The Evaluation of Interface Aesthetics

Mati Mõttus
Institute of Informatics
Tallinn University
Narva mnt 29
Tallinn
Estonia
mati@foti.ee

David Lamas
Institute of Informatics
Tallinn University
Narva mnt 29
Tallinn
Estonia
drl@tlu.ee

Maarja Pajusalu
Institute of Informatics
Tallinn University
Narva mnt 29
Tallinn
Estonia
maarja.pajusalu@gmail.com

Rui Torres
Universidade Fernando Pessoa
Praça 9 de Abril, 349 | 4249-004 Porto
Portugal
rui.torres@ufp.pt

Abstract
There are many factors that contribute towards good user experience (Roto, Law, Vermeeren and Hoonhout, 2011). These factors include the content and its organization, the functionality and features, the information and interaction design, as well as the visual design (Garett, 2002; Morville's, 2004; and Hassenzahl, 2005).

This paper builds on the contribution of visual design into user experience as grounds to tackle the assessment of visual aesthetics evaluation methods. The intention of the study is to test objective and subjective evaluation methods with the same objects for comparison. Finding out the correlations between the objective and subjective evaluation results enables the usage of computerized image analysis for the purposes of evaluating aesthetics.
The work reported in this paper thus contributes towards identifying a suitable objective method for a mathematical description of beauty.

MIDI '13, June 24 - 25 2013, Warsaw, Poland
Copyright 2013 ACM 978-1-4503-2303-1/13/06…$15.00.

Keywords
Aesthetics evaluation; objective methods; subjective methods.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI):User Interfaces - Evaluation/methodology

Introduction
Beauty is in the eye of the beholder. Aesthetics of visual image trigger emotions in person, who looks at the image. Subjective methods of measuring aesthetics, assess really the strength and polarity of those emotions. Some recent studies [4,5,6,11,12] reveal strong connection between first impression (visual stimuli) and overall judgement about the interactive product. Other studies relate aesthetics to affordances and user experience [15]. Relations between visual aesthetics and usability indicate the possible need for

more detailed design decisions and consideration of applying aesthetics evaluation methods in interface design.

The emotions that viewer gets while looking at the image is hidden in certain properties of the image. These properties can be evaluated through objective methods of measuring aesthetics. George David Birkhoff [3] introduced as early as in 1933 objectively measurable aesthetics of objects. His measures of complexity, symmetry, balance, and others have been widely used and interpreted later on. In the field of information technology and graphic screens, Ngo, Samsudin, and Abdullah [9] model mathematically the visual layout aesthetics using Birkhoff's and further measures.

Current paper reports the experiment of comparing objective and subjective methods with the purpose to reveal disadvantages of objective methods. Subjective methods were previously considered to be applicable [10] and are used as reference in comparison. This is the repetitive study that could specify how to improve the existing model. More specific study of aesthetical aspects can give a clue for improving objective methods. The suggestions for improvement are explained in conclusion.

The reported experiment is part of larger study about interface aesthetics. Goal of this study is to develop the model of interface aesthetics for practical use in interface design. The new model will be based on existing model [8]. The model of interface aesthetics will allow objective evaluation of visual interface aesthetics. Using the model will help interface designers to find better layout solutions without conducting expensive and time consuming empirical studies with many participants. The appliance of model can easily be automated, because data for calculations can be collected via automatic image analysis.

Similar Studies

Similar studies have been conducted by Altaboli and Lin in 2011 [1] and Pajusalu in her master thesis 2012 [10]. Both studies did compare objective and subjective methods to evaluate websites and concluded that objective methods can not express aesthetics adequately. Pajusalu mentioned, that objective counts based method is not applicable in interface design.

Altaboli and Lin took data from empirical study of Moshagen and Thielsh [7] - 42 webpages evaluated by 3 subjective methods: VisAWI [7], classical and expressive aesthetics [4]. Then calculated 8 objective measures of aesthetics for all 42 webpages. These 8 measures were part from 13 measures of the model by Ngo et al [8].

Pajusalu evaluated 8 websites of different art museums with subjective methods: WisAWI, classical and expressive aesthetics. 8 websites with different layouts were chosen to reveal different facets of aesthetics [7]. Then 5 measures of counts based objective method were calculated and the results of both methods were compared.

