Usability of Augmented Reality for Revealing Secret Messages to Users but Not Their Devices

Sarah J. Andrabi
UNC Chapel Hill
sandrabi@cs.unc.edu

Michael K. Reiter
UNC Chapel Hill
reiter@cs.unc.edu

Cynthia Sturton
UNC Chapel Hill
csturton@cs.unc.edu

ABSTRACT

We evaluate the possibility of a human receiving a secret message while trusting *no* device with the contents of that message, by using visual cryptography (VC) implemented with augmented-reality displays (ARDs). In a pilot user study using Google Glass and an improved study using the Epson Moverio, users were successfully able to decode VC messages using ARDs. In particular, 26 out of 30 participants in the Epson Moverio study decoded numbers and letters with 100% accuracy. Our studies also tested assumptions made in previous VC research about users' abilities to detect active modification of a ciphertext. While a majority of the participants could identify that the images were modified, fewer participants could detect *all* of the modifications in the ciphertext or the decoded plaintext.

1. INTRODUCTION

In the face of massive surveillance by nation-state-level actors (e.g., [22, 30]), including the implantation of surveillance functionality in commodity device firmware (e.g., [17]), it appears that truly private electronic communication is as challenging today as ever. At the core of this problem are the complex cryptographic operations that must be performed to encrypt and decrypt messages: These operations are too complex for humans to perform themselves, and so they must use devices to do so—perhaps devices that might have already had their privacy compromised. To achieve truly private communication, then, it would seem necessary to eliminate devices from the trusted computing base (TCB), i.e., to have humans themselves perform the cryptographic operations.

History is rife with examples of private communication performed by humans without devices, usually because capable devices were not yet available. While most of these ciphers are trivially breakable today, a notable exception is the one-time pad, which is both perfectly private and has encryption and decryption functions involving only exclusive-or operations [16]. One-time pads were a method of choice for spies in World War II for protecting communications with their home countries, performing the exclusive-or of messages manually against keys encoded on, e.g., a paper pad that they brought with them (e.g., [29]). This idea was modernized in 1994 with the invention of *visual cryptography* (VC) [26], in which keys are encoded on visual transparencies and manual exclusive-or operations are replaced with overlaying these transparencies on suitably encoded ciphertexts to reveal their contents.

The practical difficulties associated with one-time pads are well known. Most notable is that they require a quantity of key material (in the above proposals, on paper tape or transparencies) equal to the total size of all messages to be transmitted. In this paper we explore the feasibility of making the one-time pad, and specifically its implementation in VC, practical using *augmented reality* head-mounted displays (ARDs) such as Google Glass, Epson Moverio, Sony SmartEyeGlass, or Microsoft HoloLens, while still ensuring that no single device is trusted with the contents of a message to the user. We evaluate the efficacy of storing and rendering one-time pads (encoded for use in VC) in an ARD, which the user visually aligns over an encoded message rendered on another display to reveal its contents. If workable, this design should largely eliminate the problems associated with storage of one-time pads and in fact can support their generation pseudorandomly (e.g., from a secret key per sender) with little practical loss of security.

In this paper we show that VC via ARDs is in fact possible and reasonably usable, first using a small pilot study with Google Glass and then a formal user study using the Epson Moverio. For conveying messages we used individual letters and numbers encoded as images. Of the 30 participants, 26 in the formal user study decoded numbers and letters with 100% accuracy. The other four participants decoded at least 80% of the alphanumeric plaintexts accurately. The median time taken by participants to identify the decoded letters and numbers was 8.88 seconds. As indicated by our study, however, several challenges need to be addressed including the small field of view of ARDs, head tracking, and image stabilization.

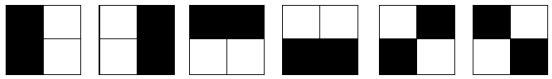
Moreover, since numerous communication protocols involving VC have made assumptions about users' abilities to detect active modification of a ciphertext or the plaintext that results from it (e.g., [25, 31, 15, 9, 6]), we also evaluate users' abilities to do so. Through the Epson Moverio study, we assess whether users were able to identify invalid decoded plaintexts resulting from malicious modifications to the ciphertext. A large majority of our participants were able

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

to detect the modifications, even though fewer were able to identify all of the changes in an image. Even in the presence of the *least* number of modifications (depending on the study) to the recovered plaintext, more than 80% of the participants detected the presence of the modifications.

We also conducted an online user study to analyze users' abilities to detect changes in ciphertexts directly. Again, a large majority of participants detected that the images were modified. We explored the extent of modification needed in order for a user to detect it, and based on our current implementation, small modifications were detected by most of the participants. Our findings indicate that augmented reality provides a promising use case for visual cryptography.

The paper is organized as follows. Section 2 presents background on visual cryptography and augmented reality. In Section 3, we present our methodology for implementing visual cryptography using augmented reality displays (ARDs). We describe our user studies in Section 4 and present the results in Section 5. Finally, we present the limitations of our work in Section 6 and the conclusion and future work in Section 7.

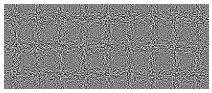


(a) Valid blocks for an image share

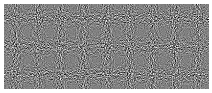
(b) Recovering a black block

(c) Recovering a white block

Figure 1: Visual cryptography block representations



(a) First share



(b) Second share

(c) Overlaid shares

Figure 2: Example of visual cryptography

2. BACKGROUND

In this section we give a brief introduction to visual cryptography (Section 2.1) and augmented reality (Section 2.2). We then discuss an approach to combine these technologies that enables message recovery without trusting any individual device with the message (Section 2.3) and that forms the basis of our implementation (which is further described in Section 3).

2.1 Visual Cryptography

Visual cryptography [26] is a cryptographic secret-sharing scheme where visual information is split into multiple shares, such that one share by itself is indiscernible from random noise. (Equivalently, one of these shares constitutes a one-time pad as discussed in Section 1, and the other represents the ciphertext of the secret message.) The human visual system performs a logical OR of the shares to decode the secret message being shared.

In visual cryptography, a message share is represented by a rectangular grid of *blocks*, each block itself being a square 2×2 grid of *regions*. Each block in a message share includes two black regions and two transparent regions, as shown in Figure 1a. The ciphertext image is decoded by overlaying two image shares on top of each other. Consider the block in Figure 1b. If one share has a block with the two top regions black and the two bottom regions transparent and the second share has the reverse—two top regions transparent and two bottom regions black—then when the two shares are overlaid the resulting block will appear solid black. This is the decoded image. Recovering a white block is done in a similar manner, as shown in Figure 1c. Note that the recovered "white" block is partially black and partially transparent. This "gray" block is used to represent white.

Consider Figure 2, in which the secret message is decomposed into two shares shown in Figures 2a–2b, such that on stacking the two, the user sees Figure 2c. Even though the contrast of the image has been degraded by 50% from regular black-on-white text due to the gray blocks, the human visual system can still identify the content of the secret message ("123") easily.

When the human visual system does the decryption, no computation is needed; it is a visual OR operation. That is, if a given region is black in either of the overlaid shares, the user sees a black region. Otherwise the user sees a white region. Even though the human eye performs a logical OR operation per region, the result is an XOR operation per block. Hence, the security of the one-time pad is retained. In particular, an individual share reveals no information about whether a specific decoded plaintext block is black or white. A disadvantage of the process is the loss in contrast of the decrypted message, which in turn affects the clarity of the recovered secret.

While an adversary in possession of only one share learns no information about the corresponding secret message, he might still compromise the integrity of the secret message if he is able to modify one of the shares in a meaningful way. A method to defend against this attack, introduced by Naor and Pinkas [25], is to use part of the secret message to convey the desired information and deliberately leave part of the secret message as a solid "color," i.e., every block consisting of two black regions and two transparent regions ("white") or every block consisting of four black regions ("black"). If the adversary modifies one of the blocks in one share by swapping the white and black regions, the result will be to change the color of the block in the decoded plaintext image. If the changed block happens to be in the dedicated non-content region, the recipient will know that one of the shares was compromised. If half of the secret message is a dedicated non-content region then the adversary has only a 50% chance of successfully modifying a share without noticeably modifying the non-content region of the decoded plaintext. One of the questions we examine in our user study is whether a user will always notice a single block of the wrong color in a non-content region of the decoded plaintext image.

