Video Manipulation Techniques for the Protection of Privacy in Remote Presence Systems

Alexander Hubers, Emily Andrulis, Tanner Stirrat, Ross Sowell, and Ruonan Zhang
Department of Computer Science
Cornell College
Mount Vernon, IA, USA
{ahubers15,eandrulis16, tstirrat15, rsowell, and rzhang16}@cornellcollege.edu

Cindy Grimm and William Smart
Department of Mechanical Engineering
Oregon State University
Corvallis, OR, USA
{cindy.grimm,bill.smart}@oregonstate.edu

ABSTRACT

Systems that give control of a mobile robot to a remote user raise privacy concerns about what the remote user can see and do through the robot. We aim to preserve some of that privacy by manipulating the video data that the remote user sees. In this paper, we explore the effectiveness of different video manipulation techniques at providing different types of privacy. Participants were asked to watch a video captured by a robot exploring an office environment and to complete a series of observational tasks under differing video manipulation conditions. We examine privacy protection as a function of condition. Our results show that using video manipulations can lead to fewer privacy violations for different privacy types. Specifically, we found that the "Can't tell" and "Can't observe" privacy types had the fewest privacy violations with the redact and replace video manipulations, and it was determined that the expectations for the "Can't discern" privacy type were best upheld with the blur video manipulation.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*evaluation/methodology*; I.2.9 [**Artificial Intelligence**]: Robotics—*operator interfaces*; I.4.3 [**Image Processing and Computer Vision**]: Enhancement—*filtering*

General Terms

Design, Experimentation, Human factors

Keywords

Privacy interfaces, remote presence systems, video manipulation

1. INTRODUCTION

What makes a mobile remote presence system more of a privacy concern than, say a Skype connection? The main thing is a shift in control. If you do not want someone to see you in the shower over Skype, then you do not take your laptop into the bathroom. You have control over what the user on the other end sees. However, with mobile remote presence systems, the remote user can pilot the system around the world, giving them control over what they see.

For the purposes of this paper, we define a remote presence system (RPS) to be a system that allows a remote operator to be virtually present in another location, and to interact with the people and things there. The telephone and Skype are two common examples of such systems. However, the focus of the work in this paper is on systems that cannot just observe their environment, but also act on it. The canonical example of such a system is the Personal Roving Presence (PRoP) project [23], a mobile robot base mounting an LCD screen, camera, and speaker on a human-height pole. The robot wandered around the UC Berkeley campus and surrounding areas under the control of a remote operator, interacting with the people that it met. A decade later, similar systems are becoming commercially available, including the Beam [3], VGo [4], and InTouch RP-VITA [1]. The caricature of these systems is that they are "Skype on a stick": a traditional video-conference interaction where the user has some control over the physical location of the system.

These mobile systems introduce new privacy concerns. With a passive teleconference system, the remote participant can only look at what the proximal participant points the camera at. On mobile systems, the remote operator can now point the camera themselves, taking control of the privacy away from the proximal user.

One method of preserving the privacy of the proximal user is to alter the visual data that the remote operator sees. The central question that we investigate in this paper is: What methods of video manipulation are most effective at ensuring a given privacy type?

1.1 Observational Privacy Types

We define a *privacy type* to be a specific restriction on the capabilities of the remote presence system. The capabilities

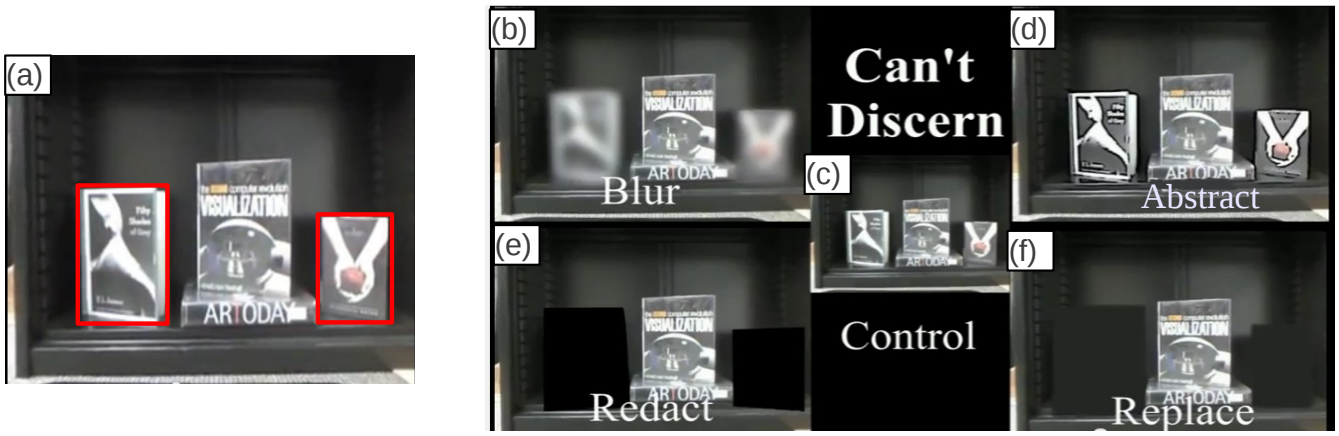

Figure 1: Video manipulation techniques. The original image is shown with the two objects to be filtered highlighted in red (a). Blur (b) Control (c), Abstract (d), Redact (e), and Replace (f) video manipulations are applied to the filtered areas.

can either be *physical*, where the system is prevented from taking some action (the RPS cannot enter/leave some area or touch some object), or *observational*, where the sensor data transmitted to the remote user are censored in some way. In this paper, we focus on techniques for ensuring different types of observational privacy as related to the video stream, since this is both the most challenging sensor modality and also the one most likely to violate privacy.

We consider implementations for three types of observational privacy: "Can't tell", "Can't discern", and "Can't observe." We frame these as restrictions on the remote user, rather than on the RPS itself. The RPS has unrestricted access to the data from its sensors, but only passes on a modified version of these data to the remote user.

