International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: http://www.tandfonline.com/loi/hihc20

Application of Object-Based Metrics for Recognition of Well-Designed Dashboards

Jiří Hynek & Tomáš Hruška

To cite this article: Jiří Hynek & Tomáš Hruška (2018): Application of Object-Based Metrics for Recognition of Well-Designed Dashboards, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2018.1518004

To link to this article: https://doi.org/10.1080/10447318.2018.1518004

Published online: 19 Sep 2018.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2018.1518004

Application of Object-Based Metrics for Recognition of Well-Designed Dashboards

Jiří Hynek and Tomáš Hruška

IT4Innovations Centre of Excellence, Brno University of Technology, Faculty of Information Technology, Brno, Czech Republic

ABSTRACT

Measuring the characteristics of visually emphasized objects displayed on a screen seems to be a promising way to rate user interface quality. On the other hand, it brings us problems regarding the ambiguity of object recognition caused by the subjective perception of the users. The goal of this research is to analyze the applicability of chosen object-based metrics for the evaluation of dashboard quality and the ability to distinguish well-design samples, with the focus on the subjective perception of the users. This article presents the model for the rating and classification of object-based metrics according to their ability to objectively distinguish well-designed dashboards. We use the model to rate 13 existing object-based metrics of aesthetics. Then, we present a new approach for the improvement of the rating of one object-based metric—Balance. We base the improvement on the combination of the object-based metric with the pixel-based analysis of color distribution on the screen.

1. Introduction

"Dashboard" is a frequently used term connected with business intelligence and management information systems. It is a favorite tool used by many organizations to comprehensively present their data for operational, analytical, or strategic purposes. It presents key performance indicators which help to evaluate the progress and benefit of business activities (Eckerson, 2006). Since dashboards support decision-making, they have become popular among a wide range of users for the management of personal activities. Lately, there have been an increasing number of web applications providing dashboard templates to visualize data gathered from common services like social networks. The rising diversity of dashboards has led designers to think about the principles of high-quality dashboard design.

The first rules which brought some clarity to dashboard characteristics were provided by Stephen Few (Few, 2006). He defined the dashboard as "a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance." He pointed out that most of the existing so-called dashboards break with this definition. They are usually not able to present information on a single screen (the user needs to interact and look for the information —e. g., scroll or go through pages). Few has provided a framework based on a knowledge of famous books regarding design and graphics (e. g., Tufte, 2001; Ware, 2012). This framework contains rules and advice for the dashboard design, including examples of well-designed dashboards.

Even more than 10 years after the release of Few's publication, we can still observe that the majority of dashboards ignore Few's rules and advice or express them in their own way. We assume that the reason might be the complexity and vague definition of the framework and the lack of other sources which would provide formal and quantitative knowledge in the area of dashboard design. The requirement of the dashboard—"present information on a single screen"—is what distinguishes dashboards from other interfaces and, also, makes them difficult to design. Designers need to focus on the design aspects such as strong simplification, the elimination of unnecessary elements, highlighting significant relationships between data, or the careful selection of graphical elements capable of comprehensively presenting a great deal of data using a small area. The dashboard designer needs to be a person with experience in human–computer interaction and capable of correctly applying the framework. The presence of users is usually required to evaluate the usability, which increases the time and expenses of the design phase.

A major challenge in improving dashboard design is that of finding measurable characteristics which would detect some of the design problems and help to distinguish well-designed interfaces from poorly designed ones. Such characteristics could be measured automatically during the early design phase without the presence of users. For example, Hynek and Hruška (2016) showed that measuring the average colorfulness of dashboards could help to distinguish dashboards designed according to Few's framework. However, it is not usually simple to describe the complex design attributes of a user interface, as they usually depend on the subjective judgment of the viewer. The metrics are usually simple, focused on simple visual attributes.

CONTACT Jiří Hynek ihynek@fit.vut.cz; hruska@fit.vut.cz IT4Innovations Centre of Excellence, Brno University of Technology, Faculty of Information Technology, Božet?chova 1/2 612 66 Brno, Czech Republic.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

One possible step in improving the metric-based evaluation is to process a screen similarly as it is perceived by human brain—not as a matrix of pixels but as a group of objects within a scene as described by Baker et al. (2009). Then, we evaluate the objects (*widgets*) in the screen (usually represented by their boundaries) and their properties (e.g., size or position) (Charfi et al., 2014). For that purpose, we use *object-based metrics*.

The main weakness of the applicability of object-based metrics is the ambiguous definition of the object. For instance, we can consider the object as a boundary of every single graphical element in a screen parsed from the structural description of a screen (Purchase et al., 2011). Also, we can use a segmentation algorithm to detect visually dominant regions representing the area of the object (Reinecke et al., 2013). Otherwise, we can specify regions manually according to our perception (Zen & Vanderdonckt, 2014). The problem with the first two approaches is that they do not usually consider objects with the same complexity as people perceive them (e.g., Gestalt laws (Ware, 2012)). The subjective perception of selected users limits the third approach. Two users will most likely specify regions in a slightly different way (Figure 1). Using two such descriptions of regions as independent inputs for an object-based metric gives us two (probably different) values describing one visual characteristic of the same dashboard. The question is: How much does the ambiguity of user perception influence the supposed agreement about dashboard characteristics measured by an object-

