Subjectivity in Aesthetic Quality Assessment of Digital Photographs: Analysis of User Comments

Won-Hee Kim
Yonsei Institute of Convergence Technology, Yonsei University
Incheon, Republic of Korea
wonheekim@yonsei.ac.kr

Jun-Ho Choi
School of Integrated Technology & Yonsei Institute of Convergence Technology, Yonsei University
Incheon, Republic of Korea
idearibosome@yonsei.ac.kr

Jong-Seok Lee
School of Integrated Technology & Yonsei Institute of Convergence Technology, Yonsei University
Incheon, Republic of Korea
jong-seok.lee@yonsei.ac.kr

ABSTRACT
While most of the existing work in aesthetic image quality assessment focuses on the overall (or average) opinion of users, this paper raises the issue of subjectivity (or taste) of aesthetic quality. We argue that subjectivity differs among different images, and investigate what causes such difference. We first analyze statistics of the user ratings of photos in a photo contest website, DPChallenge, in the viewpoint of average and standard deviation values of the ratings. Then, more importantly, we analyze the users' comments in order to identify sources contributing to subjectivity. When considering the importance of personalization in photo applications, we believe that our findings will be a valuable first step in the relevant future research.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

General Terms
Measurement, Experimentation, Human Factors

Keywords
Aesthetic quality assessment, subjectivity, user comments, digital photographic image

1. INTRODUCTION
With the rapid growth of the imaging and mobile technologies, taking digital photographs has become a daily life activity these days. The number of photographs possessed personally or available online has grown explosively. Accordingly, users' interests and demands toward aesthetically pleasing photographs have also increased significantly. People want to take, share, and view photographs that have high aesthetic quality. Therefore, automatic assessment of aesthetic quality of photographs is a promising technique in many applications related to production, enhancement, management, retrieval, and recommendation of photographs. In general, however, it is very challenging because different people have different aesthetic tastes and preferences.

Over the past decade, aesthetic image quality assessment has been studied by several researchers. Most conventional work used image features to measure the aesthetics of photos. The initial work concentrated on low level features. In the work by Tong *et al.* [1], a large set of general purpose low level features related to blurness, contrast, colorfulness, saliency, color, energy, texture, and shape is used to classify photos as professional or snap-shots. Most of these features are not designed particularly for measuring the aesthetic properties of photos, but for image classification such as indoor vs. outdoor images, city vs. landscape pictures, photos vs. paintings, etc. Therefore, they are not sufficient to accurately evaluate aesthetic photo quality.

To overcome this limitation, Ke *et al*. [2] proposed a photo quality assessment method using high level features that are designed to distinguish high and low quality photos, such as edge distribution, color distribution, hue count, and blurriness. Datta *et al*. [3] proposed an approach using high level features related to light, colorfulness saturation, hue, rule of thirds, familiarity, texture, size ratio, region composition, depth of field, and shape.

There exists some work on aesthetic quality assessment using other high level features. Wong and Low [4] used a set of salient features that characterize the subject and the subject-background relationship. Li *et al*. [5] represented the aesthetic quality of a photo with faces by using technical, perceptual, and social relationship features. Marchesotti *et al*. [6] used generic image descriptors such as bag-of-visual words, Fisher vector, and GIST. Zhang *et al*. [7] introduced photo aesthetics evaluation framework, focusing on learning the image descriptors that characterize local and global structural aesthetics from multiple visual channels.

These researches aimed at evaluating the level of aesthetic photo quality. Their main tasks are prediction and classification of average aesthetic scores. Therefore, they ignore the subjectivity in aesthetic quality. Recently, there have been a few attempts to understand the subjectivity of image assessment in the psychological perspective.

Fedorovskaya and De Ridder [8] showed the existence of subjectivity in assessment of image quality degradation. Palmer and Schloss [9] demonstrated that the gender, expertise, culture, and perceptual experience affect color preference of individuals. Chu *et al*. [10] showed that the familiarity exerts a major influence on determining the perceived interestingness of images. Fedrizzi [11] argued that the gender plays an important role in the aesthetic perception by examining the brain activity recorded

MM '15, October 26-30, 2015, Brisbane, Australia
© 2015 ACM. ISBN 978-1-4503-3459-4/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2733373.2806380

using magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI).

