Correlating Low-Level Image Statistics with Users' Rapid Aesthetic and Affective Judgments of Web Pages

Xianjun Sam Zheng
Siemens Corporate Research
755 College Road East
Princeton, NJ 08540
sam.zheng@siemens.com

Ishani Chakraborty
Siemens Corporate Research
and Rutgers University
755 College Road East
Princeton, NJ 08540
ishanic@cs.rutgers.edu

James Jeng-Weei Lin
Siemens Corporate Research
755 College Road East
Princeton, NJ 08540
jameslin@siemens.com

Robert Rauschenberger
Siemens Corporate Research
and Simon Fraser University
755 College Road East
Princeton, NJ 08540
robert.rauschenberger@siemens.com

ABSTRACT
In this paper, we report a study that examines the relationship between image-based computational analyses of web pages and users' aesthetic judgments about the same image material. Web pages were iteratively decomposed into quadrants of minimum entropy (quadtree decomposition) based on low-level image statistics, to permit a characterization of these pages in terms of their respective organizational symmetry, balance and equilibrium. These attributes were then evaluated for their correlation with human participants' subjective ratings of the same web pages on four aesthetic and affective dimensions. Several of these correlations were quite large and revealed interesting patterns in the relationship between low-level (i.e., pixel-level) image statistics and design-relevant dimensions.

Author Keywords
Low-level features, image statistics, computer vision, user interface design, aesthetic judgments, empirical methods

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User interfaces.

INTRODUCTION
Increasing numbers of research studies (e.g., [15,29,30]) and business successes show that an attractive design not only increases users' acceptance of a product or system, but also makes an important contribution to its associated user experience. The attractiveness of a design is generally determined subjectively by a split-second judgment, as demonstrated by Lindgaard, Fernandes, Dudek and Brown [17]. In that study, participants were able to make aesthetic judgments about the design of a web page within 50 milliseconds, and do so with reliability. The rapidness of this judgment suggests that it is formed on the basis of an automatic (feed-forward) analysis of the display in terms of very rudimentary, elementary features, such as those encoded early in the human visual system (e.g., [4,11,12,31, 35]).

Professional design, by contrast, is a laborious, time-intensive process that builds up artifacts skillfully using heuristics or guidelines that entail applying various elements of design, such as color, contrast or texture, and organizing or structuring them in accordance with established principles, such as symmetry, balance, unity, etc. [1,5,8,28] – many of which were later discovered to be the same principles that underlie the visual system's perceptual organization of the visual input it receives from the physical world (e.g., [13]).

Because of the extraordinarily rapid aesthetic and affective judgments users are capable of making about displays they have never before seen, it must be possible to model or predict users' judgments using pixel-level statistics that are agnostic about the existence of objects, meanings, cultural or discipline-specific conventions, etc. From what is known about the neurophysiology of the brain (see, e.g.,

[16]), this is the only type of information that will have a chance to be processed in as brief a timeframe as employed in, for example, Lindgaard et al.'s study [17] (see also [6]). What we are proposing here might therefore be considered quite innovative, or perhaps even slightly controversial: that something as seemingly sophisticated as a user's aesthetic and affective evaluation of a design may be reduced, in part, to a decomposition of the display into simple colors, textures and luminances.

In the fields of computational vision and content-based image retrieval, similar proposals, although still somewhat controversial, have gained ascendency in recent years. For example, Oliva and Torralba have shown compellingly how scene gist (an inherently semantic classification) can be derived from low-level image statistics (which they call the Spatial Envelope) [24]. Similarly, Wang, Li and Wiederhold's *SIMPLIcity* image-retrieval system, is capable of matching an image to semantic categories (e.g., indoor-outdoor, objectionable-benign, etc.) based on a decomposition into 4x4-pixel regions characterized in terms of simple features, such as color, texture, and shape [32]. On these a priori grounds, the suggestion that aesthetic and affective judgments about the design of a display could be computed from similar image statistics appears plausible.

In previous research, the present authors have shown that a supervised learning algorithm was capable of learning the correlation between low-level image statistics, such as spatial frequency, luminance entropy, etc., and users' judgments of perceived usability and subjective appearance (e.g., aggressive, expensive, masculine, self-explanatory, etc.) [34]. In fact, it was able to predict the rank ordering of car infotainment systems by users on aesthetic and usability dimensions with between 60 and 72 percent accuracy – a remarkable feat, given that the algorithm needed to get the precise order of UI examples right in order for its response to count as correct.

The present study extends this earlier work by examining whether the layout *structure* that is computed based on low-level image statistics adequately captures design principles such as symmetry, balance and equilibrium (described in detail below). As such, it significantly extends the work of Ngo and colleagues [20-22], who, in turn, elaborated on the work of Birkhoff [2]. In contrast to this earlier work, however, the present study does not base its computational assessment of the aesthetic dimensions *symmetry*, *balance* and *equilibrium* on a manually-derived object-based definition of image regions, but rather on the abstract, agnosic (i.e., object-agnostic) structural properties of the image, which, themselves, are computed from low-level statistics far removed from objects, meanings or cultural contexts, etc. In this respect, the work reported here is more radical in its formalization of the evaluation process than any of its predecessors. Indeed, Ngo et al. [22] themselves conceded that "generalizing from [their] model … to real applications may not be tenable," whereas here, we show that it is possible to develop a general purpose model that can be applied to any image data.

