Image Cropping: Collection and Analysis of Crowdsourced Data

Sally Ahn
Maneesh Agrawala, Ed.
Björn Hartmann, Ed.
Brian A. Barsky, Ed.

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-94
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-94.html

May 11, 2012

Image Cropping: Collection and Analysis of Crowdsourced Data

by Sally Ahn

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Maneesh Agrawala
Research Advisor

(Date)

Professor Bjoern Hartmann
Second Reader

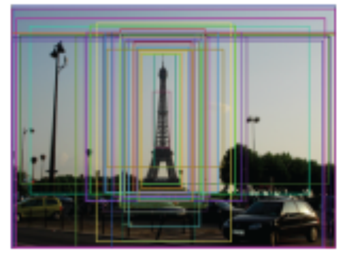
(Date)

Professor Brian Barsky
Third Reader

(Date)

Original photograph

Crowdsourcing



Cropping data

Abstract

There has been much interest in automated cropping and retargeting in the field of computer graphics and vision because composition plays a vital role in the visual appeal of photographs. Although image cropping requires minimal physical manipulation, the decisions users must make in order to complete the manipulation is complex. Understanding the pattern (if any exists) of these decisions is an important prerequisite to automated cropping that has been overlooked by current cropping algorithms. In this report, we take a step back from numerous works in automated cropping and retargeting to analyze *real* cropping behaviors of real *people*. We do this by employing crowdsourcing techniques to collect many crops for a set of 68 photographs and then analyze them with respect to three composition guidelines recommended in photography and art literature: 1) Rule of Thirds, 2) Filling the Frame, 3) and Leading Lines. We found that people most consistently followed the Rule of Thirds. While a positive correlation also existed for Filling the Frame, the findings were not conclusive. No correlation was found for between the first two guidelines and Leading Lines.

Table of Contents

1. Introduction

The composition of a photograph plays a significant role in determining its aesthetic quality. In *The Art of Photography, An Approach to Personal Expression*, Bruce Barnbaum declares “good composition” as the single element important for any artistic effort [1]. This claim is supported by a study in which Savakis et al. showed that composition was the most influential factor of image appeal in photography [2]. Photographers control the composition of their photographs through vantage point selection and framing decisions that involve what to include and what to exclude from the outermost boundaries of the final image [3]. Although the vantage point cannot be changed after the photograph has been taken, the framing can still be changed with post-shooting editing.

Since *cropping* involves removing margins of an image and thus enables the photographers to reexamine the framing decisions that may have been rushed or misguided at the time of shooting, we define a “good” crop as one that *improves* the composition of a photograph. Peter Ensenberger, former Director of Photography and author of *Focus on Composing Photos*, describes photographs with “good composition” as “well-balanced images” that achieve “visual harmony or dynamic tension.” [4]. This definition suggests that enhancing or adding visual harmony, balance, or dynamic tension can improve the composition of a photograph.

Hence, creating “good” crops requires reevaluating the factors that contribute to a good composition. Thus, despite the simplicity in the *physical* manipulation of the photograph’s frame, “good” crops require making complex *decisions*. For example, photographers must answer whether all the content in the frame contribute to the meaning they want to communicate through the image, whether any element distracts from the subject they want to communicate, and how to position and *organize* the elements to emphasize the subject and its supporting content [3]. Understanding the pattern (if any exists) of these decisions is an important prerequisite to automated cropping that has been overlooked by current cropping algorithms.

In this report, we aim to gain a better understanding of these multifaceted cropping decisions by analyzing crops made by different individuals for a set of photographs. Although previous works have tried to quantify the aesthetics of composition in existing photographs [5], [6], these methods only consider one "cropping" decision per photograph (the frame itself) and do not explore the variance of individual crops that are possible for each photograph. Rubinstein et al. conducted a comparative study of image retargeting algorithms, where participants were asked to compare two retargeted images and select the one they liked better [7]. In addition to retargeted images, they included results of manually cropped images to examine the tradeoff between deformation and content removal [7]. They found that while more recent algorithms received higher ratings (determined by the number of times the method was preferred over a different method), cropping was still among the most preferred methods [7]. The authors' conclusion that the high ratings of cropping indicate that people prefer loss of content to deformation artifacts motivates us to investigate people's cropping *behavior*.

Unlike previous works that analyze existing photographs, we collected many framing decision made by real people for each photograph. Our approach was to crowdsource cropping tasks to collect many independent crops for each photograph. We then analyzed this data to identify factors that may have been common to these individually made cropping decisions.

One challenging aspect of analyzing cropping and its effect on the photograph's aesthetic appeal is the subjectivity of what it means to be a "good" composition. In order to assess whether a crop has improved a composition, we must know how to evaluate good compositions. We consulted various books and online resources on photography and composition to address this. In the following section we describe some of our findings and how they inspired the metrics we used to analyze our cropping data.

2. Analyses of Photographic Images

Angela Faris-Belt, the author of *The Elements of Photography,* defines a "photographic image" as the product of three components within a frame: the *subject, form,* and *content* [3]. The "subject" refers to the "essence or meaning" of the image, and can be abstract ideas such as "hope," which may not be concrete objects visible in the image [3]. On the other hand, the "content" (also referred to as "subject matter") of the photograph refers to the visible elements contained in the photograph [3]. Finally, "form" refers to the *organization* of the elements, i.e. the decisions made by the artist to arrange the design elements (line, shape, value, texture, color, and more specific to photography, framing and lighting) of the image [3]. One way of describing these distinctions would be to say that the subject captures the *semantics*, or subjective message of the photograph; the content captures the physically visible *objects*; and the form captures the overall *composition* of the photograph. In this report, we focus on the components that can be analyzed objectively, namely, the *content* and *form*.

