Supporting Blind Photography

Chandrika Jayant†, Hanjie Ji‡, Samuel White‡, and Jeffrey P. Bigham‡
Computer Science and Engineering†
University of Washington
Seattle, Washington 98195
cjayant@cs.washington.edu

Computer Science, ROC HCI‡
University of Rochester
Rochester, NY 14627
{hanjie.ji,swhite,jbigham}@cs.rochester.edu

ABSTRACT

Blind people want to take photographs for the same reasons as others– to record important events, to share experiences, and as an outlet for artistic expression. Furthermore, both automatic computer vision technology and human-powered services can be used to give blind people feedback on their environment, but to work their best these systems need high-quality photos as input. In this paper, we present the results of a large survey that shows how blind people are currently using cameras. Next, we introduce EasySnap, an application that provides audio feedback to help blind people take pictures of objects and people and show that blind photographers take better photographs with this feedback. We then discuss how we iterated on the portrait functionality to create a new application called PortraitFramer designed specifically for this function. Finally, we present the results of an in-depth study with 15 blind and low-vision participants, showing that they could pick up how to successfully use the application very quickly.

Categories and Subject Descriptors: H.5 [Information Interfaces and Presentation]: User Interfaces

General Terms: Design, Human Factors

Keywords: camera, blind, visually impaired, photography

1. INTRODUCTION

Photography has been an important part of mainstream culture for over 100 years, helping people preserve memories, socialize, and express creativity. Blind people want to take photographs for the same reasons as everyone else, and blind photographers around the world serve as a testament to the importance of photography for blind people. The online presence of blind photographers is strong, with hundreds on Facebook, Flickr, blind photography websites and galleries, and blogs. Photographs can also serve as a way for blind people to get feedback on their environment, through automatic or human-powered interpretation (e.g., recognizing text, identifying products, locating objects). Applications like EasySnap and PortraitFramer, both introduced in this paper, seek to aid this process with framing and environmental information provided to the blind user to be used as they see fit.

How does a blind person take a photograph? To understand what real blind photographers are doing now, and to hear other blind and low-vision people's opinions and issues, we conducted an online survey with 118 blind people, which demonstrated the extent to which blind people are already taking photographs, and want to take more. To explore a general paradigm that can assist blind photography for a broad selection of tasks, we developed EasySnap, an audio-feedback camera application on the iPhone platform, and tested it out with six people. We also created a more in-depth portrait framing application, PortraitFramer, which helps frame and orient multiple people in a photograph, with different audio and vibrational cues. This was to get more detailed feedback related to a specific application, and to conduct visual observations with 15 subjects. The subjects expressed a positive reaction to this application, and all successfully used the PortraitFramer to take framed portraits with only a few minutes of training. We chose faces to use as the subject matter in order to simplify the computer vision problem and concentrate on interactions. This was also motivated by the popularity of blind people taking photographs of other people, as expressed in the survey results. Of the 118 people surveyed, 84 had recently taken photographs. Of those photographs, 52 (62%) were of friends and family, the majority of photos taken.

The contributions of this paper include: (i) an empirical foundation for understanding the need for accessible photography for blind people and demonstration that they already use cameras, (ii) the creation of a multi-modal application to get more consistent photo results given specific tasks, (iii) the creation of an accessible interface to take portraits, and (iv) usability studies to evaluate the camera interfaces.

2. RELATED WORK

Although there have not been explicit studies on blind and low-vision users' interaction with a camera, many projects and papers have mentioned the need for this type of research [20, 21, 4, 3, 30]. Usually the focus is on the computer vision algorithms and camera technology itself, while the more practical interaction techniques for picture composition are not discussed or analyzed. Exploring interaction techniques is a central feature of our work.

There has been a lot of research the last two decades in how to use computer vision techniques for applications for

ASSETS'11, October 24–26, 2011, Dundee, Scotland, UK.

the visually impaired. These projects include sign recognition [28], way-finding with environmental markers [13], shopping assistants like Grozi [18], currency recognition [26], text detection and optical character recognition (OCR) [12], and street sign and scene detection [30]. Other areas like robotics, not necessarily in the accessibility domain, use some of the same computer vision strategies [17, 16].

There are statistics that indicate that more than 100,000 blind and visually impaired individuals currently own an Apple iPhone, since the introduction of the VoiceOver screen-reader in 2007 [27]. Android phones have become increasingly more accessible as well. There is a growing number of accessible applications on mainstream devices that could benefit from some added accessibility with the camera. Some applications employing the camera already exist, such as the Looktel currency reader [29], the shopping and visual information tool oMoby [2], remote sighted assistance with VizWiz [11], and various OCR applications like the knfbReader Mobile [5].

