PicMe: Interactive Visual Guidance for Taking Requested Photo Composition

Minju Kim*
Graduate School of Culture Technology, KAIST
minjukim@kaist.ac.kr

Jungjin Lee*
KAI Inc.
jj.lee@kaistudio.co.kr

ABSTRACT

PicMe is a mobile application that provides interactive on-screen guidance that helps the user take pictures of a composition that another person requires. Once the requester captures a picture of the desired composition and delivers it to the user (photographer), a 2.5D guidance system, called the virtual frame, guides the user in real-time by showing a three-dimensional composition of the target image (i.e., size and shape). In addition, according to the matching accuracy rate, we provide a small-sized target image in an inset window as feedback and edge visualization for further alignment of the detail elements. We implemented PicMe to work fully in mobile environments. We then conducted a preliminary user study to evaluate the effectiveness of PicMe compared to traditional 2D guidance methods. The results show that PicMe helps users reach their target images more accurately and quickly by giving participants more confidence in their tasks.

CCS CONCEPTS

• **Human-centered computing → User centered design**.

KEYWORDS

Photography Assistance, Interactive Visual Guidance, Photo Composition, Mobile Application

ACM Reference Format:
Minju Kim and Jungjin Lee. 2019. PicMe: Interactive Visual Guidance for Taking Requested Photo Composition. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300625

*Both authors contributed equally to this work.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300625

1 INTRODUCTION

People record special moments anytime and anywhere using a mobile camera. People take pictures themselves or ask someone to take photographs of them with companions, the background, or in certain situations (Figure 1). When requesting a photo to be taken, although people provide the desired camera position and subjects to be included in the picture to others, it is often not well reflected in the photograph. According to the pre-user study, we determined that even if the camera position was set in person or photos that were previously taken were shown, it is difficult to obtain the desired pictures. The results also showed that it is difficult for a photographer who takes pictures on request to follow the intended scene of the requester precisely. In particular, we found that the composition, one of the main elements of pictures [5], is also hard to be applied as the requester intended. Therefore, in situations where people ask others to take their picture, a system that can support their communication is needed. Notably, a guide system that enables the requester to deliver the desired photo and allows the photographer to recognize and follow it accurately and easily is required.

We present PicMe, which provides an interactive visual guide to the composition that enables the user to take pictures that the requester wants. PicMe introduces a 2.5D guidance method that shows a three-dimensional composition of the target image as a virtual frame in the live camera view of the user (photographer). Once the requester takes a picture of desired composition, a virtual frame appears on the user's screen. The user can control the camera position and orientation to reach the exact three-dimensional target where the virtual frame is located. Meanwhile, PicMe continuously measures the status of the user's camera and provides additional contextual guidance and feedback. Specifically, we provide a small-size target image in an inset window as feedback and edge visualization for further detailed alignment. We implemented PicMe to work interactively in mobile environments.

We conducted two preliminary user studies. First, we evaluated the appropriate visual guidance methods for PicMe. The results showed that the virtual frame is useful to set overall direction and position, and the edge-based guide is good for detail adjustment. Second, after reflecting on the

Figure 1: Everyday situations where people ask others to take pictures.

findings of the first user study to PicMe, we evaluated the effectiveness of PicMe as a photography guide and the participants' subjective opinions regarding usability issues. Twelve subjects experimented with the target image acquisition task using PicMe compared with the edge overlay and virtual frame methods. The results show that PicMe can help participants reach the target image more accurately and quickly by giving users more confidence in their tasks. The results of usability and satisfaction also suggest that PicMe has potential as a guide application for taking desired photographs.

Our key contributions include:

(1) Conduct a pre-user study to find issues in a context where people ask and are asked to take photographs and specify design requirements.
(2) Conduct the preliminary user study to evaluate appropriate visual guidance methods for PicMe.
(3) Design and develop PicMe, which provides a virtual frame with co-located and 2.5D guidance as well as real-time feedback considering the user's context.
(4) Conduct the user study to evaluate the effectiveness and usability of PicMe, and discuss results.

The paper is structured as follows. We begin with the related study in section 2, and describe the purpose of a pre-user study, findings, and design goals in section 3. In section 4, we illustrate the whole process and results of the preliminary user study. Then, in section 5, we describe PicMe, mainly focusing on guidance methods of a virtual frame and feedback. In section 6, we evaluate PicMe and discuss results. Lastly, we present findings, limitations, and future work in section 7, and conclusion in section 8.

2 RELATED WORK

Computer-aided Task Guidance

Providing task guidance through computer-aided instruction has been studied for decades. Computer-based instructions, such as textual and graphical visual hints, can provide immediate benefits for task performance [27, 30]. While these hints are statically placed on separate screens, several researchers have also studied them using co-located and real-time guidance to offer on-demand support [14, 15]. Such systems lead users to focus their attention and get direct feedback rather than divide their attention between the instruction and task. For example, several researchers presented interactive drawing tools that assist people in creating high-quality works of art by providing guidance and corrective feedback using traditional drawing techniques [17, 19]. In a similar approach, Blackwell et al. and Fichtinger et al. explored guidance methods that enhance a surgeon's ability to perform a complicated procedure by overlaying medical images, such as CT reconstructions, in the patient's anatomy [9, 12]. In addition, several researchers projected hints using a variety of graphical representations directly on a user's body to guide a user to complete the desired movement and gesture [7, 34]. Moreover, commercial products such as Google's PhotoSphere, iPhone's Panorama, and Microsoft's Photosynth have assisted users in framing a camera shot using the image overlay approach [3, 24, 25]. With the help of visual guidance (e.g., arrows, dots, and rectangular), users can efficiently capture images to create 360-degree panoramic photos (street view) and three-dimensional models of the photos. In addition, Lucero et al. presented a service that provides users with guidance on where the next photograph could be taken by displaying virtual rectangles on the 3D map. This approach is similar to the virtual frame in PicMe [20]. Recently, Roy opted for a leader-follower approach, where a leader creates a path by walking through a route once and sends it to followers to help them reach the destination [33]. They found that the guide format reflecting the leader's intentions and important checkpoints helped followers find their way more quickly and easily.

