User's Web Page Aesthetics Opinion: A Matter of Low-Level Image Descriptors Based on MPEG-7

SILVIA URIBE, FEDERICO ÁLVAREZ, and JOSÉ MANUEL MENÉNDEZ,

Analyzing a user's first impression of a Web site is essential for interface designers, as it is tightly related to their overall opinion of a site. In fact, this early evaluation affects user navigation behavior. Perceived usability and user interest (e.g., revisiting and recommending the site) are parameters influenced by first opinions. Thus, predicting the latter when creating a Web site is vital to ensure users' acceptance. In this regard, Web aesthetics is one of the most influential factors in this early perception. We propose the use of low-level image parameters for modeling Web aesthetics in an objective manner, which is an innovative research field. Our model, obtained by applying a stepwise multiple regression algorithm, infers a user's first impression by analyzing three different visual characteristics of Web site screenshots—texture, luminance, and color—which are directly derived from MPEG-7 descriptors. The results obtained over three wide Web site datasets (composed by 415, 42, and 6 Web sites, respectively) reveal a high correlation between low-level parameters and the users' evaluation, thus allowing a more precise and objective prediction of users' opinion than previous models that are based on other image characteristics with fewer predictors. Therefore, our model is meant to support a rapid assessment of Web sites in early stages of the design process to maximize the likelihood of the users' final approval.

CCS Concepts: • **Information systems→ Web interfaces**; • **Human-centered computing→ HCI design and evaluationmethods**; **User studies;**

Additional Key Words and Phrases: Aesthetics, Web site visual appeal, search engine evaluation, first impression, image descriptors, regression

1. INTRODUCTION

Since its first definition, the aesthetics of interactive systems has been increasingly important and is now regarded as a critical factor for user acceptance [Mahlke 2008a]. Thus, some dimensions of the user experience in Web site navigation, such as satisfaction, usability, and credibility, are affected by its aesthetic level [Robins and Holmes 2008; Hassenzahl and Monk 2010], which hence becomes a new core element for Web site evaluation. According to these works, we can claim that the correctness of a solution in terms of users' early satisfaction is more tightly related to its attractive visual appeal than to other characteristics, such as its innovation and functionality. Therefore, development efforts must be focused on the aesthetic design from the earliest stages of the process to obtain a solution that meets specific requirements related to its own presentation mode. By doing so, we are closer to obtaining an easy and intuitive solution [Tractinsky et al. 2000].

Some authors have focused on inferring user preferences in Web sites by analyzing different aesthetic dimensions, such as complexity, balance, or equilibrium of the interface [Ngo et al. 2003]. However, when it comes to how fast the aesthetic perception is formed at first sight, it seems that only a few milliseconds are needed [Jacobsen and Höfel 2002; Leder and Nadal 2014]. In this way, some authors propose that after viewing a Web site for the first time, it takes less than 500ms to make an aesthetic first impression [Lindgaard et al. 2006; Lindgaard et al. 2011; Tractinsky et al. 2006; Tuch et al. 2012a], although other features like the process of understanding the semantic content may take a little bit longer [Fernandes et al. 2003].

Within this short amount of time, the user is more capable of perceiving low-level information related to the color, texture, and luminance [Oliva and Torralba 2006; Rosenholtz et al. 2005] than of considering complexity layouts. Therefore, other rudimentary elements from the human visual system must be considered [Hubel and Wiesel 1979]. In this regard, the MPEG-7 [MPEG ISO/IEC 2000] standard defines and provides a set of descriptors for visual media [Salembier and Sikora 2002] that can be helpful to model and evaluate Web sites. Its definition began in 1998 and in 2001 became an international standard, being formally known as the interface for multimedia content description. Specifically, MPEG-7 is responsible for providing the necessary infrastructure for the description of multimedia content: on the one hand, keywords and semantic meaning, and on the other hand, structural information. Thus, content and format are independent of one another. Hence, this standard provides a set of media descriptors widely used for analyzing the structure of the media, as is done in this article.

In addition, common Web site evaluation methods are applied after obtaining the final product or at least once an advanced prototype version is ready. In the usability field, this type of assessment is called *summative evaluation*, which provides important information to solve operational problems but presents high costs, both money- and time-wise, because of the late application stage. In contrast, it is advisable that new models follow *formative evaluation* principles, which provide several benefits thanks to them being applied in early stages of the Web site generation process [Redish et al. 2002]. In fact, these techniques help developers optimize the solution by the following:

- *Minimizing engineering costs*: The earlier the designer detects a developmental mistake (related to usability, aesthetics, etc.), the cheaper it is to apply the change. Additionally, the costs of iterations to improve the Web site are lower the sooner they are carried out.
- *Saving users' money and time*: An optimal visually appealing solution leads to a more usable product, which means a more efficient tool to achieve users' goals in a simpler and faster way.

Under these premises, we present a new formative aesthetic evaluation model that comprises 10 different low-level image characteristics derived from the MPEG-7 analysis of Web site screenshots. We select these parameters because they have been considered in other content-based retrieval systems, such as image recommenders, with optimal results [Sánchez et al. 2012], without including any other type of information. In addition, considering that users are able to give an evaluation of a Web site's appeal in a very short amount of time, we assume that the opinion generation process is similar regardless of the static content the user is watching. Furthermore, following a formative evaluation approach, this model can be applied at an early development stage, helping designers to reduce the number of iterations and thus the total cost of the solution.

Finally, it is important to mention that the metrics of the model have been selected so that they can be applied by nonaesthetics experts, allowing its use even in small organizations whose staff has no knowledge of visual appeal assessment methods.

This article is outlined as follows. Section 2 presents the theoretical background regarding the evaluation of Web site aesthetics and the main attributes for assessing their visual appeal. The aims of our research work are established in Section 3. Section 4 explains our new aesthetic model based on low-level image characteristics. Section 5 analyzes the obtained results and compares them to previous studies in this field, as well as its application to two different datasets, one focused on Web search engines. Our main findings are presented in Section 6, and limitations to our solution and future work are provided in Section 7. In Section 8, we offer our concluding remarks.

2. THEORETICAL BACKGROUND

2.1. Visual Appeal as a Recurrent Research Issue: The Analysis of Aesthetics in the Web Site Environment

The wide importance of aesthetics in older disciplines like philosophy, science, and arts has supported its use in the human-computer interaction (HCI) field, providing a novel research area with many possibilities. Although the term *aesthetics* is usually replaced with synonymous terms such as *beauty*, *visual appeal*, *attractiveness*, and *pleasure*, its meaning is the same and is defined "as the immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning" [Moshagen and Thielsch 2010]. Moreover, the importance of aesthetics has led to the definition of a new scientific and engineering discipline called *engineering aesthetics* [Lui 2003], which provides a new methodology for considering the different aspects of visual appeal in the implementation process based on two related research lines:

- The quantitative analysis of the critical dimensions involved in a specific aesthetic response process
- The study about how changes in these dimensions affect the aesthetics perception.

Our work contributes not only to the first goal by establishing a set of characteristics to evaluate Web sites objectively but also to the second by showing how to obtain an efficient tool to optimize Web site design.

In contrast to previous studies, which focused on the analysis of typical usability elements like efficiency and effectiveness of interactions [Nielsen 1993; Shackel 2009], recent efforts have looked into other alternative dimensions that affect user experience, such as pleasurability and psychological emotion. As can be seen in Figure 1, and according to the model defined in Thüring and Mahlke [2007] and Mahlke [2008b], the interaction between the system and the user includes two different perceptions, one related to the instrumental qualities of the product and the other to the noninstrumental qualities, where the aesthetics of the systems represent a very important factor to take into account. User experience starts with the interaction between a user and a product in a specific context of use, which mainly contains the time-spatial characteristics of the interaction. This experience represents a complex process that comprises different elements related not only to the system itself (e.g., its aim, its functionality, and the way the content is shown in the interface) but also to the users (who are influenced by factors such as experiences and expectations). According to this definition, visual appeal is also included in the definition of the system, as content presentation is one of the four main dimensions that characterize the product.

