You Can't Watch This! Privacy-Respectful Photo Browsing on Smartphones

Emanuel von Zezschwitz, Sigrid Ebbinghaus, Heinrich Hussmann, Alexander De Luca
University of Munich (LMU), Munich, Germany
{emanuel.von.zezschwitz, hussmann, alexander.de.luca}@ifi.lmu.de, sigrid.ebbinghaus@gmail.com

ABSTRACT

We present an approach to protect photos on smartphones from unwanted observations by distorting them in a way that makes it hard or impossible to recognize their content for an onlooker who does not know the photographs. On the other hand, due to the chosen way of distortion, the device owners who know the original images have no problems recognizing photos. We report the results of a user study ($n = 18$) that showed very high usability properties for all tested graphical filters (only 11 out of 216 distorted photos were not correctly identified by their owners). At the same time, two of the filters significantly reduced the observability of the image contents.

Author Keywords

Photo Browsing; Smartphones; Privacy; Obfuscation

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION

Smartphones are among the most ubiquitous computing devices of our times with the ability to store and generate a plethora of potentially sensitive data. Photos, created with or stored on these devices, are considered as very sensitive information by device owners [6], and sharing a subset of these photos, like showing a specific photo to a friend, is a common reason for sharing smartphones [11].

Oftentimes, tasks involving photos and photo sharing take place in (semi-)public settings. This means that private details can be revealed to onlookers without the device owner's agreement. For instance, scrolling through a photo gallery to show a specific photo to a friend can reveal other sensitive data to this person. Also, searching photos for sharing them remotely, can endanger the user's privacy when interacting in public (e.g. in a metro). This represents a serious privacy threat for smartphone users.

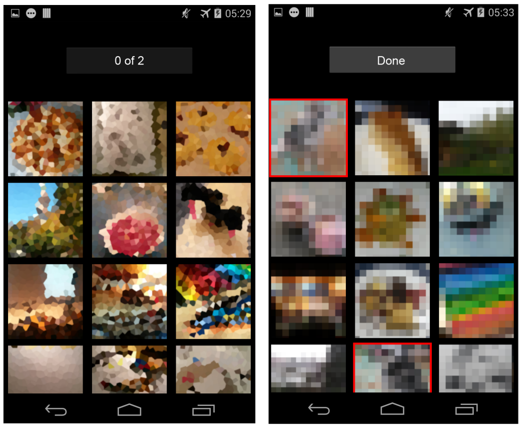

Figure 1. The user study prototype showing two different filters and strengths. Left: Crystallize (high). Right: Pixelate (high). The red borders indicate the photos selected by the study participant.

In this paper, we present an approach that solves this problem. Photos are obfuscated in a way that does not negatively influence the users' ability to correctly identify them. However, the obfuscation makes it hard for an onlooker to make sense of the photos' contents. The main challenge of this approach is to improve the privacy of the user while maintaining high comfort in using the photo browser.

To achieve this, we exploit several known phenomena from memory and visual perception research [2, 4, 7, 12]. In related work, these effects have already been successfully applied to make authentication more secure against observation attacks. Hayashi et al. [10] as well as Harada et al. [8] present image-based authentication systems, in which priming effects and image obfuscation are used to improve the systems' shoulder surfing resistance. Wang et al. [13] showed that repeated exposure of such filtered images enables users to recognize even highly degraded images.

However, these effects have never been tested in connection with privacy-related problems. We report on a user study and present the results which indicate that interaction with an obfuscated photo gallery is still easy and convenient. At the same time, the concept makes it very hard or impossible for onlookers to correctly identify photo content.

THREAT MODEL

Our threat model includes every situation, in which users are interacting with the photo galleries on their smartphones and another (potentially malicious) person is located in the vicinity with the possibility to gaze at the screen. We assume that these will mostly be instances in which a user is voluntarily

CHI'16, May 07-12, 2016, San Jose, CA, USA.
Copyright © 2016 SBN 978-1-4503-3362-7/16/05... $15.00.
DOI: http://dx.doi.org/10.1145/2858036.2858120

allowing the other person to look at the screen. A common use case is showing a specific photo to a friend. In order to do so, the users have to scroll through the photo gallery eventually giving away other photos that they might want to hide.

While malicious shoulder surfing attacks are unlikely [9], we argue that the previously mentioned use case is quite common: showing the screen to another person but still not wanting to share sensitive data. This is backed up by the data from our user study, in which all participants reported this being a common problem when interacting with their devices.