We will refer to previous studies and explore whether co-located and real-time guidance can help people with photography tasks. In particular, after allowing the requester to generate a guide image that he/she wants, we will focus on exploring how to lead the photographer to reach the target image effectively.

Guide-assisted Methods for Quality Photos

In photography, there have been many types of research that guide users to take high-quality photos. Most of them have focused on issues of how to take pictures with a good composition and how to guide users to do so [6, 23]. Accordingly, many photography applications have provided framing suggestions based on structure objectives for balance, placement, and emphasis to users after a photo is taken. For example, Bhattacharya et al. and Yao et al. presented mobile applications that provide on-site composition feedback to the user, recommending photos that have a similar composition [8, 38]. In addition, Mitarai proposed a photo-shooting guide that suggests many types of configurations, such as triangular or rule of thirds compositions [26]. Recently, several researchers have developed real-time guidance systems to assist users

to reach the desired viewpoint when capturing a photo [21]. For example, Rawat et al. developed a real-time viewpoint recommendation system to improve a user's scene composition by using publicly available images along with social media cues [28, 29]. Xu et al. also presented a photo-taking interface that provides real-time animated arrow feedback on how to position the subject according to the rule of thirds in photography [37]. In a similar approach, Bae et al. presented a guidance system to recapture an existing photograph from the same viewpoint [4]. They provided three types of visual guidance (two 2D arrows and an edge visualization) at the same time for professional re-photography utilizing a DSLR camera with a tripod and laptop. However, it would be difficult for a user to interpret those visual guidance in the casual mobile shooting environment that we targeted. Also, Bae's 2D arrows could not describe the desired 3D orientation. In addition, a commercial application called Camera51 suggested the optimal composition to users after identifying and analyzing people, scenes, and lines that appear on the screen [1].

Meanwhile, we think our study is similar to target acquisition in AR in the sense that we lead the user to the target picture. However, there are differences in the target image approach task in our mobile environment; we guide the user not only in pointing and selecting objects but also in reaching specific 3D locations and angles as precisely as possible. For example, while Fitts and Rohs et al. mainly dealt with quantitative rules among the elements of target acquisition, we focused on proposing and experimenting with various visual guides and then connecting them sequentially by considering the accuracy and speed of each guide [13, 31, 32]. Although our purpose differs from that of Rohs' research, there is a similar point; both attempt to improve the accuracy of the predictive model by dividing the interaction into multiple parts.

Unlike the previous studies that guide users according to the established rules, our system aims to take a photograph of the composition that the requester wants. In addition, previous researchers have provided 2D guidance methods to allow a user to take photographs in accordance with the guideline. Furthermore, our approach is to provide real-time 2.5D guidance to enable the user to locate the camera at the desired three-dimensional position and orientation accurately.

3 DESIGN CONSIDERATIONS FOR GUIDING PHOTO SHOOTS

In order to understand the overall context where people ask and are asked to take photographs and specify design considerations, we conducted a pre-user study.

Pre-User Study Design

Task and Procedure: The pre-user study was an online questionnaire and participants freely filled in their opinion. We used an online survey so as to enable participation from more people to discover the difficulties and needs in everyday situations where people ask and are asked to take photographs. We composed the questionnaire contents concretely to define specific problems. The questionnaire consisted of two parts, and all participants completed both surveys. The first part is a questionnaire where the participants are the requesters of a photo shoot. In the second part, the participants are the photographers who take pictures on request. First, for the case where participants are the requesters, we surveyed people to discover the frequency, context, and reasons that they ask someone to take pictures. In addition, we investigated the resulting satisfaction level with the photographs and limitations of the pictures taken by others. Second, for the case where participants are the photographers, we surveyed which elements they consider the most important when taking pictures of others. We also asked about the degree of mental burden, such as, "Do you feel that you are adequately providing what was asked of you?" and "Do you think the requester is satisfied with the results of the photograph?" Moreover, we were interested in understanding the level of physical burden the participants endured, such as the time required to take pictures and frequency of being asked to retake pictures. Online surveys were posted in schools and local communities for two weeks.

Participants: A total of 201 people (120 females) participated in the survey. The average age was 28.43, ranging from 18 to 59 years old, and most of them usually enjoy taking pictures. The entire survey lasted about 20 minutes, and we provided 30 free coffee coupons to participants through a lottery.

Findings

Survey results showed that all participants experienced being asked to take pictures as a photographer. Alternately, as a requester, people ask others to take pictures often (17%), occasionally (73.5%), or not at all (9.5%). Most participants mentioned that they ask other people when they want to take pictures with companions as well as the background and landscape through a wide angle of view. However, the main reasons they do not ask others are because they usually were unsatisfied with the results, do not trust their shooting abilities, and asking was cumbersome. Notably, although it depends on the situation, a large number of people were generally dissatisfied or neutral with the photographs taken by others (unsatisfied: 19.4%, neutral: 65.2%, and satisfied: 15.4%). Similarly, most participants serving as the photographer commented that it was not always easy to determine what picture a requester wanted, and they could not ensure

that a requester received a satisfactory one (78%). The following two issues are our main findings.

1) Lack of communication methods between people: Many participants mentioned that as a requester, although they deliver the desired camera position and subjects they want to include in the picture, it is rarely reflected in the photograph. Some participants said that it is difficult to get pictures that they want, even if they personally set the camera position or show pictures that were previously taken to others. Accordingly, it is common to ask people to take another photo, but there is no significant improvement. Participants emphasized that the lack of such communication methods between requesters and photographers resulted in low satisfaction with the results of the photograph. Moreover, many participants mentioned that as a photographer, they were not sure if he or she was doing well as requested when shooting others. 134 participants commented that it is difficult to accurately follow a requester's specific requirements such as location, composition, directions, and many others. Some of them commented that these problems could be a mental burden when taking photographs of others (Mental demand: very high (6.5%), high (28.4%), moderate (34.8%), low (22.3%), and very low (8%)). Most participants mentioned that to solve these problems, they usually take several pictures of others from different compositions and locations, or ask the requester if he/she wants any photo retaken. Nonetheless, most participants said they do not want to spend much time taking photos of others. Specifically, 84.6% of the respondents answered that they do not want to spend more than one minute taking pictures of others. They noted the need for communication methods to identify and implement the requestor's intentions easily and quickly.

