Delft University of Technology

Visual Perception with Color for Architectural Aesthetics

Bittermann, Michael; Ciftcioglu, Ozer

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Visual Perception with Color
for Architectural Aesthetics

Michael S. Bittermann

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

Özer Ciftcioglu, *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 needons 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 subtending 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 aesthetical 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 Ω_S defining the visual scope. The pdf is given by

$$f_\Omega(\Omega) = 1/\Omega_S \tag{1}$$

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 Ω. The perception is quantified by the likelihood $\mathbb{L}_G$ and obtained by

$$\mathbb{L}_G = \int_0^\Omega \frac{1}{\Omega_S} d\Omega = \frac{\Omega}{\Omega_S}, \ \Omega \le \Omega_S \tag{2}$$

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 Euclidian 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 *1931 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 Euclidian distance among the chromaticity coordinates. Explicitly, the color difference ΔE_{ab}^* between two colors in $L^*a^*b^*$ space is obtained by

$$\Delta E_{ab}^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \tag{3}$$

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 ΔE_1^* between two colors that both have chromaticity coordinates $a^* \ne 0 \wedge b^* \ne 0$, is generally larger compared to distance between their achromatic, i.e. grey, counterparts, denoted b ΔE_2^*, where $a^* = 0 \wedge b^* = 0$. 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

$$\sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} < L_2^* - L_1^* \tag{4}$$

therefore

$$\Delta E_1^* > \Delta E_2^* \tag{5}$$

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 is a deterministic quantity, it is by all means to be an essential probabilistic measure of color perception, and therefore it is considered to be as likelihood. It conforms to all the conditions to be a likelihood [25]. It is the likelihood an object is perceived due to color difference. We denote this likelihood by $\mathbb{L}_C$ and define it by

$$\mathbb{L}_C = \frac{\Delta E_{ab}^*}{\Delta E_{ab\_max}^*} = \frac{\sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}}{\sqrt{(L_{max}^* - L_{min}^*)^2 + (a_{max}^* - a_{min}^*)^2 + (b_{max}^* - b_{min}^*)^2}} \qquad (6)$$

where $\Delta E_{ab\_max}^*$ 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) $\Delta E_{ab\_max}^* = 375.6$ 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 $\Delta E_{ab\_max}^* \ll 375.6$. 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 has no color difference vis-à-vis the background anymore, but only vis-à-vis object nr 2, then $\mathbb{L}_C$ of object nr. 1 is given as before by (6). The only difference is, this time ΔE_{ab}^* is computed between the colors of the first and second object.
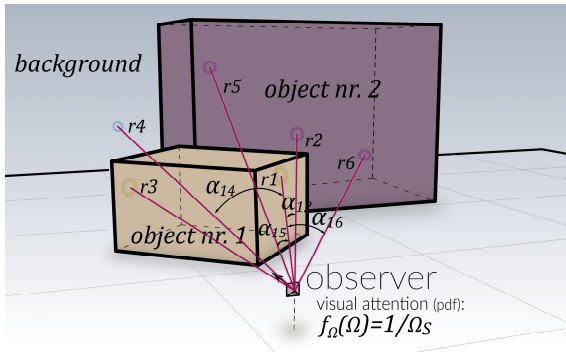

Fig. 1. Rays simulating visual perception of the color differences along the perceived limitation of object nr. 1 vis-à-vis object nr. 2, as well as object nr. 1 vis-à-vis the background.

The situation becomes more involved and interesting when the second object is located in such a way that part of the first object's perception is still vis-à-vis the background, as seen in figure 1. Such partial occlusion is the general case in everyday perception, and one notes that the two situations described just before are special cases of this general one. More explicitly, in

general an object may be partly occluded by another object, or may itself partly occlude another object. In this case multiple color differences need consideration in order to compute the likelihood the object has been perceived due to color. Every 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 $\mathbb{L}_C$. The likelihoods of perception due to geometry $\mathbb{L}_G$ given by (2) of every occlusion region belonging to the object of concern determine the weights in the computation of $\mathbb{L}_C$ as given by

$$\mathbb{L}_C = \frac{\Delta \overline{E}_{ab}^*}{\Delta E_{ab\_max}^*} = \sum_{i=1}^n w_i \frac{\Delta E_i^*}{\Delta E_{ab\_max}^*} = \sum_{i=1}^n [\frac{\mathbb{L}_{Gi}}{\sum_{j=1}^n \mathbb{L}_{Gj}} \frac{\Delta E_i^*}{\Delta E_{ab\_max}^*}] \qquad (7)$$

In (7) ΔE_i^* denotes the color difference given by (3) between the i-th object and the object of concern; $\mathbb{L}_{Gi}$ and $\mathbb{L}_{Gj}$ 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. $\mathbb{L}_{Gi} \ll \sum_{j=1}^n \mathbb{L}_{Gj}$ 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 $\mathbb{L}_C \cong 0.5(\Delta E_1^* + \Delta E_2^*)$. 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 $\mathbb{L}_G$.