Blind and low-vision people taking photographs may surprise many sighted people, but there is a whole community based around it. Blind photography is researched, explored, and celebrated in books (e.g., Deifell's 'Seeing Beyond Sight'), movies ('Proof' (1992) and 'A janela da alma' (2003)), news articles, websites, and art exhibits all over the world, not just in the US— from Ukraine to China, India to Israel [9, 10]. Communities such as Flickr, Facebook, myspace, Twitter, and more comprise over hundreds and thousands of interested people and blind photographers [7]. There are many websites dedicated to these groups or to particular artists. For decades, there have been, and still are, classes and books that teach blind people about cameras and how to use them [8, 15, 14].

Blind people who take part in photography consist of not only people with limited and highly attenuated sight, but also those with no sight or light perception at all. Blind and low-vision people have come up with some do-it-yourself ways to make their cameras more accessible, including adding tactile buttons or raised dots, using a sonified viewfinder, and using a viewfinder enlarger, to name a few mentioned in our survey. Such modifications often only make menus and buttons more accessible and have little to do the photo-taking composition and process. Adapted viewfinders can assist with composition in certain situations, but those discussed in our survey were for expensive digital cameras and used by professionals. Some prototype cameras have been created for blind people [19, 6], but they also concentrate on making the hardware accessible and changing the output (e.g. sounds, tactile prints, vibrations). The devices do not address the interactive photo-taking process. An everyday blind person should be able to quickly take snapshots on an accessible mainstream device just as sighted people have the opportunity to.

3. CAMERA SURVEY

We conducted an accessible online survey on camera usage that was sent out to various blind organizations, mailing lists, and companies, receiving 118 responses. The average age of the survey respondents was 40.0. There were 55 females and 63 males. When asked directly about their vision, 66 identified as totally blind, 15 had only a small amount of light perception, and 37 identified as generally low vision and legally blind. About half the respondents (56% of blind and 52% of low-vision respondents) carried an internet-enabled cell phone and an audio player with them on a daily basis.

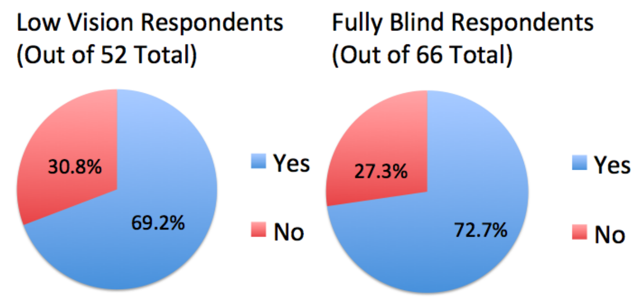


Figure 1: Percentage of how many blind and low-vision respondents had used a camera recently (out of 118 total). Of these respondents, 71% had recently taken photos.

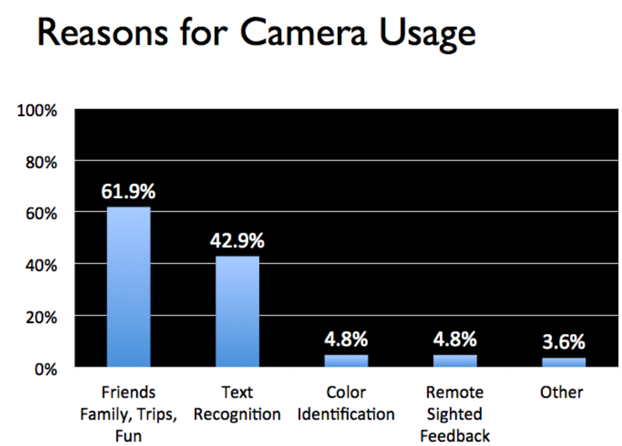


Figure 2: Reasons for recent camera usage, by percentage (out of 84 respondents who had taken photographs recently).

When asked whether being able to use a camera accurately would be useful for them, 90 respondents said yes (76.3%), two said no (1.7%), and 26 were not sure (22.0%). 84 respondents (71.2%) had used a camera (including those on cell phones) recently. 34 (28.8%) had not (18 of whom were completely blind). See Figure 1 for an overview. Some reasons cited for not using the camera included "I can't use the camera," to "I'm curious but I haven't tried." Inaccessible phone cameras were another reason to not take photos. Only two people said they did not think they could at all. Of those 84 respondents who said they had used a camera recently, the main reason for using one was to take photographs of friends, family, trips, and events (see Figure 2). Of the respondents who were totally blind, 48 had taken photos (of 66), and 18 had not. The majority of all respondents took photos as a hobby or experiment (52 of 84). Other cited reasons were for text recognition (36) and remote sighted identification (4), of which one example was taking a photo of a vending machine control and sending it to a spouse to get information about it.