2) The importance of composition in photography: When participants take photographs, they ranked the importance of photography elements in order of composition (66.7%), shooting location (15.9%), focus (10%), and other (e.g., lighting direction, exposure, or tone) (7.4%). Most participants mentioned that they hope others will take pictures where all the graphic elements in the image are well organized and balanced. In addition, when they request their photo to be taken, they consider the composition to be the most important factor (76.6%).

Design Requirements

Our research focused on developing a guiding system that helps requester get pictures of the desired composition, rather than attaining quality photos. According to the results of the online survey, the lack of communication methods between people resulted in low satisfaction and uncertainty with the picture, hesitation of asking, among others. In particular, the fact that although requesters inform the photographer about the desired camera position and subjects they want to include in the picture, it is rarely reflected in the photograph, making our case more powerful. Inspired by these findings and recommendations, we set the following design goals for PicMe.

- **For the requester:** They should be able to deliver desired photo composition accurately and get a high-quality and satisfying picture.
- **For the photographer (user):** They should be able to take the requested picture efficiently and easily.
- **A method to deliver the requester's desired photo:** The requester can specify a photo composition that he/she wants to the photographer. Instead of conveying the desired photographs verbally or showing an example to the photographer, a method that enables the photographer to consistently recognize the desired photo during the photographing process is necessary.
- **A method to precisely and efficiently take requested pictures:**
  1) Guidance and feedback for the photographer - The photographer needs to understand the requester's vision and carry it out accurately and efficiently. In addition, the photographer needs to know if he/she is doing well. Therefore, it is necessary to provide guidance to help the photographer accurately reach the target image. It is also necessary to provide real-time feedback so that the photographer can confirm photographing state and make a correct judgment.
  2) Contextual guidance for the photographer - It is necessary to provide appropriate types of guidance according to the photographer's task performance (e.g., a difference between the target image and the photographer's camera view).

We investigated various guide-assisted methods (e.g., visual, auditory, and tactile), and we applied visual guidance in the preliminary attempt. As shown in previous studies, visual advice and feedback are intuitive methods used in various studies to capture quality photos [4, 28, 29, 37].

4 INVESTIGATION OF APPROPRIATE VISUAL GUIDANCE FOR PICME

Before we present PicMe, we experimented with visualization methods appropriate for it. We conducted a user evaluation to compare the performance of four visual guide systems: three modes that have been employed for general visual guidance [15], and a virtual frame, which we propose for PicMe (Figure 2). The three methods (picture-in-picture, image overlay, and edge overlay) used in this study have been used in various others, such as photo-shooting [1], drawing [17–19], medical [9, 12], and many others recognized as useful and intuitive guidance among various visual effects.

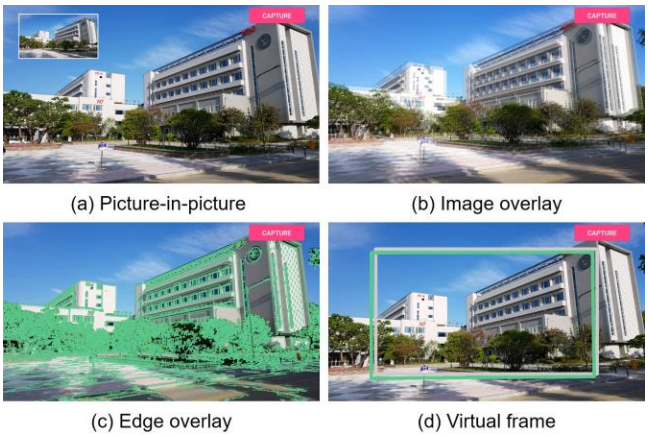

Figure 2: Four types of guides to find an appropriate visualization to apply to PicMe.

- **Picture-in-picture:** Displaying a small-size target image in an inset window in the upper left corner of the user's camera view.
- **Image overlay:** Overlaying a transparent target image on the user's camera view.
- **Edge overlay:** Overlaying a contour of the target image on the user's camera view.
- **Virtual frame:** Displaying a virtual frame that illustrates the three-dimensional composition of the target image on the user's camera in the form of 2.5D.

User Study Design

Participants: Ten participants (4 females) between the ages of 21 and 33 were recruited to participate in the study. All participants enjoy taking photographs, and all have normal vision. Each participant received a compensation of USD 10.

Task: To evaluate the performance of each visual guidance, we conducted a target image approach task. The participant performed the task of positioning the camera view towards the target image using the four guiding methods. For each task, we randomly provided target pictures of various compositions taken in advance. Once the participants reached the target image and pressed the shooting button, we measured the accuracy and task completion time. All different target images were offered to the participants. However, the distance and degree of angles between the target composition and participant's initial camera composition were kept as constant as possible. In addition, to measure the participants' task performance more precisely, the experiment was repeated five times for each mode. Accordingly, each participant performed 20 tasks (5 tasks per guide modes X 4 guide modes), and the order of all tasks and modes was counterbalanced to control the learning effect.

Procedure: We used a within-subject design with one independent variable (four guide modes) and one dependent variable (accuracy). For accuracy, we measured 1) *the mean feature error*: the average distance between the matched corresponding points that are detected from both the target image and the image taken by the user, and 2) *the image error*: the RMSE (root mean squared error) between the target image and the user's image. We also recorded task completion time, and all participants had a training time to familiarize themselves with each mode before the experiment. While more focus was placed on accuracy, the amount of time spent on others was also an important issue to consider (Section 3). Therefore, participants were asked to perform the task as accurately and quickly as possible. The entire experiment lasted about 50 minutes, and participants were observed at all times as the tasks were performed.