The question of how to convey a message using VC such that any modification to a share by an attacker will be detectable has been well studied [6, 31, 9, 15]. These proposals all make some assumptions about the capabilities of the human visual system (HVS), which we attempt to validate in our user studies. The questions we explore are: whether a

human will detect a block in a message share that does not have exactly two black regions; whether a user will notice blocks in the decoded message that have either one or three black regions; whether a user will notice blocks of the wrong color in a non-content region (as described in the previous paragraph); and whether a user will notice blocks that subtly change the semantic meaning of the decoded message.

2.2 Augmented Reality

Augmented reality (AR) systems supplement reality by combining a real scene viewed by the user and a virtual scene generated by the computer [24, 5]. The AR system superimposes a user's environment with computer-generated information, making the user's environment digitally manipulable and interactive.

The most common realization of AR is with see-through head-mounted displays (HMDs). An HMD is a device that is typically attached in close proximity to the eyes with a display system that couples a digital image with the environment [11]. HMDs enable the merging of virtual views with a user's physical environment, for example by enabling a physician to see the 3D rendering of CT images of a patient superimposed onto the patient's abdomen [4]. We refer to these augmented reality HMDs as augmented reality displays (ARDs) for brevity. ARDs have started to ship commercially, with the latest examples being Microsoft's HoloLens [23] and Google Glass [13].

The human visual system forms a core element of these near-eye ARDs, especially when considering the performance and usability of the system. The field of view (FOV) of the human eye describes the dimensions that can be seen by the eye at any given instant of time. For the human eye, this field of view is approximately 150° horizontally and 120° vertically. The binocular overlap region, within which a target is visible to both eyes, measures about 114° [11].

Many commercially available binocular ARDs like the Lumus DK-32, the Optinvent ORA-S, and the Epson Moverio BT-200/100 are limited to a FOV of less than 60°. Maimone et al. [21] developed a near-vision see-through ARD with a field of view of approximately 110° diagonally, and so far this is the maximum FOV that has been achieved with AR displays. Google Glass has a 14° monocular FOV and the Epson Moverio has an FOV of 23° [28]. The major limitation resulting from a small field of view is the inability to show finer details on ARDs.

There are several challenges in overlaying information onto a user's FOV, such as contrast sensitivity and color perception. The color perceived by users wearing an ARD is affected by the transparency of the graphics being rendered. In addition, some displays significantly diminish the brightness of the real world so that the graphics on the device's display need not be bright [11, 20]. These issues may result in changes in contrast and hues perceived by the user. For example, since black is rendered as transparent to allow the real world to be seen, dark colors may be perceived improperly, and bright colors that are intended to occlude the real-world background may not be able to do so [20].

Another challenge in head-mounted ARDs is head jitter. As a user tries to focus on an object, slight head movements translate to large movements of the displayed object. This is particularly important for our scenario, as the users are trying to align two images. However, these problems can be addressed using built-in image stabilizers, motion sensors and head tracking [8].

2.3 Using Devices to Implement VC

While our work is, to our knowledge, the first to explore using augmented reality head-mounted displays to implement VC, other works have suggested the use of specialized devices to replace the transparencies envisioned by the original VC authors. In particular, Tuyls et al. [32] suggest a transparent display device that the user holds or attaches to her computer display. To reveal a plaintext to the on looking user, the transparent display device renders one image share, and the underlying computer display renders the other. Neither the transparent display device nor the underlying computer display individually gains any information about the plaintext image that the user sees.

While the design we test replaces the transparent display device with an augmented reality display (ARD), otherwise it is conceptually similar and borrows underlying elements from the Tuyls et al. design. In particular, a sender generates and securely transmits image shares in our envisioned deployment in the same way (subject to some ARD-specific constraints, discussed in Section 3), and the receiving user can read the message without trusting either the ARD or the computer display with the contents of the message. Additionally, Tuyls et al. demonstrate how the user may respond without divulging her response to her devices.

We stress, however, that our contribution is not in the conceptual novelty of our design of a VC system using ARDs, but rather in a study of the usability of our approach for doing so. To our knowledge, we are the first to undertake such a usability study.

3. IMPLEMENTATION OF VC USING AUGMENTED REALITY

As originally envisioned, visual cryptography used printed transparencies consisting of black (opaque) and transparent regions [25, 2]. However, ARDs render images on a glass prism via light projection. Black regions are represented by the absence of illumination and are transparent, while formerly transparent regions are now rendered through the projection of white light [11, 20]. Thus, the notion of transparency and opacity in visual cryptography achieved through ARDs is inverted with respect to traditional visual cryptography using printed transparencies.

In our implementation, each share of the secret message is composed of 2×2 blocks. Each block in a share includes two illuminated (white) regions and two transparent (black) regions. When two images are overlaid, the regions combine as in Figure 3. In our initial experiments, we tried using colors other than white to render the opaque regions, but we found white was the most effective.

One share is sent to the user's ARD, and the other is sent to a display in the vicinity of the user. To superimpose them, the user views the visual share through the ARD displaying the second visual share. The two shares thus overlaid reveal the secret message. Below we describe how we designed our system to work with Google Glass and the Epson Moverio.

3.1 Google Glass

Google Glass is an Android based augmented reality device that is Wi-Fi and Bluetooth enabled. It provides a

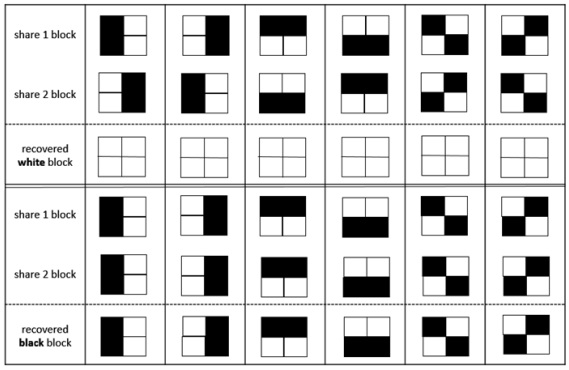


Figure 3: Construction of visual cryptography scheme for augmented reality. The secret block is obtained by superimposing the two corresponding shares, looking through an augmented-reality head-mounted display.

monocular AR experience with a small display screen for the right eye. Glass enables users to view content including text, images, and videos. It applies transformations on images that are displayed on it such that they are scaled in both the horizontal and vertical directions. To counterbalance that transformation, we scale the images to 500×600 pixels and pad them with a black background, such that our final images are 1200×1200 pixels on a desktop monitor. The display of Google Glass with a resolution of 640×360 pixels (equivalent of a 25 in (64 cm) screen from 8 ft (2.4 m) away) [13] is very small, with a field of view of 14°. We found the maximum image size for the secret message that could comfortably be displayed on Google Glass was a grid of 5×5 blocks, with each block containing two black and two white regions.

In addition to the above transformations, we added alignment markers (red borders around blocks, as shown in Figure 4) to help users align the two images. We also blacked out the areas that contained no useful information to further aid image alignment and account for head jitter.

3.2 Epson Moverio

Our second study used an Epson Moverio BT-200 Model H423A—an Android based augmented reality device, with display screens for both eyes, i.e., binocular display. It has a resolution of 960×540 pixels (equivalent to 80 inch image from 16.4 feet away) [28, 10]. The Epson Moverio has a field of view of 24° and enables 2D and 3D viewing. The 2D mode is similar to the viewing mode in Google Glass and it is the mode we used in our study. We used two image sizes containing 7×7 blocks and 9×5 blocks, respectively. These sizes were selected based on trial and error for what was a comfortable and easy image size for alignment to be done by untrained, first-time participants. The investigator for the study, because of training effects, was able to pack up to two letters or four numbers into an image and still able to align and identify the characters in the images.

Even though the information being conveyed is limited, there is a marked improvement over using Google Glass. We believe this is mostly because of the Epson's larger field of view. With a larger field of view and higher resolution more information can be conveyed through one image.

