approximately the same solid angle. Hence, for object nr.1 \( L_C \cong 0.5(\Delta E_1' + \Delta E_2') \). The situation appears to be quite different for object nr. 2. The occlusion region between this object and the background, and the one between object nr. 2 and object nr. 1, apparently involve quite different values for \( L_C \).

Due to the complexity of geometric constellations of objects in visual scope, analytical treatment of the situation is inconvenient. In particular computing the values of \( L_G \) in (7) is a problematic issue, since the solid angle of the occlusion region is ill defined. To handle the case we use a probabilistic ray tracing approach as follows. We are sending and tracing rays away from the point of observation in random directions within visual scope, as seen in the figure. One notes that due to (1) the probability density function (pdf) underlying the ray emission ought to be constant per differential solid vision angle \( \Omega \). The set of rays is termed vision rays and denoted by \( R \). In the figure \( R = \{r_1, \ldots, r_k\} \). Each vision ray intersects either object nr.1, object nr. 2, or the background as the object nearest to the observer along the ray. For the clarity of the explanation, we consider computing \( L_C \) for object nr. 1 as a generic example. For this we note two relevant subsets of \( R \). One set \( R \subset R \) contains the vision rays intersecting object 1, that we refer to as perception rays of object 1, \( R_1 = \{r_1, r_2\} \). The other set \( R_{A1} \subset R \) contains all vision rays that do not intersect the object, i.e. \( R_{A1} = R - R_1 = \{r_3, r_4, r_5, r_6\} \). For every element of \( R_1 \) we find that element of \( R_{A1} \) so that the angle subtended by the two rays is minimal. For \( r_1 \) this element is \( r_2 \), since the \( \alpha_{12} < \alpha \forall \alpha \in \{\alpha_{14}, \alpha_{15}, \alpha_{16}\} \), and for \( r_2 \) the element is \( r_3 \). We term \( r_2 \) and \( r_3 \) as occlusion rays of object 1. They form the set

![Fig. 1. Rays simulating visual perception of the color differences along the perceived limitation of object nr. 1 vis-à-vis object nr. 2, as well as object nr. 1 vis-à-vis the background.](image-url)
$R_C = \{r_2, r_3\}$ that is used to approximate $L_C$ in (7) as expressed by (8).

$$L_C = \frac{1}{\Delta E_{\infty}} \sum_{i} \sqrt{(L_i - L_0)^2 + (a_i - a_0)^2 + (b_i - b_0)^2}$$

In (8) $i$ is the cardinality of $R_C$ and equal to that of $R_1$. In our example in figure 1, $R_C = 2$. The variables $L_i, a_i, b_i$ specify the color of the object of concern, i.e. object nr. 1 in this example; the variables $L_i', a_i', b_i'$ specify the color of the object intersected by the $q$-th element of $R_C$, i.e. the $q$-th occlusion ray. As to object nr. 1 in the figure, $L_C$ is given by

$$L_C = \frac{1}{2} \sqrt{(L_1 - L_0)^2 + (a_1 - a_0)^2 + (b_1 - b_0)^2} + \sqrt{(L_2 - L_0)^2 + (a_2 - a_0)^2 + (b_2 - b_0)^2}$$

where the indices 1 and 2 refer to objects 1 and 2 respectively, and index 3 refers to the background. The accuracy of the result depends on the number of rays that are used to simulate the vision, which is in the order of $10^4$, and it can be raised to an arbitrary value, limited exclusively by the available computation time.