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 $\mathbb{L}_G$ 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 Ω. The set of rays is termed *vision rays* and denoted by R. In the figure $R = \{r_1, .., r_6\}$. 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 $\mathbb{L}_C$ for object nr. 1 as a generic example. For this we note two relevant subsets of R. One set $R_1 \subset R$ contains the vision rays intersecting object 1, that we refer to as *perception rays* of object 1. $R_1 = \{r_1, r_3\}$. The other set $R_{\neg 1} \subset R$ contains all vision rays that do not intersect the object, i.e. $R_{\neg 1} = R - R_1 = \{r_2, r_4, r_5, r_6\}$. For every element of R_1 we find that element of $R_{\neg 1}$ 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_3 the element is r_4. We term r_2 and r_4 as *occlusion rays* of object 1. They form the set

$R_C = \{r_2, r_4\}$ that is used to approximate $\mathbb{L}_C$ in (7) as expressed by (8).

$$\mathbb{L}_C = \Delta \bar{E}_{ab}^* = \frac{1}{i} \sum_{q=1}^{i} \frac{\sqrt{(L_q^* - L_p^*)^2 + (a_q^* - a_p^*)^2 + (b_q^* - b_p^*)^2}}{\Delta E_{ab\_max}^*} \tag{8}$$

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_p^*, a_p^*, b_p^* specify the color of the object of concern, i.e. object nr. 1 in this example; the variables L_q^*, a_q^*, b_q^* 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, $\mathbb{L}_C$ is given by

$$\mathbb{L}_C = \frac{1}{2} \frac{\sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} + \sqrt{(L_3^* - L_1^*)^2 + (a_3^* - a_1^*)^2 + (b_3^* - b_1^*)^2}}{\Delta E_{ab\_max}^*}$$

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 θ_i, 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_1 and O_2. 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]

$$\mathbb{L}_j = \mathbb{L}_1(\theta_1)\,\mathbb{L}_2(\theta_2) = e^{-\frac{\theta_1^2}{2\sigma_j^2}(O_1-1)^2} e^{-\frac{\theta_2^2}{2\sigma_j^2}(O_2-1)^2} \tag{9}$$

where σ_j is a constant maximizing satisfaction of the consistency condition of possibility theory. For the two-input case $\sigma_j = 0.299$. The likelihood parameters θ_1 and θ_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

$$\theta_1 = \frac{1-O_1}{(1-O_1)+(1-O_2)}, \quad \theta_2 = \frac{1-O_2}{(1-O_1)+(1-O_2)}, \tag{10}$$

so that (9) becomes

$$\mathbb{L}_j = e^{-\frac{1}{2\sigma_j^2}\left(\frac{1-O_1}{(1-O_1)+(1-O_2)}\right)^2 (O_1-1)^2} e^{-\frac{1}{2\sigma_j^2}\left(\frac{1-O_2}{(1-O_1)+(1-O_2)}\right)^2 (O_2-1)^2} \tag{11}$$

The output neuron of a fuzzy neural tree is termed as *root node* and denoted by $\mathbb{O}$. The inner nodes providing the input to the root node are instances of $\mathbb{L}_j$ in (9). They are termed as *penultimate* nodes and denoted by $\mathbb{L}_k$. 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.

$$\mathbb{O} = \sum_{k=1}^{n} w_k \mathbb{L}_k, \quad \sum_{k=1}^{n} w_k = 1 \tag{12}$$