1. **Can't tell**. The remote operator cannot tell if a particular object is there or not. The goal is to remove an object without the operator noticing that it is missing. Examples include: Not noticing any items exist on a table, unable to tell that there is a person present in the room.
2. **Can't observe**. The remote operator might be able to tell there is something there, but cannot directly perceive it. Examples include: Not being able to look into a certain room, not being able to identify a shape as a person, or not showing what types of objects are on a table.
3. **Can't discern**. The remote operator can tell that there is something there and can identify the class of the object, but not the particular instance. Examples include: Unable to read the text of documents on the table, unable to make out facial features, and unable to make out details of pictures on the walls.

1.2 Video Manipulation Techniques

In order to preserve privacy we need to alter the visual data the operator is seeing. There are a variety of methods for doing so, many of which arise from the field of non-photorealistic or artistic rendering. We discuss related work in that area here, and then we list the specific techniques that we evaluate for privacy protection in this paper.

Broadly speaking, we classify manipulation of images and video by how they change the image: blur, inpainting, abstraction, line drawings, and painterly rendering. Blurring is a straightforward image filter and is commonly used in TV to obscure people's faces. Inpainting [5, 8, 11, 17, 29, 31] allows for filling an area of an image with synthesized content that is ideally indistinguishable from its environment. Abstraction, also sometimes called image stylization [19, 33, 21, 32, 10], is similar to blurring, in that details are elided, but it differs in that strong edges are preserved. It can also involve restricting the color palate to create a cartoon-like effect. Since it is essentially a texture filter, most methods can be efficiently implemented on a GPU [34, 36]. Line drawings [9, 25, 13, 15, 16] similarly preserve edges, but eliminate color information and render the result as a pen and ink or pencil-style sketch (sometimes with shading represented as hatching [28]). Painterly rendering techniques try to mimic a particular style, such as pixelation [14], oil or watercolors [20, 22, 35, 24] and comic-style [26]. Although not always intentional, most of these techniques also result in some image simplification or loss of detail, especially with large brush sizes.

One primary concern when working with video or a moving camera is that the image can flicker because the lines and strokes change over time, or (if the strokes or lines are fixed with respect to the image plane) there is a "shower door" effect (this is not an issue for blurring and some of the simpler abstraction techniques). There are several methods for addressing this [25, 20], with the primary approach being to evolve the current stylization to the next frame, rather than starting from scratch.

For the purposes of the initial study presented in this paper, we have chosen four basic video manipulation techniques to evaluate with respect to their effectiveness at providing different types of observational privacy. We will refer to them as Redact, Blur, Replace, and Abstract. These were cho-

sen because they are representative of the different classes of techniques that we might apply to protect privacy, implementations were readily available, and they can be efficiently implemented on a GPU to achieve interactive rates.

1. **Redact**. This is the simplest and most conservative of the techniques. We hide something by removing it from the video stream, i.e., blacking it out.

2. **Blur**. Also a straightforward technique, we obscure a specified portion of the video data by applying a Gaussian blur.

3. **Replace**. We replace an object with data from its surrounding environment by applying an inpainting technique [30].

4. **Abstract**. We abstract a portion of the video by applying a combination of bilateral and meanshift filtering.

See Figure 1 for an example of each video manipulation technique.

1.3 Impact of Video Manipulation on Protecting Privacy

Various studies have done research looking into how using video manipulations may help uphold privacy expectations. When considering privacy, it is important to understand that one expects autonomy, confidentiality, and solitude to be considered [7]. Filtering out parts of an image through marker detection has been shown to better uphold privacy [27]. Using a blur filter has been shown to better balance protecting one's privacy while also allowing sufficient awareness to the user, so that any necessary and relevant information may still be gleaned from the image [18, 6]. However, in some circumstances where the privacy concerns are greater a blur filter may not be sufficient, and another technique such as redact may work more effectively [12].

In our study we examine which video manipulation techniques work best for the three different privacy types we have explicitly defined. Although research has been done on the effects of using video manipulations to uphold privacy, these studies have been conducted for usage of always on video spaces and video media spaces that have a fixed camera, which is unable to move around to explore an environment. In our study, we investigate to what degree these video manipulations uphold privacy when the camera is capable of traversing around a room and examining objects from different angles of view.

1.4 Hypothesis

In the study presented in this paper, our goal is to determine which methods of video manipulation are most effective at ensuring different types of observational privacy. To evaluate the video manipulation techniques, we test the following specific hypothesis: using the video manipulation technique to provide a given type of privacy leads to fewer violations of privacy expectations.

2. METHOD

We tested five different video manipulations under three different types of privacy expectations. Participants were asked to watch three short video clips that were captured by a robot exploring an office environment, and to respond to five questions asking them to identify objects within the environment.

2.1 Participants

140 participants were recruited through Amazon's Mechanical Turk. Participants were compensated 40¢ for their participation. Participants were told that they would be expected to "watch a clip from the perspective of a robot investigating an office and answer 5 short questions." The average time spent per participant was between 3-5 minutes.

2.2 Procedure

Participants were given the following prompt:

"A mobile robot has explored an office environment for you and acquired the videos on the following three pages. When you are ready to begin, please click "Continue to Videos" below to watch the video and then answer the following questions about what you saw. The questions will be divided into three pages, each containing a separate video. You may take notes, and you may pause, rewind, or replay the video as often as you like. However, once you begin the test, you may not exit out and come back, return to the previous page, or refresh to repeat it."

Each page provided a video clip that concerned one privacy type and used one video manipulation technique. By the end, each participant had viewed one video clip from each of the three privacy types and each clip's video manipulation had been randomly assigned.

The video clips were followed by questions that asked the participant to identify objects in the video. Each privacy type had a different set of questions it asked the participant.