These studies are limited in that they showed only the existence of subjectivity or examined effects of only certain factors on only certain aesthetic aspects. In this paper, we go one step further by considering more fundamental questions regarding subjectivity and examining much larger number of data, in order to obtain more comprehensive understanding about subjectivity in aesthetic photo quality assessment. We analyze of evaluations (ratings and comments given by online users) for a significantly large number of photos. We aim at answering questions such as which factors in photos contributes to aesthetic subjectivity, how the average aesthetic quality level is related to subjectivity, how the image content influences subjectivity, etc.

2. DATA SET DESCRIPTION

We use photo and comment data available in DPChallenge (http://www.dpchallenge.com), which is an online photograph community website. It hosts weekly digital photography contests called challenges for particular topics. Users of the community can rate a photo submitted to a challenge from 1 (worst) to 10 (best), and can also write comments. At the end of the week, the photos in the challenge are ranked by their mean rating values and the winner is announced. Naturally, photos submitted to a challenge significantly vary in terms of aesthetic quality. The characteristics of the users also vary significantly, i.e., from amateurs to professionals.

Total 307,132 photos and their ratings in 1,971 challenges (from January 2002 to August 2014) were collected for our study. Figure 1 shows the histogram of the mean rating scores, which is close to a normal distribution. The mean and standard deviation values of the distribution are 5.43 and 0.73, respectively; the minimum and maximum values are 1.81 and 8.60, respectively.

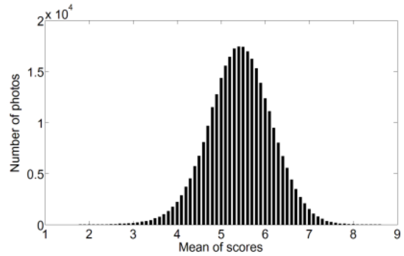


Figure 1. Histogram of the mean ratings

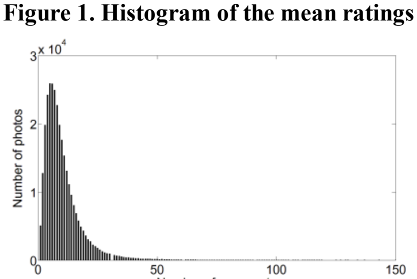


Figure 2. Histogram of the number of comments

In addition, the comments for the photos were collected. We considered only the comments made during the challenge period in our analysis, since comments after the challenge period may be biased due to the announced ranking result. The number of such valid comments was 2,828,525. The maximum number of comments per photo is 145, and the average number of comments per photo is 9.4. Figure 2 shows the histogram of the number of comments. 69.7% of the photos have less than 10 comments and 92.6% photos have less than 20 comments.

In order to analyze subjectivity based on the content of comments, we set six broad categories, each of which contains several challenges sharing the same topic. In total, 18,033 photos and the 182,368 comments for them are analyzed. Table 1 summarizes the statistics of each category, i.e., the number of photos, total number of comments, total number of words in the comments, average of mean ratings (***M***), and average of standard deviations of ratings (***S***).

Table 1. Summary of the data set used for comment analysis

category	#photos	#comments	#words	avg(*M*)	avg(*S*)
portrait	6374	80352	1339132	5.49	1.41
landscapes	3242	29257	460592	5.53	1.38
architecture	1335	9723	151708	5.52	1.38
animal	3665	37541	577239	5.58	1.35
plant	2135	14794	229205	5.44	1.31
street	1282	10701	176637	5.51	1.31

3. SUBJECTIVITY IN ASSESSMENT

The mean value of ratings (denoted as ***M***) of a photo, shown in Figure 1, represents the aggregated overall level of aesthetic quality of the photo. The standard deviation of ratings (denoted as ***S***), on the other hand, represents the level of disagreement or diversity in rating across users, which reflects subjectivity in aesthetic quality evaluation. A low value of ***S*** means agreement among raters about the aesthetic quality of the photo, whereas a large value of ***S*** indicates that some users rate the photo high and some other users rate it low, and thus individual preference significantly differs among them.