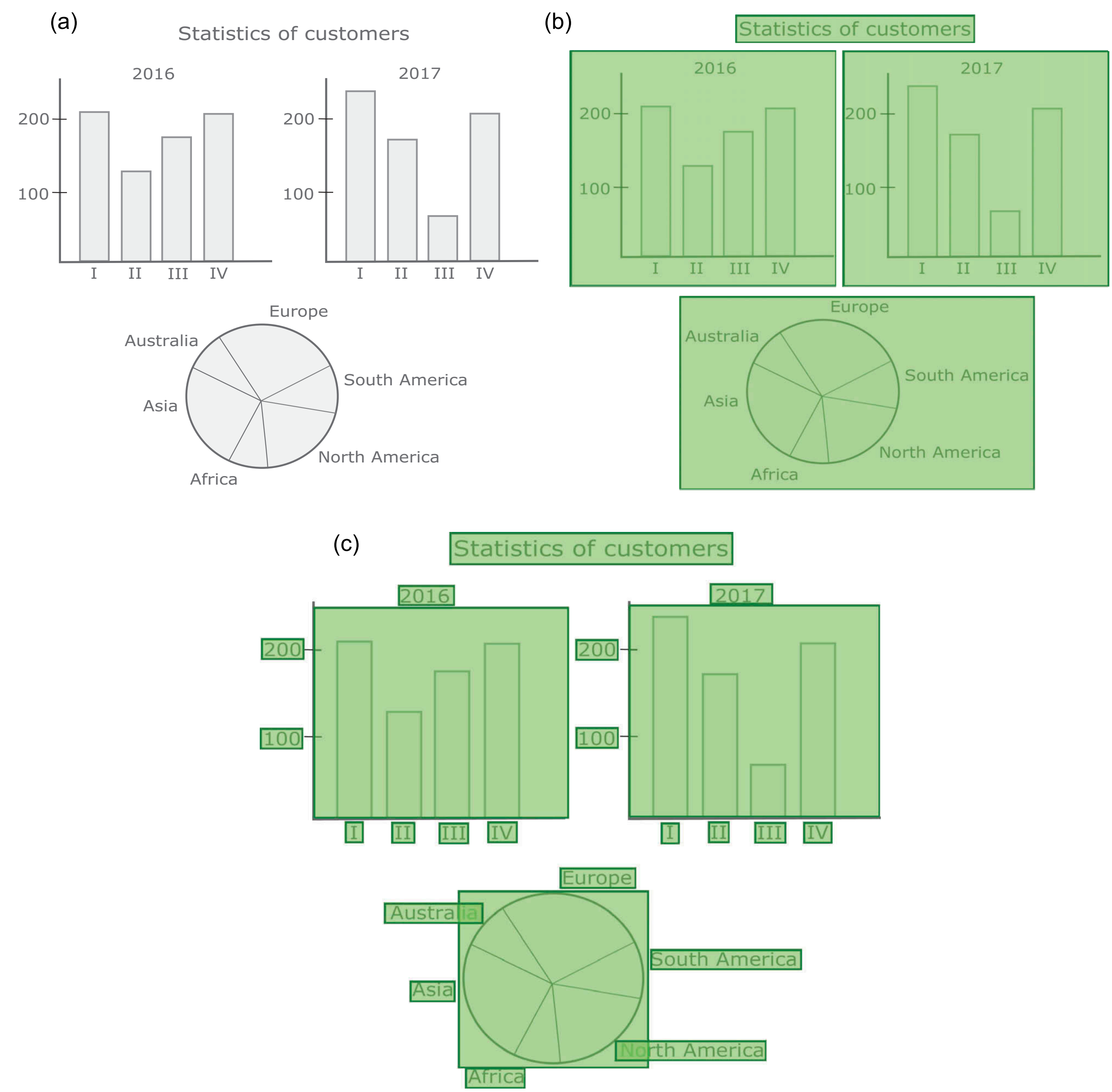

Figure 1. Example of the two different ways (b, c) of subjective perception of objects in the dashboard (a). The perceived objects are specified by the rectangular boundaries (regions), which are used as inputs for object-based metrics (such as Balance or Symmetry).

based metric? Since dashboards usually consist of complex widgets, we expect the ambiguity to be non-negligible. The hypothesis is that there are object-based metrics which should not be used for the evaluation of dashboard quality because they are not able to sufficiently consider the perception ambiguity.

The goal of this research is to explore and confirm this hypothesis and propose a solution to the problem of perception ambiguity. The research provides a brief state of the art regarding the perception of interface objects and existing object-based metrics. We focus on the 13 frequently used metrics of aesthetics presented by (Ngo, Teo, & Byrne, 2000). Then, we present three experiments. The first experiment collects data describing the user perception of the objects in dashboards and observes their similarities and deviations. The next experiment uses the data to analyze the applicability of Ngo's metrics for measuring the dashboard quality. For this purpose, we have established a model used for the evaluation and classification of metrics. The last experiment proposes an improvement of selected object-based metrics based on the combination of the object-based approach with the pixel-based approach, demonstrated on the *Balance* metric. The results of this research should help to improve the evaluation of the dashboard quality and the screen design in general.

2. Visual perception of the dashboard screen

The purpose of the dashboard defined in the Introduction is to display relevant information. However, the dashboard itself is not represented by information but only by data. The user is the one who connects the data with meaning and creates information (Eckerson, 2006). Data exploration tasks characterize the process of the data transformation. In the beginning, the viewer reacts to the light by visual receptors—the rod and cone cells located in the human eye. The light is then transformed into an electrical signal which is transferred to the brain by the optic nerves (Gibson, 1950). The brain initially perceives the visual signal and constructs an image of the perceived data (recognition of objects such as points, edges, shapes, or patterns and the comparison thereof). Then, the brain tries to comprehend the recognized objects, organize them, and add meaning to them. The second part is called *sensemaking*, and Baker et al. (2009) define it as "the ability to comprehend complex information, assimilate it, create order from it, and develop a mental model of the situation as a precursor to responding to the situation."

According to Baker et al. (2009), a visual representation improves sensemaking in data exploration tasks when it supports consistency with the viewer's knowledge, analogical reasoning, strong Gestalt properties, and the four basic visual perceptual approaches—association, differentiation, ordered, and quantitative perception. Viewers usually try to associate a perceived view with a previous experience or with a similar problem. Otherwise, they try to create a new interpretation of the perceived view and store it in their long-term memory. The quick recognition and comparison of objects can improve the sensemaking. The detection of the differences and similarities between the perceived objects plays a role in object ordering and grouping, which helps to simplify the perceived view.

Figure 2 points out that there are situations when the differences are perceived very quickly—without attention. These situations are consequences of *preattentive processing*—the perceptual task of object recognition—which is performed very quickly without the user's attention (in less than the time it takes the human eye to move, which takes about 200 milliseconds) (Ware, 2012). According to Healey, Booth, and Enns (1996), there are 17 preattentively perceived features which can be, according to Ware (2012), classified into four categories—color, form, spatial position, and motion. The appropriate usage of preattentive features can significantly decrease the time of dashboard sensemaking as shown by Few (2006).

Gestalt psychology describes the problem of object ordering and grouping (Few, 2006; Gibson, 1950; Ware, 2012). It helps us to understand the principles according to which people recognize objects (visual patterns) and cluster them into larger visual groups. It provides several laws (e.g., the laws shown in Figure 3). However, a missing mathematical model of Gestalt laws complicates conversion of the laws into computer algorithms to automatically predict how a user perceives the displayed screen. The problem of quantitative description of Gestalt laws is still the aim of researchers (Jäkel et al., 2016). In this research, we consider the fact that a viewer will finally group simple objects into more complex visually emphasized objects.

Visually emphasized objects together with background elements (larger scale, solid surfaces, and structures) make a scene of visual representation (Henderson & Hollingworth, 1999). Every object within the scene can be described by its visual characteristics (Baker, Jones, & Burkman, 2009). An appropriate choice and arrangement of objects within the scene are crucial for the interpretation of data by the viewer. They can emphasize various relations between data, yet they can skew or hide other facts (examples in (Tufte, 2001)). Hence, an analysis of the object characteristics within the scene can be useful during the design phase of a dashboard and user interface in general.

3. Measuring object characteristics

Measuring object characteristics became significant with the evolution of graphical user interfaces. In 1980s, designers used metrics to evaluate textual user interfaces (Smith & Mosier, 1986; Tullis, 1984). In the 1990s, they applied metrics in tools for the automatic design of user interfaces (Ivory & Hearst, 2001). Examples of the tools were presented by several researchers (Bodart et al., 1994; Kim & Foley, 1993; Mahajan & Shneiderman, 1997; Sears, 1993, 1995; Shneiderman et al., 1998). The usual goal of the tools was to analyze simple layout properties. Vanderdonckt and Gillo (1994) have published advanced techniques for the evaluation of screen layouts, divided into five groups: physical, composition, association (and dissociation), ordering, and photographic techniques. Since they described the techniques qualitatively, it was difficult to convert some of them to an algorithm.

x	y[u]	x	y[u]
1.	0.606	6.	7.985
2.	1.248	7.	5.549
3.	3.276	8.	3.276
4.	7.627	9.	1.595
5.	8.925	10.	0.457

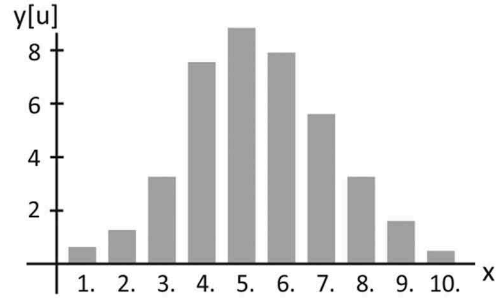

Figure 2. Two different visualizations of the same data. It is easier to compare the values if they are presented graphically (right) because of preattentive processing.

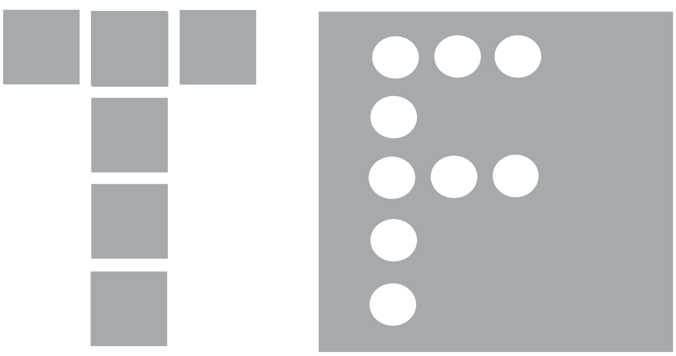

Figure 3. Example of Gestalt laws. The viewer will most likely see the two letters even though the letters consist of several shapes. This is the consequence of the Gestalt laws of continuity, similarity, and proximity.