Should it be determined that, as in our previous work, relatively high-level aesthetic attributes can be abstracted from an image using pixel-level statistics, and these attributes, in turn, correlate highly with users' subjective evaluations of aesthetic and affected dimensions of a UI, our method could potentially aid designers by providing a highly efficient surrogate for cost- and time-intensive user studies during early design iterations. Furthermore, it would shed light on the brain processes that permit ultra-rapid subjective judgments of the aesthetic and affective qualities of a scene or image, demonstrating that these are the results of the processing of the low-level features of an image in early visual areas of the brain.

In the following, we provide a literature review, followed by a description of the methods employed in our study. Finally, we close with a discussion of our principal findings, and concluded the paper with a deliberation of future directions and design implications.

Related work

Rapid aesthetic judgments

The idea that affective judgments can be made at a precognitive level – indeed *are* made at a precognitive level – goes back at least to Zajonc, who showed that people's preferences for particular stimuli are developed within only a brief exposure time – as little as a few milliseconds [33]. Subsequently, Norman theorized that the first stage of users' reactions toward designs is at the visceral level [23]: The response is immediate, holistic, and physiological, not ponderous, analytic or cognitive-deliberative. More recently, Lindgaard et al. conducted several studies to investigate how quickly users form aesthetic impression of web page designs [17]. They concluded, as expressed most succinctly and pointedly in the title of their paper: "Attention web designers: You have 50 ms to make a good first impression!" This same idea has also recently captured the imagination of non-expert audiences in the form of a popular book by Malcolm Gladwell, "Blink: The Power of Thinking Without Thinking" [9].

Computational aesthetic measures

In order to improve design effectiveness and efficiency, it is important to model the design process and to formalize the aesthetic measures that inform this process. In his seminal book, Birkhoff [2] formalized the aesthetic measure of an object as a ratio between its relative degree of order and its complexity ($M = O/C$, where M is aesthetic measure or value, O is aesthetic order, and C is complexity). Birkhoff's approach can be applied to examine various visual objects, such as polygons. Because, however, computer interface or graphic displays often comprise dozens, or even hundreds, of separate objects, rather than just a single geometric shape, Ngo and his colleagues [20-22] elaborated Birkhoff's work by formalizing fourteen aesthetic measures especially for evaluating interface design.

Biologically inspired computer vision

Over the last 20 years, the field of computational image analysis has achieved significant progress. The goal of computer vision is to extract meaningful information, such as patterns or objects, from images by computing their statistical features. The biologically inspired variant of computer vision simulates aspects of the human brain by focusing on those image features that are perceptually relevant. More recently, this approach has been applied to understand how users perceive graphical displays or interfaces. Rosenholtz et al. [26], for example, showed that statistical features of such images (e.g., color or texture entropy) correlate significantly with users' perceptual ratings of the clutter immanent in a design.

METHODS

Users' aesthetic and affective judgments

Participants

Twenty-two full-time staff and student interns from Siemens Corporate Research naïve to the goals of the study volunteered to participate without compensation. These included 16 men and 6 women, with an average age of 26.3 (*SD* = 4.2). All participants had normal or corrected-to-normal vision, and were free of color blindness.

Materials

Thirty web pages were selected from various sources on the Internet. The web pages differed in their visual design quality: Examples of good design included the winners of the Webby Awards (www.webbyawards.com), and examples of bad design were selected using Google searches with key words like "bad/poor visual design," etc. To minimize any preconceptions the participants might have had about any of the web pages, popular web pages or high traffic sites (e.g., apple.com, facebook.com) were excluded; web pages containing emotional objects – such as a baby's face – or familiar objects – such as the iPhone – were likewise excluded. Apart from these exclusions, the remaining pages were included indiscriminately (on a quasi-first-come-first-serve basis), without any particular bias toward one attribute or another, because the nature of the present study is largely exploratory.

For the same reason, the pages included in the study were not further edited, which meant, for example, that they were not cropped to fill out the browser window completely. Instead, it was assumed that most users currently employ a 1280 x 1024 screen resolution, and that contemporary web pages are designed for a maximized browser window. Any deviation from this standard resulted in a potentially "unbalanced" layout, with vacant space to the right of a left-justified web page not optimized for viewing at 1280 x 1024 (see, e.g., Figure 1, bottom left image). Because these design decisions were presumed to be deliberate and because the inclusion of pages in the study was without reference to any aesthetic selection criteria, as described above, we elected not to "correct" the layout of these web pages, which we would have viewed as an indefensible tampering with the original stimulus material. Moreover, several of the web pages included in the study that did not fill the entire maximized browser window did take pains to incorporate the resulting vacant margins in an aesthetically pleasing manner (e.g., Figure 1, top left image). In other words, there are no accidents in design – on this premise, the decision not to balance out sub-optimally designed pages seemed entirely warranted.

Apparatus

A 21-inch Dell monitor was used to display the web pages with a resolution of 1280 x 1024 pixels. The program for displaying stimuli and collecting participants' ratings ran on a Dell Latitude D610 laptop. The program was custom written in Matlab using functionalities from the Psychtoolbox [3,25].

Procedure

Participants were tested individually. After reading an instruction sheet and providing demographic information, each participant was seated in front of a computer, at a distance to the display of about 70 cm.

