2018-12-01

Cake Cam: Take Your Photo and Be in It Too

Candice Lynn Davis
Brigham Young University

Cake Cam: Take Your Photo and Be in It Too

Candice Lynn Davis

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Michael D. Jones, Chair
Kevin Seppi
Casey Deccio

Department of Computer Science

Brigham Young University

ABSTRACT

Cake Cam: Take Your Photo and Be in It Too

Candice Lynn Davis
Department of Computer Science, BYU
Master of Science

In this thesis we explore a different kind of computer-mediated collaboration (CMC) in which two people collaborate to perform a single task in a time-sensitive asymmetric situation wherein one person is more invested in the outcome than the other is. Our conjecture is that interactive computing can quickly communicate intent, which may help mediate time-sensitive asymmetric tasks between two people.

We explore this idea specifically in the context of asking a stranger to take one's picture at a tourist site. Consider a tourist handing a cell phone to a complete stranger and asking the stranger to take the tourist's picture in front of a landmark. This collaboration often leaves the tourist unhappy with the final image. The tourist cannot quickly communicate the intended framing to the stranger who is, quite likely, in a hurry and, frankly, not interested in the final picture's quality. We explore mediation of this interaction with a mobile app, titled *Cake Cam*. Cake Cam is a computing tool used to communicate intent in time-sensitive asymmetric tasks wherein one collaborator is more interested in the outcome's quality than the other. To use the app, a tourist takes a photo of the scene, carefully framed to his or her desires. The tourist then hands the phone to a stranger, asks the stranger to take the tourist's picture, and moves into the frame. Augmented reality alignment markers guide the stranger into taking the photo the tourist initially framed, producing the intended photo. We found that Cake Cam was more effective and efficient in guiding users into replicating a photo than verbal descriptions were. On average it took almost 3 more tries for the participant to take the correct photo without Cake Cam, than with it. Additionally, with Cake Cam, the final photo was closer to the intended framing by an average $9cm$. Participants that used Cake Cam found the process to be less difficult and were more confident they had captured the intended photo than the participants that used the normal camera app.

Keywords: Human Computer Interaction, Computer Mediate Collaboration, Collaborative Photography, Real-Time Aesthetic Guidance

ACKNOWLEDGMENTS

I would like to thank all of the people that helped make this possible. First, I would like to express my sincere gratitude to my wonderful advisor, Dr. Michael Jones. I am so grateful he allowed me to work on this project and his continuous support and guidance throughout this process was very appreciated. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

Besides my advisor, I would like to thank Dr. Kevin Seppi for his support and insightful comments and encouragement.

I would also like to thank my fellow lab mates for their feedback, assistance, and of course friendship. It was great to work with all of you guys!

Finally, I am very grateful to my husband for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without him. Thank you.

Table of Contents

List of Figures **vii**

List of Tables **ix**

1 Introduction **1**

1.1 CMC2 1

1.2 Communicating the Intended Photo 3

1.2.1 Usage Scenario: Current Method 4

1.3 Communicating Intent With Cake Cam 6

1.3.1 Usage Scenario: Cake Cam 7

1.4 Mobile Collaboration 9

1.5 Contribution 10

2 Related Work **12**

2.1 Collocated Computer Mediated Collaborative Work 12

2.1.1 Mobile Collaborative Work 12

2.1.2 Collaborative Photography 13

2.2 Real-Time Aesthetic Guidance 14

2.3 Computer Vision 15

3 Cake Cam **16**

3.1 Cake Cam Development 17

3.1.1 User Experience and Design 17

3.1.2 Visual Inertial Odometry 20
3.1.3 Distance Calculator 22

4 Study Design **23**
4.1 Study 1: User Interface Testing 23
4.1.1 Design 23
4.1.2 Analysis 24
4.1.3 Demographics 24
4.1.4 Results 24
4.2 Study 2: Transmitting and Receiving Intent 26
4.2.1 Design 27
4.2.2 Analysis 28
4.2.3 Demographics 29
4.3 Pose Error Analysis 29
4.3.1 Translational Error 30
4.3.2 Rotational Error 30

5 Results **32**
5.1 Statistical Summary 32
5.2 Themes in Participant Responses 37
5.2.1 Worry 38
5.2.2 Communication 38
5.2.3 Resignation 41
5.2.4 Ease 42
5.3 Combined Results 42

6 Discussion and Conclusion **44**
6.1 Limitations 44
6.2 Explanation of Results 45

6.3 Future Work 46
6.3.1 Cake Cam 46
6.3.2 Generalizing 47
6.4 Conclusion 48

References **50**

List of Figures

1.1 Example of a poorly framed tourist photo. 2

1.2 *CMC2 tasks* are computer-mediated collaborative tasks between two collocated people. All other computer-mediated collaborative tasks we have classified as *CMC*, and they will not be studied in this thesis. 3

1.3 The poorly framed, crooked photo on the left was taken by a local in a hurry. The properly exposed, well framed photo on the right was taken by the tourist's dad, who could invest the time and effort in capturing the intended photo. . 5

1.4 The photograph on the left was taken by the tourist using Cake Cam. This photo represents the composition the tourist intended. The photograph on the right was taken by the local. Cake Cam guided the local into taking the same photograph set by the tourist, but the tourist is now standing in the frame. . 8

3.1 The tourist captures the intended photo by taking the picture they want to be in (frames 1 and 2). The app communicates the intended framing by guiding the local to align the red and white markers (frame 3). When the markers are aligned, the local snaps a photo of the tourist (frame 4). 17

3.2 We tested various markers to understand which ones best guided the local into taking the picture in the correct position. 18

3.3 Final 3D marker design 19

3.4 We tested overlapping the initial image and the current camera feed. The initial image is placed on top with a lower opacity allowing the local to view both images. The goal is to align the two images; however, depending on the complexity of the scene, this idea's success rate varies. 20
3.5 Visual inertial odometry combines the feature matching from the camera with the physical motion from an inertial measurement unit to create a more robust 3D position and orientation of the camera. 21

4.1 Each coordinate frame represents the relative pose error between the reference image and final image for each participant of study. 25
4.2 Comparison of images pairs between two participants with a high and low pose error, respectively. The pose error of P1 40 was below the average, whereas the pose error of P1 37 was just above the average. 26

5.1 Box plots of the first and last pose error for both Cake Cam and the da app. 34
5.2 The top graph shows the relative pose error between the reference image and final image for each participant that used Cake Cam. The bottom graph shows the relative pose error between the reference image and final image for each participant that used the normal camera app.Each coordinate frame represents the relative pose error for one participant. 35
5.3 The number of photos taken with both Cake Cam and the normal camera, charted against the participants perceived difficulty in taking them. 36
5.4 Reference photo and final photo taken with Cake Cam. This photo was captured in one try by participant 12 (P12). 37
5.5 This participant took four tries to capture the correct photo. The photograph on the left is the initial reference photograph. The four photographs on the right are each new attempt at capturing the correct photograph. 40

List of Tables

1.1 Classification of Cake Cam using the framework designed by Lundgren et al. [20] 10

4.1 The translational and rotational mean error and standard deviation for study 1. 25

5.1 Translational error for the normal camera app and Cake Cam. Cake Cam users had smaller translation errors than normal camera uses on both the first and the last picture taken. P-values represent comparison between mean error for that row and the row below. 32

5.2 Rotational error for the normal camera and Cake Cam using the scalar representation of rotational error in [16]. The difference between rotational errors was not significantly different for the last picture taken using Cake Cam compared to the normal camera. 32

5.3 Statistical summary for number of photos taken, confidence in having captured the intended photo, and difficulty of capturing the intended photo for study 2. 34

Chapter 1

Introduction

It has been said that "the best camera you have is the one that is with you" [13]. In most cases, that camera is on a cell phone. Tourists often hand their cell phones to complete strangers and ask them to take the tourists' pictures in front of a landmarks. However, the end result is often unsatisfying for the tourist. The image may include framing problems, such as cutting off the landmark or the tourist in the image. The fundamental issue is that the tourist is far more interested in the quality and composition of the image than the stranger who took the picture is. For the tourist, the picture is an important souvenir from a once-in-a-lifetime experience. For the impromptu photographer, taking the picture is, at best, a momentary distraction. This can lead to poorly framed photos of the tourist, as shown in Figure 1.1.

1.1 CMC2

The problem of asking a stranger to take one's picture in front of a landmark falls into the broad category of computer-mediated collaboration (CMC). CMC is collaboration in which a computer system provides critical connections between the people collaborating to perform some task. CMC has been studied in many contexts ranging from classic work in education [31] to disaster relief [27] and sharing photos [18]. Primarily this work has involved large groups of people [23], [21], [19], [8], [15].

In this thesis, we propose a different kind of CMC problem, in which two people collaborate to perform a single task. We call this a two-person CMC problem and abbreviate

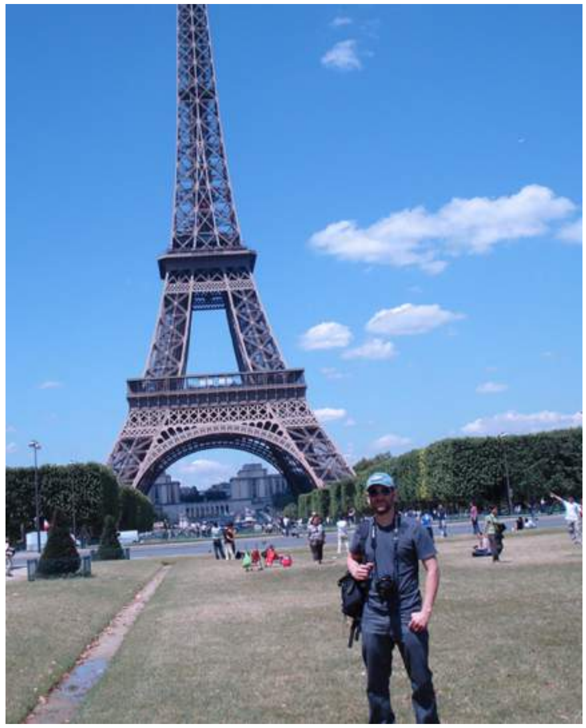

Figure 1.1: Example of a poorly framed tourist photo.

it as CMC2. We will look specifically at time-sensitive asymmetric CMC2 tasks. A time-sensitive task is a task that needs to be completed quickly. An asymmetric task is one in which one person is very concerned with the quality of the outcome and the other person is not concerned with the quality of the outcome.

Time-sensitive asymmetric CMC2 tasks are important because they occur in many formal and informal contexts: for example, negotiating with an airline gate agent or asking a stranger for directions. Completing these kinds of CMC2 tasks is important for both people involved; the person who cares more wants a good outcome (which varies with the nature of the task) and the person who cares less wants to get the collaborative task completed using minimal time and effort. Better understanding of computer-mediation in CMC2 tasks may lead to better outcomes for all involved.

Figure 1.2 illustrates these two groups of collaborative tasks. We classify CMC2 tasks as any computer-mediated collaboration between two collocated people. All other tasks, collocated groups, distributed groups, or two distributed people, we classify as CMC.

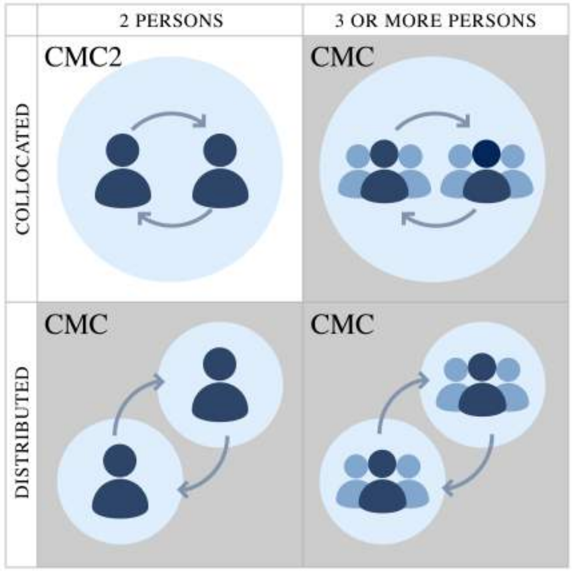

Figure 1.2: *CMC2 tasks* are computer-mediated collaborative tasks between two collocated people. All other computer-mediated collaborative tasks we have classified as *CMC*, and they will not be studied in this thesis.

