Adaptive Photographic Composition Guidance

Jane L. E[1], Ohad Fried[1], Jingwan Lu[2], Jianming Zhang[2], Radomír Měch[2], Jose Echevarria[2], Pat Hanrahan[1], James A. Landay[1]
[1]Stanford University, Stanford, CA, USA
[2]Adobe Research, San Jose, CA, USA
{ejane, ohad, hanrahan, landay}@cs.stanford.edu, {jlu, jianmzha, rmech, echevarr}@adobe.com

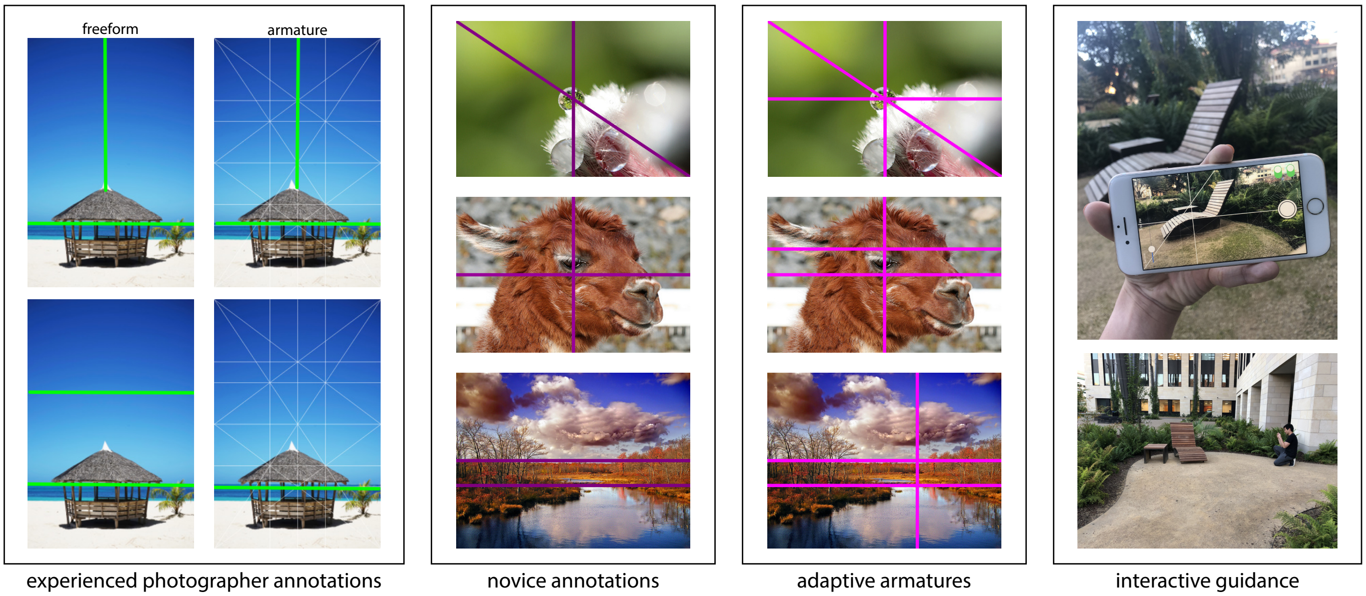

Figure 1. To design our interactive composition guidance interface, we were interested in better understanding people's ability to recognize composition and to annotate them on a composition grid. Left to right: We collected annotations from both *experienced* photographers as well as *novices* on Mechanical Turk. Inspired by these results, we developed an algorithm for heuristically computing these lines, or *adaptive armatures*. We display these adaptive armatures as an overlay in an in-camera composition *guidance* tool and study how it impacts how people take photos. Photos by Sharon Mollerus, yugoQ, and Nicholas_T. Creative Commons Attribution License.

ABSTRACT

Photographic composition is often taught as alignment with composition grids—most commonly, the rule of thirds. Professional photographers use more complex grids, like the harmonic armature, to achieve more diverse dynamic compositions. We are interested in understanding whether these complex grids are helpful to amateurs.

In a formative study, we found that overlaying the harmonic armature in the camera can help less experienced photographers discover and achieve different compositions, but it can also be overwhelming due to the large number of lines. Photographers actually use subsets of lines from the armature to explain different aspects of composition. However, this occurs mainly offline to analyze existing images. We propose bringing this mental model into the camera—by adaptively highlighting relevant lines to the current scene and point of view. We describe a saliency-based algorithm for selecting these lines and present an evaluation of the system that shows that photographers found the proposed adaptive armatures helpful for capturing more well-composed images.

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00
DOI: https://doi.org/10.1145/3313831.3376635

Author Keywords

photography; camera interfaces; composition; dynamic symmetry

CCS Concepts

•Human-centered computing → Graphical user interfaces; •Computing methodologies → Graphics systems and interfaces; *Computational photography*

INTRODUCTION

Cameras are becoming smarter, but currently provide limited aid in helping with creative decisions. However, as they become more pervasive, people are increasingly interested in

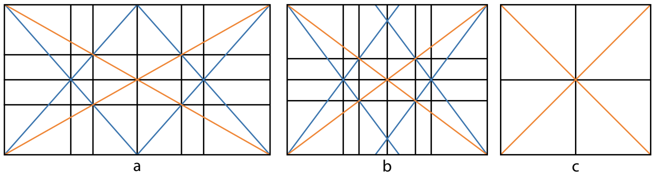

Figure 2. The armatures are designed around main diagonals (orange) and shorter, reciprocal diagonals (blue). Horizontal and vertical lines extend from intersection points of these diagonals. Our Mechanical Turk task presents three forms of armatures for different aspect ratios: (a) *wider/taller than root 2*: based on proportional design—reciprocal diagonals always connect to center resulting in horizontal/vertical lines at thirds and quarters, (b) *between root 2 and square*: equivalent to (a) at a specific aspect ratio, the reciprocal diagonals are perpendicular to the longer (main) diagonals, (c) *square*: (b) collapses to this set of lines.

learning to be better content creators. While recent developments in graphics and vision have been helpful in producing technically and aesthetically improved images, much of this work has focused primarily on automatic workflows, where the user has limited input in and understanding of the creative choices made.

An important artistic aspect of photography is composition—the relationships between the different elements in an image and their properties (i.e., size, position, color, texture, etc.). Composition is used to direct the viewer's gaze through a photo by establishing visual hierarchies and rhythm [4, 19, 20, 47]. In this paper, we focus on the spatial relationships of elements within the frame. These spatial relationships have been studied over the centuries in fields like painting, photography, cinema and graphic design. They are often explained through the use of grids of lines, with popular examples like the rule of thirds, or the golden ratio. These grids rely on the assumption that aligning important visual elements along their lines and intersections will tend to generate images that are more visually appealing to the human eye [18, 23, 47].

A common grid used by photographers like Henri Bresson-Cartier and Annie Leibovitz [19, 23, 28] is the harmonic armature (Figure 2), which dates back to Pythagoras [18, 23]. While some experienced photographers do not need to see and strictly follow such grids, others directly place overlays on their camera view for guidance [12, 21]. Many cameras provide the option of directly overlaying a variety of different grids on the viewport. In any case, not all the lines in a grid are meaningful for a given scene, so it is up to the photographer to determine which ones to use and how. While at first glance human observers seem to intuitively perceive relationships between the elements in an image and an overlaid grid, it actually takes practice to find and focus on the more relevant ones. Books, blog posts, etc. often illustrate composition with composition grids overlaid and individual lines highlighted to emphasize certain aspects of the artist's composition choices [43]. However, this is an offline process—photographers manually annotate lines to better explain existing images.

Inspired by this behavior of manually highlighting lines in existing images, we propose a novel in-camera guidance that aims to automatically highlight the relevant lines to the composition of the current camera image (see Figure 1).

In this paper, we present:

- an **algorithm** for automatically identifying adaptive armatures based on image saliency,
- an **interactive in-camera app** that shows these adaptive armatures to users as composition guidance, and
- a **user evaluation** of this app that shows that participants do in fact prefer the composition of photos they took using the adaptive guidance versus using static composition guidance.