Although some authors focused on importing visual principles from visual design to user interfaces long ago [Vanderdonckt and Gillo 1994; Mullet and Sano 1996; Galitz 2007], the beginning of the analysis of Web aesthetics can be dated back to the study

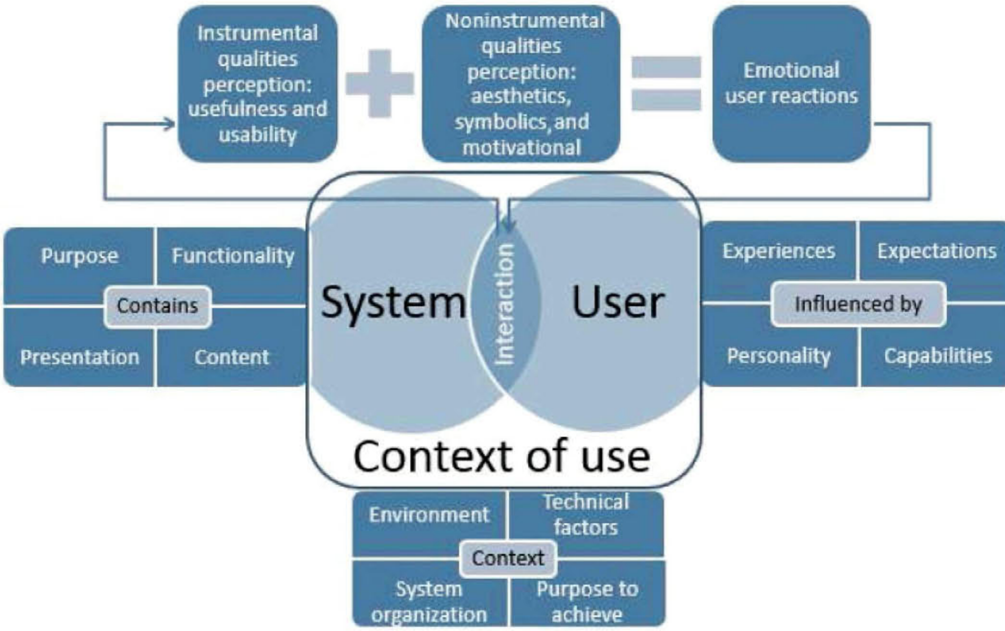

Fig. 1. User experience framework and components.

by Tractinsky et al. [2000], who presents empirical evidence of interface aesthetics mainly based on visual principles. Moreover, this author also connected visual appeal and subjective usability evaluation by extending some previous works [Kurosu and Kashimura 1995; Tractinsky 1997]. The most important conclusion of his work is the fact that beautiful designs are usually considered as usable solutions. On the contrary, Tuch et al. [2012b] found the reverse relationship under specific conditions, explaining that it is the usability that has an effect on postuse perceived aesthetics. Either way, the connection between usability and aesthetics is a wide field of study.

Continuing this trend, the next step focused on establishing the main dimensions influencing users' aesthetics perception to model it. In this regard, Lavie and Tractinsky [2004] define two main elements:

- *Classical aesthetics* are related to an organized and clear design, which is close to the main usability and interface design rules. This is quite relevant because improving the classic aesthetic level may revert to obtain a higher perceived usability level for the user.
- *Expressive aesthetics* are linked to the creativity and originality of the solution.

Once the concept was defined for the HCI field, several studies focused on evaluating the influence of aesthetics on users' perception and interaction behavior, confirming its importance in the design and implementation of optimal solutions. The relationship between visual appeal and perceived usability has been widely analyzed by Hassenzahl [2008] and Hartmann et al. [2008]. Following these works, recent studies have evolved to establish that Web site aesthetics are closely linked to other dimensions of users' interaction and perception about a site, such as the evaluation of its content [Palmer 2002; De Angeli et al. 2006; Nebeling et al. 2011; Thielsch et al. 2014], its trustworthiness [Cyr et al. 2005], its capability of improving performance [Moshagen et al. 2009;

Sonderegger and Sauer 2010], and even the customer's loyalty and intention to revisit the site [Mahlke 2002; Cober et al. 2003; Cyr et al. 2006]. From these relationships, we can infer that helping to improve the aesthetic general level of a Web site may also help to improve other relevant features of the Web site environment

That said, one of the most common mistakes when talking about user experience is the comparison between usability and aesthetics, as both terms are focused on improving the interaction between the user and the device. However, as shown in Figure 1 and following the guidelines set by various authors [Thüring and Mahlke 2007; Roto et al. 2011], the correct option is to say that usability is one of the pillars on which a good user experience lies, as are other aspects like aesthetics, design, functionality, and interaction. In fact, and following this line of thought, Hassenzhal and Tractinsky [2006] ensure that the main challenge for HCI is to design for pleasure rather than for lack of difficulty, which comes to justify further differentiation between the two aspects, although they are interlinked dimensions of the user experience.

Finally, another important issue to study in this field is related to the generation of users' opinion looking at the given stimuli, as assessment prediction must consider every aspect of the evaluation process. As mentioned previously, some researchers emphasized the fact that first impressions about appeal, usability, and trust are usually obtained in a few milliseconds [Lindgaard et al. 2006; Papachristos and Avouris 2011], supporting the idea that a good first impression is vital for users to accept the site. We consider this result to be an essential characteristic of opinion generation, because in this short amount of time, the user is only capable of evaluating a static image of the stimuli. This leads us to propose the premise of our model, which is based on applying an evaluation methodology for static Web site aesthetics looking at image characteristics and low-level parameters similar to those used in other research fields such as image recommendation [Barrilero et al. 2011].

2.2. Psychobiological Model for Visual Perception and Aesthetic Low-Level Parameters

As in the previous case, visual perception also refers to a complex concept with different definitions depending on the field of study considered. From a biological point of view, it can be defined as the active process that makes the transformation from the luminance information to an object recreation. From a computer vision approach, it is more related to the ability to interpret the light information of an external stimulus to provide a specific output. The most influential model in psychological aesthetics so far is Berlyne's collative-motivational model, which evolved from 1950 to 1971 [Berlyne 1970]. This model was built over the influence of two general concepts, namely hedonic tone and arousal, and it determines the main variables that activate the perception process. Based on this work, Mankeliunas [1980] defines three basic models for visual perception analysis—physical, physiological, and computational—depending on the involved science. More recently, several research works have defined other models for aesthetics perception, like the one proposed in Leder et al. [2004] and reviewed in Leder and Nadal [2014], where the authors present a complex framework for modeling aesthetic experience from a psychological point of view, or Silvia's appraisal-of-interest model [Silvia 2006], which focuses on assessing the differences in aesthetics evaluation according to the user's interest.

Regardless of the definition taken into account, the process performed by the human eye is the same and can be defined by two main dimensions [Xie 2003]:

- The elements' appeal, characterized by three main factors: color (chrominance), light (luminance), and texture information
- The elements' geometry, composed of postural, movement, and shape factors.

As such, the use of low-level parameters related to the chrominance, luminance, and texture information to obtain a final model for Web site visual appeal represents a new approach that is consistent with the factors involved in the visual perception model.

2.3. Attributes for Evaluating Web Page Aesthetics in Previous Studies

To address the lack of a standard methodology for aesthetics analysis in HCI since its first definition [Norman 2004], current efforts are especially focused on the development of specific methods, guidelines, and tools for breathing life into the so-called aesthetics engineering. For this reason, one of the most important objectives is to obtain an evaluation method that can be applicable without the need for special equipment or software (as opposed to commonly used solutions in the usability analysis context, which use eye-tracking tools or expensive image analysis solutions, etc.), while at the same time providing aesthetics measures as soon as possible so as to not increase the application costs.

Today, aesthetics evaluation methods can be classified into two main groups (objective and subjective) depending on the way assessment results are obtained. Objective methods are based on the analysis of the rules of visual appeal derived from empirical studies. Subjective methods make use of users' opinion and feedback. According to this classification, in this work we focus on objective methods, as they provide faster and more efficient evaluation methods in terms of costs and time. Nevertheless, advanced studies have confirmed the correlation between the results of both methods [Altaboli and Lin 2011; Donyaee et al. 2006], helping researchers to infer one from the other.

Based on the study provided by Pajusalu [2012] and extended with new models, Table I presents the main objective evaluations methods in the literature.

As can be seen, aesthetics in HCI is an interesting research area with different problems that remain to be solved. In this regard, two of the biggest issues are related to the human supervision of the generation process and the use of a high number of descriptors to model the interface, which increases the associated costs of the implementation. Therefore, our method improves existing aesthetics research by reducing these problems. To that end, we apply a set of MPEG-7 descriptors in an objective way without considering any other kind of information, such as the commonly used objective aesthetics measures in Ngo et al. [2000] or Nebeling et al. [2011], as a big effort must be made to obtain them. Previous works focus on content analysis of the Web site layout, but our model is based on the early perception process evaluation, which is more tightly related to the basic dimensions described in the preceding section.

Finally, with regard to the number of parameters, our method involves a smaller set of descriptors than some previous works [Reinecke et al. 2013], with a higher level of correlation to users' opinion, making it a better approach.