Furthermore, we assume that the attacker has a similar viewing angle as the device owner and the distance is equal or only slightly larger. That is, the view of both persons on the screen is comparable which would give away the content of private and/or sensitive photos to the attacker. In such a scenario, even privacy foils do not provide appropriate protection.

The system proposed in this paper increases the device owner's privacy in such situations by hiding content in a manner that it is still easily understandable by the owner. Our study shows that for attackers without knowledge about the original photographs, the distorted photos do hardly reveal private information.

CONCEPT AND PRE-STUDY

The main idea presented in this paper is to use specific graphical filters to obfuscate photos on a smartphone. Two examples of an obfuscated photo browsing app, as used in the study, can be found in figure 1. Our approach exploits the human ability to recognize known images, objects [4, 8] and faces [2] even when they are distorted, as visual perception is strongly influenced by what we know and what we have seen before [7]. This effect is even stronger if the images are created by the person (e.g. photos made by the device owner) [12].

In order to give away as few information as possible, the idea is to not only obfuscate sensitive photos but the complete photo gallery. Only obfuscating specific photos would already give away potentially sensitive information to an onlooker like "*the user does not want me to see this specific photo so there is something about it*". Even if blurring all images might still communicate mistrust we assume that such problems are minimized and become obsolete when concepts like this become standard (similar to lock screens).

To identify appropriate filters, we firstly performed a theoretical analysis based on related work. The resulting filters (in different strengths) were then evaluated in a pre-study with 24 participants [5]. The task was to find and select privacy sensitive photos with specific content (e.g. nudity or drug use) within a set of twelve images. The photos were provided to the participants two weeks before the actual study took place. Every participants received the same instructions on how to get familiar with the photos and had to confirm the training via questionnaire. To investigate how the filters worked for unknown photos, the actual pre-study comprised both familiar photos and unknown images of the same sensitive content. The pre-study details are described in [5]. The important part, which influenced the main study, is the resulting set of three filters that were identified as suitable candidates for the final prototype and the concept in general. The filters which were chosen based on a literature review and the pre-study results are depicted in figure 2: Pixelate, Crystallize, Oil Paint.

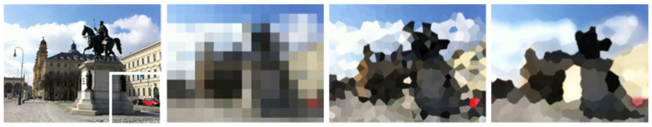

Figure 2. Unfiltered image and the filters used in the final prototype and the user study: Pixelate, Crystallize, Oil Paint (from left to right). The filter strength used in this example was "high". The white borders in the unfiltered photo indicate an example of the used photo snippets.

USER STUDY

The results of the pre-study influenced the main study in the following ways: a) Three appropriate filters at suitable strengths were identified and implemented for the final study. b) The main limitation of the pre-study was the use of photos that we provided to the participants. Thus, we opted for photos created by the participants in the final study.

User Study Design

The study was conducted using a repeated measures factorial design with two independent variables: *Filter Type* (Oil Paint, Crystallize, Pixelate) and *Filter Strength* (none, medium, high). Filter strengths was specified in the pre-study [5]. For instance, the high setting was based on what Hayashi et al. found as the maximum that their users could identify [10]. Each combination with filter strength "none" represented the control condition for the experiment.

A 9 × 9 Latin square design was used to counterbalance the variables and minimize learning effects. Participants were randomly selected and came in teams of two. Both team members acted as "attackers" and "device owners" and the roles were switched after each content assessment task. The term "device owner" will be used for the participant who owns and knows the original photos. The term "sattacker" will be used whenever the original photos are unknown.

The study was conducted in an isolated room at our premises and all participants used the same device. Only the two participants as well as the experimenter were present during the experiment. The study was filmed with a video camera for further analysis.

Procedure

Before the lab study was conducted, participants were randomly recruited in teams of two. We asked each participant to provide 216 photos prior to the study following specific rules. All photos were manually checked by the experimenter and in case one of the following rules was violated, they had to be replaced.

The photos had to be: a) not older than 1 year; b) photographed by the participants; c) taken with their smartphones; d) clearly distinguishable (i.e. no more than one photo of the same subject); e) understandable by strangers;

In addition, the selected photos had to be unknown to the other participant (team partner).

These rules were designed in order to have photo galleries in the study that resembled realistic or "normal" photo galleries as could be found on smartphones nowadays. However, it should not be ignored that rule d) represents a deviation from this assumption as not allowing sequences of very similar images clearly reduces ecological validity. This decision was made to keep the set as diverse as possible. We were afraid that most participants would otherwise simply send a set of the same event (or even of the same object).