One limitation of the two headsets used, or some of the others that are commercially available, such as the Optinvent ORA-S, Recon Jet, Vuzix M100, and Lumus, is absence of any form of built-in head tracking and image stabilization. This was one of the major challenges that we faced during the Google Glass study: even slight head movement misaligned the images. We elaborate more on this as we describe the user studies.

3.3 Threats introduced by ARDs

As discussed in Section 1, our goal is to implement VC in such a way that realistic attackers can access (to read or modify) only one of the two visual shares needed to reveal the secret message. While our focus in this paper is on the usability of ARDs for this purpose, we pause to consider challenges in ensuring this property with the two ARDs that we have used in our implementations.

One type of attacker that poses challenges with either of our ARDs is a third party in proximity of the user who might photograph both the physical share (i.e., the share in the user's physical proximity) and, with high definition, the share displayed in the ARD. Since this attacker would not see these two images aligned as the intended message recipient does, he would then need to reconstruct the secret message offline using optical decoding techniques or manual effort. Such attacks, however, are not far-fetched: the possibility of observing the image displayed in Google Glass is well-known (e.g., "bystanders who see the tiny display screen above the right eye" [19] and "the screen is just visible from the outside" [12]) and similarly intricate optical decodings have been demonstrated (e.g., [18]). That said, this threat is equally present—if not more so—in traditionally envisioned deployments of VC that use, e.g., physical transparencies to reveal a secret message. Using ARDs, it is presumably easier for the user to ensure that her ARD is not being observed (e.g., by shading it with her hands) during message recovery.

A related concern is that several ARDs (e.g., Google Glass, HoloLens, Epson Moverio BT-200, and Vuzix M100) have a front-facing camera. (Several others, like the Epson Moverio BT-100 and Laster See-Thru, do not.) As such, if an ARD with a front-facing camera is compromised by malware, it could potentially use the front-facing camera to photograph the other share and combine it with the share given to the ARD to display, revealing the private message. It is therefore necessary that, for an ARD with a front-facing camera such as Glass, the camera lens be covered physically while the physical share is displayed. We did not incorporate this step into our user study but would recommend doing so in actual deployment. Similarly, if the physical share is displayed on a computer display, that computer should not have a camera facing the user that might capture the share displayed in the user's ARD (or else that camera should be physically covered).

Finally, ARDs like Google Glass are not really standalone devices; rather, they share their information with the service provider to which they are tethered—Google in the case of Glass. There is a risk that this service provider could both extract image shares from the ARD and combine them with the corresponding other image shares that it receives otherwise (e.g., sent to the user's gmail account). Addressing this risk presumably involves the sender conveying ciphertexts to the user via a different service provider than the

one with which the ARD shares its contents, or to adopt an ARD that does not so freely share its contents with a service provider.

4. USER STUDIES

We conducted three user studies: a pilot user study using Google Glass, an improved formal study using the Epson Moverio ARD, and an online user study using Mechanical Turk. All our user studies were reviewed and approved by our university's IRB. In the pilot user study we gauged participants' ability to align visual shares and discern the overlaid information from noise. In the Moverio user study, participants decoded letters and numbers, and we also investigated their ability to recognize modifications to the decoded plaintexts. In the Mechanical Turk study we assessed users' ability to detect modifications to individual image shares. The pilot user study and the formal user study were conducted over a period of two weeks each. The formal user study was conducted three months after the pilot user study.

4.1 Participant Demographics

The pilot study had 31 participants: 22 male and 9 female. Out of the 31 participants, 19 were aged between 18-25, 9 between 26-35 and 3 between 36-45. Of the 31 participants, 8 wore prescription glasses. In the formal study, there were 30 participants, of which 23 were male and 7 were female. The age groups included 17 participants between 18-26, 11 between 26-35, and 2 between 36-45. Of the 30 participants, 15 wore prescription glasses. Both participant groups consisted mostly of university students, staff, and faculty members. Recruitment was done through emails to department and student group mailing lists. We recruited participants this way because participants had to come to our office location (on our university campus) for participating in the study.

Given the nature of these studies, we believe the most important bias in our participant pools is that toward younger participants. While we do not claim that these participant groups enable us to generalize our results to the general population, we are hopeful that they should be representative of populations represented by these age groups. Results would vary for other population groups and can be explored as part of future work.

For the Mechanical Turk study, we had 50 participants, with 28 male and 22 female participants. The age groups were 14 between 18-25, 20 between 26-35, 12 between 36-45, and 4 between 46-55.

4.2 Pilot User Study

4.2.1 Training

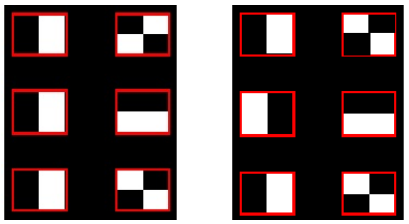
To acclimate users to the experience of using Google Glass, each user first underwent a brief training session on how to perform image navigation and alignment on the device. The training set comprised two stages: First, three photos not related to the study were displayed to demonstrate the act of navigation between pictures in Google Glass' image browser. Then, three images containing blocks (see Figure 4) were presented to the user as practice for the actual task. The users were asked to align the blocks on Glass and the monitor screen and report which blocks were white. During this training phase, the participants were provided guidance if they responded incorrectly.

In the training phase and in task 1 below, each recovered message was a Braille character. We used Braille characters because they fit well in a 5×3 block grid we used with Glass. A Braille character is a three-row, two-column grid with raised "dots" in some cells that can be felt. The locations of these dots in the grid indicate a character (e.g., see [1]). We used a white block to correspond to a raised dot and a black block to correspond to the absence of a raised dot. Our choice of Braille for this pilot study was primarily due to the simplicity of this mapping of Braille characters to VC, and of course not because we intend for blind persons (the primary users of Braille) to make use of VC or AR. The participants were not told that the recovered messages should be Braille characters and no knowledge of Braille was required.

4.2.2 Task Descriptions

After training, the participants were given three tasks. In each task, each participant was given a set of image pairs (one in Glass and one on the computer monitor) and asked to overlay and align the images on Glass and the monitor to figure out which blocks were completely white. The participant then had to note this down on a sheet of paper before moving on to the next pair of images in the set. A 3×2 grid was drawn on the paper for each image pair, and the participants marked the observed white blocks on the grid. Each participant filled out a questionnaire after completing each of the three tasks and then a final questionnaire after all tasks were completed.

Each participant was timed for each image pair presented. The timed duration included the time taken to navigate to the image on Google Glass as well as the monitor, align the two image shares, identify the white blocks, and note their locations on a sheet of paper.

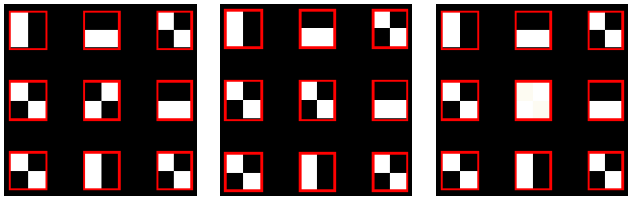


(a) First share (b) Second share

Figure 4: Image shares for Braille character "i"

Task 1: Task 1 had five image pairs; an example image pair is shown in Figure 4. The image pairs in this task were shares of randomly chosen English Braille characters. All the participants were presented a different set of image pairs. The participants had to write down which blocks in the recovered plaintext were white.

Task 2: This task was similar to the previous task but the images were shares of a 3×3 grid of blocks with at most one white block, e.g., see Figure 5. The location of the white block (if any) was chosen uniformly at random. There were five image pairs in this task, and the participants were asked to write down which block in each recovered plaintext (if any) was white by writing the white block's number, with the upper left corner being block one, the lower-right corner being block nine, and numbers incrementing along rows (like a telephone keypad). If there was no white block, the



(a) First share (b) Second share (c) Decoded plaintext

Figure 5: Image shares for the number 5 in pilot study (Section 4.2)

participant was instructed to write a zero. We chose this task design because the majority of the population is familiar with this number-pad pattern, and given the limitations of using VC with Google Glass in its current form, this was a viable way to convey numbers.