3. THE NEED FOR A NEW AESTHETICS EVALUATION MODEL: MAIN OBJECTIVES OF OUR RESEARCH WORK

According to the application environment and the improvement possibilities identified by previous works, we have defined a new model for evolving the analysis of Web site aesthetics to obtain a user's opinion prediction system based on a set of some MPEG-7 low-level image parameters. This solution may help to characterize the visual appeal of Web sites in a manner different from previous models, providing developers with innovative options to improve their interfaces. The main contributions of this research are related to four different aspects:

- *Identification of a new set of parameters for modeling Web sites*: Based on the MPG-7 standard, we identify 14 descriptors related to the texture, color, and luminance information of Web pages to characterize their visual appeal. These parameters

Table I. Main Aesthetics Objective Evaluation Methods

Characteristics Method	Elements of Evaluation	Tools	Advantages	Disadvantages
Application of visual techniques to traditional and multimedia layouts [Vanderdonckt and Gillo 1994]	Based on application of five sets of visual techniques to optimize interface design	No special tool; techniques can be classified into five groups: physical, composition, association, ordering, and photographic	Uses commonly accepted visual principles to suggest the arrangement of layout components; presents some guidelines for layout design	Does not provide automatic evaluation method
Application of design rules and techniques to graphical user interfaces [Mullet and Sano 1996]	Six major areas: elegance and simplicity; scale, contrast and proportion; organization and visual structure; module and program; image and representation; style	No special tool; proposes to apply techniques from design aesthetics in modern graphic design, industrial design, interior design and architecture	Includes catalog of common errors from existing GUI interfaces	Does not provide automatic evaluation method
Layout appropriateness (LA) [Sears 1993]	Based on analysis of widgets used in the interface, sequence of actions users perform, and how frequently each sequence is used	An algorithm for determining layout appropriateness and computational cost	Proposes specific layouts according to LA-optimization	Is only based on widget position and task costs
Objective aesthetics measures for graphic screens [Ngo et al. 2000; Ngo et al. 2003]	14 elements: balance, equilibrium, symmetry, sequence and order, complexity, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm	Some partial development of several elements, like aesthetics measurement application tool [Zain 2008]	Good definition of parameters; comparable score among different screens	Shortage of complete software implementing the measurements (e.g., Lok et al. [2004] and Zen and Vanderdonckt [2014]); complicated screen division to diagram elements; several parameters, with no establishment of the most important ones
Objective measures for Web pages [Purchase et al. 2011], as an adaptation of the previous model	Same 14 elements as in preceding entry; color also considered	Specific software that analyzes HTML code of Websites	HTML code clearly and unambiguously represents each visual element	Inability to analyze components within other components

(Continued)

Table I. Continued

Characteristics Method	Elements of Evaluation	Tools	Advantages	Disadvantages
Objective measures for the complexity of Web pages [Fu et al. 2007]	Four components: size complexity, local density, measure of group, and measure of alignment	No special tool	Smaller set of measurements	Complicated screen division to diagram elements; does not consider other dimensions, like color; does not determine the exact weight for each measure according to users' opinion
Aesthetic coloring system [Zhang et al. 2009]	Colors of layouts areas	Specific software	Offers selection of aesthetic combinations to specific layout	Does not consider other aesthetics dimensions
Low-level image statistics [Zheng et al. 2009]	Based on three steps: computation and discretization of low-level statistics (color, intensity, and texture), decomposition into regions of minimum entropy, and evaluation on aesthetics dimensions (symmetry, balance, equilibrium, and number of decomposition regions)	Mathematical calculations	No need to divide screen into elements before evaluation; considers different dimensions of Web site aesthetics	Complicated quadtree decomposition; depends on Ngo et al. [2000] definition of aesthetics measurements
Counts-based measure [Altaboli and Lin 2011]	Number of different sizes of visual objects, number of images, number of different font types, JPG file size of screenshot	No special tool	Simple method; well suited for informative and task-oriented sites	Can be time consuming with complex design
Physiological measurements [Strebe 2011]	Reactions based on eye movement, breathing, heart rate, skin conductance, etc.	Specific equipment	Actual users' reactions	Specific technology and field studies with users
Content ratios [Nebeling et al. 2011]	Document-window ratio, content-window ratio, wide text ratio, small text ratio, visible text ratio, visible links ratio, media-content ratio	No special tool	Simple method; considers different viewing contexts	Does not consider color; depends on content elements
Complexity and colorfulness model [Reinecke et al. 2013]	A complexity model based on content information and a colorfulness model	Specific software	Considers different dimensions of aesthetics; elements can be obtained with no human supervision	Contains many elements and thus can be complex and time consuming

provide a new analysis environment, differing from the ones used in previous research works that are mainly focused on the complexity and colorfulness of the interface.

- *Definition of a new prediction model to infer users' opinion of Web site aesthetics*: This model comprises a set of only 10 parameters from the 14 initially defined. They have been selected according to their correlation to users' evaluation, obtained from a regression process between a dataset of users' opinion of Web site appeal [Reinecke and Gajos 2014] and the selected low-level parameters. Hence, the obtained model is more accurate than those of previous works [Reinecke et al. 2013] using a smaller set of descriptors.
- *Construction of a new dataset on visual appeal focused on search engines*: In the dataset, 120 users evaluated different aesthetic characteristics of six different Web sites.
- *Application of this model to two different datasets to validate it and then analyze different appeal dimensions*:
  - In the dataset provided in Moshagen and Thielsch [2010], participants had to evaluate the Web aesthetics from two points of view: classic and expressive. Thanks to this, we can ensure that this new model infers the expressive level of the image more accurately than the classic one.
  - In the new search engine dataset, users were asked to assess the expressive and general level of aesthetics of the Web sites. In this case, the results confirm that for this environment, the proposed model represents a better approach to infer the general aesthetic level rather than the expressive one.

In the following section, we present the obtained model, describing the low-level characteristics considered, the used datasets, and the process for generating the model and its validation.

4. A NEW AESTHETIC MODEL BASED ON LOW-LEVEL PARAMETERS DERIVED FROM MPEG-7

4.1. Low-Level Characteristics

Our model is strictly based on the analysis of low-level image characteristics related to the main MPEG-7 parameters without considering any other Web site information. This standard's descriptors can be classified into five different groups (color, texture, shape, motions, and others), but since we are considering static images, our model focuses on the color and texture parameters, removing the ones referred to motion (we are not considering temporal changes) and to shape (we are considering only low-level values). Moreover, in our model, these parameters have been obtained by means of the MPEG-7 low-level feature extraction command line tool provided by Bastan et al. [2009], and they are related to three different aspects: luminance, texture, and chrominance. The final characteristics vector is composed by 14 different parameters:

$$\text{LowLevel\_Vector} = (L_m, L_v, E1, E2, E3, E4, N, VC_{intra}, VC_{inter}, S_m, S_v, SC, L_e, L_h), \quad (1)$$

and their definition is explained in Table II [Sánchez et al. 2012].

4.2. Datasets

To confirm the validity of our model, we have to test it using a large dataset that considers a wide range of different Web pages evaluated by many users. In this regard, we make use of the dataset provided by Reinecke and Gajos [2014], composed of the rating of 430 Web sites' screenshots by almost 40,000 users on a 1 to 9 Likert scale, as well as the results of their visual evaluation for each Web site according to their model. With these values, this dataset meets the condition of the minimum sample size

Table II. Initial List of the Model's Visual Descriptors

Visual Dimension	Descriptor Name	Description	Parameter
Luminance	Direct luminance	- It is derived from the MPEG-7 color layout descriptor [Salembier and Sikora 2002]. - It determines the image mean saturation to distinguish between dark and bright interfaces. - It is composed of two values: mean and variance luminance	L_m L_v
	Bit plane distribution entropy	- According to Shan and Hai-tao [2008], this descriptor determines the bit plane distribution entropy across the four highest bit planes of the image, where the majority of the structural information is contained.	E1 E2 E3 E4
Color	Chromatic variety	- It analyzes the color variety along the image by detecting the dominant colors and is composed of three values: the number of dominant colors, the variance among different tonalities of the same color, and the variance among different colors.	N VC_{intra} VC_{inter}
	HSV color space	- It analyzes the image saturation mean and variance in the HSV color space.	Sm Sv
	Spatial coherence	- This descriptor, extracted from the MPEG-7 dominant color descriptor [Salembier and Sikora 2002], analyzes the color continuity in the Web site image.	SC
Texture	Line energy	- It is extracted from the MPEG-7 edge histogram descriptor [Salembier and Sikora 2002]. - It measures the total line density in the image according to its energy level across different directions. - It helps us to distinguish images with no transitions from those with full edges.	L_e
	Line homogeneity	- It is also extracted from the MPEG-7 edge histogram descriptor [Salembier and Sikora 2002]. - It establishes the image line continuity by calculating the line variance distribution along 4×4 neighboring blocks.	L_h

needed to obtain a valid linear regression model based on the number of parameters referred [Stevens 1996; Tabachnick and Fidell 2007].

The Web sites contained in this dataset represent a wide range of genres with a large variety of image characteristics and are classified into three different groups according to their origin: 350 English Web sites, 60 foreign Web sites, and 20 sites nominated for Webby Awards [Webby Awards 2015] in recent years. Concerning the users' rating of the selected Web pages' visual appeal, ratings were collected over 1 year, and to guarantee evaluation consistency, users were asked to give their assessment in two different stages, removing those evaluations that present a high deviation. For our model, we consider the added value for each Web site and then the mean of the two given ratings. Once we selected the dataset to be used, we carried out a first data filtering stage to prepare the information. We removed 15 records where some information was missing (a model parameter, an evaluation, etc.), which only represents 3.4% of the dataset.