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 $(w_1', w_2', \dots, w_n')$ is the one that is aligned to the feature vector $(\mathbb{L}_1, \mathbb{L}_2, \dots, \mathbb{L}_n)$. 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 $w_k' = c\mathbb{L}_k$, $\forall k \in \{1, 2, \dots, n\}$ 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 $\boldsymbol{w}'$ sum up to unity as stipulated in (12). In this case the root node output is given by [27]

$$\mathbb{O} = \sum_{k=1}^{n} \mathbb{L}_k^{\,2} \Big/ \sum_{k=1}^{n} \mathbb{L}_k \tag{13}$$

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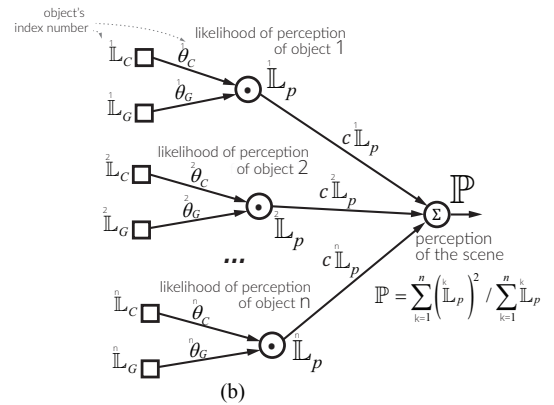

(b)

Fig. 2.   Fuzzy neural tree modeling visual perception of a scene

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-

ron are assumed to be independent of each other, (9) applies. The output of an inner node expresses the likelihood a scene object is perceived. This quantity is denoted by $\mathbb{L}_p$ and obtained by

$$\mathbb{L}_p = \mathbb{L}_1(\theta_c)\,\mathbb{L}_2(\theta_G) = e^{-\frac{\theta_1^2}{2\sigma_j^2}(\mathbb{L}_C-1)^2}\,e^{-\frac{\theta_2^2}{2\sigma_j^2}(\mathbb{L}_G-1)^2} \tag{14}$$

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

$$\theta_C = \frac{1-\mathrm{L}c}{(1-\mathrm{L}c)+(1-\mathrm{L}_G)} \tag{15}$$

$$\theta_G = \frac{1-\mathrm{L}_G}{(1-\mathrm{L}_C)+(1-\mathrm{L}_G)} \tag{16}$$

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

$$\begin{aligned}\mathbb{L}_p &= \mathbb{L}_1(\theta_C)\,\mathbb{L}_2(\theta_G)\\ &= e^{-\frac{1}{2\sigma^2}\left[\frac{1-\mathbb{L}_C}{(1-\mathbb{L}_C)+(1-\mathbb{L}_G)}\right]^2(\mathbb{L}_C-1)^2}\\ & \quad e^{-\frac{1}{2\sigma^2}\left[\frac{1-\mathbb{L}_G}{(1-\mathbb{L}_C)+(1-\mathbb{L}_G)}\right]^2(\mathbb{L}_G-1)^2}\end{aligned} \tag{17}$$

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

$$\mathbb{P} = \sum_{k=1}^{n}\left(\overset{k}{\mathbb{L}}_p\right)^2 \Big/ \sum_{k=1}^{n}\overset{k}{\mathbb{L}}_p \tag{18}$$

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].

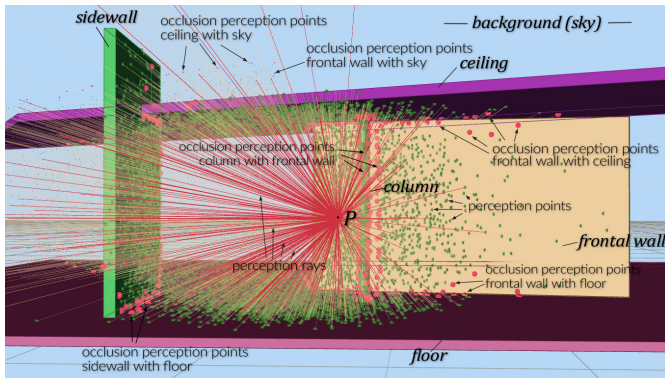An exemplary scene subject to perception computation is shown in figure 3.