We additionally contribute:

- a **dataset of crowdsourced composition annotations** that capture the common knowledge of what people believe are the most relevant lines to an image's composition within the constraints of the harmonic armature grid,
- a **user evaluation** comparing the use of static composition guidance versus no guidance that shows participants are more confident in their compositional ability and feel more creative using the guidance.

RELATED WORK

Composition Analysis. Previous work has focused on low-level image components that affect photographic composition. Zhou et al. [63] estimate vanishing points and use them to retrieve similar images. Lee et al. [33] detect dominant geometric elements and use convolutional neural networks to classify photographic composition rules in outdoor scenes.

He et al. [24] find and highlight triangles on images to create compositional awareness and promote creativity. Other works classify an image to retrieve similar high quality examples for inspiration [35, 58, 62]. We focus on the placement of salient features with respect to composition grids, rather than specific image components or types of composition [60].

Composition Enhancement. Although several image recomposition methods have been proposed [27], cropping is still a very straightforward way to improve composition. Some have focused on explicit attention and aesthetics models [8, 50], but the most popular approach is to learn it directly from examples from professional photographers or user annotated samples [3, 9, 10, 22, 34, 37, 52, 57]. This approach tends to work well for automated workflows, although performance may vary depending on the dataset used for benchmarks [6]. Fang and Zhang [14] trained a network to find good crops in 360° images. Zeng et al. [59] proposed grid anchors to reduce the number of crop candidates to evaluate. Example-based cropping has been proposed for general images [7, 38], and portraits [61]. Rather than automatically finding good crops, we instead try to help users achieve good compositions actively by changing their point of view during capture.

Capture-time Guidance. Modern cameras come with some visual aids to help the photographer make technical and creative decisions during capture. Some typical examples are zebra patterns for over/underexposure, histograms, focus peaking, levels or static composition grids. Mitarai et al. [45] detect relevant elements in the current frame to classify the composition and show overlays providing specific guidance on how to refine them, limiting the user's creative freedom. Lo et al. [39] compute an aesthetics score straight from the camera feed in real-time; Lujun et al. [40] perform image assessment and cropping suggestions; Ma et al. [42] use a view proposal network to suggest different crops, plus an interface that learns user preferences for future proposals. We are interested in composition guidance that creates awareness about different creative options, while providing a mental model for the user to reason about them. These interfaces provide limited feedback for building that mental model.

Other related work on capture-time guidance have focused on helping to re-photograph previous photos [2], or reference photos taken while traveling [17, 30]. Rawat and Kankanhalli [49] proposed a point of view recommendation system based on previous photos taken in popular locations. Other work recommends where to stand when posing for souvenir photos [41, 48, 51, 55]. More creative guidance has focused on posing for better selfies [15, 25], better portrait lighting [13, 36], or more general pose-based portrait recommendations [16]. Extending into video, the work of Kumano et al. [32] helps teach videography by analyzing camera work styles.

Closely related to our work, Xu et al. [56] devised a 3-camera system that provides real-time instructional feedback to users, so they can learn to compose better portraits using the rule of thirds. For that, they compute a measure of the alignment between the regions of interest and the rule-of-thirds grid, and show arrows on the viewfinder for the users to follow to improve the alignment. Apart from focusing on composition beyond portraits, our guidance also uses composition grids, but aims for a less constrained user experience: instead of using arrows to make the user follow the criteria of the algorithm, our systems highlights potentially interesting compositions using adaptive armatures. Our user experience is then closer to the *smart guides* implemented in many design tools [44]: when salient elements show interesting spatial relationships between them, such relationships are highlighted for the user, so she can decide whether to follow and refine them or not.

IN-CAMERA GUIDANCE

With the goal of enabling non-expert photographers to leverage composition grids in their photography, we aim to understand how to effectively provide early compositional feedback in-camera. We asked: Are people interested in in-camera guidance? If so, how should it work? More specifically, is composition a good concept to target for in-camera guidance?

Photography Practice Survey

Our first question involved understanding if people wanted in-camera guidance at all, and if so, what type of feedback would be helpful or not to their current photography practice.

To answer this question, we designed a survey asking about people's existing photography practice and past learning experiences. We surveyed adults with an interest in photography and received a total of 127 responses from participants (74 male, 51 female, 1 non-binary, 1 preferred not to say), 19 to 63 years old with a range of photography experience.

Photography Practice Survey Results

Many of our participants had had some sort of photography learning experience (only 23 had never used any resources such as classes, videos, books, etc. to learn photography)—several of the ones who had taken in-person classes (22 of 45) mentioned the effectiveness of the "direct, immediate feedback in the moment" (P7), that those provided. Specifically, many mention the benefits of the feedback being "individualized" (P111) and "hands-on" (P73) for making immediate adjustments and correcting mistakes. Without this guidance, it can be difficult to "transfer knowledge to other conditions" (P10). Thus we saw benefit in pursuing the direction of trying to provide contextually-based in-camera guidance, so users can make better creative decisions during capture.

When asked why they would or would not use capture-time guidance in their camera, many mentioned concern with an app being too "distracting" (P51), "disruptive" (P53), or "intrusive" (P93). This was due to either concern about missing a moment (P12), intruding on others' time when in a group (P14), or feeling less artistic freedom (P86). Guidance should thus prioritize being minimally distracting and more about providing suggestions than insisting on specific artistic choices.

Our survey respondants supported the idea of focusing on composition feedback. We coded their responses to an open-ended question asking what they want to improve in their photography for mentions of popular photography concepts. Many explicitly mentioned composition (18); the other most popular concepts included camera settings (24) and lighting (22). Additionally, some responses mention higher level goals such as being more creative or having more "professional"-looking photos (11), which may also implicitly include composition. Composition is important to consider in the camera because the range of possible changes is limited once the photographer has left the scene. On a Likert scale from 1 (strongly disagree) to 5 (strongly agree), participants rated their editing practices as significantly more around small framing adjustments (Mdn = 4, IQR = 3-5) than substantial cropping to change the image's composition (Mdn = 3, IQR = 2-4) [Wilcoxon signed-rank test [54] $V = 2217$, $p < .001$], which further supports the idea of providing composition guidance at capture time to reduce the need for more drastic changes while editing later.

Overall, people have a preference for getting feedback on the framing of their current photo (Mdn = 4, IQR = 3-4) rather than thinking of possible photos to take (Mdn = 3, IQR = 2-4) [Wilcoxon signed-rank test $V = 2183$, $p < .001$]. Thus, we should additionally focus our guidance on helping people refine their current image's composition. Of our survey respondents, 89% are willing to spend up to 5 seconds on capture-time guidance to get a high-quality result.

Experienced Photographer Interviews

To go into more depth on what types of guidance could be helpful in photography practices, we interviewed nine experienced photographers about their current photography practices and tools. All had formal training in photography, and 5 additionally had teaching experience.

Interviews were structured around the following questions:

- Describe your typical photography process(es).
- What photography tools do you use and what guidance does it provide?

- Would composition guidance be helpful for you?

Experienced Photographer Interview Results

Six of the interviewees expressed that they already consistently use overlays of sorts such as focus dots, light meters, levels, or composition grids (primarily the rule of thirds). This suggests that photographers are okay with some amount of their camera view being obstructed when that information is useful to their overall ability to take better photographs.

When asked what guidance might be helpful for them, some (5) expressed interest themselves in something that might give them new perspectives, like an "experienced photographer on my shoulder saying try this, try this" (P7), or even providing random composition guidelines to swipe through and try (P0). In general, these experienced photographers were very open to having any feedback that might help them try out different ideas in taking a photo in order to have more choices to pick from when going back to edit in the future.

Some suggested they are able to use guidance as needed while maintaining creative freedom to break the rules and follow their intuition. For example, explicitly disregarding the composition grid: "might have a rule-of-thirds overlay, but don't follow it super closely. I gravitate towards the bottom two eyes in the rule-of-thirds..." (P1). However, others mentioned conforming to a certain style, whether it be due to the prominence of the rule of thirds overlays, realizing an unexpected theme across many photos, or just having a "tendency to view the world in a specific way, but someone else might be different, always open to try a different type of shot or idea" (P6).

In describing their own photography processes, these photographers recounted both a process of searching for good compositions, as well as pre-composing and waiting for a shot—the latter was mentioned as particularly important in street photography. One described this waiting as a "necessary tension" in street photography (P5).