The evolution of the internet and spreading of user interfaces among a broad spectrum of users have increased the importance of subjective feeling, satisfaction, and even the first impression of users. This aspect of an interface is usually perceived very quickly before a viewer fully understands the content of a user interface (Lindgaard et al., 2006). It often plays a significant role in the acceptance of a whole product. Moreover, it may improve interface usability (Kurosu & Kashimura, 1995; Tractinsky, Katz, & Ikar, 2000). Several articles (Kristeller, 1951; Lavie & Tractinsky, 2004; Tractinsky, 2004) describe this factor as *aesthetics* (from the Greek *aisthanesthai*—to perceive). The dictionary (Merriam-Webster, 2004) explains aesthetics as "pleasurable to the senses" or "attractive." Neilsen's usability framework categorizes the aesthetically designed user interface as "subjectively pleasing" (Nielsen, Usability engineering, 1994; Nielsen, Usability inspection methods, 1994).

In the early 2000s, Ngo et al. attempted to formally describe aesthetics (Ngo & Byrne, 2001; Ngo et al., 2000; Ngo, Teo, & Byrne, 2003). They presented the 13 quantitative object-based metrics of aesthetics described in Table 1. The metrics strongly correspond with the selected techniques published by Vanderdonckt and Gillo (1994). They analyze a screen as a set of rectangles (*regions*) representing the boundaries of interface objects. The regions are described only by their dimensions (size and position). The metrics use no further object characteristics (like color or shape). They analyze areas of regions, the distribution of regions in a screen, the aspect ratios of regions, and a level of screen granularity (number of regions, aligned points, or different sizes). The result of every metric is a value of the range $\langle 0, 1\rangle$. It represents a rate of an aesthetic factor. Figure 4 demonstrates an example of the Balance metric.

Table 1. List of Ngo's metrics of aesthetics (definitions used from (Ngo, 2001)). We classified the metrics according to their dependency on the following characteristics of a screen: AD—the area of regions and distribution of regions in the screen; AR—aspect ratio of regions, G—granularity of a screen (e.g., the number of regions).

Measure	Simplified definition used from Ngo (2001)	Dependence
Balance	Difference between total weighting of components on each side of horizontal and vertical axis	AD
Equilibrium	Difference between the center of mass of the displayed components and the physical center of the screen	AD
Symmetry	Extent to which the screen is symmetrical in three directions: vertical, horizontal, and diagonal	AD
Sequence	Measure of how information in display is ordered in a hierarchy of perceptual prominence corresponding to the intended reading sequence	AD
Cohesion	Extent to which the screen components have the same aspect ratio	AR
Unity	Extent to which visual components on a single screen all belong together	AD, G
Proportion	Comparative relationship of the dimensions of components to certain proportional shapes	AR
Simplicity	Extent to which component parts are minimized and relationships between the parts are simplified	G
Density	Extent to which the percentage of component areas on the entire screen is equal to the optimal level	AD
Regularity	Extent to which the alignment points are consistently spaced	G
Economy	Extent to which the components are similar in size	G
Homogeneity	Measure of how evenly the components are distributed among the quadrants	G
Rhythm	Extent to which the components are systematically ordered	AD

4. Ambiguity of object recognition

Numerous researchers have evaluated the applicability of Ngo's metrics to the present time, especially for website interfaces. They have usually based the evaluation of metrics on a comparison of the measured results with the reviews of p users who rated n user interfaces. Their results depend on a selected group of users, analyzed user interfaces and approaches to the specification of interface regions. We have

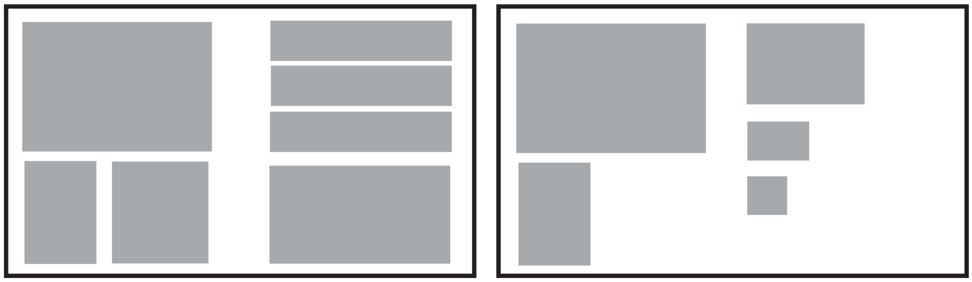

Figure 4. Example of two screens which can be compared using the Balance metric. The left screen is balanced, since the weight of the regions is uniformly distributed among screen sides. The right screen is unbalanced due to the greater weight of its left side.

found four approaches of region recognition described in following four paragraphs.

The first approach generates its own layouts containing exact descriptions of the regions. The primary purpose is to simulate specific situations used for the comparison of user perception with the results given by a metric. Altaboli and Lin (2011) generate screens containing four black squares with different dimensions to test extreme values of the metrics Balance, Unity, and Sequence. Then, they demonstrate the correlation between these metrics and the user perception of aesthetics ($n = 8, p = 13$; users rated aesthetics on a 10-point scale). Salimun, Purchase, Simmons, and Brewster (2010) generate layouts comprised of triangles. They confirm the effect of the metrics Cohesion, Economy, Regularity, Sequence, Symmetry, and Unity. However, they also point out that users prefer interfaces with a medium level of aesthetics ($n = 15, p = 72$; the users compared aesthetics between pairs of screens). Bauerly and Liu (2008) replace black squares with random images to make the displays look realistic. They show that a high number of interface objects decreases the aesthetic appeal ($n = 27, p = 16$).

The second approach is based on an analysis of the structural description of real interfaces. Purchase et al. (2011) use a browser extension to parse the document object model (DOM) of web pages to specify regions. They analyze most of Ngo's metrics (except Equilibrium, Symmetry, and Rhythm) and confirm the correlation between the metrics and user perception ($n = 15, p = 21$). However, their results show that the aesthetics does not match the interface usability, which contradicts the findings of Kurosu and Kashimura (1995) and Tractinsky et al. (2000). They pointed out that the approach of DOM processing does not consider the visual content of an identified component.

The third approach uses raster screenshots and tries to detect regions automatically, using image processing methods. It considers the visual aspect of a screen compared to the previous approaches. Zheng et al. (2009) use the algorithm of iterative decomposition of a screen into quadrants of minimum entropy (*Quadtree decomposition*) based on low-level image statistics. They evaluate Balance, Symmetry, Equilibrium, and the number of quadrants and compare their influence on the judgment of the users ($n = 30, p = 22$). According to the results, the influence is not always the same (Balance has the highest influence, in contrast to Equilibrium). Reinecke et al. (2013) evaluate the same metrics as Zheng et al. They focus on the prediction of the visual complexity of interface ($n = 450, p = 548$). They use Quadtree decomposition *and Space-based decomposition* (decomposition of a screen by separating the components along the vertical and horizontal spaces in the screen).

The fourth approach depends on the manual selection of regions by the users. Zain, Tey, and Goh (2011) describe an application for the manual dragging of interface objects combined with further image processing. They use the application to confirm the correspondence between the expected ranking of metrics and values calculated from regions gathered by the users dragging objects ($n = 12$, $p =$ unspecified) using Balance, Equilibrium, Symmetry, Sequence, and Rhythm. Mazumdar et al. (2015), base on this model, extend it with Cohesion and Unity and use it to evaluate the aesthetics of one type of interface—semantic web tools ($n = 11$, $p =$ unspecified). The measured values are similar for most of the analyzed interfaces. Recent research (Zen & Vanderdonckt, 2014) provides the QUESTIM tool, which enables the loading of a website screenshot and lets users manually specify the regions representing the input for Ngo's metrics. According to their results, evaluating all 13 metrics ($n = 4, p = 25$, 5-point Likert scale), only 5 of 13 metrics (Balance, Equilibrium, Density, Economy, and Proportion with the best results) correspond to the user reviews. However, they point out the small set of interface samples and the problem of the subjective selection of regions. They also suggest a possible improvement of metric thresholds determining what is aesthetically efficient.