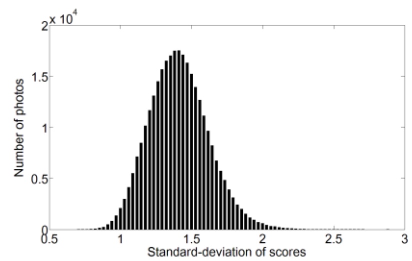
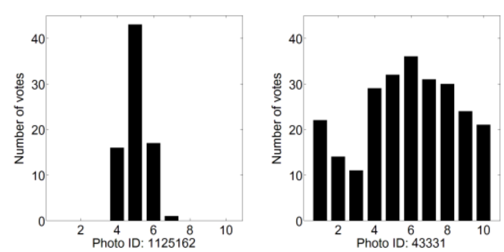


Figure 3. Histogram of the standard deviations of the ratings



(a) **(b)**

Figure 4. Example histograms of ratings. (a) A case with low subjectivity (*S*=0.69) (b) A case with high subjectivity (*S*=2.62)

Figure 3 shows the histogram of ***S*** for the 307,132 photos. The mean and standard deviation values of the distribution are 1.41 and 0.21, respectively; the minimum and maximum values are 0.69 and 2.89, respectively. As in Figure 1, the histogram of ***S*** follows approximately a normal distribution. Figure 4 shows rating distributions of two particular photos, one with a low ***S*** value and another one with a high ***S*** value. In Figure 4(a), most of the ratings are concentrated between 4 and 6, but in Figure 4(b), the whole rating range (from 1 to 10) is used by raters.

We examine the correlation between ***M*** and ***S***, but they do not have significant correlation, with a Pearson correlation coefficient of 0.086.

4. ANALYSIS OF COMMENTS

In this section, aesthetic photo quality assessment and its subjectivity are analyzed based on the users' comments.

The relationship between the number of comments and the ratings is shown in Figures 5 and 6. Figure 5 presents the scatter plot of the number of comments with respect to ***M***. Overall, it is observed that photos having higher ***M*** values tend to have more comments. In addition, the distribution of the number of comments is rather uniform in the extremely low range of ***M*** (lower than 3), while it is concentrated in the low range (below about 30 comments) for the photos in the mid range of ***M*** (between 4 and 5). Therefore, it seems that the commenters are more interested in high or low quality photos than the mid-level quality photos.

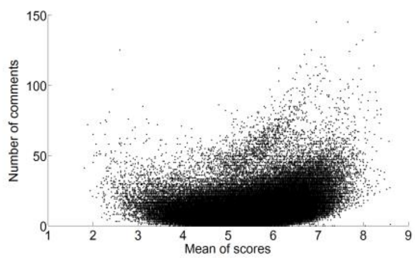


Figure 5. Scatter plot of the number of comments with respect to the mean rating values (*M*)

Figure 6 shows the scatter plot of the number of comments with respect to ***S***. A weak positive correlation is observed on the whole. It is also interesting to see that photos having extremely low ***S*** values (i.e., having almost perfect agreement among raters) received very low numbers of comments. Therefore, it can be said that photos incurring higher subjectivity tend to receive more comments.

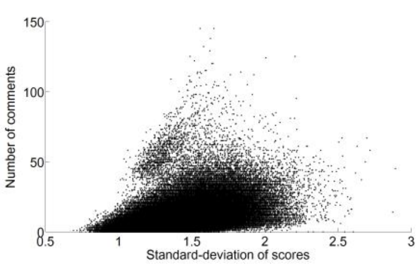


Figure 6. Scatter plot of the number of comments with respect to the standard deviation of ratings (*S*)

In order to analyze subjectivity from the content of comments, we identify words conveying users' opinions and categorize them into six groups as shown Table 2. The words were manually classified into six groups. The POSI group is composed of words representing positive feedback. Especially, the word 'ribbon' means the markers given to the photos ranked first, second, and third. The words in the CREA group evaluate the originality and creativity of photos. The NEGA group consists of words with negative and critical meanings. PERS is the group consisting of words used to describe personal and individual opinions. The words in the BOTH group indicate that the comments simultaneously include contrasting positive and negative opinions (for instance, "The photo has good color balance **but** the composition is poor."). Finally, the words in the ELEM group represent particular elements of photo aesthetics. For each photo, the frequency of the words listed in the table is counted for each of the six groups for further analysis.