Fusion of the geometric and color perception information is accomplished in this work using the likelihood-based neural tree method known as fuzzy neural tree (FNT) [26]. The rationale to use the approach is that the imprecision of the perceptual information necessitates treatment by means of a soft computing methodology. The categorization into geometric and color aspects in perception is an act of human linguistic abstraction that it is best dealt with the methods of soft computing. In a fuzzy neural tree, the output of $i$-th terminal node is denoted by $x_i$ and it is introduced to a non-terminal node. The detailed views of node connections from terminal node $i$ to internal node $j$ and from an internal node $i$ to another internal node $j$ are shown in the publication [26]. A connection weight between two nodes is shown as $w_{ij}$ in both cases. In the neural network terminology $w_{ij}$ is the synaptic strength between the neurons. Both, terminal and non-terminal node outputs have interpretation as likelihood. Accordingly the weights denoted by $w_{ij}$ in [26] are shown as the likelihood parameter of node $j$ and the output $O_j$ of inner node $j$ in [26] are shown as $L_j$ in the following equations and figures. We consider a non-terminal node $j$ that has two inputs, which are the outputs of two previous nodes denoted by $O_i$ and $O_k$. As the two inputs to a neuron are assumed to be independent of each other, the fuzzy memberships at the inputs can be thought to form a joint two-dimensional fuzzy membership. In this case $L_j$ is computed by [26]

$$L_j = L_i(\theta) L_k(\theta) = e^{-\frac{\theta_i^2}{2\sigma_1^2}(1-1)^2} e^{-\frac{\theta_k^2}{2\sigma_2^2}(1-1)^2}$$

where $\sigma_1$ is a constant maximizing satisfaction of the consistency condition of possibility theory. For the two-input case $\sigma_1 = 0.299$. The likelihood parameters $\theta_1$ and $\theta_2$ are selected commensurate to the amount of information each of them conveys via the respective connection. The selection is in accordance with Shannon’s information theorem. Further, the likelihood parameters must sum up to unity for defuzzification in the rule-chaining process from node to node. Due to these stipulations, the likelihood parameters in (9) are given by

$$\theta_1 = \frac{1 - O_i}{(1 - O_i)(1 - O_k)}, \quad \theta_2 = \frac{1 - O_k}{(1 - O_i)(1 - O_k)},$$

so that (9) becomes

$$L_j = e^{-\frac{1}{2\sigma_1^2}(1-1)^2} e^{-\frac{1}{2\sigma_2^2}(1-1)^2}$$

The output neuron of a fuzzy neural tree is termed as root node and denoted by $\Omega$. The inner nodes providing the input to the root node are instances of $L_j$ in (9). They are termed as penultimate nodes and denoted by $L_{nk}$. The root node output is obtained via the weighted summation given by (12), which represents the final defuzzification of the information processed in the neural tree.

$$\Omega = \sum_{i=1}^{n} w_i L_{i\Omega}, \quad \sum_{i=1}^{n} w_i = 1$$

In (12) $n$ denotes the number of scene objects. A bias regarding the relative importance of the information coming from the penultimate nodes may be absent. Particularly, this is the condition for visual perception in the aesthetical context. In this case an important weight vector $(w_1, w_2, ..., w_n)$ is the one that is aligned to the feature vector $(L_1, L_2, ..., L_n)$. It maximizes the output of the defuzzification operation with the fuzzy logic principles, taking the information from each input into account commensurate with the information’s relative fuzziness. That is, the influence of a root node’s input on the node’s output is proportional to the likelihood associated with the input, namely $w_k = c L_k, \forall k \in \{1, 2, ..., n\}$ where $c$ is a scale factor and a constant. The aligned defuzzification corroborates with common human vision experience. An object’s attributes influences the perception of the same attributes as property of the scene commensurate with the perception of the object.

Fulfilling the conditions of defuzzification, $c$ is to be selected in such a way that the components of $w$ sum up to unity as stipulated in (12). In this case the root node output is given by [27]

$$\Omega = \frac{\sum_{i=1}^{n} L_i^j}{\sum_{i=1}^{n} L_i}$$

Figure 2 shows the FNT modeling visual perception based on the above computations.