2.3 Control Videos and Privacy Expectations

The videos used in our experiment were captured from the video stream of the Kinect sensor on a TurtleBot 2 [2] that was navigated via teleoperation through a staged office environment. Care was taken to ensure that all objects relevant to the tasks were clearly visible at some point during the videos. All videos used in our study are available at `http://www.youtube.com/channel/UCy1QigehZabjia0OlW_PZPA`

Three different video clips were designed to typify when each type of privacy expectation might be pertinent in the real world. We describe each video scene by the type of privacy expectation it attempts to uphold and protect.

- A "Can't tell" privacy type was applied in the video clip that recorded the robot exploring the office environment and passing some valuable items along the way. In order of appearance, the valuables in the scene were a flat screen television on a shelf, a desktop computer monitor in the corner, and two ASUS laptops on chairs in the middle of the room. In particular, this privacy expectation dictated that the user would

not be able to correctly identify the number of valuable items, but instead should report viewing zero total valuables. Additionally, the user should not be able to identify the locations of any valuable items nor any details about them. In short, the remote user should not be able to tell that these items exist in the room. To test this, participants were asked both, "How many computers or electronic valuables (costing more than $50) are there in the room?" and "Name the valuables and the their locations in the room."

- A "Can't observe" privacy type was applied in the video clip that recorded the robot viewing out a door into a hallway that contained one cardboard box in the center. In this clip the proximal user has specified that the user must not be able to see outside the room, so the user was not allowed to see anything past the open doorway into the hall. Although the cardboard box was the object of interest, it was expected that the user would not be able to see the cardboard box nor anything else in the hallway. After viewing this video clip, participants were asked "Could you see anything in the hallway? If so, please list anything that you saw."
- A "Can't discern" privacy type was applied in the video clip that recorded the robot viewing an office bookshelf and particularly the four books on the bookshelf. These four books were "ARToday," "Visualization: The Second Computer Revolution," "Fifty Shades of Grey," and "Twilight." For this privacy expectation the proximal user does not mind that the books are visible, but does not want to reveal his affinity for embarrassing romance novels. For this reason, the remote user is expected to identify correctly that there are four books, but not be able to read the titles of "Twilight" or "Fifty Shades of Grey." However, the user should still be able to identify "ARToday" and "Visualization: The Second Computer Revolution." Participants were instructed to answer "How many books are on the bookshelf?" and "Name as many titles of the books as you can."

2.4 Video Manipulation Independent Variable

All participants watched each of the three video clips, one for every privacy type. Each of the clips was randomly assigned from the five video manipulations that were applied to every privacy type. A total of 15 videos were created: five distinct video manipulations (Redact, Replace, Abstract, Blur, and the control) applied to each of the three privacy types (Can't tell, Can't observe, and Can't discern). In the manipulated video clips the objects that were filtered were the four valuables (two laptops, one flat screen television, and one desktop computer) in the "Can't tell" scene, the doorway in the "Can't observe" scene, and the two books ("Twilight" and "Fifty Shades of Grey") in the "Can't discern" scene. Although multiple items may have been filtered in one scene, all the filters were the same type of video manipulation within a single video clip. These video manipulations were all applied using Adobe After Effects on the original control video.

2.5 Measures

Our primary source of data is the participants' responses to the survey questions. We also record the time that it takes for participants to answer all questions before going on to the next page. We then coded the recorded answers based on the privacy type.

We measure the degree to which privacy was protected in a distinct manner for each privacy condition, based on the expectations for that type of privacy. For more information on what types of answers were considered privacy violations and which were not please refer to the appendix "Coding Scheme".

The expectation for the "Can't tell" privacy type is that the user will not be able to identify any of the four valuables (desktop, television screen, and two laptops) as the robot explores the environment. In parsing through the data for the "Can't tell" privacy type, one privacy violation was awarded for each correctly identified valuable the participant recorded viewing in the scene.

In measuring the degree of privacy violation for "Can't observe" our main concern was that the participants should not record seeing anything that was in the hallway. Although the main object in the hallways was a cardboard box, a privacy violation was recorded if the participant recorded seeing anything at all that was in the hallway, and that would have been unobservable if the door to the hallway had been closed.

With the "Can't discern" privacy scenario our expectation is that the user is able to correctly identify the two unfiltered books on the shelf, and recognize that there are two more books whose titles should not be distinguishable. We analyzed this data in three ways. First, we recorded the number of privacy violations that occurred and found the average number for each video manipulation. Two privacy violations were possible: one for identifying "Twilight" and another for "Fifty Shade of Grey." Ideally, there should be no privacy violations for this privacy expectation to be upheld. Additionally, we wanted the user to identify that there were four books on the bookcase, so we analyzed the number of books recorded and graphed the averages based on the video manipulation. With this privacy type the participant should have recorded four books as the number of books on the shelf. For task performance, we also wanted to make sure the participant could correctly identify the two unfiltered books on the book shelf. Participants were expected to identify two book titles, "ARToday" and "Visualization: The Second Computer Revolution," correctly.

3. RESULTS

The degree to which each privacy expectation was upheld was analyzed and recorded in different ways for each privacy type. Figure 2 shows the percentage of privacy violations that were recorded on average for each privacy type based on the video manipulation.

Abstract had the most privacy violations for the "Can't tell" privacy condition, with the control video clip following so closely behind it that no significant difference was found between these two video manipulations. Although the blur video manipulation created a lower average number of privacy violations than abstract and control, the difference between these three was not found to be significant. However,