From these interviews, we learned that even experienced photographers could benefit from feedback that encourages them to try new ideas. In certain scenarios, they need to capture a shot immediately and thus rely on their instincts to quickly frame the photos, while in other scenarios they are willing to spend more time and devote both screen real-estate and their attention to a tool that helps them achieve a higher quality image. We aim to support both of these scenarios, for novices and experts alike.

In-Camera Guidance Design Goals

According to our survey and interviews, we came up with three design goals for our guidance. Guidance should be:

- **Context-aware.** Visual guidance should adapt to the current image and appear overlaid on the viewfinder.
- **Minimally intrusive.** It should be easily disregarded to allow photographers to pursue other creative choices.
- **Support exploration and refinement.** It should help photographers discover new ideas as well as execute existing intentions.

COMPOSITION GUIDANCE

Our next question was how to design in-camera composition guidance. Many cameras provide options for a number of composition grid overlays, but are these effective for achieving well-composed images? Do they follow our in-camera guidance design principles?

As we learned in our interviews, even experienced photographers tend to restrict themselves to options proposed in front of them—e.g. following the rule of thirds guidelines if those are visible on the camera screen. Thus, we were interested in using a more versatile grid, the harmonic armature, that allows for more compositional diversity. The harmonic armature provides different sets of orthogonal and diagonal lines. Diagonal lines are usually employed for dynamic symmetry [23], while from their intersections, the rule of thirds and golden ratio grids emerge.

As noted in the introduction, photographers currently manually highlight subsets of lines on existing images to describe how the lines are relevant to the image composition; and they instinctively do the same during capture, as surfaced in our interviews. We wanted to be able to provide similar feedback in real-time, in-camera. However, we first wanted to understand whether through training, experienced photographers develop a consistent view of which lines are relevant, and whether novice users are similarly able to produce such annotations.

Experienced Photographer Annotations

To understand whether experienced photographers had a consistent view of what lines described a composition, we ran a study where we had participants annotate photos.

We also performed an annotation study with 8 of our interview participants. The annotation task was broken into two parts. First, participants were asked "For each image, please freeform draw a set of lines that describe the composition of the physical elements of the image. Explain why you drew those lines." Next, the same images are displayed one at a time with the harmonic armature overlaid (see Figure 2b). Participants were asked to perform the same task, but to only select lines from the armature. For this task, we chose 5 images of varying complexity in composition for all participants to annotate. We additionally asked them to bring in 5-6 of their own photos to annotate. We asked them to choose photos that they believed had particularly nice/interesting composition.

Experienced Photographer Annotation Results

In watching these experienced photographers annotate images, we learned that experienced photographers are able to recognize lines in image compositions both freeform as well as using a composition grid. Given the constraints of a composition grid, they were also generally comfortable with selecting lines that might not be perfectly aligned with features in the image to approximately represent these features, as shown in Figure 3a. For images with prominent distinct elements, participants did frequently draw lines through or bordering these elements to highlight alignment of elements.

However, while we saw some consistent line annotations, we also saw many unique annotations. This is likely partially due

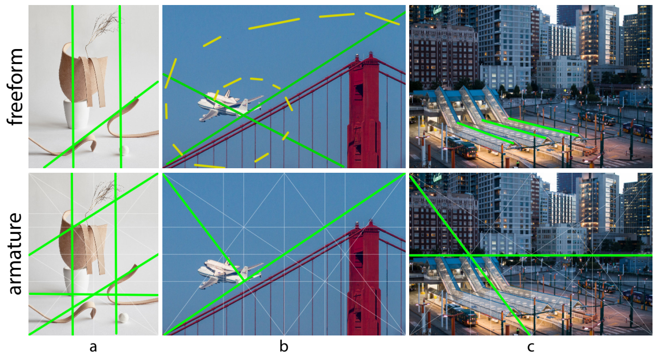

Figure 3. Comparing experienced photographers' freeform annotations with armature-based annotations: (a) participant selects nearest lines to freeform lines that does not perfectly exist in the armature, (b) participant draws golden spiral to represent the location of a salient object in the freeform annotation, finds nearest lines to other freeform lines in the armature that also highlight the object location, and (c) participant selects new alignments that were surfaced by overlaying the armature.

to the small number of participants and images—to respect our participants' time, we tried to keep both the interview (see Experienced Photographer Interviews Section) and annotations to within one hour. The additional inconsistency is due to the differences in our participants' artistic styles and their views on photographic composition. Lines are important for a variety of concepts related to composition, and experienced photographers already have many of these engrained in their mind. Despite the instruction to focus on composition of physical elements in the image, in addition to these alignments, participants annotated leading lines, strong hard edges, soft "edges" between regions (e.g. of different color), etc.

Nonetheless, many of these participants found the annotations to be a useful and fun exercise. In fact, while doing these studies, many participants indicated that they enjoyed the activity of identifying the lines within the fixed composition grid. Compared to drawing freeform, they noticed more unexpected alignments (both at intersection points and along edges) that they otherwise would not have seen. Some even mentioned possibly adjusting their images to align with lines that they highlighted rather than their current composition (we later saw this behavior in our user studies, see the Evaluation and Discussion Section). While these adjustments were not always deemed improvements, this suggests that even for experts who already have training in envisioning composition grids in their mind, searching for these alignments can change their perspectives in ways that might impact their photography.

Mechanical Turk Annotation

Given that experienced photographers could identify relevant line alignments from a composition grid to a given image, we wondered if novices could also use the same mental model and leverage a composition grid to describe the composition of an image. In addition, with more annotations, we wondered if we would be able to find some consistency in their annotations to suggest a set of the "most relevant" lines to an image's composition. To collect the data to answer this question, we crowdsourced annotations on Mechanical Turk.

Mechanical Turk Task

We constructed a dataset of 500 photos for our Mechanical Turk studies. The MIRFLICKR-25000 dataset includes 25000 images from Flickr that are open under the Creative Commons license and selected based on their interestingness rating [26]. From these, we selected the 500 most aesthetically pleasing images according to Kong et al.'s photo aesthetics ranking network [31]. We removed those with extreme aspect ratios ($< .5, > 2$) or with inappropriate content, and replaced them with images from the next 100 most aesthetic images.

We created a Mechanical Turk task in which users were shown 10 of these images at random and instructed to "Please select 1-4 lines that best describe the composition of the physical elements in the photo." The interface was an image with the composition grid overlaid and users used their mouse to hover over and click to select/deselect a line.

The first time a worker does a task, they are presented with a short description of the task and of composition. To avoid the task feeling too subjective and risky, we provide three example image annotations along with short explanations. Finally workers are given a short interactive tutorial on how to use the annotation interface. To complete the tutorial, users are required to select lines that match the example selection.

After the trial task, the first two annotations are easier tasks with simple compositions as further training. For these images, they are given the choice to see sample annotations and explanations. We used these tasks as validation to remove careless annotators. Per annotation task, workers would annotate 12 images (2 validation, 10 for dataset). We estimated this to take around 4 minutes, and paid $1 per task to match the minimum wage of $15/hr and used Fair Work to allow workers to inform us if the task took longer than anticipated [53].

Mechanical Turk Annotation Results

We collected a total of 1004 annotation task responses across 582 workers, totaling 11961 non-validation image annotations (an average of 22 annotations per image). From these annotations, we filtered out task responses where the worker was unable to get both validation tasks correct—we counted a task as "correct" if they selected one of the lines provided in the sample annotation. This resulted in the removal of 218 tasks, for a resulting set of 9133 image annotations (~17 annotations per image).

From this dataset, we aimed to determine a set of "most relevant" composition lines for each image as our ground truth annotation. Images vary in how well they can be represented by a composition grid, therefore also in consistency in annotations. To select our ground truth lines, for a given image, we defined the score of each line as the percentage of worker annotations that included that line. After an initial batch of annotations, we empirically determined a score threshold of 0.4 for whether a line should be selected as one of the ground truth lines—this resulted in an average number of lines selected as ground truth being close to the average number of lines selected per image in the annotations. Given this threshold, 10 images had no ground truth lines. These were also removed from the dataset, for a final set of 8966 image annotations

(~17 annotations per image). For this final dataset, the average number of ground truth lines per image is 1.82 and the average number of lines annotated per image is 2.17.

