Real-time Guidance Camera Interface to Enhance Photo Aesthetic Quality

Yan Xu[1], Joshua Ratcliff[1], James Scovell[2], Gheric Speiginer[3], Ronald Azuma[1]
[1]Intel Labs   [2]Intel Corporation   [3]Georgia Institute of Technology
{yan.xu, joshua.j.ratcliff, james.j.scovell, ronald.t.azuma}@intel.com, gheric.speiginer@gatech.edu

ABSTRACT

This paper explores whether it is effective to use real-time on-screen guidance to help users take better photos with mobile devices. Using a three-camera array, we developed a photo-taking interface that provides real-time feedback on *how* to position the subject-of-interest according to a photography composition rule: rule-of-thirds. We conduct a user study to compare the aesthetic quality of photos taken with our real-time guidance interface against a static gridline interface common to existing digital cameras. Expert photographers and Mechanical Turk workers rate the aesthetic quality of these pairs of photos. Results indicate the photos taken with our real-time guidance interface have significantly higher aesthetics scores. This study shows the potential in using camera array, computational photography, and real-time guidance interface to help non-expert users take better photos.

Author Keywords

Camera User Interface; User Evaluation; Photo Aesthetics; Camera Array.

ACM Classification Keywords

H.5.2 User Interfaces: *Evaluation*

INTRODUCTION

Digital cameras on mobile devices enable users to take photos whenever and wherever they want. However, consistently taking good quality photos requires knowledge and skill that most casual users lack. Rather than requiring users to read books or take photography classes, what if a mobile device could analyze the scene and instruct a user, in real time, on how to compose a good quality photo? Would such an interface improve the aesthetic quality of photos taken by non-expert users?

These questions are timely and important because a new generation of mobile devices with multiple cameras and advanced computational photography capabilities is just

now reaching the market: examples include the *HTC One M8* smartphone (dual cameras) and the *Dell Venue 8 7000 series* tablets (three cameras). Such devices make this hypothetical camera guidance interface feasible by enabling a whole new set of scene analysis capabilities. Therefore it is now feasible and critical to test the usefulness of these interfaces. In this paper, we aim to answer the following question, *is real-time guidance an effective approach to increase the aesthetic quality of the photos?*

To answer this question, we designed and developed a camera user interface that provides real-time guidance. We focus on one photography rule: *rule-of-thirds*. We chose to implement this composition rule because it is widely adopted in photography, and it is implemented as static gridlines in many smartphone and digital cameras. Rule-of-thirds proposes that a photo should be imagined as divided into nine equal parts by two equally spaced horizontal lines and two equally spaced vertical lines, and that 1) important compositional elements should be placed along these lines or their intersection [11], and 2) the proportion of the subject of interest should be roughly one third of the total image space [12] (pp. 15–17). Our system analyzes every image frame using multiple camera inputs and gives real-time feedback according to rule-of-thirds.

We devised a user study that evaluates the effectiveness of this interface by comparing the end result—the aesthetic quality of photos taken with the real-time guidance interface and with the static gridline interface, in the specific scenario of single-person portraiture. Based on empirical data, this work reveals the potential of using camera arrays, computational photography, and real-time guidance to assist users in taking better photos.

RELATED WORK

Although photo aesthetics are subjective ("*beauty is in the eyes of the beholder*"), there are many rules and lessons from photography experts [9,11,12], which serve as a starting point for beginners to learn how to take better photos. Researchers from the image processing community leverage these photography rules, such as rule-of-thirds, simplicity, color compatibility, etc., and train classifiers using large sets of user-rated photos to distinguish between good and bad photos based on multiple aspects of aesthetic quality [2,3,5,6,7]. While most of the work is done on large datasets offline, two research projects stood out as most relevant to our work, as both gave nearly real-time feedback on photo aesthetics. Yao et al. built a mobile application

that gives three kinds of feedback after a photo is taken, including the aesthetic quality score, color combination, and other example photos that have similar composition which is automatically selected from a high-quality photo database [13]. Pedro and Church's system maps musical feedback to the aesthetic score of smartphone camera image frames [10]. The user feedback in both systems depends on the aesthetic scores calculated from a pre-trained model. Other than the scores, there is little instructional feedback. Users must use trial and error to determine *how* to improve the photo. In contrast, we provide real-time instructional feedback through our system. Moreover, although the above two projects reported user studies, the main question remain unanswered, does the aesthetic quality of the photos increase due to the interactive feedback on photo aesthetics?