The next portion of the survey posed an open-ended ques-

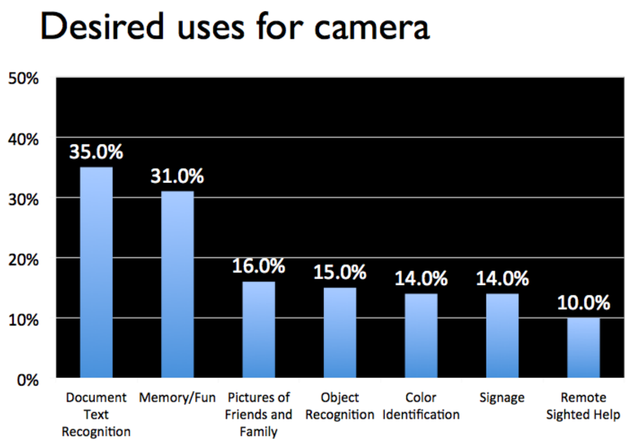


Figure 3: Imagined and desired uses for a camera in daily life, shown by percentage of respondents (118 total).

tion asking the respondents if they could use a camera, what they would use the camera for in their daily lives. Results are shown in Figure 3. The top two uses desired were for document text recognition and for fun, memory, and creativity. Next, respondents were asked to check off at most three daily tasks from a list that they could use help with in their daily life (in terms of priority and usefulness). Reading digital appliances and street signs were the top choices, at 66% and 61%, respectively.

Next, respondents were asked what type of cues they would expect to be most useful, if there was a program to help them position a camera. Choices were: Phone Vibrations (28), Audio Instructions (e.g., "Move Left," "Move Up") (55), and Audio Tones (Pitch) (29). Some open-ended suggestions given included using a complex 3D tone system for the three axes, combining some or all of the mentioned three methods, using a focal plane meter, and simply having the camera auto-shoot when "enough" of the desired object is in view (the knfbReader tells the user what percent of the page is in the view-frame, and could automatically shoot if it got to the correct percentage). One subject noted that it would really depend on what he is using the camera to do. Another wanted to be able to choose what information she received and how, and have different modes (e.g., meeting mode).

The survey results show a large desire for blind and low-vision people to take photographs, even in the absence of accessible cameras. Many people with disabilities are known to create their own interesting do-it-yourself workarounds that show a lot of creativity and reflect user diversity [23]. A surprisingly large percent of the survey respondents, even those totally blind, had taken photographs recently. Taking personal photographs (e.g., family, friends, vacations, pets) was the top reason blind people were taking photographs. This interest has not been addressed by camera phone technology. The survey gives a glimpse into the excitement, curiosity, and creativity of blind people with photography. By using the camera on mainstream phones along with computer vision, accessible cues, and/or remote human or automatic services, many daily issues of concern could be resolved or at least addressed; it seems that blind people would be very receptive to try out such applications judging by their responses.

4. EASYSNAP

Bearing the needs of blind photography in mind, we developed EasySnap, an iPhone application that assists with blind photography and provides an accessible photo album that helps users review and share pictures non-visually. It provides non-visual support to help with image framing, exposure, and blur detection. EasySnap successfully achieves two goals: (i) real-time feedback while taking a picture and (ii) generality in assisting a broad definition of photographs.

4.1 EasySnap Application

EasySnap has three modes: "Freestyle", "People", and "Object". The simplest mode is "Freestyle" mode, which functions like an ordinary camera, providing no audio feedback. With no constraints, simply by point-and-shoot, users are given the most freedom in taking pictures. Users are still given feedback regarding blur and darkness, which earlier work has shown are common problems [11]. "People" mode and "Object" mode provide a real-time status report of the person or object that one wants to take a picture of while moving the camera to frame the view.

"People" mode is specifically designed to take pictures of a person. It detects whether there is a face in the view of the camera, and tells users its location and size. Once the mode is activated, users move the camera slowly around the general direction of the person that they would like to get in the photo. If there exists a face in the frame, the real-time feedback reports how much the face takes up the screen, its location in the frame, and how the phone is angled. Otherwise, it reports "Searching" every two seconds. In this case, the generic face detection algorithm from the OpenCV library is used and well-tuned– using a bigger search window and limiting the smallest face that can be found– for high speed performance on the iPhone. The algorithm uses a cascade of boosted classifiers working with Haar-like features, which are trained with pictures of front faces as positive examples [1].

"Object" mode is designed to help take pictures of objects in the environment (e.g., a book, a cup, a piece of furniture). In this mode, users first take a picture of the object up close, and then EasySnap will provide audio feedback to help make sure the object stays in the frame as users move back to frame or change the point of view of the camera. The feedback, which is reported every three seconds, consists of three parts: the current position of original view, how much the original view is taking up the screen, and the phone orientation with respect to gravity. Here is an example of the feedback: "Bottom right, 60 percent, slightly angled down." The feedback functions as a useful input to the users instead of an explicit instruction so that they can autonomously adjust their framing according to their artistic or practical needs.