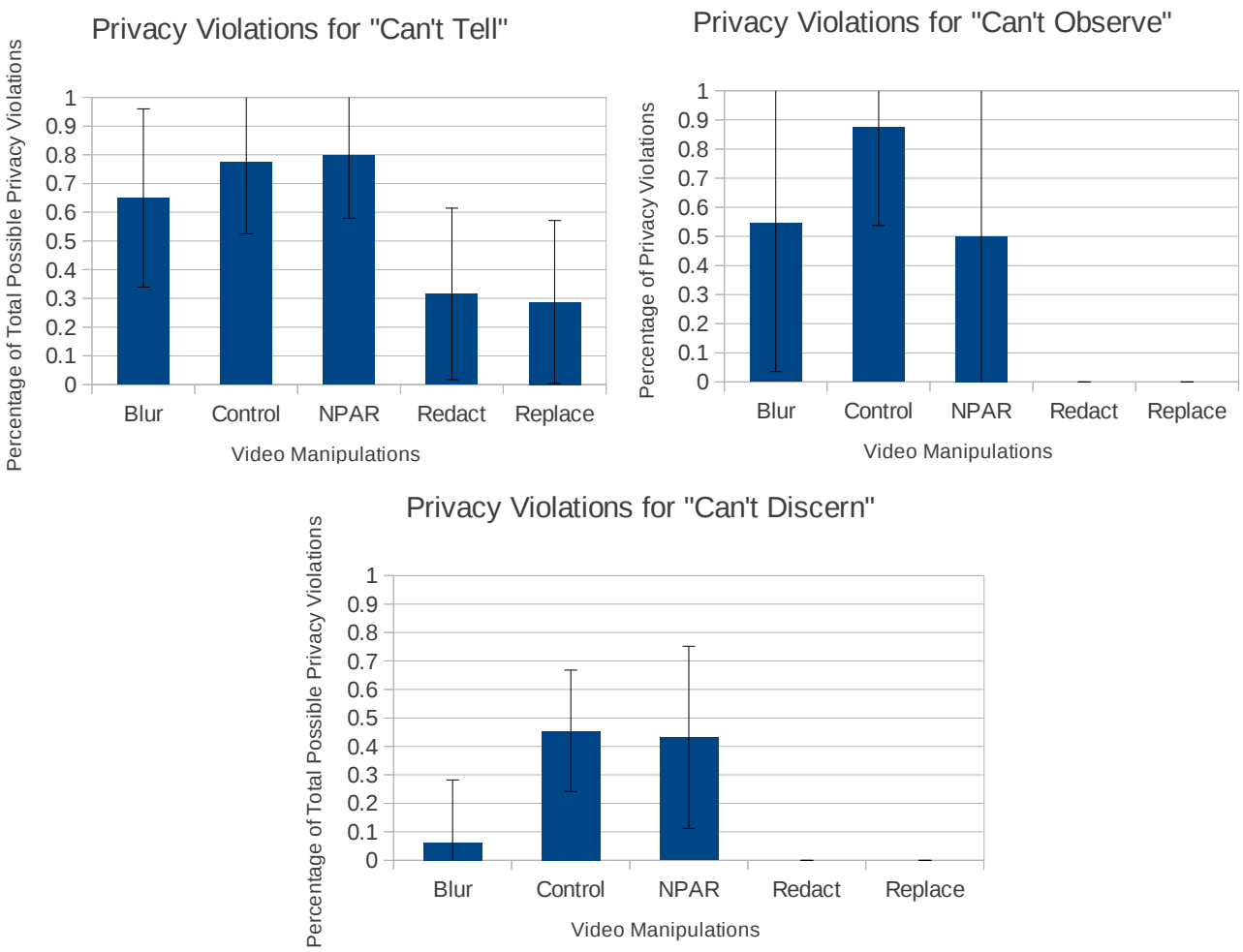

Figure 2: Privacy violation results. These three charts show the percentage of privacy violations that occurred for each video manipulation. Each privacy type had a different number of possible privacy violations: four for "Can't tell," one for "Can't observe," and two for "Can't discern." The data shown illustrates the percentage of total possible privacy violations that occurred on average for each video manipulation. The ideal percentage of privacy violations is always zero. The black bars represent the standard deviation for each video manipulation.

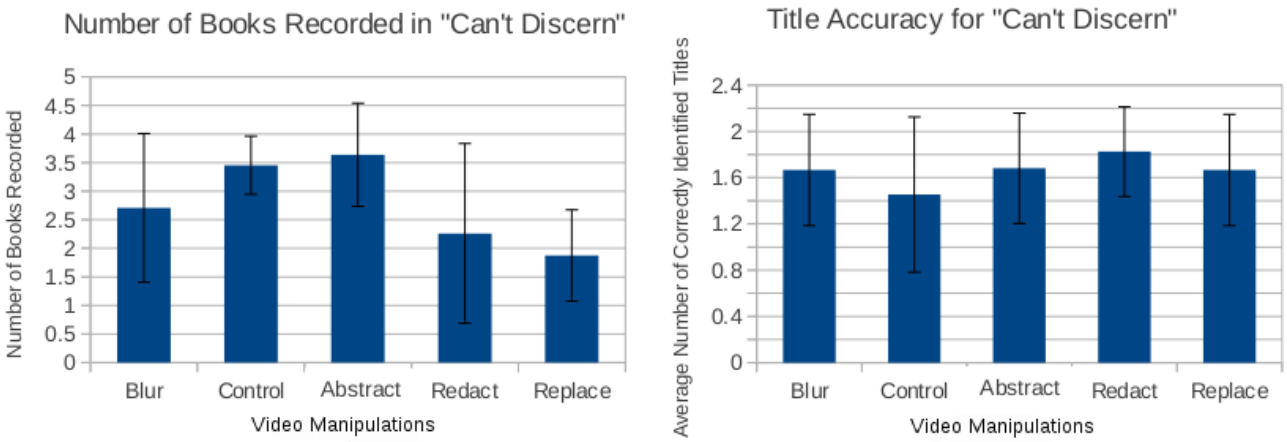

Figure 3: "Can't discern" accuracy results. The "Can't discern" privacy expectation, along with expecting no privacy violations, stated that the participant should be able to accurately identify the unfiltered books and identify the correct number of total books. For this privacy expectation to be upheld, the participants should have identified four total books, and two unfiltered titles. Black bars represent the standard deviation for each video manipulation.

the drop in privacy violations between the blur and redact video manipulations was significant, and redact and ultimately replace had the lowest number of privacy violations. The difference between redact and replace was not significant.