We will study the role of interactive computing in time-sensitive asymmetric CMC2 tasks by looking specifically at the problem of asking a stranger to take one's picture in front of a tourist landmark.

We will refer to the person who frames and appears in the picture as the *tourist* and the person who takes the final picture as the *local*. This CMC2 problem involves a tourist asking a local to take a picture: the local is in a hurry, the tourist is more concerned about the picture's outcome, and the local is less concerned. While we will use these terms for clarity, in some cases the "tourist" might be local to the area, and, conversely, the "local" asked to take the picture might be a tourist.

1.2 Communicating the Intended Photo

Having a picture taken by a complete stranger is inherently an asymmetrical task one person is invested in the results and has the vision while the other person controls the process. The asymmetry forces collaboration between the users to perform a common task and achieve mutually satisfactory results. However, as mentioned above, this collaboration is often unsuccessful because the stranger has no way of knowing the other person's vision for the desired photo. Communicating the intended photo, specifically the tourist's desired composition, is one of the key challenges in this interaction. The local's individual photography skills can also impact the final outcome, but we view this as a secondary challenge. Because the tourist has no knowledge of the local's photography skills, it bears no effect on the tourist's desired composition of the photo.

Currently when tourists would like photographs of themselves, they simply stop locals and ask the locals to take the tourists' photos via a few quick verbal instructions. Often these scenarios play out as described below:

1.2.1 Usage Scenario: Current Method

A tourist, Teresa, is visiting San Francisco with her parents for the first time. She is excited to get her picture taken in front of the iconic Golden Gate Bridge. Teresa would like a photo with her parents and herself, with the bridge extending across the photo in the background. Teresa needs to find someone to take her picture and then describe to that person how to frame the photograph.

The composition is an important aspect of the photo, because it will lead the camera's auto-exposure system to use the settings Teresa has in mind. The exposure will be important because the sky is bright and the sun is behind the people, casting the bridge and people into shadow. Teresa has in mind a photo with the sky vastly overexposed so the people and bridge are well-lit. She knows that for her camera, the auto-exposure system will get the exposure she wants if the photo is framed with the people and bridge occupying more of the frame than the sky.

Teresa stops Larry, a local to San Francisco, and asks him to take her picture. Larry is about to start a jog across the bridge and is in a hurry to get started. However, he is willing to quickly snap a photo for a tourist.

Figure 1.3: The poorly framed, crooked photo on the left was taken by a local in a hurry. The properly exposed, well framed photo on the right was taken by the tourist's dad, who could invest the time and effort in capturing the intended photo.

Teresa quickly describes the photo she wants to Larry and then moves into place next to her parents. After Larry snaps the picture, he hurries on his way. Larry's picture is shown on the left side of Figure 1.3[1] Teresa sees that, while the picture on her phone does have the Golden Gate Bridge extending into the background, it is not as prominent as she would have liked. At first glance it would be easy to miss the landmark completely. Additionally the photo is very underexposed because the composition did not activate the auto-exposure system as desired. The white sky is the dominant feature in the photo, which coloration caused the camera's automatic exposure settings to adjust for it and underexpose the rest of the image and left Teresa and her parents in shadow.

[1]Images of Teresa and her family are used with permission. This scenario is based on the experience of a friend of the author, but names have been changed.

Teresa is not from California, and being here at the Golden Gate Bridge is a significant event that she would like to remember with a good photo. She would like another photo taken, but does not think she can describe exactly what she is looking for to another stranger. Teresa decides to simply have her dad take the photo of her and her mom.

The picture taken by Teresa's dad is shown on the right side of figure 1.3. The composition of this photo is as Teresa desired, with the bridge extending across the photograph and into the background. Overall the image is properly exposed, as the stark white sky does not dominate the image. Teresa's dad was more willing to invest the effort required to understand and capture the intended photograph, using the common method of verbally describing the exact framing desired.

From Teresa, the tourist's, perspective, the transmission of intent is necessary to capture the picture he or she wants. The tourist is willing to put in the effort required to communicate intent, and consequently capture the desired photograph. From the perspective of Larry, the local, the reception of intent is necessary to capture the picture the tourist wants. However, the local is not concerned with the quality of the outcome and may be in hurry, so the local needs to quickly receive intent. Unfortunately, as Teresa's usage scenario outlines, there is currently no convenient way to communicate intent in a quick and precise manner that satisfies the needs of both the local and the tourist.

1.3 Communicating Intent With Cake Cam

In this thesis we explore CMC2 tasks using a set of studies design to investigate how strangers collaborate in time-sensitive asymmetric CMC2 collaboration by mediating the interaction. In order to carry out these studies we designed and implemented a new mobile interaction that helps invested party quickly communicate his or her intent to the uninvested party. To do this we developed an app, titled Cake Cam, which lets the tourist take his or her photo and be in it too. The name is a play on the English proverb "you can't have your cake and eat it too," implying that a person cannot have the best of both worlds i.e., that he or she cannot both control the framing of the photo and be in it as well. Cake Cam, however, removes this paradox and allows the user to experience the best of both worlds when it comes to having his or her photo taken.

To use the app, a tourist takes a photo of the carefully framed scene. The tourist then hands the phone to a local, asks the local to take their picture, and moves into the frame. Alignment markers guide the local to correctly frame the picture with the tourist in it, producing the intended photo. (See figure 3.1). The following scenario depicts another tourist-local interaction, this time mediated by Cake Cam.

1.3.1 Usage Scenario: Cake Cam

A tourist and university student from the United States, Taylor, is visiting Ecuador for the first time in her life. It may be the only time she will get to visit Ecuador. She is excited to get her picture taken in front of the Monument to the Equator at Mitad del Mundo. Taylor has in mind a particular framing of the picture, which will capture her unique experience. Taylor needs to find someone to take her picture and then convey to that person how to frame the picture.

Taylor opens Cake Cam and composes the picture she would like to be in. She imagines herself in a picture shot from a low angle with with the full monument behind her. She frames the picture so that the full monument will be within the frame.

Taylor then stops Luis, a local Ecuadorian, and asks him to take her picture. Luis is ending his lunch break and is in a hurry to get back to his office. However, Luis is willing to quickly snap a photo for a tourist.

Taylor tells Luis to align the markers and snap the picture. Luis quickly figures out that moving and rotating the phone moves the markers. Luis moves the phone down and angles the phone up to capture Taylor's intended composition.

After Luis snaps the picture, he hurries on his way. Taylor sees that the picture on the phone is a close match to the one she intended to capture, and both parties are satisfied.

Figure 1.4: The photograph on the left was taken by the tourist using Cake Cam. This photo represents the composition the tourist intended. The photograph on the right was taken by the local. Cake Cam guided the local into taking the same photograph set by the tourist, but the tourist is now standing in the frame.

Figure 1.4[2] illustrates this scenario. The photograph on the left was the initial framing photograph taken by the tourist, representing the tourist's intended composition of the photo. The photograph on the right is the photo captured by the local using the app Cake Cam. The two photos are nearly identical in composition with the major difference being that the tourist is now standing in the frame. By simply aligning the markers, the local was able to replicate and capture the unique angle the tourist wanted, without needing further instruction than to align the markers. The app thus communicated intent and successfully mediated this interaction.

[2]Images of Taylor are used with permission. This scenario is based on the experience of a friend of the author, but names have been changed.

In comparison to the current method usage scenario described in the first scenario, the app communicated the intended framing faster and more accurately than a verbal description could have. The tourist was able to get the unique photo they were looking for without taking up as much of the local's time.

1.4 Mobile Collaboration

Others have studied the usage of mobile apps in collocated collaboration ([4], [8], [18]). Lundgren et al. [20], for example, developed a design framework that guides the creation of new mobile experiences for collocated interaction, and analyzes existing interactions. Their framework provides four relational perspectives for designing collocated mobile interaction: the social situation in which it takes place, the technology used, the physical environment, and the temporal elements of design. We used this framework to classify our app.

The app Cake Cam was created to communicate intent when a tourist and a local collaborate to take a picture at a tourist location i.e. the FRAMING, the main social situation where the activity takes place, is public and the FOCUS of the app is communication. The tourist has a specific image in mind while the local person does not know what photograph the tourist intends making both INFORMATION SYMMETRY and INTERACTION ABILITY asymmetrical. As the process of taking a photo progresses the INFORMATION DISTRIBUTION unfolds. Moving the phone closer to the intended location results in a proximity-based EVENT TRIGGER that informs the local to take the photograph. The full classification can be found in table 1.1.

While this framework is helpful in comparing collocated collaborative apps, it does not account for key elements typical of asymmetric CMC2 tasks. Motivation is an essential part of these CMC2 tasks and is not addressed in this framework; so too is duration, which has a big impact on the type of interaction, unaddressed. Additionally granularity of the location is not accounted for. With Cake Cam, the margin of error for the device location must be

Social	
FOCUS: COORDINATION OF ACTION: FRAMING:	communication combined public
Technological	
INFORMATION SYMMETRY: INTERACTION ABILITY: INFORMATION DISTRIBUTION: EVENT TRIGGERS:	asymmetrical asymmetrical unfolding proximity-based
Spatial	
PROXIMITY: LOCATION: MOVEMENT:	people none on the go
Temporal	
SYNCHRONIZATION: ENGAGEMENT: PACING:	user-driven combined user-paced

Table 1.1: Classification of Cake Cam using the framework designed by Lundgren et al. [20]

within centimeters of the desired location to achieve any degree of satisfaction, whereas other mobile collaboration tasks often have a much larger granularity.

Of note, Lundgren et al. outline how the framework was used to redesign an app, Automics [8], to encourage more collaboration. The significant difference in the resulting app, InstaCampus [10], was changing the INFORMATION SYMMETRY from symmetrical to asymmetrical. Lundgren found that this change encouraged more collaboration between participants. Other research [9], [28] has also found that introducing asymmetry encourages collaboration. Information asymmetry can lessen the cognitive load, enabling participants to complete the task faster. It also forces collaboration when sharing information is critical. In our task INFORMATION SYMMETRY and INTERACTION ABILITY is naturally asymmetrical. Cake Cam was created to facilitate this naturally asymmetric collaboration.

1.5 Contribution

This thesis explores the hypothesis that an augmented reality mobile app can mediate completion of time-sensitive asymmetric CMC2 tasks by communicating a tourist's intent when a local and tourist collaborate to take picture in front of a landmark. Our conjecture is that interactive computing can quickly communicate intent, which may help mediate time-sensitive asymmetric CMC2 tasks. In the case of a tourist handing his or her cell phone to a local to take a picture, the cell phone might guide the local to take the tourist's intended picture. This guidance can be given more quickly and with greater accuracy than giving a verbal description of the intended framing of the picture can.

We built a mobile app called Cake Cam as a means for communicating intent in picture-taking collaboration. Cake Cam uses augmented reality markers to guide the local into taking the photo as intended by the tourist. We first validated the usability of Cake Cam with a user study and found that participants were able to replicate the intended photo by moving the phone within $9cm$ of the initial location with a rotational error of 5.8 degrees. Finally we validated the claim that the mobile app can mediate the interaction between a tourist and local by communicating intent. We found that the app Cake Cam was better at guiding a user into taking the photo a tourist intended, and that participants found the process of capturing the correct photo to be easier with Cake Cam than with the current verbal method.

Based on evaluation of these users' studies, we discuss future work and possible implication for other asymmetric CMC2 problems. Specifically we believe that communicating abstract representation of intent is more effective than communicating exact intent.

Chapter 2

Related Work

This work builds on results from four research communities: computer-mediated collaboration, collaborative photography, real-time aesthetic guidance, and computer vision. The following sections outline the techniques and insights that have contributed to this work.