At the beginning of the lab study, the procedure was explained in detail to the participants. This also included an explanation of the different roles (attacker and device owner) and how they were alternated during the study. The device owner's task was to identify two specific photos in a subset of 24 photos displayed on two scrollable gallery pages. For each task those 24 photos were randomly selected out of the 216 photos provided by the participant. Overall, all 216 photos were used in the study but not all at the same time.

An important aspect was how to communicate to the participant which photos to search for. There are two problems related to this: a) Providing 216 unique photo descriptions would have added significant workload to the participants without a guarantee that the participants would recognize their descriptions and b) we could not show the actual photos to the participants immediately before searching them as this would have introduced priming effects. Instead of showing the complete photos, we displayed small fractions of the photos for instruction. For this purpose, we automatically cropped the bottom-right corner of the images as indicated in figure 2, left. If the device owner was not 100% certain about which photo the snippet represented, a new snippet was randomly chosen and displayed. During this phase, the attacker had to turn away, which was checked by the experimenter. Then, the attacker was positioned behind the device owner in a location that provided a perfect view on the interaction area. When both team members felt ready, the selection (and observation) task was performed.

To find out whether the shoulder surfing attack was successful, a photo content assessment was performed after each round. For this, the attacker turned around to a laptop in the study room that allowed them to enter a description for the photos selected by the device owners. Participants were asked to describe the observed photos as detailed as possible and to take as much time as required.

After the study, the device owner and a neutral third person were asked to rate the accuracy of the attacker's descriptions (e.g. "*Red thing. Maybe a car.*") using 5-point scales from 1 (not at all the image content) to 5 (perfect description). The exact question they answered was "*How accurately does the following text describe the photo?*". The undistorted photo was displayed next to the attacker's description.

Based on the Latin square design, this was repeated 9 times. As already mentioned, both participants acted as attackers and as device owners and the roles were switched after each content assessment task. Overall, the study took around 60 minutes per team. Each participant received 15 EUR.

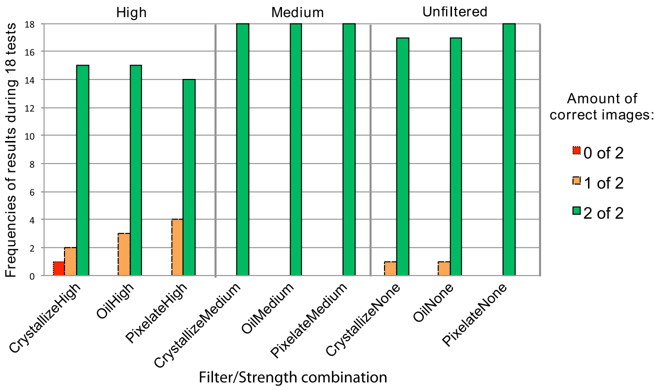

Figure 3. Number of images found and selected by the device owners.

Participants

We recruited 18 participants through mailing lists, social networks and word-of-mouth. None of them participated in the first study. The average age was 25 (20-30; 9 female, 9 male). All of them owned a smartphone for at least one year. All participants reported to frequently use their smartphones for taking and browsing photos. As the study required identifying photos, we ensured optimal conditions for the participants. This included the requirement to wear their glasses during the experiment in case they had any visual impairments.

Results

With 18 participants and all acting both as device owners and attackers, the results are based on $18 \times 9 = 162$ selection tasks and $18 \times 9 = 162$ observation attacks.

Identification Rates

From the device owners' point-of-view, the most important aspect of our approach is to protect their privacy by still offering a convenience level comparable to standard photo browsing apps. In this study, the performance of the device owner's role is defined by the amount of errors made during the selection tasks. As two photos had to be selected for each task, the participant could make zero, one or two mistakes per task.

Figure 3 shows the results for the performance measurement. Overall, error rates were very low. With a medium filter strength, the device owners made no mistakes. Using a high filter strength, 4 images were incorrectly selected with the Pixelate filter, 3 with the Oil Paint filter and 4 with the Crystallize filter (2 in the same task). Please note that two photos were incorrectly selected in the control condition. The video material revealed that this was due to the participants not remembering which photos they should select. The errors in the other conditions could not be drawn back to this reason. Friedman's tests indicated no significant influence of filter strength on the amount of correct image selection (all $p > .05$).