Each inner node is associated to one scene object, and it has two inputs that are the likelihoods given by (2) and (8). As the color difference and geometric inputs to visual perception neu-
is given by fuzzification of the perceptual likelihoods of the objects. This is obtained via aligned defuzzification of the perceptual likelihoods of the objects. This is given by

\[
\mathbb{L}_p = \mathbb{L}_s(\theta_e) \mathbb{L}_s(\theta_c) e^{-\frac{\theta_c^2 (L_c-1)^2}{2\sigma_c^2}} e^{-\frac{\theta_e^2 (L_e-1)^2}{2\sigma_e^2}}
\]  

Due to (10) the likelihood parameters of the perception FNT are given by

\[
\theta_c = \frac{1 - L_c}{(1 - L_c) + (1 - L_o)}
\]

\[
\theta_e = \frac{1 - L_o}{(1 - L_c) + (1 - L_o)}
\]

Analog to (11), using (15) and (16) in (14) yields

\[
\mathbb{L}_p = \mathbb{L}_s(\theta_e) \mathbb{L}_s(\theta_c) e^{-\frac{1}{2\sigma_c^2} \left( \frac{1 - L_c}{(1 - L_c) + (1 - L_o)} \right)^2 (L_c-1)^2}
\]

\[
\mathbb{L}_p = \mathbb{L}_s(\theta_e) \mathbb{L}_s(\theta_c) e^{-\frac{1}{2\sigma_e^2} \left( \frac{1 - L_o}{(1 - L_c) + (1 - L_o)} \right)^2 (L_e-1)^2}
\]  

quantifying the output of an inner tree node. It expresses the likelihood of perceiving one of the scene objects. The root node output of the perception FNT models the perception of the scene, which is the probability the scene has been seen. It is denoted by \(\mathbb{P}\). Analog to (13) \(\mathbb{P}\) is obtained via aligned defuzzification of the perceptual likelihoods of the objects. This is given by

\[
\mathbb{P} = \sum_{k=1}^{n} \mathbb{L}_p^k / \sum_{k=1}^{n} \mathbb{L}_p
\]  

Due to (13), where we take the perception of each object into account commensurately with the relative fuzziness characterizing each perception. The perception of chromatic properties of the objects and scene are detailed in [28].

An exemplary scene subject to perception computation is shown in figure 3.

It consists of five building elements labeled accordingly in the figure, as well as a background. The intersections of the rays with the objects of the scene are shown as green dots in the figure, and they are referred to as perception points. Those among the perception points that represent perception of an occlusion region are shown by red dots and referred to as occlusion perception points. The rendering of the same scene produced from the viewpoint \(P\) shown in figure 3 is shown in figure 4a, where a random color composition has been assigned to the scene objects. The colors assigned to the objects and the likelihoods of perception for each of them (17) are given in table 1. Figure 4b shows the same scene as figure 4a, except the frontal wall’s color has been changed to white.

\[
\begin{align*}
\mathbb{L}_p &= \mathbb{L}_s(\theta_e) \mathbb{L}_s(\theta_c) e^{-\frac{\theta_c^2 (L_c-1)^2}{2\sigma_c^2}} e^{-\frac{\theta_e^2 (L_e-1)^2}{2\sigma_e^2}} \\
&= e^{-\frac{1}{2\sigma_c^2} \left( \frac{1 - L_c}{(1 - L_c) + (1 - L_o)} \right)^2 (L_c-1)^2}
\end{align*}
\]

For figure 4a the scene perception is \(\mathbb{P} = 0.149\), whereas for figure 4b \(\mathbb{P} = 0.142\).

The differences as to \(\mathbb{L}_p\) among the objects are seen from figure 5. Let us consider these differences in the light of the difference between the \(\mathbb{P}\) values of the two scenes. One notes that although the perception of several objects is severely diminished when the color of the frontal wall is changed to white, the scene perception is only slightly diminishing, namely by 5%. This is due to the alignment of the defuzzification taking place at the root node given by (18). Reduction of the scene perception is attenuated, due to the commensurate emphasis of perceptual information stemming from objects that remain rather well perceived in the scene. Such an object is the ceiling in the case in figure 4.