Instead of using complex and un-robust computer vision methods such as image segmentation, a light-weight tracking algorithm is employed to generate real-time framing feedback of the present view. Specifically, the close-up picture is captured as the initial view, from which a set of SURF feature points is calculated and continuously tracked in the subsequent frames by the Lucas-Kanade optical flow method. A bounding box of the tracked points is generated in each

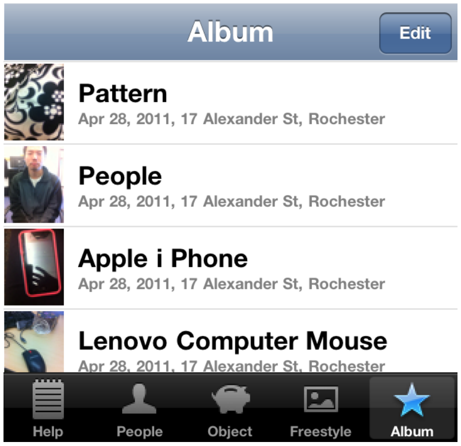


Figure 4: An example of part of EasySnap's photo album - photos are tagged with a descriptive label, and the time and location where they were taken.

frame. A percentage that indicates the area ratio of the original bounding box to the new bounding box is reported to users. From this number users are able to infer how far they have gone away from the subject and how much the object is taking up the screen. The object location is calculated based on the coordinates of the bounding box and is reported back to users in one of the following position feedback: "top left", "top", "top right", "left", "center", "right", "bottom left", "bottom", "bottom right". The application also warns users when the phone is angled down or angled up with respect to gravity using the accelerometer built in the iPhone. This series of operations runs at 4-5 frames per second.

The final images captured in "Object" and "People" mode are checked for proper exposure and sharpness. Exposure detection takes precedence and works by creating a grayscale histogram to check for any large concentrations of dark pixels indicating insufficient lighting. Sharpness is estimated by computing the mean and standard deviation of an image from its binary map and evaluating these values using a set of pre-built covariance matrices created from images known to be blurry or sharp [24]. At the user interface level, users are warned with an audio alert after successfully capturing a photo if it may be too blurry or too dark, and are requested to retake the picture or continue.

EasySnap also implements an accessible album (See Figure 4). After each image is taken, the GPS coordinates of the phone are sent to Google Maps to fetch the location, and the image is sent to IQEngine [2] to recognize its content. Both location and content are leveraged to automatically label the images in a timely fashion, which can greatly help people with vision loss to browse the album.

4.2 Study

To explore the effectiveness of EasySnap in assisting blind photography, we conducted a study with six people (three blind people, two visually impaired, and one low vision) ranging in age from 19 to 60. Before the study, the participants were briefed on the idea of the application and its interface, tried out iPhone gestures and VoiceOver (if the user had never used iPhone before), and familiarized themselves with each mode by shooting two pictures in each mode, respectively. In the formal study, they were asked to take three photos in each of "People" and "Object" mode, along with corresponding pictures with "Freestyle" mode for comparison. In "Object" mode, three objects of three different sizes were randomly picked from the environment. To alleviate short-term memory of the position where they just took the pictures, the picture taking order is randomized to create unbiased comparison pairs. See Figure 6 for an example.

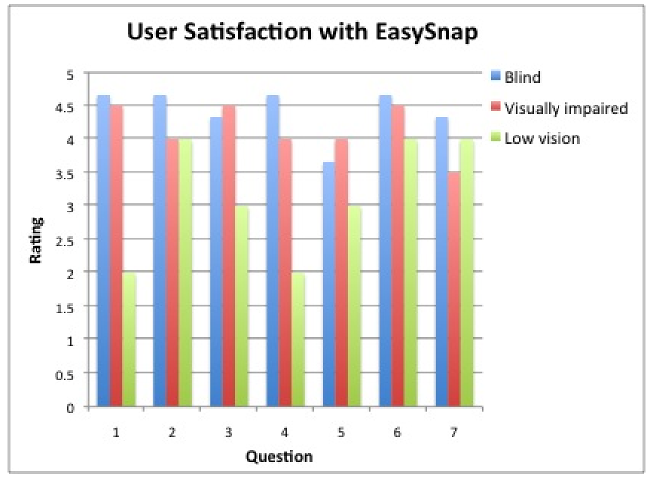


Figure 5: Likert scale results from 1 (disagree strongly) to 5 (agree strongly). (1) It is difficult for me to take pictures normally. (2) EasySnap helped me take better pictures. (3) I understood the directions that EasySnap gave me. (4) EasySnap helped me take pictures of objects. (5) EasySnap helped me take pictures of people. (6) The photo album was easy to use. (7) I would use EasySnap if it was available. Results demonstrate participants found EasySnap useful in assisting photography (2,3,4,7).

4.3 Results and Discussion

At the end of the study, participants were asked to take a short survey (See Figure 5). The results show that most of the participants agreed that EasySnap helped their photography and found it easy to use, and that the two blind participants, in particular, thought so more than others. In addition, one of them left the following comment, "I have no idea what [it] is going to be when I walked in, but it actually works. It feels like having a 'cane' while taking pictures."