Participants were instructed to fixate a small black cross at the center of the screen and press the space bar when ready to begin. Then, a web page would be shown for 150 ms. After the disappearance of the web page, a mask screen, which was a randomization of all the pixels from the previous screen, would appear for 500ms, followed immediately by a blank screen for 1,000 ms. (Examples of the masks are shown in Figure 1. These types of random-pattern masks are quite common in experimental psychology; they maintain the color and luminance

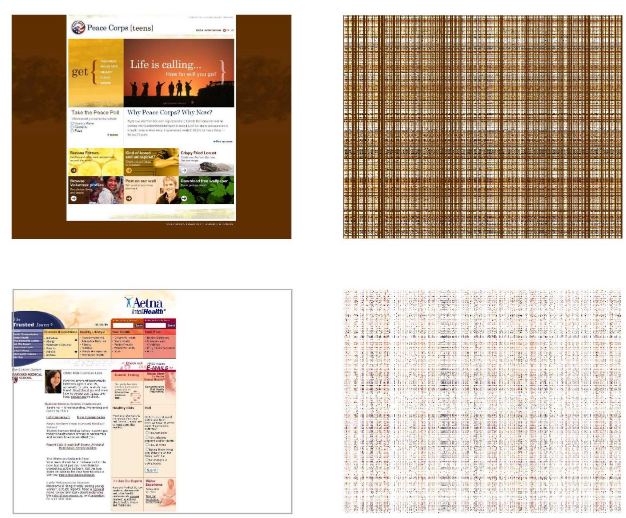

Figure 1. Two examples of masks used in the study (right), along with the images that preceded them (left).

histograms of the randomized image but do not engage any object-processing centers of the visual brain [14].) Following this delay, a screen with four rating criteria would appear. For each criterion, participants used a mouse to select a score on a Likert scale from one to seven, with one representing the most negative value on the corresponding dimension and seven the most positive value. Participants repeated these steps for a total of thirty web pages. The presentation order of the each web pages was randomized for each participant by means of a Latin Square. The total duration of the study was about 20 min.

The four rating criteria were selected from the AttrakDiff questionnaire, a fairly commonly employed scale developed by Hassenzahl, Burmester and Koller [10] to evaluate the attractiveness of a product or a user interface design systematically. The questionnaire comprises four dimensions: Attractiveness, Pragmatic Quality, Hedonic Quality: Identity, and Hedonic Quality: Stimulation (see [10] for details). We selected the most representative contrast from each dimension. These were, in order of the dimensions listed above: *repelling* vs. *appealing*, *complicated* vs. *simple*, *unprofessional* vs. *professional*, and *dull* vs. *captivating*. As described above, participants simply placed a tick-mark along a seven point scale that spanned from, for example, *repelling*, on one end, to *appealing*, on the other.

Aesthetic metrics based on low-level image statistics

Materials
The same thirty web pages rated by the human participants were used for the algorithmic analysis of the low-level image statistics.

Computing procedure
All images were analyzed in three steps (see Figure 2): First, low-level image statistics (i.e., color, intensity, and texture) were computed and discretized. Second, based on the histogram results generated in the first step, images were decomposed into regions of para-threshold image statistics (e.g., regions of minimum entropy); this was done via quadtree decomposition [7]. Lastly, the resulting quadtree was evaluated on the aesthetic dimensions (adapted from [22]): *symmetry*, *balance*, *equilibrium*, and number of decomposition regions. A more detailed description of the quadtree algorithm and definitions of the three aesthetic dimensions follow. More precise, mathematical definitions are given in the Appendix (also, see [22]). An example of a quadtree decomposition is shown in Figure 3, overlaid on the decomposed image. The decomposition uses the same algorithm as the present study; but the image shown was excluded from the study because the MSN site is too well known. It illustrates particularly vividly, however, how certain image regions are broken down further whereas others are not.

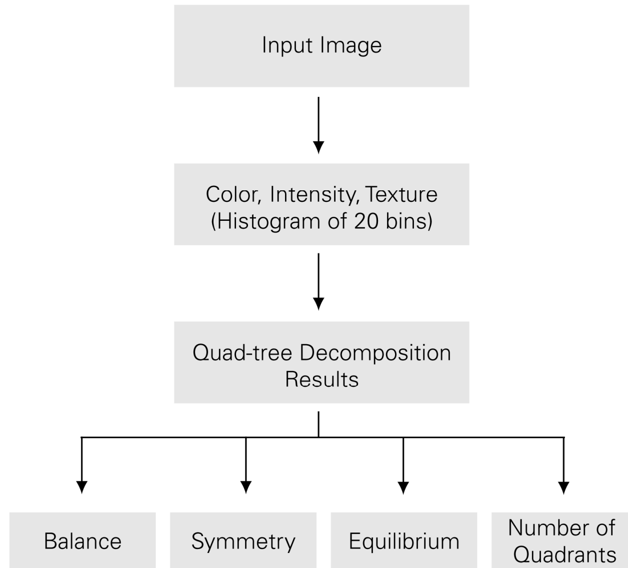

Figure 2. General procedure for deriving the structural dimensions from low-level image statistics (see text for details).

Low-level image statistics
Color was measured in the RGB color space, which was projected into a continuous RGB color cube with 20 discrete cubic bins using the minimum variance quantization method (see www.mathworks.com – Image Processing Toolbox: Color: Reducing the Number of Colors in an Image). That is, a three-dimensional space with volumetric bins was constructed with the three dimensions of the RGB scale as its three orthogonal axes. Color entropy was then calculated from the resulting color histogram, just as it would have been from a conventional one-dimensional histogram.