To represent the *content* of the photograph, we manually created *content masks* (Figure 1) that segments *focal point* objects in the photograph. A focal point is the "central point of interest" that will "draw the eye of viewers" and provides a "resting place" for the eye in the image [8]. The author of *Tony Northrup's DSLR Book: How to Create Stunning Digital Photography*, also notes that focal points are not obvious for "landscape, nature, and architectural photography," and we did indeed run into this difficulty of identifying a "point of interest" for some photographs where the main subject was ambiguous, such as landscape or candid photographs. For such photos, the content masks included the entire frame of the photograph.

Figure 1: Examples of our original photographs in our photo set (top) and their content masks (bottom). Image source: http://www.flower-pictures.org (middle); http://www.petsfoto.com (right)

To analyze the *form*, or composition, of the individual crops, we chose two popular guidelines for composition guidelines found in photography and art literature: Rule of Thirds and Filling the Frame. We chose these "rules" because they were among the most frequently mentioned guidelines for photograph compositions, but it is important to keep in mind that compositions that break these guidelines can also be aesthetically pleasing if done with purpose and skill [9]. For the purpose of analyzing our crop data, however, these guidelines provide a widely accepted basis for defining quantifiable measures for comparative analysis. We gathered these guidelines from various photography books and online resources [1], [3], [4], [8–12].

The Rule of Thirds is one of the most commonly taught and used composition techniques for photographers [11]. The principle behind the Rule of Thirds is related to dividing the frame of the photograph into harmonious proportions. The frame of a photograph can be implicitly divided by any object, even if an explicit line is not present [10]. The Rule of Thirds to simplifies the process of achieving harmonious proportions in composition. The main idea is to divide the frame of the photograph into thirds, both horizontally and vertically, which creates four *power points* where the lines intersect (Figure 2). The lines divide the frame into regions that closely approximate the Golden Section, which has been known to produce "harmonious" division since the Renaissance [10]. The photographer can achieve a pleasing composition by positioning an object of interest falls on one of the power points.

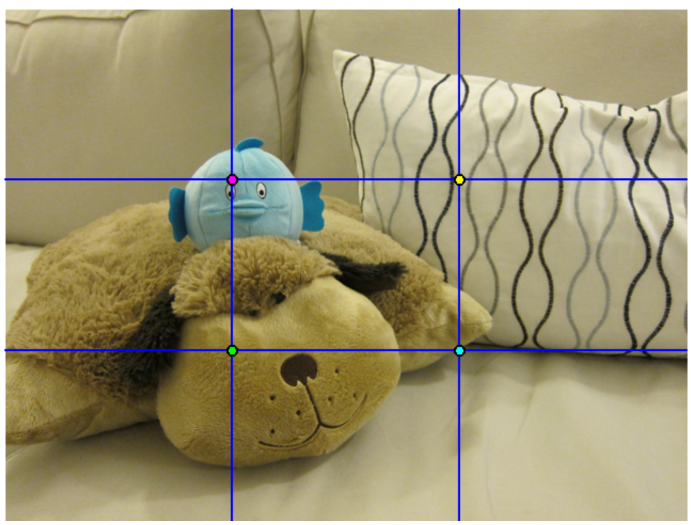


Figure 2: Power points are the four intersecting points of the Rule of Thirds grid. The colors indicate which of the four corners the power point belongs to (magenta = upper left, yellow = upper right, cyan = lower right, green = lower left).

Another advice we encountered in nearly every photography resource was to *fill the frame* with the content element that best conveys the subject of the photograph. Some sources refer to such frame-filling as "simplifying" [11]. By filling the frame with the most relevant object, the photographer removes distracting elements (Figure 3). However, distracting elements can be removed by other ways, such as by defocusing the space around the subject (also known as "negative space"). Skillful integration rather than removal of negative space can also improve the overall composition by defining and emphasizing the main subject [12].

Figure 3: The image on the right exemplifies the fill the frame guideline. Image source: http://www.photosshow.com

3. Data Collection

We gathered all of our data through the crowdsourcing platform Mechanical Turk, where we posted many independent *cropping tasks*. These tasks presented the worker with an embedded online cropping interface implemented with JCrop[1]. The cropping task asked workers to "crop the image...so that it looks the most visually appealing." Clicking and dragging on the photograph allowed users to freely position and transform a new

[1] JCrop: http://deepliquid.com/content/Jcrop.html

frame over the original photograph. Regions outside the new frame were darkened to provide better visualization of the potential crops before users submitted their final crop. The collection of 68 original photographs posted for these cropping tasks is a mix of photos taken from the Internet[2] and a personal photo library. Most photographs were 640 pixels by 480 pixels in size, with a few exceptions that were slightly smaller. All but three were of horizontal orientation because portrait orientations were more rare.

These online tasks were posted to Mechanical Turk in two batches, i.e. two sets of 30 and 38 photographs (the latter set was deployed several months after the completion of the first batch). At the end of both batches, our data consisted of 2308 crops with 18 to 59 crops for each photograph. The difference in the number of crops per photograph was due to randomizing the selection of the photograph to present to each user as well as pre-filtering unusable data that were submitted by some workers (e.g. crops with width or height of zero).

We note that although online platforms like Mechanical Turk provide a quick and relatively cheap way to collect large amounts of data, they also demand higher quality control than onsite experiments. Researchers from various fields have studied different techniques for creating incentives for workers to reduce lazy workers who "game" the system [13–16]. Specifically, Shaw et al. found that among several incentive influences, "Punishment Agreement," (workers were told that their payment would hinge upon the evaluation of other workers) was the most significant factor for improving the workers' performance [15]. Therefore, we modeled our cropping task after this design by telling them that their crop would be "evaluated by other workers who will compare your cropped photograph with the original photograph to choose which image looks better."

---

[2] Image sources for original photograph included various web photo sites, including: http://www.photosshow.com, http://www.flower-pictures.org, http://www.petsfoto.com, http://visionanimal.freetzi.com

For each crop that was submitted, we also asked five different individuals to choose between the cropped photograph and the original based on their aesthetic appeal. Figure 4 shows visualizations of all crops and crops that received majority "votes" (i.e., 3 or more of the voters chose the crop over the original photograph). In general, voters seemed to reject crops that formed close-fitting bounding boxes around the visually salient object. This discrepancy may have resulted from the opposing guideline for negative space mentioned in Section 2. For our analysis, we considered all crops because some photographs yielded very few voter-approved crops. Finally, we asked all participants (both croppers and voters) to answer a short background survey that asked for their age, gender, photography experience, and art experience. Figure 5 shows the interface for the cropping and voting tasks.