Task 3: The images used in this task had the same layout as in task 2. Each participant was given three pairs of images and asked to indicate which block (if any) of the recovered plaintext was white. However, unlike in task 2, all three visual shares were displayed side-by-side at the same time on the monitor. The participants still had to navigate the image shares on Glass, however, and overlay the shares on Glass with the shares on the monitor from left to right until completing the task. We chose this task design to test the ease of alignment when more than one physical share was displayed on the monitor. For this task all the participants were given shares for the same three numbers, which (for no particular reason) was the room number of the room where the study was conducted.

4.2.3 Results

Figure 6 presents the distribution of average plaintext recovery time per participant, i.e., averaged over all image pairs in the task indicated on the horizontal axis. A timer running in the background captured the time spent per image. As soon as a participant navigated to the next image, the timer for that image started.

Each box in this plot consists of three horizontal lines, which indicate the 75th, 50th (median), and 25th percentiles of the average time per participant. Whiskers extend to cover the lowest (highest) point within 1.5× the interquartile range of the lower (upper) quartile. There was no significant decrease in time spent per image pair as a participant progressed through an individual task.

There is noticeable variation in the amount of time spent by users per image pair, ranging from a few seconds to as large as 40 seconds. For identifying the white blocks in the recovered plaintext, the users not only had to attune themselves to what they are looking for but also use a new technology—all of the users were first-time users of Google Glass.

There is no relationship between error rates and timing data. Participants who identified the decoded plaintext incorrectly may or may not have spent little or more time on those visual shares. Participants who had a higher error rate did not on average take longer than participants who had no errors. A total of 9 participants out of 31 decoded at least one plaintext incorrectly. Table 1 shows the number of

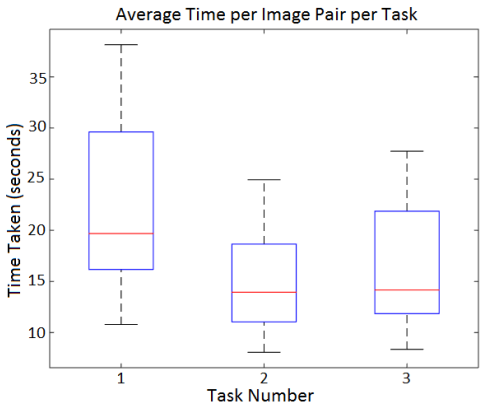


Figure 6: Box plot showing average plaintext recovery time per participant, i.e., averaged over the number of plaintext recoveries in the task on the horizontal axis, in the pilot study (Section 4.2).

	Task 1	Task 2	Task 3
All correct	22	28	27
1 incorrect	6	3	4
2 incorrect	1	0	0
3 incorrect	0	0	N/A
4 incorrect	2	0	N/A
5 incorrect	0	0	N/A

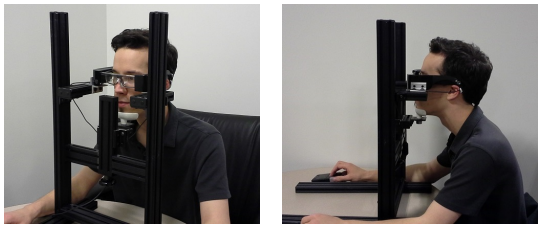
Table 1: Number of participants per error number in pilot study (Section 4.2)

participants that made the number of errors indicated in the leftmost column. A total of 403 image pairs were presented to the participants, out of which 16 occurrences were incorrectly identified. One interesting observation is that if the participant made no error in the first task, then they made no errors in the subsequent tasks.

It was harder for participants wearing prescription glasses to clearly see images on Glass and align them, as indicated by them on the questionnaires. Some participants reported discomfort using Google Glass as they were left-eye dominant. We moved to the Epson Moverio in the second study to address these issues. Many participants reported that the alignment was hard because their heads would not stay completely still and they were thus unable to figure out if a particular block of the recovered plaintext was white or not. As there is no form of head tracking or image stabilization in Google Glass, this was a problem. We modified the design of the second user study based on this information. Some users also reported that they took longer on images because they had to recall which blocks were white while marking their answers down on the sheet. Based on this, we changed the design of the next study so that the users did not have to write down what they see.

4.3 Formal User Study

For the second user study we used an Epson Moverio BT-100, an Android-based ARD [10]. Unlike Google Glass, the Moverio presents content for both eyes, i.e., it is has a binocular display. However, we found the alignment to be tedious in a binocular setting; therefore, we used the Moverio only



(a) Front view (b) Side view

Figure 7: Chin rest setup for the formal user study.

as a monocular headset, but allow the users to select either the left- or the right-eye display.

Similar to Google Glass, the Epson Moverio does not have image stabilization or head tracking capabilities to account for head jitter. To compensate for this, we provided a chin rest to the participants (see Figure 7). The chin rest stand also situated the ARD on a small platform for added support.

One visual share was displayed on the Epson Moverio and the other was displayed on a monitor. To enable image alignment, participants could move and scale the visual share displayed on the monitor. An Xbox controller was used for this purpose. The participants reported their observations (what they were asked to report varied based on the task) and the investigator noted them. At the end of the study, participants filled out a questionnaire. The questionnaire aimed to capture their confidence levels as the study progressed, as well as participant demographics.

4.3.1 Training

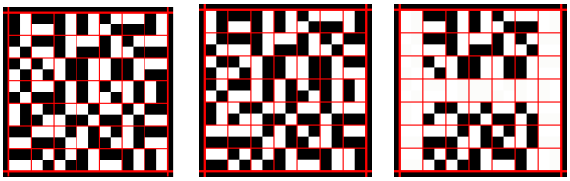
As in the pilot user study, the participants underwent a short training. Participants were given examples of the images they would be viewing and descriptions of the tasks. Through a presentation and practice session on the setup, participants were familiarized with the tasks. The decoded plaintext consisted of individual letters and numbers that resemble a 14-segment LED display font, as in Figure 8. After the training, participants took a 1-2 minute break. Each task was described again as participants started the tasks.

Figure 8: 14-segment LED font used for displaying letters and numbers in the study described in Section 4.3

4.3.2 Task Descriptions

There were three tasks in the study. Each task involved a set of ten image pairs. In each of the tasks, participants aligned the image pairs (visual shares) and identified the decoded plaintext character. They also reported other characteristics of the decoded plaintext for tasks 2 and 3. The participants were timed for each image pair presented; this included the time taken to navigate to the next image pair, identify the characteristics of the decoded plaintext requested in each task and report their observations. A participant had to align only the first image pair of each task; the rest of the image pairs would retain that alignment, unless the user shifted her position.

Task 1: In task 1, users simply identified the decoded characters. Each participant was given ten image pairs that encoded random characters—five numbers and five letters. Upon overlaying the two visual shares on the Epson's display and the monitor, the letter or number was revealed, as in Figure 9.



(a) First share (b) Second share (c) Decoded plaintext

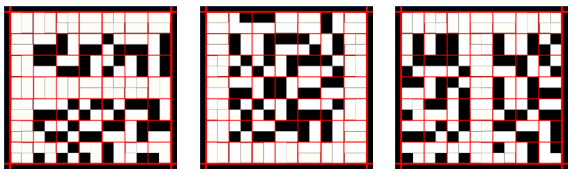
Figure 9: Image shares for letter 'H' (Section 4.3)

Task 2: In the second task, we first reiterated to the participant what constitutes a *legal* decoded plaintext image, namely one in which the blocks that form a character have all regions white (illuminated) and all other blocks have exactly two black and two white regions. All participants were given ten image pairs, where the decoded plaintext of each pair was a letter. For each image pair, we asked participants to identify the letter in the decoded plaintext, identify whether the plaintext is legal, and if not, to count the number of blocks that made it illegal—i.e., blocks that are part of the letter but not wholly white, or blocks not part of the letter that are not half-black and half-white. Here we call such blocks *nonconforming*. All decoded plaintexts in task 2 had between one and six nonconforming blocks, though we did not inform the participants that all decoded plaintexts in the task were illegal.

All participants were given the same image pairs in this task, and the plaintexts decoded from these image pairs were chosen to include "confusing" letters. For example, nonconforming blocks at particular locations can cause confusion especially between 'F' & 'E', 'O' & 'U', and 'T' & 'I'. Six of the ten image pairs provided to participants were intentionally chosen to yield one of these letters, and nonconforming blocks were positioned in their decoded plaintexts so as to maximize confusion. Examples of decoded plaintexts for these three letter pairs are shown in Figure 10.