Moreover, for the validation stage, we used two additional datasets. The first was the generic dataset provided in Moshagen and Thielsch [2010], composed of the assessment

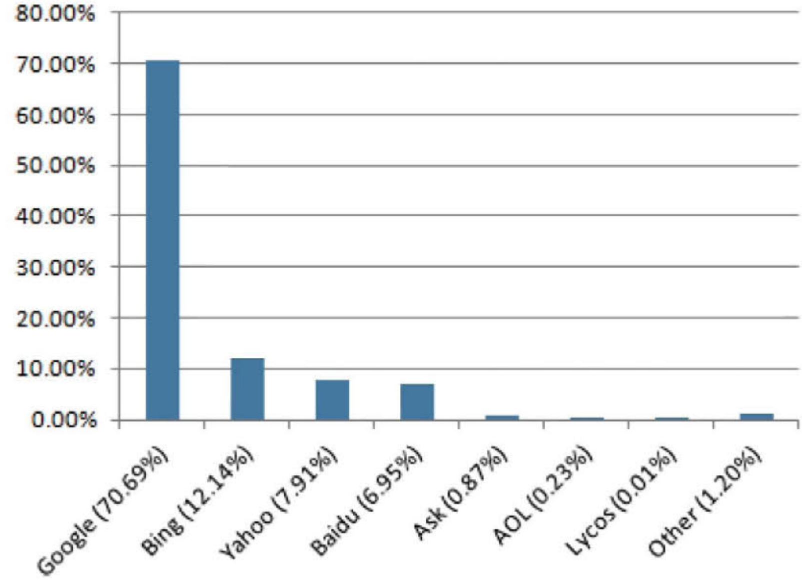

Fig. 2. Search engine market share. (From NetMarketshare (n.d.)

information of 42 Web sites by 512 users (67.8% female, participants between 15 and 82 years old (mean, 30.5; SD, 10.61)), rated on a 1 to 7 Likert scale. The second dataset is detailed in the following.

NEW DATASET ON SEARCH ENGINE VISUAL APPEAL

We designed an experiment to obtain this new dataset with two main goals in mind: to validate the proposed model with a different dataset than the initial one, and to determine whether the users' expectation and perception about aesthetics are the same in specific fields, such as the search engine in the general case. Moreover, the reason to select the search engine environment to make this experiment is its importance in the current multimedia Internet environment, based on the fact that almost 81% of Internet browses start from this kind of Web site [TNS 2008], as well as its associated impact in Internet economy [Buguin et al. 2011]. To enable replication and extensions of our results, the information about this dataset can be downloaded from http://www.gatv.ssr.upm.es/~sum/dataset.

Materials

Our stimuli consist of six different search engine Web site screenshots as a current representative set for Internet browsing. They were selected according to two different criteria:

- First, their selection was based on the market share (Figure 2). We have considered some of the most-used solutions without considering foreign languages to avoid cultural bias (all participants are mostly Spanish speakers).
- Second, we considered stimuli with a wide diversity of characteristics to analyze different design possibilities. For this reason, we only took into account the most different designs among the Web sites from that market set and then we also included two additional solutions which are very different from the previous ones.

Based on this, the selected search engines are presented in Table III.

Table III. Web Search Engines for the New Dataset

Name	Main Characteristics	Explanation of Selection
Google	- Contains textual information and images - High information density in layout - Limited number of fonts and background colors - Linear structure, 2 columns, simple	The information is presented as textual and visual data. In addition, users are familiar with this solution.
Yahoo	- Contains textual information and images - Low information density in layout - Limited number of fonts and background colors - Linear structure, 2 columns, simple	The information is presented as textual and visual data in a different way compared to the others.
Ask	- Only contains textual information - Low information density in layout - Limited number of fonts and background colors - Linear structure, 1 column, simple	Interface is very simple, with low information density and without graphical data.
Lycos	- Contains textual information and images - Low information density in layout - Limited number of fonts and background colors - Linear structure, 2 columns, simple	The information is presented as textual and visual data in a different way than the others.
Blackle	- Only contains textual information - Low information density in layout - Limited number of fonts and black background - Linear structure, 1 column, simple	It presents a black background, which is unusual in this field.
Hotbot	- Only contains textual information - Low information density in layout - Larger number of fonts and background colors - Linear structure, 2 column, simple	It presents a different color combination for the background.

Procedure

The survey was published on the Internet to allow users easy access. Moreover, it contains three different parts to obtain all necessary data from the users:

- *User profile*: The profile contains the main user's characteristics about age, gender, educational level, profession, and consumption habits.
- *Contextual information*: This part contains the main factors around which the user consumed the content and which can influence the process. It focuses on information about navigation devices used in general to access the Web.
- *Subjective opinion about Web search engine aesthetics*: This part contains 12 questions about each Web site, some derived from those defined in the VisAWI questionnaire [Moshagen and Thielsch 2010], and some referring to the multimedia component of the interface. We selected only a subset of the VisAWI questionnaire to avoid user overload. Indeed, our set includes the four main questions indicated in the short version of the questionnaire given by Moshagen and Thielsch [2013], due to its reliability when capturing a single dimension of perceived visual aesthetics.

As such, according Moshagen and Thielsch [2010], different components of this questionnaire are highly correlated to different dimensions of aesthetics. Particularly, simplicity, which comprises aspects related to unity, clarity, and balance, is high correlated to classic aesthetics, whereas diversity, which comprises dynamics, novelty, and creativity, is more related to the expressive dimension of Web site aesthetics. Therefore, we used them to analyze these different components of aesthetic impression.

Since the VisAWI-S cannot obtain enough information to characterize each specific aesthetic facet [Moshagen and Thielsch 2013], we also included two additional questions for these two main aspects—simplicity and diversity—derived from the full version of the VisAWI. Thus, we avoid overloading users and facilitate the process by eliminating only those questions that are negatively keyed items.

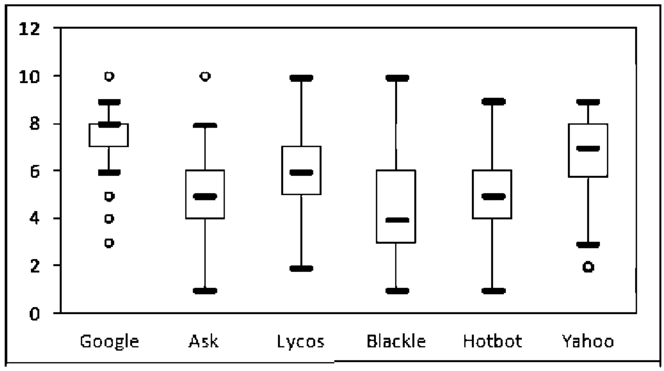

Fig. 3. Search engine aesthetic level evaluation.

Table IV. Runtimes for Each Image Obtaining Imaging Descriptors

Complete Execution	Tool Execution [Bastan et al. 2010]	Descriptors Extraction Execution	Results Output
5.9 sec	1 sec	3.47 sec	1.43 sec

Finally, users were asked to provide a specific rating for the aesthetic visual appeal of each Web site. The results for this question are shown in Figure 3.

Participants did not receive monetary compensation for being part of the experiment.

Users

We report on data collected for 2 months. During this time, 110 volunteers (54% male) completed the experiment. Participants were between 15 and 68 years old (mean, 34.5; SD, 10.9), and 86% had at least a college degree. Most participants (90%) made use of a computer to complete the survey, and almost 86% used Web search engines for Internet browsing.

4.3. Obtained Model

To establish our model, we first obtain a representation of the dataset Web sites in our new low-level characteristics dimension space. Thus, each Web site is represented by its corresponding low-level parameters vector, which contains its associated descriptors for texture, color, and luminance. This first step is implemented in Matlab and uses the command line tool described in Bastan et al. [2009], which provides the dominant color, color layout, and edge histogram MPEG-7 descriptors associated with each input by using a library of feature extraction. The said tool uses low-level C++, and its compilation is carried out with MinGW [MinGW 2015]. In terms of performance, the breakdown of runtimes per image, obtained on an Intel Core i7 processor at 3.40GHz with 8GB of RAM, is shown in Table IV.

Once this is accomplished, the next step is to determine the most important low level-characteristics to predict users' ratings. This is achieved by applying a stepwise multiple regression algorithm [Hintze 2007], which combines the forward and backward selection techniques to determine only the descriptors that are statistically significant for the model.

To avoid overfitting problems in the regression model extraction, we apply a 5-fold cross-validation process by dividing the entire dataset into five complementary subsets. Then we perform the regression process to the training set, created by merging four different subsets, and finally we carry out the validation to the remaining subset. We repeat this process until all subsets have been used as validation sets, then we select

Table V. Regression Model for Low-Level Parameters

		Estimated Coefficients			
Parameter	**b**	**SE**	β	**tStat**	p**-Value**
(Intercept)	4.746	.081	.202	2.493	.013
L_e	123.467	.091	–.182	–2.000	.046
L_m	2.413	.057	.012	.210	.823
E1	1.923	.087	.269	3.091	.002
E2	.327	.079	.192	2.430	.015
E3	.953	.062	.222	3.580	<.001
N	.916	.086	–.005	.058	.947
VC_{intra}	4.127	.120	.343	2.853	.004
VC_{inter}	–7.895	.073	–.600	–8.219	<.001
SC	–3.615	.133	.674	5.067	<.001
S_v	–19.754	.066	–.020	–.303	.760
L_e:N	–33.622	.058	–.294	–5.068	<.001
$L_{e:}$ S_v	967.185	.079	.253	3.202	.001
E2:E3	2.122	.073	.147	2.013	.046
E1:SC	–3.445	.075	–.208	–2.773	.006
E2: S_v	–16.307	.053	–.214	–4.037	<.001
VC_{inter}:SC	8.421	.055	.348	6.327	<.001
L_m ^2	–4.274	.052	–.164	–3.153	.002

Number of observations	332	R^2 adj.	0.37
Error degrees of freedom	318	F-statistic vs. constant model	12.5
RMSE	.79	p-value	<.001
R^2	.404		

E1, first bit plane entropy; E2, second bit plane entropy; E3, third bit plane entropy; E4, fourth bit plane entropy.

the model that provides the most accurate results for each test subset and apply it to the entire dataset.