Intensity was measured as the luminance of the image in the L*A*B* color space (CIELAB). The continuous luminance values were likewise transferred into 20 discrete intensity-level bins. Intensity entropy was then computed from the resulting intensity histogram.

Texture was measured in terms of a texton histogram [18] that was derived by extracting texture features from the image by applying a set of linear orientation and spatial frequency filters, then clustering the resulting features into 20 levels of textons.

Quadtree decomposition
The quadtree decomposition partitions an image into regions that contain equal amounts of information in the Information Theoretic sense [27]. Quadtrees, which have been used since the mid-70s, have found common application in a variety of different domains, such as image representation, spatial indexing, collision detection,

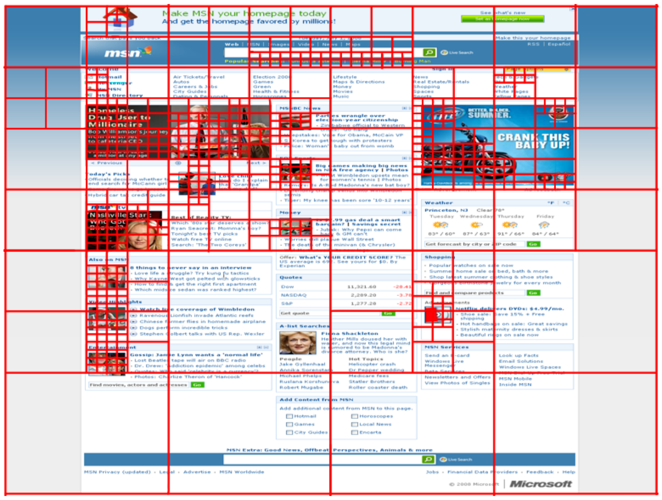

Figure 3. Example of a quadtree decomposition, overlaid on the original image. (See text for details.)

the storage of sparse data, and computational fluid dynamics, and are not new to our study. The quadtree algorithm proceeds to divide an image recursively into quadrants until all resulting quadrants do not permit further division by virtue of having reached some criterion on a particular dimension. For example, if the criterion is minimum color entropy (color entropy is a measure of the "predictability" of a color in a given region), the algorithm will first divide the image into four quadrants, then subdivide each of those quadrants further into quadrants that still evidence a certain degree of unpredictability in their color composition. A homogeneously blue image would, therefore, never be subdivided into quadrants to begin with. Figure 3 illustrated how the recursive partitioning of an image into quadrants leaves some regions undivided – such as the bottom row – and parses other regions further and further. Evidently, the upper regions of the image exhibited a higher degree of texture entropy than the remainder of the image. As becomes apparent from this example, quadtree decomposition provides a representation of the spatial layout of information in an image. That is, regions with the greatest density of information will contain the largest number of "leaves" (child quadrants). The formal algorithm for the quadtree decomposition is given in the Appendix.

Aesthetic dimensions

Balance is defined as the distribution of optical weight in an image, whereby optical weight is the perception of the "heaviness" of the elements of a scene or image. An unbalanced scene has an unequal weight distribution between two concomitant image pairs – top versus bottom or left versus right. *Balance* can be conceptualized as a comparison of the inventory of elements on either side of the vertical or horizontal meridian of an image. As such, it is related to axial symmetry, but less stringent in its requirements. In Figure 4, for example, the left and right sides of the image comprise the same elements, albeit in a different spatial ordering, from top to bottom. Owing to the concordance in composition between the two sides of the image, the image would be regarded as high on the dimension *balance*.

Symmetry is measured in terms of two components: axial and radial symmetry. Axial symmetry is a function of the two major axes – vertical and horizontal – whereas radial symmetry considers the symmetry across any other axis orientation with two or more axes that intersect at the central point of the image. In Figure 4, the well balanced image is only moderate in symmetry because the imperfect correspondence between elements on either side disrupts the symmetry between these elements.

Equilibrium measures the centering of an image around its midpoint. It is computed as the difference between the center of mass of the elements in the image and the physical center of the image. Simply put, *equilibrium* can be regarded as a measure of the "lopsidedness" of an image. In Figure 1 (bottom left image), the example image is low on the *equilibrium* dimension because the center of mass of the elements in the image is clearly to the top left of the physical midpoint of the image. As the Figure indicates, *equilibrium* is strongly affected by distance from the physical midpoint, so that two disproportionately massive objects at unequal distances from the midpoint might still generate high *equilibrium*.

RESULTS

Pearson correlations were calculated to examine the degree of relationship between each of the image structural dimensions (i.e., *balance*, *symmetry*, etc.) and the four aesthetic dimensions (i.e., *repelling-appealing*, *complicated-simple*, etc.). Prior to the correlation analysis, diagnostic tests were performed to validate the consistency of the participants' aesthetic judgments for each of the thirty web pages. Three web page designs, which generated controversy amongst different participants (values were 2 SD above the group mean), were eliminated from a further analysis of the data. For the ensuing correlation analysis with the remaining 27 web pages, the significance level was set to α=.05, two-tailed (to remain agnostic about the direction of the correlation). In the following, only significant correlations are reported.

The overall pattern of results is summarized in Figure 5 and will be discussed in detail in the next section (DISCUSSION);

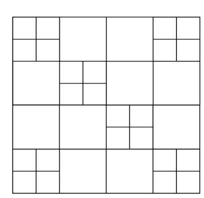

Figure 4. A pattern high in *balance* but moderate in *symmetry*.