Results and Discussions

Figure 3 shows the mean feature error (left) and RMSE (middle) across all four modes. A repeated measures ANOVA statistic demonstrated a significant difference between the four guide modes (mean feature error: $F_{3,36} = 11.020$, $p < 0.01$, RMSE: $F_{3,36} = 13.366$, $p < 0.01$). Regarding the mean feature error, the PIP mode achieved the highest value of 17.43px, followed by Image with 9.25px, VF with 8.01px, and finally Edge with 6.89px. Post-hoc pairwise comparisons (Bonferroni corrected) showed that there was a significant difference between PIP and the other three modes ($p < 0.05$), but there were no significant differences between the three guide modes. In addition, the RMSE for the PIP mode was highest, at 42.03, followed by Image with 34.63, VF with 32.55, and finally Edge with 31.05. Edge had the highest image matching rate, but there was no significant difference between the modes except for PIP. Regarding the task completion time, statistical differences were observed between the guide types ($F_{3,36} = 12.515$, $p < 0.01$) (Figure 3, right). PIP mode was the fastest, at 38.29s, followed by VF with 64.82s, Edge with 86.81s, and finally Image with 98.70s. Post-hoc pairwise comparisons showed a significant difference between all modes ($p < 0.05$).

We focused more on the accuracy than the task completion time for people taking requested pictures. Overall, the participants performed fastest with PIP, but the accuracy was the lowest in this case. No statistical differences were observed between the accuracies of the other three guides, but people spent the longest time when using the image overlay mode. In particular, several participants mentioned that the image overlay guide was familiar, because it was visually similar to the target image. However, P4 said that sometimes it was difficult to distinguish the reference (target) image, and when the camera view became close to the target image, they were disturbed, as if seeing stereoscopic images.

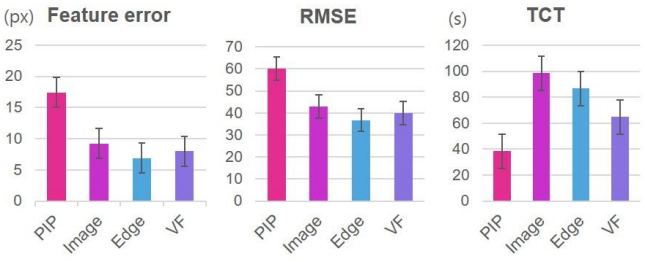

Figure 3: The results of the mean feature error (left), RMSE (middle) and task completion time (right) across picture in picture (PIP), image overlay (Image), edge overlay (Edge), and virtual frame (VF).

In contrast, participants performed the task most accurately using Edge, but spent a long time. Some participants said that they spent the most time looking for the direction and position where the target image was located. However, they mentioned that once they determined the approximate location of the target image, they were able to fit more detail into the target image. In addition, it was easy to understand the guide image by visually distinguishing it from the camera view. With VF, the participants were able to access the target image relatively quickly and accurately. Some participants said that they were intuitively able to determine in which direction and location the target image was located. However, most of them mentioned that they were uncertain of the details, because the information about the target image only appears in frames. Eight participants chose VF, and two participants chose Edge as their preferred guide mode, and none selected PIP or image overlay.

Through this experiment, we were able to identify the characteristics of each guiding method, and obtained insight that could be advantageous in certain situations. We concluded that the following three approaches could be appropriate for PicMe. We can use PIP to roughly grasp the target image quickly, and VF to help the user approach the three-dimensional position and orientation of the target image. Furthermore, we can use Edge to allow the user access to the precise position and orientation of the target image.

5 PICME GUIDANCE DESIGN

We present PicMe, which provides an interactive visual guide that allows the user (photographer) to take a requested photo composition (Figure 4). To design PicMe, we referred to the general photo-taking process that can be divided into two stages: 1) identifying and moving to a potential photograph location and 2) framing a detailed shot from the current location [11]. The key design concept is to allow the user to accurately and efficiently take pictures of others. Based on the previously identified features of each guide (accuracy, speed, users' comments), we divided the PicMe guide into three steps, each connected sequentially. PicMe provides the three guides in phases, depending on the similarity between the current camera image and the target image, without making the user select each guide separately. Accordingly, we designed PicMe using PIP for identifying desired composition, VF for rapid framing and edge-based visualization for precise adjustment. The characteristics and roles of each type are as follows, and we utilize them in a hybrid way while PicMe is running. We implemented PicMe to work fully in mobile environments.

- **Picture-in-picture:** Picture-in-picture mode provides the location information of the target at a high level to allow the user to scan the scene and move to a potential photograph location. In the visualization, a small-size target image in an inset window is displayed in the upper left corner of the user's camera view. PIP appears when the user first receives a guide for taking a picture and when the user loses the target image on screen (i.e., the difference between the target image and the user's camera view is large) during the PicMe operation. Conversely, when the user's camera position approaches the target image, the PIP disappears from the screen. Therefore, the user can have a sense of where to move the camera eye while referring the PIP according to their sutuation.
- **2.5D virtual frame:** This provides 2.5D guidance that allows the user to grasp and frame a shot of the exact three-dimensional position and orientation of the target image. When the user's camera screen and target composition begin to overlap with each other through the PIP process, a virtual frame representing the 3D composition of the target image (green-colored frame of VF in Figure 4) and a current view frame (grey-colored frame) representing the user's live view are simultaneously displayed on the screen. To reach the target image, the user can adjust the camera's position and orientation until their current camera frame is correctly aligned with the virtual frame. At this time, the user can determine how to frame the shot by getting feedback on the virtual frame that changes its shape and size according to the user's control.
- **Edge-based visualization:** This provides guides at a low level to allow the user to focus on further alignment of the detail elements from the current location. When the matching error between the target image and the current user's camera view becomes smaller through virtual frame guidance, the edge of the target image appears superimposed on the camera screen. The user can match the details, such as the range and position of the subject, to the background to be included in the photograph more precisely while comparing the edge image with the camera view of the user. In particular, our experimental results showed