Current study is continuation of Pajusalus work (the previous study). The same 8 websites of art museums will be evaluated again. The results of subjective methods can be improved by using larger sample instead of six test users. The methods will be reviewed in order to have more precise result. The hypothesis of previous study was: Applicable methods exist to evaluate visual aesthetics.

Subjective methods were previously proved to be applicable, therefore the new hypothesis is: Applicable objective methods can be developed to evaluate visual aesthetics.

The research question for current paper is: How to improve the existing model of objective evaluation of visual aesthetics?

Review of Methods

Following are previously missed methods, that could be used for improving the results and applying some of these may help to prove the hypothesis.

Interface criticism

Interface criticism [2], is subjective method, which includes also aesthetical measures. Interface criticism

is a method based on literary and art criticism traditions. It is the only method (apart from specific questionnaires and experiments) concentrating on higher level constructs as representations, genres, and stylistic references. It also suggests analysing the use of standards and developmental potential of interface and supports further developments of the guidelines.

Modelling of interface aesthetics
Modelling of interface aesthetics [8] includes objective assessing measures, that give out aesthetical value 0...1, where 0 means "not aesthetical" and 1 means "absolutely aesthetical". 14 measures were brought up: balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm, order and complexity. All those measures describe, how objects are placed in composition and the size of the objects.

Aesthetic colouring system
Aesthetic colouring system is a method for producing aesthetically pleasing colour schema for complex layout (e.g. mobile phone) [14]. Coloured layouts can be automatically produced using an optimization tool that assists decision making in identifying optimal or near optimal solutions for problems with large search space. This method proves the possibility to objectively evaluate aesthetics of selected colour combinations for various types of interfaces.

Measuring the physiological changes
Measuring the physiological changes of users enables assessing their reaction to the experienced aesthetics. Strebe [12] suggests research of affective reactions in order to apply it on evaluating website aesthetics. To evaluate the impact of affectively effective aesthetics of websites on approach and avoidance behaviour, she suggested using screen saving, facial and eye movement tracking. An interesting tool to measure facial reactions is based on the evidence that different levels of aesthetics in websites results in proved facial reactions – movement of a muscle above eyebrows and a muscle on the cheek [12].

Objective methods need to be revised for proving the hypothesis, because not all proposed methods were used in previous study and some methods were just partly used.

Choosing the Methods

The same subjective methods were chosen as in previous study, because the method proved to be applicable. For better results the empirical study was repeated with 41 participants instead of 6 in previous study.
Previously applied objective method was counts-based method which uses the number of objects in layout to describe the aesthetics of visual image. Besides number of objects, also other image properties are relevant concerning aesthetical value. More aesthetical properties of interface layout are involved in model of visual aesthetics by Ngo et al [8]. This model already includes similar measures that are used in counts-based method. Therefore is no need to use counts-based method again. Altaboli and Lin used 8 measures of visual aesthetics in their study [1], but the model contains 13 aesthetic measures. In following experiment all 13 measures will be used to calculate aesthetical value of web pages.
Remaining methods in previous list are Interface Criticism, Aesthetic Colouring and Physiological Changes. Interface Criticism as subjective method was not preferred in main study, because it provides qualitative data which is difficult to interpret for correlation analysis. Aesthetic Colouring is not actually the method for evaluating aesthetics but rather for generating aesthetical layout. Physiological Changes could not be measured in this experiment because the lack of special equipment and experts.

Applying the Objective Method

The model of visual aesthetics (the model) consists of 13 measures which express mathematically aesthetic value of visual image. The appliance of model means

Figure 1. Website of Amoda

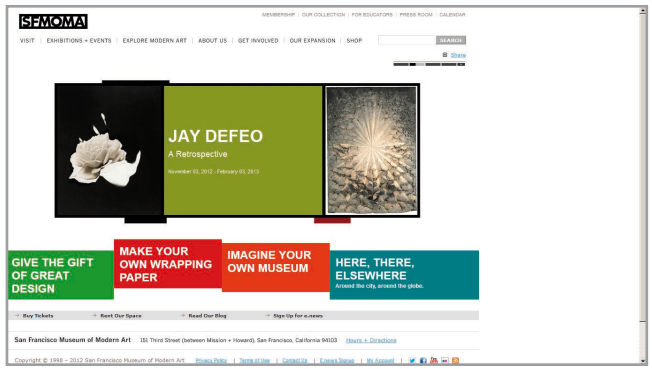


Figure 2. Website of SFMOMA

that objects of layout will be counted and its properties will be measured. Image pixels will be used as units when measuring objects width, height, area and location. Next subsections explain how the layout objects were extracted from the image.