5. Focus of the research

The goal of our research was to analyze the possibility of the application of Ngo's metrics for the evaluation of dashboard quality concerning the subjective perception of the users. We followed the approach of Zen and Vanderdonckt (2014), who work with manually selected regions. We let the participants manually select objects in a screen rather than let them rate an interface directly. Selected regions were used to measure the ambiguity of user perception (Experiment 1) and the impact of the ambiguous perception on the objectivity of Ngo's metrics and the ability of the metrics to distinguish well-designed dashboards (Experiment 2). Finally, we proposed the improvement of Ngo's metrics based on the combination of subjectively skewed object-based metrics with an objective pixel-based analysis of a screen. We evaluated the approach using the Balance metric (Experiment 3).

We analyzed 130 dashboard samples divided into two groups: 9 well-designed and 121 randomly chosen. The group of well-designed dashboards consisted of the samples designed according to Few's framework (Few, 2006). The second group was composed of randomly chosen dashboard samples found on the internet. The reason we chose the samples based on Few's knowledge to be well-designed was the lack of other samples based on similarly reliable knowledge as (Eckerson, 2006; Tufte, 2001; Ware, 2012). We did not perform user testing since the evaluation of dashboard quality should be based not only on the first impression of the users but also on an in-depth analysis of interface usability as it was provided by Few. Moreover, it was not the aim of this research to evaluate the correctness of Few's framework or to establish another one. Another reason was to explore Few's framework from an aesthetic point of view. We use the label "well-designed" in the following text. However, readers of this article should consider the limitation of this label.

Choosing a dashboard as the only interface type that interested us was supported by results of Mazumdar et al. (2015), which suggested a possible regularity in the ratings of interfaces with a similar purpose. We consider this strategy as better for finding optimal metric thresholds because of findings of Salimun et al. (2010), which suggest that well-rated interfaces do not necessarily need to be rated by high values of Ngo's metrics.

The following three experiments provide a summarized overview of the results. More detailed results (including all descriptions of the regions and the statistics based on these descriptions) are available online.[1]

6. Experiment 1: Analysis of user perception

The goal of the first experiment was to analyze how users recognize the visually emphasized objects of dashboards. We gathered 130 image samples of dashboards and distributed them uniformly among 251 users, who provided us with descriptions of regions representing their subjective perception of the objects within a dashboard. Then, we used the descriptions to rate the ambiguity of visual perception.

6.1. Gathering the region descriptions

We selected the users among third-year students of the Information Systems course at the Brno University of Technology, Faculty of Information Technology. We dedicated one lecture to familiarizing the students with the term "dashboard" and the fundamental principles of data visualization and visual perception. Then, we assigned optional homework to the students. The students specified the regions of 20 dashboards according to their subjective perception. They used a simple Java application to load a dashboard, draw the perceived regions, and generate an XML description of the specified regions (Figure 5). The application did not allow them to specify regions hierarchically (regions within regions). A total of 251 users provided us with 5,020 descriptions of regions in total (approximately 39 descriptions for every dashboard).

6.2. Measuring perception ambiguity

First of all, we took descriptions of the regions of the same dashboard and combined them into one description representing the probabilities of region occurrences in every pixel of the dashboard. Figure 6a shows a visualization of such a description in the grayscale color space. Then, we used the descriptions of region probabilities to measure the entropy of the dashboard—a value representing the rate of user disagreement about the distribution of regions.

We calculated the binary entropy of every pixel according to the following formula:

$$E_{p_i} = -\left(p_i \log_2 p_i + (1 - p_i)\log_2(1 - p_i)\right) \tag{1}$$

where $p_i \in \langle 0, 1\rangle$ represents the probability of region occurrence in the i-th pixel position ($i = [x, y]$) in the matrix and $E_{p_i} \in \langle 0, 1\rangle$. An example of visualization of entropy values can be seen in Figure 6b. Then, the entropy of dashboard d was calculated as the average binary entropy of all the pixels in the dashboard:

$$E_d = \frac{\sum_{i=0}^{n} E_{p_i}}{n} \tag{2}$$

where n is the number of all the pixels in dashboard $d \in D$ and $E_d \in \langle 0, 1\rangle$. We measured the average entropy μ_E with its standard deviation σ_E for the set of all dashboards.

In addition to the entropy, we analyzed the number of regions in the dashboard and the user disagreement about this value. We calculated the average number of regions μ_{r_d} with its standard deviation σ_{r_d} and the coefficient of variation $c_v\left(\mu_{r_d}, \sigma_{r_d}\right)$ for every dashboard d. Then, we analyzed the average coefficient of variation $\mu\left(c_v\left(\mu_{r_d}, \sigma_{r_d}\right)\right)$ for the set of all dashboards (simply c_{v_r}).

6.3. Results

The average entropy of all dashboards μ_E was 0.262 ($\sigma_E = 0.109$). This means that the value p_i of every pixel i was 0.955 on average (95.5% of the users agreed on the logical value of the pixel). The average entropy can therefore be considered as low. Visualization of entropy matrices then indicated that high entropy was detected on the borders of regions (the black borders of white rectangles in Figure 6b). We expect that it was caused by a different precision of the users during the specification of regions rather than different perception. We also noticed that some logical parts of dashboards were more ambiguous than others—e.g., menus or toolbars (the left and top parts of Figure 6b). Some users considered these areas as solid regions, other users split them into smaller logical regions (such as buttons and labels).

The average coefficient of variation c_{v_r} was 0.78. This means that the standard deviations of the number of regions were relatively high compared to the average numbers of regions. This revealed the fact that the users usually agreed about the location of regions but disagreed about their quantity. They segmented the screen with a different granularity, as demonstrated in Figure 1.

In conclusion, the experiment confirmed the fact that users similarly recognize visually emphasized objects, which corresponds to the Gestalt laws. Based on the results of average entropy μ_E, we expect that a designer should be able to create a description of visually dominant regions which will cover a similar area of a screen as the average

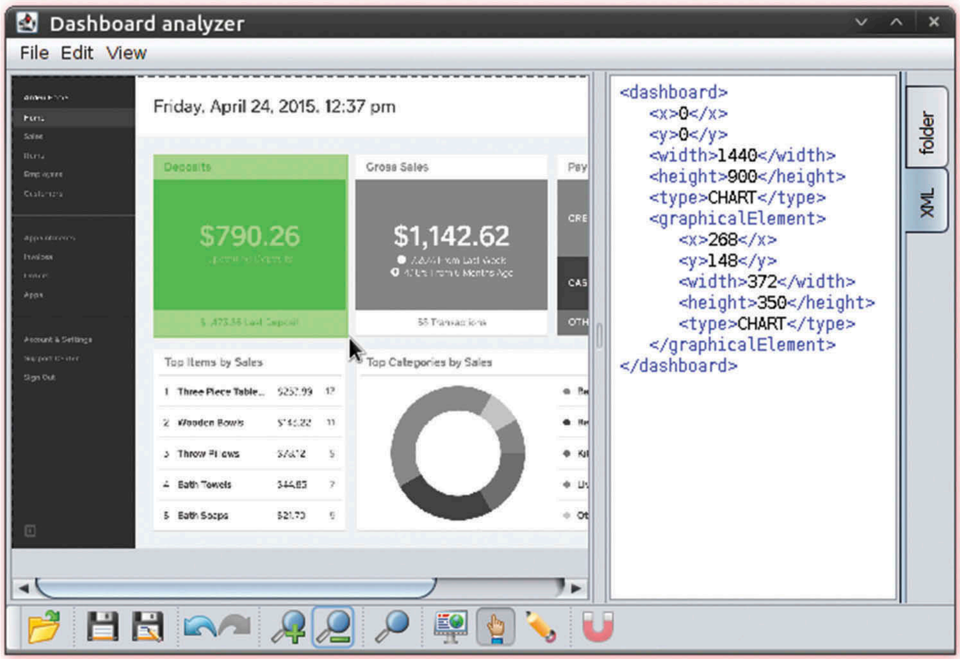

Figure 5. An example of the specification of regions using the Java application. The green area represents a selection of a visual region drawn by a user. The XML description presented on the right is re-generated with every change of regions in the canvas. It contains the specification of the dashboard and one region of the dashboard.

