LMU

Emanuel von Zezschwitz, Alexander De Luca
and Heinrich Hussmann (Editors)

Privacy-Respectful Photo Browsing for Smartphones

Filter Selection and Evaluation

Sigrid Andrea Ebbinghaus

Technical Report
LMU-MI-2015-3, November, 2015
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Emanuel von Zezschwitz, Alexander De Luca and Heinrich Hussmann (Editors)

Privacy-Respectful Photo Browsing for Smartphones Filter Selection and Evaluation

Sigrid Andrea Ebbinghaus

UNIVERSITY OF MUNICH (LMU)
Department of Computer Science
Media Informatics Group
Prof. Dr. Heinrich Hußmann

Technical Report

Privacy-Respectful Photo Browsing for Smartphones: Filter Selection and Evaluation

Sigrid Andrea Ebbinghaus
sigrid.ebbinghaus@gmail.com

1 Introduction

The public use of smartphones is common in urban societies by now. In the subway, at the supermarket checkout or at the restaurant - people use their mobile phones in almost every context or situation, mostly without bothering who is watching them. But when it comes to specific kinds of data and information, they are usually aware that those are private and not meant to be seen by strangers. Although personal photos taken with the smartphone often belong to this category of data, people risk exposing them when using their smartphone galleries in public. For instance, in order to send a specific picture via email, one might have to browse several pages of photographs before finding the correct picture (see figure 1.1, left-hand side). In the process, an observer or attacker is potentially able to view dozens of private photos without the owner noticing it. Even when dealing with friends (see figure 1.1, right-hand side) there is the risk of people accessing parts of the gallery or specific images they were not supposed to see, while the smartphone is passed around or the user searches for pictures to show the others. Thus, the users of all current smartphone photo galleries again and again compromise their privacy and, until now, there is little they can do about it.

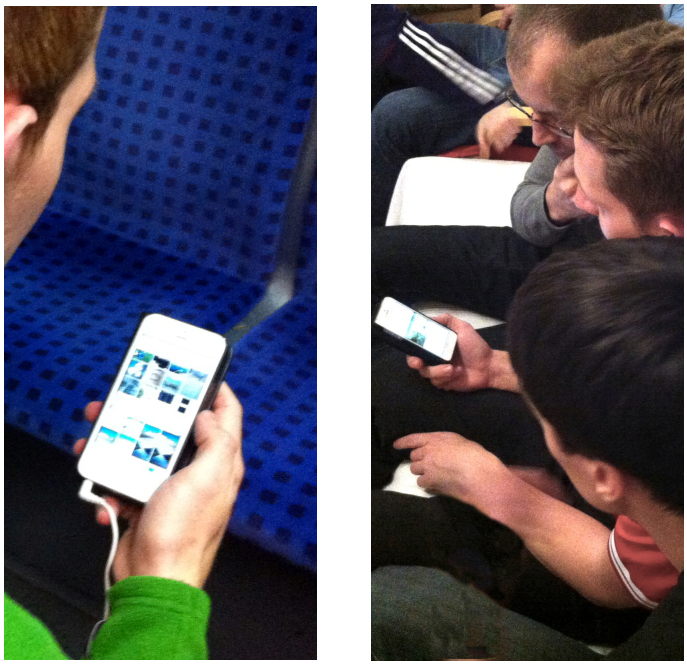

Figure 1.1: Two scenarios in which the privacy of smartphone gallery users can be compromised: browsing through the gallery in the subway and showing photographs to friends on one's device

Considering the described privacy risks involved with mobile photo browsing it becomes evident that smartphone owners have no distinct control over who is able to look at their photos and recognize content on them and who is not. Would it not be pleasant to know, that only the owner of a photo gallery is able to recognize her photos, while strangers cannot understand their content? The human brain and especially our visual perception are strongly influenced by what we know and what we have seen before [16, 6]. Research has shown that images of familiar objects [5, 7] and faces [13, 3] can be recognized due to their familiarity - even when they are distorted or degraded. If a person knows the original, undistorted picture, it is easier for her to make sense of a distorted version and thus recognize it (see figure 1.2). A stranger, on the other hand, who has no knowledge of the original image, might not be able to make sense of the image content anymore, if the distortion or degradation is strong enough.