For the "Can't observe" privacy condition, there was only one possible privacy violation. The control video had the highest percentage of privacy violations, with almost 90% of participants violating the privacy expectation. From there the percentage of privacy violations dropped significantly with the blur effect causing only a slight majority to experience a privacy violation, and the abstract video manipulation that had privacy violations recorded exactly half the time. No significant different was found between the blur and abstract video manipulations. There was a significant difference though between the blur and redact video manipulations. Both redact and remove video manipulations recorded zero total privacy violations, so no difference was found between them and they both recorded the least amount of privacy violations.

Although there were two possible privacy violations for the "Can't discern" privacy type, none of the video manipulations averaged more than one privacy violation. The control and abstraction video manipulation had the highest percentage of total possible privacy violations with 45.5% and 43.2% of total possible privacy violations occurring respectively, and no significant difference was found between the two. There was significant decrease from there, since the next highest percentage of total possible privacy violations came from the blur video manipulation with only 6.25% of the possible privacy violations occurring on average. Even fewer was the number of privacy violations for both the redact and replace video manipulations since these two video manipulations resulted in no recorded privacy violations. However, there was no significant difference found between the three video manipulations with the fewest privacy violations.

The "Can't discern" privacy expectation also dictated that aside from expecting no privacy violations the participants should be able to accurately identify the number of books and the titles of the two unfiltered books. The correct number of books on the bookshelf was four, and the closest average number of books recorded was with the abstract video manipulation, which recorded more than 3.5 books on average. The control video followed closely behind with just under 3.5 books recorded on average, and no significant difference was found between the two video manipulations. However, there was a significant difference between the control and the blur video manipulation, which averaged almost halfway between 2.5 and 3 books. Redact averaged just over two books identified on the bookshelf, while the replace video manipulation averaged just below two books recorded being on the bookshelf. The differences between the blur and redact video manipulations and the redact and replace video manipulations were not significant, yet there was a significant difference between the blur and replace video manipulations.

For accuracy in identifying the two book titles that were unfiltered all video manipulations averaged at least 1.45 book titles that were correctly identified of the two possible. Video clips using the redact video manipulation had the highest average of correctly identified book titles, and there was a significant difference between this and the control video, which had the lowest average of correctly identified book titles. Abstract, replace, and blur video manipulations all had very similar averages and no significant difference was found between these averages.

Although time spent was analyzed, no significant informa-

tion was revealed in analyzing the data. The difference in time spent for each privacy condition among the different video manipulations was insignificant and the standard deviations for these average times were disproportionally large. Therefore, it was concluded that no significance could be gathered from analyzing the recorded times it took participants to complete each section of the study.

Our data we collected supports our hypothesis that certain privacy types can be best upheld through the use of certain video manipulations. The control video never had the lowest number of privacy violations, which leads to our conclusion that video manipulations can help better uphold privacy expectations.

4. DISCUSSION

In analyzing the results we discuss each privacy type separately to better consider the full significance of all the recorded results.

4.1 Can't Tell

When trying to determine which video manipulation works best for each privacy type, the first thing to look at it is the percentage of total possible privacy violations that occurred for each video manipulation. The highest number of average privacy violations occurred with the abstract and control video manipulations. This could be due to the way that the control video gives a clear view of the image and all it's objects and the abstract video manipulation emphasizes the outside lines around an object, which could make the item more noticeable if it is recognizable by its shape like most electronic valuables are.

Replace had the lowest average number of privacy violations for the "Can't tell" privacy type, but it was very closely followed by the redact video manipulation. Although one might hypothesize that replace would be significantly more useful in hiding an object completely from the screen and making it seem as though the object was never there, the data does not support that, since no significant difference was found between replace and redact video manipulations. There were many other factors that could have led participants to correctly identify more valuables in the scene. One such variable was the fact that although the valuable items themselves may have been filtered, other related items (e.g. cords from laptops, CPU attached to desktop computer monitor, speakers surrounding television) were not filtered and could be clearly identified around the filtered areas. It is possible that seeing these items was enough context for the participant to infer what valuables were being filtered out and correctly identify the number of valuables and their locations. Even with the redact video manipulation users seemed to use the surrounding environment to deduce the locations and details of the valuables. A snippet from a typical response exemplifies this problem: "it looked like 2 laptops on chairs blocked by black squares." This is also a problematic response because ideally with the "Can't tell" privacy type we would not want users to know anything is being hidden at all.

Another issue that consistently arose with the video clips for "Can't tell" was that there seemed to be much confusion on what the valuables to be recorded really were. Many people mentioned speakers that were not supposed to be considered one of the valuables recorded. Also, it appears that a good number of participants considered the painting that was sitting against the wall near the desk to be a desktop computer. One response in particular seems to exemplify the confusion: "There are three different computers. One is immediately in front of you as you enter the room, while a second is sitting on a desk in the far left corner. If you look to the left, you'll see what appear to be two tablets sitting on chairs and a desktop computer on a desk." The "computer" that is noted as being directly in front is actually a television screen, but the second one mentioned is a computer and is recorded as the desktop computer. Likewise, the two tablets on the chairs are actually laptops. The most concerning misidentification was the desktop computer on the desk, which was in reality only a painting and not an electronic valuable at all. Many responses recorded there being two desktop computers, but very few correctly identified the flat screen television. This was true across the different video manipulations, including those who had viewed the control clip.

For this reason, the number of total valuables recorded was not analyzed. It was apparent that there were too many responses that included objects that were not one of the four valuables in their total count. On top of that, many answers to the total number of valuables did not correspond to their next answer which described the recorded valuables. A participant would answer 0 for total number of valuables, but would then describe two of the four valuables well in the next answer. These inconsistencies led us to believe that solely examining the average number of privacy violations would give a better analysis of how well the "Can't tell" privacy expectation was upheld.