Extracting objects

Cutting out layout objects for measuring is not precisely explained when describing model of interface aesthetics [8], some samples were explained, but interfaces can have wide range of complex layouts e.g. text areas with unclear borders and animations. In such cases the objects will be determined by subjective decision.

The restriction of the model is that objects in layout must be in rectangular shape. Which means that all non rectangular objects have to be approximated into rectangular shape.

Extracting objects is most questionable process, because complicated designs can be differently interpreted. For example the main article area of Amoda site (fig.1) can be taken as one object or each article separately. Decision is subjective and can be arbitrary. The result definitely depends on that choice, but in case of doubt, the method can be applied repeatedly with all possible combinations.

The main principle of extracting objects used in this

Figure 3. Website of SFMOMA with extracted objects

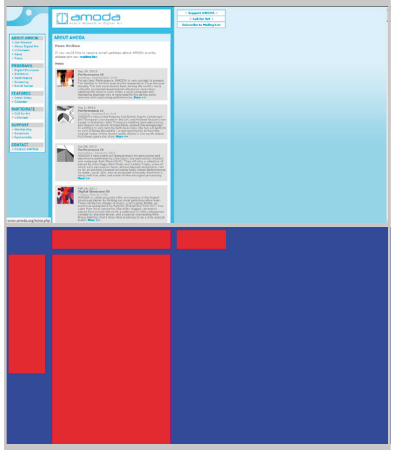
study is simple. If one-piece visual object has different horizontal or vertical alignment points in some part of it, then initial object will be divided into different objects (see SFMOMA banners just below centre of image fig.2,3).

Overlayed objects

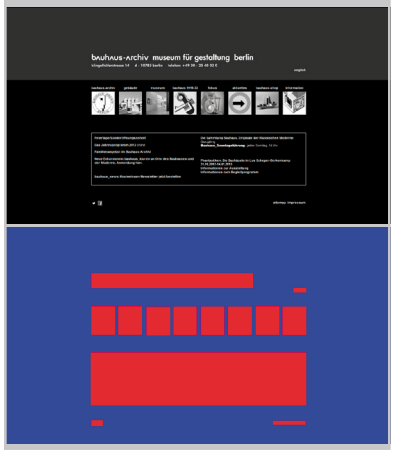
If two or more objects are overlayed then two choices can be made - take those objects as one (see SFMOMA centre of image fig.2,3) or as separate non-overlaying

All evaluated websites and its extracted objects.

Amoda



Bauhaus



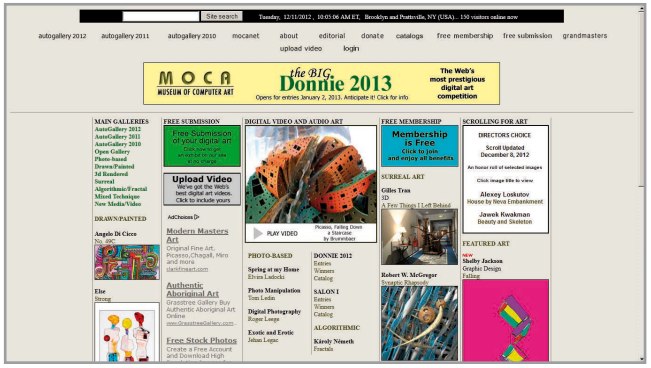


Figure 4. Website of MOCA

Figure 5. Website of Musee Psyche

objects (see SFMOMA banners just below centre of image fig.2,3).

Animated objects
Animated objects were freezed at the normal end of animation or in some subjectively chosen point of animation loop. Then handled as usual still objects.