Task 3: In task 3, each decoded plaintext had a "content" and "non-content" part, as suggested by Naor and Pinkas [25] to increase the likelihood of detecting adversarial manipulation of an image share (see Section 2.1). The content part carries the message, in this task a number. The non-content part contains no meaningful information and is made up of blocks of all the same "color", i.e., blocks that all have exactly two black and white regions (i.e., "black") or blocks that all have four white regions (i.e., "white"). For the purposes of our study, in a *legal* decoded plaintext, each block of the non-content part has exactly two black and two white regions. Thus, an illegal plaintext is simply one containing white blocks in its non-content part. These are the *nonconforming* blocks as shown in Figure 11.

The content part of a plaintext could be either on the left or the right of the plaintext image. There was a shift from

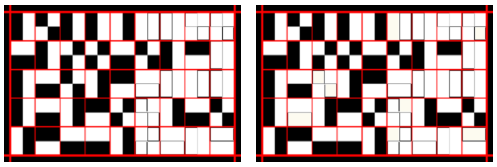


(a) 'F' with four nonconforming blocks in bottom row, or 'E' with six nonconforming blocks in bottom row

(b) 'O' with three nonconforming blocks in top row, or 'U' with three nonconforming blocks in top row

(c) 'T' with five nonconforming blocks in bottom row, or 'I' with six nonconforming blocks in bottom row

Figure 10: Decoded plaintexts for three image pairs with nonconforming blocks to maximize confusion between (a) 'F' & 'E'; (b) 'O' & 'U'; or (c) 'T' & 'I'; used in task 2 of study described in Section 4.3



(a) Legal decoded plaintext

(b) Illegal decoded plaintext

Figure 11: Decoded plaintexts obtained upon overlaying visual shares in task 3.The content part is on the right side and the non-content is on the left. (a) shows a *legal* image with non-content region containing only blocks with two black and two white regions; (b) shows an *illegal* image with malformed white-blocks in the non-content region.

right to left once during the task. The participant was told that a switch would be signaled by an all-white non-content region (a few nonconforming blocks notwithstanding); i.e., this was a signal that the next decoded plaintext would have its non-content region on the other side from its present location. If the non-content region was all-black (again, aside from a few nonconforming blocks), then the non-content region would be on the same side in the next decoded plaintext. The participants were asked to report for each decoded plaintext whether the content was on the left or the right, report the alphanumeric character they decoded, and report if the non-content part was legal and, if not, the number of nonconforming blocks it contained.

In doing so, the task evaluated users' ability to detect nonconforming blocks in a non-content area. The number of nonconforming blocks in this task ranged from 0 to 10. We also observed the minimum number of nonconforming blocks needed in the non-content part of a decoded plaintext in order for users to be able to discern the plaintext as illegal.

4.4 Mechanical Turk Study

An adversary in possession of an image share might attempt to modify blocks in the image share to change the meaning of the decoded plaintext. For example, by flipping the black and white regions in the image-share block, the adversary changes whether the corresponding block in the decoded plaintext is black or white. However, to predict

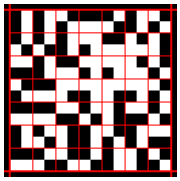


Figure 12: Image share with nine malformed block shares (Section 4.4)

the resulting block color in the decoded plaintext, he must know the color of that block in the original decoded plaintext, i.e., prior to flipping the black and white regions. If he does not know the color of the original decoded plaintext block, he can nevertheless increase the chances that it is turned to white by modifying the block in his image share to have three (or four) white regions, instead of only two. In prior implementations of VC, detecting such malformed blocks may go undetected in an individual share because of the small sizes of the blocks. However, in our implementation, detecting these malformed blocks in an image share (i.e., blocks that do not have exactly two black regions and two white regions) is feasible and can be an an important step to detecting an adversary's manipulation of that share.

We conducted an online user study on Amazon Mechanical Turk to evaluate users' ability to do so. Each participant was given the same 20 image shares in a random order and was asked to identify whether each contains any malformed blocks and, if so, how many. Note that each image was an image share, and so there was no other meaningful information in the image share. An example is shown in Figure 12. This image share has nine malformed blocks that are all next to each other in a group. The malformed blocks could also be positioned randomly in the image share.

One of the goals of the study was to observe the differences in identifying randomly positioned malformed blocks and grouped malformed blocks. Hence, the set of images given to the users was a mix of both: there were eight image shares with randomly positioned malformed blocks, eight image shares with grouped malformed blocks, one image share with one malformed block, and three image shares with no malformed blocks. We also explored the least number of malformed blocks (both grouped and random) that enabled participants to reliably discern that an image share is illegal. The presence of grouped malformed blocks is a characteristic of the attacks described by Chen et al. [6].

At the end of the study the participants filled out a questionnaire to capture demographics and their confidence as the study progressed.

5. RESULTS

This section presents the results and analysis from the formal user study (Section 5.1) and the Mechanical Turk user study (Section 5.2).

5.1 Formal User Study Results

5.1.1 Timing data

Figure 13 shows the average time for decoding plaintext taken by the users in each of the tasks. Of particular interest is the first task since it reflects the amount of time taken to identify the letters or numbers. For tasks 2 and 3, the

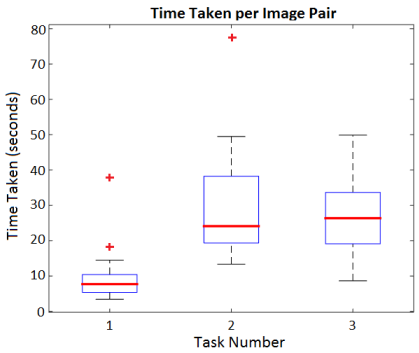


Figure 13: Boxplot showing the average time for plaintext decoding (excluding the first) per participant, in each of the three tasks as described in Section 4.3

Relation between	Correlation coefficient	p-value
Tasks 1 and 2	0.7285	5.0144×10^{-06}
Tasks 2 and 3	0.6679	5.50604×10^{-05}
Tasks 1 and 3	0.3867	0.0348

Table 2: Correlation between task times based on Pearson correlation in user study described in Section 4.3

reported times include the time taken to count the number of nonconforming blocks. The actual time to identify the characters and simply the presence of nonconforming blocks in the decoded plaintext is lower. Figure 13 does not take into consideration the time taken for the first image in each task. For the first image in each task, users spent a considerable amount of time initially aligning the image shares, yielding an outlier that dramatically skews the averages per task. The time taken to initially align the image shares ranged from 18.49 to 313.32 seconds in the first task; 11.94 to 303.16 seconds in the second task; and 27.4 to 280.79 seconds in the third task.

There is a correlation between the time taken in the different tasks on a per-user basis. Table 2 shows the Pearson correlation coefficients and corresponding p-values for each of the timing relations. Participants who took longer on either task 1 or 2 were likely to spend more time in task 2 or 3. A stronger correlation was observed between timing in consecutive tasks, i.e., between tasks 1 and 2 and between tasks 2 and 3. We also found a correlation between time taken in task 3 and image clarity (see Section 5.3.1).

5.1.2 Error Rates

In each task the participants reported the characters they observed. In task 1, 26 out of 30 participants identified all images correctly, and the remaining 4 participants identified 9 out of 10 images correctly. In task 2, 28 participants identified the letters in all the images correctly, one participant identified 9 images correctly, and one participant identified 8 images correctly. In task 3, 27 participants identified the number in the images correctly, two participants identified 8 images correctly, and one participant identified 9 images correctly. This indicates that users were able to clearly identify the letter or number being conveyed. The participants who made mistakes identifying characters in task 2 had also made errors in task 1. Two out of the three participants who made errors in identifying the numbers in task 3 also decoded the plaintexts incorrectly in tasks 1 and 2.

In tasks 2 and 3, we also evaluated users' ability to recognize *illegal* decoded plaintexts using the definitions of illegal given in Section 4.3.2. In these tasks, the participants were asked to identify whether the decoded plaintexts were illegal and, if so, report the number of observed nonconforming blocks.