4.2 Can't Observe

The results from the "Can't observe" privacy type resembled very closely the results that were expected. The control video had the most privacy violations since participants could clearly view the entire hallway with nothing obstructing their view. Both blur and abstract video manipulations greatly reduced privacy violations, yet still led to violations a majority or half of the time respectively. Redact and replace completely upheld this privacy expectation since neither allowed any user to describe anything in the hallway. While both appear to work equally well, in practice the user might be inclined to choose one over the other based upon which manipulated video they find more appealing or realistic.

Although our results did not appear to be skewed, there was a discrepancy in the replace video manipulation that was used during the "Can't observe" scenario. In all other video clips the filters were placed over the entire door frame to fully block out the hallway. However, in using the replace video manipulation a filter was placed only over the cardboard box. Our main concern with this was that there might be more privacy violations since participants could still mention other details about the hallway (e.g. the stairs, banister, exit sign). Since no privacy violations occurred, it can be deduced that replacing only the box was enough to take attention away from the hallway, even though the question asked about it directly. In future implementations of this video manipulation the replace filter should be placed

over the whole door frame to fully protect from privacy violations, and not rely on only taking attention away from the scene and hoping the user will not note anything else in the background of the view. It is also hypothesized that if the filter had been placed over the entire doorway the view through the doorway would not have been as clear and understandable. It is likely that the replace video manipulation would have given the appearance that there was a second door, and the user may have been confused by what they saw in the image. For this reason, a redact video manipulation might be preferable because it gives an obvious cue to the user that there is something beyond the doorway, but they are not allowed to view it.

4.3 Can't Discern

The "Can't discern" privacy type was a bit more complex than the other two because it had multiple components to verify that it was both being upheld and still not overstepping its goal and hindering awareness that need not be blocked.

In examining the average number of privacy violations that occurred per video manipulation, it is important to note that, similar to the findings in the "Can't tell" scenario, both the control video and the video using the abstract manipulation had the highest number of privacy violations. Again, one might infer that the large number of privacy violations using the abstract video manipulation occurs because this video manipulation emphasizes heavy lines and borders like those that make up the text which reveal the book's title, while the control clip shows the text as clearly as one would see it without using a robot.

The blur video manipulation only experienced an eighth of one privacy violation on average. Even though this was a low average number of privacy violations, it is still not as low as the zero privacy violations that the replace and redact video manipulations incurred. At the same time, this difference in privacy violations was not found to be significant. One possible reason that the blur video manipulation may have still allowed users to discern the titles of the filtered books could be due to the intensity of the blur filter. It is possible that a blur filter with a higher intensity would better protect the titles of the novels. Another factor that played a role in identifying the filtered books was the fame and notoriety associated with these particular books. Recently both of our filtered novels have gained fame in the media and our society, with both novels having sold over 100 million copies worldwide and each either has been or is in the process of being made into a movie. Due to the overwhelming prevalence of these books in our society, it is possible that participants did not need to be able to read the title to recognize the book. With their iconic cover art, it is possible the participant could deduce which book was being filtered out as long as they could still get a sense of what the illustration on the cover looked like. In future work, one could further investigate how a stronger blur filter might work effectively if it not only blurred the text, but also the cover art to the point where it was unrecognizable.

As far as privacy violations are concerned it is clear from our results that the blur, redact, and replace video manipulations work best with making privacy violations impossible. However, the "Can't discern" privacy type also expects that the user can still correctly identify the quantity of books on the shelf, including the filtered books. In this category, replace scored the lowest averaging less than two books identified. The redact filter did not do significantly better than replace, but the blur filter did have a significantly higher number of books that were identified on average as compared to replace. Although the difference between blur and redact was not significant, this does point to the fact that blur would be preferable to replace, at least in upholding this aspect of the privacy expectations. Abstract and control clips averaged the most number of books recorded with answers right around 3.5. Therefore, it seems that the more privacy violations that the video manipulation had on average, the better the video manipulation also did with correctly identifying the number of books on the shelf.

As for accuracy with distinguishing the titles of the non-filtered books, all five video manipulations seemed to do relatively well. However, amongst them control had the lowest average, perhaps because participants focused too much on the other filtered titles and they did not as easily notice the two unfiltered ones. Abstract, blur, and redact video manipulations were incredibly similar in their performance, which shows that although they do not yield perfect results, there was definitely no distraction that took away from the unfiltered titles when filtering the other two books. Overall, the redact video manipulation had the highest average of correctly identified unfiltered titles, and there is a significant difference between the redact and control video manipulations in this category. However, there was no significant difference found between redact, replace, or blur in distinguishing titles of the non-filtered books, so nothing can be concluded from this as to which is the best of the three. Therefore, the only category in which one of the three was significantly better than another was in determining the correct number of books on the shelf. For this reason, one could make the case that blur would be preferable to replace. Although the redact and blur video manipulations had no significant difference between them in any measurement of upholding the "Can't discern" privacy type, we have concluded that for most "Can't discern" privacy type situations blur might be preferable since it has a more subtle appearance. One can hypothesize that using a redact filter might draw unwanted attention to the region that is to be indiscernible, while a blur filter could be written off as simply bad video quality and would not be alarming or confusing as a black box that blocks the user's view.

The case for redact in this video clip is also very unique in that the boxes used to redact the books from the bookshelf were a very similar color to the bookshelf itself. This led to an effect that looked quite similar, although not identical, to replace. Perhaps if the boxes used had been a color such as red that created a greater contrast against the back drop of the bookshelf, or if the bookshelf itself had been a bright color to contrast with the boxes used for redacting, the results would more likely show a greater difference between results for the replace and redact video manipulations. Also, the shape of the boxes was similar enough to the shape of a standard book that it could be deduced that the boxes were hiding books as well. Varying the shapes of the redaction filter used could lead to less participants viewing

the box as a representation for the book.