SYSTEM DESIGN

This system is tailored for our user study where users take single-person portraiture photos. We choose this scenario both because it is one of the most common scenarios for mobile device photos [4], and as it is technically feasible to implement. Our system has three main components: 1) to find the region of interest (ROI); 2) to calculate how much the ROI aligns with rule-of-thirds division lines and intersections, and whether its proportion is close to one-third; 3) to render a real-time interface that dynamically suggests which direction users should move.

To find the ROI, which is the human subject in our user study, face detection and block matching stereo disparity calculation are combined. A three-camera array (left, right, and center) is attached to the back of an Ultrabook PC (Sony VAIO Duo 11, i7-3517U). We use the stereo correspondence algorithm from OpenCV [1] to get a disparity map from the left and right cameras and rectify it to the center camera. We detect faces by Haar Feature-based Cascade Classifiers from OpenCV [1]. We segment the pixels with similar disparity values to the detected face region as the ROI. The above method fits the needs of our user study, but non-human subjects would require different approaches, such as depth-based segmentation enabled by camera arrays.

For the second step, we first multiply the ROI with a bitmap mask highlighting the rule-of-thirds grid and crossings (see Figure 1). The numerical result measures how much the subject person aligns with the rule-of-thirds grid and crossings. Secondly, we calculate the proportion of ROI by dividing the width of human subject to the image width.

The real-time guidance user interface went through several iterations. In the final version, semi-transparent arrows are rendered and animated on top of the image and on two sides. These arrows direct users to move six directions—up, down, left, right, closer, and further away. We smooth the animation of the arrows to avoid jerkiness. The arrows on the two sides direct user to move closer to or further from the subject. They start to appear when the proportion

deviates from one-third beyond a small margin. The color saturation and speed of the side arrows depends on how much the difference is.
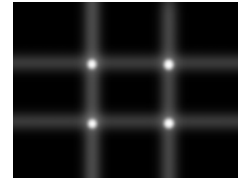
Figure 1. The bitmap mask for calculating the alignment between subject-of-interest and rule-of-thirds
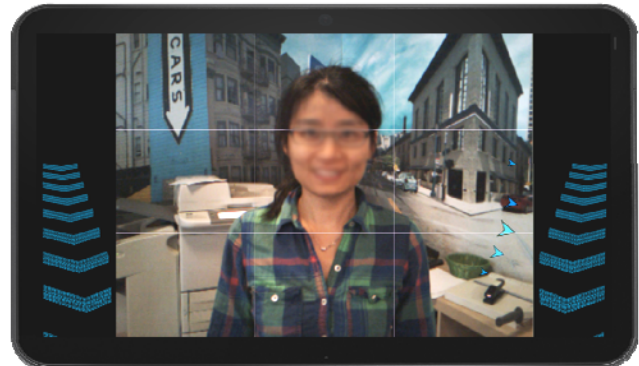
Figure 2. Real-time interface with animated arrows that guide users to follow Rule-of-Thirds. Here these arrows indicate that the user should move to the lower right and further away.

USER STUDY

The goal is to find whether the real-time guidance interface can effectively enhance photo aesthetics. We hypothesize that there is a significant difference in photo aesthetic quality between photos taken with real-time guidance interface and those taken with the static gridline interface.

Procedure and Participants

We recruited forty non-expert users through several company mailing lists. These participants had not previously taken classes or read books on photography. Each used both the real-time guidance tool and a static gridline tool to take photos of a friend, as it is closer to a real-life scenario than taking photos of a stranger. The order of interfaces was alternated to counterbalance order effects. Each pair of photos shared the same subject and similar backgrounds, which helped to control for the content-related aesthetic differences.

To avoid the problem of unintentional photo snapping, each user took up to five photos and chose his or her favorite one. To avoid the variance among photo takers who may or may not know rule-of-thirds before the study, we showed users an example photo that illustrates rule-of-thirds and explained its meaning beforehand. Users were given up to 5 minutes of learning time for each interface. We could not control the subjects' facial expression, body posture, or