Fig. 3.   An exemplary scene subject to perception computation

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.
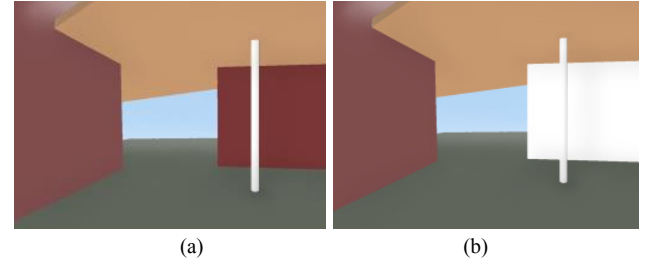

| (a) | (b) |

Fig. 4.   Two scenes subject to perceptions; scene nr. 1, where $\mathbb{P}=0.149$ (a); scene nr. 2, where $\mathbb{P}=0.142$ (b)

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

TABLE I
PERCEPTION OF THE SCENE SHOWN IN FIGURE 4A

object	$L^*a^*b^*$	$\mathbb{L}_G$	$\mathbb{L}_C$	$\mathbb{L}_P$
frontal wall	{0.48, 0.21, 0.21}	.19	.23	.17
side wall	{0.50, 0.30, 0.30}	.09	.16	.12
ceiling	{0.78, 0.60, 0.44}	.18	.18	.15
floor	{0.37, 0.40, 0.36}	.27	.16	.17
column	{1.0, 1.0, 1.0}	.03	.28	.10

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.
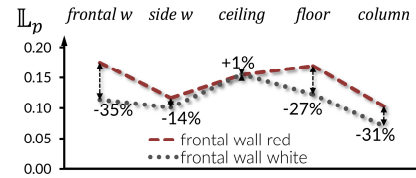

Fig. 5.   Differences in object perceptions for the scenes in figure 4a and 4b

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 aesthetical 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 intensely 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 scene is. The aesthetics of a man-made environment can be appreciated 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. When 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 accidentally 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 possible compositions, requiring significant conscious effort to resolve the conflict inherent to the objectives of high perception and chromaticity parsimony. Accordingly, scenes representing 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 $L^* = L_r^*$; $a^* = 0, b^* = 0$. The minimal difference is considered in two senses. In one sense, the difference is with respect to the lightness dimension L^*. In the other sense the difference concerns *CIELAB chroma* denoted by C_{ab}^* and obtained by [20]

$$C_{ab}^* = \sqrt{a^2 + b^2} \tag{19}$$

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 reference color in the chromatic sense. The likelihood of chroma absence is denoted by $\mathbb{L}_{C_{ab}^*}$ and given by

$$\mathbb{L}_{C_{ab}^*} = 1 - \frac{C_{ab}^* - C_{ab\_min}^*}{C_{ab\_max}^* - C_{ab\_min}^*} = 1 - \frac{\sqrt{(a^{*2}-0^2)+(b^{*2}-0^2)}}{\sqrt{(a_{max}^{*2}-0^2)+(b_{max}^{*2}-0^2)}} = 1 - \frac{\sqrt{a^{*2}+b^{*2}}}{\sqrt{a_{max}^{*2}+b_{max}^{*2}}} \tag{20}$$

In case one would consider chroma for all visible colors, then $C_{ab\_max}^* \approx 181$. In the ensuing computer experiments in this work $C_{ab\_max}^* = \sqrt{a_{max}^{*2}+a_{max}^{*2}} = 134$, which is due to the restricted gamut formed by the colors that can be displayed on a standard computer monitor.

Second, we consider the lightness aspect of the color parsimony. The likelihood, the observer perceived an object's lightness to be the same as the scene's reference lightness, is denoted by $\mathbb{L}_{L^*}$ and given by

$$\mathbb{L}_{L^*} = 1 - \frac{|L_r^* - L^*|}{L_{max}^* - L_{min}^*} = 1 - \frac{|L_r^* - L^*|}{L_{max}^*} \tag{21}$$

Based on the above considerations the fuzzy neural tree for color parsimony is shown in figure 6. Each inner node is associated 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

$$\theta_{C_{ab}^*} = \frac{1 - \mathbb{L}_{C_{ab}^*}}{\left(1 - \mathbb{L}_{C_{ab}^*}\right) + \left(1 - \mathbb{L}_{L^*}\right)} \tag{22}$$
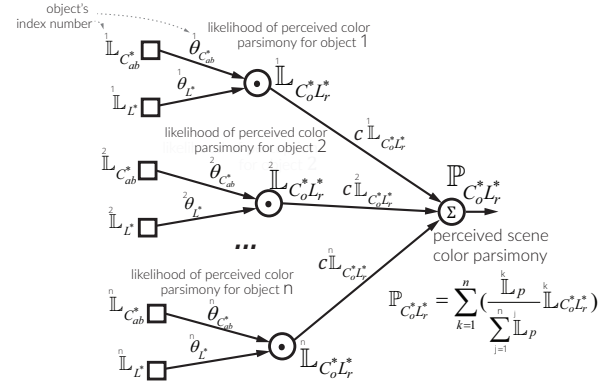