4.4 Limitations

As this is an initial step in determining which video manipulations most effectively uphold different privacy types, there is much more work that can be done to further support these findings. One limitation of this study was that each privacy type was manipulated using all five video manipulations, but there was only one video clip used for each specific condition and manipulation combination. To help generalize these findings, one could experiment with different conditions (e.g. lighting condition, number of objects/focal points in a scene, settings or calibrations for the video manipulations) to create more video clips and better analyze which video manipulation would be most appropriate for most environments that attempt to uphold a certain privacy type. There are also other kinds of video manipulations that were not used in this study that could be used and tested to see if they would do any better with upholding certain privacy types. Another idea would be to try some video manipulations in combination to see if their combined effect was any more effective in reducing privacy violations for certain privacy types. This study also made mention of how some video manipulations might be preferable for certain privacy types due to the appearance they gave to the user, and further investigation could be done to examine which type the proximal user prefers to show the remote user when multiple video manipulations are equally effective in protecting from privacy violations.

In examining the number of correctly identified non-filtered book titles, our study only hinted at how implementing these video manipulations might affect task performance. There is still work to be done to determine to what extent these video manipulations will impact task performance, particularly when recognizing filtered objects is not the primary task. Our study was also limited in that participants could only passively view a video clip instead of controlling the RPS on their own to further investigate a scene and search for answers. An extrapolation of this study would be to use these video manipulations in a live video implementation where users could teleoperate a RPS and examine filtered objects in real time. For this study we began creating software to help us accomplish this task, but due to major concerns with localization, which was creating filters that were not as accurate as they should be, it was decided that using Adobe After Effects would more effectively suit the needs of this study.

4.5 Design Implications

Our results show that using video manipulations with RPS can help protect privacy. Specifically, it would be recommended that the user take advantage of the redact or replace video manipulations if the object to be filtered out was of the "Can't tell" or "Can't observe" privacy types. The blur video manipulation should be used for best results with an object that has "Can't discern" privacy expectations, such as text or details that shouldn't be identifiable on an object. Although our scenarios only tested one privacy type per video clip, our results suggest that in a real scene where multiple privacy types are being used there would be multiple video manipulation techniques needed for different objects or areas.

5. CONCLUSION

In this study we examined four different video manipulations along with a control to create five video manipulations that we applied to three different scenarios. Each scenario had a goal to uphold a certain privacy type. The data we collected supported our hypothesis that video manipulations are effective in obscuring details to protect privacy expectations. We found that "Can't tell" was best upheld by either a redact or replace video manipulation. Similarly, "Can't observe" was well met when using the redact and replace video manipulations. In the "Can't discern" scenario the blur, replace, and redact video manipulations worked well in different ways, but we concluded that blur performed best overall for protecting this privacy expectation.

6. REFERENCES

[1] InTouch Technologies, Inc. InTouch telemedicine system, August 2012. http://www.intouchhealth.com/.

[2] Open Source Robotics Foundation. TurtleBot 2, August 2012. http://turtlebot.com/.

[3] Suitable Technologies, Inc. Beam remote presence system, August 2012. https://www.suitabletech.com/.

[4] VGo Communications, Inc. VGo robotic telepresence for healthcare, education, and business, August 2012. http://www.vgocom.com/.

[5] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3):24:1–24:11, July 2009.

[6] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, pages 1–10, New York, NY, USA, 2000. ACM.

[7] M. Boyle, C. Neustaedter, and S. Greenberg. Privacy factors in video-based media spaces. In S. Harrison, editor, *Media Space 20 + Years of Mediated Life*, Computer Supported Cooperative Work, pages 97–122. Springer London, 2009.

[8] A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro. A comprehensive framework for image inpainting. *Trans. Img. Proc.*, 19(10):2634–2645, Oct. 2010.

[9] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella. Suggestive contours for conveying shape. *ACM Trans. Graph.*, 22(3):848–855, July 2003.

[10] D. DeCarlo and A. Santella. Stylization and abstraction of photographs. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, pages 769–776, New York, NY, USA, 2002. ACM.

[11] I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. *ACM Trans. Graph.*, 22(3):303–312, July 2003.

[12] A. Edgcomb and F. Vahid. Privacy perception and fall detection accuracy for in-home video assistive monitoring with privacy enhancements. *SIGHIT Rec.*, 2(2):6–15, Sept. 2012.

[13] E. Eisemann, H. Winnemöller, J. C. Hart, and D. Salesin. Stylized vector art from 3d models with region support. In *Proceedings of the Nineteenth Eurographics conference on Rendering*, EGSR'08, pages 1199–1207, Aire-la-Ville, Switzerland,

Switzerland, 2008. Eurographics Association.

[14] T. Gerstner, D. DeCarlo, M. Alexa, A. Finkelstein, Y. Gingold, and A. Nealen. Pixelated image abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, NPAR '12, pages 29–36, Aire-la-Ville, Switzerland, Switzerland, 2012. Eurographics Association.

[15] A. A. Gooch and P. Willemsen. Evaluating space perception in npr immersive environments. In *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, NPAR '02, pages 105–110, New York, NY, USA, 2002. ACM.

[16] B. Gooch, E. Reinhard, and A. Gooch. Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph.*, 23(1):27–44, Jan. 2004.

[17] J. Herling and W. Broll. Pixmix: A real-time approach to high-quality diminished reality. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR '12, pages 141–150, Washington, DC, USA, 2012. IEEE Computer Society.

[18] H. H. J. Kim, C. Gutwin, and S. Subramanian. The magic window: Lessons from a year in the life of a co-present media space. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, GROUP '07, pages 107–116, New York, NY, USA, 2007. ACM.

[19] J. E. Kyprianidis. Image and video abstraction by multi-scale anisotropic kuwahara filtering. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, pages 55–64, New York, NY, USA, 2011. ACM.

[20] J. Lu, P. V. Sander, and A. Finkelstein. Interactive painterly stylization of images, videos and 3d animations. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*, I3D '10, pages 127–134, New York, NY, USA, 2010. ACM.

[21] D. Mould. Texture-preserving abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, NPAR '12, pages 75–82, Aire-la-Ville, Switzerland, Switzerland, 2012. Eurographics Association.

[22] S. Olsen and B. Gooch. Image simplification and vectorization. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, pages 65–74, New York, NY, USA, 2011. ACM.