With a relatively high threshold of 0.4, we see that these average numbers of lines in the ground truth annotations are quite close to those in the workers' initial annotations, meaning most lines are reaching this score threshold. Additionally, only 10 images resulted in no lines above this threshold. This suggests that there is consistency in the notion of what are the "most relevant" lines in a composition grid of a given image and that novices can perceive and annotate this relationship.

Our Mechanical Turk task also included an optional text box for feedback on the task. We received a lot of positive feedback on the task. Many workers mentioned they found the task fun (30), enjoyable (19), and/or interesting (29), and would like to keep doing more of these tasks (9). In particular, one mentioned it being helpful for their learning: "I am working on my photography skills and this exercise was helpful for me to better understand this concept even if I wasn't entirely helpful to you."

The code for this annotation task and anonymized Mechanical Turk annotations along with analysis code (validation, filtering, etc.) are provided in supplemental materials.

Annotation Insights

From our experienced and novice composition grid-based annotations, we learned:

- **Performing annotations can be useful.** Experienced photographers described noticing new alignments. Novice photographers found them helpful for understanding composition. Both groups found the task enjoyable.
- **Annotations are approachable.** Both experienced and novice photographers were able to perform the annotation task with reasonable consistency. Thus, we chose to pursue this path for our in-camera composition guidance.

ADAPTIVE ARMATURES

Since our formative studies suggested that there exists a notion of prominent lines for a given image, we set out to automatically detect a set of relevant lines to provide these annotations in camera. We call this method *adaptive armatures*.

Heuristic Algorithm

Composition can be posed as aligning visually important elements in an image with the lines/intersection points of a composition grid [56, 60]. We use saliency to represent this visual importance. Thus, the adaptive armature should capture the set of lines that best align with the saliency map. Given an image, we score the candidate lines for the image armature such that the higher the score of a line, the better the elements of visual importance in the image are aligned to that line, and the more "relevant" they are to the image's composition.

For a given image, we compute an attention-based saliency map using Apple's built-in beta Vision library [1] (but any performant saliency estimation method could work). This map is used to vote for lines in the full armature grid. For each point

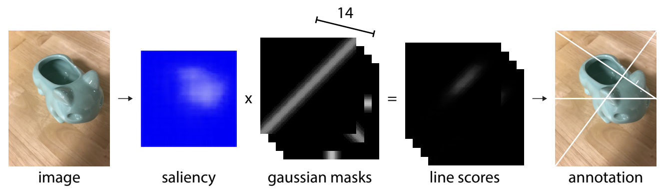

Figure 4. Computing the line scores to determine the *adaptive armature*. We compute the saliency map of the image and multiply that by a Gaussian mask per line. The sum of the resulting image is the line's score. The top 3 scoring lines become the adaptive armature.

p in the saliency map with saliency value **S**, its contribution to the score of a composition line **L** is:

$$score = \mathbf{S} \cdot Gaussian(distance(\mathbf{p}, \mathbf{L})). \qquad (1)$$

for a Gaussian kernel with a size of 1/10 of the image's longer dimension and a sigma of 1/4 of the kernel size. The score per line is then normalized by the length of the line. We select the top 3 scoring lines as our resulting heuristic annotation (see Figure 4), or *adaptive armature*.

Mobile Implementation

Our app runs on an iPhone 7 running the iOS 13.0 public beta. We built a basic camera app where saliency and the adaptive annotations are constantly computed in the background for the current camera image. The app has a camera shutter button and flashes the screen white when a photo is taken. For the purposes of our studies it has no other camera functionality to encourage participants to focus solely on composition.

To reduce computation time, we precompute the Gaussian maps per line and store them as images. The score per line can then be computed by simply multiplying the saliency image by the corresponding precomputed Gaussian map and computing the sum [56]. To maintain an interactive experience, we compute line scores at discrete 0.1 second intervals. Since users need time to process the highlighted annotations, this interval strikes a balance between perceptually-constant visual update and non-distracting digestible user experience.

Mechanical Turk versus Heuristic

We wanted to evaluate our heuristic results against the ground truth we obtained through Mechanical Turk annotations. For each image, we computed the average number of lines at the intersection of heuristic annotation and the ground truth lines divided by the number of lines at the union of these two sets of lines. The average across images of this metric is 0.38. For reference, the average is 0.11 for a random sampling of 10000 annotations—to sample these annotations, a number of lines is uniformly sampled from [1, 4] (matching the instructions for the Mechanical Turk workers) and then each line is selected uniformly. This metric for individual worker annotations as compared to the ground truth lines is 0.47. It is reasonable and expected that this metric is lower for our heuristic method as the ground truth is computed based on the worker annotations.

STUDY DESIGN ITERATION

We conducted two small pilot studies and a formative user study with our tool as a part of an iterative design process.

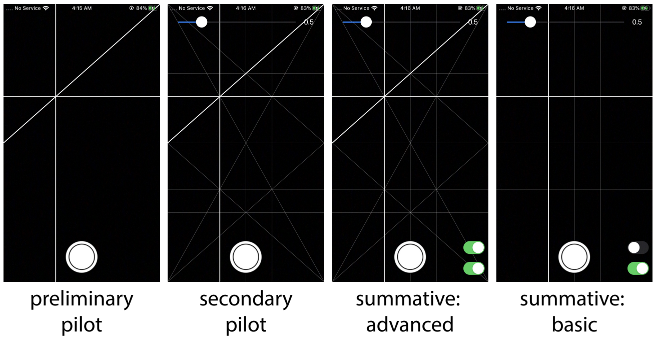

Figure 5. Iterations of the adaptive composition guidance. *Preliminary Pilot*: displays only the adaptive selected lines, updating every 0.1 sec. *Secondary Pilot*: displays non-selected lines at 0.3 opacity, slider allows user adjustment of update speed (defaults to 0.5 sec). *Summative*: iteration after pilot studies, top toggle switches between advanced and basic modes (just center and thirds, only two selected lines), bottom toggle turns composition visualization on or off.

Figure 6. Corresponding photo tasks at the two study locations. Tasks are designed to be similar in subject and complexity and to increase in scale: (1) small object on a solid surface, (2) chair/table structure, and (3) facades of a building.

These studies helped us answer some questions about different potential interaction conditions of an in-camera composition guidance tool as well as refine our interface for our summative user study, which followed.

Pilot Studies

Preliminary Pilot Study: No Guidance and Adaptive Lines

We ran a pilot study (n = 3) to evaluate our initial study design comparing two interface conditions: first no guidance, and then our adaptive composition guidance.

Participants were provided with a sheet of paper describing some basic composition guidelines to guarantee that they had a baseline perspective on how to consider composition. They were told both by the experimenter and on the sheet that they are not required to follow these guidelines. Participants were then asked to complete 6 photo tasks, focusing on composition. The first two participants did the same 3 tasks twice, once for each tool; the final did 6 different tasks, 3 per tool. After each condition, participants filled out a survey asking a variety of Likert questions (on a 7 point scale) related to their experience, and measuring the Creativity Support Index (CSI) (0 to 100) [11]. At the end of the study, they were asked to select their favorite photo per task and rate them on composition. Finally, they answered open-ended questions about what they liked/disliked about the tool and how it influenced their photo capture process. Studies were screen recorded and participants were asked to think aloud.

Preliminary Pilot Study Interface

The no guidance interface had a single interactive button for photo capture and no other interface elements displayed on the camera view. The adaptive interface additionally dynamically displayed the 3 adaptive armature lines. A line was only visible when it was selected, other lines in the grid were transparent (see Figure 5).

Preliminary Pilot Study Results

We learned that repeating the same photo tasks resulted in learning effects, such that the second time around participants already had explored the space of options and knew what photo they liked and wanted to take. They ended up using our tool primarily to refine the shot they already had in mind from the first condition. For the third pilot, we picked a second location and designed 3 tasks in the two locations that were comparable based on scale and environment (see Figure 6).