changes in the background that was caused by different camera positions.
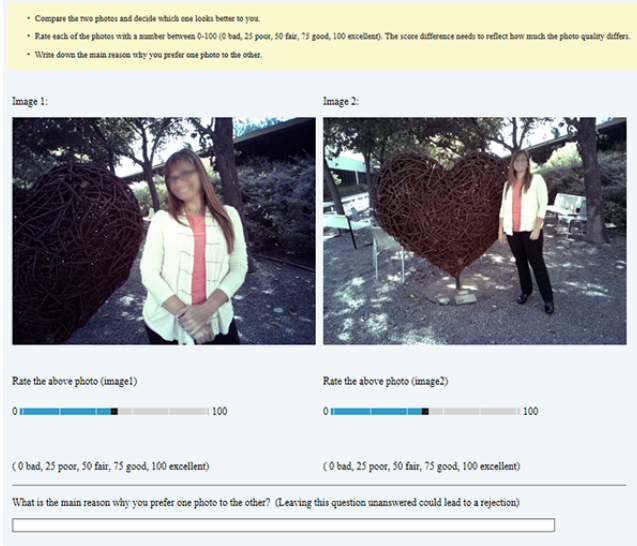
Figure 3. The web page for expert photographers and Mechanical Turk workers to rate photos and write comments

After these photos were taken with both interfaces, we recruited professional photographers and Mechanical Turk workers [8] to rate these 40 pairs of photos side-by-side using a 0-100 scale (Figure 3). Raters were also asked to write the main reason why they preferred one photo over the other. We recruited 24 professional photographers who specialize in portraiture with at least 2 years of full-time experience through an agency. On Mechanical Turk, we encouraged raters to review all forty pairs of photos by giving a 2-dollar bonus. We reject ratings completed in less than 10 seconds. In total, 48 of Mechanical Turk raters rated all 40 pairs of the photos. We randomize the order the photos to counterbalance any order effect.

We use within-subjects design so that individual differences in score ranges and evaluation criteria would not obscure the difference in photo aesthetic scores.

Findings

Our study shows that the aesthetic quality ratings for photos taken with the real-time guidance (RG) interface is significantly higher than those taken with the static gridline (SG) interface, according to both expert photographers and Mechanical Turk workers. Table 1 presents the means and standard deviations.

There were two independent variables, camera interface and photo-taker, and one dependent variable, photo's aesthetic rating. Using the two-factor repeated measures ANOVA, we found that expert photographers and Mechanical Turk workers (MT) rated the photos taken by RG interface to be significantly better than those taken by SG interface (expert: $F = 7.62$, $p < .05$, $\eta^2_{partial} = .249$); MT: $F = 20.41$, p

$< .01$, $\eta^2_{partial} = .303$). Photo-taker is a significant factor for photo ratings (expert: $F = 5.88$, $p < .01$; MT: $F = 10.45$, $p < .01$). There is a moderate positive correlation between the average photo ratings from experts and MT raters (Pearson product-moment correlation coefficient $R = 0.68$, $p < .01$).

Raters	Real-time Guidance UI	Static Gridline UI
Expert photographers	M = 43.95 SD = 22.99	M = 40.60 SD = 22.08
Mechanical Turk workers	M = 65.98 SD = 21.87	M = 60.92 SD = 22.42

Table 1. The means and standard deviation for the photo aesthetic rating scores (range: 0-100)

Our data analysis show that users conform better to rule-of-thirds with the RG interface than with the SG interface:

- Users align the subject to the rule-of-thirds grid better with the RG interface than the SG interface (average diff = 31.76 (on a 0-250 scale), $p < .05$, one tailed paired t-test)

- The proportion of human subject's width is significantly closer to 1/3 when using the RG interface compared to the SG interface (average difference = 6%, $p < .01$, one tailed paired t-test). Users tend to have smaller subject/image ratios when using the SG interface.

There is no significant correlation between the alignment to the rule-of-thirds grid and the aesthetic quality score. There is a moderate negative correlation between the deviation from one-thirds proportion and the aesthetic quality score (Pearson product-moment correlation coefficient $R = 0.43$, $p < .01$)

To further understand the criteria that raters used, we coded user comments into different categories. 51.7% of the expert comments were about the global features of the photo, including composition (balance, camera angle and distance, rule-of-thirds etc.), lighting (exposure, evenness, shadows, etc.), and color (white balance, saturation, contrast, true to reality etc.) In contrast, only 29.8% of the Mechanical Turk raters' feedback fell into the above categories. 26% of experts' comments were about the face and body of the subject, including genuine/natural smile, mood, eye contact, glass reflection, lighting effect on the skin tone, aliveness, naturalness, movement and action, head/body proportion. 44.8% of the Mechanical Turk raters' feedback fell into those categories.