Table V shows the obtained regression model parameters for information in the dataset, according to the method explained previously (note the Wilkinson notation [Wilkinson and Rogers 1973]).

As can be seen, Table V presents important values used to determine the accuracy of the model, which can be interpreted as follows:

- The estimated values (Estimated) for each coefficient of the regression model.
- The standard error (SE) for these estimated coefficients.
- The t-statistics (tStat) and p-value (p-Value) of each regressor of the model, which helps to establish which descriptors are statistically influential for the model.
- The root mean squared error (RMSE). This parameter determines the difference between the actual values and the ones predicted by our model. The obtained value (0.79) represents a low deviation of the error distribution according to the range of the evaluations (Likert scale from 1 to 9).
- The coefficient of determination (R2), which determines the variability explained by the model. In this case, the model explains 40.4% of the data variability.
- The adjusted coefficient of determination (R2 adj.), as a modified version of the previous parameter which compensates for the inclusion of more variables to the model. It is the parameter used to compare different regression models.
- The F-statistics and the p-value for the entire model, which establish that the model is representative (p-value < 0.05 as the significance level).

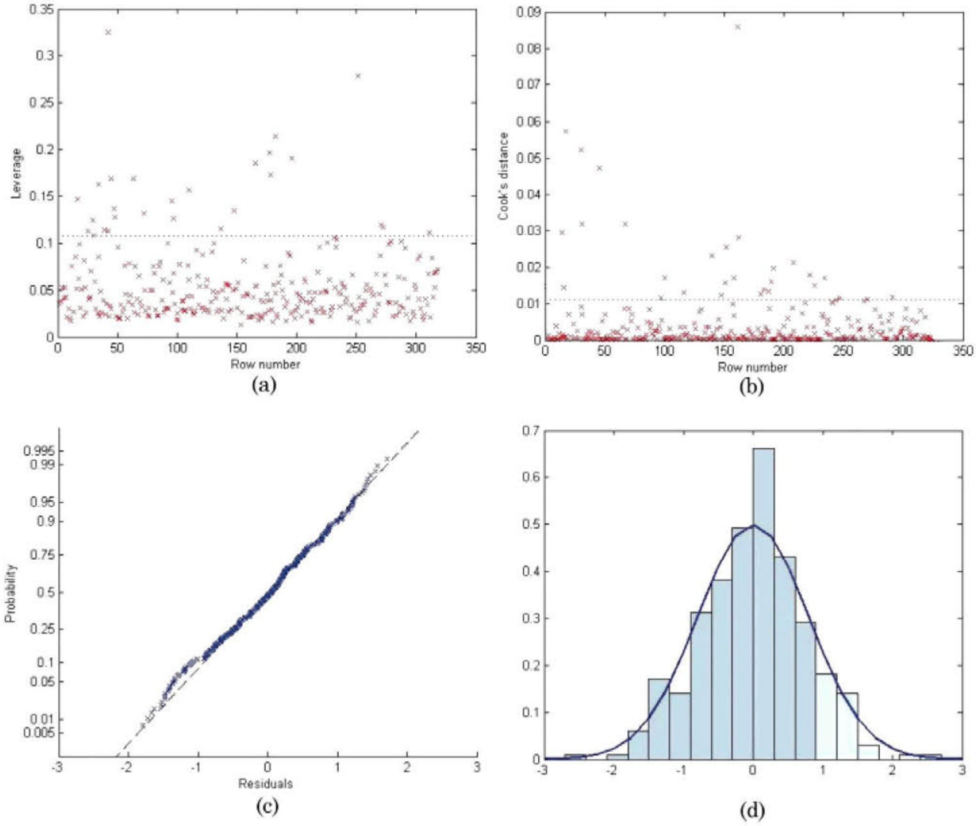

Fig. 4. Graphical diagnosis of the model. (a) Case order plot of leverage. (b) Case order plot of Cook's distance. (c) Normal probability plot of residuals. (d) Histogram of residuals.

Considering these results, only 10 of the 14 initial parameters are taken into account for the model, as only they make an important statistical contribution to the outcome.

According to the model definition, we can establish that the most influential coefficients are the ones related to the color continuity (SC) and the variability between different colors (VC_{inter}), which gives an idea of how relevant color information is for users when evaluating a Web site, as some previous research has shown [Cyr et al. 2010].

The next step is to find possible influential observation points that can affect the fitting of the model to remove them and to obtain a final model with no outliers. This diagnosis is made by means of a graphical study of the model's data, starting with a leverage plot of the data and model (Figure 4(a)), which helps us determine those observations that are far away from the average predicted value. As can be seen in the Figure 4(a), there are two points with a clear high value, but we have to continue the analysis to confirm whether they are influential points in the model. To this aim, we apply Cook's distance [Cook 1977], which is a measure of influence that considers both the location of the point in our 10 space dimensions and the response variability. According to Figure 4(b), there is only one point with a relatively high Cook's distance, but according to previous works, it is far from the common threshold established by Cook and Weisberg [1982], so it cannot be considered an influential point for our model.

Once we have discarded the presence of influential points, the next step is to determine the existence of outliers. To this end, we provide two additional plots:

- The residuals' histogram (Figure 4(c)), which represents the range of the residuals and their frequencies. In the figure, the line is close to straight, which means a reasonable fit to normally distributed residuals.
- The probability plot (Figure 4(d)), which examines the distribution of the residuals compared to a normal distribution. The histogram looks fairly symmetric, so we can discard the presence of outliers.

Based on this analysis, we can determine that there are no a clear outliers or influential observations. Hence, we do not remove any data from the training subset to recalculate the regression model.

5. RESULTS FOR OUR NEW AESTHETICS EVALUATION MODEL

Throughout this section, we explain the results obtained for our new prediction model via a comparative analysis with a previous model [Reinecke et al. 2013] to establish the advantages of our proposal. The comparison is done with the said research work, as it provides a complete aesthetic model based on the correlation between the users' opinion for the same dataset and objective metrics divided into two different classes:

- The visual complexity level of the Web sites, determined by a regression model composed of seven elements of the image: text area, nontext area, number of leaves, number of text groups, number of image areas, colorfulness (according to the definition given in Tuch et al. [2012a]), and hue level
- Their color information, which is calculated by means of a combination of 13 image attributes: gray, white, maroon, green, lime, blue, teal, saturation, colorfulness (according to the definition in Hasler and Suestrunk [2003]), number of image areas, number of quadtree leaves, text area, and nontext area.

As explained by the authors, each model has been established separately according to the users' evaluation regarding their impression on the visual complexity and colorfulness of the Web pages. Then the final model is obtained by applying a new regression model that combines these two partial solutions in relation to the users' visual appeal perception. At this point, it is important to note that we are making the comparison with the general model obtained by the combination of these two factors (visual complexity and color information), but we are not considering the final results given by the authors regarding the personalization of the model with the users' characteristics, as we are proposing a better general model.

The comparative analysis is especially focused on two important aspects of the obtained model: the improved correlation between objective evaluation and users' assessment, and the use of a smaller set of parameters based on different aesthetic parameters of the Web site's image. The combination of both objectives must lead us to a more effective predictive model that provides better results with a lower processing load, which can reduce the associated cost.

We provide a comparative study focused on two different analyses:

- In the first analysis, the obtained error is reduced (4.5%) by making use of only 10 descriptors, as opposed to the 17 descriptors used in the other model (41.1% reduction).
- In the second analysis, we compare the obtained results for the training subset of the generic dataset used for obtaining the evaluation method. As shown in Table VI, our model obtains better results in terms of error, and it also improves the explanation of the data variability (31%).

Table VI. Comparison Analysis for Both Models for the Training Subsets

Parameter	Reinecke's Model	Our Model
RMSE	0.90	0.861
R^2	0.267	0.404
R^2 adj.	0.254	0.332
p-value	<.001	<.001
Number of descriptors in model	17	10

Table VII. Validation Results of the Model with a General Dataset

	Classic Aesthetics	Expressive Aesthetics	General Aesthetic Level
RMSE	1.336	1.154	1.28

Table VIII. Validation Results of the Model with a Specific Dataset (Web Search Engines)

	Classic Aesthetics	Expressive Aesthetics	General Aesthetic Level
RMSE	0.9	1.01	1.18

Table VI shows that our model presents an improvement over the existing research in two aspects: obtaining a slightly more accurate result by using a smaller set of parameters.

Once the regression model for the Web aesthetic evaluation has been obtained, the next step is its validation over a different dataset. As such, we used two additional datasets (see Section 4.2).

Validation over an Additional Generic Dataset

The first dataset (obtained from Moshagen and Thielsch [2010]) provides users' evaluation data according to the two main aesthetics dimensions in HCI design: classic, which is related to a more conventional definition of the concept based on a clear design, and expressive, which is related to the artistic sense of the design (creativity and originality). It also provides a general evaluation for the visual appeal of each Web site according to the VisAWI questionnaire, which we considered as a reference for the aesthetic level.

For this validation process, the first step was to obtain the low-level descriptors for each Web site. Then the new model was applied and the obtained results were compared to the mean rating for each dimension.