Besides the direct positive feedback from the study participants, we were also interested to see what "unbiased" viewers would think about the effectiveness of EasySnap. To this end, we put the pictures taken by the participants onto a web page and presented them in random order to 31 people, who judged which of each of the pairs of photographs they preferred based on the framing criteria (i.e., the picture is better centered, the picture is taken at a better distance). In total, for the 36 pairs of images (18 for "Object", 18 for "People") we collected 1116 evaluations, 58% of which are better with EasySnap's feedback, 29% are better without feedback, and 12% are neutral. With a close look at the results we found that both "Object" and "People" modes with feedback have achieved around a 60% success rate in assisting photography (See Figure 7).

(a) With "Freestyle" mode. (b) With "Object" mode.

Figure 6: An example of two photographs of the same object (a stuffed panda) taken with 2 modes of EasySnap. a) Photo of object taken in "Freestyle" mode. b) Photo of object taken in "Object" mode.

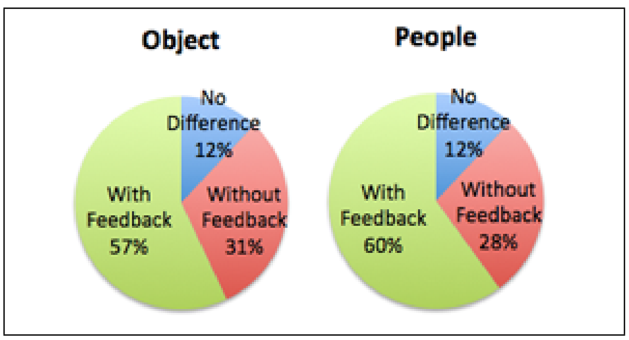
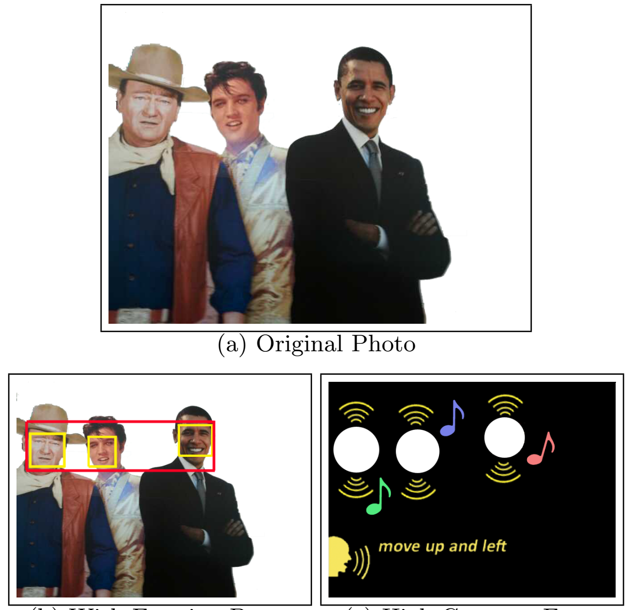


Figure 7: EasySnap evaluation results.



(a) Original Photo

(b) With Framing Boxes (c) High Contrast Faces

Figure 8: Steps of PortraitFramer application. a) Photo taken (background removed for clarity). b) Faces found. c) If there are faces, PortraitFramer announces the number of faces, and vibrates where those faces are on the touchscreen while also showing them as high contrast circles. When each face circle is touched, it plays a short, distinct pitch. Instruction mode can be turned on to provide directional instructions. The user can accept this photo and save it, or try again, by swiping left or right respectively.

5. PORTRAITFRAMER

Motivated by the results of the aforementioned survey, along with EasySnap we concurrently designed a group portrait application for phones running on the Android platform (1.6+) because of Android's text-to-speech API. The interface is a self-voiced portrait taker that provides interaction cues for a blind or low-vision person as an additional tool to help shoot "well-framed" photographs of people. Face detection uses fairly straightforward computer vision compared to identifying general objects or reading text, therefore allowing us to hone in on the interaction issues. While EasySnap works with one person, PortraitFramer works with groups of people as well.

The application starts by asking the user to click the camera button to take a photo. The user is told how many faces are in the camera's sight. Concurrently, there is a screen that comes up which is black, with the face areas in white (Figure 8(c)). The visible cues are suitable only for those with some limited vision. Touching the "face areas" causes the phone to vibrate, so the user can explore the touchscreen and get a sense of where the faces are in their camera view frame. Users might just want access to the information and decide themselves where they want to position the people in the photograph, using audio and/or haptic cues. No directions are given about the z-axis, or distance to the scene; however, the user can feel or see how large the white "face areas" are to get a sense of their distance from the subject. Other information that can be accessed through the interface is the face borders (including overall framing border, see Figure 8(b)) and the actual size of the face areas in relation to the screen view, similar to EasySnap's "People" mode.

5.1 First Study Setup

To get user feedback on the design of PortraitFramer, formative user studies were conducted with eight blind and low-vision people. After ten minutes of training, the participants were asked to take three photos using the PortraitFramer application. They were asked to go through their steps verbally. Three cardboard cutouts were set up in an office, to simulate people (See Figure 8(a)). Naturally, the human to human interaction part of the photo-taking experience is extremely important, with communication between the photographed and the photographer possibly changing both the location of the people or the aim of the photographer. However, for this initial formative study, we wanted to have all the participants taking photos of the same setup to more easily notice differences. The task was to try to take a centered photo of the three cutouts. All photographs with face framing were saved to the phone to see the progression of localizing the faces (See Figure 8 for one example). After about 20 minutes of taking photos, we asked their overall thoughts about the software, if it seemed useful, and what problems and suggestions they had.