DISCUSSION

Our study demonstrates that the real-time guidance about rule-of-thirds generates significantly higher aesthetic quality portraiture photos than the static gridline interface. Moreover, the photos taken by the RG interface follow rule-of-thirds better. This shows that even when users know rules-of-thirds through instruction, the UI still makes a

difference in whether users follow the rule or not. They tend not to ignore feedback when it responds to their real time motions. These findings point to a fruitful new direction to explore. By leveraging the increased scene analysis capabilities and RG interface enabled by camera arrays and computational photography, future mobile platforms can guide casual users into taking photos that look like ones taken by experts.

While this study addresses the question of *whether* RG interfaces can enhance photo aesthetic quality, the *how* and *why* questions need to be answered in future work. Our study reveals complex and intertwined factors that influence photo aesthetic quality. We find subject/image proportions close to 1/3 correlate with better aesthetic scores. This is consistent with [9] (pp.3), which points out that new photographers tend to shoot too far away from the subject and include too many distractions. But we also find that aligning the subject of interest with rule-of-thirds gridlines does not correlate with better photo aesthetics. It can be caused by other aspects of photography that are not supported by our interface. For example, a user may solely focus on rule-of-thirds when using our tool but accidentally introduce new distractions in the background. In contrast, experienced photographers attend to multiple aspects of photo-shooting at the same time. Therefore, to analyze photo aesthetics, we need to analyze multiple aspects of the scene in real time, such as depth, lighting, segmentation, semantics, etc. Full potential of this research direction will be only be unlocked with sensors that enable a greater understanding of the captured scene, both in depth and semantics. As depth sensing mobile platforms are just now reaching the market, they open up new opportunities for computational photography to change the process and outcome of photo-taking.

We collected ratings from both experts and Mechanical Turk workers. Even though they focus on different aspects of photo aesthetics, both groups rated photos taken with our real-time guidance interface to be better. They are moderately correlated with roughly a 20-point difference (with experts being stricter). This shows that although photo aesthetics are subjective, there is agreement among different user groups to some extent.

Limitations
Rule-of-thirds is just one photo composition rule. Our implementation of rule-of-thirds is one simple interpretation of the rule. Incorporating multiple rules is future work. Also, our work does not encourage creative photographic techniques that break standard composition rules.

CONCLUSION
This paper contributes to the literature by developing and evaluating a real-time photo-taking guidance interface. We present empirical evidence that this interface can enhance photo aesthetics in a constrained user scenario. It suggests that a new generation of mobile devices with camera arrays may enable breakthroughs in mobile user experiences through advanced computational photography algorithms and real-time guidance feedback.

REFERENCES
1. Bradski, G. OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

2. Datta, R. and Wang, J. ACQUINE: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. *Proc. of Multimedia Information Retrieval*, ACM, (2010), 421–424.

3. Khan, S. and Vogel, Evaluating visual aesthetics in photographic portraiture. In *Proc. of Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, ACM (2012), 1-8.

4. Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. The Ubiquitous Camera: An In-Depth Study of Camera Phone Use. *IEEE Pervasive Computing*, *4*, 2 (2005), 42–50.

5. Li, C. and Gallagher, A. Aesthetic quality assessment of consumer photos with faces. *Proc. of Image Processing* IEEE (2010), 3–6.

6. Luo, Y., & Tang, X. Photo and video quality evaluation: F1ocusing on the subject. *In Proc. of Computer Vision– ECCV*, Springer (2008), 386–399.

7. Luo, W., Wang, X., & Tang, X. Content-based photo quality assessment. *International Conference on Computer Vision*, IEEE (2011), 2206-2213

8. Mason, W. and Suri, S. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods 44*, 1 (2012), 1–23.

9. Northrup, T. (2012). *How to create stunning digital photography*. Mason Press.

10. Pedro, J.S. and Church, K. The sound of light: induced synesthesia for augmenting the photography experience. *CHI'13 Extended Abstracts,* ACM (2013), 745–750.

11. Peterson, B. F. (2003). *Learning to see creatively*, Amphoto Press.

12. Smith, J. T. (1797). *Remarks on rural scenery*. Nathaniel Smith ancient Print. Cited by the Wikipedia page about rule of thirds: http://en.wikipedia.org/wiki/Rule_of_thirds (retrieved September 10, 2014)

13. Yao, L., Suryanarayan, P., and Qiao, M. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International journal of Computer Vision*, 96, 3, Springer (2012), 353-383.