III. MULTI-OBJECTIVE NATURE OF COLOR AESTHETICS

One of the important application areas of human visual perception involving color is architectural design. The colors of the architectural objects forming an environment, such as walls, columns, floor and ceiling, have a significant influence on the visual perception of the environment and thereby on an aesthetical judgment made about the environment. Aesthetical judgment is assessing the magnitude of pleasure caused directly by perception. The directness refers to the absence of rea-
soning for estimating utility or other abstract qualities of
the object. According to its strict definition, aesthetical judgment
is devoid of individual or collective preferences [29, 30]. In
the aesthetic context the two-folded influence of color on
visual perception is to point-out. From one side, as shown
above, the color composition of a scene determines how in-
tensely each object is perceived and thereby how intensely the
scene as a whole is perceived. However, color is not merely
involved in the event that one has seen an object and a scene.
It also determines what one sees, i.e. what the nature of a sce-
ne is. The aesthetics of a man-made environment can be ap-
preciated directly from visual perception under the following
conditions. The scene is to judge as aesthetical, when a certain
degree of scene perception is present, while for this degree,
color is involved as parsimoniously as possible. Then we look
at a landscape during foggy weather, then the perception of
the scene is bound to have a low intensity. This is obviously
fulfilling one’s perceptual expectation, due to the generally
low color difference among weakly chromatic colors, so that it
will not strike an observer as an extraordinary perception.
Conversely, however, it is a rare and hence notable event, if
low chroma coincides with rather high scene perception at the
same time. The likelihood of such an event to occur acci-
dentally is very low, because each color of a scene object is a
point in an extensive three-dimensional color space. Therefore,
the combination of several objects implies a vast space of pos-
sible compositions, requiring significant conscious effort to
resolve the conflict inherent to the objectives of high percep-
tion and chromaticity parsimony. Accordingly, scenes repre-
senting solutions to the scene perception-color parsimony-
problem are aesthetical ones.

The color parsimony mentioned above is in the sense that
the scene is perceived to minimally differ from its achromatic
reference color that has the CIELAB coordinates \( L^* = L_1^* \);
\( a^* = 0, b^* = 0 \). The minimal difference is considered in two
senses. In one sense, the difference is with respect to the light-
ness dimension \( L^* \). In the other sense the difference concerns
CIELAB chroma denoted by \( C_{ab} \) and obtained by [20]
\[
C_{ab} = \sqrt{a^* + b^*}
\] (19)
As to the chromatic aspect of the parsimony, it concerns the
extent by which a color is without chroma, i.e. \( C_{ab} = 0 \). If this
condition is fulfilled, then the color resembles the scene refer-
ce color in the chromatic sense. The likelihood of chroma
absence is denoted by \( L_{C_{ab}} \) and given by
\[
L_{C_{ab}} = 1 - \frac{C_{ab}}{C_{ab, max}} = 1 - \frac{\sqrt{a^{*2} + b^{*2}}}{\sqrt{(a_{max}^{*2} - 0) + (b_{max}^{*2} - 0)}} = 1 - \frac{\sqrt{a^{*2} + b^{*2}}}{\sqrt{(a_{max}^{*2} + b_{max}^{*2})}}
\] (20)
In case one would consider chroma for all visible colors, then
\( C_{ab, max} = 181 \). In the ensuing computer experiments in this
work \( C_{ab, max} = \sqrt{a_{max}^{*2} + b_{max}^{*2}} = 134 \), which is due to the re-
stricted gamut formed by the colors that can be displayed on a
standard computer monitor.

Second, we consider the lightness aspect of the color parsim-
ony. The likelihood, the observer perceived an object’s
lightness to be the same as the scene’s reference lightness, is
denoted by \( L_{L^*} \) and given by
\[
L_{L^*} = 1 - \frac{|L^* - L^*_s|}{L_{max} - L_{min}} = 1 - \frac{|L^* - L^*_s|}{L_{max}}
\] (21)