Figure 4: Visualizations of the popular regions determined by all submitted crops (left) and crops approved by at least 3 out of 5 voters (right). Most voters rejected crops that centered a close-fitting bounding box on the main element (the flower bush).

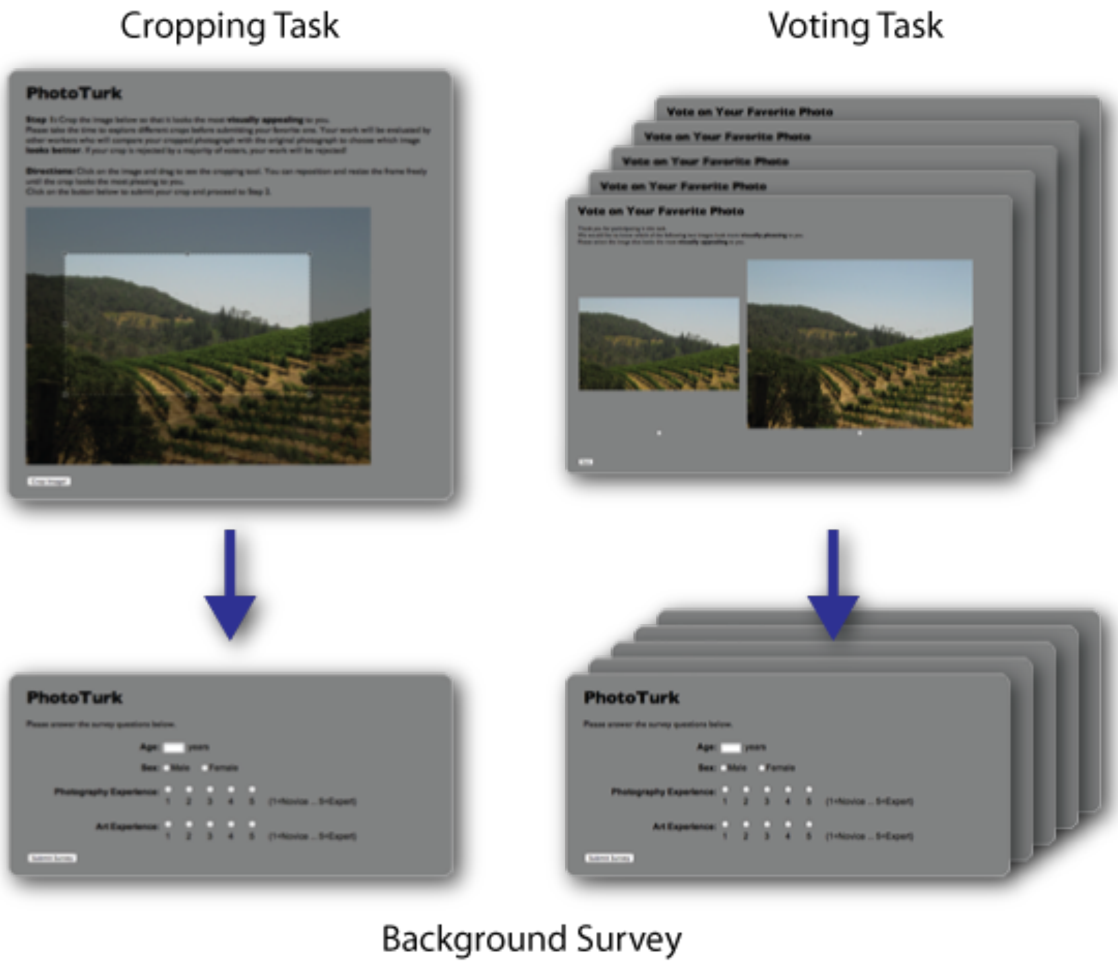


Figure 5: The web interface for the cropping task allowed worker to freely transform and reposition the cropping window before submitting their final choice. When a worker submitted a crop, five voting tasks were automatically posted for that crop. The order of the cropped and original photograph was randomized for each voter to remove bias. Both tasks also involved answering a short background survey that asked for the worker's age, sex, photography experience, and art experience.

Early in our analysis, we found that one worker had submitted an obviously bogus crop despite our experiment design (Figure 6). To minimize undue skewing of our data from such crops, we filtered out crops whose width and height were less than one tenth of the original photograph's width and height, respectively (68 by 48 pixels for most photographs). This restriction removed 29 crops from the original set. Our final dataset consisted of 2279 crops with 16 to 59 crops for each photograph.

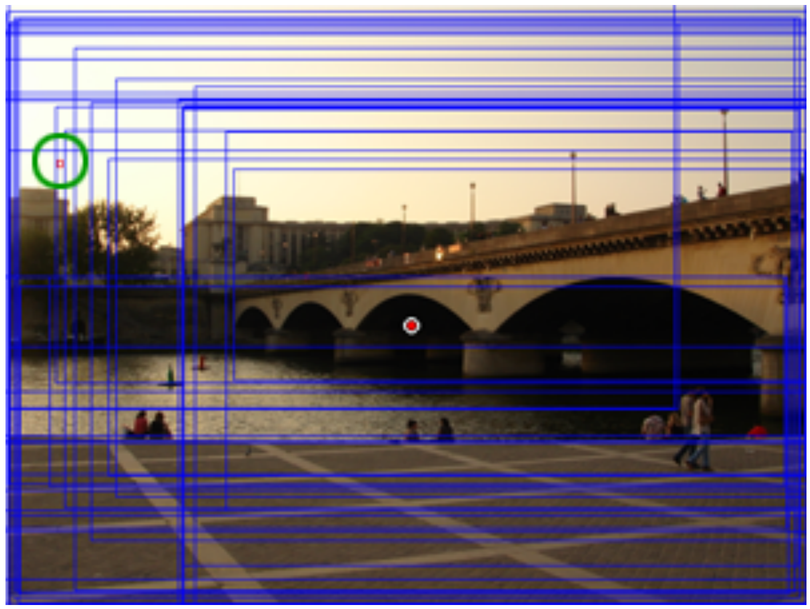


Figure 6: An outlier crop (red box circled in green) submitted by one worker. We removed such crops from our data set by restricting each dimension of all crops to be at least one tenth of the original photograph's.