Table VII presents the RMSE obtained by inferring the users' opinion using our new evaluation method. As can be seen in the table, we can ensure that this new dataset does not represent a subset of the previous one since the related error is higher than the training one. Moreover, our model presents a higher accuracy for the expressive aesthetics level. This is related to the fact that our method evaluates the Web composition in terms of low-level characteristics (luminance, chrominance, and texture), which are directly related to the image perception process and thus to the expressive dimension. On the contrary, we are not considering the common parameters to evaluate aesthetics in terms of complexity and clear design (balance, symmetry, unity, etc.) that are used in the classic dimension.

Validation over a Specific Dataset

As done in the previous case, for this validation the model results are compared to the users' rating along three different aesthetic dimensions: classic, expressive, and generic visual appeal level. The obtained results are provided in Table VIII.

In this case, our model provides higher accuracy for the classic aesthetic level than for the expressive one, which may be because users do not consider variety or originality as

Table IX. Validation Results According to Different Web Search Engines

RMSE	Classic Aesthetics	Expressive Aesthetics
Web search engine with multimedia component	0.82	0.78
Web search engine without multimedia component	0.72	0.73

relevant features in the evaluation of search engine Web pages. Nevertheless, this leads us to believe that there may be some other influential parameters in the Web search environment that must be addressed. To analyze this, we repeat the validation by distinguishing between interfaces with and without multimedia components, obtaining the results shown in Table IX. As can be seen from the table, the model fits better for the expressive aesthetics in those interfaces that have multimedia elements in their design. This is mainly because these interfaces are more similar to normal Web layouts, so the model behaves as it does at a general level.

6. DISCUSSION OF THE IMPLICATIONS

With regard to the main purpose of this study, throughout this work we have presented a new model for an objective aesthetics evaluation of static Web sites by means of a limited set of 10 low-level parameters relating to three different image aspects: luminance, texture, and chrominance. Moreover, as explained previously, this new model to predict the users' opinion improves on previous works, providing an easier and less costly evaluation method. This improvement can be classified into three different aspects:

- First, it achieves a better prediction accuracy (less RMSE and an explanation of high data variability) with fewer parameters, which means that our solution is more efficient than previous models.
- Second, it is independent from the image content distribution, avoiding the structure analysis dependency that other studies present [Zheng et al. 2009]. An additional strength of the current approach is that it provides an entirely automated solution without containing any subjective stage in the analysis process.
- Third, the users' opinion about the expressive aesthetics can also be inferred by this method, with the exception of the search engine field, where users do not consider it in the same way as in other types of Web sites.

In another respect, it is also important to analyze the model composition itself. We can establish that the most important parameters for the users' first impression are the ones related to the chrominance of the image (VC_{inter} and SC), which can be explained by the features in the human visual system [Oliva and Torralba 2006; Rosenholtz et al. 2005]. With regard to image texture information, the evaluation only depends on the line energy, as it is directly related to the image transitions. Therefore, it can be proposed as a new and simpler way to analyze the complexity of the image.

Our model is composed of some low-level parameters considered separately in addition to some combinations that allow minimization of negative effects or maximization of positive ones. Another interesting contribution is the study of these combinations to give them the correct visual meaning according to the perception process. Thus, we can interpret the following combinations:

- The coefficient composed of a second and third bit plane entropy (E2 and E3) can be construed as a high-level parameter related to the homogeneity in the image, which contributes to the positive perception of the Web site.
- The coefficients corresponding to the line energy (L_e), number of dominant colors (N), and color variance (S_v) reveal the relationship between color and texture information, one with a positive effect (the one related to the variance, as it confirms the

difference among colors), and one with a negative effect (preventing the use of many different colors). This can be explained because a higher complexity level reveals higher entropy, which allows for the presence of several colors. According to this interpretation, these two effects are correlated and they affect in the same way to the final model. Moreover, it confirms the idea that a medium level of complexity is needed for users to positively score the Web page [Comber and Maltby 1997].

- The coefficients that combine spatial coherence (SC) and interchromatic variance (VC_{inter}) can be interpreted similarly: if the variance between different colors is high (which means a more complex interface in terms of colors), the color continuity is low, so both effects go hand in hand.
- The coefficients corresponding to the first bit plane entropy (E1) and spatial coherence (SC), as well as the second bit plane entropy and color variance (S_v), reveal the relationship between the luminance descriptors and the color information of the image.
- Finally, the coefficient that contains the mean luminance (L_m) reveals that the importance of the light information shows a quadratic effect.

Based on these conclusions, Web site designers can derive some important implications. It can be said that users mostly perceive Web site aesthetics by means of the chrominance and luminance levels of the image, so a special effort must be made to optimize their influence and ensure a positive perception. For this reason, this optimization can be achieved by maximizing the model relationships among the low-level descriptors detected.

Finally, as stated previously, our model provides a determination coefficient of 40.4%, which means that this model explains a 40.4% of the variability (which is reduced to 37% when considering the adjusted determination coefficient, which is in charge of compensating the inclusion of more variables to the model). This result can be considered a good value, as in this domain there are many uncontrollable, undeterminable, and unknown subjective factors influencing the aesthetics-level impression. Nevertheless, some authors [Achen 1982; Moksony 1990] claim that the use of different error coefficients, such as the standard error or the RMSE coefficient, seem to be more robust and could be preferred over the determination coefficient as a measurement of goodness of the model.

7. LIMITATIONS AND FUTURE WORK

Our solution shows some promising results, although it does have some limitations. In this regard, our model uses objective characteristics to evaluate Web sites and predict the users' first impression on visual appeal to the general public without segmentation. However, it would be interesting to analyze the influence of users' demographic background on their perception. Therefore, the first line of research that we propose is related to the study of the interaction between these demographic variables and the low-level descriptors of the model to achieve a more personal prediction solution.

Another limitation of our model is the evaluation of only static images of Web pages that fit on a single screen, as users are not able to scroll through the page, although this is mostly relevant for usability analysis than for aesthetics [Thielsch et al. 2015]. Nevertheless, and with the purpose of evolving our solution, it would be interesting to analyze the change of the model for Web sites that are more complex and dynamic or have an interactive layout, where actions like scrolling the page and navigating different links produce a new user behavior (as explained in Weinreich et al. [2008]). Moreover, although our research is focused on the analysis of computer interfaces, it could also be interesting to expand the evaluation to other scenarios, such as smartphones, given their current importance.

Another interesting issue for future work is the combination with models evaluating other user experience dimensions to obtain a more complex solution for the Web site design environment. In this regard, research works like the one shown in Speicher et al. [2015] represent an interesting starting point to achieve this aim. In it, the authors define a suite for usability optimization in the search engine Web interfaces environment using two elements: an evaluation tool based on usability scores and a set of best practices for improvement.

In another aspect, as stated earlier, our model's aim is to provide a new method for the evaluation of the first impression about aesthetics in the Web site domain, which can be related to the perceptual analysis proposed by Leder et al. [2004]. Nevertheless, it would be interesting to evolve it to find a relationship or even to analyze later dimensions of the psychological model given by Leder et al. [2004] to provide a final evaluation methodology of the entire process.

The potential evolution of our model by means of these research lines can help us design an innovative tool for the automatic adaptation of Web site layouts. This would increase users' acceptance by considering their Web site aesthetics preferences in their context of use.

8. CONCLUSIONS

Throughout this article, we presented a new model to evaluate Web sites and predict the users' first impression based on a set of 10 objective image descriptors derived from MPEG-7 and an analysis of a static image of the interface in three essential dimensions: luminance, chrominance, and texture.

The precision of the model is proven by the obtained results and confirms the importance of aesthetics in first impressions from Web sites. Moreover, these results help us confirm our initial idea that relates rapid users' evaluation with low-level features above other aspects of the usage scenario. Thanks to that, we can confirm it is possible to evaluate Web pages in an objective way by applying specific computer algorithms that facilitate the rating prediction process.

Our model can be considered as the starting point of an effective tool to evaluate Web site design, as it represents an easier solution than similar previous ones. This means that it can help reduce costs and allow for its application in the early stages of the creation process, because it can be applied not only to functional Web sites but also to prototypes. As explained in Section 1, this model can be applied by nonaesthetics experts, although the results will be better interpreted if they have some image analysis skills to define visual changes to obtain a more aesthetic Web site. Furthermore, it can be seen as a cheaper substitute for empirical evaluations with human users and provide quantitative feedback for the designers, thus obtaining optimized interfaces in a reduced amount of time.

In another respect, we also want to emphasize that according to the relationship between Web site aesthetics and other Web site features such as usability, loyalty, and intention to revisit (see Section 2.1), the obtaining and development of this model may also help to find a way to improve these other features, which represent an important collateral advantage for designers.

Finally, the practical implications of our research work can be summed up as follows:

- We propose a new model based on a set of 10 low-level parameters for Web site characterization. These descriptors are derived from the MPEG-7 standard and have never been used to provide this kind of solution, although they have been proven to yield optimal results for image recommendation.