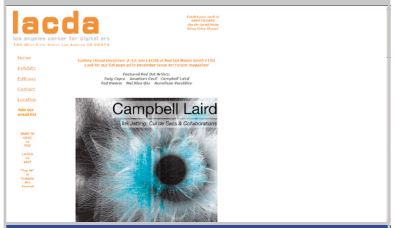
Objects in quadrants
Objects in quadrants of screen. How the objects belong to quadrant is not strictly explained by Ngo et al (8), but there is some hints to it in their previous publication (9). According to this, the object belongs to certain quadrant, if centre point of object belongs there. The cases with objects exactly on quadrant boundary were not actual in this study, but must be reconsidered subjectively, if it happens, because weight of object depends also on color lightness and shape of the objects. Color, lightness and shape are the measures not included in present method, therefore subjective decision is recommended.
The area of objects on quadrant is the area of actual part of object on that quadrant. The distances between object and the frame edge are measured from centre point of the object to the nearest edge, either horizontally or vertically.

Layout size
Layout size is the minimum bounding box for all objects in layout. For example Psyche (fig.5) has layout of 469x738px, but MOCA (fig.4) has layout of 1218x758px. Screen size for this experiment was fixed at 1366x768px and full screen view was used for empirical study. The Screen size was used instead of layout size, when normalizing the dimensions-based measures, because full screen view is one option, that designers must test for proper looking. Screen size does not influence aesthetical value, but screen aspect ratio does. Therefore is important to use screens with the same aspect ratio with both objective and subjective methods.

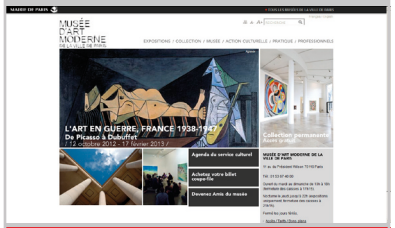
Alignment points
Alignment points are top left points of the positioned object. Subject might be misunderstood, because alignment points are said to be for alphanumeric data only, but in layout may be pictures and graphical shapes as well. Objects have the same horizontal alignment point when alignment points of those objects have the same x-coordinate. Similarly, the same vertical alignment point is, when y-coordinate of different objects alignment points is the same. The

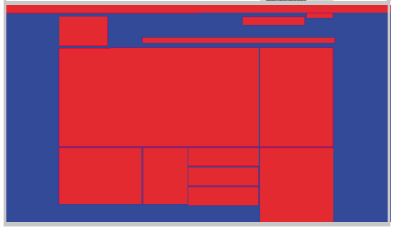
All evaluated websites and its extracted objects.

Lacda



MAM Paris





term "different spacing points" is explained as different distances between alignment points [8]. Both different horizontal and vertical spacing points will be counted.

All necessary measurements which were taken for calculations [8] are listed here:

Counted numbers:
- Objects total
- Different object's sizes
- Objects on quadrants (upper, lower, left, right)
- Alignment points (horizontal, vertical)
- Different spacing points

Coordinates, size, area in pixels:
- Screen center
- Screen size
- Layout size
- Objects center
- Objects size
- Objects area in quadrant (upper, lower, left, right)

Results and analysis

The normalized results with objective method are shown in tab.1. All 13 measures separately, overall measure and rank for 8 web pages. Overall measure is arithmetic average of all measures. Rank shows the order of websites by aesthetical value where "1" is most aesthetical.
The normalized results with subjective methods are shown in tab.2. All facets of VisAWI, Classical and Expressive aesthetics separately, average and rank.
The correlations between objective measures and subjective facets are shown in tab.3.

Correlations R>0,5 were between
- balance measure and expressive facet (R=0,502)
- symmetry measure and diversity facet (R=0,52)
- symmetry measure and expressive facet (R=0,54)
- unity measure and simplicity facet (R=0,52)

Negative correlations R<-0,5 were between
- sequence measure and simplicity facet (R=-0,56)
- rhythm measure and classical facet (R=-0,501)
- rhythm measure and expressive facet (R=-0,548)

Though, all correlations above remain statistically insignificant.

