Entropy and Compression Based Analysis of Web User Interfaces

Egor Boychuk, **Maxim Bakaev**
(bakaev@corp.nstu.ru)
Novosibirsk State Technical University

International Conference on Web Engineering (ICWE)
Daejeon, Korea, 2019

Introduction – Motivation

- Automated evaluation of web UIs:
  - gains popularity in WE, as number and diversity of WUIs increase, while update cycles shorten and budgets tighten
  - quantitative analysis (UI metrics) and prediction of users' subjective impressions (visual complexity, aesthetics, etc.)
  - code-based analysis: fast and accurate, but can be poor at evaluating quality-in-use and visual perception-related metrics
  - image-based analysis: robust, but can be computationally expensive (image elements recognition)
- Popular and fast metrics for images:
  - various information theoretic (entropy) metrics
  - byte sizes in compressed formats (JPEG and PNG)
  - color-related metrics (RGB, HSV, LAB)
  - Zipf's law, Subband Entropy, grids, edges, etc.

Image-Based WUI Analysis

- PNG File Size (in bytes): 636471
- JPG File Size (in bytes): 138924
- Mean Lightness: 80.29, Mean A (Green-Red Space):1.56, Mean B (Yellow-Blue Space) -3.19
- Edge Density: 0.14, Number of Alignment Lines: 121

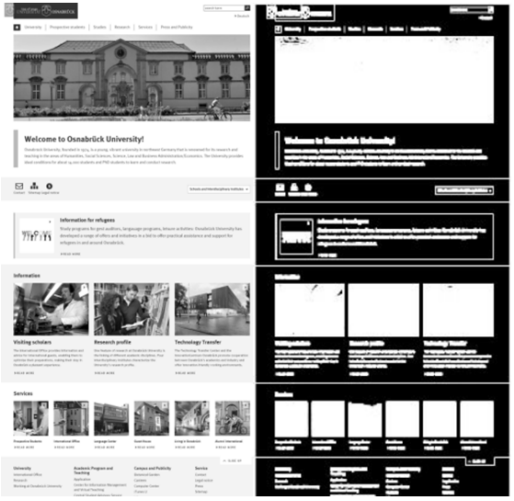
Metrics provided by AIM service (Aalto University)

RESEARCH PROBLEM:

Explore applicability of simple image compression and entropy metrics for predicting impressions of WUI users.

METHODS AND TOOLS:

- Subband Entropy metric for raster images, as implemented in Matlab (negates graphic content)
- JPEG-100 and PNG-24 metics
- Collecting subjective assessments from human subjects in an experimental survey



The Experimental Survey

- Hypotheses:
  1. Metrics based on compression algorithms (that are already know to work for images) are predictive of web UI's visual perception.
  2. The metric based on information entropy can further improve the predictive power.
- Material:
  - Homepages of 497 websites belonging to universities and colleges (from 10639 screenshots collected by our Python web crawler)
  - Full web pages, not just the part above the fold or fixed size
  - Relatively unknown organizations, all content in English
- Main independent variables:
  - size of the screenshot file in PNG-24: *PNG_size*
  - size of the screenshot file in JPEG-100: *JPEG_size*
  - Matlab's entropy value for the .png file: *M_Entropy*

The Experimental Survey

- Participants:
  - students of NSTU and IT professionals
  - altogether 70 people (43 females, 27 males)
  - age from 18 to 29 (mean 20.86, SD = 1.75)
- Dependent variables (1-7 Likert scales):
  - how visually complex the WUI appears: *Complex*
  - how aesthetically pleasant the WUI appears: *Aesthetic*
  - how orderly the WUI appears: *Orderly*
- Procedure:
  - online questionnaire showing the screenshots (fixed sizes for all)
  - random order of presenting the stimuli (from the 497 screenshots)
  - all the 3 scales are mandatory
  - the default number of evaluated screenshots: 50 + another 50 opt.