[23] E. Paulos and J. Canny. Social tele-embodiment: Understanding presence. *Auton. Robots*, 11(1):87–95, July 2001.

[24] P. L. Rosin and Y.-K. Lai. Towards artistic minimal rendering. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, NPAR '10, pages 119–127, New York, NY, USA, 2010. ACM.

[25] S. Rusinkiewicz, D. DeCarlo, and A. Finkelstein. Line drawings from 3d models. In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, New York, NY, USA, 2005. ACM.

[26] C. Sauvaget and V. Boyer. Comics stylization from photographs. In *Proceedings of the 4th International Symposium on Advances in Visual Computing*, ISVC '08, pages 1125–1134, Berlin, Heidelberg, 2008. Springer-Verlag.

[27] J. Schiff, M. Meingast, D. Mulligan, S. Sastry, and K. Goldberg. Respectful cameras: detecting visual markers in real-time to address privacy concerns. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 971–978, Oct 2007.

[28] M. Singh and S. Schaefer. Suggestive hatching. In *Proceedings of the Sixth international conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'10, pages 25–32, Aire-la-Ville, Switzerland, Switzerland, 2010. Eurographics Association.

[29] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum. Image completion with structure propagation. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 861–868, New York, NY, USA, 2005. ACM.

[30] A. Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9(1):23–34, 2004.

[31] M. Vijay Venkatesh, S.-c. S. Cheung, and J. Zhao. Efficient object-based video inpainting. *Pattern Recogn. Lett.*, 30(2):168–179, Jan. 2009.

[32] H. Winnemöller. Xdog: advanced image stylization with extended difference-of-gaussians. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, pages 147–156, New York, NY, USA, 2011. ACM.

[33] H. WinnemöLler, J. E. Kyprianidis, and S. C. Olsen. Special section on cans: Xdog: An extended difference-of-gaussians compendium including advanced image stylization. *Comput. Graph.*, 36(6):740–753, Oct. 2012.

[34] H. Winnemöller, S. C. Olsen, and B. Gooch. Real-time video abstraction. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 1221–1226, New York, NY, USA, 2006. ACM.

[35] H. Xu, N. Gossett, and B. Chen. Abstraction and depiction of sparsely scanned outdoor environments. In *Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'05, pages 19–27, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[36] H. Zhao, X. Jin, J. Shen, L. Shen, and R. Pan. Fast shape-simplifying image abstraction using graphics hardware. In *Proceedings of the 4th International Conference on E-Learning and Games: Learning by Playing. Game-based Education System Design and Development*, Edutainment '09, pages 390–398, Berlin, Heidelberg, 2009. Springer-Verlag.

7. APPENDIX: CODING SCHEME

To make sense of the recorded data we received it was necessary that the data be sorted and categorized, particularly into which answers should be recorded as privacy violations and which should not. As the data was analyzed guidelines were put into place on which responses would be considered privacy violations and which would not.

For the "Can't tell" privacy type, privacy violations were recorded for any correct identification of one of the four valuables. One privacy violation was recorded for any mention of a television or LED screen. Another one privacy violation was recorded for any mention of a desktop computer, computer, or CPU. Up to two privacy violations were recorded for any mention of laptops or personal computers. The number of privacy violations for laptops was dependent on the number of laptops mentioned in the recorded text, and then further relied upon the total number of valuables recorded when the text was unclear. The ideal number of privacy violations is zero for this privacy type. It is important to note that any mention of a monitor was not counted as a privacy violation for the television, but was recorded as a privacy violation for the desktop if no other mention of the desktop was made. Also, many of the responses had more items listed than the valuables that were considered privacy violations. A frequent object listed was the desktop/monitor/computer/laptop on the far left that was actually just a painting on a shelf leaning against a wall.

In the "Can't observe" privacy condition scenario, data was recorded about what, if anything, participants could see down a hallway. Most typically a privacy violation was recorded when any mention to the cardboard box was made. This box was also referred to several times as something being made of wood, a carton, and a package. An atypical response that was considered a privacy violation was the rare instance where the stairs were heavily implied in the recorded response when the participant mentioned that this was recorded on the second floor of a house. Since this answer demonstrated that the participant must have taken note of the stairs, which are in the hallway, we considered this a privacy violation.

There were a good deal of commonly recorded items that were not seen as privacy violations since they were not located in the hallway, but rather inside the room: a chair, folding tables, a door stopper, the door itself, and the carpet. It is important to note that any mention of a door was interpreted as a reference to the main wooden door that opens to the hallway. It was unclear whether any of the responses were referring to the door shape that could be seen at the end of the hallway, and due to the lack of specificity in responses it was determined that there was a greater likelihood that participants would have been referring to the wooden door.

Another pertinent detail in recording privacy violations was that we did not consider any mention of a black box as a privacy violation when the redact video manipulation was being used because the assumption was made that this was referring to the filter instead of the cardboard box. This black box was also referred to as a black mass, black screen, and stone. While rare, there were a good number of items that were identified in participants answers that could not be traced back to anything in the video clip. Examples include an inverted cup, a glass bowl, a sofa, and a metallic button, and none of these were considered privacy violations since they did not exist or resemble anything that did exist in the hallway.

With video clips concerning the "Can't discern" privacy type, a privacy violation was recorded once for any mention of "Twilight" and once for any mention of "Fifty Shades of Grey." It should be noted that spelling did not have to be exact, but rather as long as the title was fairly close or had any of the main words from the title (i.e. "fifty" "shades" "grey" "twilight") the privacy violation was recorded. However, there was a unique case where a participant incorrectly identified a title as "Breaking Dawn," which is another novel in the "Twilight" saga, and this was not recorded as a privacy violation because the title was clearly a guess and incorrect at that. In assessing accuracy in reading the unfiltered titles a correctly identified book title was recorded for any mention or relatively close spelling to "ARToday" and another was recorded for any mention of any of the words in the title or relatively close spellings to "Visualization: The Second Computer Revolution."