Figure 6. An example of a description of region probabilities (a) and a visualization of pixel entropies (b) represented in the grayscale color space. The higher color intensity represents a higher probability (a) and higher entropy (b) of region occurrence. The pixels representing medium probabilities of region occurrence ($p_i \sim 0.5$) are represented by higher values of entropy. Such pixels usually create borders of visually dominant objects. They can also be found in management areas (toolbars, menus) on the borders of a screen.

region description made by a sufficient number of instructed users. On the other hand, the subjective factor of visual perception will always be present ($\mu_E > 0$). Another designer will most likely create a slightly different description. Two designers using their subjective descriptions to evaluate one user interface can end up with different results. Hence, they should use only sufficiently robust metrics which are able to consider certain differences caused by subjective perception. Specifically, they should not use object-based metrics, which are highly dependent on the number of objects (due to the high value of c_{v_r}).

7. Experiment 2: Analysis of metric characteristics

The goal of the second experiment was to analyze the impact of the subjective perception of the users on the applicability of the 13 object-based metrics of aesthetics designed by Ngo et al. (2000) for the detection of well-designed dashboards. We used the descriptions of regions gathered in the first experiment as inputs for the metrics and calculated values for all the descriptions. Then, we analyzed the ambiguity between the values measured for the descriptions of the same dashboards. Finally, we compared the metrics and classified them.

7.1. Measuring metric characteristics

Figure 7 visually explains how we processed the descriptions of regions. First of all, we calculated a set of values $V_{(d,m)}$ measured by a metric m for every analyzed dashboard d (i.e., for the set of its descriptions of regions provided by users). Then, we removed 10% of the values in $V_{(d,m)}$ with the highest

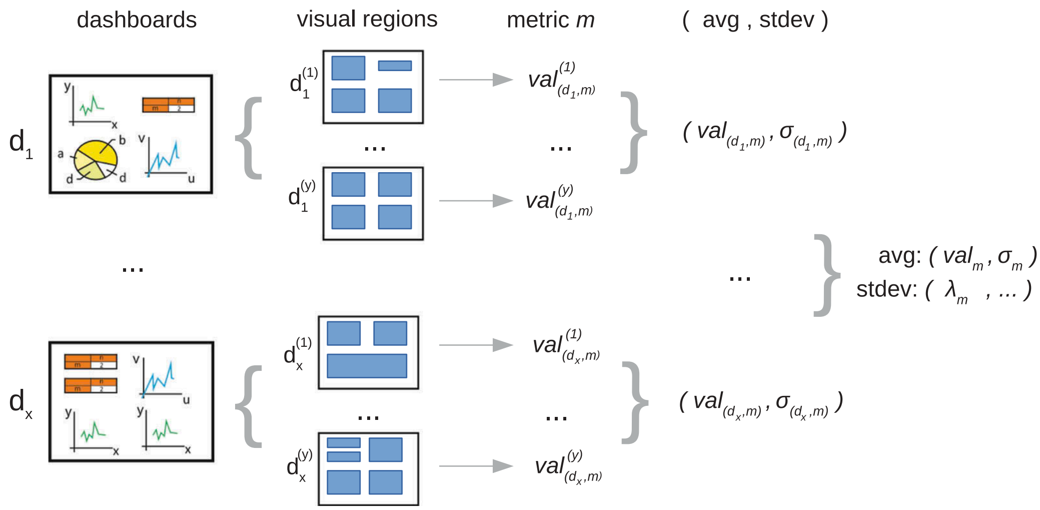

Figure 7. The process of measuring the values of dashboard regions ($x = 130$, $y \sim 39$). Every visual region was used as the input for a metric m . The measured values were used to create average values and standard deviations.

distance from the average value of $V_{(d,m)}$. The reason was to filter the values calculated from the most extreme descriptions of the regions. Then, we used the filtered sets of values to calculate 130 average values $\mu(V_{(d,m)})$ (for further purposes, simply $val_{(d,m)}$) with their standard deviations $\sigma_{(d,m)}$. Finally, we used these 130 average values and standard deviations to calculate 1 average value val_m and 1 average standard deviation σ_m. Similarly, we calculated the standard deviation of 130 average values $val_{(d,m)}$. We labelled it λ_m. The procedure described in this paragraph was repeated for the subsets of well-designed and random dashboards described in the chapter titled "Focus of the Research."

After we processed the descriptions of the regions, we analyzed the aggregated variables σ_m and λ_m to rate the influence of subjective perception on the applicability of the metrics:

- σ_m: This measures the average impact of subjective perception on the precision of the metric m. If the value of σ_m rises, there is more likely to be a greater difference between the values measured by the metric m for two independent descriptions of regions of one dashboard. We named this characteristic *metric volatility* (the opposite of *metric stability*).
- λ_m: This measures the ability of metric m to distinguish dashboards. If the value of λ_m rises, there is more likely to be a greater difference between values measured by the metric m for the descriptions of regions of two different dashboards. We named this characteristic *metric scalability*.

Metric stability together with metric scalability represents the characteristic *metric subjectivity* $= \frac{\sigma_m}{\lambda_m}$, which measures the average impact of subjective visual perception on the precision of the metric m relative to the range of the most frequently measured values. This means that a high value of metric volatility can be compensated by a high value of metric scalability.

To rate the ability of metrics to distinguish one group of user interfaces from another (e.g., well-designed from random dashboards), we established the variable $\gamma_m = \text{overlap}\left(val_m^{(A)}, \lambda_m^{(A)}, val_m^{(B)}, \lambda_m^{(B)}\right) \in \langle 0, 1\rangle$. The overlap function measures the overlapping coefficient of two normal distributions (of the groups A and B) represented by a mean val_m and standard deviation λ_m. If the value of the overlapping coefficient rises, it will be more difficult to distinguish these two groups by the metric m.

Finally, we established the overall rates of the metric m as:

$$\text{objectivity}_m = \text{subjectivity}_m{}^{-1} = \frac{\lambda_m}{\sigma_m} \tag{3}$$

$$\text{decisiveness}_m = \gamma_m{}^{-1} \tag{4}$$

The more objective (stable and scalable) the metric is, the less subjectively skewed results the metric provides. The more decisive the metric is, the greater the difference between the two groups the metric can find.

7.2. Metric classification

The purpose of Experiment 2 was not to observe particular metric values of objectivity and decisiveness, since they depend on the group of users and the set of analyzed samples chosen for this research. Instead, we categorized and compared the metrics with each other. We established a classification for this purpose.

- **Class 0**: The metric m which can quantify a particular aspect of a user interface according to a specified formula.
- **Class 1**: The metric m of Class 0 with *a high value* of objectivity$_m$ which is able to consider the subjectivity of visual perception to a specified extent.

- **Class 2**: The metric *m* of Class 1 with *a high value* of decisiveness$_m$ which is able to distinguish two kinds of user interfaces to a specified extent.

The definitions of Class 1 and Class 2 do not intentionally contain specifications as to what the high values of objectivity and decisiveness are because they might be different for another experiment. For this research, we set the limit of both high values to be 2.0 (λ_m will be at least 2 times higher than σ_m; γ_m will be lower than 0.5). We chose rather weak limits.[2] However, these limits might be modified for future experiments.

7.3. Results

The first results, presented in Figure 8, describe the metric objectivity. It was the first characteristic we analyzed because we wanted to distinguish the metrics of Class 0 and Class 1. The values of objectivity correlate with the categorization of Ngo's metrics made in Table 1.

The metrics based on the analysis of screen granularity (M_G = {Unity, Simplicity, Regularity, Economy, and Homogeneity}) have low values of objectivity, close to 1.0. We expected a low rate of objectivity because of the results of Experiment 1 (the users segmented the screen with a different granularity). Hence, it might be complicated to use the metrics of M_G for a comparison of dashboard aesthetics. We classified the metrics of M_G as members of Class 0.