Participants gave us the feedback that in using our tool, without the context of the composition grid, the lines that appeared did not seem to have any particular structure. Additionally, the lines would appear and disappear too quickly for them to react effectively, causing frustration in using the tool. Due to the feedback from this preliminary pilot, we chose to keep the full composition grid always visible, but faintly to avoid over-cluttering the camera view (see Figure 5). We added a slider that allowed users to adjust the speed of the line updates from 0.1 seconds to 3 seconds per update (at intervals of 0.5 seconds). We set the slider default to 0.5 second updates—it appeared to still be relatively responsive to movement of the camera, at slower speeds the updates felt laggy.

Additionally, participants expressed that the adaptive highlighting helped them come up with ideas that they previously would not have thought of. We realized that this could be due to a number of factors, seeing lines of any sort overlaid on the camera view, or the adaptive nature of the lines. We hence decided to try two additional conditions: showing a static composition grid, and randomly selected lines to highlight rather than those chosen by our heuristic algorithm.

Secondary Pilot Study: 4 Conditions

We next ran another pilot study (n = 6), this time testing out all 4 conditions: no guidance, static composition guidance, random adaptive composition guidance, and heuristic adaptive composition guidance.

Participants were again provided with a sheet of paper describing some basic composition guidelines and told that they were not required to follow these guidelines. We used 4 different locations, with 3 comparable tasks per location (see Figure 6). All participants saw the tool with no guidance, followed by static guidance, and then half saw the algorithmic adaptive guidance first while the other half saw the random adaptive guidance. Locations were manually approximately counterbalanced. To reduce study length, we did no intermediate surveys, and participants were interviewed at the end about what they liked/disliked about the 4 guidance interfaces and how they influenced their photo capture process. Studies were screen recorded and participants were asked to think aloud.

Secondary Pilot Study Interface

The no guidance interface had a single interactive button for photo capture on the camera view. The static composition guidance additionally displayed the full composition grid at 0.3 opacity overlaid on the camera view. Our adaptive interface additionally dynamically displayed the 3 lines from the current annotation overlaid on the camera view. These lines were either randomly or algorithmially selected. The adaptive interface provides the user with a slider control to adjust the speed of annotation updates (see Figure 5).

Secondary Pilot Study Results

Some participants could not differentiate between the heuristic and random adaptive guidance, but those who could, expressed that the random highlighting was distracting and hard to ignore. P2 described it as "unusable," and P1 said that at first, the random highlighting "reduced trust" in the tool. While this participant expressed later that the random highlighting allowed for more creative discovery due to the more drastic changes in suggestion, we chose not to pursue this condition further due to the otherwise negative response.

Two participants expressed that the composition grid was too complicated given their limited compositional knowledge. In particular, they were unable to use the diagonal lines due to a lack of understanding of how they are commonly used in composition. Some (3) also expressed that having a better mental model of how the adaptive highlighting algorithm selected lines would allow them to better use the tool.

Formative Study: No Guidance and Static Lines

The secondary pilot helped us eliminate the random highlighting condition, resulting in 3 conditions to consider: no guidance, static composition guidance, and our adaptive composition guidance. Many cameras provide the option to overlay static composition grids of different types. However, there is little understanding of how/if these help users compose better photos or if they limit users' creativity. We were curious to better understand how such static composition overlays impact how users capture photos and how that differs from how they capture photos without an overlaid composition grid, and chose to run another formative study with two experimental conditions: no guidance and static composition guidance.

We recruited 12 participants for this formative user study: 5 self-identified as novice photographers, 4 as amateur, and 3 as intermediate. All participants went to two locations to complete 6 total photo tasks, and we counterbalanced locations and conditions. All participants performed the same tasks, conditions permitting (as we were using a public space, there were a few (5) occasions where one task was slightly adjusted due to other people occupying the space). For each task, we provided specific instructions on what should be the subject of the image. We also chose to constrain the space in which participants were allowed to move around the photo subjects. These were only stated when participants tried to walk outside of the boundaries. The experimenter brought the participant to the same starting location for each task, and tasks were always completed in the same order at a given location.

Participants were again provided with a sheet of paper describing some basic composition guidelines and told that they were not required to follow these guidelines. They were told they will be completing 2 sets of 3 photo tasks and to focus on composition—for each task, they should have a photo that they believe is well-composed. After each condition, participants filled out a survey asking them to rate their confidence in their ability to capture well-composed photos, and measuring the Creativity Support Index (CSI) [11]. At the end of the study, they were additionally asked to select their favorite photo per task and rate them composition (Likert 1-7). Finally, they were interviewed about what they liked/disliked about the tool and how it influenced their photo capture process. Studies were screen recorded and participants were asked to think aloud.

Formative Study Interface

The no guidance interface had a single interactive button for photo capture on the camera view. The static composition guidance additionally displayed the full composition grid at 0.3 opacity overlaid on the camera view.

Formative Study Results

We found that while the static composition guidance didn't improve users' opinions on the composition of their photos, it made participants feel more confident in their ability to compose photos (Mdn = 5, IQR = 4.75-5.25), no guidance (Mdn = 3.5, IQR = 2.75-4) [Wilcoxon signed-rank $V = 0$, $p = .005$]. We saw this boost in confidence in our interviews as well. P5 described that having the "guidelines [made] it much easier to have the 'correct' composition." P9 felt "more comfortable taking fewer photos" while subconsciously holding "myself to a higher standard." This was because the lines allowed the participant to precisely refine alignments before taking the photo. P10 described the lines as being "useful because it gives you greater confidence that you are taking a good photo" because they provided a concrete way of describing why it was better.

The static guidance also had a higher CSI than no guidance, suggesting that it provides better support for creativity [$V = 13$, $p = .04$]. During the study, P10 (who first used the tool with no guidance) expressed that "I already feel more creative" while completing the first task using the guidance. The individual factors that significantly improved were Exploration, Results Worth Effort, and Enjoyment (see Table 1).

However as in our secondary pilot study, we again heard feedback that the overlaid grid was too complex, "there were so many [lines] and I only knew how to use a few of them" (P5).

SUMMATIVE EVALUATION AND DISCUSSION

With our insights on the benefits of static composition guidance, we wondered how the dynamic nature of our adaptive armatures might similarly or differently influence how participants composed and captured photos.

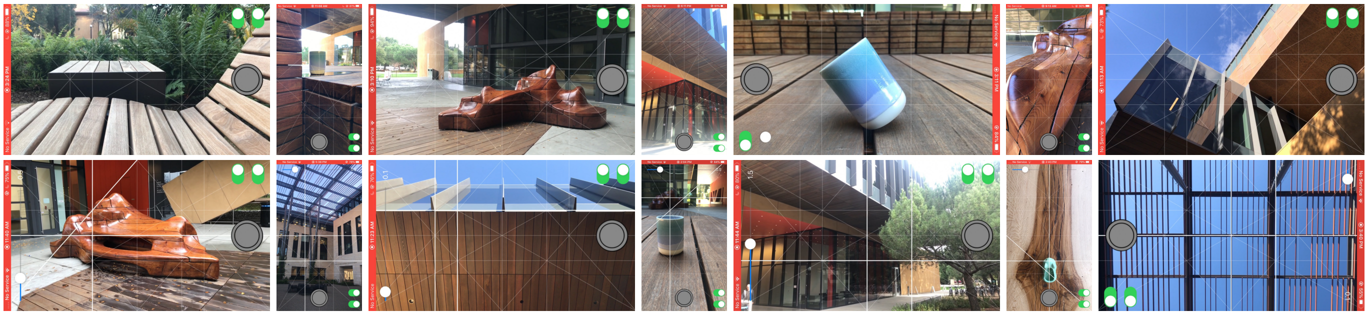

Figure 7. Some participant photos from summative user study. These are shown with the overlays that the participants saw while doing the user study. *Top*: static guidance, *bottom*: adaptive guidance. Participants often found grid lines to align to elements or edges in the image, but also sometimes used them as looser guidelines for leveling the image, splitting the image into regions (e.g., thirds), or occasionally even disregarded the grid lines.