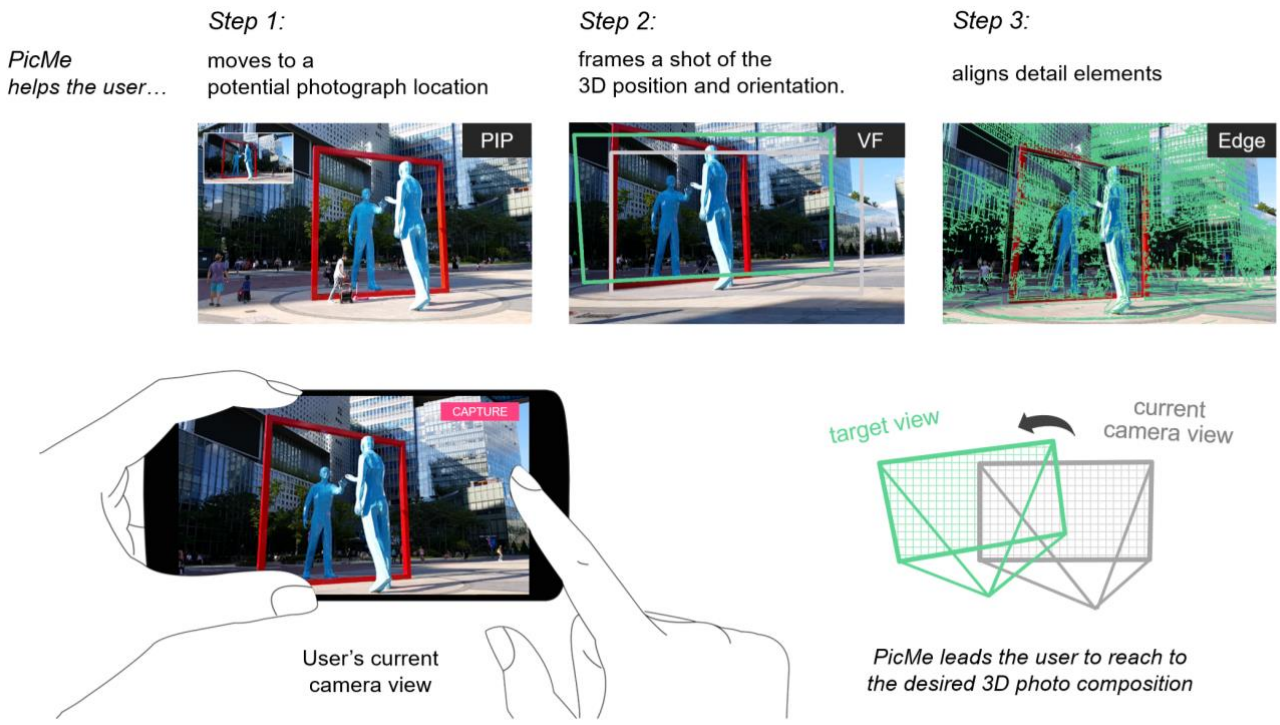

Figure 4: Three steps to help the user locate the camera at the desired composition in real-time, 1) Picture-in-picture, 2) Virtual Frame, and 3) Edge overlay.

that the edge-based guideline enables the user to reach the target position easily by visually separating the target image from the current camera view.

Once a requester captures a photo composition and delivers it to the user, PicMe provides three visualization guides depending on the process in which people take pictures and its context (Figure 5). In addition, when the user closely matches the camera screen to the position of the target image, PicMe provides feedback to let the user take a picture. When the color of the camera button changes to pink, the user is then ready to take a photo similar to the requested image.

Software Implementation

We implemented our guidance system on the Android platform using well-established computer vision techniques. The first step is to determine how well the current camera view matches with the target view. To implement this, we detected the feature points of both target image and the current camera image respectively by using AKAZE detector [2]. We then matched each feature points with the brute-force search. At this time, the fundamental matrix is estimated using the RANSAC algorithm from the corresponding feature points in order to filter out outliers [16]. In addition, we utilized inlier ratio and the average distance between inlier matches as the evaluation metrics for switching guidance types. If inlier ratio is less than 0.5, we activate PIP mode, or the virtual frame is provided to the user. To generate the virtual frame, we first estimated homography transformation between the inlier matches using RANSAC. Then, the homography transformation is applied to four corners of the camera frame to reproduce a perspective view of the target image. We used a cropped frame to visualize the frame boundary effectively. In addition, we applied temporal box filtering to each transformed corner position in order to remove the temporal jittering which is caused by instability of the feature matching results. If the average distance is smaller than 15 pixels, PicMe changes the guide type to edge-based visualization. At this time, we used Canny's edge detector [10] to produce the detailed edges of the target image. Finally, if the average distance is under 5 pixels, PicMe changes the color of the camera button to pink to let the user know the user know he/she is ready to take a picture.

In this paper, we used the results of feature point matching instead of direct pixel discrimination to switch the guide mode reliably. (e.g., when calculating the inlier ratio of feature point matching to trigger VF, the requester's region in

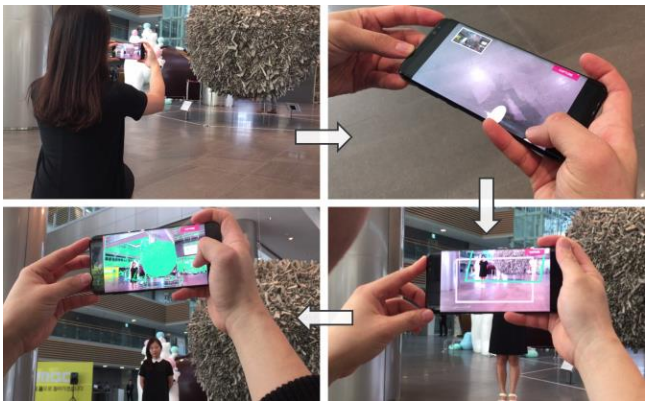

Figure 5: Real user scenario of PicMe. A requester captures a photo composition and delivers it to the photographer (user), the user grasps the target image roughly (PIP), aligns the current camera frame and a virtual frame (VF), and adjusts the detail elements (Edge).

the image has little effect on the result because there are few matching feature points). A feature point matching method is a widely used and reliable method for detecting and tracking objects in the computer vision field. We confirmed that it works reliably in our system because, in most cases, when someone requests a photograph to be taken, the requester (foreground) does not cover the target image (background) entirely as in wide-angle or full-body shots.