2.1 Collocated Computer Mediated Collaborative Work

The area of collocated collaborative work is a well-studied field within the HCI and CSCW communities; however, much of the research has focused on collocated groups of people rather than two individuals, as presented in this thesis. Work within the field has also not yet explored the unique asynchronous aspect presented in this thesis.

2.1.1 Mobile Collaborative Work

Lundgren et al. [20] developed a design framework that guides the creation of new mobile experiences for collocated interaction, and analyzed existing interactions. Their framework provides a way to classify and compare collocated collaborative mobile apps. We used this framework to classify Cake Cam in chapter 1; however, this framework does not classify some characteristics of collocated collaboration: duration, motivation, and granularity, which are all particularly important for CMC2 tasks.

Various examples of technical systems have been developed and researched for collocated computer mediated collaborations: Lucero et al. [18] developed a phone-based application that allows a small group of collocated people to share photos using the metaphor of passing paper photos around. They found that people were willing to share their devices to support collaborative interactions. MobiPhos [6] is a photo-sharing application wherein users need to be collocated to share and explore each other's photo collections. MobiComics [19] introduced an application that allows a group of collocated people to create comic strips using their mobile phones and share to public displays. The mobile app TalkBetter [11] was designed to reinforce everyday parent-child conversation specifically for children with language delay. Mobile Stories [9] is an app that allows young children to collaboratively read and create stories. Fails et al. found collaboration increased when they implemented two co-design methods: content splitting and space sharing. Content splitting is simply dividing content between each device, which is synonymous with role assignment. This idea is similar to asymmetric INTERACTION ABILITY which, as discussed above, Lundgren et al. also introduced to increase collaboration in Automics [8].

Common to all of these works is the matter that the mobile application was designed to facilitate group collaboration, often with multiple devices. In contrast, our work is focused on collaboration between two individuals with only one mobile device.

2.1.2 Collaborative Photography

Existing research has explored various aspects of collaborative photography; however, prior work in this area does not address the problem of communicating a photo's intended specifications with a stranger who then takes the picture. Cake Cam is a platform for exploring communication of intent in this form of interaction.

In InstaCampus, Fischer et al. [10] researched collaborative photography and notification management in small groups. In another project Jarusriboonchai et al. [12] compared traditional photography methods with an asymmetrical collaborative method that supported the social collaborative experience of taking photos. Kim et al. later presented LetsPic [14], a group-based photoware, as a way to support group awareness for in-situation collaborative photography over a large physical space.

Other research has explored specifically tourist-centered photo collaboration. Brown et al. [4] studied photo collaboration between participants who were separated by distance with a mobile app that allowed tourists to instantly share pictures with their remote friends and family. The instant photo sharing in the system allowed immediate remote feedback that led the friends and family to request particular photos, rejecting taken photos and asking for new ones. Another project, a mobile application called Yousies [29], allowed for its user to get his or her photo taken by a stranger, another user, without needing to pass the device around. The photo would be taken with the stranger's camera and then be instantly shared with the tourist.

2.2 Real-Time Aesthetic Guidance

Previous to Cake Cam, work has been done on real-time aesthetic guidance of photography and video. However, this body of work focuses on using photography principles to guide the framing of the photo, rather than following the personal preferences of the photographer. In contrast, Cake Cam takes no position on aesthetics but enables tourists to capture pictures of themselves based on their own preferences.

McAdam et al. [22] researched guidance for low-level features like exposure, luminance, and motion blur. Brewster et al. [3] developed a system that guides subject positioning within a landscape photograph, using a visualized rectangle. However, this system was not implemented on a smartphone or tested for interactive use. NudgeCam [5] is a smartphone app that provides real-time feedback based on standard media capture heuristics to encourage higher-quality video. It places text and colored boxes around the face on the image to indicate problems with the video feed, including face size, face location, scene brightness, tilt, and stability. Li et al. [17] developed a system that interactively guides the user into taking a better self-portrait (selfie). Directional arrows guide the user into taking a better photo based on empirical models of three parameterized composition principles: face size, face position, and lighting direction. Xu et al. [30] developed a photo-taking interface using a three-camera array that provides real-time feedback on how to position the subject of interest according to the rule of thirds.

2.3 Computer Vision

Within the field of computer vision, Davison et al. [7] proposed MonoSLAM, one of the first real-time 3D monocular localization and mapping frameworks. Since then, many improvements have been contributed from various research groups. Specifically in mobile, Li et al. [17] implemented a monocular visual inertial state estimation for robust camera localization on a smartphone for mobile augmented reality. Shelley [26] implemented visual inertial odometry (VIO) on a smartphone. VIO combines feature tracking with data from an inertial measurement unit (IMU) to create a robust estimate of the camera's pose in 3D space. Recently, further research has been done to improve the accuracy of VIO. ROVIO (Robust visual inertial odometry) [2] uses pixel intensity errors of image patches to achieve accurate tracking results. Bloesch et al. [1] uses an iterated extended Kalman filter to further increase the accuracy.

Chapter 3

Cake Cam

The mobile app Cake Cam was developed to mediate the interaction between two parties. In this particular CMC2 problem, computing (Cake Cam) is used to quickly communicate one party's intended picture framing to an impromptu photographer who will take the photo.

To perform the interaction, the tourist follows these four steps summarized in figure 3.1:

1. The tourist takes and approves a photo of the carefully framed scene.
2. The tourist then hands the phone to a local and asks the local to take his or her picture.
3. The tourist moves into the frame.
4. Alignment markers guide the local to correctly frame the picture with the tourist in it, producing the intended photo.

Step one of this process will be familiar to anyone that has taken a photo with a mobile camera. We intend step two to cause no changes in how the tourist usually asks a local to take their photo. Step four introduces a new interaction using augmented reality. This is the core functionality of the app. The alignment markers are used to communicate the tourist's intended photo composition to the local. For the app to be effective, this process needs to be quick and easy to understand. The user experience and underlying computer vision algorithms used to align the markers are discussed in the next subsections.

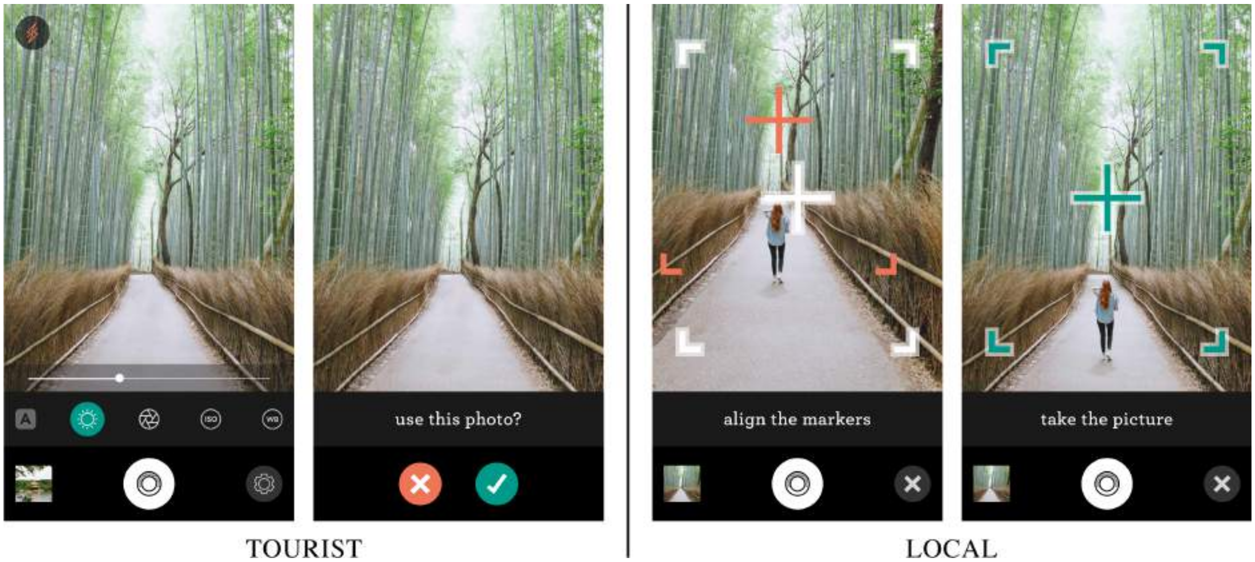

Figure 3.1: The tourist captures the intended photo by taking the picture they want to be in (frames 1 and 2). The app communicates the intended framing by guiding the local to align the red and white markers (frame 3). When the markers are aligned, the local snaps a photo of the tourist (frame 4).

3.1 Cake Cam Development

Development of the app can be broken down into three main components: user experience and design, visual inertial odometry, and distance calculator. The details of each will be discussed in the following subsections.

3.1.1 User Experience and Design

Cake Cam introduces a new interaction in mobile photography. Because of this new component, the user experience is a critical part of the app development and a key part of validating our thesis. Because the problem is time-sensitive for the local, the interaction needs to be quick, clear, and language-agnostic, enabling the stranger to use the app with no explanation.

The challenge is making the experience intuitive for the local, who is likely in a hurry and not concerned about the quality of the photograph. This led us to create a paper prototype with transparent "screens" showing different arrangements of arrows, circles, and crosshairs to guide the local to the intended framing 3.2. We then conducted user testing

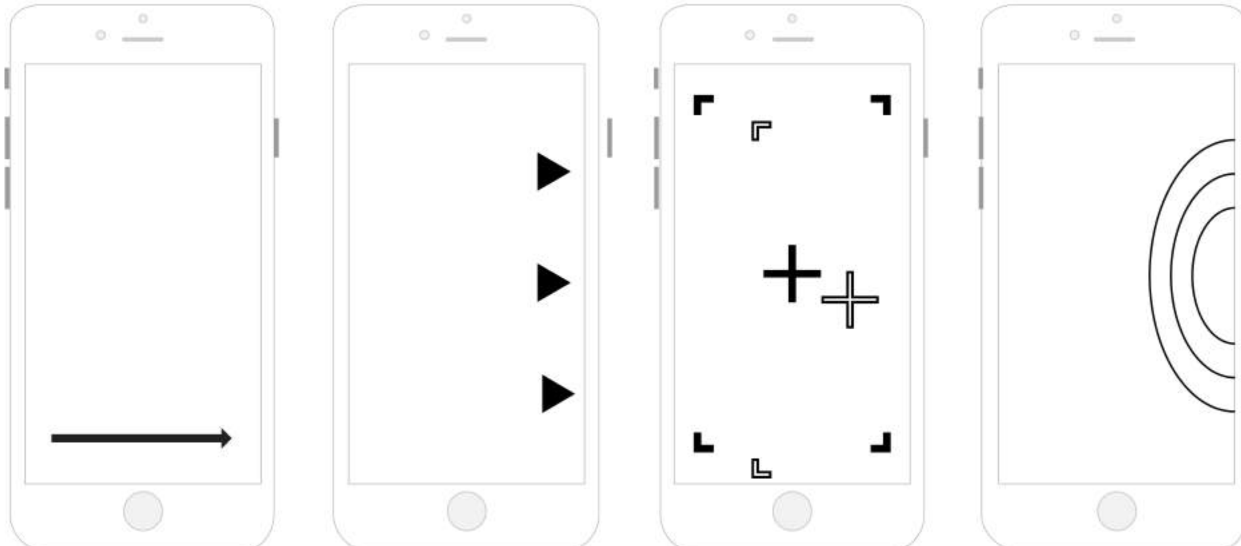

Figure 3.2: We tested various markers to understand which ones best guided the local into taking the picture in the correct position.

with a small group of people to determine which markers best guided people into taking the picture in the correct position.

We found that a crosshair in the center with corner markers, as shown in figure 3.1, was the most intuitive setup. Users instinctively aligned the two sets of crosshair markers during testing. Although the users in our testing instinctively aligned the markers, we later added the instruction align the markers at the bottom of the screen, for further clarification.

Additionally, we found that making the crosshair markers 3D better guided the local to the correct camera orientation. We also added an error bound around the camera's position and orientation so that the user did not get caught up trying to perfectly align the markers. The final marker design can be seen in figure 3.3.