	Formative		Summative	
	no guidance M (SD)	static M (SD)	static M (SD)	adaptive M (SD)
Exploration	8.8 (2.9)	*11.9 (3.3)	13.3 (4.4)	**13.8 (4.4)
Expressiveness	10.9 (3.9)	11.4 (1.9)	13.0 (4.3)	13.0 (4.5)
Results Worth Effort	10.4 (2.0)	*13.2 (1.7)	13.7 (4.3)	**14.3 (4.0)
Immersion	10.5 (3.2)	9.3 (4.6)	9.7 (5.3)	9.3 (6.1)
Enjoyment	7.9 (3.1)	*12.5 (3.1)	14.0 (5.0)	**13.8 (5.2)
Collaboration	5.2 (4.5)	6.1 (4.5)	5.5 (4.6)	6.1 (4.1)
Overall	**49.7 (11.0)**	***59.1 (10.7)**	**64.8 (20.1)**	****65.4 (20.0)**

* significant improvement over no guidance using Wilcoxon signed-rank test (within subjects)
** significant improvement over no guidance using Wilcoxon-Mann-Whitney test
(only correlation as this is between two different study populations)

Table 1. CSI breakdown for each study (overall score: 0 to 100, individual factors: 0 to 20). Factors are ordered in order of importance based on pairwise comparisons—orders matched for the two studies.

Summative Study: Static and Adaptive Lines

We recruited another 24 participants for our summative user study: 2 self-identified as novice photographers, 16 as amateur, and 6 as intermediate. These participants experienced two experimental conditions: static composition guidance and heuristic adaptive composition guidance.

Our summative study procedure mostly matched that of our formative study with the following changes. Since we repeatedly heard that complexity of the grid and the sheer number of lines was overwhelming, we added a short tutorial of the app that described interface elements and explained the proportions in the composition grid and showed more examples that use lines other than the rule of thirds. This was shown immediately after the composition guidelines to all participants. We also wrote another tutorial that explained the adaptive algorithm at a high-level due to the suggestion that having a better mental model of the tool would help participants use the tool in a more informed manner. This was shown to participants in addition to the description of the static composition grid before the algorithmic adaptive highlighting condition.

The experimenter instructions, surveys, interview questions, tutorials, etc. are included in our supplemental materials.

Summative Study Interface

In our final tool, to further address the complexity of the composition grid, both the static and adaptive interfaces have toggles to switch between basic (centers and rule of thirds) and advanced modes (full armature), and to turn the lines on and off. The static composition guidance displays the full composition grid at 0.3 opacity overlaid on the camera view. The adaptive interface additionally dynamically highlights the 3 adaptive armature lines (2 lines in basic mode) by increasing the line stroke and opacity of the selected lines. The adaptive interface provides the user with a slider control to adjust the speed of annotation updates (see Figure 5). See video figure for a live demonstration of the tool.

Summative Study Results

We found that users believed their photos were more well-composed when using the adaptive composition guidelines (Mdn = 5.5, IQR = 5-6) compared to the photos taken using the static composition guidance (Mdn = 5, IQR = 4-6) [Wilcoxon signed-rank test $V = 247.5$, $p = .003$]. In particular, the adaptive tool helped improve self-assessed composition of task 3 photos (static: Mdn = 4, IQR = 4-5; adaptive: Mdn = 5, IQR = 5-6) [$V = 24$, $p = .02$], arguably the most complex task due to the scale and number of lines naturally present in buildings. This was also reflected in participant interviews—when asked when they would be most likely to use the tool, many specifically mentioned more complex scenes like landscapes, architecture, and with multiple objects, etc. (11).

While we did not see a significant difference in the CSI (or for any individual factors) for the adaptive guidance over static guidance, qualitative feedback during interviews suggested it allowed for more creativity due to its encouragement of exploration due to the generation of more ideas (15). P7 said the tool had a "whimsical" quality that made interacting with the tool more fun—the tool "made me more experimental, more immersed, willing to try more things," whereas using the static guidance was more about coming up with an idea and just achieving it. P23 described that the "bright lines help you think about specific lines, making it more helpful than static lines because it gives you ideas." P0 noted that using the adaptive tool, "I noticed more perspectives that I wasn't aware of or hadn't thought of." In fact, P1 expressed usually relying too heavily on a specific type of aesthetics, and liked that the tool "would allow me to explore more by challenging me to try putting salient objects in other locations." P11 noted the "diagonal lines gave you a chance to try different angles from traditional capture—angles that traditionally might look weird, given the perspective of the lines, instead look like a nice novel way of looking at the scene."

Nonetheless, a few (3) did note that the grids made them focus more on aligning rather than coming up with new creative ideas. Photos from the study (Figure 7) show that with both static and adaptive guidance, participants were often inclined to find an alignment between the image and the grid lines. However, in some cases the image was not served by the specific composition grid and they disregarded it.

We did not see a significant difference in the participants' confidence in composing images, but again we saw support in our interviews. The tool provided "validation, while helping reduce small motor changes" (P1), served as good "secondhand confirmation that their photo is actually nice" (P5), and "reinforced notions that I already had about composition" (P14). P6 felt like the tool was "guiding me," making the capture of well-composed images less stressful.

Comparing No Guidance and Adaptive Lines

Our summative study did not directly compare no guidance and adaptive lines. Thus results here compare our adaptive interface results from the summative study to the no guidance results from our formative study—it is to be noted that these results can only suggest correlation as we are comparing between different populations, neither of which saw both conditions. Here we found similar results to static composition guidance versus no guidance from our formative study.

The adaptive guidance increased participants' confidence in their ability to take well-composed photos (no guidance: Mdn = 3.5, IQR = 2.75-4; adaptive: Mdn = 6, IQR = 5-6) [Wilcoxon-Mann-Whitney $W = 26.5$, $p < .001$]. Compared to no guidance, adaptive guidance also had higher CSI with the same creativity factors significantly improving as well: Exploration, Results Worth Effort, and Enjoyment (see Table 1).

User Scenarios

We found some common ways in which our participants used the static and adaptive guidance. Here, we walk through two user scenarios that detail some of these behaviors. Please see our video for specific examples from the studies.

While traveling, a user wants to capture an artistic photo of a unique landmark. She pulls up the tool on her camera and scans the space looking for interesting compositions.

- **Static guidance.** The user checks all grid lines as she scans the space. She eventually notices two vertical thirds lines nicely aligning with parts of the structure and decides to try out this idea. She slightly adjusts the camera to place a foreground object at the bottom right thirds intersection point. She shifts the camera horizontally/vertically, placing the object at different intersection points to test out a few compositions. She picks one, refines the alignment of the verticals, and takes a photo. She has captured this idea and continues scanning the space looking for different ideas.
- **Adaptive guidance.** The user follows the highlighted grid lines as she scans the space. Two vertical thirds lines light up, nicely aligning with parts of the structure and she decides she wants to try out this idea. As she adjusts the camera to place a foreground object at the bottom right thirds intersection point, she notices a different line highlight. After taking a photo capturing her current idea, she decides to try aligning to this newly highlighted line, and takes another photo. Again, a new line highlights, making her notice an alignment with the structure that she previously hadn't seen, giving her yet another idea.

LIMITATIONS AND FUTURE WORK

Heuristic Algorithm. So far we have built our system around the idea of composition guidance solely based on salient regions. As the end goal of composition is to achieve some sort of visual balance (or lack of), it would be interesting to take into account additional high-level cues [29]. We explored the use of edge detection in our algorithm, but this didn't improve our consistency and made understanding the result harder. Thus, we chose to go with a simpler mental model for participants so as to better study the interaction in this initial paper. As we did observe participants aligning to edges in many cases, we would be interested in further pursuing this direction in future iterations.

Interface Customization. Participants expressed interest in having more control over the set of composition grid lines. For example, some mentioned wanting to reflect personal stylistic preferences, while others mentioned setting a specific template overlay to constrain or match a given composition. This is a feature that would be straightforward to integrate via interactive selection of lines, or automatic extraction of templates from images using our current heuristics. These suggestions reinforce the usefulness of this mental model and the engagement it creates as an interactive tool.

Additional Applications. Apart from in-camera guidance, we believe our adaptive composition lines could be useful for interactive cropping tools. Likewise, appropriate saliency models [5] coupled with other grid systems [46] could provide a similar experience in graphic design tools. Other domains we are interested in include composition-based image retrieval, and other perceptually meaningful analyses of images.