4. Data Analysis Method

For our analysis, we defined two measures, M1 and M2, to evaluate each crop with respect to the composition guidelines described in Section 2. We explored the data through various visualizations and then tried to spot quantitative correlation between the crops and the composition guidelines by computing M1 and M2 for all crops. We describe these two measures below.

4.1 Rule of Thirds

A cropping decision complies with the Rule of Thirds if the locations of the power points coincide with the location of the subjects of the photograph. Therefore, we manually produced *content masks* for each photograph (Figure 1), as described in Section 2. We first checked whether the power points lied within or outside the subject mask.

Figure 7 (c) shows a visualization of this for one photograph. Crops that center the subject i.e., does *not* follow the Rule of Thirds, have higher number of power points within the subject mask. Thus, the number of power points that fall into the mask is not a good indication of a crop that follows the Rule of Thirds.

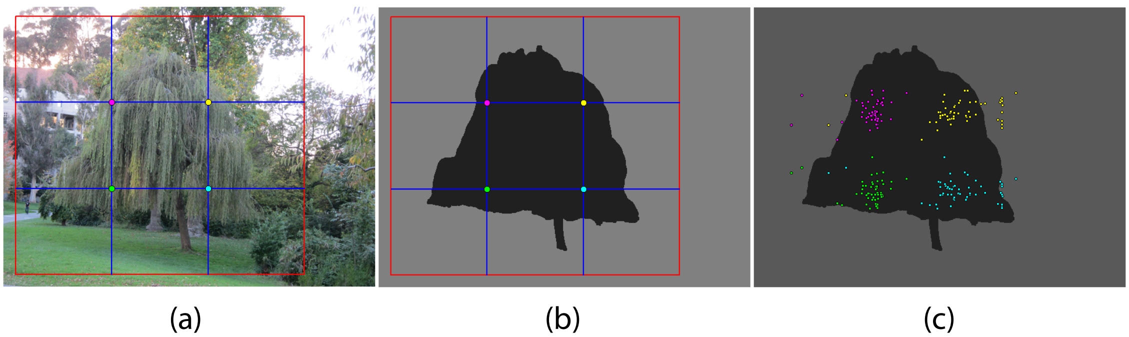


Figure 7: (a) Power points of the Rule of Thirds grid of ***one*** *crop (red border). (b) The power points for a single crop shown over the content mask. (c) Power points of all crops (shown without their borders and grid for visual clarity) for this photograph.*

To rectify this, we computed the centroids of each blob in the content masks and found the offset distance between the centroid (representing the location of the subject matter) and each of the four power points in every crop. For M1, we normalized this distance by the diagonal length of the original photograph used the minimum (i.e. the "best" score) of these four distances as the Rule of Third measure for the crop. Figure 8 visualizes the chosen power point and its M1 distance for all crops. For photographs with more than one blob in the content mask (i.e., multiple centroids, as in Figure 8), we compared all four power points with each centroid and chose the one with the minimum distance.

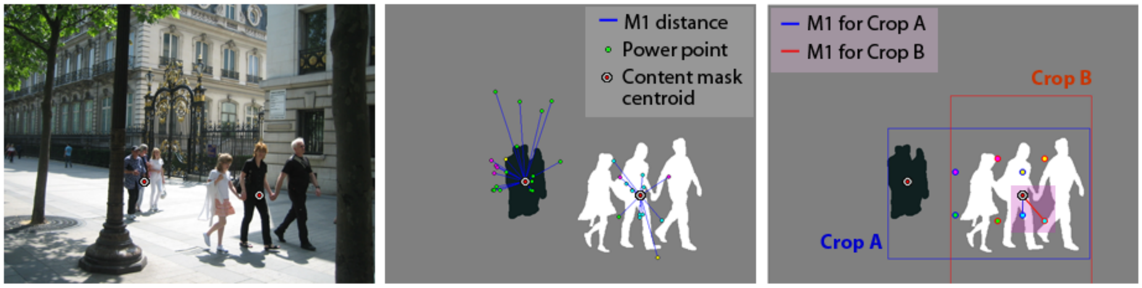
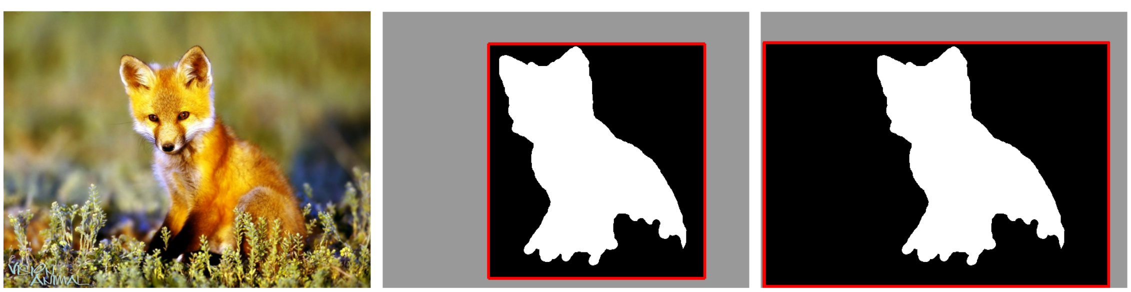


Figure 8: To calculate M1, we found the power point with the minimum distance to any of the content mask centroids. For photographs with multiple centroids, we chose the centroid with the minimum distance to any of the crop's centroids. The middle panel shows the M1 distances and power points found for all crops. The right panel shows that crops with shorter M1 distance more closely follows the Rule of Thirds (highlighted in pink).