Based on the above considerations the fuzzy neural tree for
color parsimony is shown in figure 6. Each inner node is asso-
ciated to one scene object, and it has two inputs that are the
FNT terminal nodes given by (20) and (21). Analog to (10) the
likelihood parameters of the color parsimony FNT are given by
\[
\theta_L = \frac{1 - L_{L^*}}{1 - L_{L^*_s} + (1 - L_{L^*_s})}
\] (22)
The output of an inner node of the FNT represents the likeli-
hood the associated scene object has the same achromatic
character and lightness level as the reference color of the sce-
ne. We term this quantity as the likelihood of perceived color
parsimony and denote it by \( L_{C_{ab} L^*} \). Due to (11) it is obtained by
\[
L_{C_{ab} L^*_n} = L_n\left(\theta_L \right) L_{L^*} \left( \theta_L \right) = e^{-\frac{1}{2\sigma^2} \left[ \frac{1 - L_{C_{ab} L^*_n}}{(1 - L_{L^*_s})} \right] \left( \frac{1 - L_{C_{ab} L^*_n}}{(1 - L_{L^*_s})} \right)^2}
\] (24)
The root node output of the color parsimony FNT models the
perception of the color parsimony of the scene. It is denoted by
\( P_{C_{ab} L^*_n} \). Analog to (13), it is obtained via aligned defuzzi-
cation of the color parsimony likelihoods associated by every
object
\[
P_{C_{ab} L^*_n} = \sum_{i=1}^{n} L_{C_{ab} L^*_i}^{(n)} = \sum_{i=1}^{n} \frac{\sum_{j=1}^{m} L_{L^*} \left( \theta_L \right)}{\sum_{j=1}^{m} L_{L^*} \left( \theta_L \right)}
\] (25)
In (25) the parameter \( n \) denotes the number of scene objects.
In the equation the vector \( \mathbf{w} \) is not aligned to the vector
\( \left( L_{C_{ab} L^*_1}, L_{C_{ab} L^*_2}, ..., L_{C_{ab} L^*_n} \right) \) consisting of the objects’ color parsimo-
ny likelihoods from (24). The alignment is to the vector
\( \left( L_{L^* 1}, L_{L^* 2}, ..., L_{L^* n} \right) \) consisting of the objects’ perception likeli-
hoods from (17). This is done so that an object’s chromaticity
influences the perception of a scene’s chromatic properties
commensurate with the object’s likelihood of perception.

Using the mathematical terms above, the verbal definition of
an aesthetical judgment of a scene’s color composition is
rendered more precisely as follows. It is measurement of the
effectiveness in resolving the perceptual conflict between
maximizing (18) and (25) at the same time. It is to emphasize
that both objectives are conflicting with each other, and this is the case independent of what kind of aesthetics is pursued in the design, i.e. independent of the achromatic reference color $L_r^*$. The conflict can be seen as follows. Equation (20) yields greater values for $L_{ab}^*$ for smaller values of $a^*$ and $b^*$, contributing to a greater value of $P_{C_{ab}}^*$ in (25) in this case. Low values of $a^*$ and $b^*$, however, generally imply a low degree of color difference $\Delta L_{ab}^*$ in (3). This contributes to lower object perception $L_p$ in (17) and hence lower scene perception $P$ in (18). An architectural scene is an aesthetic one, when it satisfies the condition of Pareto optimality for the two conflicting objectives given by (18) and (26) that are both subject to maximization for $L_r^* \in [L^*: 0 \leq L^* \leq 100]$. Given a certain scene, in the case there exists no other color composition that at the same time yields a greater scene perception AND greater color parsimony, then the scene is to be judged as an aesthetic one. If this condition is fulfilled for $L_r^* \approx 100$, then the scene should be further specified as beautiful; if it is fulfilled for $L_r^* \approx 0$, the scene should be labeled as sublime.