On the other hand, the values of objectivity of the metrics based on the analysis of the aspect ratios of the regions (M_{AR} = {Cohesion and Proportion}) are higher than 2.0 . It appeared that the subjective perception of the users had a low impact on the metrics of M_{AR} . Hence, we consider the metrics of M_{AR} as members of Class 1.

The remaining six metrics based on the analysis of the area and distribution of regions in a screen (M_{AD} = {Balance, Equilibrium, Symmetry, Sequence, Density, and Rhythm}) appeared to be more objective than the metrics based on the analysis of screen granularity. The results correspond to the low average entropy measured in Experiment 1. However, except for Rhythm, the values of their objectivity are lower than 2.0, which makes them members of Class 0.

The next results, presented in Figure 9, describe the metric decisiveness. Since we categorized only three metrics as members of Class 1—Cohesion, Proportion, and Rhythm—only these metrics could become members of Class 2. However, as shown in Figure 9, the values of decisiveness are low except for one metric: Density. Thus, it would be complicated to use Ngo's metrics for the detection of well-designed samples.

One possible reason for the low rates of decisiveness might be the insufficient number of well-designed samples. In addition, the group of randomly chosen dashboards may contain well-designed dashboards which would make it harder to distinguish known well-designed samples. Finally, we need to consider the possibility that the aesthetics of dashboards does not relate to the appearance of the selected samples.

In conclusion, the experiment pointed out the problem regarding the low objectivity of several object-based metrics. We should use the metrics based on the analysis of screen granularity with close attention. On the other hand, the metrics based on the analysis of the aspect ratio of regions (Cohesion and Proportion) or the distribution of regions in a screen (Rhythm) seem to be more immune to the subjective perception of the users. However, their use for the detection of well-designed dashboard samples is highly questionable.

8. Experiment 3: Improvement of metric characteristics

The goal of the last experiment was to find a possible approach which would decrease the impact of the subjective perception of the users on the characteristics of metrics explained in Experiment 2 (metric objectivity and

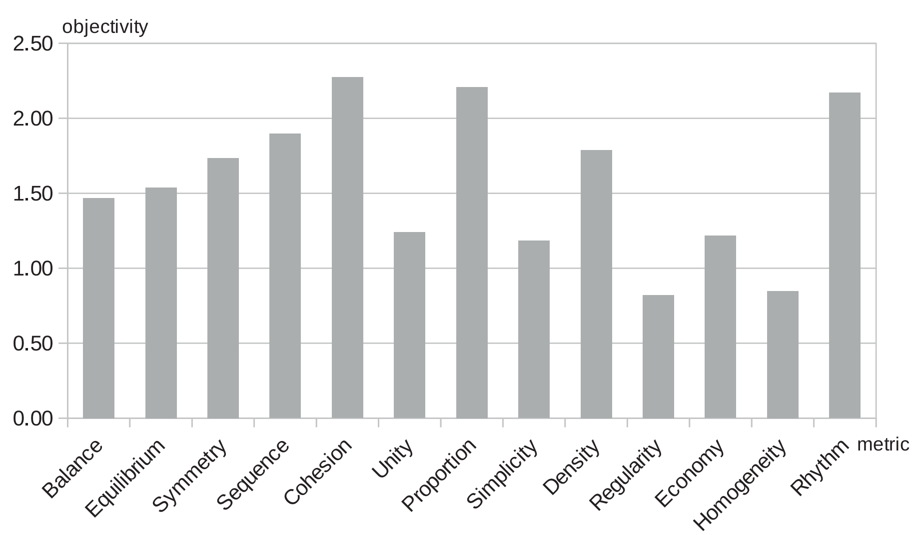

Figure 8. Values of metric objectivity measured for all Ngo's metrics.

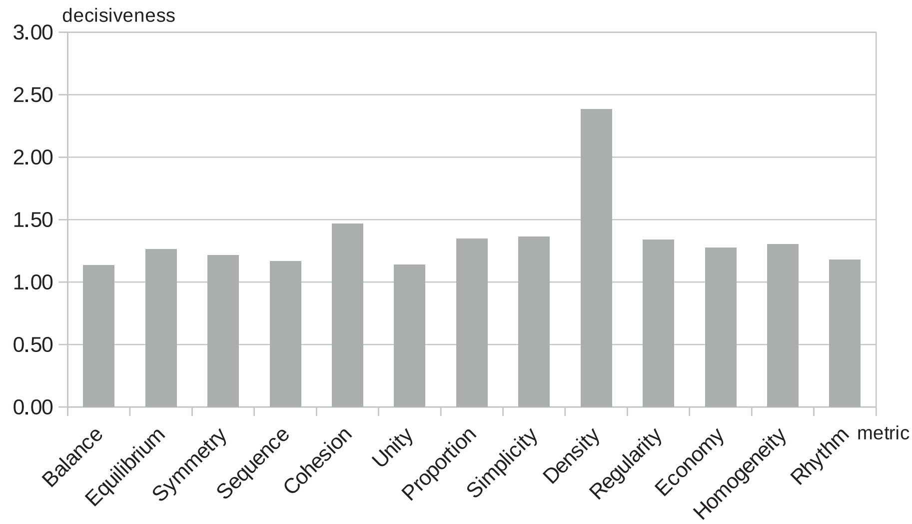

Figure 9. Values of metric decisiveness measured for all Ngo's metrics.

decisiveness). We focused on the group of metrics based on the analysis of the distribution of regions in a screen—particularly on the Balance metric which we rated as the least objective metric of this group. We improved this metric by the combination of the metric with the objective pixel-based approach of analysis of the color distribution in a screen.

8.1. The Idea of balance improvement

Ngo computes Balance as the difference between the total weighting of the components on each side of the horizontal and vertical axes:

$$\mathrm{BM} = 1 - \frac{\mathrm{BM}_{\mathrm{vertical}} + \mathrm{BM}_{\mathrm{horizontal}}}{2} \in \langle 0, 1 \rangle \tag{5}$$

$$\mathrm{BM}_{\mathrm{v.}} = 1 - \frac{w_L + w_R}{\max(|w_L|,\ |w_R|)},\ \mathrm{BM}_{\mathrm{h.}} = 1 - \frac{w_T + w_B}{\max(|w_T|,\ |w_B|)} \tag{6}$$

where w_j is a weighting of side $j \in \{L,\ R,\ T,\ B\}$ (left, right, top, bottom) containing n_j regions:

$$w_j = \sum_i^{n_j} a_{ij} d_{ij} \tag{7}$$

where $a_{ij}d_{ij}$ is a weighting of a region i in a quadrant j calculated as a product of the region area and its distance from the centre (more in (Ngo et al., 2000)). Figure 4 shows an example of balanced and unbalanced weightings of sides.

The problem of the weighting w_j is that it works with ambiguous values a_{ij} and d_{ij}. The users usually agreed about the approximate area and distribution of regions in a screen but they did not usually specify these regions with exactly the same precision. Our idea was to include objective information about the color of subjectively specified regions to objectively affect their weightings. Hence, we modified the formula of the Balance weighting:

$$w_j{}' = \sum_i^{n_j} a_{ij} d_{ij} C_{ij} \tag{8}$$

where C_{ij} is the *coefficient of color* of a region i in a quadrant j representing the colorfulness of the region. Since two sides of a screen are always compared to each other, there is no problem in modifying the weightings of each side by adding C_{ij} to the formula and keeping the range of the formula: $\langle 0, 1 \rangle$. We explored several approaches to measuring the coefficient of color using different color spaces:

- $C_r^{(1)} \in \langle 0, 1 \rangle$: The average color intensity of a region r represented in the 8-bit grayscale color space converted from the RGB color space.
- $C_r^{(2)} \in \langle 0, \infty \rangle$[3]: The average colorfulness of a region r inspired by Yendrikhovskij, Blommaert, and Ridder (1998) and Reinecke et al. (2013): $C_r = S_r + \sigma_r$ where S_r is a value of the average saturation of a region r in the CIE Lch color space and σ_r is its standard deviation.
- $C_r^{(3)} \in \langle 0, 1 \rangle$: The average value of all pixel values in a region r calculated as $1 - (b_i - b_i s_i)$ where $s_i \in \langle 0, 1 \rangle$ is the saturation and $b_i \in \langle 0, 1 \rangle$ is the brightness of the i-th pixel of the region in the HSB color space (based on the suggestion of Ngo et al. (2000) that users might assign visual importance to pixels with high saturation or low brightness).