Table 2. Word groups for comment analysis

group	words
POSI	amazing, beautiful, excellent, fabulous, fantastic, top, favorite, good, gorgeous, great, like, love, nice, perfect, please, sweet, wonderful, awesome, best, ribbon, well
CREA	creative, interesting, novel, origin, concept, idea, new, unique, unusual, not usual
NEGA	artifact, lack, miss, need, poor, require, bad, blurry, noise, sorry, why, out of focus
PERS	personal, taste, because, me, my, think, thought
BOTH	although, but, however, nevertheless, notwithstanding, nonetheless, though, despite that, even so, even though,
ELEM	align, balance, color, curve, detail, drama, emotion, expose, light, line, mood, object, rotation, simple, tone, texture, angle, background, brightness, composition, contrast, foreground, harmony, hue, perspective, pose, ratio, saturation, symmetry, viewpoint, depth of field

Table 3 shows the Pearson correlation coefficients between the total word frequencies and the mean rating values (***M***). The POSI group has a high correlation with ***M***, which is obviously expected. The ELEM group also has a relatively high correlation. The *portrait* category receives more influence of the words in the CREA group than any other categories. Since most people have seen a lot of portraits of many different kinds, originality and creativity tend to affect positively to evaluate the aesthetic quality. In contrast, CREA is not related to the evaluation of the *landscapes* and *animal* categories. It means that the assessment criteria of the categories are rather conservative and classical.

Table 3. Correlation coefficients for *M*

category	POSI	CREA	NEGA	PERS	BOTH	ELEM
portrait	0.656	0.202	-0.169	0.318	0.170	0.499
landscapes	0.663	0.039	-0.247	0.209	0.021	0.457
architecture	0.628	0.152	-0.186	0.353	0.088	0.483
animal	0.658	0.003	-0.205	0.276	-0.002	0.462
plant	0.711	0.027	-0.151	0.353	0.167	0.552
street	0.752	0.163	-0.130	0.483	0.209	0.502

Table 4 presents the Pearson correlation coefficients between the word frequencies and the standard deviation values of ratings (***S***). In comparison to Table 3, one can observe significantly increased correlations of ***S*** with the word frequencies in the CREA, PERS, and BOTH groups. Existence of the words in the CREA group indicates that the photo is unusual in aesthetic quality. The creativity and unusualness act positively to some users but negatively to some other users. When a commenter uses a word in the PERS group, he/she already knows that his/her opinion is subjective and thus may not be accepted by some users. Furthermore, use of a word in the BOTH group indicates mutually contradictory opinion within a comment; depending on personal preference, the positive or negative side plays a more important role, which results in subjectivity. The photos in *landscapes* show a relatively high correlation between the frequency of the ELEM words and ***S***. This means that subjectivity of the landscape photos is largely evaluated by low level aesthetic criteria (e.g., composition, color, and lighting) rather than high level perceptual aspects such as emotion and feeling. Exceptionally, the *street* category has low correlations for all groups. This is probably due to the fact that the photos in this category significantly vary in terms of their content, ranging from faces of people in streets, cars and buildings in streets, to landscapes containing streets.

Table 4. Correlation coefficients for *S*

category	POSI	CREA	NEGA	PERS	BOTH	ELEM
portrait	0.376	0.321	0.186	0.453	0.412	0.255
landscapes	0.514	0.325	0.219	0.456	0.442	0.458
architecture	0.495	0.326	0.186	0.465	0.412	0.392
animal	0.455	0.367	0.193	0.484	0.429	0.278
plant	0.507	0.322	0.107	0.429	0.391	0.367
street	0.379	0.223	0.091	0.346	0.361	0.246