Even if there were low correlations between single items and between final results, the 1st rank was the same for both methods and subjective methods 2nd was objective methods 3rd. Surprisingly objective methods 2nd had last rank (8th) with subjective method and subjective method 3rd was 6th with objective method. Lets look at the VisAWI and classical/ expressive questionnaire to find out the reason, why the results are still so different.

The questions of VisAWI

Simplicity:
1. The layout appears too dense (r)
2. The layout is easy to grasp
3. Everything goes together on this site
4. The site appears patchy (r)
5. The layout appears well structured

Diversity:
1. The layout is pleasantly varied
2. The layout is inventive
3. The design appears uninspired (r)
4. The layout appears dynamic
5. The design is uninteresting (r)

Colorfulness:
1. The color composition is attractive
2. The colors do not match (r)
3. The choice of colors is botched (r)
4. The colors are appealing

Craftmanship:
1. The layout appears professionally designed
2. The layout is not up-to-date (r)
3. The site is designed with care
4. The design of the site lacks a concept (r)

All evaluated websites and its extracted objects.

Mbar

MOCA

Site / measure	Amoda	Bauhaus	Lacda	MAM Paris	Mbar	MOCA	Psyche	SF MoMa
balance BM	0,167	0,687	0,454	0,968	0,691	0,890	0,832	0,812
equilibrium EM	0,933	0,931	0,938	0,998	0,947	0,988	0,981	0,956
symmetry SYM	0,304	0,856	0,649	0,902	0,871	0,874	0,750	0,931
sequence SQM	0,750	0,750	0,750	1,000	1,000	1,000	1,000	0,750
cohesion CM	0,467	0,458	0,377	0,539	0,480	0,467	0,254	0,404
unity UM	0,466	0,716	0,477	0,179	0,252	0,227	0,520	0,383
proportion PM	0,835	0,854	0,735	0,771	0,717	0,868	0,692	0,668
simplicity SMM	0,333	0,100	0,143	0,103	0,086	0,079	0,200	0,097
density DM	0,722	0,471	0,651	0,806	0,635	0,832	0,508	0,869
regularity RM	0,438	0,445	0,415	0,674	0,529	0,619	0,506	0,571
economy ECM	0,250	0,167	0,143	0,100	0,091	0,063	0,143	0,100
homogeneity HM	0,500	0,023	0,067	0,038	0,015	0,000	0,000	0,000
rhythm RHM	0,374	0,065	0,386	0,056	0,220	0,160	0,094	0,267
Overall measure OM (average)	0,503	0,502	0,476	0,549	0,503	0,544	0,499	0,524
Rank	4	6	8	1	5	2	7	3

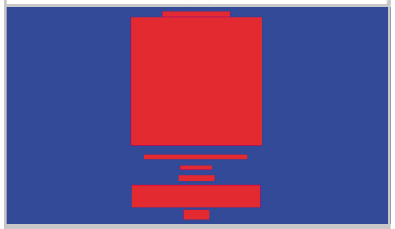
Table 1. The normalized results of evaluation. Objective method of Modelling Interface Aesthetics.

	Amoda	Bauhaus	Lacda	MAM Paris	Mbar	MOCA	Psyche	SF MoMa
simplicity	0,572	0,759	0,651	0,683	0,502	0,294	0,693	0,722
diversity	0,394	0,700	0,528	0,738	0,454	0,377	0,648	0,712
colourfulness	0,604	0,744	0,545	0,773	0,491	0,391	0,718	0,733
craftsmanship	0,522	0,747	0,558	0,772	0,450	0,383	0,626	0,731
classic	0,578	0,780	0,581	0,747	0,452	0,353	0,696	0,761
expressive	0,376	0,699	0,481	0,724	0,450	0,340	0,644	0,706
Average	0,508	0,738	0,557	0,739	0,467	0,356	0,671	0,728
Rank	6	2	5	1	7	8	4	3

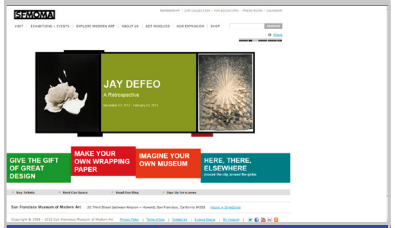
Table 2. The normalized results of evaluation. Subjective methods VisAWI, Classical and Expressive aesthetics.