	Repelling vs. Appealing	Complicated vs. Simple	Unprofessional vs. Professional	Dull vs. Captivating
Balance	✓		✓	✓
Symmetry	✓			✓
Equilibrium				
# of Quadrants		✓	✓	

Figure 5. Schematic of overall pattern of results. Check marks indicate statistically significant correlations.

Figure 6 provides further detail by showing the majority of the significant correlations and presenting examples of representative images for select points along both axes. Figure 6 importantly illustrates that any significant correlations reported in the following were not inflated by a potential bimodal distribution of the data along one axis; instead, the data points appear to be distributed relatively evenly along both axes.

The data revealed that *balance* (computed based on color entropy) was positively correlated with the *repelling-appealing* dimension [$r(25) = .571$, $p = .002$], with the *unprofessional-professional* dimension [$r(25) = .459$, $p = .016$], and with *dull-captivating* dimension [$r(25) = .630$, $p < .001$]. *Balance* (computed based on intensity entropy) was likewise positively correlated with the *repelling-appealing* dimension [$r(25) = .602$, $p < .001$], with the *unprofessional-professional* dimension [$r(25) = .486$, $p = .014$], and with *dull-captivating* dimension [$r(25) = .666$, $p < .001$]. Finally, *balance* (computed based on texture entropy) showed the same pattern of correlations as the previous two derivations of *balance* (based on color and intensity entropy, respectively). That is, the it correlated positively with the *repelling-appealing* dimension [$r(25) = .631$, $p < .001$], with the *unprofessional-professional* dimension [$r(25) = .530$, $p = .004$], and with the *dull-captivating* dimension [$r(25) = .715$, $p < .001$].

The structural dimension *symmetry* (computed based on color entropy), in turn, correlated positively with the *repelling-appealing* dimension [$r(25) = .451$, $p = .018$], as well as the *dull-captivating* dimension [$r(25) = .550$, $p < .003$]. *Symmetry* (computed based on intensity entropy) was positively correlated with the *dull-captivating* dimension [$r(25) = .456$, $p = .023$]. Finally, *symmetry* (computed based on texture entropy) was positively correlated with the *dull-captivating* dimension [$r(25) = .386$, $p = .047$].

Whereas the structural dimensions were all correlated positively with the aesthetic dimensions, the final number of 'leaves' following a complete quadtree decomposition (labeled as "Number of Quadrants" in Figure 5) yielded only negative correlations. Specifically, it was significantly correlated (when based on color entropy) with the dimension *complicated-simple* [$r(25) = -.441$, $p = .021$] and with the dimension *unprofessional-professional* [$r(25) = -.384$, $p = .048$].

DISCUSSION

What is encouraging about the results presented above is that the correlations observed were as high as they were (and, moreover, statistically significant). It is therefore possible to begin interpreting what each of these correlations might signify to a designer. What is furthermore encouraging about these results is that – as reflected in Figure 5 – the abstract structure that was extracted from the images by way of the quadtree composition mapped onto the various aesthetic dimensions in a consistent manner, independent of the featural dimension in which this structure was expressed in the original image. In other words, it seemed to matter little whether an image was balanced because this balance emerged from, say, the spatial distribution of luminance entropy in the image or from the spatial distribution of color entropy. The different featural origins of *balance*, *symmetry*, etc., could therefore be aggregated into a single row in Figure 5 for each of these dimensions.

As can clearly be seen by the positive correlations shown in Figures 5 and 6, the results were quite systematic, with *balance* correlating with the largest number of aesthetic dimensions and *equilibrium* with none of them. Why *equilibrium* might fail to correlate with any of the low-level featural dimensions should become clear if one considers that all the image materials used for the study were web pages. Because most web pages are – almost inherently, but at least by overwhelming convention – not organized around the midpoint of the screen, any faint traces of *equilibrium* immanent in the design of the study images was largely irrelevant to users' judgment of these images in