The idea of the improvement corresponds with Ngo et al. (2000) who suggested considering highly colorful regions as visually heavier than regions with a lower colorfulness. Pastushenko, Hynek, and Hruška (2018) evaluated this theory and showed that users rated the same interfaces varying in the colors of the widgets by different Balance values. On the other hand, this improvement does not capture the image complexity (e.g., shapes of widgets), which affects user perception as well. The inclusion of the image complexity might be the next step in improving the quality of metrics in future.

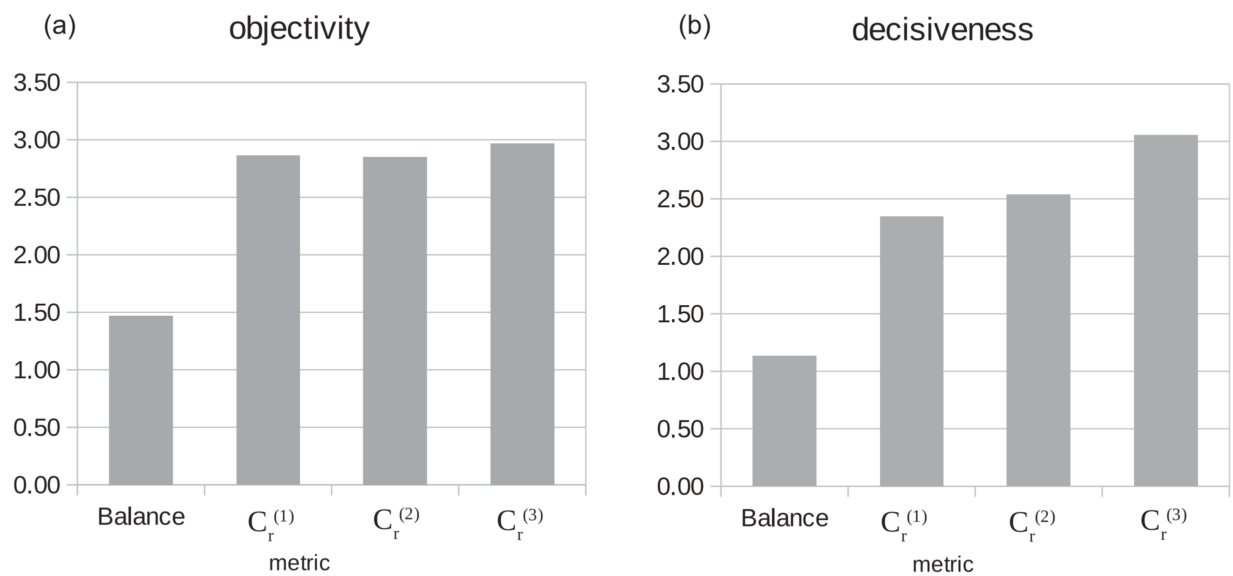

Figure 10. Change of Balance objectivity (a) and decisiveness (b) for dashboards and dashboard bodies using the coefficients of color $C_r^{(1)}$, $C_r^{(2)}$, and $C_r^{(3)}$.

8.2. Results

We can see a significant improvement in all kinds of the coefficient of color in Figure 10. The values of objectivity and decisiveness are higher than 2.0, which makes Balance a member of Class 2. We received the best results for $C_r^{(3)}$ calculated according to the formula using the rate of saturation and brightness in the HSB color space, followed by the results for $C_r^{(2)}$ considering the colorfulness calculated in the CIE Lch color space. From a practical point of view, the easiest method improving Balance is to use the coefficient $C_r^{(1)}$ based on the color intensity, since the color intensity can be simply calculated from the RGB color space. The color intensity might also correspond better with the perception of color blind people (Few, 2006). In addition, the infinite range of $C_r^{(2)}$ might cause problems with the modification of some metrics.

Table 2 presents the average values and standard deviations of Balance measured for well-designed and randomly chosen dashboards. The well-designed dashboards are more balanced than the randomly chosen ones for all types of Balance, including the modified ones. We can see the decrease in $val^{(rand)}$ for the modified versions of Balance. This indicates that the modified versions of Balance are stricter than the original Balance. Since the original Balance rated certain dashboards as balanced, the modified versions of Balance rated the dashboards as unbalanced because of their unbalanced distribution of color in a screen.

Table 2. The average values of Balance (*val*) with its standard deviations (λ) for groups of well-designed (*well*) and random dashboards (*rand*).

	$val^{(well)}$	$\lambda^{(well)}$	$val^{(rand)}$	$\lambda^{(rand)}$
Balance	0.873	0.107	0.842	0.107
$Cr^{(1)}$	0.830	0.086	0.639	0.196
$Cr^{(2)}$	0.819	0.072	0.641	0.181
$Cr^{(3)}$	0.845	0.068	0.648	0.190

In conclusion, the proposed modification of Balance seems to be the correct way to improve the objectivity and decisiveness of the metric. We expect that the idea of modification could be used to improve other metrics based on the area and distribution of regions in a screen in future (namely, Equilibrium, Symmetry, Sequence, and Rhythm).

9. Discussion

The modified version of Balance using the coefficient of color can be used for the improvement of the tools designed for metric-based evaluation of user interfaces. Since existing tools apply different approaches to detect regions, it might be appropriate to use the metric which considers the possible ambiguity of the inputs. In future, we would like to design a segmentation algorithm for the automatic detection of regions based on the average user perception analyzed in Experiment 1. However, we are aware that it might not be possible to design an algorithm providing an entirely objective segmentation of screen. Objective metrics will help us to reduce the impact of this limitation. We will try to improve Ngo's remaining metrics similarly to Balance.

There are some limitations of the results which we should consider in future work. The first limitation is caused by the chosen set of dashboard samples. The set of well-designed dashboards contains a small number of samples since not many examples are available, but several samples are based only on Stephen Few's work. However, the goal of this work was not to evaluate existing dashboard design frameworks. We are developing a generator using recommended design principles which will help us increase the set of well-designed dashboards in future (Pastushenko et al., 2018).

The second limitation of the results is caused by the chosen sample of users who provided us with descriptions of perceived regions. We worked with a relatively large number of users compared to other evaluations described in the section titled "Ambiguity of Object Recognition." However, the users were mainly technical students. We expect that there may be slight, but interesting deviations between the perception of people of different specializations (e.g., persons having skills in the arts).

Conclusion

The research pointed out the fact that an ambiguous definition of interface objects can complicate the application of object-based metrics for the evaluation of dashboard quality (and the quality of a user interface in general). A different recognition method, even a different user, can specify interface objects differently. We showed that users tend to perceive visually emphasized objects of a dashboard in a similar but not the same manner. Objects are usually composed of several simple graphical shapes clustered preattentively by the human brain, making logical parts of a screen (as described by the Gestalt laws). The level of screen granularity was usually the main subject of disagreement between users. It complicates the application of object-based metrics for the evaluation of a user interface—especially those which depend on the number of objects.

This article has described the analysis and solution of the problem. We established the model to rate the impact of subjective perception on metric ability to objectively distinguish chosen well-designed dashboards from others (metric objectivity and decisiveness). Then, we used this model to evaluate 13 metrics designed by Ngo et al. (2000) for measuring aesthetic quality. We classified the metrics and found that none of them meets our requirements. For this reason, we designed an improvement which combines object-based metrics with a pixel-based approach measuring the colorfulness of the interface regions. We demonstrated the approach on the Balance metric. The improved metric was rated as objective and able to recognize well-designed dashboards. We believe that the proposed model can be generalized and applied for the evaluation of other metrics with a combination of other kinds of user interfaces.