CONCLUSION

Composition is an important aspect of photography. The use of composition grids is common for teaching composition. Many cameras also provide support for overlaying a range of composition grids. However, there is little support for whether composition grids are a mental model that could easily be grasped and leveraged by novices. We found that novices were relatively consistent in annotating how they perceived composition with respect to such grids. We also found that having such guidance on a camera can help users feel more confident and creative.

We explored composition guidance by creating a new kind of photographic composition guidance around our idea of adaptive armatures. We found that adaptive armatures support users both in exploring new composition ideas as well as in refining composition at capture time, aiding them in producing photos that they believe are better composed.

ACKNOWLEDGEMENTS

We would like to thank Mitchell Gordon for his helpful discussions around the study designs and analysis, Matthew Cong for his help with figures and the video, and our participants and the experts we consulted for their time and insight. Our research is supported by Brown Institute for Media Innovation, the Microsoft Research Dissertation Grant, and the Hasso Plattner Institute Design Thinking Research Program.

REFERENCES

[1] Apple. 2019. Framework: Vision. (2019). https://developer.apple.com/documentation/vision

[2] Soonmin Bae, Aseem Agarwala, and Frédo Durand. 2010. Computational Rephotography. *ACM Trans. Graph.* 29, 3, Article 24 (July 2010), 15 pages. DOI: http://dx.doi.org/10.1145/1805964.1805968

[3] S. Banerjee and B. L. Evans. 2007. In-Camera Automation of Photographic Composition Rules. *IEEE Transactions on Image Processing* 16, 7 (July 2007), 1807–1820. DOI: http://dx.doi.org/10.1109/TIP.2007.898992

[4] Bruce Block. 2013. *The Visual Story: Creating the Visual Structure of Film, TV and Digital Media*. Routledge.

[5] Zoya Bylinskii, Nam Wook Kim, Peter O'Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, and Aaron Hertzmann. 2017. Learning Visual Importance for Graphic Designs and Data Visualizations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 57–69. DOI:http://dx.doi.org/10.1145/3126594.3126653

[6] Luigi Celona, Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2019. Autocropping: A Closer Look at Benchmark Datasets. In *Image Analysis and Processing – ICIAP 2019*, Elisa Ricci, Samuel Rota Bulò, Cees Snoek, Oswald Lanz, Stefano Messelodi, and Nicu Sebe (Eds.). Springer International Publishing, Cham, 315–325.

[7] Hui-Tang Chang, Yu-Chiang Frank Wang, and Ming-Syan Chen. 2014. Transfer in Photography Composition. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 957–960. DOI: http://dx.doi.org/10.1145/2647868.2654976

[8] J. Chen, G. Bai, S. Liang, and Z. Li. 2016. Automatic Image Cropping: A Computational Complexity Study. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 507–515. DOI: http://dx.doi.org/10.1109/CVPR.2016.61

[9] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. 2017. Learning to Compose with Professional Photographs on the Web. *CoRR* abs/1702.00503 (2017). http://arxiv.org/abs/1702.00503

[10] Bin Cheng, Bingbing Ni, Shuicheng Yan, and Qi Tian. 2010. Learning to Photograph. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 291–300. DOI: http://dx.doi.org/10.1145/1873951.1873992

[11] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 21.

[12] James W Cowman. 2019. Dynamic Symmetry Art. (2019). https://www.dynamicsymmetryart.com/

[13] Jane L. E, Ohad Fried, and Maneesh Agrawala. 2019. Optimizing Portrait Lighting at Capture-Time Using a 360 Camera as a Light Probe. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 221–232.

[14] Hui Fang and Meng Zhang. 2017. Creatism: A deep-learning photographer capable of creating professional work. *CoRR* abs/1707.03491 (2017). http://arxiv.org/abs/1707.03491

[15] Naihui Fang, Haoran Xie, and Takeo Igarashi. 2018. Selfie Guidance System in Good Head Postures. In *Joint Proceedings of the ACM IUI 2018 Workshops co-located with the 23rd ACM Conference on Intelligent User Interfaces (ACM IUI 2018), Tokyo, Japan, March 11, 2018.* http://ceur-ws.org/Vol-2068/symcollab2.pdf

[16] Farshid Farhat, Mohammad Mahdi Kamani, and James Z. Wang. 2018. CAPTAIN: Comprehensive Composition Assistance for Photo Taking. *CoRR* abs/1811.04184 (2018). http://arxiv.org/abs/1811.04184

[17] Juliano Franz, Anderson Maciel, and Luciana Nedel. 2019. Instant Rephotography. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 208–227.

[18] Matila Costiescu Ghyka. 1977. *The Geometry of Art and Life*. Courier Corporation.

[19] Tavis Leaf Glover. 2014. *Canon of Design: Mastering Artistic Composition*. Tavis Leaf Glover.

[20] Tavis Leaf Glover. 2016. *Photography Composition & Design: A Fun Approach for Serious Artists*. Tavis Leaf Glover.

[21] Tavis Leaf Glover. 2019. Dynamic Symmetry Grids: Unity, Movement, Rhythm & Strength. (2019). https://ipoxstudios.com/dynamic-symmetry-grids-for-photographers-and-painters/

[22] G. Guo, H. Wang, C. Shen, Y. Yan, and H. M. Liao. 2018. Automatic Image Cropping for Visual Aesthetic Enhancement Using Deep Neural Networks and Cascaded Regression. *IEEE Transactions on Multimedia* 20, 8 (Aug 2018), 2073–2085. DOI: http://dx.doi.org/10.1109/TMM.2018.2794262

[23] Jay Hambidge. 1967. *The Elements of Dynamic Symmetry*. Courier Corporation.

[24] S. He, Z. Zhou, F. Farhat, and J. Z. Wang. 2018. Discovering Triangles in Portraits for Supporting Photographic Creation. *IEEE Transactions on Multimedia* 20, 2 (Feb 2018), 496–508. DOI: http://dx.doi.org/10.1109/TMM.2017.2740026

[25] Chuan-Shen Hu, Yi-Tsung Hsieh, Hsiao-Wei Lin, and Mei-Chen Yeh. 2019. Virtual Portraitist: An Intelligent Tool for Taking Well-Posed Selfies. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 1s, Article 12 (Jan. 2019), 17 pages. DOI: http://dx.doi.org/10.1145/3288760

[26] Mark J. Huiskes and Michael S. Lew. 2008. The MIR Flickr Retrieval Evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*. ACM, New York, NY, USA.

[27] Md Baharul Islam, Wong Lai-Kuan, and Wong Chee-Onn. 2017. A survey of aesthetics-driven image recomposition. *Multimedia Tools and Applications* 76, 7 (01 Apr 2017), 9517–9542. DOI: http://dx.doi.org/10.1007/s11042-016-3561-5

[28] Michel Jacobs. 1926. *The Art of Composition: A Simple Application of Dynamic Symmetry*. Doubleday, Page & Company.

[29] Baris Kandemir, Zihan Zhou, Jia Li, and James Z. Wang. 2017. Beyond Saliency: Assessing Visual Balance with High-level Cues. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017 (Thematic Workshops '17)*. ACM, New York, NY, USA, 26–34. DOI: http://dx.doi.org/10.1145/3126686.3126712

[30] Minju Kim and Jungjin Lee. 2019. PicMe: Interactive Visual Guidance for Taking Requested Photo Composition. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 395, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300625

[31] Shu Kong, Xiaohui Shen, Zhe Lin, Radomír Měch, and Charles Fowlkes. 2016. Photo Aesthetics Ranking Network with Attributes and Content Adaptation. In *ECCV*.