We also tested a design that simply overlaid the first image taken by the tourist and the current camera feed. The first image would have a lower opacity, allowing the local to view both images. The idea was that the local could then simply align the two images and take the photo. However, as shown in figure 3.4, overlapping the two images creates a busy image with visual clutter. The image on the left is a more complicated image with similar colors throughout the image, this type of scenery would be extremely difficult for a

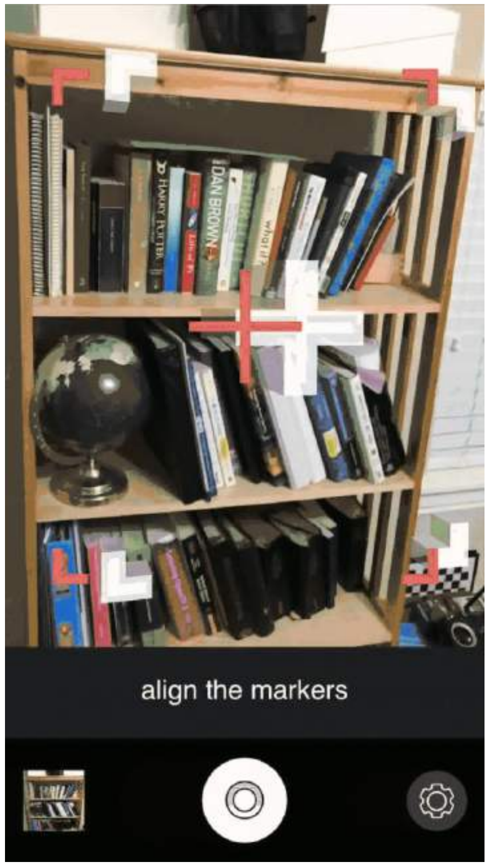

Figure 3.3: Final 3D marker design

local to align as there are no distinct features to match. However, even when the image is simpler with distinct features, as in the image on the right, it still is not quickly obvious what direction the local needs to move the camera to align the two images.

We found that while this idea works for very simple images with one prominent feature, overall it was too hard for participants to understand the concept and align the two images. One resulting observation is that while this method provides all the information, i.e., the exact photo the tourist wants, it was ultimately too complicated for the local to quickly interpret and match the image.

The rest of the app interaction was designed to work similarly to the standard iOS camera app. This gives the users a sense of familiarity while introducing a new experience.

Figure 3.4: We tested overlapping the initial image and the current camera feed. The initial image is placed on top with a lower opacity allowing the local to view both images. The goal is to align the two images; however, depending on the complexity of the scene, this idea's success rate varies.

3.1.2 Visual Inertial Odometry

The key technical challenge in systems such as Cake Cam's is detecting the camera pose. A camera pose is the combination of its position and orientation in world coordinates. The original camera pose is important because it marks the location where the user intended the camera to be positioned and oriented as the photo is taken. Given the camera's original pose, we also need to know the new camera pose in order to compute the difference between the two. We use this pose difference to guide the local into the intended camera position.

We need to know the current camera pose in real time because this allows us to give visual feedback to the local photographer and guide him or her into the correct camera pose to take the desired photo. Given the current camera pose and the initial camera pose, we

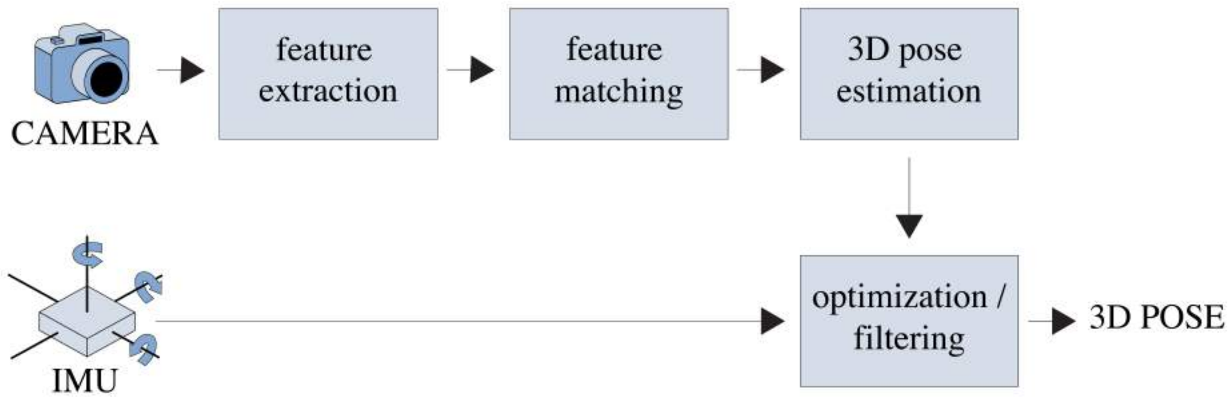

Figure 3.5: Visual inertial odometry combines the feature matching from the camera with the physical motion from an inertial measurement unit to create a more robust 3D position and orientation of the camera.

can compute the error between the two poses. This allows us to provide immediate feedback about the direction the local photographer needs to move the camera.

The operational mechanism for obtaining the camera pose in real time is visual inertial odometry (VIO). As shown in figure 3.5, this algorithm uses a monocular camera coupled with linear acceleration and angular velocity from an inertial measurement unit (IMU) to create a robust estimate of the camera's pose in 3D space. This is done by tracking feature points between image frames. Using the feature points, the algorithm calculates the essential matrix, which is a 3x3 matrix that relates images from two perspectives, and then decomposes it to obtain the relative poses between image pairs. VIO gives a robust estimate of the camera's pose due to the combination of the monocular camera and the IMU data. Alone, a monocular camera can estimate only relative position without absolute scale, which is required to know the actual size or distance in meters. The IMU sensors provide absolute scale; however, the sensors are prone to substantial drifts in position estimates compared with the camera's estimates. VIO successfully blends together the two measurement methods to achieve sensor fusion and improve the robustness of the camera pose estimation.

Smartphones are ideal platforms for VIO because the camera and IMU are tightly integrated into a single device. However, the algorithm performs poorly in low light, against moving backgrounds, or by blank walls, because the algorithm cannot find and match feature points. Furthermore, the algorithm fails to account for sudden large movements.

When the user finishes framing his or her picture, the VIO algorithm is initialized and the current camera position and orientation are marked as the initial pose. As the smartphone is handed to a local, the error between the current pose of the phone and the initial pose increases. Visible alignment markers guide the local to minimize pose error and correctly frame the picture. Because this interaction must be responsive, we use optimized VIO routines provided by Apple's iOS AR Kit.

3.1.3 Distance Calculator

When the local moves the phone to the correct location, the alignment markers turn green indicating the photo should be taken. This is done by calculating the distance between the initial camera pose and the current camera pose. Cake Cam uses a combination of the difference in Euler angles and the difference in the x, y, z axes to calculate the error between the initial camera pose and the current camera pose. The margin of acceptable error varies in the x, y, and z axes: moving the camera up, down, left, and right has a more dramatic effect on the framing of the photo, compared to simply moving the camera forward and backward. Specific limits are set in the x, y, and z directions with the x and y directions having a tighter error threshold than z.

Additionally when the translational error is greater, it becomes more critical that the orientation of the phone is closer to the initial camera orientation. As the phone is moved closer to the intended position there is more room for error and the threshold for roll ϕ, pitch θ, and yaw ψ increases. Roll ϕ has a tighter error threshold because this too has a more dramatic effect on the framing of the photo. Pitch θ and yaw ψ have the same error threshold.

Chapter 4

Study Design

We validated our thesis with two separate user studies. The first study was designed to test the usability of the app's interface. The second study explored how much intent was communicated by the app, and the experience of receiving intent in this manner.

These studies attempt to answer the question, "Can technology be used to communicate intent between two people, specifically in the context of having a photo taken at a tourist landmark?" In the process of addressing this question, we also hope to provide insight into the general experience of communicating intent in the stated manner.

4.1 Study 1: User Interface Testing

Study 1 was designed to test Cake Cam's user interface. This study was used to establish that the app's user interface can effectively guide participants into reproducing a photograph. By first establishing that the designed interface is effective, we can focus Study 2 solely on how intent is communicated with and without the app and not on whether the algorithm and interface are successful in guiding the participant.

4.1.1 Design

The research team member took the initial framing photograph and then passed the camera to the participant. The participant was asked to align the markers and take a photograph once they felt the markers were aligned. The participant was not taking a photograph of a person so the focus was on aligning the markers. This concluded the study.

4.1.2 Analysis

The app stored the 3D pose of the camera for each photograph. This data was then used calculate the pose error, which was broken up into the translational error and rotational error, between the two images. The mean rotational and translation error between all participants was calculated as described in section 4.3.This information gave us an average error to expect in later studies. Had the error range been too great, the range would have indicated that the interface was not accurately communicating intent and would need to be refined.

4.1.3 Demographics

The study was run with 40 participants. Of the participants, 21 were female and 19 were male, with a mean age of 21.25 years. The sampling method was a convenience sample, wherein the research team stopped random people on a university campus: BYU, specifically.

4.1.4 Results

This study was designed to test the effectiveness of the user interface. We measured the effectiveness by calculating the pose error between the image taken by the research team member and the image taken by the participant. The pose error was calculated as described in section 4.3. In our analysis of study 1, we found that participants were able to use Cake Cam to replicate a photo with an average translational error of $14cm$ and the mean rotational error was 0.017 or 3.54 degrees. Additionally participants were able to complete the study in average of 37 seconds from the time the initial reference photo was taken to when the research team member was handed the phone back. This includes the time to give the short instruction, "align the markers and take a photo." Table 4.1 summarizes the results from the rotational and translational error for study 1.

We graphed the relative pose error between the reference image and final image for each participant in figure 4.1. Each coordinate frame in the graph represents one participant.

	Mean	Std. Dev
Translational Error	14.5	7.30
Rotational Error	3.54	1.20

Table 4.1: The translational and rotational mean error and standard deviation for study 1.

As illustrated by the figure, the majority of coordinate frames are clustered around the origin with a few outliers.

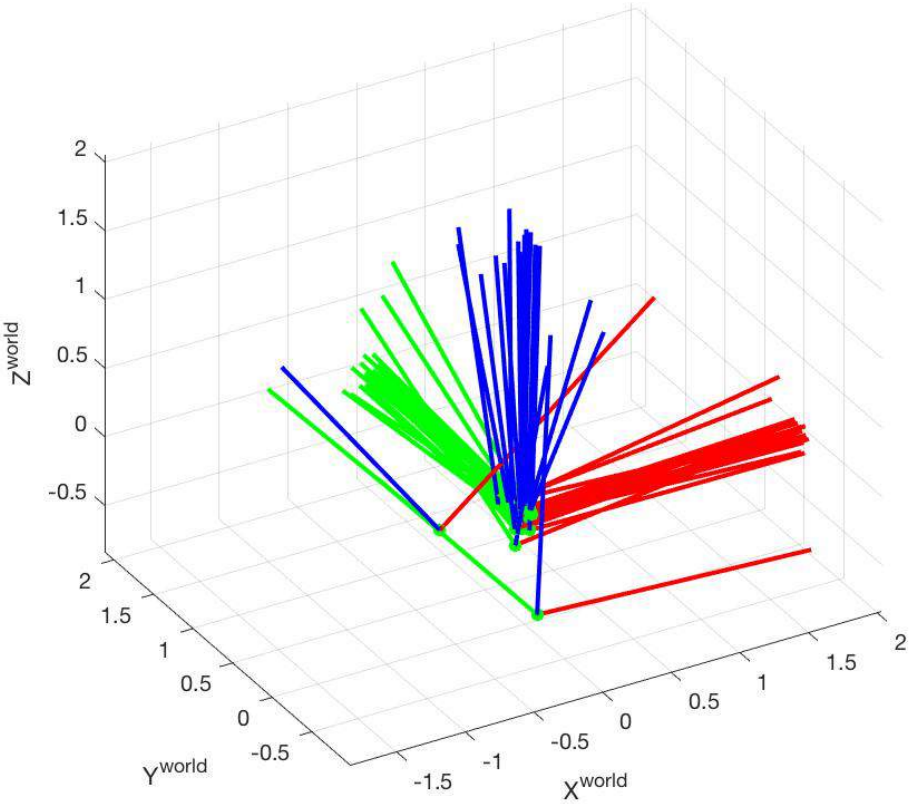

Figure 4.1: Each coordinate frame represents the relative pose error between the reference image and final image for each participant of study.