5.2 Results

Our subject pool consisted of four males and four females, and the average age was 48.1. Three of the participants were fully blind. For those with low-vision, vision problems included having poor light/dark perception, only being able to see blurry shapes, having no peripheral vision, and being born blind and regaining some sight. Five participants had used a camera before. Participants ranged from low to high-tech, in terms of the devices they carried around with them. Four had previously used a camera. Participants were asked what would be motivating uses of the camera for them, if any. Three mentioned OCR for text documents, street and office signs, and unfamiliar documents while on the go (e.g., airplane magazine, menu). One man regretted not chronicling the last 10 years of his life. Two participants wanted to be able to send a photo to a remote human service or family member to get it identified. Four were eager to take photographs of their friends and family. Beauty, creativity, and art (e.g., architecture, sunset) were reasons expressed by four participants. One woman wanted to take photographs to study up close later to come up with characters for her novels. One subject said she only saw using a camera for OCR, but then mentioned that she had wished she could take photographs at her family reunion to show her mother.

All eight study participants successfully centered the faces after getting vibration and overlay cues from the application. Figure 8 shows the succession of photos taken from one subject. The average amount of time to take a successful photograph was 13 seconds. All participants took a successful photograph within three attempts. Success was measured by having the bounding box of three faces touch all the quadrants of the screen, but not overflow it.

5.2.1 *Space and Pose Estimation*

Space issues were the most difficult to predict or understand. Five participants started off with a discussion of basic camera skills and their understanding of how a camera works. Three people were confused that distance made a difference; they had to think for awhile before understanding that the further away they took the picture, the more people could fit into the photo. Blind people have a very different sense of space than sighted people, especially having to do with perspective [22]. Three participants had an excellent sense of distance. Two had a little trouble holding the camera vertically straight. All but two moved the phone along the x-y plane when trying to take a more accurate picture according to the program's feedback; the other two tilted the phone (from the same center point) around the x and y axes. It was difficult for one of the blind participants to understand the mapping between the vibrations and the physical scene they were photographing. We had to explain that we were only simulating localized vibration, and that the whole phone would actually vibrate when the faces were touched. One subject had no touchscreen or smartphone experience but had used the Optacon, which allowed her to easily grasp the concept of the screen layout relating to the physical world (the Optacon raises the shapes of letters that it sees through the camera [25]). For two participants, it was difficult to distinguish between the various spots that caused vibrations, especially if two face areas were close. Study participants could tell the difference between small and larger face areas by sight or vibration.

5.2.2 *Suggested Cues*

There were many suggestions given by the participants in terms of what cues they would like from the software. Two participants suggested using musical tones or different types of vibrations instead of the same vibrations to indicate which part of screen has a face, in order to find the spots which were hard to localize for two participants. One subject suggested using volume to indicate the size of the head and proximity of the person. Three participants strongly preferred to have speech directions. One was concerned about having too many noises when they were out in public, but acknowledged that often earbuds were used regardless while using the phone (one in, one out). Two low-vision participants wanted options for seeing a more contrasted framing overlay over the actual video screen, and wanted zoom ability with holistic context. One person wanted the camera to beep when it was at the "right" distance away from the subject and automatically take the photo then.

5.2.3 *Reactions and Public Attitudes*

Two participants said they would be uncomfortable with obvious verbal cues spoken out in public from the application. One subject felt the opposite, and thought that it could explain to people what was going on, and spur interesting conversations. Reactions were overall extremely positive. Only one subject was skeptical and said he did not see himself excelling in taking pictures of people. One was visibly excited about the practicality and potential of this application and wanted to use it right away. "This is something I can imagine blind and low-vision people using right out of the box. People who decided pictures were no longer for them would say "Now I can take a picture"." Other comments were that the application was "cool," "cute," and "very easy to use." In response to the statement "I liked this application," on a Likert scale from 1 to 7 (1 being strongly disliked it, 7 being loving it), the average response was 5.5. In response to the statement "I would use this application," (1 being never, 7 being very often), the average was 3.9. One participant thought his vision was not bad enough to need the application, and one subject mentioned she only wanted to use OCR and never take photographs of people. When asked about what reactions sighted people might have, seeing a blind person using a camera, two people were worried that it would seem they were faking being blind; with both a cane and a camera, one woman felt very self-conscious and encouraged us to talk to sighted people about their opinions.

5.3 Second Study

Based on the results of this study, we changed a few parts of the PortraitFramer application for a second iteration to make its likelihood of adoption potentially higher. Changes included adding in quick swiping gestures for saving photos or trying again, adding in pitches to the vibrations of the faces (each face would give emit a different pitch when touched, to distinguish between them all), a tilt correction option, and an option to have the program explicitly tell you how to move the camera in order to center the faces.