Nonconforming blocks represent an attacker's active modifications observable in the decoded plaintext (see Section 2.1). In task 2, all decoded plaintexts were illegal, and in task 3, there were seven illegal plaintexts out of ten. Figure 14 shows the percentage of participant responses identifying a decoded plaintext as illegal as a function of the number of nonconforming blocks present in the plaintext. For both tasks 2 and 3, as the number of nonconforming blocks increases, more participants indicated that a decoded plaintext was illegal. Participants were able to discern that an image was illegal even when only one nonconforming block was present in the plaintext. In task 3, 97% of user responses correctly identified the *legal* plaintexts.

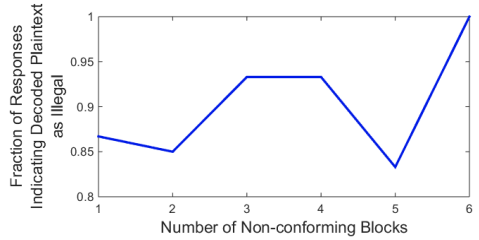
In task 2, participants were given (among others) images with the letters 'F', 'O' and 'T'. As discussed in Section 4.3, we suspected that the presence of nonconforming blocks could easily cause these letters to be confused with other letters. All participants identified these letters or plausible alternatives: specifically, participants identified each 'O' as either 'O' or 'U', each 'T' as 'T', and each 'F' as 'F', 'E', or 'P'. However, in reporting our results, we considered a participant's response as correct only if she also reported the presence of nonconforming blocks (i.e., that the plaintext was illegal).

As seen in Figure 14a, when there were five nonconforming blocks in task 2, there was a decrease in the number of responses that correctly identified a plaintext as illegal. In task 2, the only letter with five nonconforming blocks was the letter 'K'. The nonconforming blocks did not obscure the letter, nor did they cause the 'K' to closely resemble another character. This may have led participants to conclude that the image was legal.

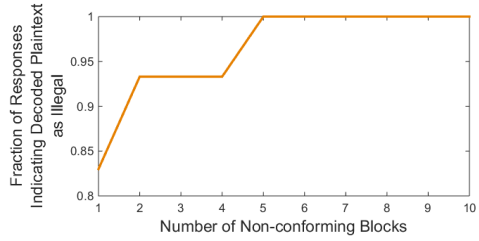
In task 3, participants were also asked to report whether the content was on the left or the right. There was a shift from right to left once during the task, which all 30 users were able to correctly ascertain.

Figure 15 presents the number of nonconforming blocks reported versus the actual number of nonconforming blocks in the decoded plaintext. In Figures 15a–15b, the thicker red line indicates the median value. The bottom and top bars represent the 1st and the 3rd quartiles, respectively. A single line indicates that these quartiles are the same. A median value that is the same as the 1st quartile or the 3rd quartile indicates a skew in the responses.

The plots indicate that the participants erred toward reporting fewer nonconforming blocks than actually were in the plaintexts, and as the number of nonconforming blocks increased, the reported nonconforming blocks also increased. It is important to note that despite not being able to identify all the nonconforming blocks, 89.33% of decoded plaintexts in task 2 and 96% in task 3 were correctly identified as illegal by the participants. In a real-world scenario, this ability to distinguish between illegal and legal plaintexts can be considered more important than identifying the exact number of nonconforming blocks.

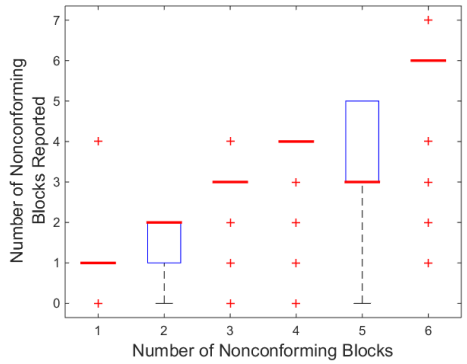


(a) Task 2

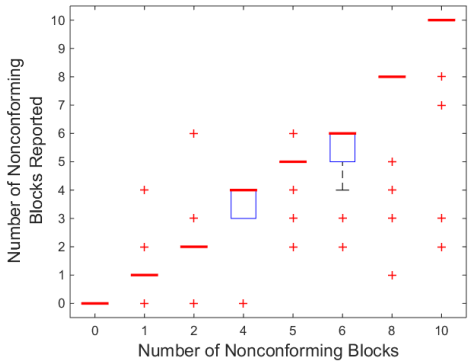


(b) Task 3

Figure 14: Fraction of responses indicating plaintexts as legal or illegal based on the number of nonconforming blocks present in plaintext (Section 4.3)



(a) Task 2



(b) Task 3

Figure 15: Number of nonconforming blocks reported by participants versus actual number of nonconforming blocks present in plaintext (Section 4.3)

The number of participants who were able to correctly identify nonconforming blocks in task 3 is more than that in task 2 as indicated by Figure 16. This is not surprising: In task 3, participants were asked to look for white blocks in a region expected to have all black blocks (or vice versa), while in task 2, the nonconforming blocks could potentially be part of the character. This highlights the importance of using a non-content region as an aid for detecting potential modifications made by an adversary.

Based on the observations from our user study, use of visual cryptography with the aid of augmented reality displays seems promising, despite the challenges that need to be overcome. Our participants were able to discern the characters being conveyed to them, and they were also able to recognize active modification of the decoded plaintexts. It appears the assumptions of the visual cryptographic literature about the capabilities of the human visual system hold true in an augmented reality setting.

5.2 Mechanical Turk Results

In this study, participants were asked to detect malformed blocks in an image share, i.e., blocks that do not have exactly two black regions and two white regions. All 50 participants were able to correctly identify legal image shares, i.e., with no malformed blocks. However, none of the participants were able to detect illegal image shares containing only one malformed block, except for one user who reported that there were four malformed blocks in that image share. All participants correctly identified the remaining 25 illegal image shares as illegal. There was one participant who reported a large number of malformed blocks for most image shares; perhaps the participant did not clearly understand the instructions and was counting the number of regions within malformed blocks, as opposed to the number of malformed blocks.

All the participants correctly identified all image shares with grouped malformed blocks as illegal. However, not all participants detected that an image share was illegal for the case of randomly positioned malformed blocks, as indicated by Figure 17. In this case, though, as the number of malformed blocks increased, the number of participants who indicated that the image shares were illegal also increased. Mechanical Turk users typically spent less than 45 seconds per image.

The number of randomly positioned malformed blocks in an image share was 3, 4, 5, 6, 7, 9, 10, or 12. The number of grouped malformed blocks in an image share was 4, 5, 6, 7, 9, 10, 12, or 13. Figure 18 shows the number of participants who reported all the malformed blocks in an image share, for a given number of malformed blocks. It shows that participants were better at detecting malformed blocks that were grouped together in the image share, as opposed to randomly positioned malformed blocks. However, there is a slight decrease in the number of participants who were able to detect all the randomly positioned malformed blocks as the number of malformed blocks increased. Figure 18 reports the number of participants who reported the exact number of malformed blocks. The number of participants who reported 12 ± 2 malformed blocks for the image share with 12 malformed blocks is 47. So even though the participants reported the incorrect number of malformed blocks, they were within a close range.