Figure 1.2: Merely looking at the original photograph (right) once is enough, to help our perception to better recognize and understand the distorted version of it (left)

In other words there are certain ways of altering and distorting photographs that might be able to prevent attackers or unauthorised people from understanding the content of personal photos, while the owner of the pictures is still able to recognize them. This technical report evaluates the feasibility of applying such distortion filters to a smartphone photo gallery in order to protect the users privacy. While several researchers have already implemented similar methods of obfuscating image content, mainly in the area of graphical authentication (e.g. [8, 15, 7]), this technical report focuses on the selection and evaluation of suitable distortion filters.

First, related work is reviewed to derive a set of suitable filters (section 2). Next, we determine sensitive image subjects to be able to test the obfuscation abilities of the filters (section 3). In section 4, we report on a user study which evaluates the performance of the selected filters by implementing them into a prototype gallery consisting of both privacy sensitive and insensitive photographs. Finally, section 5 discusses the results.

2 Filter Selection and Filter Strength

This section discusses the research concerning image filters. At this point, it should be noted that this report does not claim to have found the best suitable graphical filters for the task of obfuscating images from strangers while still being recognizable by the owner. The goal of this research was to find suitable filters, that are feasible for the use case. This does not imply that there are no other filters, which are equally or better qualified for the task. The choice of graphical distortion filters is almost unlimited and there is even the possibility of developing a new filter only for this purpose.

2.1 Filters in Related Work

Although the use case of the following studies is graphical authentication instead of mobile photo browsing, they are all relevant related work as they leverage image filters to protect pictures (pass-images) from Shoulder Surfing and similar attacks.

Harada et al. [7] propose a user authentication system that utilizes unclear pass-images to improve robustness against observing attacks and leakage (leakage means that the user gives away her password or information about it, by telling someone). In an enrollment phase, the legitimate user is shown original images and their degraded counterparts (see figure 2.1) in order to later be able to make sense of the degraded image and recognize it. The authors argue that the attacker, who only sees the degraded image, is not able to make sense of the content and thus is not able to simply observe and steal the unlock pattern. The study results show that the method was able to decrease attacker success rates while not preventing the legitimate user from authenticating.

Figure 2.1: Example for original image and degraded counterpart used by Harada et al. [7]. The user is allowed to see (and learn) both images in order to later be able to make sense of only the degraded image. The attacker on the other hand, without having seen the original version, is not able to do so

Figure 2.2: Example for original images and degraded counterparts used by Denning et al. [5]. The user is primed to the complete images of different objects to improve her ability to recognize the degraded versions of those images during an authentication

Denning et al. [5] present an authentication system based on the user's implicit memory of images and the recognition of their incomplete counterparts (see figure 2.2). Their main goal is to reduce the burden of memorizing complicated or difficult authentication patterns. Hence, they analyze the viability of a priming effect. Although their results indicate that the priming effect for the tested images (line drawings of objects) was not very strong and that the authentication time was rather long, the authors are still confident that implicit memory based authentication is promising due to usability and security improvements compared to other solutions.

Wang et al. [15] improve the strength of a graphical authentication mechanism gradually over time by applying an edge detection filter to the pass-images and reducing the amount of visible edges over time (see figure 2.3). Due to repeated exposure to the filtered images, the legitimate user is still able to recognize them in a highly degraded state while the attacker's ability to understand the images decreases more and more. The study indicates that the authentication method provides protection from various attacks. The authors also make recommendations on the selection of good pass-images. For instance, if all pass-images are similar to each other, it might be easier for an attacker to detect a pattern.