Finally, there are open questions which arise from this article:

- Is the ambiguity of the user perception of visually dominant regions generally different in specific parts of a user interface? Experiment 1 suggested that some logical parts of dashboards were usually more ambiguous than the rest of the dashboard—e.g., menus or toolbars. We expect that the reduction in such parts may influence the usability of a user interface.
- Does a high value of the ambiguity of the user perception have a negative influence on the usability and quality of the user interface?
- Is it possible to improve the objectivity and decisiveness of Ngo's remaining metrics without a radical change of their characteristics? We expect that the metrics based on the analysis of the distribution of regions in a screen could be modified similarly to the Balance metric.
- Is it possible to use all formulas of the coefficient of color $C_r^{(1)}$, $C_r^{(2)}$ and $C_r^{(3)}$ to improve Ngo's other metrics based on the analysis of the distribution of regions? A comparison of these approaches should be done for other metrics.
- Are there any other visual characteristics (e.g., image complexity, shapes of widgets) which can be used for the improvement of metrics similarly to color?
- What is the optimal level of objectivity and decisiveness? We established limits for the purposes of this research but we expect that these levels might be different for various types of user interfaces and users. Further experiments should therefore be done.

The listed items can be considered as suggestions for future work.

Notes

1. Available at: http://www.fit.vutbr.cz/~ihynek/dashboards/ijhci-2018.
2. For instance, we need to consider that the group of randomly chosen dashboards might also contain well-designed dashboards, which might increase the value of γ_m.
3. Highly colorful values are >2.

Acknowledgments

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project 'IT4Innovations excellence in science – LQ1602'.

ORCID

Jiří Hynek http://orcid.org/0000-0002-7292-6094
Tomáš Hruška http://orcid.org/0000-0001-5454-6592

References

Altaboli, A., & Lin, Y. (2011). Investigating effects of screen layout elements on interface and screen design aesthetics. *Advances in Human-Computer Interaction*, *2011*, 1–10. doi:10.1155/2011/659758

Baker, J., Jones, D. R., & Burkman, J. (2009). Using visual representations of data to enhance sensemaking in data exploration tasks. *Journal of the Association for Information Systems*, *10*, 533.

Bauerly, M., & Liu, Y. (2008). Effects of symmetry and number of compositional elements on interface and design aesthetics. *Intl. Journal of Human-Computer Interaction*, *24*, 275–287.

Bodart, F., Hennebert, A.-M., Leheureux, J.-M., & Vanderdonckt, J. (1994). Towards a dynamic strategy for computer-aided visual placement. *Proceedings of the workshop on Advanced visual interfaces*, (pp. 78-87).

Charfi, S., Trabelsi, A., Ezzedine, H., & Kolski, C. (2014). Widgets dedicated to user interface evaluation. *International Journal of Human-Computer Interaction*, *30*, 408–421.

Eckerson, W. W. (2006). *Performance dashboards: Measuring, monitoring, and managing your business*. Hoboken, NJ, USA: John Wiley & Sons.

Few, S. (2006). *Information dashboard design*. North Sebastopol, CA, USA: O'Reilly Media.

Gibson, J. J. (1950). *The perception of the visual world.*Boston, MA, USA: Houghton Mifflin.

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*, 107–135.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Hynek, J., & Hruška, T. (2016). Pixel-based analysis of information dashboard attributes. *East European Conference on Advances in Databases and Information Systems*, (pp. 29-36).

Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys (CSUR)*, *33*, 470–516.

Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision Research*, *126*, 3–8.

Kim, W. C., & Foley, J. D. (1993). Providing high-level control and expert assistance in the user interface presentation design. *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, (pp. 430-437).

Kristeller, P. O. (1951). The modern system of the arts: A study in the history of aesthetics part I. *Journal of the History of Ideas*, *12*(4), 496–527.

Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. *Conference companion on Human factors in computing systems*, (pp. 292–293).

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, *60*, 269–298.

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour\& Information Technology*, *25*, 115–126.

Mahajan, R., & Shneiderman, B. (1997). Visual and textual consistency checking tools for graphical user interfaces. *IEEE Transactions on Software Engineering*, *23*, 722–735.

Mazumdar, S., Petrelli, D., Elbedweihy, K., Lanfranchi, V., & Ciravegna, F. (2015). Affective graphs: The visual appeal of Linked Data. *Semantic Web*, *6*, 277–312.

Merriam-Webster. (2004). *Merriam-Webster's collegiate dictionary*. Springfield, MA, USA: Author.

Ngo, D. C. (2001). Measuring the aesthetic elements of screen designs. *Displays*, *17*(2), 73–78.

Ngo, D. C., & Byrne, J. G. (2001). Application of an aesthetic evaluation model to data entry screens. *Computers in Human Behavior*, *17*, 149–185.

Ngo, D. C., Teo, L. S., & Byrne, J. G. (2000). Formalising guidelines for the design of screen layouts. *Displays*, *21*, 3–15.

Ngo, D. C., Teo, L. S., & Byrne, J. G. (2003). Modelling interface aesthetics. *Information Sciences*, *152*, 25–46.

Nielsen, J. (1994a). *Usability engineering*. San Diego, CA, USA: Academic Press.

Nielsen, J. (1994b). Usability inspection methods. *Conference companion on Human factors in computing systems* (pp. 413–414). ACM.

Pastushenko, O., Hynek, J., & Hruška, T. (2018). Generation of test samples for construction of dashboard design guidelines: impact of color on layout balance. *World Conference on Information Systems and Technologies*, (pp. 980-990).

Purchase, H. C., Hamer, J., Jamieson, A., & Ryan, O. (2011). Investigating objective measures of web page aesthetics and usability. *Proceedings of the Twelfth Australasian User Interface Conference-Volume 117*, (pp. 19-28).

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 2049–2058).

Salimun, C., Purchase, H. C., Simmons, D. R., & Brewster, S. (2010). Preference ranking of screen layout principles. *Proceedings of the 24th bcs interaction specialist group conference*, (pp. 81-87).

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

Sears, A. (1995). AIDE: A step toward metric-based interface development tools. *Proceedings of the 8th annual ACM symposium on User interface and software technology*, (pp. 101-110).

Shneiderman, B., Chimera, R., Jog, N., Stimart, R., & White, D. (1998). *Evaluating spatial and textual style of displays*. Tech. rep.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software*. Bedford, MA, USACA, USA: The MITRE Corporation.

Tractinsky, N. (2004). Toward the study of aesthetics in information technology. *ICIS 2004 Proceedings*, 62.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*, 127–145.

Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire, CT, USA: Graphics Press.

Tullis, T. S. (1984). *Predicting the usability of alphanumeric displays*. Ph. D. dissertation, Rice University.

Vanderdonckt, J., & Gillo, X. (1994). Visual techniques for traditional and multimedia layouts. *Proceedings of the workshop on Advanced visual interfaces*, (pp. 95-104).

Ware, C. (2012). *Information visualization: Perception for design*. San Francisco, CA, USA: Elsevier.

Yendrikhovskij, S. N., Blommaert, F. J., & Ridder, H. (1998). Optimizing color reproduction of natural images. *Color and Imaging Conference, 1998*, pp. 140-145.

Zain, J. M., Tey, M., & Goh, Y. (2011). Probing a self-developed aesthetics measurement application (SDA) in measuring aesthetics of mandarin learning web page interfaces. *arXiv preprint arXiv:1101.1606*.

Zen, M., & Vanderdonckt, J. (2014). Towards an evaluation of graphical user interfaces aesthetics based on metrics. *Research Challenges in Information* Science *(RCIS), 2014 IEEE Eighth International Conference on*, (pp. 1-12).

Zheng, X. S., Chakraborty, I., Lin, J. J.-W., & Rauschenberger, R. (2009). Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1-10).

About the Authors

Jiří Hynek is a student in a Ph.D. degree program at the Brno University of Technology, Faculty of Information Technology. He focuses on issues connected with the evaluation of user interface quality.

Tomáš Hruška is a graduate of the Brno University of Technology, Czech Republic. Prof. Hruška received his CSc. (Ph.D.) in Computer Science and Engineering from the Brno University of Technology, Czech Republic. He has worked in the field of Information Systems and Visualisation.