6 USER STUDY OF PICME

We conducted the user study to evaluate the effectiveness of PicMe as a photography guidance method. We wanted to evaluate the performance and subjective level of usability of PicMe compared with the Edge and VF methods, where the latter was designed in a previous study to have similar functionality. Especially in the aspect of the performance, our goal was to enable the photographer to take a photo with a designated composition as accurately as possible. In this context, when the matching rate between a target picture and one taken by the photographer was higher, then the quality and level of satisfaction with the picture were also higher. Therefore, in this study, we focused on measuring the user's performance of a target image approach task.

User Study Design

Participants: Twelve participants (5 females, average age: 26.5) between the ages of 20 and 35 were recruited to participate in the study. All have normal vision, and most of them usually enjoy taking pictures. They received a compensation of USD 10 each.

Task: As in the previous experiment, we conducted a target image approach task. Each participant performed 15 tasks (5 tasks per guide mode X 3 guide modes). For each task, we provided target pictures of various compositions randomly and measured two types of accuracy (the mean feature error, RMSE) and task completion time. At this time, to evaluate the accuracy precisely, the shooting location and initial position of participants were kept as constant as possible by providing target pictures. In addition, we excluded a feedback function that tells the user to take a picture from PicMe. Furthermore, a questionnaire was conducted regarding the participants' feelings on the usefulness of the guided methods and their understanding of how to take a guided picture [22].

Figure 6: Participants using PicMe for evaluating the effectiveness and usability as a photography guide.

Procedure: We used a within-subject design with one independent variable (three guide modes) and one dependent variable (accuracy). We also recorded task completion time and obtained several subjective user ratings via questionnaires. The order of the guide conditions was counterbalanced with each of the 6 possible orderings. Participants conducted a training phase, and they were asked to perform the task as accurately and quickly as possible. Also, they filled out a questionnaire upon completing five tasks of each mode. Each of the four issues had one question, and each question was rated on a 7-point Likert scale. (Q1: Through the visual guide, I was able to take the requested picture accurately. Q2: The visual guide helped me understand how to access the requested picture. Q3: I thought the visual guide was easy to use. Q4: I felt very confident using the visual guide and I would like to use this system frequently.) The entire experiment lasted about 45 minutes, and participants were observed at all times as the tasks were conducted (Figure 6).

Results and Discussions

Figure 7 shows the mean feature error (left), RMSE (middle), and task completion time (right) between all guide modes. A repeated measures ANOVA statistic demonstrated a significant difference between the three guide modes (mean feature error: $F_{2,33} = 5.346$, $p < 0.01$, RMSE: $F_{2,33} = 9.212$, $p < 0.001$). Regarding RMSE, the PicMe mode achieved the lowest value of 31.90px, followed by Edge with 35.56px, and VF with 42.40px. In terms of the mean feature error for each mode is as follows: PicMe 6.01px, Edge 7.21px, VF 8.81px.

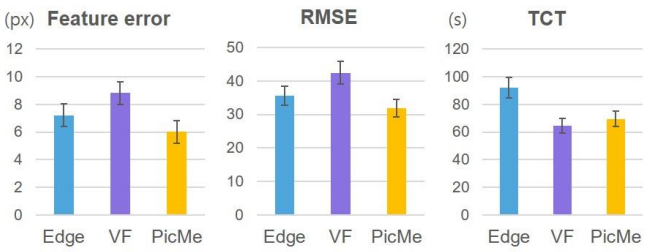

Figure 7: The results of the mean feature error (left), RMSE (middle) and task completion time (right) across edge overlay (Edge), virtual frame (VF), and PicMe.

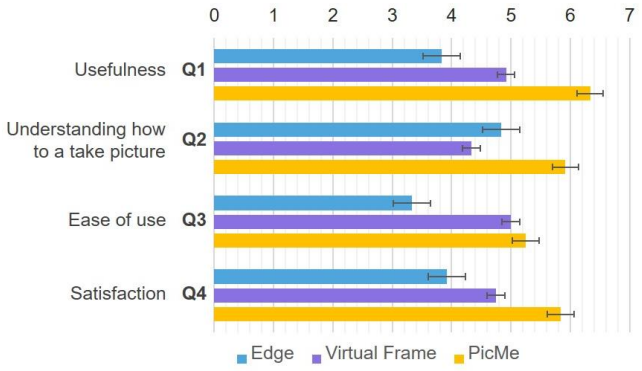

Figure 8: Average subjective responses from the participants on 7-point Likert scale. 1=Strongly disagree and 7 = Strongly agree.

Both Edge and PicMe completed the task using a stroke overlay guide, so there were no significant differences in performance (RMSE: $p > 0.581$, mean feature error: $p > 0.62$). However, PicMe required considerably less time ($F_{2,33} = 5.777$, $p < 0.007$), and the Post-hoc pairwise comparisons exhibited meaningful differences between Edge and PicMe ($p < 0.041$). For this reason, participants said that they could first get a hint regarding the 3D composition through a virtual frame, and then fit the camera view in detail.

In addition, participants also reached the target image more accurately through PicMe than with VF (mean feature error: $p < 0.05$, RMSE: $p < 0.05$). P2 and P5 stated that PicMe helped them to feel confident that they were approaching the target image effectively by providing specific guidance from the overlaid contour image. However, in terms of the task completion time, there was no significant difference between VF and PicMe ($p > 0.05$). Several participants said that even though PicMe has the advantage of providing two guides according to the context, sometimes they felt confused, especially when the guide went over from Edge to VF again. Especially, P6 stated that it would be better to impose a setting not to return to VF once the user has finished framing a shot. Similarly, P11 mentioned that he felt the alignment changed slightly as the guide type changed. He said it would be necessary to provide additional guides to bring the user's camera view to a more accurate position before making the user move to the Edge guide.