Fig. 6.   Fuzzy neural tree modeling perception of a scene's color parsimony

$$\theta_{L^*} = \frac{1 - \mathbb{L}_{L^*}}{\left(1 - \mathbb{L}_{C_{ab}^*}\right) + \left(1 - \mathbb{L}_{L^*}\right)} \tag{23}$$

The output of an inner node of the FNT represents the likelihood the associated scene object has the same achromatic character and lightness level as the reference color of the scene. We term this quantity as the *likelihood of color parsimony* and denote it by $\mathbb{L}_{C_o^* L_r^*}$. Due to (11) it is obtained by

$$\begin{aligned}
\mathbb{L}_{C_o^* L_r^*} &= \mathbb{L}_1\left(\theta_{C_{ab}^*}\right) \mathbb{L}_2\left(\theta_{L^*}\right) = e^{-\frac{1}{2\sigma^2}\left[\frac{1-\mathbb{L}_{C_{ab}^*}}{(1-\mathbb{L}_{C_{ab}^*})+(1-\mathbb{L}_{L^*})}\right]^2 \left(\mathbb{L}_{C_{ab}^*}-1\right)^2} \\
&\quad -\frac{1}{2\sigma^2}\left[\frac{1-\mathbb{L}_{L^*}}{(1-\mathbb{L}_{C_{ab}^*})+(1-\mathbb{L}_{L^*})}\right]^2 \left(\mathbb{L}_{L^*}-1\right)^2 \\
&= e
\end{aligned} \tag{24}$$

The root node output of the color parsimony FNT models the perception of the color parsimony of the scene. It is denoted by $\mathbb{P}_{C_o^* L_r^*}$. Analog to (13), it is obtained via aligned defuzzification of the color parsimony likelihoods associated by every object

$$\mathbb{P}_{C_o^* L_r^*} = \sum_{k=1}^{n} w_k \mathbb{L}_{C_o^* L_r^*}^k = \sum_{k=1}^{n} \left(\frac{\mathbb{L}_p^k}{\sum_{j=1}^n \mathbb{L}_p^j} \mathbb{L}_{C_o^* L_r^*}^k\right) \tag{25}$$

In (25) the parameter n denotes the number of scene objects. In the equation the vector $\boldsymbol{w}$ is not aligned to the vector $\left(\mathbb{L}_{C_o^* L_r^* \_1}, \mathbb{L}_{C_o^* L_r^* \_2}, \dots, \mathbb{L}_{C_o^* L_r^* \_n}\right)$ consisting of the objects' color parsimony likelihoods from (24). The alignment is to the vector $\left(\mathbb{L}_{p\_1}, \mathbb{L}_{p\_2}, \dots, \mathbb{L}_{p\_n}\right)$ consisting of the objects' perception likelihoods from (17). This is done so that an objects' 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 $\mathbb{L}_{C_{ab}}$ for smaller values of a^* and b^*, contributing to a greater value of $\mathbb{P}_{C_o^* L_r^*}$ in (25) in this case. Low values of a^* and b^*, however, generally imply a low degree of color difference ΔE_{ab}^* in (3). This contributes to lower object perception $\mathbb{L}_p$ in (17) and hence lower scene perception $\mathbb{P}$ in (18). An architectural scene is an aesthetical 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 aesthetical 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-domination 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 $\eta c = 10$, mutation probability 0.05, and polynomial mutation parameter $\eta m = 30$. 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 solu-

tions 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.
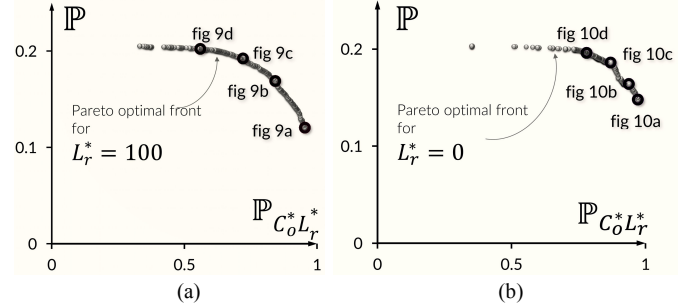