Descriptive Statistics

- In total, 4235 full evaluations per the 3 scales each
- Each WUI evaluated by 8-10 participants (mean 8.52)
- Each participant on average evaluated 60.1 WUIs
- Valid WUIs: 493 (99.2%), while 4 were discarded
- Shapiro-Wilks: normality hypothesis had to be rejected for *Orderly* (p = 0.002), but not for *Complex* (p = 0.622) and *Aesthetic* (p = 0.085).

Variable	Range	Mean (SD)
width	1440-2862	1448 (80.89)
height	900-16571	2586 (1379.95)
JPEG_size	0.31-10.11	2.00 (1.08)
PNG_size	0.13-11.20	1.95 (1.28)
M_Entropy	1.78-7.80	5.05 (0.95)
Complex	1.75-5.78	3.58 (0.65)
Aesthetic	1.71-6.25	4.12 (0.86)
Orderly	2.11-6.13	4.44 (0.64)

Highest and Lowest Complexity

Correlation Analysis

- Pearson's and Kendall's (tau-b, non-parametric statistic for ordinal scales), highest correlations per scales highlighted
  - No correlation between *Complex* and *Aesthetic* (unexpected)
  - High correlation between *Aesthetic* and *Orderly* (domain-specific?)
  - *M_Entropy* correlated with *JPEG_size* and *PNG_size* (expected)
  - *M_Entropy* correlated with *Orderly* (unexpected for frequency entropy)
  - *JPEG_Size* most correlated with *Complex* (will use as the baseline)

	JPEG_size	PNG_size	M_Entropy	Complex	Aesthetic	Orderly
JPEG_size	1	r = 0.951 p < 0.001	r = 0.315 p < 0.001	**τ = 0.143** **p < 0.001**	τ = 0.257 p < 0.001	τ = 0.193 p < 0.001
PNG_size		1	r = 0.391 p < 0.001	τ = 0.102 p = 0.001	**τ = 0.325** **p < 0.001**	**τ = 0.233** **p < 0.001**
M_Entropy			1	τ = -0.025 p = 0.413	τ = 0.247 p < 0.001	τ = 0.186 p < 0.001
Complex				1	τ = -0.030 p = 0.329	τ = -0.078 p = 0.011
Aesthetic					1	τ = 0.574 p < 0.001
Orderly						1

Regression Analysis

- Baseline models for the subjective scales:

$Complex = 3.316 + 0.133 \times JPEG\_size$ R^2=0.05, p<0.001, AIC=2602

$Aesthetic = 3.609 + 0.254 \times JPEG\_size$ R^2=0.103, p<0.001, AIC=2855

$Orderly = 4.218 + 0.109 \times JPEG\_size$ R^2=0.034, p<0.001, AIC=2604

- Extended models for the subjective scales:

$Complex = 3.504 + 0.504 \times JPEG\_size - 0.316 \times PNG\_size - 0.063 \times M\_Entropy$

  - R^2=0.105, p<0.001, AIC=2576, highest Beta: *JPEG_size*

$Aesthetic = 2.731 - 0.373 \times JPEG\_size + 0.503 \times PNG\_size + 0.229 \times M\_Entropy$

  - R^2=0.248, p<0.001, AIC=2772, highest Beta: *PNG_size*

$Orderly = 3.541 - 0.188 \times JPEG\_size + 0.225 \times PNG\_size + 0.166 \times M\_Entropy$

  - R^2=0.127, p<0.001, AIC=2558, highest Beta: *PNG_size*

- The extended models are better for all the scales

Conclusions

- Compressed and entropy based metrics are reasonably applicable for WUIs
  - there are domain-specific findings (like *Complex* vs. *Aesthetic*)
- We proposed extended regression models for Complex (+110%), Aesthetic (+141%), Orderly (+274%)
  - only fast to calculate image-based metrics were used
- We introduced frequency-based entropy metric for analysis of WUI perception in humans
  - *M_Entropy* significant in all the models at 0.052
  - entropy increases together with orderliness, but decreases with complexity perception (unexpected, but we find it the second time)
- Limitations: low R^2s, low fidelity of the subjective scales
- Prospects: integration of code and image based metrics