To better represent what this pose error means for the composition of the photograph, figure 4.2 shows a sample of image pairs with a high and low pose error, respectively. The rotational and translation error of participant 40 (S40) was lower then the overall mean error, while the rotational and translational error of participant 37 (S37) was very close to the mean pose error. It can be seen that the two image pairs are very similar. The main difference is that the photo captured by S37 is slightly lower than intended, which difference caused greater translational and rotational error. Overall, however, both images are very close to the intended photograph.

Figure 4.2: Comparison of images pairs between two participants with a high and low pose error, respectively. The pose error of P1 40 was below the average, whereas the pose error of P1 37 was just above the average.

From these results we have concluded that the user interface of Cake Cam is effective in guiding photo replication by positioninfg the phone in the correct location for the user. These results enabled us to continue with study 2 feeling confident that a poor user interface would not compromise the results.

4.2 Study 2: Transmitting and Receiving Intent

The second study explored how intent was communicated to the local photographer using a computer. This mixed-method study was designed to answer the following questions:

1. Does the app communicate intent more quickly to the local compared to using a phone without the app to take a picture?
2. What is the local's experience while receiving intent through a computer?

Answering these two questions was a crucial part of understanding the app's limitations and the user experience from the local's perspective.

4.2.1 Design

In this study, the participants took on the role of the local, and a member of the research team acted as the tourist. Half of the people used the Cake Cam app, and the other half used a camera app that was designed to be similar to the standard iOS camera app. We will refer to this app as the normal camera app. The normal camera app works exactly like the standard iOS camera app, but it also stores the pose data. We went with a between-subjects design to prevent bias caused by the learning effect. A learning effect is caused when participants perform better under the second condition because of prior practice or knowledge they gained from the first condition. With this study, if participants were asked to take a second photo using a second method, they would already have a better idea of the type of photo we were trying to capture, which knowledge would skew the results

Each participant was asked to take a picture of the research team member in front of the Centennial Carillon Bell Tower on BYU campus. The research team member was looking for a specific photo with requirements on where the subject should be placed and how the photo should be framed. Both apps also displayed the translational and rotational error for the research team member to view and was used to determine when the photo was close enough to the intended photo.

After the first photo was taken, the participant was asked question 1, of the following list. The research team member then looked at the photo. If the photo met the requirementsa translational error of < 0.26 and a rotational error of < 0.02, the participant was asked questions 24, and the study concluded. We choose a threshold of 0.26 and 0.02 as it was the maximum error that did not cut of specific elements, the bell tower and trees, in the photo. If the photo did not meet requirements, the participant was given one additional piece of instruction to improve the photo and was asked to take the photo again. The instruction came from a list of possible instructions to give to the participant such as, "place the bell tower on the right of the photo." This process was repeated until the translational and rotational error was within the specified range and the correct photo was achieved. The participant was then asked questions 23, and the study was finished.

1. On a scale of 1 to 10, how confident are you that this is the photo I wanted?
2. On a scale of 1 to 10, how difficult was it to get the photo I wanted?
3. Please describe your experience.

4.2.2 Analysis

From the questionnaire, we used question 1 to understand how much intent is initially communicated to the local with and without the app. Question 2 helped us understand whether participants preferred capturing the intended photo with the Cake Cam app or the normal camera app. Furthermore, it illuminated the difficulty levels participants experienced in this communication with and without the app.

We chose to use an unlabeled 10-point scale in questions 1 and 2 rather than 5-point Likert items because, in preliminary trials of this protocol, we found that people compressed their rating to the top end of a 1 to 5 scale. We anticipated that most participants would answer in the range of 5 to 10, so having more values between 5 and 10 would help us better differentiate the participants' experiences. This is consistent with the distribution of scores on the Software Usability Scale (SUS) test which also skews high [24].

From the information gathered from question 3, we compared the average number of photographs taken before getting the desired photo for both Cake Cam and the iPhone camera app. We anticipated that fewer photographs would be taken using Cake Cam, because Cake Cam more clearly communicates intent than a regular cell phone camera.

We used a two-tailed t-test to determine if the responses to the Likert items and were statistically significant. Additionally, we used Spearman's Rho Test for correlation between the number of photos taken and the perceived difficulty by the participant.

We then used thematic analysis on answers to the open-ended question (question 4) as a means to gain insight and knowledge from data gathered. Our thematic analysis consisted of two members of the study team reading and re-reading all participant responses three times and then meeting to discuss themes that emerged from the comments. The goal was to identify a set of themes that were interesting and relevant, and also described most of the responses. This analysis helped us gain insight into the user experience from the local's perspective. An open-ended thematic analysis was appropriate because this was a new kind of user experience and we did not know what themes to evaluate. Future studies may confirm or refute themes identified in this part of the study.

4.2.3 Demographics

We used a sample size of 40 participants, with 20 using each app. The participant group size was determined using a power analysis with a critical difference of 70% between mean rating scores using confidence intervals that assumed the direct comparison of proportions using the standard Wald version of the confidence interval formula [25].

Of the participants, 22 were female and 18 were male. The average age was 21 years old. The sampling method was a convenience sample, and the research team randomly stopped people who were walking near the bell tower on BYU campus.

4.3 Pose Error Analysis

To better elaborate on how much closer Cake Cam was able to direct the user into replicating the first photo, the app stored the 3D pose of the camera for each photograph. The pose of a rigid object in 3D space is parameterized by 6 variables: 3 for position and 3 for rotation, and is given by this formula:

$$\mathbf{x} = \begin{bmatrix} \mathbf{p} & \boldsymbol{\Theta} \end{bmatrix}^\top = \begin{bmatrix} x & y & z & \phi & \theta & \psi \end{bmatrix}^\top . \tag{4.1}$$

This data was then used to calculate the relative pose error between the two images. Although the origin of the camera coordinate frame is the location where the app was first opened, we are only interested in relative pose error. The mean rotational and translational error between all participants was calculated therefore.

4.3.1 Translational Error

Once the reference photo was taken in the app, an initial pose $\mathbf{x_i}$ was stored. Upon taking the final picture, the final pose x_f was used to calculate the Euclidean translational error:

$$\mathbf{t_e} = ||\mathbf{p_f} - \mathbf{p_i}|| = \sqrt{(x_f - x_i)^2 + (y_f - y_i)^2 + (z_f - z_i)^2}. \tag{4.2}$$

4.3.2 Rotational Error

To properly account for the geometry of rotations, Euclidean metrics (as used for translational error) cannot be used, because rotational space is not defined in Euclidean space. Instead, we compute rotation matrices from the captured Euler angles $\mathbf{\Theta_i}$ and $\mathbf{\Theta_f}$. We use the 3-2-1 Euler angle sequence [2], [26] and calculate the orientation matrix as follows:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\phi) & sin(\phi) \\ 0 & -sin(\phi) & cos(\phi) \end{bmatrix} \begin{bmatrix} cos(\theta) & 0 & -sin(\theta) \\ 0 & 1 & 0 \\ sin(\theta) & 0 & cos(\theta) \end{bmatrix} \begin{bmatrix} cos(\psi) & sin(\psi) & 0 \\ -sin(\psi) & cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.3}$$

$$= \begin{bmatrix} c_\theta c_\psi & c_\theta s_\psi & -s_\theta \\ s_\phi s_\theta c_\psi - c_\phi s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & s_\phi c_\theta \\ c_\phi s_\theta c_\psi + s_\phi s_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi & c_\phi c_\theta \end{bmatrix}, \tag{4.4}$$

Where $c_* \triangleq cos(*)$ and $s_* \triangleq sin(*)$. A property of rotation matrices are that their inverse is the same as their transpose. Therefore, given any rotation matrix R, we have the identity

$$RR^\top = I. \tag{4.5}$$

This gives us a computationally efficient and geometrically sensible way to compare two rotationsif R_1 and R_2 are the same rotation, the product $R_1R_2^\top$ will be the identity matrix I. Although there are 9 values associated with this orientation comparison, we can encode rotational error into a single number by using the following metric [16]:

$$R_e = \frac{1}{2} tr\left(I - R_i R_f^T\right), \tag{4.6}$$

Where the matrix trace $tr(\cdot)$ is defined as the sum of the main matrix diagonal. The values that this error metric can have range continuously from 0 to 2, where $R_e = 0$ represents no rotation error and $R_e = 2$ represents 180° error about the principal rotation axisthe axis about which all rotation occurs. We used this calculation as it is simple and accurately measures what matters for this problem comparing two rotations.

Chapter 5

Results

We found that with Cake Cam, the correct photo was taken in less time and the final image was overall closer to the intended photo than with the normal camera app. Through this study we also learned more about receiving intent through an app. Further discussion of these results can be found in the following sections.

5.1 Statistical Summary

Translational Error				
	Mean (cm)	Std. Dev	*p*-value	*t*-value
First: Cake Cam	9.2	6.0	¡ 0.00001	5.6
First: Normal	115	81		
Last: Cake Cam	9.2	6.0	0.00083	3.6
Last: Normal	17	6.4		

Table 5.1: Translational error for the normal camera app and Cake Cam. Cake Cam users had smaller translation errors than normal camera uses on both the first and the last picture taken. *P*-values represent comparison between mean error for that row and the row below.

Rotational Error				
	Mean (deg)	Std. Dev	*p*-value	*t*-value
First: Cake Cam	0.0050	0.0050	0.0024	-3.3
First: Normal	0.021	0.020		
Last: Cake Cam	0.0042	0.0040	0.79	-0.27
Last: Normal	0.0039	0.0039		

Table 5.2: Rotational error for the normal camera and Cake Cam using the scalar representation of rotational error in [16]. The difference between rotational errors was not significantly different for the last picture taken using Cake Cam compared to the normal camera.

The translational pose error between the first photo taken with Cake Cam and the first photo taken with the normal camera were statistically significant ($p < 0.00001$). As expected, after the first photo was taken, the difference between the pose error of Cake Cam and the normal camera was quite large. The mean translation error for Cake Cam was $9.2cm$, while the mean translational error for the normal camera was $115cm$. The mean rotational error for Cake Cam was .0050 where $2 = 180$ degrees, while the mean rotational error for the normal camera was 0.021. With guidance, the final pose error of the normal camera decreased with a translational error of $17cm$ and a rotational error of 0.0039. These results are closer to the results achieved by Cake Cam. However, even with verbal guidance, the normal camera app was still not as accurate as Cake Cam was in achieving the desired photo. Furthermore, there is still a statistically significant difference between the two translational errors ($p = 0.00083$).

These results are summarized in table 5.1. However, while the mean rotational error was smaller with Cake Cam, there was no statistical difference between the rotational pose error between Cake Cam and the normal camera app. We theorize that this correspondence may be due to the standard eye-level orientation we choose and is a limitation on our conclusions discussed further in chapter 6. These results can be seen in table 5.2. The shape of the distribution can be seen in the box plots provided in Figure 5.1.

In figure 5.2 we graphed the relative pose error between the reference image and the final image for the first photograph taken by each participant. Each coordinate frame represents the relative pose error for one participant. The graph on the top shows the relative pose error for Cake Cam. The graph on the bottom shows the relative pose for the normal camera app. Theses graphs highlight the extreme difference between the pose error of the two groups.