Figure 8 shows the average subject responses from the participants for the three types of a guide. The participants' subjective ratings indicate that most participants found PicMe more helpful (Q1) in performing the tasks rather than a guide with only VF and Edge ($F_{2,33} = 5.714$, $p < 0.007$). P7 and P9 mentioned that providing a contextual guide from the global to local level allowed them to conduct the task more efficiently. In Q2 (Understanding how to take pictures), there was a significant difference among all guides ($F_{2,33} = 5.267$, $p < 0.01$), especially between VF and PicMe ($p < 0.01$). Several participants emphasized that providing a certain level of detailed guidance had a positive effect on their work. The responses on Q3 (Ease of use) showed the highest preference for PicMe, but there was no significant difference between VF and PicMe ($p > 0.05$). For that reason, P4 and P6 said that they could access to the target image to some degree by using only a simple VF mode rather than PicMe that provides two types of guides. Finally, nine participants chose PicMe, two chose VF, and one selected Edge as their highly satisfied guide mode (Q4). As a matter of visual confusion when switching guide modes, two participants preferred to use only one guidance mode, either Edge or VF.

In terms of ecological validity of our user study, similar to the process followed in previous studies [17, 36], participants first underwent a training phase for the same time period for each guidance mode. However, according to the result regarding the subjective level of usability (Figure 8), the average scores of the usefulness and ease of use of PicMe were observed to be higher than those of Edge, which is a general visual guide, and there were meaningful differences between two modes. Furthermore, most participants expressed high satisfaction because PicMe helped them feel confident when they took requested photos. In this context, we can expect PicMe to be practically feasible.

7 DISCUSSIONS AND LIMITATIONS

The results of this study support our approach to guiding a user's camera view. With our system, users can perform requested photo composition more accurately than in the conventional way. However, there are remaining issues to consider for further development of our system.

Balance between level of guidance and clarity

At several modes of guidance, we made the choice to avoid visual clutter by focusing on displaying the main visual direction on the screen. However, according to our experimental

results and the user's needs, it could be necessary to provide additional guidance and feedback on the screen, to guide the user more efficiently. For example, when the users tried to align a camera view to a virtual frame, they quickly and accurately recognized the demand for moving to the left, right, bottom, and top. However, when there was a need to move forward or backward, the users showed a tendency to realize where to place the camera view after several attempts. Previous studies on providing guidance on drawing, photo shooting, and manufacturing faced similar limitations and attempted to give additional visual hints to lead the user better. With a similar approach, our system may provide arrows to give the users directional hints more intuitively. At this time, we need to consider the balance between the visual complexity and level of guidance. Alternatively, we could apply multi-sensory feedback such as auditory (e.g., speech, tone, and silent) or vibration using other sensors in the mobile phone to help users set their camera to the appropriate direction and position [18, 35].

Whole scene guide vs. representative subject guide

Our system uses well-known visualization techniques such as picture-in-picture, edge, and image overlay from previous studies, and suggests a virtual frame to guide the users in a three-dimensional composition. Through our experiment regarding the virtual frame, we found that the participants use the whole visualization of the virtual frame to match the overall composition of the target image, as we intended. However, as for the edge guide, according to the participant's observation during the experiment, the participants tended to align the current camera view and the visual guide image by fitting the edge of representative subjects (e.g., a monument/ building outline in our experiment) in the target image, rather than trying to fit the edge of the whole scene. This behavior is similar to thinking about where to position the desired subject on the camera screen when people take a picture. Accordingly, we could further develop our system by finding the main subject in the target image and then providing a visual guide only relating to it.

Additional considerations for designing PicMe

In the current stage, PicMe provides a user with three visual guides, where each guide mode switches according to the user's photo-shooting performance. However, the pre-user study results showed that participants were visually confused when modes changed automatically, especially, when they were at the boundaries of each mode (e.g., PicMe offers VF → Edge → again VF). This made them want to choose a specific mode in certain situations. Accordingly, on-demand assistance must be provided. For example, each guidance appears when the users holds or touches the screen with their thumb. In addition, apart from offering visual guide modes, exploring the meaningful interaction between the system and the users is also critical. In the future, after completely understanding the needs and preferences of both requesters and photographers, it will be possible to try various functions (e.g., functions that enable requesters to create a guide picture including themselves, using photographs taken by an expert at specific location as guide images, or allowing photographers to select whether they want to use the guide).

Improvement of software system

While our study has used simple and robust computer vision algorithms to extract visual guides, these algorithms may fail to detect the desired features in some scenes or places. Consequently, our system can sometimes miss visual guides or feedback that could help users reach the requested composition. For a similar reason, some participants said they were disturbed while taking pictures when an unstable virtual frame was displayed during the experiment. There are still limits to the operation of real-time feature tracking and displaying visualization on a mobile device; thus, many researchers used additional equipment or high-end mobile devices to solve these problems. Displaying our virtual guide over the live view of a real-world environment is similar to the visualization in augmented reality. In our system, combining object recognition and tracking technology, which have been studied in terms of augmented reality for a long time, can be one approach to drive our system more stably and precisely.

Exploration of the requester and photographer

This paper focused on evaluating visual guidance for guiding photography by the photographer who takes pictures on request. Therefore, we focused on evaluating the photographer's performance and measuring subjective opinions regarding usability issues. We also measured the participants' subjective opinions as a photographer regarding usability issues. Furthermore, in the future, it is necessary to evaluate both the requester and photographer in the context of using PicMe, wherein the two interact with each other. Additional user studies will enable us to explore more specific research issues from the perspective of both users, such as additional functions and necessary design considerations when interaction occurs. It is also necessary to investigate the requesters' satisfaction and expected level of accuracy.