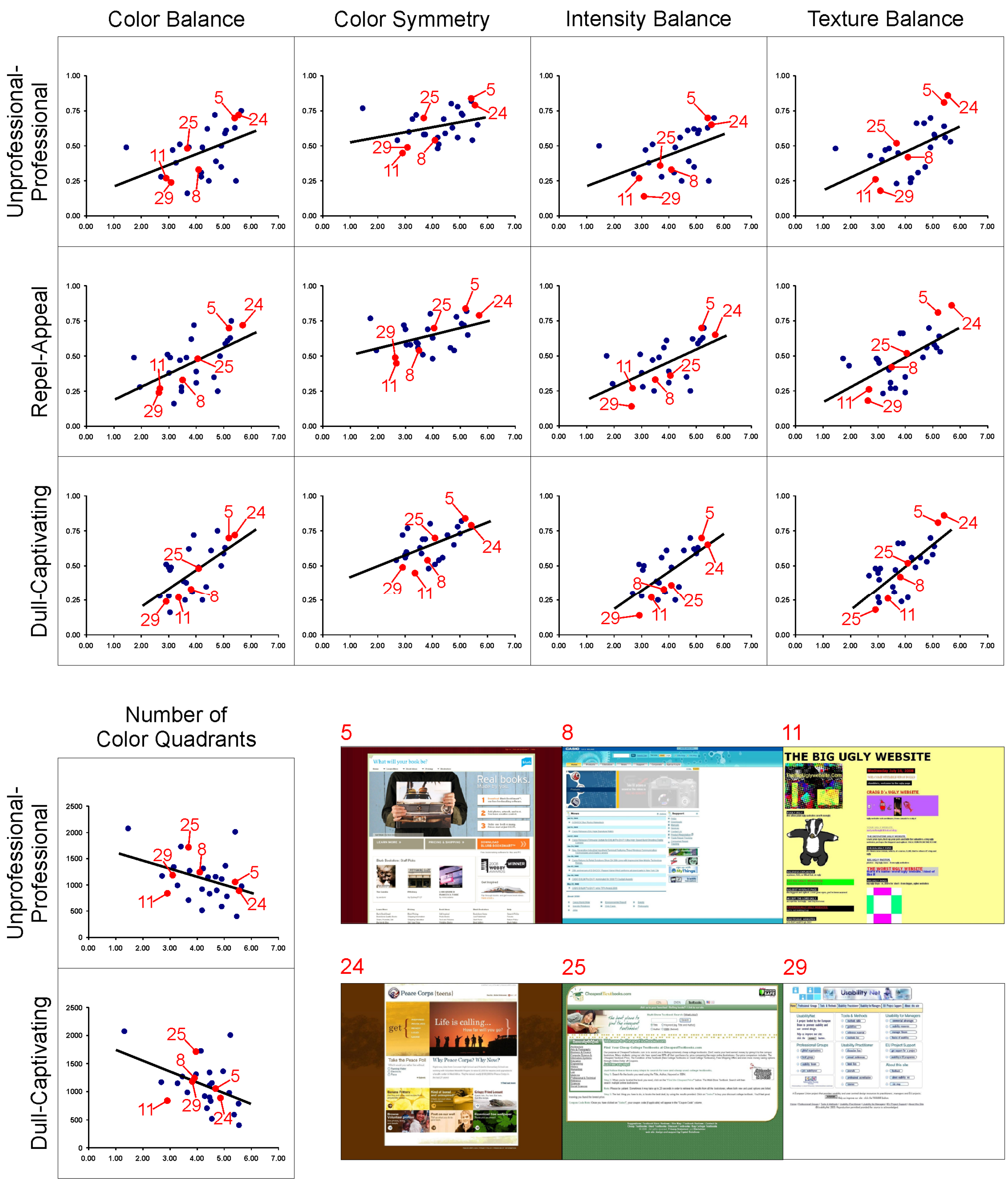

Figure 6. Individual correlations (scatter plots) and representative images for six data points.

terms of their appeal, simplicity, professionalism or capacity to captivate.

Balance, by contrast, could reasonably be expected to have had a large impact on these same judgments. As Figure 6, #11 demonstrates, which shows the image judged to be the least balanced or symmetrical, a web page that does not distribute its visual information content (in the Information Theoretic sense) evenly across the image simply, and quite evidently, *looks* unprofessional and unappealing. For comparison purposes, Figure 6, #24 shows the image judged to be the most balanced and symmetrical. This image adheres to the principles of good design and organizes its image contents in a systematic and balanced fashion, giving it an appealing, professional and captivating appearance. The fact that this image, or other images in which these aesthetic dimensions were judged as high, is not also viewed as simple argues that a moderately – but not overly – complicated image will still be regarded as appealing and professional while not sacrificing its ability to captivate. Owing to these competing requirements for a professional and appealing, yet captivating, web page to hover somewhere between simplicity and overwhelming visual complexity, none of the correlations with either *balance*, *symmetry*, or *equilibrium* are, therefore, predictably significant.

Unlike *balance*, *symmetry* appears to be less important for the judgment of a web page as professional or unprofessional. The possible reason can easily be understood by reminding oneself that *symmetry* is a more stringent requirement than *balance*, and that these two dimensions are in some ways related. Whereas *symmetry* could be argued to be a largely aesthetic attribute, somewhat pedantically matching left side *literally* with right side, and top *literally* with bottom, *balance* merely ensures that the perceptual (or, optical, in Ngo's terms [20-22]) weight of an image is equally distributed across the different corresponding quadrants of the display (left-right, top-bottom). As such, it reflects certain deliberateness and thoughtfulness and level of effort to balance out the information content of a web page, rather than somewhat carelessly juxtaposing high (visual) information content on one side of the display with an essentially informationally vacant opposite side of the display.

Whereas *symmetry*, *balance* and *equilibrium* speak to the distribution of (visual) information in the image, "Number of Quadrants" in Figure 5 is simply the raw count of final number of 'leaves' following a completed quadtree decomposition, independent of the spatial distribution of these 'leaves'. Unlike the spatially-specific dimensions, Number of Quadrants, therefore, not surprisingly, correlates (negatively) with the *complicated-simple* dimension, such that images with a greater number of 'leaves' are regarded as more complicated. Correspondingly, they are also viewed as less professional. At first glance, this results may appear to contradict the lack of correlation of any of the other structural dimensions – exempting *balance* – with the *unprofessional-professional* valuation, on the one hand, and the lack of correlation of these valuations with the *complicated-simple* dimension, on the other. Yet, herein lies precisely the solution to making designs both captivating *and* professional: It is not the raw informational complexity of the design per se that leads a design to be rejected as unprofessional or complicated or repelling. As long as the designs are still balanced and/or symmetrical (potentially high on *equilibrium* – although see above), they will continue to be perceived as professional, simple, etc. However, at the same time, the high number of 'leaves' – and therefore raw informational content – allows these designs not to be dull.