This study was similar to the first one, in which it was a formative study and participants were asked to take photos of the three faces and center them. We tested this out using instructions as well as the freeform option (no instructions, but still with vibration, pitch, and overlay cues).

5.4 Results

There were seven subjects for this study, with an average age of 40.1. Subjects did not overlap between studies. There were five females and two males. One subject was completely blind, five had severe low-vision with barely perceptible light perception, and one was low-vision but could see shapes. Five had used a camera before. One said that when taking portraits, "I aim it at the sound." Another said, "When I take pictures of my grandkids, my daughter tells me "Higher, higher!" or helps with where to aim." Using the instruction mode, all participants took a successfully centered photo within five seconds (average 3.2) and three tries (average 2.8).

When showed the tilt option, no one in this study found that they would want to use it, having no trouble keeping the phone straight up and down. All of the participants seemed to be good at layout concepts (e.g., face is in the left of screen, so move the camera left, or whether to move forward or backward to change size of faces). When touching the screen to find the faces, one user was extremely methodical (up/down then horizontal), one used a seemingly arbitrary approach of exploring the screen, and the remaining five lay somewhere in between.

Six participants liked having the addition of pitch to the vibration of the faces, for added verification and for situations like crowds. Three preferred using the instruction option: "As much detail as possible is great" said one. Two liked explicit verification of success: "Great, take a photo!" Another subject felt the opposite saying the less information he could get away with, the better, for the sake of simplicity, speed, and autonomy.

There were four user-suggested interaction techniques that came out in the post-study interview. One of them would announce the general size of each face when it was touched on the screen (i.e. "small," "medium," "large"). Another suggestion was similar but would give even more detail, speaking the percentage each face took up of the screen. I had been using random pitches for each of the faces when touched, instead of ordering them left to right or up to down in terms of pitch, which was suggested. Another suggestion was to give each face a spoken number from left to right when touched, so that the user would have a more absolute understanding of where each face was in the screen, and not relative in terms of pitch. One participant wanted an option to toggle between "point and shoot" mode and "more detailed instruction" mode. The two youngest participants expressed a desire for being able to use this application to help them tag their friends in Facebook pictures, and two participants wanted to be able to either tag already existing photographs, or to be able to name and annotate photographs before saving them. Three other participants expressed a desire for facial recognition to be built into this application as well.

Reactions to the application were all positive and excited, even with one subject who came into the study quite skeptical. "It's cool, I like it!" "It's fun. I would use it on my phone." The "app makes me feel confident that i didn't chop the heads off." "I could take pictures of my friends for fun and put it on my Facebook". We asked the same questions as in the first study based on a Likert scale of 1 to 7. For the first statement, averages were around 5.5 for the first version and 6.2 for the second version of PortraitFramer. These numbers were quite high, and didn't have large standard deviations. However, the second statement resulted in much more varied responses. The first version had an average score of less than 4, while the second version was closer to 6. Not only that, but in the first version, four participants gave very low scores, while in the second one, everyone scored at least a 5. The change in results for statement 2 was significant, using a standard t-test ($t = 2.15, DF = 13, p < .05$). This is promising in terms of adoption of this application in real life.

6. DISCUSSION

The fact that so many blind people already take photographs was surprising to us, even though we knew there was a growing community. We had also expected most photographs to be for practical matters, such as OCR or barcode scanning. A lot of practical task applications using cameras and computer vision have not gotten good enough to be consistently used and trusted by blind and low-vision people, and our EasySnap results show promise for aiding in the framing process.

The desire and excitement for taking photographs of people in their lives from the survey was a strong motivation to come up with an application to do so, and does not require heavy computer vision. All of the fifteen blind and low-vision study participants picked up on how to use PortraitFramer within mere minutes.

The importance of customization and user preference options became quite clear through our studies. There were some strong reactions about sighted people's potential reactions to seeing a blind person taking a photograph, which begs more research as mainstream technology is becoming more and more accessible and universal. Themes of security and privacy, responsibility, independence and autonomy, convenience, personal expression, and social acceptance need to be considered in the design process and will result in better technology and experiences for all, not just blind people. Balancing an effort between the user and technology, with the constraints of the users' preferences in mind, is a large but important task for future human computer interaction. Customization is key, and the user can and should decide how much or how little information they want and how they want it presented to them.

While the target user base in this work is blind and low-vision people and has been designed with them in mind, many of these ideas can be considered for different user groups in different situations. Novel input and output methods for people with different situational impairments, preferences, and abilities should be studied in general.