Participants took an average of 3.05 more photos with the normal camera app compared to Cake Cam. The number of photos taken to get the right photo with the normal camera

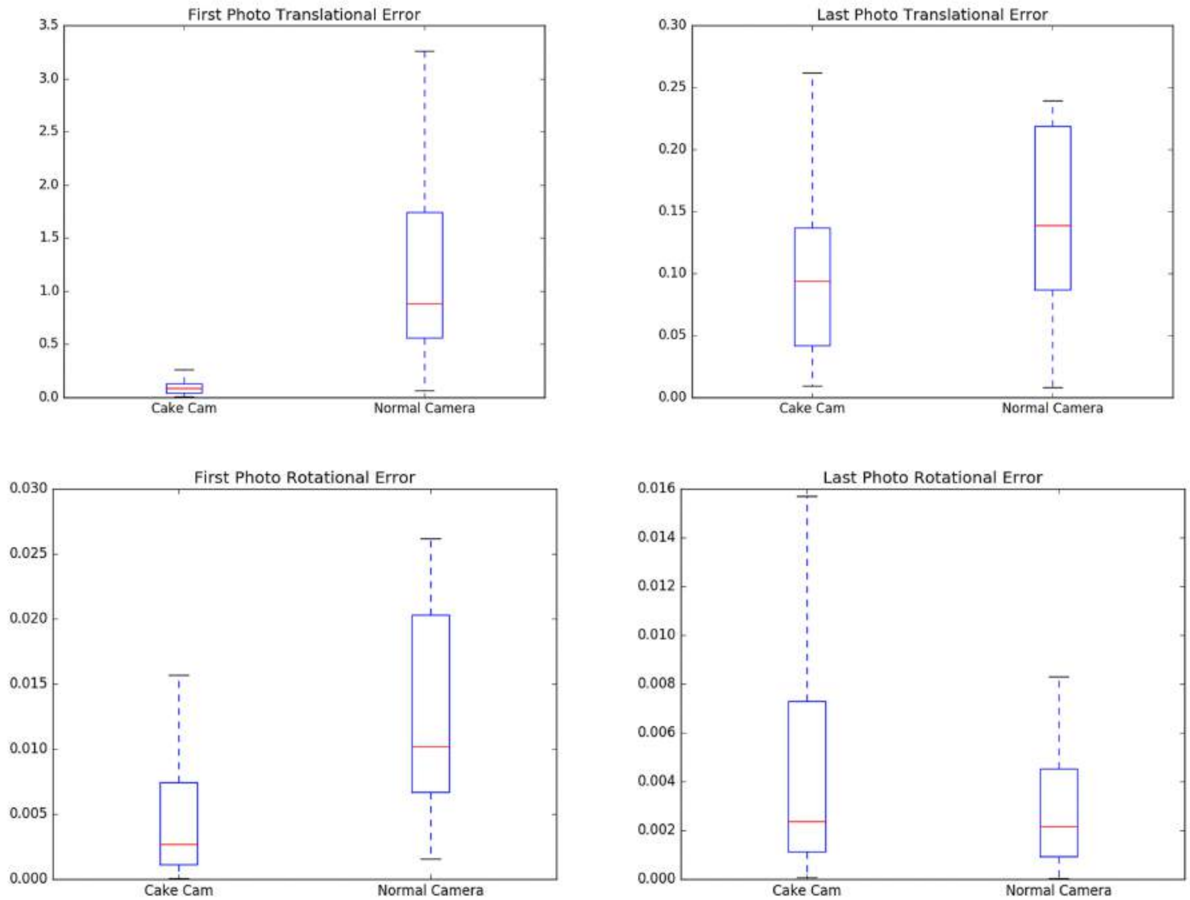

Figure 5.1: Box plots of the first and last pose error for both Cake Cam and the da app.

	Camera	Mean	Std. Dev	p-value	t-value
Photos Taken	Cake Cam	1.05	0.22	¡ 0.00001	-7.5
	Normal	4.1	1.8		
Confidence	Cake Cam	8.1	1.4	0.00041	3.9
	Normal	5.9	2.1		
Difficulty	Cake Cam	2.2	1.4	0.000052	-4.6
	Normal	5	2.4		

Table 5.3: Statistical summary for number of photos taken, confidence in having captured the intended photo, and difficulty of capturing the intended photo for study 2.

had a mean of 4.1. Comparatively, Cake Cam had a mean of 1.05 photos taken: a statistically significant difference of $p < .00001$. Exact values can be found in table 5.3.

After taking the first photo, participants using Cake Cam rated their confidence in having taken the correct photo as 8.1, on average. Participants using the normal camera rated

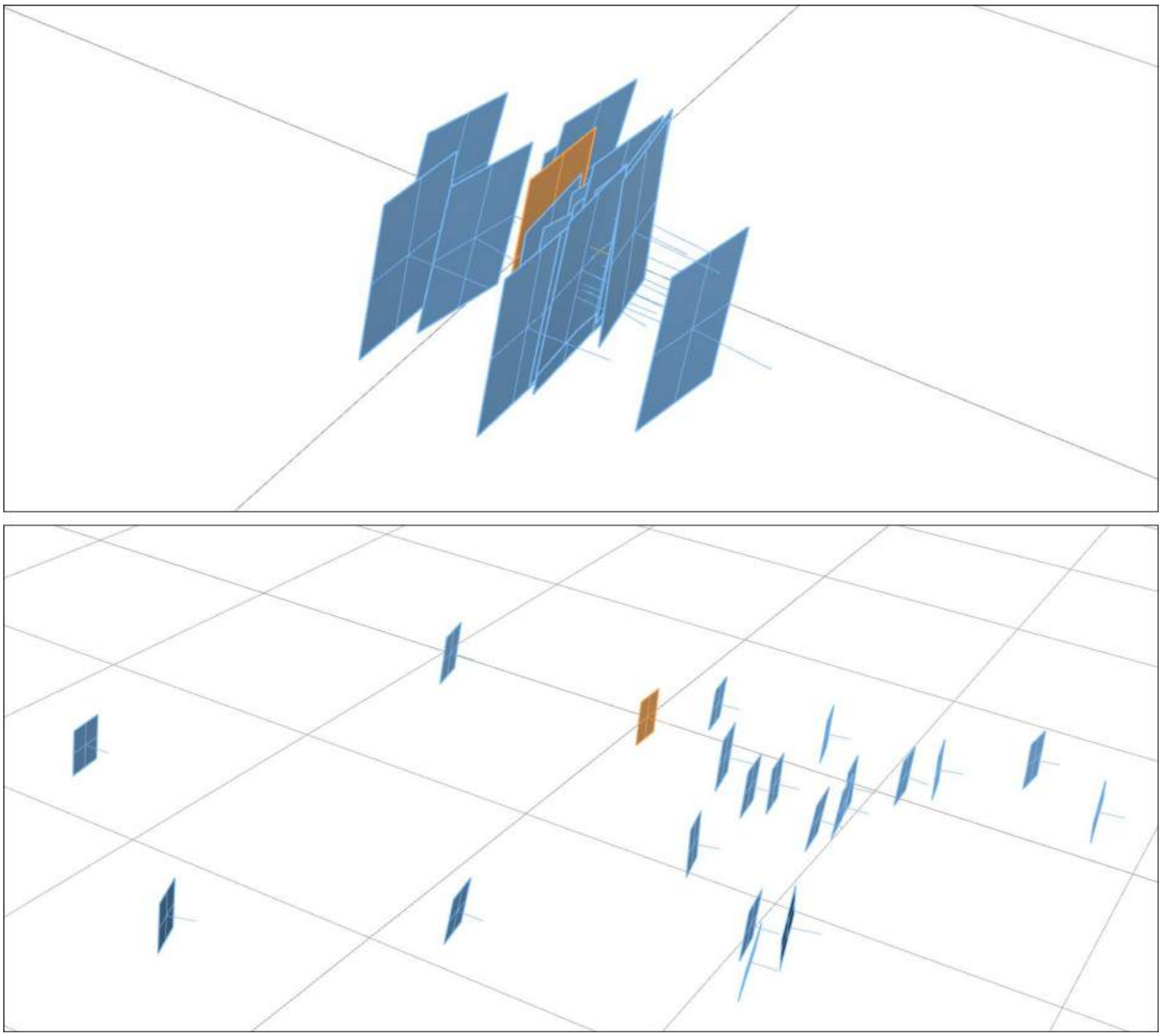

Figure 5.2: The top graph shows the relative pose error between the reference image and final image for each participant that used Cake Cam. The bottom graph shows the relative pose error between the reference image and final image for each participant that used the normal camera app.Each coordinate frame represents the relative pose error for one participant.

their confidence as 5.9, on average. The difference between users' confidence is statistically significant: a value of $p = 0.00041$. As a side note, we find it interesting that participants using the normal camera rated their confidence so high when, on average, it took almost three more ties to capture the intended photo. This contradiction will be explored further in section 5.3.

Once the participant had taken the correct photo, he or she was asked to rate the difficulty in capturing it. Participants that used Cake Cam rated the difficulty as 2.2, compared to a rating of 5 for participants that used the normal camera. This difference is also statistically significant, at $p = 0.000052$.

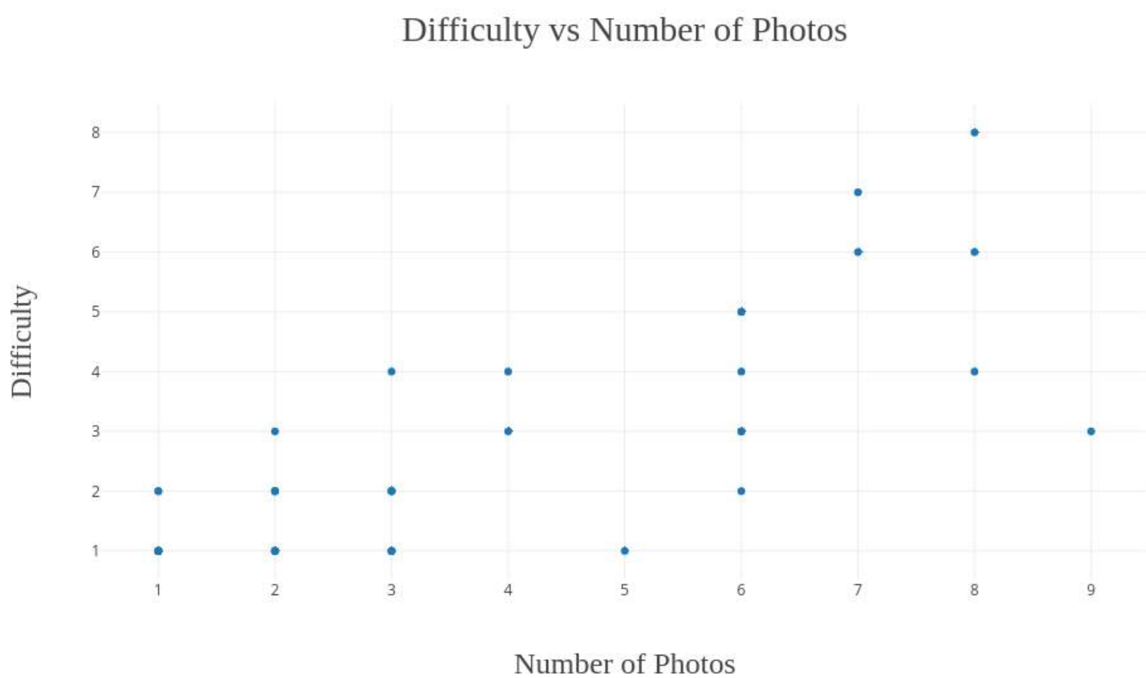

Figure 5.3: The number of photos taken with both Cake Cam and the normal camera, charted against the participants perceived difficulty in taking them.

Using Spearman's Rho Calculation, we found there was a correlation between the number of photos taken with both Cake Cam and the normal camera and how difficult the participant perceived the experience to be (with $p = 0.01$ and $R = .72$). Figure 5.3 shows the number of photos taken, charted against the participant's perceived difficulty in taking them. While it is not surprising that taking multiple photos is more difficult then taking one photo,

Figure 5.4: Reference photo and final photo taken with Cake Cam. This photo was captured in one try by participant 12 (P12).

these results give more insight into the current method of having a photo taken at a tourist location. We believe there is a social faux pas against asking a local to take multiple photos, or even giving more instructions. This theory will be discussed further in section 5.3.