- These descriptors can be modeled to predict the users' first rating about Web site aesthetics due to the importance of these low-level features for the human visual system.
- We have tested this model over a new dataset based on Web search engines, concluding that it can infer the users' opinion on aesthetics with high accuracy.
- Finally, this model can be considered as a foundation for an objective tool that provides quantitative feedback about the visual appeal of Web sites in the design process, helping developers obtain an interface that can be easily accepted by users.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Kaharina Reinecke of the University of Michigan and Krzysztof Z. Gajos of Harvard University for providing the first public dataset of Web sites' visual appeal, and especially Dr. Reinecke for providing the specifications.

REFERENCES

C. H. Achen. 1982. *Interpreting and Using Regression*. Sage Publications, Beverly Hills, CA, 61–64.

A. Altaboli and Y. Lin. 2011. Objective and subjective measures of visual aesthetics of website interface design: The two sides of the coin. In *Proceedings of the 14th International Conference on Human-Computer Interaction: Design and Development Approaches, Volume Part I (HCII'11)*. 35–44.

A. De Angeli, A. Sutcliffe, and J. Hartmann. 2006. Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS'06)*. ACM, New York, NY, 271–280. DOI:http://doi.acm.org/10.1145/1142405.1142446

M. Barrilero, S. Uribe, M. Alduán, F. Sánchez, and F. Álvarez. 2011. In-network content based image recommendation system for content-aware networks. In *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS'11)*. 115–120, DOI:http://dx.doi.org/10.1109/INFCOMW.2011.5928791

M. Bastan, H. Cam, U. Güdükbay, and O. Ulusoy. 2009. An MPEG-7 compatible video retrieval system with integrated support for complex multimodal queries. *IEEE Multimedia* PP, 99, 1. DOI:http://dx.doi.org/10.1109/MMUL.2009.74

D. E. Berlyne. 1970. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, New York, NY.

J. Buguin, L. Corb, J. Manyika, O. Nottebohm, M. Chui, B. Muller Barbat, and R. Said. 2011. *The Impact of Internet Technologies: Search*. Technical Report. McKinsey & Company.

R. T. Cober, D. J. Brown, P. E. Levy, A. B. Cober, and L. M. Keeping. 2003. Organizational Web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection and Assessment* 11, 158–169.

T. Comber and J. R. Maltby. 1997. Layout complexity: Does it measure usability? In *Human-Computer Interaction (INTERACT'97)*. IFIP—The International Federation for Information Processing. Springer, 623–626. DOI:http://dx.doi.org/10.1007/978-0-387-35175-9_109

R. D. Cook. 1977. Detection of influential observation in linear regression. *Technometrics, American Statistical Association and American Society for Quality* 19, 1, 15–18.

R. D. Cook and S. Weisberg. 1982. *Residuals and Influence in Regression*. New York. Chapman & Hall, New York, NY.

D. Cyr, C. Bonanni, J. Bowes, and J. Ilsever. 2005. Beyond trust: Website design preferences across cultures. *Journal of Global Information Management* 13, 4, 25–54.

D. Cyr, M. Head, and A. Ivanov. 2006. Design aesthetics leading to m-loyalty in mobile commerce. *Information & Management* 43, 8, 950–963. DOI:http://dx.doi.org/10.1016/j.im.2006.08.009

D. Cyr, M. Head, and H. Larios. 2010. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal on Human-Computer Studies* 68, 1–2, 1–21. DOI:http://dx.doi.org/10.1016/j.ijhcs.2009.08.005

M. Donyaee, A. Seffah, and J. Rilling. 2006. Exploring the correlations between predictive usability measures and user test. In *Proceedings of the International Conference on Software and Product Measurement (MENSURA'06)*.

G. Fernandes, G. Lindgaard, R. Dillon, and J. Wood. 2003. Judging the appeal of Web sites. In *Proceedings of the 4th World Congress on the Management of Electronic Commerce*. 15–17.

F. Fu, S. Y. Chiu, and C. H. Su. 2007. Measuring the screen complexity of Web pages. In *Proceedings of the 2007 Conference on Human Interface: Part II*. 720–729.

W. O. Galitz. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques* (3rd ed.). Wiley.

J. Hartmann, A. Sutcliffe, and A. De Angeli. 2008. Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction* 15, 4, Article No. 15. DOI:http://doi.acm.org/10.1145/1460355.1460357

D. Hasler and S. E. Suesstrunk. 2003. Measuring colorfulness in natural images. In *Proceedings of SPIE/IS&T Human Vision and Electronic Imaging*, Vol. 5007. 87–95. DOI:10.1117/12.477378

M. Hassenzahl. 2008. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction* 19, 4, 319–349. DOI:http://dx.doi.org/10.1207/s15327051hci1904_2

M. Hassenzahl and A. Monk. 2010. The inference of perceived usability from beauty. *Human-Computer Interaction* 25, 3, 235–260. DOI:10.1080/07370024.2012.500139

M. Hassenzahl and N. Tractinsky. 2006. User experience—a research agenda. *Behaviour and Information Technology* 25, 2, 91–97.

J. L. Hintze. 2007. Stepwise regression. In *NCSS Statistical System User's Guide III: Regression and Curve Fitting*. NCSS, Kaysville, UT, 197–207.

D. H. Hubel and T. N. Wiesel. 1979. Brain mechanism of vision. *Scientific American* 241, 3, 150–162.

T. Jacobsen and L. Höfel. 2002. Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills* 95, 755–766.

M. Kurosu and K. Kashimura. 1995. Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. In *Conference Companion on Human Factors in Computing Systems (CHI'95)*. ACM, New York, NY, 292–293. DOI:http://doi.acm.org/10.1145/223355.223680

T. Lavie and N. Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of Web sites. *International Journal of Human-Computer Studies* 60, 3, 269–298. DOI:http://dx.doi.org/10.1016/j.ijhcs.2003.09.002

H. Leder, B. Belke, A. Oeberst, and D. Augustin. 2004. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology* 95, 489–508.

H. Leder and M. Nadal. 2014. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—developments and challenges in empirical aesthetics. *British Journal of Psychology* 105, 4, 443–464. DOI:10.1111/bjop.1208

G. Lindgaard, C. Dudek, D. Sen, I. Sumegi, and P. Noonan. 2011. An exploitation of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer Human Interaction* 18, 1, 1–30.

G. Lindgaard, G. Fernandes, C. Dudek, and J. M. Brown. 2006. Attention Web designers: You have 50 milliseconds to make a good first impression! *Behaviour and Information Technology* 25, 2, 115–126. DOI:http://dx.doi.org/10.1080/01449290500330448.

Y. Liu. 2003. Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. *Ergonomics* 46, 13–14, 1273–1292. DOI:http://dx.doi.org/10.1080/00140130310001610829

S. Lok, S. Feiner, and G. Ngai. 2004. Evaluation of visual balance for automated layout. In *Proceedings of the International Conference on Intelligent User Interaction*. 101–108.

S. Mahlke. 2002. Factors influencing the experience of Website usage. In *Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems (CHI EA'02)*. ACM, New York, NY, 846–847. DOI:http://doi.acm.org/10.1145/506443.506628

S. Mahlke. 2008a. Visual aesthetics and the user experience. In *The Study of Visual Aesthetics in Human-Computer Interaction*, M. Hassenzahl, G. Lindgaard, A. Platz, and N. Tractinsky (Eds.). Schloss, Dagstuhl, Germany, 1–6.

S. Mahlke. 2008b. *User Experience of Interaction with Technical Systems*. Ph.D. Dissertation. Berlin University of Technology, Berlin, Germany.

M. V. Mankeliunas. 1980. Los modelos en psicofísica. En *Psicología de la percepción*, A. Ardila (dir.). Mexico.

MinGW. 2015. Minimalist GNU for Windows: A Native Windows Port of the GNU Compiler Collection (GCC). Retrieved January 14, 2016, from http://sourceforge.net/projects/mingw/.

F. Moksony. 1990. Small is beautiful. The use and interpretation of R2 in social research. *Szociologiai Szemle* (Special Issue) 130–138.

M. Moshagen, J. Musch, and A. S. Göritz. 2009. A blessing, not a curse: Experimental evidence for beneficial effects on visual aesthetics on performance. *Ergonomics* 52, 10, 1311–1320. DOI:http://dx.doi.org/10.1080/00140130903061717

M. Moshagen and M. T. Thielsch. 2010. Facets of visual aesthetics. *International Journal of Human-Computer Studies* 68, 10, 689–709. DOI:http://dx.doi.org/10.1016/j.ijhcs.2010.05.006

M. Moshagen and M. T. Thielsch. 2013. A short version of the visual aesthetics of Websites inventory. *Behaviour and Information Technology* 32, 12, 1305–1311. DOI:10.1080/0144929X.2012.694910

MPEG ISO/IEC. 2000. *ISO/IEC CD 15938 (1-6) Information Technology. Multimedia Content Description Interface: Part 1-6 ISO/IEC JTC/SC29/WG11/N3701-N3706.* MPEG, La Baule, France.

K. Mullet and D. Sano. 1996. Designing visual interfaces. *ACM SIGCHI Bulletin* 28, 2, 82–83. DOI:http://dx.doi.org/10.1145/226650.570118

M. Nebeling, F. Matulic, and M. C. Norrie. 2011. Metrics for the evaluation of news site content layout in large-screen contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11).* ACM, New York, NY, 1511–1520. DOI:http://doi.acm.org/10.1145/1978942.1979164

NetMarketshare. n.d. Home Page. Retrieved November 12, 2015, from https://netmarketshare.com.