Conclusions

The present results illustrate vividly that the conjecture, based on psychophysical evidence (e.g., [24,26,35]), that even seemingly high-level judgments about the aesthetic and affective aspects of an image are likely computed by the brain using low-level features that are agnostic about objects, meanings, cultural contexts, etc. This result is both sobering from a humanist perspective and promising from a design perspective: It means that, contrary to Ngo et al.'s pessimism about the generalizability of their formal model, it is possible to develop computer algorithms that are capable of delivering quick and dirty evaluations of subjective aspects of any (image-based) design.

These evaluations can help to reduce cost significantly in early iterations of the design process by serving as a substitute for full-fledged empirical evaluations with human users. They are also capable of providing guidance to designers and afford them the freedom to explore various design alternatives without the fear of cost overrun or painful delays in delivery. The algorithm described in the present paper, as well as future refinements and expansions, are not intended to replace careful evaluations of designs by human users, nor are they intended to undermine the profound experience and consummate skill of a well trained designer.

ACKNOWLEDGMENTS

We thank CHI, PDC and CSCW volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. We also thank Charles Law for the creation of figures, and Sarah Pearson for assisting with the collection of the empirical data.

REFERENCES

1. Arnheim, R. *Art and visual perception*. University of California Press, Berkeley, USA, 1954.
2. Birkhoff, G. D. *Aesthetic measure*. Harvard University Press, Cambridge, MA, USA, 1933.
3. Brainard, D. H. The Psychophysics toolbox. *Spatial Vision 10*, 4 (1997), 433-436
4. Di Lollo, V., Enns, J. T. and Rensink, R. Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General 129*, 4 (2000), 481-507.
5. Dondis, D. A. *A primer of visual literacy*. MIT Press, Cambridge, MA, USA, 1973.
6. Fabre-Thorpe, M., Delorme, A., Marlot, C. and Thorpe, S. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neurosciences 13*, 2 (2003), 171-180.

7. Finkel, R. A. and Bentley, R. A. Quad trees: A data structure for retrieval on composite keys. *Acta informatica 4*, 1 (1974), 1-9.
8. Galitz, W. O. *The essential guide to user interface design: An introduction to GUI design principles and techniques*. Wiley Technology, Indianapolis, IN, USA, 2007.
9. Gladwell, M. *Blink: The Power of Thinking Without Thinking*. Little, Brown and Company, New York, NY, USA, 2005.
10. Hassenzahl, M., Burmester, M. and Koller, F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [A questionnaire for the measurement of perceived hedonic and pragmatic attributes]. In *Mensch & Computer 2003*, B. G. Teubner (2003), 187-196.
11. Hubel, D. H. and Livingstone, M. S. Segregation of form, color, and stereopsis in primate area 18. *Journal of Neuroscience 7*, 11 (1987), 3378-3415.
12. Hubel, D. H. and Wiesel, T. N. Brain mechanisms of vision. *Scientific American 241*, 3 (1979), 150-162.
13. Koffka, K. *Principles of Gestalt psychology*. Lund Humphries, London, UK, 1935.
14. Kourtzi, Z. and Kanwisher, N. Cortical regions involved in perceiving object shape. *The Journal of Neuroscience 20*, 9 (2000), 3310-3318.
15. Kurosu, M. and Kashimura, K, Apparent usability vs. inherent usability. In *CHI 1995 Conf. Companion*, ACM Press (1995), 292-293.
16. Lamme, V. A. F. and Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences 23*, 11 (2000), 571-579.
17. Lindgaard, G., Fernandez, G., Dudek, C. and Brown, J. Attention web designers: You have 50 milliseconds to make a good impression. *Behavior & Information Technology 25,* 2 (2006), 115-126.
18. Leung, T. and Malik, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision 43*, 1 (2004), 29-44.
19. Mathworks Image Processing Toolbox. www.mathworks.com – Image Processing Toolbox: Color: Reducing the Number of Colors in an Image
20. Ngo, D. C. L. *Visit: Visitor information system implementation tool*. Ph.D. Thesis, Trinity College, Dublin, Ireland, 1994.
21. Ngo, D. C. L. and Byrne, J. G. Aesthetic measures for screen design. In *Proc. OzCHI 1998*, IEEE Computer Society (1998), 64-71.
22. Ngo, D. C. L., Teo, L. S. and Byrne, J. G. Modelling interface aesthetics. *Information Sciences 152* (2003), 25-46.
23. Norman, D. *Emotional design: Why we love (or hate) everyday things*. Basic Books, New York, NY, USA, 2004.
24. Oliva, A and Torralba, A. Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research: Visual perception 155*, (2006), 23-36.
25. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision 10*, 4 (1997), 437-442.
26. Rosenholtz, R., Li, Y, Mansfield, J. and Zhenlan, J. Feature congestion: A measure of display clutter. In *Proc. CHI 2005*, ACM Press (2005), 761-770.
27. Shannon, C. E. A mathematical theory of communication. *The Bell Systems Technical Journal 27* (1948), 379-423, 623-656.
28. Tidwell, J. *Designing interfaces: Patterns for effective interaction design*. O'Reilly Media, Sebastopol, CA, USA, 2005.
29. Tractinsky, N. Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In *Proc. CHI 1997*, ACM Press (1997), 115-122.
30. Tractinsky, N., Katz, A. S. and Ikar, D. What is beautiful is usable. *Interacting with Computers 13,* 2 (2000), 127-145.
31. Treisman, A. M. and Gelade, G. A feature-integration theory of attention. *Cognitive Psychology 12*, 1 (1980), 97-136.
32. Wang, J. Z., Li, J. and Wiederhold, G. SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 9 (2001), 947-963.
33. Zajonc, R. Feeling and thinking: Preferences need no Inferences. *American Psychologist 35*, 2 (1980), 151 – 175.
34. Zheng, X. S., Chakraborty, I., Lin, J. J. W. and Rauschenberger, R. Developing quantitative metrics to predict users' perceptions of interface design. In *Proc. HFES*, HFES (2008), in press.
35. Zipser, K., Lamme, V. A. F. and Schiller, P. H. Contextual modulation in primary visual cortex. *The Journal of Neuroscience 16*, 22 (1996), 7376-7389.