Overall analysis of the pose error and data from the questionnaire indicates that the two data sets were created from distinct populations. Further we can confidently say that Cake Cam was more accurate at guiding participants into taking the correct photo with less difficulty, and gave the participants more confidence in capturing the intended photo.

5.2 Themes in Participant Responses

At the conclusion of the study, we asked the participants to describe their experience of taking the requested photo. We used thematic analysis on the responses and identified four themes: worry, communication, resignation, and ease of use. We discuss each of them in the following subsections.

5.2.1 Worry

Many participants expressed a sense of worry (or lack thereof) about taking the photo the tourist wanted. Participants that used Cake Cam gave feedback such as, It was awesome! I wasn't stressed about getting what you wanted in the frame (P11). In contrast, participants that used the normal camera app expressed a more negative experience: I felt bad because I wasn't getting it how I was supposed to. It did get a little frustrating, as I knew you had something in mind but I wasn't getting it (P38).

We were pleased to note that our app not only solved the pain of communicating the right photo, but also eased the burden of the local taking the photo. We were surprised that so many of the people playing the role of the local-turned-photographer commented that they never knew how to get the photo the other person wanted and this app took away all such worries.

5.2.2 Communication

Participants that used Cake Cam made comments about how nice it was that the app communicated the exact photo desired. As one participant said, It was cool how you were able to tell me exactly how you wanted the photo. You're never sure what angle they are looking for (P25). And again, we see that this communication help ease the worry of one participant when he said, It made it really easy. Didn't have to guess what you wanted because of the markers. I'm not good at taking photos, so usually I take 20 to make up for that. (P21)

Comparatively participants that used the normal camera app often indicated that more communication and direction would have been helpful. Participants made such comments as, If you want it a specific way I need more direction, since I can't read your mind (P36). One participant, however, noted that a general lack of communication is common when someone has his or her photo taken by a local when he said, this is something that happens all the time. It's hard to get exactly what they want if they don't communicate it (P29).

These comments seem to indicate that from the local's perspective, more communication would be better. The lack of communication might explain the lack of worry on the part of the standard-camera app participants, which was mentioned previously: a local cannot worry about failing to meet requirements that he or she is not aware exist. Cake Cam allows the tourist to communicate his or her requirements without giving any instruction, while the normal camera app and a simple statement, take my photo with the bell tower leave determining the composition of the photo to the local. When the tourist does not indicate specifics about the photo they want taken, the local must rely on his or her own judgment. This burden causes the local to feel stress, and worry that the tourist will be unhappy with the final image.

Another point of interest regarding communication involves participant 37, who was removed from the final study results. Initially this participant misheard the first instruction. The participant thought the instruction was to put the bell tower on the researcher's right, instead of the right of the photo. The participant took five photos with the same instruction between each photo, "Can you take the photo with the bell tower on the right side of the photo?" before she realized her mistake. In this scenario verbal communication is not always guaranteed to be interpreted correctly. And the problem would be even further exacerbated in places where the tourist does not speak the native language. Participant 37 believed she was taking the photo correctly and had a hard time overcoming the initial, misheard instruction. Because of this communication error, we ultimately decided to remove her results from the final study analysis.

Many participants experienced a similar struggle. After being given a piece of instruction to correct the photo, such as,"Can you take the photo with the bell tower on the right side of the photo?" The participant would then correct this aspect of the photo, but

Figure 5.5: This participant took four tries to capture the correct photo. The photograph on the left is the initial reference photograph. The four photographs on the right are each new attempt at capturing the correct photograph.

simultaneously introduce a new mistake, such as moving in closer. This struggle can be seen in figure 5.5. In the first photo, the participant centered the bell tower. The participant was then given instruction to place the bell tower on the right of the photo. The participant complied, but then also moved closer to the research team member. For the next photo, the participant moved back, but she again center-aligned the bell tower. After the fourth and final try, the participant was able to correctly capture the intended image. We found that many participants similarly circled around the intended photo until they finally captured the desired image.

Overall we believe that participantsspecifically those equipped with the standard camera appwanted more communication about the intended photo; however, current mechanisms for communicating this intent are unreliable and takes too much time.

5.2.3 Resignation

Another theme we identified in our results was a sense of resignation to producing a "tourist"-quality photo. As one participant commented, "I think when you ask someone to take your photo, you have to have 'tourist photo' expectations" (P41). Such a statement implies that, as another participant said, "when you ask someone [to take your photo] you have to assume it's not going to be what you want" (P3). We found one reason for this resignation to be caused by the social anxiety of asking a local to take multiple photos, as was done in this study. One participant commented, "It's okay to ask those kinds of instructions of a friend, but to ask someone you don't know is kind of strange" (P39). Asking someone to take multiple photographs breaks a social norm and is not something people are comfortable doing. This convention leaves the tourist with no option but to accept the poor photograph.

This struggle was particularly well illustrated in our experience with participant 40, who was removed from the final results as he did not complete the study. This participant gave up after taking five photos. Frustrated, he proclaimed that the photo was fine and that he was done. The participant was agitated by the experience and refused compensation.

Other participants commented that they were fine taking multiple photos because they knew it was for a study, but would otherwise be annoyed if someone asked them to do so.

5.2.4 Ease

Many participants that used Cake Cam commented on how easy it was to get the intended photo. Cake Cam participants made comments such as, "It made it a lot easier: all I had to do was line it up. Knowing what they wanted was way easier!" (P17). Other participants made very similar general comments about the usability of the app, "It was super easy! As soon as I looked at the screen it was very self-explanatory" (P9).

While some of these comments may be due to a novelty effect, a limitation that we discuss more in the chapter 6, the effect cannot fully account for the overwhelmingly positive experiences reported. We believe the experiment supports our claim that the Cake Cam interface communicates the intended photo in a manner that is more effective, efficient, and natural than the current verbal method for a standard camera app.

5.3 Combined Results

Combing the statistical results with the thematic analysis on the open-ended questions provided some interesting insights about communicating intent through a mobile app.

We found that participants were more confident that they had taken the right photo when they used Cake Cam. This statistic aligns with the theme of worry we identified. Participants that used Cake Cam frequently commented about the lack of worry they experienced, compared to when they normally take photos for others. The app provided more direction on where the tourist wanted the photo taken, giving the locals more confidence that they were capturing the intended photo. One participant commented, "I didn't have to guess what you wanted, because of the markers" (P18). Surprisingly, participants using the normal camera rated their confidence relatively high, considering that it took an average of almost three more tries to capture the intended photo. We believe this higher confidence rating ties to the resignation theme identified previously. Participants suggested that tourists commonly resign themselves to a less-than-perfect photo, so the participants may have been more confident that they got the correct photo based on the assumption that this study's tourist would be similarly resigned.

The theme of resignation also ties into the correlation we discovered between difficulty and the number of photos taken before capturing the intended image. Asking a local to take multiple photos causes them to spend more time on the task and increases the burden. While someone people, like participant P31, will have no problem taking multiple photos, as she explained, "In today's world of social media it's kind of expected that you'll have to take a few photos" (P31). Other people will not have this same attitude as we saw with participant P40 who got upset at the request for multiple photos and left before finishing the study.

The introduction of social awkwardness is where we started to sense a sort of paradox when it comes to having a photo taken at a tourist location. While the participants acting in the role of the local were actively worried about failing to meet the tourist's expectations and indicated a desire for more communication, they did not want to spend the time needed to receive this communication and potentially take multiple photos. The tourist has an equally daunting struggle: asking a local to retake an unsatisfactory photograph breaks a social norm and makes the whole experience more difficult for the local. Based on the information our participants presented along these lines, it seems that these limitations have been simply accepted as insurmountable, and that no photo resulting from such an interaction is likely to be satisfactory. It is, therefore, our great delight to introduce Cake Cam as the solution to this seemingly hopeless obstacle.

Overall, we found that Cake Cam helped mediate the interaction between a tourist and a local collaborating to take a photo with unprecedented efficiency. Cake Cam improved the entire experience for both parties involved: it provided the tourist with the photo they wanted, and made the process easier for the local.

Chapter 6

Discussion and Conclusion

In this thesis we have established that Cake Cam can effectively guide users into replicating a photo. Further, we learned that augmented reality markers were more effective and efficient in guiding users into replicating a photo than verbal descriptions were. On average it took 3 more tries for the participant to take the correct photo without Cake Cam than with it. Additionally with Cake Cam, the final photo was closer to the intended framing by $9cm$.

6.1 Limitations

This thesis focuses on the particular CMC2 task of asking a stranger to take one's picture in front of a landmark. This limited focus may be seen by some as a limitation; however, even though this work is limited to one type of CMC2 task, the concepts examined here can adapt into more general functions, as outlined in the Future Work section following.

Another limitation that arises in this thesis affects the results of the second study, wherein we had all participants take a typical, eye-level picture of the tourist in front of a landmark. Because we chose a standard orientation for the camera, we did not fully test Cake Cam's ability to guide users into taking a photo at an extreme orientation. There was no significant difference between the final orientation error of photos taken with Cake Cam and photos taken with verbal description. This is a limitation of our study, and further experimentation would need to be done to conclusively say Cake Cam is better at guiding users into replicating photos at extreme orientations as well as the more typical varieties.

A third limitation comes into play with users' descriptions of their Cake Cam experience. As acknowledged previously, their enthusiastic reviews and the ratings they gave may have been influenced by the *novelty effect* of the new app. The novelty effect describes the short-term increase in performance when a participant is introduced to a new technology or system. Thus, this phenomenon and its effects needed to be considered when we analyzed the participants' optimistic reviews.

6.2 Explanation of Results

Based on the results of this work, we proposed the following two explanations for our results:

1. Cake Cam increases users' ability to replicate photos, and
2. AR markers are more effective and efficient than a verbal description is at quickly communicating intent.

As discussed in chapter 5, the use of Cake Cam requires on average 3.05 fewer tries to replicate a photo, which is a statistically significant figure. Furthermore, Cake Cam lowered the perceived difficulty in performing the task by 56%.

We also found that AR markers communicate intent quicker than a simple verbal description can. We hypothesize that there are two elements to this conclusion. First, AR markers provide continuous feedback to the local who is attempting to replicate the photo. This continuous feedback contrasts with the limited verbal descriptions given before and after each picture is taken, which method leaves greater room for error. Furthermore, we learned from our thematic analysis that participants find it to be socially awkward to ask or be asked to take multiple photos or to involve too much instruction; however they conversely indicated a desire for more instruction on how to take the photograph. These two contradicting needs can be brought into alignment with continuous feedback. The local receives the information they wantwhat photograph to take without having to take the time to verify the photograph is correct and possibly retake it.

The second element in the benefits of AR-marker-guided communication is that a low-order representation of the goal is better than a highly-detailed representation. AR markers provide an abstract idea of the intended photo, as opposed to showing the desired picture and asking the user to replicate it without further instruction. As detailed before, we originally used a superimposed copy of the desired image over the screen as a guide so that the user could faintly see it while attempting to align the photos. However, this was confusing to the user and took longer to achieve the desired photo compared with using AR markers. This lag is likely caused by the requirement that the user must process the highly detailed visual instruction, requiring them to reverse engineer the camera's pose from the image before taking one themselves. By communicating abstract instructions and salient features, the user does not need to process so much information. The AR markers, therefore, communicate what needs to done to achieve the intended resultwhere to position the phoneinstead of communicating the exact intended results, the desired photograph. We hypothesize that abstract representation of intent is more effective for communicating intent than the more detailed alternative.

Our hypothesis that abstract representations allow for more effective and prompt communication of intent is further discussed in the Future Work section that follows.