4.2 Filling the Frame

To measure a crop's fulfillment of the "filling the frame" guideline, we defined M2 as the ratio of the area of the content mask to the area of the background. According to this guideline, crops with higher M2 values are deemed to be "better," since the subject occupies more of the frame (Figure 9). We did not distinguish the different regions for photographs that had multiple content mask centroids.



Original Photograph M2 = 0.4066 M2 = 0.2434

Figure 9: This figure compares the M2 values of two crops (red border) for the same image. The M2 value corresponds to the ratio between the subject (white) and the background (black). Image source: http://visionanimal.freetzi.com

4.3 Leading Lines

In photography literature, there is great emphasis on the importance of lines as a design element in composition (e.g. to direct the gaze of the viewer or to evoke different sentiments with different types of lines). However, while the Rule of Thirds and subject to background ratios are general and applicable to most photographs, many photographs do not have lines. Hence, we manually categorized our photo set into its respective line types plotted each photograph on a plane defined by their M1 and M2 measures to see whether a correlation existed between these three guidelines.

The categories we used to define line types were the following: no lines, diagonal lines, horizontal lines, and vertical lines. This categorization is based on the guidelines found in photography books: "Horizontal lines give a sense of quiet and peace. Vertical lines feel powerful, solid, and permanent. Diagonal lines are more dynamic, conveying movement and change" [11]. For most photographs, the presence or absence of lines was obvious. We identified ambiguous cases (e.g. conflict among lines belonging to multiple categories) with the "no lines" category. Figure 10 shows example photographs for each of these categories.

Figure 10: Examples of the line categories for photographs. Each column represents one category. Some photographs, like the one in the lower left corner, contained lines belonging to several categories (in this case, diagonal and horizontal), but no category dominated the other. Such photographs were labeled as having no lines.

5. Data Analysis Results

5.1 Rule of Thirds

Our initial exploration of the M1 data was to compare the values across the different photographs. Figure 11 shows a stem plot of the minimum distance found for each photograph (i.e. the M1 distance of the crop that scored "best" in this metric). The graph shows that the minimum M1 measures varied widely across different photographs. Figure 12 shows the M1 measures of *all* crops for each photograph as box-plot summaries. Values that were more than 1.5 times the interquartile range away from the 25th and 75th percentiles of the samples (top and bottom of the boxes) are shown as outliers. We can see that for some photographs (e.g., 28 and 45) the minimum M1 distance can be misleading due to a wide spread of crops for one particular photograph. Therefore, we used the median values and compared the M1 measure with that of the original photograph. Figure 13 shows the difference found by subtracting the crops' median M1 distance from the M1 distance of the photograph.

The majority of the bars are positive (blue), meaning that the M1 value was greater for the original photograph. Since M1 represents an undesired offset from the ideal case (M1 is 0 when the power point coincides exactly with the subject centroid), the blue bars indicate that the crops adhered more closely to the Rule of Third than the original photograph. For some photographs, however, this trend was significantly reversed (red). Upon inspection, we found that these photographs (1, 6, 7, 10, 16, 26, 48, 66, 68) already closely followed the Rule of Thirds (with respect to their content mask centroids). Figure 13 shows the corresponding photographs for the three most negative bars compared with one high-scoring photograph (67).

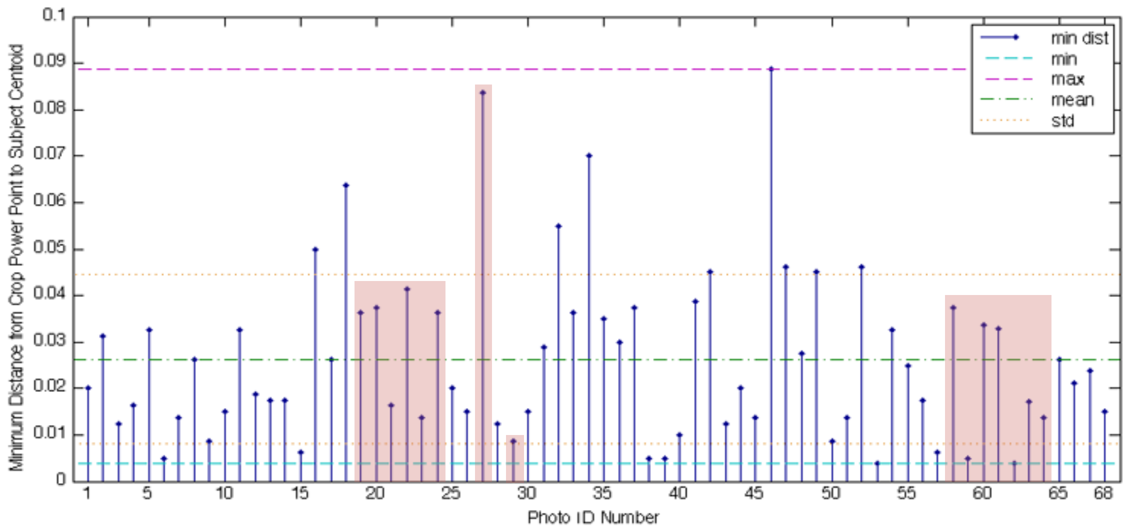


Figure 11: The minimum M1 found among all crops for each photograph. Photos 19-24, 27, 29, and 58-64 are highlighted red to indicate that their content masks encompassed the entire frame (with a single content mask centroid at the frame center). We wondered if this might affect the M1 measures but found no apparent correlation.