Observation Attacks

To judge the success of the observation attacks, the ratings of the photo content assessment were analyzed. We focus on the device owners' perspectives. However, the ratings of the neutral third person strongly correlated with the owners' ratings ($r = 0.91$).

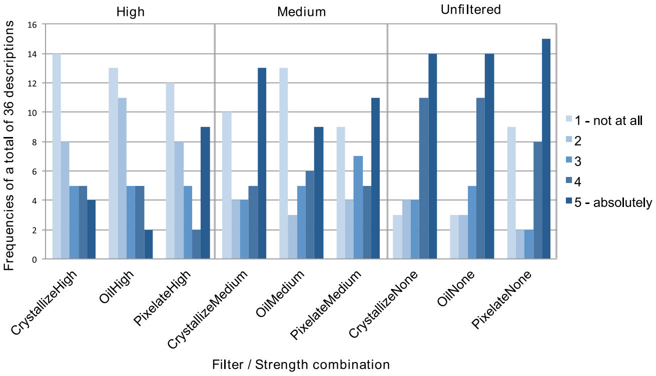

Figure 4. Accuracy of the attackers image descriptions rated by the device owners.

Figure 4 reports the rating frequencies sorted by filters and strengths. The attackers perform very well in the baseline with the majority of descriptions being absolutely accurate or very accurate. With increased filter strength, description accuracy decreases. At medium strength, the amount of bad or completely wrong descriptions is balanced with good descriptions. In the high filter strength the results are not balanced. While only 5 out of 72 photos had a perfect description when using the Oil Paint and Crystallize filters, this was the case for 9 out of 36 photos in the Pixelate filter condition.

Friedman's test for the different filters at different strengths revealed significant differences for the Oil Paint filter ($\chi^2(2) = 15.085; p < .001$) and the Crystallize filter ($\chi^2(2) = 13.244; p < .001$) and none for the Pixelate filter ($p > .05$). Bonferroni corrected Wilcoxon post tests showed significant differences between the baseline and medium and high strength levels for the Oil Paint filter and between none and high conditions of the Crystallize filter (all $p < .025$).

User Opinion

The study video material was used to identify specific trends in the reactions of the participants to the concept. There were two main trends that will be reported in this section:

1) The participants highly underestimated their ability to correctly identify the photos. The first time they saw a distorted gallery with filter strength high, all of them expressed doubts about their ability to solve the tasks. After performing the tasks, they were astonished "*how easy it was*".

2) The feedback with respect to the general concept before, during and after the study was highly positive in all cases. The participants (including the pre-studies) expressed high interest in using a distorted photo browsing gallery and agreed on the addressed observations being a problem. One participant even asked if it was possible to extend the prototype software so that she could use the app on her personal device.

DISCUSSION AND LIMITATIONS

The results of this work indicate that the concept has potential to solve many privacy threats related to photo use on smartphones. The system was able to significantly reduce observability of the personal photographs of the study participants. At the same time, the added privacy should not add additional burden to the users as this would make it likely that they will circumvent the mechanism [1]. Based on our findings, we argue that our approach does not add any noteworthy burden to the users and does therefore fulfill this necessary criterion.

The results furthermore revealed that the selected filters highly influence the system's performance. While the identification rates are stable over all filters, Oil Paint and Crystallize clearly outperformed the Pixelate filter with respect to the achieved degree of privacy. This might be due to properties that distinguish them from Pixelate. These properties seem desirable when developing such a system. The main properties are: a) slight distortion of the photos' color proportions; b) smudging edges within the photos; c) overlapping colors; d) fast removal of small details.

Even though the study showed that the system works well, we argue that the obfuscation should only be activated when needed. There are situations in which it would be desirable to allow the users control over the filters and the filter strengths. For instance, a user might want to share a specific gallery with a friend who is not familiar with the photos and thus needs unfiltered access. This could be done either by explicit activation (e.g. button) or the process could partially be automated taking into account context information or the image's metadata [3].

Finally, there are inherent limitations in our approach which need to be addressed. First, the limited sample size (n=18) does not allow the generalization to a specific population. Secondly, even if the study design was adequate to prove the general feasibility of the concept at this early stage, it excluded important real-life factors. For example, we did not test the scalability of the approach. While the study showed that the concept works well for small galleries of distinguishable images, we cannot make any claims about galleries with several hundred potentially very similar photos.

CONCLUSION AND FUTURE WORK

In this paper, we presented a concept to reduce privacy risks when browsing photos on smartphones based on distortion of the respective photos. Our results show that the selected graphical filters were able to significantly reduce observability of the photo content. At the same time, the error rate of the system stays stable when compared to an unfiltered baseline.