D. C. L. Ngo, A. Samsudin, and R. Abdullah. 2000. Aesthetics measures for assessing graphic screens. *Journal of Information Science and Engineering* 16, 1, 97–116.

D. C. L. Ngo, L. S. Teo, and J. G. Byrne. 2003. Modeling interface aesthetics. *Information Sciences* 152, 1, 25–46. DOI:http://dx.doi.org/10.1016/S0020-0255(02)00404-8

J. Nielsen. 1993. *Usability Engineering*. Morgan Kaufmann, San Francisco, CA.

D. A. Norman. 2004. Introduction to this special section on beauty, goodness, and usability. *Human-Computer Interaction* 19, 4, 311–318. DOI:http://dx.doi.org/10.1207/s15327051hci1904_1

A. Oliva and A. Torralba. 2006. Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research* 155, 23–36.

M. Pajusalu. 2012. *The Evaluation of User Interface Aesthetics*. Master's Thesis. Institute of Informatics, Tallinn University, Tallinn, Estonia.

J. W. Palmer. 2002. Web site usability, design, and performance metrics. *Information Systems Research* 13, 2, 151–167. DOI:http://dx.doi.org/10.1287/isre.13.2.151.88

E. Papachristos and N. Avouris. 2011. Are first impressions about Websites only related to visual appeal? In *Proceedings of the 13th IFIP International Conference on Human-Computer Interaction, Volume Part I (INTERACT'11).* 489–496.

H. C. Purchase, J. Hamer, A. Jamieson, and O. Ryan. 2011. Investigating objective measures of Web page aesthetics and usability. In *Proceedings of the 12th Australasian User Interface Conference, Volume 17 (AUIC'11)*. 19–28.

J. Redish, R. G. Bias, R. Bailey, R. Molich, J. Dumas, and J. M. Spool. 2002. Usability in practice: Formative usability evaluations—evolution and revolution. In *Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems (CHI EA'02).* ACM, New York, NY, 885–890. DOI:http://doi.acm.org/10.1145/506443.506647

K. Reinecke and K. Z. Gajos. 2014. Quantifying visual preferences around the world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14).* ACM, New York, NY, 11–20. DOI:http://dx.doi.org/10.1145/2556288.2557052

K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K. Z. Gajos. 2013. Predicting users' first impressions of Website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13).* ACM, New York, NY, 2049–2058. DOI:http://dx.doi.org 10.1145/2470654.2481281

D. Robins and J. Holmes. 2008. Aesthetics and credibility in Web site design. *Information Processing and Management* 44, 1, 386–399. DOI:http://dx.doi.org/10.1016/j.ipm.2007.02.003

V. Roto, E. Law, A. Vermeeren, and J. Hoonhout. 2011. User experience white paper. In *Proceedings of the Dagstuhl Seminar on Demarcating User Experience.*

R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. 2005. Feature congestion: A measure of display clutter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05).* ACM, New York, NY, 761–770. DOI:http://doi.acm.org/10.1145/1054972.1055078

P. Salembier and T. Sikora. 2002. *Introduction to MPEG-7: Multimedia Content Description Interface*. B. S. Manjunath (Ed.). John Wiley & Sons, New York, NY.

F. Sánchez, M. Barrilero, S. Uribe, F. Álvarez, A. Tena, and J. M. Menéndez. 2012. Social and content hybrid image recommender system for mobile social networks. *Mobile Networks and Applications* 17, 6, 782–795. DOI:http://dx.doi.org/10.1007/s11036-012-0399-6

A. Sears. 1993. Layout appropriateness: A metric for evaluating user interface widget layouts. *IEEE Transactions on Software Engineering* 19, 7, 707–719.

B. Shackel. 2009. Usability—context, framework, definition, design and evaluation. *Interacting with Computers* 21, 5–6, 339–346. DOI:http://dx.doi.org/10.1016/j.intcom.2009.04.007

Z. Shan and W. Hai-Tao. 2008. Image retrieval based on bit-plane distribution entropy. In *Proceedings of the 2008 International Conference on Science and Software Engineering*. 532–535. DOI:http://dx.doi.org/10.1109/CSSE.2008.270

P. J. Silvia. 2006. *Exploring the Psychology of Interest*. Oxford University Press, New York, NY.

A. Sonderegger and J. Sauer. 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* 41, 3, 403–410. DOI:http://dx.doi.org/10.1016/j.apergo.2009.09.002

M. Speicher, A. Both, and M. Gaedke. 2015. S.O.S.: Does your search engine results page (SERP) need help? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. 1005–1014.

J. Stevens. 1996. *Applied Multivariate Statistics for the Social Sciences* (3rd ed.). Lawrence Erlbaum Associates, Mahwah, NJ.

R. Strebe. 2011. Visual aesthetics of Websites: The visceral level of perception and its influence on user behaviour. In *Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries: Research and Advanced Technology for Digital Libraries (TPDL'11)*. 523–526.

B. G. Tabachnick and L. S. Fidell. 2007. *Using Multivariate Statistics* (5th ed.). Pearson, Boston, MA.

M. T. Thielsch, I. Blotenberg, and R. Jaron. 2014. User evaluation of Websites: From first impression to recommendation. *Interacting with Computers* 26, 1, 89–102.

M. T. Thielsch, R. Engel, and G. Hirschfeld. 2015. Expected usability is not a valid indicator of experience usability. *PeerJ Computer Science* 1, e19.

M. Thüring and S. Mahlke. 2007. Usability, aesthetics and emotions in human technology interaction. *International Journal of Psychology* 42, 4, 253–264. DOI:10.1080/00207590701396674

TNS. 2008. Digital World, Digital Life: Snapshots of Our Online Behavior and Perspectives Around the World. Retrieved January 14, 2017, from http://www.tnsglobal.com/_assets/files/TNS_Market_Research_Digital_World_Digital_Life.pdf.

N. Tractinsky. 1997. Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*. ACM, New York, NY, 115–122. DOI:http://doi.acm.org/10.1145/258549.258626

N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi. 2006. Evaluating the consistency of immediate aesthetic perceptions of Web pages. *International Journal of Human-Computer Studies* 64, 11, 1071–1083. DOI:http://dx.doi.org/10.1016/j.ijhcs.2006.06.009

N. Tractinsky, A. S. Katz, and D. Ikar. 2000. What is beautiful is usable. *Interacting with Computers* 13, 2, 127–145. DOI:http://dx.doi.org/10.1016/S0953-5438(00)00031-X

A. N. Tuch, E. E. Presslaber, M. Stöcklin, K. Opwis, and J. A. Bargas-Avila. 2012a. The role of visual complexity and prototypicality regarding first impression of Websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies* 70, 11, 794–811. DOI:http://dx.doi.org/10.1016/j.ijhcs.2012.06.003

An N. Tuch, S. P. Roth, K. Hornbæk, K. Opwis, and J. A. Bargas-Avila. 2012b. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior* 28, 5, 1596–1607. DOI:http://dx.doi.org/10.1016/j.chb.2012.03.024

J. Vanderdonckt and X. Gillo. 1994. Visual techniques for traditional and multimedia layouts. In *Proceedings of the Workshop on Advanced Visual Interfaces (AVI'94)*. ACM, New York, NY, 95–104. DOI:http://dx.doi.org/10.1145/192309.192334

M. Xie. 2003. The basics of visual perception. In *Fundamentals of Robotics: Linking Perception to Action*. World Scientific, River Edge, NJ, 446–450.

Webby Awards. 2015. Home Page. Retrieved January 14, 2017, from http://www.webbyawards.com/.

H. Weinreich, H. Obendorf, E. Herder, and M. Mayer. 2008. Not quite the average: An empirical study of Web use. *ACM Transactions on the Web* 2, 1, Article No. 5. DOI:http://dx.doi.org/10.1145/1326561.1326566

G. N. Wilkinson and C. E. Rogers. 1973. Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 22, 3, 392–399. DOI:http://dx.doi.org/10.2307/2346786

J. M. Zain, M. Tey, and G. Ying Soon. 2008. Using aesthetic measurement application (AMA) to measure aesthetics of Web page interfaces. In *Proceedings of the 2008 4th International Conference on Natural Computation, Volume 6 (ICNC'08)*. IEEE, Los Alamitos, CA, 96–100. DOI:http://dx.doi.org/10.1109/ICNC.2008.764

M. Zen and J. Vanderdonckt. 2014. Towards an evaluation of graphical user interfaces aesthetics based on metrics. In *Proceedings of the IEEE 8th Conference on Research Challenges in Information Science (RCIS'14)*. IEEE, Los Alamitos, CA, 1–6.

Q. Zhang, W. Kang, C. Zhao, and X. Ming. 2009. Aesthetic coloring for complex layout using genetic algorithm. In *Proceedings of the 2009 WRI Global Congress on Intelligent Systems (GCIS'09)*. IEEE, Los Alamitos, CA, 406–410. DOI:10.1109/GCIS.2009.15

X. S. Zheng, I. Chakraborty, J. Jeng-Weei Lin, and R. Rauschenberger. 2009. Correlating low-level image statistics with users—rapid aesthetic and affective judgments of Web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, New York, NY, 1–10. DOI:http://doi.acm.org/10.1145/1518701.1518703