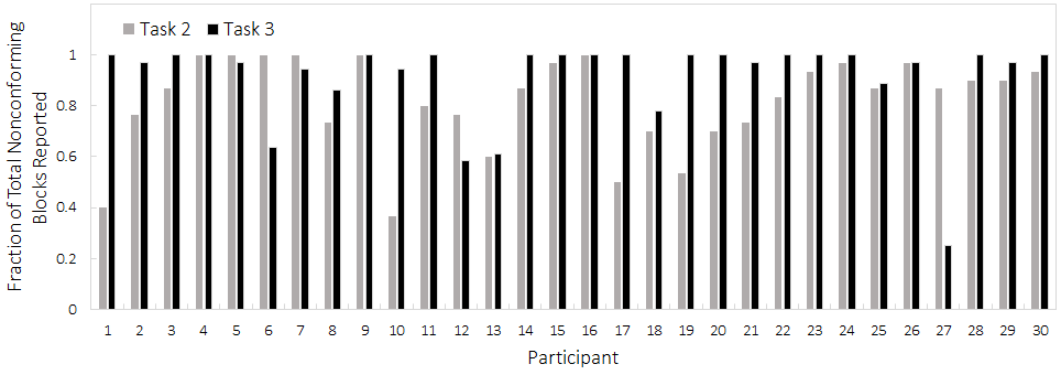


Figure 16: Fraction of nonconforming blocks reported by each participant in task 2 and 3, described in Section 4.3

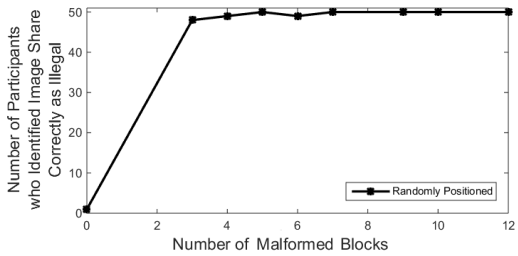


Figure 17: Number of participants who indicated that image shares with randomly positioned malformed blocks were illegal (Section 4.4)

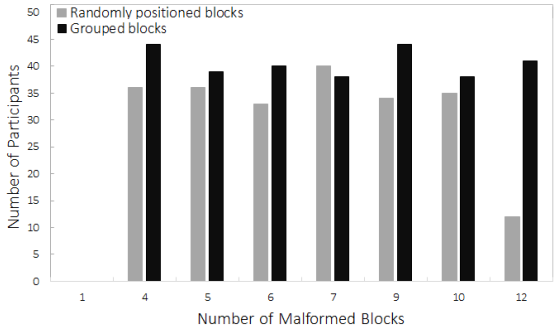
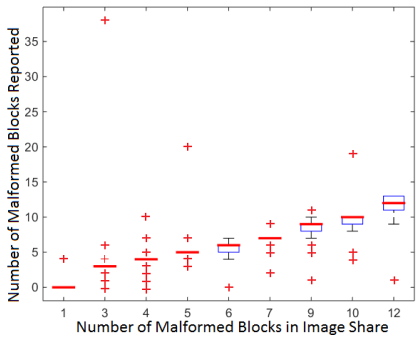
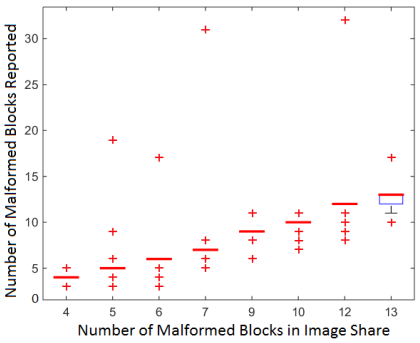


Figure 18: Number of participants who correctly identified all the malformed blocks in an image, whether the malformed blocks were randomly positioned or grouped as described in Section 4.4



(a) Randomly positioned malformed blocks



(b) Grouped malformed blocks

Figure 19: Number of malformed blocks reported versus number of malformed blocks present in the image share (Section 4.4)

Figure 19 shows the number of malformed blocks reported by participants versus the actual number of malformed blocks present in an image share, for both randomly positioned and grouped malformed blocks. For both cases, there was an increase in reported blocks as the actual number of malformed blocks increased. Figure 19 also shows that the participants erred on the side of reporting fewer malformed blocks than were actually present in the image. Again, participants could recognize illegal image shares even if they could not recognize all the malformed blocks. In order to detect an attacker's modification of one share, it would suffice for a user to recognize the image share as illegal, rather than count the exact number of malformed blocks. However, we explored the ability of participants to detect and count different number of malformed blocks, to determine the minimum number of malformed blocks needed to identify an image share that has been modified.

5.3 Questionnaire Responses

The formal user study and the Mechanical Turk survey posed questions to participants about demographics and their perceptions about their own performance during the tasks (see Appendix A for the questionnaires).

5.3.1 Formal User Study

We measured reported confidence, character identification in decoded plaintexts, nonconforming block recognition in decoded plaintexts, and image clarity with regard to the decision process. Participants also indicated if they used prescription glasses during the study and, if so, how glasses impacted the ease-of-use of the ARD and image alignment.

Half of our participant population wore prescription glasses during the study. During the pilot study with Google Glass, prescription glasses had proved to be a hindrance for image clarity, alignment, and ease of use. In the formal user study that was not the case. We attribute this improvement to the Epson's form factor, which makes the ARD easier to use, as compared to Google Glass, while wearing glasses. In the formal user study we found no correlation between reported image clarity and use of prescription glasses. Participants wearing prescription glasses performed neither better nor worse in terms of accuracy or timing, compared to other participants.

However, a majority of the users indicated that clarity of the images was a challenge. In the study using Epson, 11 participants indicated that the images were somewhat clear, 2 indicated that they were neither clear nor blurred, 13 indicated that the images were somewhat blurred, 2 participants indicated that the images were blurred, and only 2 participants indicated that the images were clear. These can be explained by the challenges faced by ARDs, some of which are described in Section 2.2.

Participant perception of image clarity did play a role in timing in task 3: We found a correlation between reported image clarity and the total time taken by participants in task 3 (Pearson correlation coefficient of 0.4050, p=0.026). Larger time values corresponded to lower image clarity. However, we found no such significant relationship with the times taken in tasks 1 or 2, leading us to suspect that the participants' responses to the final questionnaire may have been influenced primarily by their performance during the last task. We found no other significant relationships between timing results, accuracy, and other information gathered in the questionnaires.

5.3.2 Mechanical Turk User Study

Perceived confidence levels and improvement in detection of illegal images were captured by the questionnaire at the end of the user study. We found no relationships between accuracy, the information gathered in the questionnaires, and demographics.

6. LIMITATIONS

Our participant pool was composed primarily of university students and younger population groups. This could have introduced a bias in our results—our participants were the type that understand basic concepts of augmented reality and secret messaging. Our results may not generalize to the entire population. Future work should examine whether our findings would persist for wider population groups.

We conducted the experiments in a dimly lit room, which is the ideal lighting for perceiving bright white light in the ARD as occluding the background. Strongly lit environments could possibly interfere with the user's recognition of the decoded plaintext. As the augmented reality community overcomes this challenge, lighting should no longer be a limiting factor for the use of VC with ARDs.

Due to hardware limitations, we were unable to do image stabilization for reducing head jitter, and so the formal user study was conducted using a chin rest in order to minimize head jitter. Though it does not completely eliminate jitter, the chin rest seemed to substantially improve the ease of aligning visual shares. As such, we believe that head jitter is a challenge that can be overcome with augmented reality displays (ARDs) that include head tracking and image stabilization [27].

The amount of information conveyed per image is limited. Our efforts to increase the block density of the images were constrained by the small field of view and limited resolution of both of the devices. With a 9° increase in field of view, from Google Glass to the Epson Moverio, we were able to increase the block density and convey individual letters and numbers. As ARDs improve their fields of view, we expect users to be able to decode multiple characters simultaneously, which would improve the bandwidth at which multi-character messages could be decoded. That participants invested a median of roughly 8 seconds to decode and recognize a plaintext character in our formal user study suggests that such advances (and user training) will be necessary for messaging via our approach to become practical for any but the most sensitive messages.

7. CONCLUSION AND FUTURE WORK

Using one-time pads for highly sensitive communications (e.g., in intelligence or diplomatic contexts) has a long history. In this paper we sought to modernize this technique by coupling visual cryptography (VC) with augmented reality displays (ARDs). Specifically, we explored how VC and ARDs can be leveraged to enable a user to receive a secret message without revealing that message to her ARD (or any other device acting in isolation). Our evaluation focused on the ability of users to decode VC-encrypted plaintexts using their ARDs. Through an initial pilot study and subsequent formal study, we demonstrated that users were able to effectively leverage VC via ARDs to decode a single character at a time, provided that head jitter can be addressed—which we did using a chin rest in our second study, but which should be addressable using image stabilization—and that the block density was not too high. Future studies involving ARDs with image stabilization would be useful.