6.3 Future Work

We suggest two research directions for future work. The first is more work involving Cake Cam and having a photograph taken at a tourist location. The second is generalizing the principles employed in Cake Cam's development to other CMC2 problems

6.3.1 Cake Cam

Further studies would need to be done to clearly identify why the AR marker method of transmitting intent was effective. We hypothesize that abstract representations allow for more effective and prompt communication of intent than other methods described previously, but further work would need to be done to confirm this hypothesis. We envision a study wherein the participants would first be shown the correct photo and then asked to replicate it. Because the participant would be shown the exact intended photo, we could then compare this method of communicating intent to Cake Cam's method of abstracting intent with AR markers. We believe

Additionally, there is more work to be done with the design of the app Cake Cam. While overall the crosshair marker design was effective in guiding users, some participants did not realize they needed to align the four corner markers as well as the center markers. Further testing could be done to improve this issue and better communicate how the local needs to move the phone to capture the intended photo. We also see value in conducting more user studies that use an extreme orientation, either a high or low angle, to test Cake Cam's ability at communicating orientation as well as position.

6.3.2 Generalizing

Our work suggests avenues for future research on CMC2 tasks and we believe this idea of abstracting intent could be extended to more CMC2 tasks. We only considered the specific problem of asking a stranger to take one's picture in front of a tourist landmark, but there are many other CMC2 tasks that could be mediated with computing. One potential direction, for instance, might be asking a stranger for directions.

When a tourist stops a local and asks for directions to a specific location, the tourist is much more invested in getting to right directions than the local. The local will move on with his or her day completely unaffected by whether the tourist eventually finds the destination in question. This interaction typically happens on the street, with limited time for the tourist to explain where he or she wants to go and to receive the directions. In addition, there is the complication that tourists can have different goals when seeking a location: for instance, one might want the quickest route, the most scenic route, or the safest route to a destination. Other times, a tourist may not have specific destination in mind, but is instead looking for something general like a nearby park, a good restaurant, or the closest restroom. These goals can be hard to quickly communicate in the limited time allowed for this interaction. Additional complications may arise when asking for and receiving the directions, such as language barriers and lack of spatial knowledge. Overall a tourist is left trying to communicate and receive a lot of information in a short amount of time. Having a way to abstract the information so the local receives only the critical information is key to making this interaction smooth.

6.4 Conclusion

In this thesis we explored communication of intent specifically in the context of having one's photo taken at a tourist location. We designed and developed an app, Cake Cam, to validate the claim that an augmented reality mobile app can mediate completion of time-sensitive asymmetric CMC2 tasks by communicating a tourist's intent when a local and tourist collaborate to take a picture in front of a landmark. We ran two user studies to validate our thesis and answer the question, Can technology be used to communicate intent between two people, specifically in the context of having a photo taken at a tourist landmark?

The first study found that the user interface of Cake Cam is effective in guiding the user into replicating a photo by positioning the phone in the correct location. From the second study, we learned that Cake Cam was more effective and efficient at communicating intent compared to verbal descriptions, the current method of communicating intent. Photos taken with Cake Cam were closer to the intended photo by $9cm$. Participants that used Cake Cam found the process to be less difficult and they were more confident that they had captured the correct photo.

We used thematic analysis on answers to the open-ended question at the end of the second study, and identified four themes: worry, communication, resignation, and ease of use. These themes helped us better understand the statistical data and how it related to participants' attitudes towards communicating intent with a mobile device.

These results lead us to believe that abstract representations of intent, like Cake Cam's AR markers, allow for more effective and prompt communication of intent, because they present only the vital information needed to complete the task, streamlining the interaction into a pleasant experience for all involved parties.

References

[1] Michael Bloesch, Michael Burri, Sammy Omari, Marco Hutter, and Roland Siegwart. iterated extended kalman filter based visual-inertial odometry using direct photometric feedback.

[2] Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. Robust visual inertial odometry using a direct EKF-based approach. IEEE International Conference on Intelligent Robots and Systems, 2015-Decem:298–304, 2015. ISSN 21530866. doi: 10.1109/IROS.2015.7353389.

[3] Stephen A. Brewster and Jody Johnston. Multimodal interfaces for camera phones. In Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '08, pages 387–390, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-952-4. doi: 10.1145/1409240.1409295. URL http://doi.acm.org/10.1145/1409240.1409295.

[4] Barry Brown, Matthew Chalmers, Marek Bell, Malcolm Hall, Ian MacColl, and Paul Rudman. Sharing the square: Collaborative leisure in the city streets. In Proceedings of the Ninth Conference on European Conference on Computer Supported Cooperative Work, ECSCW'05, pages 427–447, New York, NY, USA, 2005. Springer-Verlag New York, Inc. ISBN 978-1402040221. URL http://dl.acm.org/citation.cfm?id=1242029.1242051.

[5] Scott Carter, John Adcock, John Doherty, and Stacy Branham. Nudgecam: Toward targeted, higher quality media capture. In Proceedings of the 18th ACM International Conference on Multimedia, MM '10, pages 615–618, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-933-6. doi: 10.1145/1873951.1874034. URL http://doi.acm.org/10.1145/1873951.1874034.

[6] James Clawson, Amy Voida, Nirmal Patel, and Kent Lyons. Mobiphos: A collocated-synchronous mobile photo sharing application. In Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, Mobile-HCI '08, pages 187–195, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-952-4. doi: 10.1145/1409240.1409261. URL http://doi.acm.org/10.1145/1409240.1409261.

[7] Andrew Davison, Ian Reid, Nicholas Molton, and Olivier Stasse. MonoSLAM: real-time single camera SLAM. Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on, 29(6):1052–67, 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.1049. URL http://www.ncbi.nlm.nih.gov/pubmed/17431302.

[8] Abigail Durrant, Duncan Rowland, David S. Kirk, Steve Benford, Joel E. Fischer, and Derek McAuley. Automics: Souvenir generating photoware for theme parks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 1767–1776, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979199. URL http://doi.acm.org/10.1145/1978942.1979199.

[9] Jerry Alan Fails, Allison Druin, and Mona Leigh Guha. Mobile collaboration: Collaboratively reading and creating children's stories on mobile devices. In Proceedings of the 9th International Conference on Interaction Design and Children, IDC '10, pages 20–29, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-951-0. doi: 10.1145/1810543.1810547. URL http://doi.acm.org/10.1145/1810543.1810547.

[10] Joel E. Fischer, Stuart Reeves, Stuart Moran, Chris Greenhalgh, Steve Benford, and Stefan Rennick-Egglestone. Understanding Mobile Notification Management in Collocated Groups, pages 21–44. Springer London, London, 2013. ISBN 978-1-4471-5346-7. doi: 10.1007/978-1-4471-5346-7_2. URL https://doi.org/10.1007/978-1-4471-5346-7_2.

[11] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. Talkbetter: Family-driven mobile intervention care for children with language delay. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '14, pages 1283–1296, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2540-0. doi: 10.1145/2531602.2531668. URL http://doi.acm.org/10.1145/2531602.2531668.

[12] Pradthana Jarusriboonchai, Thomas Olsson, Sus Lundgren Lyckvi, and Kaisa Väänänen. Let's take photos together: Exploring asymmetrical interaction abilities on mobile camera phones. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '16, pages 529–540, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4408-1. doi: 10.1145/2935334.2935385. URL http://doi.acm.org/10.1145/2935334.2935385.

[13] Chase Jarvis. The Best Camera Is The One That's With You: iPhone Photography. New Riders, San Francisco, CA, USA, 1st edition, 2009. ISBN 9780321684783.

[14] Auk Kim, Sungjoon Kang, and Uichin Lee. Letspic: Supporting in-situ collaborative photography over a large physical space. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pages 4561–4573, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025693. URL http://doi.acm.org/10.1145/3025453.3025693.

[15] Joel Lanir, Alan J. Wecker, Tsvi Kuflik, and Yasmin Felberbaum. Shared mobile displays: An exploratory study of their use in a museum setting. Personal Ubiquitous Comput., 20(4):635–651, August 2016. ISSN 1617-4909. doi: 10.1007/s00779-016-0931-y. URL http://dx.doi.org/10.1007/s00779-016-0931-y.

[16] Taeyoung Lee, Melvin Leok, and N. Harris Mcclamroch. Geometric Tracking Control of a Quadrotor UAV on SE(3). In Conference on Decision and Control, pages 5420–5425, 2010. ISBN 9781424477449. doi: 10.1002/asjc.567. URL http://doi.wiley.com/10.1002/asjc.567{%}5Cnhttp://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5717652.

[17] Qifan Li and Daniel Vogel. Guided selfies using models of portrait aesthetics. In Proceedings of the 2017 Conference on Designing Interactive Systems, DIS '17, pages 179–190, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4922-2. doi: 10.1145/3064663.3064700. URL http://doi.acm.org/10.1145/3064663.3064700.

[18] Andrés Lucero, Jussi Holopainen, and Tero Jokela. Pass-them-around: Collaborative use of mobile phones for photo sharing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 1787–1796, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979201. URL http://doi.acm.org/10.1145/1978942.1979201.

[19] Andrés Lucero, Jussi Holopainen, and Tero Jokela. Mobicomics: Collaborative use of mobile phones and large displays for public expression. In Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services, MobileHCI '12, pages 383–392, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1105-2. doi: 10.1145/2371574.2371634. URL http://doi.acm.org/10.1145/2371574.2371634.

[20] Sus Lundgren, Joel E. Fischer, Stuart Reeves, and Olof Torgersson. Designing mobile experiences for collocated interaction. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15, pages

496–507, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675171. URL http://doi.acm.org/10.1145/2675133.2675171.

[21] Thomas W. Malone, Jeffrey V. Nickerson, Robert J. Laubacher, Laur Hesse Fisher, Patrick de Boer, Yue Han, and W. Ben Towne. Putting the pieces back together again: Contest webs for large-scale problem solving. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, pages 1661–1674, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4335-0. doi: 10.1145/2998181.2998343. URL http://doi.acm.org/10.1145/2998181.2998343.

[22] Christopher McAdam, Craig Pinkerton, and Stephen A. Brewster. Novel interfaces for digital cameras and camera phones. In Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '10, pages 143–152, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-835-3. doi: 10.1145/1851600.1851625. URL http://doi.acm.org/10.1145/1851600.1851625.

[23] Jarno Ojala, Kaisa Väänänen-Vainio-Mattila, and Arto Lehtiniemi. Six enablers of instant photo sharing experiences in small groups based on the field trial of social camera. In Proceeding of the 10th International Conference on Advances in Computer Entertainment - Volume 8253, ACE 2013, pages 344–355, New York, NY, USA, 2013. Springer-Verlag New York, Inc. ISBN 978-3-319-03160-6. doi: 10.1007/978-3-319-03161-3_25. URL http://dx.doi.org/10.1007/978-3-319-03161-3_25.

[24] Jeff Sauro. A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices. CreateSpace Independent Publishing Platform, April 2011.

[25] Jeff Sauro and James R. Lewis. Quantifying the User Experience: Practical Statistics for User Research. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2012. ISBN 9780123849687, 9780123849694.

[26] Michael Andrew Shelley. Monocular Visual Inertial Odometry on a Mobile Device. Thesis, page 92, 2014.

[27] Kate Starbird and Leysia Palen. Working and sustaining the virtual "disaster desk". In Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13, pages 491–502, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1331-5. doi: 10.1145/2441776.2441832. URL http://doi.acm.org/10.1145/2441776.2441832.

[28] Zachary O. Toups and Andruid Kerne. Implicit coordination in firefighting practice: Design implications for teaching fire emergency responders. In Proceedings of the SIGCHI

Conference on Human Factors in Computing Systems, CHI '07, pages 707–716, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240734. URL http://doi.acm.org/10.1145/1240624.1240734.

[29] James Wen and Ayça Ünlüer. Redefining the fundamentals of photography with cooperative photography. In Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, MUM '15, pages 37–47, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3605-5. doi: 10.1145/2836041.2836045. URL http://doi.acm.org/10.1145/2836041.2836045.

[30] Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. Real-time guidance camera interface to enhance photo aesthetic quality. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pages 1183–1186, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702418. URL http://doi.acm.org/10.1145/2702123.2702418.

[31] Soobin Yim, Dakuo Wang, Judith Olson, Viet Vu, and Mark Warschauer. Synchronous collaborative writing in the classroom: Undergraduates' collaboration practices and their impact on writing style, quality, and quantity. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, pages 468–479, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4335-0. doi: 10.1145/2998181.2998356. URL http://doi.acm.org/10.1145/2998181.2998356.