Hayashi et al. [8] examine the application of an oil-painting filter (see figure 2.4) on graphical password images in order to make Shoulder Surfing and social engineering attacks more difficult. They work with personal photographs of the smartphone owner and argue that only she is allowed to learn the combination of original and filtered photo and thus later able to recognize the filtered photo. The results show that participants of their user study performed well at recognizing the distorted images and even had fun with the task. In a subsequent publication, Hayashi et al. [9] tested their authentication method against educated guessing attacks. As the distortion prevents

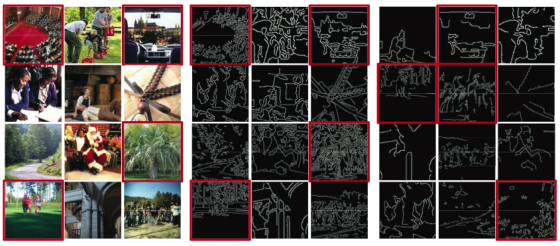

Figure 2.3: Example for original image and two stages of degraded counterparts used by Wang et al. [15]. An edge detection filter is applied to the original image and the amount of displayed edges is reduced over time to increase the strength of the authentication method

someone with no knowledge of the original image from recognizing its content, an attacker with knowledge about the victim is also prevented from guessing the right pass-images. Usually a dear person, a pet or certain things the user is fond of, are likely choices for a pass-image and thus can be guessed easily. But if the attacker does not recognize the dear person, the pet or the object, due to distortion this advantage is lost.

Figure 2.4: Example for original image and filtered counterpart used by Hayashi et al. [8]. A similar Oil Paint filter is used, among others, within the prototypes of this project

2.2 Filter Selection

In order to find suitable filters the described research is now examined in the light of the defined use case. Wang et al. [15] utilize repeated exposure to a filtered image (edge detection filter) and thus allow for degrading an image more and more over time. In the beginning the image is still recognizable, even for a stranger and in the end the image is very obfuscated (thus not recognizable, even if known, without having been exposed to prior filtering stages). Although the user of a photo gallery is also repeatedly exposed to her pictures, the high filter setting could only be used for older, well-known pictures while newly taken photographs would remain unprotected. Harada et al. [7] degrade images in a way that only conserves the meaning of the image to someone who learned to connect the original image with the strongly degraded counterpart. While this might work for a few, distinct images for graphical authentication, it is not feasible for larger collections (galleries) of photographs, especially as all color information are deleted. Finally, Denning et al. [5] work with line drawings of certain objects rather than actual photographs. Those drawings are degraded by taking away lines and thus reducing the contours. However, this is not an actual graphical filter that can be applied to photographs, at least not one-to-one, and hence there is no proof-of-concept for a filter that just works similarly.

Oil Paint Filter

From the already presented research only the oil-painting filter Hayashi et al. used [10, 8] is viable for use in a mobile photo gallery. It distorts photographs by reducing their degree of detail and hence making it difficult for someone who has never seen the original image to recognize the content. All filter effects were applied on the images for the prototype using Adobe Photoshop

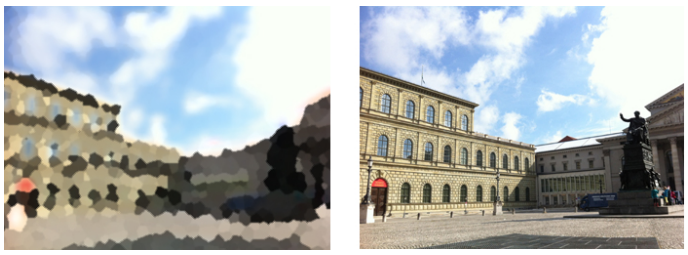

Figure 2.5: Original sample image and its distorted counterpart with an applied Oil Paint filter at strength medium

CS5. The Oil Paint filter was created using the effect Palette Knife [1]. It reduces the image's details to create the look of a canvas painting. It was used with Details set to 1 and Smoothing set to 10. The strength levels high, medium and low were achieved using a stroke of 20, 10 and 5. Figure 2.5 shows an example for the used Oil Paint filter at strength medium.