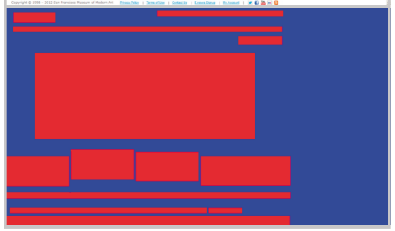
All evaluated websites and its extracted objects.

Psyche



SF MoMa





measure	simplicity	diversity	colourfulness	craftsmanship	classical	expressive
balance BM	-0,028	0,505	0,197	0,279	0,141	0,502
equilibrium EM	-0,267	0,199	0,051	0,046	-0,070	0,183
symmetry SYM	0,035	0,507	0,133	0,282	0,137	0,506
sequence SQM	-0,468	-0,105	-0,242	-0,303	-0,388	-0,087
cohesion CM	-0,284	-0,140	-0,136	-0,014	-0,174	-0,136
unity UM	0,577	0,248	0,381	0,333	0,481	0,258
proportion PM	-0,426	-0,405	-0,303	-0,256	-0,306	-0,410
simplicity SMM	0,103	-0,320	0,094	-0,114	0,050	-0,303
density DM	-0,389	-0,148	-0,207	-0,112	-0,232	-0,176
regularity RM	-0,294	0,219	0,043	0,105	-0,066	0,216
economy ECM	0,362	-0,125	0,269	0,123	0,280	-0,108
homogeneity HM	-0,063	-0,463	-0,058	-0,192	-0,089	-0,448
rhythm RHM	-0,146	-0,536	-0,404	-0,406	-0,317	-0,559

Table 3. Correlations between subjective facets and objective measures.

Classical aesthetics asks if visual design is:
1. Aesthetical
2. Clear
3. Pleasant
4. Clean
5. Symmetrical

Expressive aesthetics asks if visual design is:
1. Creative
2. Fascinating
3. Original
4. Sophisticated

All questions were answered on 7-point scale, where 1 means "not agree at all" and 7 means "agree very much". The questions with (r) are reversed prior analysing data so, that final result reflects the aesthetical coefficient in correct way (higher score is more aesthetical). All questions were averaged inside the facet group before output. All results were normalized before comparison with objective measures.

In list of questions are few, that could by description be directly related to objective measures. For example:
- simplicity 1 relates to density
- classical 5 relates to symmetry

Some of them might be related to several measures:
- simplicity 2 might relate to simplicity, unity, regularity, density and economy
- simplicity 3 might relate to regularity and unity
- simplicity 4 might relate to density and homogeneity
- simplicity 5 might relate to sequence, regularity, homogeneity and rhythm
- classical 2 might relate to simplicity and economy
- classical 4 might relate to simplicity and economy

Other questions can not be directly related to any of the objective measures, because the answers are based on emotions. For example, if layout appears pleasantly varied (diversity 1) for user, it does not give a clue, on what basis the decision was made.

Conclusion and Further Study

The experiment has proved that existing model is not applicable in interface design process, but it has revealed several ideas for making improvements to the model. Following subsections describe the issues for further study and answer the research question: how to improve the model of visual interface aesthetics?

Missing measures.

Both compared methods have measures, that are not included in another method. Thielsch et al [7] concluded that use of color has more importance than other aspects when evaluating visual aesthetics, but the model of objective evaluation did not use any measure of lightness or color at all. Adding the measures of color and lightness will be the most important task for improving the model.

The additional empirical study

The questions of subjective method reveal that measures of balance, equilibrium, cohesion and proportion are not directly covered in subjective method. In objective method the weight of those parameters is ca 0,3 (4/13) of overall weight. Therefore was highly predictable, that final results of both methods will have weak correlation. Additional empirical study can support the main study to see if this problem is relevant. Additional study might be conducted in two parts: the questionnaire and interface critics by experts. The questionnaire of additional study will contain just one question: is the viewed website aesthetical? This question together with interface critics provide the aesthetic value directly, without the use of its aspects. Correlation analysis of the results in main study and additional study can reveal if objective methods measures of equilibrium, cohesion and proportion are relevant.