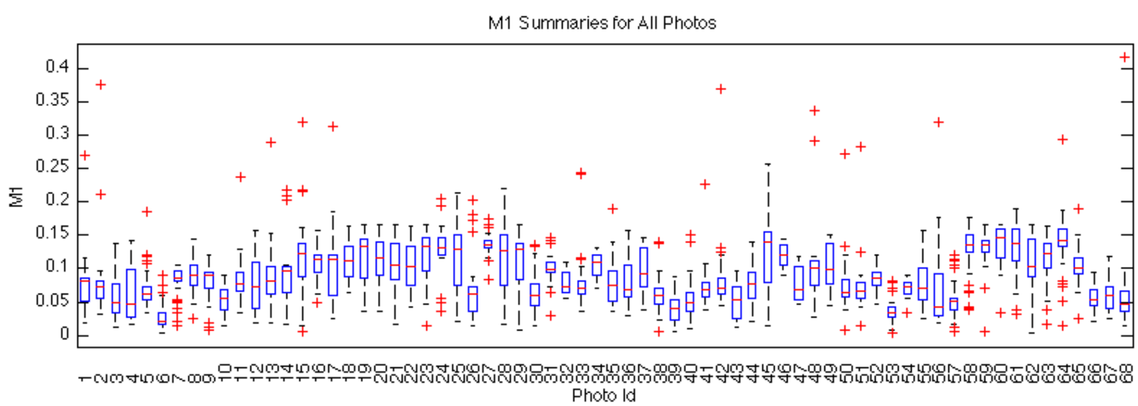


Figure 12: The box plot summaries of M1 for all crops in each photograph shows that variability of this measure differed considerably across photographs. Therefore, we chose the median values rather than the minimum for inter-photograph comparisons.

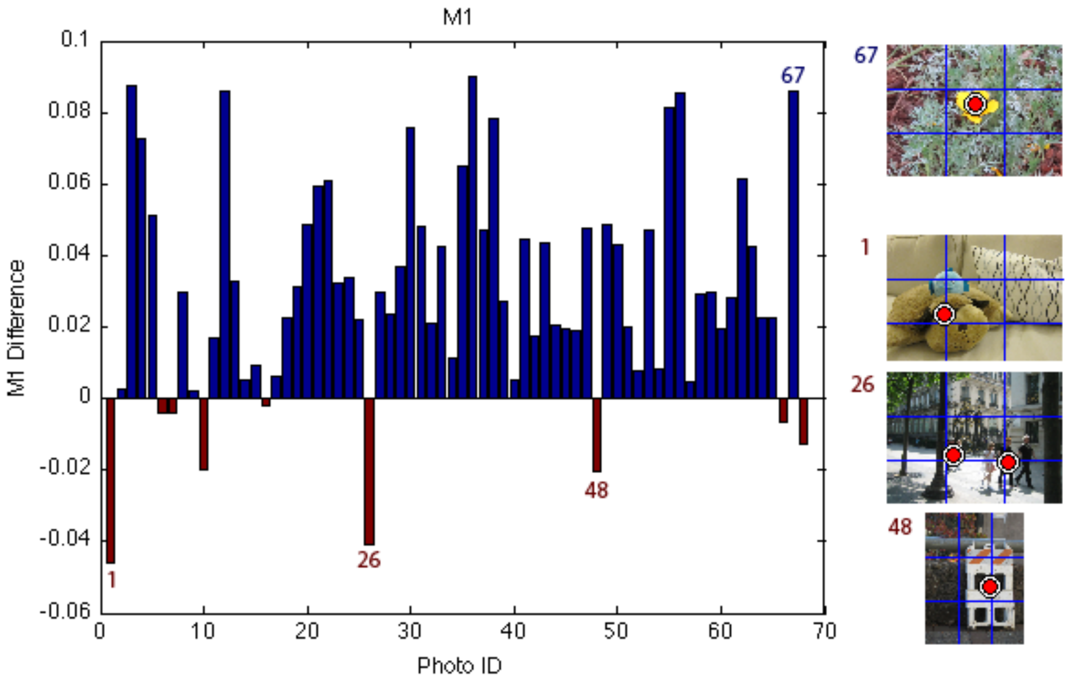


Figure 13: This bar graph shows the difference between the original photograph's M1 value and the median M1 value for all of its crops. In most cases, the original photograph's M1 values were higher than the crops' median (blue), suggesting that crops more closely followed the Rule of Third than the original photograph. For some photographs, that already followed the Rule of Thirds, this trend was significantly reversed (red).

5.2 Filling the Frame

Analyzing the subject to background ratios (M2) for all crops revealed that this measure yielded greater variability than M1. Figure 14 shows the box plot summaries for all crops. Note that photos 19-24, 27, 29, and 58-64 had no clear main subject and thus their content mask encompassed the entire frame. The data for these photographs were ignored since every crop for those photographs would produce equivalent M2 measures.

Compared with the subject-to-background ratios of the original photograph, the M2 crop measures were all higher (Figure 15). Although this finding is in agreement with the fill the frame guideline, the results are not conclusive because the act of cropping is inherently biased towards reducing the area of the background. In the aftermath, we realized that one way of minimizing this confounding factor would have been to present the cropper with a smaller predefined frame over the original photograph, which the cropper can then expand as well as reduce.