Fig. 7. Pareto front of aesthetical color compositions; **beautiful** compositions, i.e. $L_r^* = 100$ (a); **sublime** color compositions, i.e. $L_r^* = 0$ (b)
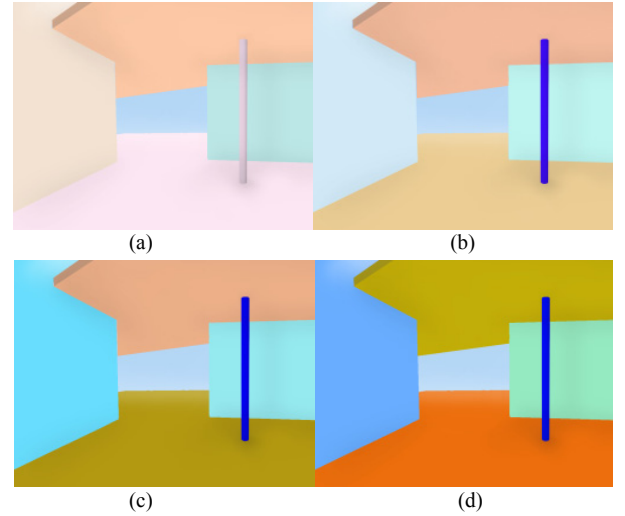

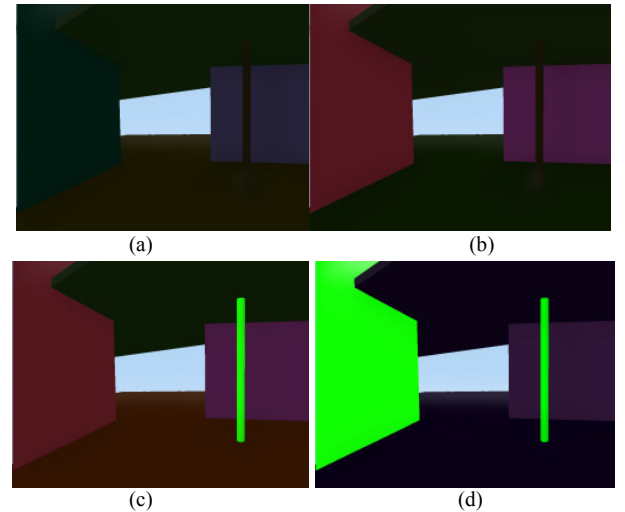Fig. 8. The four **beautiful** color compositions highlighted in figure 7a

Fig. 9. The four **sublime** color compositions highlighted in figure 7b

CONCLUSIONS

Computer-based visual perception and aesthetical 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 aesthetical ones in this work. Beautiful and **sublime** color combinations are identified as subsets of the aesthetical ones, namely particular choice of achromatic reference colors. The definition of aesthetical designs as non-dominated solutions in a two-dimensional objective function space imply that theoretically there are infinitely many beautiful as well as **sublime** color compositions for a given scene geometry. Selection 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 aesthetical 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 aesthetical judgment during a design process is a significant step. In this work color aesthetics is placed on a computational ground, reducing the imprecision in conventional aesthetical 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

[1] A. Amos, "A computational model of information processing in the frontal cortex and basal ganglia," Journal of Cognitive Neurosciences, vol. 12, pp. 505-519, 2000.

[2] L. Hemmen van, J. Cowan, and E. Domany, Models of neural networks iv: Early vision and attention. New York: Springer, 2001.

[3] D. Marr, Vision. San Francisco: Freeman, 1982.

[4] T. A. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature, vol. 317, pp. 314-319, 1985.

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, pp. 1254-1259, 1998.

[6] J. Bigun, Vision with direction: Springer Verlag, 2006.

[7] Y. Yu, J. Gu, G. K. I. Mann, and R. G. Gosine, "Development and evaluation of object-based visual attention for automatic perception of robots " Automation Science and Engineering, IEEE Transactions on vol. 10 pp. 365 - 379, 2013

[8] L. Xing, X. Zhang, C. C. L. Wang, and K.-C. Hui, "Highly parallel algorithms for visual-perception-guided surface remeshing," Computer Graphics and Applications, IEEE vol. 34 pp. 52 - 64, 2014

[9] T. F. Chan and L. A. Vese, "Active contours without edges," IEEE Trans. on Image Processing vol. 10, pp. 266 - 277, 2001.