Extraction of layout objects

The objects with non rectangular shape were approximated to rectangular - this can alter the initial designs aesthetics. Extracting objects of layout is still subjective and results of assessing can return different results depending on who did the extraction. Describing more strictly the rules for objects extraction will improve the model. Automating the extraction of objects will make the model more objective. The suggestion is to use automated image analysis for extraction of objects.

Color, lightness and shape of the object.

All objective measures in this experiment are based on 10 year old study [8], but many principles of design have been changed since. Modern design looks more like real life photo or video, not like drawing on white paper. Thus colored object is not always heavier than black and white and neither is dark object neccessarily heavier than light in balance of layout. Objects in layout must be analysed together with negative space in composition - the area between the objects. Dark background gives more weight to light objects, affecting balance and equilibrium. Contrast between objects and background affects also clarity measure. The definition of objects weight in measure of balance and equilibrium will change when eliminating the restriction of objects rectangularity. Suggestion is to use optical weight of objects in addition to geometrical weight. Optical weight depends on objects color, lightness and contrast to background.

The weight of measures

The problems occurred when comparing objective and subjective methods of evaluation: Single facets of subjective method can not be compared with single measures of objective method, because of overlapping aspects across the methods. The additional empirical study is recommended to reveal how important are single aspects of aesthetics and then correct the model accordingly. The model allows to correct all measures independently with weight constants. Suggestion to further study is to calculate the values of weight constants.

References

1. Altaboli, A., & Lin, Y. Objective and Subjective Measures of Visual Aesthetics of Website Interface Design: The Two Sides of the Coin. *Human-Computer Interaction, Part I, HCII 2011, 35–44.*
2. Bardzell, J. Interaction Criticism and Aesthetics. *CHI 2009 April 4–9, 2009, Boston, MA, USA.*
3. Birkhoff, G. Aesthetic Measure. *Cambridge, Harvard University Press (1933).*
4. Lavie, T., & Tractinsky, N. Assessing Dimensions of Perceived Visual Aesthetics of Websites. *International Journal of Human-Computer Studies, 60(3), 269-298 (2003).*
5. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., & Noonan, P. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction, 18(1), 1-30. (2011).*
6. Mahlke, S., Lemke, I., & Thüring, M. The Diversity of Non-instrumental Qualities in Human-Technology Interaction. *MMI-Interaktiv (13), 55-64. (2007).*
7. Moshagen, M. & Thielsch, M. Facets of visual aesthetics. *International Journal of Human-Computer Studies, 68(10), 689-709 (2010).*
8. Ngo, C., L., N, Teo, L., S., Byrne, J., G. Modelling interface aesthetics. *Information Sciences 152 (2003), Elsevier*
9. Ngo, D. C. L., Samsudin, A., & Abdullah, R. Aesthetic measures for assessing graphic screens. *Journal of Information Science and Engineering, 16(1) (2000).*
10. Pajusalu, M. The Evaluation of User Interface Aesthetics. *Tallinn University Institute of Informatics (2012) http://www.cs.tlu.ee/teemad/get_file.php?id=202*
11. Papachristos, E., & Avouris, N. Are first impressions about websites only related to visual appeal? *Human-Computer Interaction–INTERACT 2011, 489–496. Springer.*
12. Strebe, R. Visual aesthetics of websites: the visceral level of perception and its influence on user behaviour. *Research and Advanced Technology for Digital Libraries, 523–526. Springer. (2011).*
13. Zain, J. M., Tey, M., & Soon, G. Y. Using Aesthetic Measurement Application (AMA) to Measure Aesthetics of Web Page Interfaces. *2008 Fourth International Conference on Natural Computation, 96-100. Ieee. doi:10.1109/ICNC.2008.764*
14. Zhang, Q., Kang, W., Zhao, C., & Ming, X. Aesthetic Coloring for Complex Layout Using Genetic Algorithm. *2009 WRI Global Congress on Intelligent Systems, 406-410.*
15. Xenakis I., Arnellos A., Darzentas J. The functional role of emotions in aesthetic judgment. New Ideas in Psychology 30 (2012) 212–226