[32] M. Kumano, K. Uehara, and Y. Ariki. 2006. Online Training-Oriented Video Shooting Navigation System Based on Real-Time Camerawork Evaluation. In *2006 IEEE International Conference on Multimedia and Expo*. 1281–1284. DOI: http://dx.doi.org/10.1109/ICME.2006.262772

[33] Jun-Tae Lee, Han-Ul Kim, Chul Lee, and Chang-Su Kim. 2018. Photographic composition classification and dominant geometric element detection for outdoor scenes. *Journal of Visual Communication and Image Representation* 55 (2018), 91 – 105. DOI: http://dx.doi.org/https://doi.org/10.1016/j.jvcir.2018.05.018

[34] D. Li, H. Wu, J. Zhang, and K. Huang. 2019. Fast A3RL: Aesthetics-Aware Adversarial Reinforcement Learning for Image Cropping. *IEEE Transactions on Image Processing* 28, 10 (Oct 2019), 5105–5120. DOI: http://dx.doi.org/10.1109/TIP.2019.2914360

[35] Jia Li, Lei Yao, and James Z. Wang. 2015. *Photo Composition Feedback and Enhancement*. Springer International Publishing, Cham, 113–144. DOI: http://dx.doi.org/10.1007/978-3-319-24702-1_5

[36] Qifan Li and Daniel Vogel. 2017. Guided Selfies Using Models of Portrait Aesthetics. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 179–190. DOI: http://dx.doi.org/10.1145/3064663.3064700

[37] Z. Li and X. Zhang. 2019. Collaborative Deep Reinforcement Learning for Image Cropping. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*. 254–259. DOI: http://dx.doi.org/10.1109/ICME.2019.00052

[38] Y. Liang, X. Wang, S. Zhang, S. Hu, and S. Liu. 2018. PhotoRecomposer: Interactive Photo Recomposition by Cropping. *IEEE Transactions on Visualization and Computer Graphics* 24, 10 (Oct 2018), 2728–2742. DOI: http://dx.doi.org/10.1109/TVCG.2017.2764895

[39] Kuo-Yen Lo, Keng-Hao Liu, and Chu-Song Chen. 2013. Intelligent Photographing Interface with On-Device Aesthetic Quality Assessment. In *Computer Vision - ACCV 2012 Workshops*, Jong-Il Park and Junmo Kim (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 533–544.

[40] Chen Lujun, Yao Hongxun, Sun Xiaoshuai, and Zhang Hongming. 2012. Real-Time Viewfinder Composition Assessment and Recommendation to Mobile Photographing. In *Advances in Multimedia Information Processing – PCM 2012*, Weisi Lin, Dong Xu, Anthony Ho, Jianxin Wu, Ying He, Jianfei Cai, Mohan Kankanhalli, and Ming-Ting Sun (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 707–714.

[41] S. Ma, Y. Fan, and C. W. Chen. 2014. Finding your spot: A photography suggestion system for placing human in the scene. In *2014 IEEE International Conference on Image Processing (ICIP)*. 556–560. DOI: http://dx.doi.org/10.1109/ICIP.2014.7025111

[42] Shuai Ma, Zijun Wei, Feng Tian, Xiangmin Fan, Jianming Zhang, Xiaohui Shen, Zhe Lin, Jin Huang, Radomír Měch, Dimitris Samaras, and Hongan Wang. 2019. SmartEye: Assisting Instant Photo Taking via Integrating User Preference with Deep View Proposal Network. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 471, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300701

[43] Adam Marelli. 2013. The Great Compositions of Photographer Alfred Eisenstaedt. http://petapixel.com/2013/04/15/the-great-compositions-of-photographer-alfred-eisenstaedt/. (2013).

[44] Microsoft. 2013. PowerPoint 2013 Smart Guides. https://www.youtube.com/watch?v=fouyvhVtxWg. (2013). [Online; accessed 19-September-2019].

[45] Hiroko Mitarai, Yoshihiro Itamiya, and Atsuo Yoshitaka. 2013. Interactive Photographic Shooting Assistance Based on Composition and Saliency. In *Computational Science and Its Applications – ICCSA 2013*, Beniamino Murgante, Sanjay Misra, Maurizio Carlini, Carmelo M. Torre, Hong-Quang Nguyen, David Taniar, Bernady O. Apduhan, and Osvaldo Gervasi (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 348–363.

[46] J. Müller-Brockmann. 1996. *Grid Systems in Graphic Design: A Visual Communication Manual for Graphic Designers, Typographers and Three Dimensional Designers*. Niggli. https://books.google.com/books?id=MLYiAQAAIAAJ

[47] Tony Northrup. 2014. *Tony Northrup's DSLR Book: How to Create Stunning Digital Photography*. Tony Northrup.

[48] Yogesh Singh Rawat and Mohan S. Kankanhalli. 2015. Context-Aware Photography Learning for Smart Mobile Devices. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 1s, Article 19 (Oct. 2015), 24 pages. DOI: http://dx.doi.org/10.1145/2808199

[49] Y. S. Rawat and M. S. Kankanhalli. 2017. ClickSmart: A Context-Aware Viewpoint Recommendation System for Mobile Photography. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 1 (Jan 2017), 149–158. DOI:http://dx.doi.org/10.1109/TCSVT.2016.2555658

[50] W. Wang, J. Shen, and H. Ling. 2019. A Deep Network Solution for Attention and Aesthetics Aware Photo Cropping. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 7 (July 2019), 1531–1544. DOI:http://dx.doi.org/10.1109/TPAMI.2018.2840724

[51] Yinting Wang, Mingli Song, Dacheng Tao, Yong Rui, Jiajun Bu, Ah Chung Tsoi, Shaojie Zhuo, and Ping Tan. 2015. Where2Stand: A Human Position Recommendation System for Souvenir Photography. *ACM Trans. Intell. Syst. Technol.* 7, 1, Article 9 (Oct. 2015), 22 pages. DOI: http://dx.doi.org/10.1145/2770879

[52] Zijun Wei, Jianming Zhang, Xiaohui Shen, Zhe Lin, Radomír Měch, Minh Hoai, and Dimitris Samaras. 2018. Good View Hunting: Learning Photo Composition From Dense View Pairs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[53] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair Work: Crowd Work Minimum Wage with One Line of Code. (2019).

[54] RF Woolson. 2007. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007), 1–3.

[55] Pengfei Xu, Hongxun Yao, Rongrong Ji, Xian-Ming Liu, and Xiaoshuai Sun. 2014. Where should I stand? Learning based human position recommendation for mobile photographing. *Multimedia Tools and Applications* 69, 1 (01 Mar 2014), 3–29. DOI: http://dx.doi.org/10.1007/s11042-012-1343-2

[56] Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. 2015. Real-time Guidance Camera Interface to Enhance Photo Aesthetic Quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1183–1186. DOI:http://dx.doi.org/10.1145/2702123.2702418

[57] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. 2015. Change-Based Image Cropping with Exclusion and Compositional Features. *International Journal of Computer Vision* 114, 1 (01 Aug 2015), 74–87. DOI: http://dx.doi.org/10.1007/s11263-015-0801-5

[58] Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z. Wang, and Jia Li. 2012. OSCAR: On-Site Composition and Aesthetics Feedback Through Exemplars for Photographers. *International Journal of Computer Vision* 96, 3 (01 Feb 2012), 353–383. DOI: http://dx.doi.org/10.1007/s11263-011-0478-3

[59] Hui Zeng, Lida Li, Zisheng Cao, and Lei Zhang. 2019. Reliable and Efficient Image Cropping: A Grid Anchor Based Approach. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[60] F. Zhang, M. Wang, and S. Hu. 2013. Aesthetic Image Enhancement by Dependence-Aware Object Recomposition. *IEEE Transactions on Multimedia* 15, 7 (Nov 2013), 1480–1490. DOI: http://dx.doi.org/10.1109/TMM.2013.2268051

[61] X. Zhang, Z. Li, M. Constable, K. L. Chan, Z. Tang, and G. Tang. 2019. Pose-Based Composition Improvement for Portrait Photographs. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 3 (March 2019), 653–668. DOI: http://dx.doi.org/10.1109/TCSVT.2018.2817231

[62] Y. Zhang, X. Sun, H. Yao, L. Qin, and Q. Huang. 2012. Aesthetic composition represetation for portrait photographing recommendation. In *2012 19th IEEE International Conference on Image Processing*. 2753–2756. DOI: http://dx.doi.org/10.1109/ICIP.2012.6467469

[63] Z. Zhou, F. Farhat, and J. Z. Wang. 2017. Detecting Dominant Vanishing Points in Natural Scenes with Application to Composition-Sensitive Image Retrieval. *IEEE Transactions on Multimedia* 19, 12 (Dec 2017), 2651–2665. DOI: http://dx.doi.org/10.1109/TMM.2017.2703954