8 CONCLUSION

We present PicMe, a mobile application that provides an interactive visual guide that allows the user (photographer) take a requested photo composition. PicMe provides three visualization guides depending on the process in which people take pictures and its context. We summarize our contributions as follows. First, we conducted a pre-user study

to understand the context where people ask and are asked to take photographs and specify design requirements. Second, we designed and implemented PicMe, which provides an interactive visual guide that allows the user to take a requested photo composition in real-time. Third, we conducted two user studies to evaluate appropriate visual guidance for PicMe, and the effectiveness and usability of PicMe as a photography guide. Our results highlight that PicMe not only allows participants reach the desired three-dimensional composition of a requested photo accurately and efficiently but also gives them more confident in their tasks. We believe our study illuminates the potential of an interactive visual guide to enable people to take and get desired photographs each other.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments and wish to pay sincere gratitude towards all of the participants for valuable feedback in user studies.

REFERENCES

[1] 2014. Camera51. Retrieved September 9, 2017 from https://www.camera51.com/

[2] Pablo F Alcantarilla and T Solutions. 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell* 34, 7 (2011), 1281–1298.

[3] Apple. [n. d.]. How to shoot on iPhone. Retrieved September 11, 2018 from http://www.apple.com/iphone/photography-how-to/

[4] Soonmin Bae, Aseem Agarwala, and Frédo Durand. 2010. Computational rephotography. *ACM Transactions on Graphics (ToG)* 29, 3 (2010), 24.

[5] Serene Banerjee and Brian L Evans. 2007. In-camera automation of photographic composition rules. *IEEE Transactions on Image Processing* 16, 7 (2007), 1807–1820.

[6] William Bares. 2006. A photographic composition assistant for intelligent virtual 3d camera systems. In *International Symposium on Smart Graphics*. Springer, 172–183.

[7] Olivier Bau and Wendy E Mackay. 2008. OctoPocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. ACM, 37–46.

[8] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. 2010. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 271–280.

[9] Mike Blackwell, Constantinos Nikou, Anthony M DiGioia, and Takeo Kanade. 2000. An image overlay system for medical data visualization. *Medical Image Analysis* 4, 1 (2000), 67–72.

[10] John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.

[11] Michael Dixon, Cindy M Grimm, and William D Smart. 2003. Picture composition for a robot photographer. (2003).

[12] Gabor Fichtinger, Anton Deguet, Ken Massamune, Emese Balogh, Gregory Fischer, Herve Mathieu, Russell Taylor, James Zinreich, and Laura Fayad. 2005. Image overlay guidance for needle insertion in CT scanner. *IEEE transactions on biomedical engineering* 52, 8 (2005), 1415–1424.

[13] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[14] Matthew Flagg and James M Rehg. 2006. Projector-guided painting. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM, 235–244.

[15] Michihiko Goto, Yuko Uematsu, Hideo Saito, Shuji Senda, and Akihiko Iketani. 2010. Task support system by displaying instructional video onto AR workspace. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 83–90.

[16] Richard Hartley and Andrew Zisserman. 2003. *Multiple view geometry in computer vision*. Cambridge university press.

[17] Emmanuel Iarussi, Adrien Bousseau, and Theophanis Tsandilas. 2013. The drawing assistant: Automated drawing guidance and feedback from photographs. In *ACM Symposium on User Interface Software and Technology (UIST)*. ACM.

[18] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 203–210.

[19] Jérémy Laviole and Martin Hachet. 2012. PapARt: interactive 3D graphics and multi-touch augmented paper for artistic creation. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*. IEEE, 3–6.

[20] Andrés Lucero, Marion Boberg, and Severi Uusitalo. 2009. Image space: capturing, sharing and contextualizing personal pictures in a simple and playful way. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*. ACM, 215–222.

[21] Chen Lujun, Yao Hongxun, Sun Xiaoshuai, and Zhang Hongming. 2012. Real-time viewfinder composition assessment and recommendation to mobile photographing. In *Pacific-Rim Conference on Multimedia*. Springer, 707–714.

[22] Arnold M Lund. 2001. Measuring usability with the use questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

[23] Yiwen Luo and Xiaoou Tang. 2008. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*. Springer, 386–399.

[24] Google Maps. [n. d.]. PhotoSphere. Retrieved September 5, 2018 from g.co/photosphere

[25] Microsoft. 2012. PhotoSynth. Retrieved September 5, 2018 from http://photosynth.net/

[26] Hiroko Mitarai, Yoshihiro Itamiya, and Atsuo Yoshitaka. 2013. Interactive photographic shooting assistance based on composition and saliency. In *International Conference on Computational Science and Its Applications*. Springer, 348–363.

[27] Susan Palmiter and Jay Elkerton. 1991. An evaluation of animated demonstrations of learning computer-based tasks. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 257–263.

[28] Yogesh Singh Rawat and Mohan S Kankanhalli. 2015. Context-aware photography learning for smart mobile devices. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 12, 1s (2015), 19.

[29] Yogesh Singh Rawat and Mohan S Kankanhalli. 2017. ClickSmart: A Context-Aware Viewpoint Recommendation System for Mobile Photography. *IEEE Trans. Circuits Syst. Video Techn.* 27, 1 (2017), 149–158.

[30] Lloyd P Rieber. 1990. Animation in computer-based instruction. *Educational technology research and development* 38, 1 (1990), 77–86.

[31] Michael Rohs and Antti Oulasvirta. 2008. Target acquisition with camera phones when used as magic lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1409–1418.

[32] Michael Rohs, Antti Oulasvirta, and Tiia Suomalainen. 2011. Interaction with magic lenses: real-world validation of a Fitts' Law model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems. ACM, 2725–2728.

[33] Quentin Roy, Simon T Perrault, Shengdong Zhao, Richard C Davis, Anuroop Pattena Vaniyar, Velko Vechev, Youngki Lee, and Archan Misra. 2017. Follow-My-Lead: Intuitive Indoor Path Creation and Navigation Using Interactive Videos. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5703–5715.

[34] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188.

[35] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 95–102.

[36] Sean White, Levi Lister, and Steven Feiner. 2007. Visual hints for tangible gestures in augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 1–4.

[37] 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*. ACM, 1183–1186.

[38] 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 (2012), 353–383.