[10] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, pp. 898 - 916, 2011.

[11] A. Agarwal and B. Triggs, "Recovering 3d human pose from monocular images " IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 44 - 58, 2006.

[12] A. Saxena, M. Sun, and A. Y. Ng, "Make3d: Learning 3d scene structure from a single still image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 824 - 840, 2009.

[13] D. C. Knill, D. Kersten, and P. Mamassian, "Implications of a bayesian formulation for visual information for processing for psychophysics," in Perception as bayesian inference, ed Cambridge: Cambridge, 2008, pp. 239-286.

[14] M. S. Bittermann and O. Ciftcioglu, "Architectural design computing supported by multi-objective optimization," Proc. IEEE Congress Evolutionary Computation - CEC 2015, Sendai, Japan, 2015.

[15] M. S. Bittermann, I. S. Sariyildiz, and Ö. Ciftcioglu, "Visual perception in design and robotics," Integrated Computer-Aided Engineering, vol. 14, pp. 73-91, 2007.

[16] W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colours," Transactions of the Optical Society vol. 30, pp. 141-164, 1928.

[17] J. Guild, "The colorimetric properties of the spectrum," Philosophical Transactions of the Royal Society of London, vol. A230, pp. 149-187, 1931.

[18] H. Grassmann, "Zur theorie der farbmischung," Poggendorfs Annalen der Physik, vol. 89, p. 69, 1853.

[19] G. Wyszecki and W. S. Stiles, Color science, 2nd ed. New York: Wiley, 1982.

[20] C. I. d. l'Eclairage, "Joint iso/cie standard: Colorimetry — part 4: Cie 1976 l*a*b* colour space ", ed. Vienna, Austria: CIE Central Bureau, 2007.

[21] C. I. d. l'Eclairage, "Iso 11664-3:2012(e)/cie s 014-3/e:2011: Joint iso/cie standard: Colorimetry - part 3: Cie tristimulus values ", ed. Vienna, Austria: CIE Central Bureau, 2011.

[22] D. L. MacAdam, "Visual sensitivities to color differences in daylight," Journal of the Optical Society of America, vol. 32, pp. 247-274, 1942.

[23] C. I. d. l'Eclairage, "Joint iso/cie standard: Colorimetry — part 5: Cie 1976 l*u*v* colour space and u', v' uniform chromaticity scale diagram," ed. Vienna: CIE Bureau, 2009.

[24] C. I. d. l'Eclairage, "Iso/cie 11664-6:2014(e): Joint iso/cie standard: Colorimetry — part 6: Ciede2000 colour-difference formula ", ed. Vienna: CIE Bureau, 2014.

[25] Y. Pawitan, In all likelihood: Statistical modelling and inference using likelihood. New York: Clarendon Press Oxford, 2001.

[26] O. Ciftcioglu and M. S. Bittermann, "A fuzzy neural tree based on likelihood," Proc. 2015 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE 2015, Istanbul, Turkey, 2015.

[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.

[28] O. Ciftcioglu and M. S. Bittermann, "Computational cognitive color perception," Proc. IEEE World Congress on Computational Intelligence - WCCI 2016, Vancouver, Canada, 2016.

[29] I. Kant, Kritik der urteilskraft - analytik der ästhetischen urteilskraft. Darmstadt Wissenschaftliche Buchgesellschaft, 1983.

[30] T. W. Adorno, Aesthetic theory. London: Athlone Press, 1997.

[31] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: Nsga-ii," IEEE Transactions on Evolutionary Computation, vol. 6, pp. 182-197, 2000.

[32] I. E. Commission, "Iec 61966-2-1:1999," in Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB ed. Geneva, Switzerland: IEC, 1999, p. 51.

[33] C. I. d. l"Eclairage, "Iso 11664-2:2007(e)/cie s 014-2/e:2006: Joint iso/cie standard: Colorimetry — part 2: Cie standard illuminants for colorimetry ", ed. Vienna: CIE Bureau, 2007.

[34] C. I. d. l'Eclairage, "Iso 11664-1:2007(e)/cie s 014-1/e:2006: Joint iso/cie standard: Colorimetry — part 1: Cie standard colorimetric observers ", ed. Vienna: CIE Bureau, 2007.