APPENDIX

Quadtree decomposition

Step 1: Select a decomposition feature X and threshold T on a 2D rectangular image I.

Step 2: Compute the feature for image X(I).

Step 3: Decretize X(I) into a predetermined number of bins $\{X_1, X_2, \ldots X_B\}$.

Step 4: Compute the associated entropy H (X(I)).

$$H(X(I)) = \sum_{i=1}^{B} p_i * \log(1/p_i)$$

$$P_i = P(x_i \in X_i) = \#(x_i \in X_i)/n$$

where n is the number of pixels in the image.

Step 5: Compare the feature entropy to the threshold. If H (X(I)) > T, decompose I into four rectangular quadrants. For each quadrant, repeat from step 2.

Balance

$$B = 1 - \frac{IB_{top-bottom} + IB_{left-right}}{2} \in [0,1]$$

$$IB_{top-bottom} = \frac{w_{top} - w_{bottom}}{\max(|w_{top}|, |w_{bottom}|)}$$

$$IB_{left-right} = \frac{w_{left} - w_{right}}{\max(|w_{left}|, |w_{right}|)}$$

where

$$w_j = \sum_{i}^{n_j} a_{ij} d_{ij}\ , j = left, right, top, bottom$$

Symmetry

$$SYM = 1 - \frac{ASYM_{vert} + ASYM_{horiz} + ASYM_{rad}}{3}$$

$$ASYM_{vert} = (X_{U(l-r),L(l-r)} + Y_{U(l-r),L(l-r)} + H_{U(l-r),L(l-r)} + \\ + B_{U(l-r),L(l-r)} \theta_{U(l-r),L(l-r)} + R_{U(l-r),L(l-r)})/12$$

$$ASYM_{horiz} = (X_{(U-L)l,(U,L)r} + Y_{(U-L)l,(U,L)r} + H_{(U-L)l,(U,L)r} + \\ + B_{(U-L)l,(U,L)r} \theta_{(U-L)l,(U,L)r} + R_{(U-L)l,(U,L)r})/12$$

$$ASYM_{rad} = (X_{(U-L)l-r,(U-L)r-l} + Y_{(U-L)l-r,(U-L)r-l} + \\ + H_{(U-L)l-r,(U-L)r-l} + B_{(U-L)l-r,(U-L)r-l} + \\ + \theta_{(U-L)l-r,(U-L)r-l} + R_{(U-L)l-r,(U-L)r-l})/12$$

where, if T is any term

$$T_{U(l-r),L(l-r)} = |T_{Ul} - T_{Ur}| + |T_{Ll} - T_{Lr}|$$

$$T_{(U-L)l,(U,L)r} = |T_{Ul} - T_{Ll}| + |T_{Ur} - T_{Lr}|$$

$$T_{(U-L)l-r,(U-L)r-l} = |T_{Ul} - T_{Lr}| + |T_{Ur} - T_{Ll}|$$

To avoid confusion in the above notations, Upper and Lower shall be denoted as U and L, respectively, and left and right as l and r, so that subscript Ul means upper left, Lr means lower right, etc.

$$X_j = \sum_{i}^{n_j} |x_{ij} - x_c|, \quad j = Ul, Ur, Ll, Lr$$

$$Y_j = \sum_{i}^{n_j} |y_{ij} - y_c|, \quad j = Ul, Ur, Ll, Lr$$

$$H_j = \sum_{i}^{n_j} h_{ij}\ , \quad j = Ul, Ur, Ll, Lr$$

$$B_j = \sum_{i}^{n_j} b_{ij}\ , \quad j = Ul, Ur, Ll, Lr$$

$$\theta_j = \sum_{i}^{n_j} \left| \frac{y_{ij} - y_c}{x_{ij} - x_c} \right|, \quad j = Ul, Ur, Ll, Lr$$

$$R_j = \sum \sqrt{(x_{ij} - x_c)^2 + (y_{ij} - y_c)^2}\ , \quad j = Ul, Ur, Ll, Lr$$

Equilibrium

$$EM = 1 - \frac{|IE_x| + |IE_y|}{2} \in [0,1]$$

$$DE_x = \frac{2\sum_i^n a_i (x_i - x_c)}{nb_{frame} \sum_i^n a_i}$$

$$DE_y = \frac{2\sum_i^n a_i (y_i - y_c)}{nh_{frame} \sum_i^n a_i}$$

Where (x_i, y_i) *and* (x_c, y_c) *are the coordinates of the object center and image center, respectively, and* a_i *is the object area.*