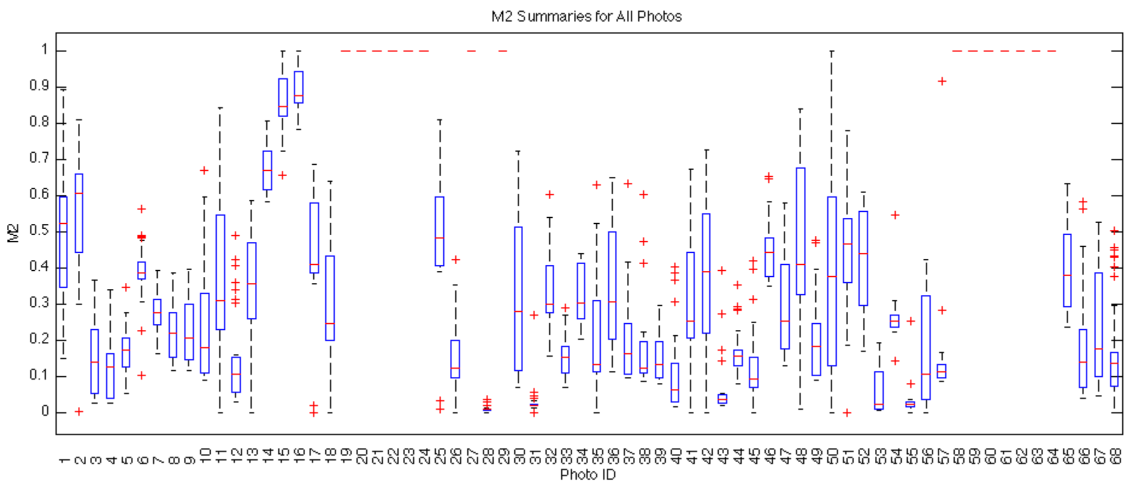


Figure 14: The box plot summaries of M2 for all crops in each photograph. Variability for this measure between photographs was higher than those for M1. Some photographs (19-24, 27, 29, and 58-64) do not show any distribution because their focal point was ambiguous (e.g. landscapes) and so their content mask consisted of the entire frame (M2 of 1).

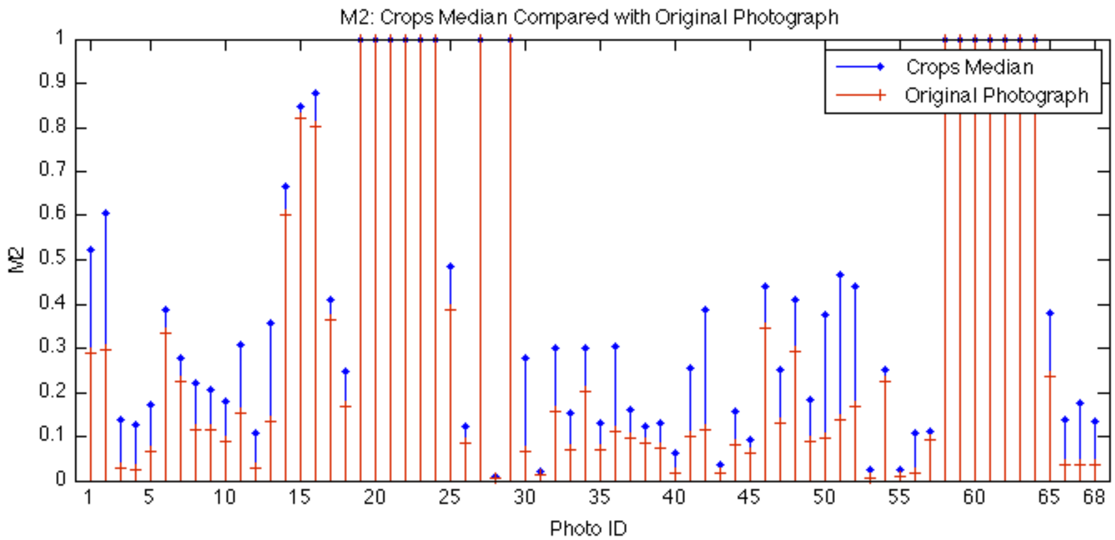


Figure 15: The median M2 measures for each photograph (blue) were consistently higher than those of the original photograph (red). Photographs 19-24, 27, 29, and 58-64 should be disregarded because they lacked clear focal points on which the content masks were based.

5.3 Leading Lines

To seek possible trends between M1, M2, and leading lines, we plotted each photograph according to its median M1 and M2 measures and labeled each with its line category (Figure 16). Since M1, M2, and leading lines are all guidelines for improving composition, we thought that plotting the photographs along the M1 and M2 dimensions might show clusters of similar line types. As Figure 16 shows, the plot showed no such correlation between line types, M1, and M2. However, few photographs fell into any of our line type categories, so further investigation is needed before we can make any definite conclusions.

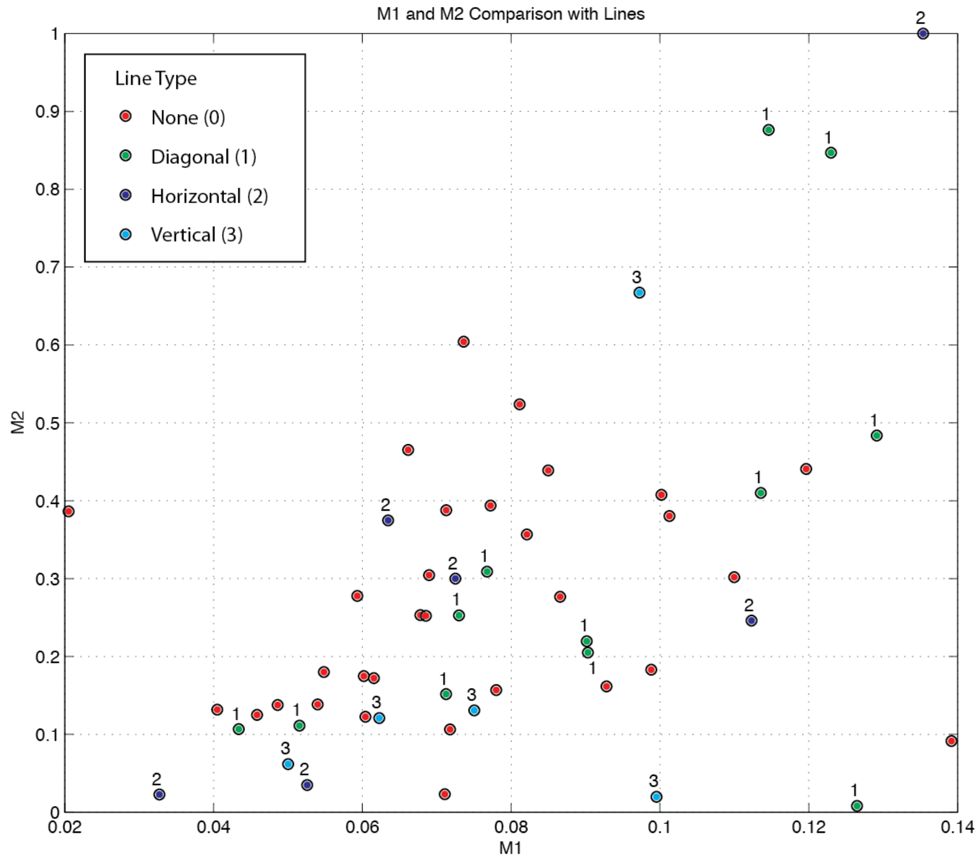


Figure 16: Each photograph is plotted according to its crops' median M1 and M2 measures. The numbers correspond the Line Type categories (only 1-3 are labeled for visual clarity). Photographs 19-24, 27, 29, and 58-64 were omitted because they had no content masks and cluttered the top of the plot.