Linear regression is performed to examine to which extent the overall aesthetic quality level (i.e., ***M***) and subjectivity (i.e., ***S***) are explained by the word frequencies of the six groups. The total word frequencies of the groups, X_i (i=1,…,6), are considered as independent variables. And, the dependent variable, Y, is ***M*** or ***S***. We aim to find the regression coefficients β_i for each X_i of the linear regression model given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \qquad (1)$$

where ε means the prediction error. Table 5 shows the results of the regression analysis. For ***M***, the correlation coefficients are higher than 0.7 in most cases. A lower correlation for the *architecture* category is mainly due to a weak correlation of the POSI group in Table 3. For ***S***, correlation coefficients up to about 0.6 are obtained, while the category-dependence is higher than for ***M***. A lower correlation for the *street* category than the other categories is due to the low correlations shown in Table 4. The highest correlations are obtained for the *animal* and *landscape* categories (0.595 and 0.581, respectively). These observations indicate that the extent to which creativity and unusualness (CREA), personal preference (PERS), and simultaneous existence of aesthetically positive and negative aspects (BOTH) explain subjectivity depends on the image content.

Table 5. Regression results in terms of Pearson correlation coefficient

category	*M*	*S*
portrait	0.718	0.497
landscapes	0.744	0.581
architecture	0.684	0.560
animal	0.732	0.595
plant	0.739	0.578
street	0.761	0.448

5. CONCLUSION

In this paper, we investigated subjectivity in aesthetic photo quality assessment. By analyzing the words in users' comments for photos in DPChallenge, we investigated what contributes to aesthetic subjectivity. It was found that creativity/unusualness and simultaneous existence of positive and negative aspects are notable sources provoking subjectivity. It was also shown that the image content plays an important role in determining the acceptability of unusualness and the amount of influence of basic low level aesthetic elements to subjectivity.

6. ACKNOWLEDGMENTS

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the "IT Consilience Creative Program" (IITP-2015-R0346-15-1008) supervised by the IITP (Institute for Information & Communications Technology Promotion).

7. REFERENCES

[1] Tong, H., Li, M., Zhang, H. -J., He, J., and Zhang, C. 2005. Classification of digital photos taken by photographers or home users. In *Proceedings of Pacific Rim Conference on Advances in Multimedia Information Processing*, 198-205.

[2] Ke, Y., Tang, X., and Jing, F. 2006. The design of high-level features for photo quality assessment. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 419-426.

[3] Datta, R., Joshi, D., Li, J., and Wang, J. 2006. Studying aesthetics in photographic images using a computational approach. In *Proceedings of European Conference on Computer Vision*, 288-301.

[4] Wong, L. -K. and Low, K. –L. 2009. Saliency-enhanced image aesthetics class prediction. In *Proceedings of IEEE International Conference on Image Processing,* 997-1000.

[5] Li, C., Andrew, G., Loui, A. C., and Chan, T. 2010. Aesthetic quality assessment of consumer photos with faces. In *Proceedings of IEEE International Conference on Image Processing,* 3221-3224.

[6] Marchesotti, L., Perponnin, F., Larlus, D., and Csurka, G. 2011. Assessing the aesthetic quality of photographs using generic image descriptors. In *Proceedings of IEEE International Conference on Computer Vision,* 1784-1791.

[7] Zhang, L., Gao, Y., Zimmermann, R., Tian, Q., and Li, X. 2014. Fusion of multichannel local and global structural cues for photo aesthetics evaluation. *IEEE Transactions on Image Processing, 23*, 3, 1419-1429.

[8] Fedorovskaya, E. A. and De Ridder, H. 2013. Subjective matters: from image quality to image psychology. In *Proceedings of IS&T/SPIE Electronic Imaging*, 86510O-1-86510O-11.

[9] Palmer, S. E. and Schloss, K. B. 2010. Human preference for individual colors. In *Proceedings of IS&T/SPIE Electronic Imaging*, 752718-1-752718-12.

[10] Chu, S. L., Fedorovskaya, E., Quek, F., and Snyder, J. 2013. The effect of familiarity on perceived interestingness of images. In *Proceedings of IS&T/SPIE Electronic Imaging*, 86511C-1-86511C-12.

[11] Fedrizzi, L. 2012. Beauty and its perception: historical development of concepts, neuroaesthetics, and gender-differences. *Rendiconti Lincei, 23*, 3, 259-269.