IV. COMPUTER EXPERIMENTS

Architectural design involves search for aesthetically pleasing environments. As to color, this means identification of appropriate color composition fulfilling the sublimity or beauty conditions described above. One notes that due to combinatorial explosion the number of possible color compositions is enormous, even for a moderate number of objects in a scene. Therefore, identification of beautiful or sublime compositions by stochastic search is appropriate. In the following two computer experiments are carried out for a scene with a fixed geometry; in one experiment beautiful compositions are sought and sublime compositions in the other one. For this a multi-objective genetic algorithm is used, namely NSGA-II [31]. This is a popular multi-objective genetic algorithm. The popularity is presumably due to its minimal number of algorithm parameters, which is achieved by a parameter less technique for determining the degree of non-dominance of a solution. The technique is based on structuring the population by passing multiple surfaces through the population in the objective function space, discretizing the degree of non-dominance of the population members. Due to the particularity of the Pareto ranking scheme, conveniently also elitism and crowding distance computation remain without parameters. In the experiments the algorithm parameters were selected as the following standard values: crossover probability 0.9, simulated binary crossover parameter $\nu_c = 10$, mutation probability 0.05, and polynomial mutation parameter $\nu_m = 30$. In the experiments the color of the scene background is $L^* = 81.5$, $a^* = -2.5$, $b^* = -24.0$ and the colors of the scene objects are restricted by the standard RGB (sRGB) gamut [32]. The conversions of the sRGB color coordinates into CIE $L^*a^*b^*$ space are based on the CIE D65 illuminant [33] and 2° CIE Standard Observer [34]. Figures 7a and 7b show the resulting Pareto frontiers of aesthetically colored scenes in objective function space. The difference between the two figures is the value of the reference lightness used during the genetic search process; in figure 7a $L_r^* = 100$ and in figure 7b $L_r^* = 0$. That is, the solutions in figure 7a represent beautifully colored scenes, whereas the ones in 7b represent scenes having sublime color. Four solutions among the Pareto-optimal ones are highlighted in each figure, and the corresponding renderings are shown in figures 8 and 9 respectively. Metaphorically the character of the beautiful color compositions ranges from gentle beauty in figure 8a to intense beauty in 8d. The character of the sublime ones ranges from uncanny obscurity in figure 9a to fierce discernment in 9d.

![Fig. 7. Pareto front of aesthetical color compositions; beautiful compositions, i.e. $L_r^* = 100$ (a); sublime color compositions, i.e. $L_r^* = 0$ (b)](image)

![Fig. 8. The four beautiful color compositions highlighted in figure 7a](image)

![Fig. 9. The four sublime color compositions highlighted in figure 7b](image)
CONCLUSIONS

Computer-based visual perception and aesthetic judgment of a scene’s color composition are presented. Color aesthetics is identified to be the resolution of conflict that exists between two properties of a scene being perceived. The first property is that the scene should have some high likelihood of being perceived and hence remembered. The second property is the color character of the scene should be perceived to be similar to an achromatic reference color for perceived color parsimony. The parsimony should be as great as possible for a certain given intensity of scene perception. Scenes that fulfill both conditions in non-dominated manner are defined as aesthetic ones. Scenes that fulfill parsimony. The parsimony should be as great as possible for a similar to an achromatic reference color for perceived color aesthetics is identified to be the resolution of conflict that exists among them depends on designer’s preference with respect to color parsimony versus intensity of scene perception. Such preference further specifies the kind of aesthetics at hand. For instance in the case of beauty, when the emphasis is on color parsimony we can term the beauty as a gentle kind of beauty. Conversely when scene perception is of primary interest the beauty can be characterized as an intensive kind. This corroborates with the common understanding of architects, that generally there exist multiple, equivalently valid solutions within the same aesthetic category. Relevant computer experiments have been set up. The validity of the theoretical considerations is verified by general acceptance. By means of the computational color perception, scenes with beautiful as well as sublime color compositions for a given scene geometry. Selection of them among depends on designer’s preference with respect to color parsimony versus intensity of scene perception. Such preference further specifies the kind of aesthetics at hand. For instance in the case of beauty, when the emphasis is on color parsimony we can term the beauty as a gentle kind of beauty. Conversely when scene perception is of primary interest the beauty can be characterized as an intensive kind. This corroborates with the common understanding of architects, that generally there exist multiple, equivalently valid solutions within the same aesthetic category. Relevant computer experiments have been set up. The validity of the theoretical considerations is verified by general acceptance. By means of the computational color perception, scenes with beautiful as well as sublime color aesthetics are established. Computational form of aesthetic judgment during a design process is a significant step. In this work color aesthetics is placed on a computational ground, reducing the impression in conventional aesthetic judgment. This is a contribution to the theoretical bases of disciplines that are dealing with aesthetics, such as architectural design, product design, and urbanism.

REFERENCES

[27] M. S. Bittermann, "Intelligent design objects (ido) - a cognitive approach for performance-based design," PhD, Department of Building Technology, Delft University of Technology, Delft, The Netherlands, 2009.