6. Conclusion

In this report, we described an investigation into people's cropping behavior through collection and analysis of crowdsourced data. Our goal was to seek trends in the real decisions people made for cropping and verify whether they coincided with the widely accepted rules of thumb for cropping. We outlined some quality control techniques that are needed conducting such experiments on crowdsourcing platforms. We also provided background insight on three compositional guidelines that dominated the literature on photography: 1) The Rule of Thirds, 2) Filling the Frame, and 3) Leading Lines. We defined two quantities for measuring the crops with respect to the first and second guidelines (M1 and M2), and then explored these measures to see if any patterns suggested correlation with the third guideline, which was harder to quantify for all photographs.

We found that M1 produced the strongest positive connection with its guideline (the Rule of Thirds), although photographs that already followed this rule reversed this affect. There was greater variability in M2 measures. Although the majority of the crops for all photographs produced a favorable increase in the subject to background ratio, our experiments did not account for the bias for background removal inherent to the task of cropping. Future experiments that present pre-cropped photographs that can be expanded as well as reduced may produce more accurate data for evaluating M2. Finally, we did not see any correlation between M1, M2, and the types of lines (i.e. visual paths) present in the photographs. The results are not conclusive due to the small fraction of photographs that contained each type of line.

Our investigation into people's cropping patterns was motivated by recent publications on automated cropping and retargeting techniques [7], [17–21]. In the course of our research, however, we found concurrent experiments on cropping that were being conducted in the field of psychology. From their studies, McManus et al. concluded that cropping is an ideal paradigm for measuring people's aesthetic experiences because participants held different but consistent preferences for crop positions [22]. Such findings are encouraging and indicate that further investigations could yield more conclusive findings for people's cropping behavior.

References

[1] B. Barnbaum, *The art of photography : an approach to personal expression*. Santa Barbara, CA; Sebastopol, CA: Rocky Nook ; Distributed by O'Reilly Media, 2010.

[2] A. Savakis, S. Etz, and A. Loui, "Evaluation of image appeal in consumer photography," 2000.

[3] A. Faris-Belt, *The elements of photography : understanding and creating sophisticated images*. Amsterdam; Boston: Focal Press, 2008.

[4] P. Ensenberger, "Improving your photography: Composition," *Photo Composition Articles*. [Online]. Available: http://photoinf.com/General/Peter_Ensenberger/Improving_Your_Photography_Lesson_Two_Composition.htm. [Accessed: 03-May-2012].

[5] R. Datta, J. Li, and J. Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," U.S. Patent App. 12/116,578May-2008.

[6] D. Mao, R. Kakarala, D. Rajan, and S. L. Castleman, "Understanding Photographic Composition Through Data-Driven Approaches," *VISAPP*, vol. 18, no. 4, 2010.

[7] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in *ACM Transactions on Graphics (TOG)*, 2010, vol. 29, p. 160.

[8] D. Rowse, "Using Focal Points in Photography," *Digital Photography School*. [Online]. Available: http://digital-photography-school.com/using-focal-points-in-photography. [Accessed: 04-May-2012].

[9] D. Rowse, "Rule of Thirds," *Digital Photography School*. [Online]. Available: http://digital-photography-school.com/rule-of-thirds. [Accessed: 24-Apr-2012].

[10] M. Freeman, *The photographer's eye composition and design for better digital photos*, 1st ed. Focal Press, 2007.

[11] T. Northrup, *Tony Northrup's DSLR Book: How To Create Stunning Digital Photography*. CreateSpace, 2011.

[12] "Understanding and Using 'Negative Space' in Photography | Photography Mad."[Online]. Available: http://www.photographymad.com/pages/view/understanding-and-using-negative-space-in-photography. [Accessed: 05-May-2012].

[13] J. J. Horton, D. G. Rand, and R. J. Zeckhauser, "The online laboratory: Conducting experiments in a real labor market," *Experimental Economics*, vol. 14, no. 3, pp. 399–425, 2011.

[14] W. Mason and D. J. Watts, "Financial incentives and the performance of crowds," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 100–108, 2010.

[15] A. D. Shaw, J. J. Horton, and D. L. Chen, "Designing incentives for inexpert human raters," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011, pp. 275–284.

[16] P. G. Ipeirotis, "Analyzing the amazon mechanical turk marketplace," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 17, no. 2, pp. 16–21, 2010.

[17] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs, "Automatic thumbnail cropping and its effectiveness," in *Proceedings of the 16th annual ACM symposium on User interface software and technology*, 2003, pp. 95–104.

[18] J. Luo, "Subject content-based intelligent cropping of digital photos," in *Multimedia and Expo, 2007 IEEE International Conference on*, 2007, pp. 2218–2221.

[19] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 771–780.

[20] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Transactions on Graphics*, vol. 28, no. 3, p. 1, Jul. 2009.

[21] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *ACM Transactions on Graphics (TOG)*, 2007, vol. 26, p. 10.

[22] I. C. McManus, F. A. Zhou, S. l' Anson, L. Waterfield, K. Stöver, and R. Cook, "The psychometrics of photographic cropping: The influence of colour, meaning, and expertise," *Perception-London*, vol. 40, no. 3, p. 332, 2011.