7. CONCLUSION AND FUTURE WORK

In this paper, we have contributed an empirical foundation for understanding the need for accessible photography for blind people and demonstrating that they already use cameras, the creation of two accessible interfaces to take pictures of objects and people by blind people, and usability studies to evaluate these interfaces. Many blind and low-vision people are already using what resources they have to be a part of the world of photography. Given the desire for more accessible camera applications and the prevalence of relatively cheap, accessible mainstream phones, we should leverage the opportunity to include future blind users into the design of novel interactive user interfaces for taking photographs. Taking portraits is one of the more popular reasons blind people already use cameras, and creating an accessible interface on mainstream technology to help do so as well as help framing other practical task-oriented applications, has the potential to have a large effect, even bringing in those blind people who did not think they could ever take a photograph before.

More in-depth studies are presently being conducted on what types of interaction cues on mobile devices work best for different tasks in terms of speed and preference. A plethora of specific task-oriented applications (e.g., recognizing street signs, qr codes, text, and faces) could benefit from better framed photographs taken by their blind and low vision users. The rise of programmable cameras comes at the perfect time where we can make sure to include blind and low vision users in the design loop and provide more universal customization options for future users.

8. ACKNOWLEDGMENTS

The authors would like to thank all the survey and study participants and Professor Ladner at the University of Washington. This work has been supported by Google and NSF Award #IIS-1049080.

9. REFERENCES

[1] http://opencv.willowgarage.com/wiki/facedetection.

[2] http://www.iqengines.com/.

[3] *A Video Based Interface to Textual Information for the Visually Impaired*. IEEE Computer Society, 2002.

[4] *Trinetra: Assistive Technologies for Grocery Shopping for the Blind*, 2007.

[5] knfbreader mobile, knfb reading technology, inc. http://www.knfbreader.com/products-mobile.php, 2008.

[6] Touch sight, camera for the blind. http://www.yankodesign.com/2008/08/13/this-camera-is-outta-sight/, 2008.

[7] Blind photographers. http://blindphotographers.org/, 2010.

[8] Blind with camera school of photography. http://blindwithcameraschool.org/, 2010.

[9] Quiet-light photography. http://www.quietlightphoto.com/, 2010.

[10] Tim o'brien's photos. http://www.timobrienphotos.com/, 2010.

[11] J. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. Miller, A. Tatrowicz, B. White, S. White, and T. Yeh. Vizwiz: Nearly real-time answers to visual questions. *UIST 2010*, 2010.

[12] X. Chen and A. Yuille. Detecting and reading text in natural scenes. In *IEEE Computer Vision and Pattern Recognition*, pages 366–373, 2004.

[13] J. Coughlan and R. Manduchi. Color targets: Fiducials to help visually impaired people find their way by camera phone. *EURASIP Journal on Image and Video Processing*, 2007.

[14] G. Covington. Let your camera do the seeing: the world's first photography manual for the legally blind, 1981.

[15] G. Covington. *Access to Photography*, pages 26–30. National Endowment for the Arts, 1989.

[16] B. Deville, G. Bologna, M. Vinckenbosch, and T. Pun. Guiding the focus of attention of blind people with visual saliency. In *Workshop on Computer Vision Applications for the Visually Impaired*, 2008.

[17] M. Dixon, C. Grimm, and W. Smart. Picture composition for a robot photographer. In *Technical Report WUCSE*, 2003.

[18] G. Foo. Summary 2009 grocery shopping for the blind/visually impaired. National Federation of the Blind, 2009.

[19] N. Haidary. Camera for the blind. http://nadeemhaidary.com/camera.html, 2009.

[20] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 159–166, 1999.

[21] J. Ivanchenko V., Coughlan and H. Shen. Crosswatch: A camera phone system for orienting visually impaired pedestrians at traffic intersections. In *Computers Helping People with Special Needs*, pages 1122–1128, 2008.

[22] B. Jones. Spatial perception in the blind. *British Journal of Pyschology*, 66(4):461–472, 1976.

[23] S. Kane, C. Jayant, J. Wobbrock, and R. Ladner. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. ACM, 2009.

[24] J. Ko and C. Kim. Low cost blur image detection and estimation for mobile devices. In *Proc. of the 11th international conference on Advanced Communication Technology - Volume 3*, ICACT'09, pages 1605–1610, Piscataway, NJ, USA, 2009. IEEE Press.

[25] J. Linvill and J. Bliss. A direct translation reading aid for the blind. volume 54, pages 40–51, 1966.

[26] X. Liu. A camera phone based currency reader for the visually impaired. pages 305–306, Halifax, Nova Scotia, Canada, 2008. ACM.

[27] U. B. of Engraving and Printing. Bureau of engraving and printing launches eyenoteapp to help the blind and visually impaired denominate us currency, 2011.

[28] P. Silapachote, J. Weinman, A. Hanson, M. Mattar, and R. Weiss. Automatic sign detection and recognition in natural scenes. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 3. IEEE Computer Society, 2005.

[29] J. Sudol, O. Dialemah, C. Blanchard, and T. Dorcey. Looktel: A comprehensive platform for computer-aided visual assistance. *CVAVI Workshop Proceedings of CVPR*, 2010.

[30] M. Vazquez and A. Steinfeld. An assisted photography method for street scenes. *Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision*, 2011.