Future work needs to explore the real-life utility of the concept in a longitudinal field study. For this purpose, the concept should be tested with a large set of potentially similar images. Furthermore, the influence of distortion on interaction speed should be investigated as we were interested in performance related to recognition rate and thus told the participants to take their time while searching for the photos. Finally, since image obfuscation should only be activated when needed, feasible interaction concepts need to be found.

In addition, we were only able to test a limited set of filters and strengths. While we did our best to identify the optimal candidates using a thorough literature review and pre-studies, we might have missed a better filter. Especially filters with properties like smudging edges that seemed to be beneficial in our study should be further explored.

REFERENCES

1. Anne Adams and Martina Angela Sasse. 1999. Users Are Not the Enemy. *Commun. ACM* 42, 12 (Dec. 1999), 40–46. DOI: http://dx.doi.org/10.1145/322796.322806

2. A. Mike Burton, Stephen Wilson, Michelle Cowan, and Vicki Bruce. 1999. Face Recognition in Poor-Quality Video: Evidence From Security Surveillance. *Psychological Science* 10, 3 (1999), 243–248. DOI: http://dx.doi.org/10.1111/1467-9280.00144

3. Daniel Buschek, Moritz Bader, Emanuel von Zezschwitz, and Alexander De Luca. 2015. Automatic Privacy Classification of Personal Photos. In *Human-Computer Interaction INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Lecture Notes in Computer Science, Vol. 9297. Springer International Publishing, 428–435. DOI: http://dx.doi.org/10.1007/978-3-319-22668-2_33

4. Tamara Denning, Kevin Bowers, Marten van Dijk, and Ari Juels. 2011. Exploring Implicit Memory for Painless Password Recovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2615–2618. DOI: http://dx.doi.org/10.1145/1978942.1979323

5. Sigrid Andrea Ebbinghaus. 2015. *Privacy-Respectful Photo Browsing for Smartphones Filter Selection and Evaluation*. Technical Report LMU-MI-2015-3. University of Munich (LMU), Department of Computer Science, Media Informatics Group, Munich, Germany. www.medien.ifi.lmu.de/forschung/publikationen/detail?pub=ZezschwitzTR_Privacy

6. Adrienne Porter Felt, Serge Egelman, and David Wagner. 2012. I'Ve Got 99 Problems, but Vibration Ain'T One: A Survey of Smartphone Users' Concerns. In *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM '12)*. ACM, New York, NY, USA, 33–44. DOI: http://dx.doi.org/10.1145/2381934.2381943

7. Richard L. Gregory. 1997. Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 352, 1358 (1997), 1121–1127. DOI: http://dx.doi.org/10.1098/rstb.1997.0095

8. Atsushi Harada, Takeo Isarida, Tadanori Mizuno, and Masakatsu Nishigaki. 2006. A User Authentication System Using Schema of Visual Memory. In *Biologically Inspired Approaches to Advanced Information Technology*, AukeJan Ijspeert, Toshimitsu Masuzawa, and Shinji Kusumoto (Eds.). Lecture Notes in Computer Science, Vol. 3853. Springer Berlin Heidelberg, 338–345. DOI: http://dx.doi.org/10.1007/11613022_28

9. Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, 213–230. https://www.usenix.org/system/files/conference/soups2014/soups14-paper-harbach.pdf

10. Eiji Hayashi, Rachna Dhamija, Nicolas Christin, and Adrian Perrig. 2008. Use Your Illusion: Secure Authentication Usable Anywhere. In *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS '08)*. ACM, New York, NY, USA, 35–45. DOI: http://dx.doi.org/10.1145/1408664.1408670

11. Amy K. Karlson, A.J. Bernheim Brush, and Stuart Schechter. 2009. Can I Borrow Your Phone?: Understanding Concerns when Sharing Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1647–1650. DOI: http://dx.doi.org/10.1145/1518701.1518953

12. Hikari Kinjo and Joan Gay Snodgrass. 2000. Does the generation effect occur for pictures? *The American journal of psychology* 113, 1 (2000), 95. http://www.jstor.org/stable/1423462

13. Zhan Wang, Jiwu Jing, and Liang Li. 2013. Time Evolving Graphical Password for Securing Mobile Devices. In *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security (ASIA CCS '13)*. ACM, New York, NY, USA, 347–352. DOI: http://dx.doi.org/10.1145/2484313.2484358