We also measured the ability of users to detect active modification of a VC share by an adversary, both in our formal user study (by detecting nonconforming blocks in decoded plaintexts) and in a Mechanical Turk study (by detecting malformed blocks in image shares). Our studies showed that detecting such active attacks is feasible for most users, though not perfectly. To our knowledge, we are the first to verify the capabilities that are assumed of the human visual system by past work on modification detection in VC.

The limited bandwidth of VC using ARDs unfortunately would appear to limit its use to only the most sensitive contexts. However, there are various other VC schemes (e.g., [2, 3, 7, 14, 33]) that might offer different usability characteristics when used with ARDs. Exploring these possibilities, as well as new devices such as HoloLens and augmented-reality contact lenses, appear to be fruitful directions for new research.

Acknowledgements

We are grateful to Prof. Henry Fuchs for useful discussions and for lending us the Epson Moverio device. We express our gratitude to Jim Mahaney for helping build the chin rest for our studies. We are also grateful to Andrew Maimone, True Price, and Kishore Rathinavel for useful discussions and their assistance with documenting the studies.

8. REFERENCES

[1] American Foundation for the Blind. Braille Alphabet and Numbers. http://braillebug.afb.org/braille_print.asp/.

[2] G. Ateniese, C. Blundo, A. De Santis, and D. R. Stinson. Extended capabilities for visual cryptography. *Theoretical Computer Science*, 250(1):143–161, 2001.

[3] B. Borchert. Segment-based visual cryptography. 2007.

[4] L. D. Brown and H. Hua. Magic lenses for augmented virtual environments. *Computer Graphics and Applications, IEEE*, 26(4):64–73, 2006.

[5] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the 25th Hawaii International Conference on*, volume 2, pages 659–669. IEEE, 1992.

[6] Y.-C. Chen, D.-S. Tsai, and G. Horng. A new authentication based cheating prevention scheme in naor–shamir's visual cryptography. *Journal of Visual Communication and Image Representation*, 23(8):1225–1233, 2012.

[7] S. Cimato, A. De Santis, A. L. Ferrara, and B. Masucci. Ideal contrast visual cryptography schemes with reversing. *Information Processing Letters*, 93(4):199–206, 2005.

[8] P. Daponte, L. De Vito, F. Picariello, and M. Riccio. State of the art and future developments of the augmented reality for measurement applications. *Measurement*, 57:53–70, 2014.

[9] R. De Prisco and A. De Santis. Cheating immune threshold visual secret sharing. *The Computer Journal*, 53(9):1485–1496, 2010.

[10] Epson. Epson Moverio BT 100. http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H423020.

[11] B. Furht. *Handbook of augmented reality*, volume 71. Springer, 2011.

[12] S. Gibbs. Google Glass review: useful - but overpriced and socially awkward. http://www.theguardian.com/technology/2014/dec/03/google-glass-review-curiously-useful-overpriced-socially-awkward, Dec. 2014.

[13] Google. Google Glass. http://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354.

[14] Y.-C. Hou. Visual cryptography for color images. *Pattern Recognition*, 36(7):1619–1629, 2003.

[15] C.-M. Hu and W.-G. Tzeng. Cheating prevention in visual cryptography. *Image Processing, IEEE Transactions on*, 16(1):36–45, 2007.

[16] D. Kahn. *The Codebreakers: The Story of Secret Writing*. The New American Library, Inc., New York, NY, 1973.

[17] Kaspersky Labs' Global Research & Analysis Team. Equation: The death star of the malware galaxy. http://securelist.com/blog/research/68750/equation-the-death-star-of-malware-galaxy/, Feb. 2015.

[18] B. Laxton, K. Wang, and S. Savage. Reconsidering physical key secrecy: Teleduplication via optical decoding. In *15th ACM Conference on Computer and Communications Security*, Oct. 2008.

[19] M. Liedtke. Review: 1st peek through Google Glass impresses. http://www.scmp.com/lifestyle/technology/article/1295459/review-1st-peek-through-google-glass-impresses, Aug. 2013.

[20] M. A. Livingston, J. L. Gabbard, J. E. Swan II, C. M. Sibley, and J. H. Barrow. Basic perception in head-worn augmented reality displays. In *Human Factors in Augmented Reality Environments*, pages 35–65. Springer, 2013.

[21] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, page 20. ACM, 2014.

[22] Mandiant. APT1: Exposing one of China's cyber espionage units. http://intelreport.mandiant.com/, 2013.

[23] Microsoft. Microsoft HoloLens. http://www.microsoft.com/microsoft-hololens/en-us.

[24] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications*, pages 282–292. International Society for Optics and Photonics, 1995.

[25] M. Naor and B. Pinkas. Visual authentication and identification. In *Advances in Cryptology–CRYPTO'97*, pages 322–336. Springer, 1997.

[26] M. Naor and A. Shamir. Visual cryptography. In *Advances in Cryptology–EUROCRYPT'94*, pages 1–12. Springer, 1995.

[27] Nels Dzyre. Ten Forthcoming Augmented Reality and Smart Glasses You Can Buy. http://www.hongkiat.com/blog/augmented-reality-smart-glasses/.

[28] Optivent. Augmented Reality HMD Benchmarks. http://optinvent.com/HUD-HMD-benchmark#benchmarkTable.

[29] M. J. Ranum. One-time-pad (Vernam's cipher) frequently asked questions. http://www.ranum.com/security/computer_security/papers/otp-faq/, 1995.

[30] M. Schwartz. Who can control N.S.A. surveillance? *The New Yorker*, Jan. 2015.

[31] D.-S. Tsai, T.-H. Chen, and G. Horng. A cheating prevention scheme for binary visual cryptography with homogeneous secret images. *Pattern Recognition*, 40(8):2356–2366, 2007.

[32] P. Tuyls, T. Kevenaar, G.-J. Schrijen, T. Staring, and M. van Dijk. Visual crypto displays enabling secure communications. In *Security in Pervasive Computing*, pages 271–284. Springer, 2004.

[33] Z. Zhou, G. R. Arce, and G. Di Crescenzo. Halftone visual cryptography. *Image Processing, IEEE Transactions on*, 15(8):2441–2453, 2006.

APPENDIX

A. USER STUDY QUESTIONNAIRES

The questionnaires presented to the participants for the formal user study using Epson Moverio and the online Mechanical Turk study are presented in this section. In the paper we referred to modifications of blocks as *'nonconforming'* in the formal user study and as *'malformed'* in the online user study. However, in the questionnaires and while explaining the tasks to the participants in the study, we used the term *'illegal'* to refer to the same. The terms *nonconforming* and *malformed* are used in the text to distinguish between the blocks as seen in the decoded plaintext and blocks as seen in one image share.

A.1 Formal User Study Questionnaire

Please answer the following questions

1. My confidence in my responses as I progressed through the tasks became:
   Higher
   Somewhat Higher
   Neither Higher nor Lower
   Somewhat Lower
   Lower
2. As I progressed through the study, my ability to identify characters/numbers became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
3. As I progressed through the study, my ability to detect illegal images became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
4. The clarity of the overlaid images was:
   Clear
   Somewhat Clear
   Neither Clear nor Blurred
   Somewhat Blurred
   Blurred
5. Had you used any AR technology before taking part in this study?
   Yes No
6. How would you describe your interaction with technology in your day-to-day life?
   Low Moderate High
7. Do you wear prescription glasses?
   Yes No
   (a) If yes, wearing prescription glasses made using the AR glasses:
   Easy
   Somewhat Easy
   Neither Easy nor Difficult
   Somewhat Difficult
   Difficult
   (b) Wearing prescription glasses made aligning the images:
   Easy
   Somewhat Easy
   Neither Easy nor Difficult
   Somewhat Difficult
   Difficult
8. Gender:
   Female Male Other
9. Age Range:
   18-25
   26-35
   36-45
   46-55
   56 and above

A.2 Mechanical Turk Questionnaire

Please answer the following questions:

1. My confidence in my responses as I progressed through the tasks became:
   Higher
   Somewhat Higher
   Neither Higher nor Lower
   Somewhat Lower
   Lower
2. As I progressed through the study, my ability to detect illegal images became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
3. How would you describe your interaction with technology in your day-to-day life?
   Low Moderate High
4. Please select your gender:
   Female Male Other
5. Please select your age range:
   18-25
   26-35
   36-45
   46-55
   56 and above