Crystallize Filter

Apart from research papers presented, Tripp [14] investigated a research question similar to the one in this thesis in 2013, with the aim of improving privacy for mobile photo browsing. Although the results cannot be used here, one of the filters used in his thesis showed promising tendencies for the task at hand. This filter is referred to as stained glass or crystallize or mosaic filter. It will be called Crystallize in this report. It was created in Photoshop using the Crystallize effect [2]. It condenses areas of pixels into a polygon shape of solid color. The strength levels high, medium and low were achieved using a Cellsize of 20, 10 and 5. Figure 2.6 shows an example for the used Crystallize filter at strength medium.

Figure 2.6: Original sample image and its distorted counterpart with an applied Crystallize filter at strength medium

[1] Reference of graphical effects and filters within Adobe Photoshop: Adobe Photoshop. Filter-effects-reference `https://helpx.adobe.com/photoshop/using/filter-effects-reference.html` Accessed on 2014/12/21.

[2] Reference of graphical effects and filters within Adobe Photoshop: Adobe Photoshop. Filter-effects-reference `https://helpx.adobe.com/photoshop/using/filter-effects-reference.html` Accessed on 2015/12/21.

Pixelate Filter

When researching privacy protection and obfuscation techniques, one quickly reaches the topic of privacy filters for hiding faces (e.g. in surveillance footage or in the media). A common and well known technique of distorting faces to hide a person's identity is pixelization [12, 4]. Although this privacy filter is still widely implemented, there is evidence that pixelization does not succeed very well in obfuscating familiar faces [2, 3]. Hence, pixelization (or Pixelate filter as referred to in this report) shows promising tendencies for obfuscating unknown content while not being able to hide familiar content. Although there is no actual proof-of-concept for photographs in general yet, especially with its very low computational demands, Pixelate seems promising and completes the set of three suitable filter types. The Pixelate filter was created in Photoshop using the Mosaic

Figure 2.7: Original sample image and its distorted counterpart with an applied Pixelate filter at strength medium

effect [3]. Despite the name, the Mosaic effect in Photoshop represent a simple pixelization with interpolation of the color of the summarized pixels. The strength levels high, medium and low were achieved using a Cellsize of 20, 10 and 5. Figure 2.7 shows an example for the used Pixelate filter at strength medium.

2.3 Filter Strength

Looking at the way humans understand images [1] it becomes evident that finding the right filter strength for a privacy enhanced photo gallery is a challenge. It is possible to take away a certain amount of information (like contours of forms and objects) before an image is so degraded that not even the owner is able to recognize it anymore. Seeing the described distortion filters as a way of reducing image information, it is crucial to find the right trade-off so that a stranger is not able to recognize image content anymore, but the owner still is. From the three selected filter types, chosen from related research, only the Oil Paint filter - taken from Hayashi et al. [8] - comes with instructions on how they chose a good level of strength. The authors performed a pretest to determine a filter strength that allowed the users to only just be able to recognize the familiar image while attackers do not. We approximated the filter strength "*high*" (pixel radius of 20) to this level of strength as we assumed that within a photo gallery with many distorted pictures, this is the maximum filter strength the user will be able to cope with. From this level, the strength levels were gradually lowered (by half) to a pixel radius of 10 and 5 (medium and low). To test whether these same settings of pixel radius (or cellsize) create comparable distortions for all three filters, a low-fidelity test with several sample images was performed. A high-pass filter was applied to the images and the resulting histograms at the different levels of filter strengths were compared, to see whether the reduction of information from the image is comparable. Figure 2.8 illustrates an example. Although the results varied slightly for different types of images, they were stable enough to accept the three described filter strengths as comparable.

[3] Reference of graphical effects and filters within Adobe Photoshop: Adobe Photoshop. Filter-effects-reference `https://helpx.adobe.com/photoshop/using/filter-effects-reference.html` Accessed on 2015/12/21.

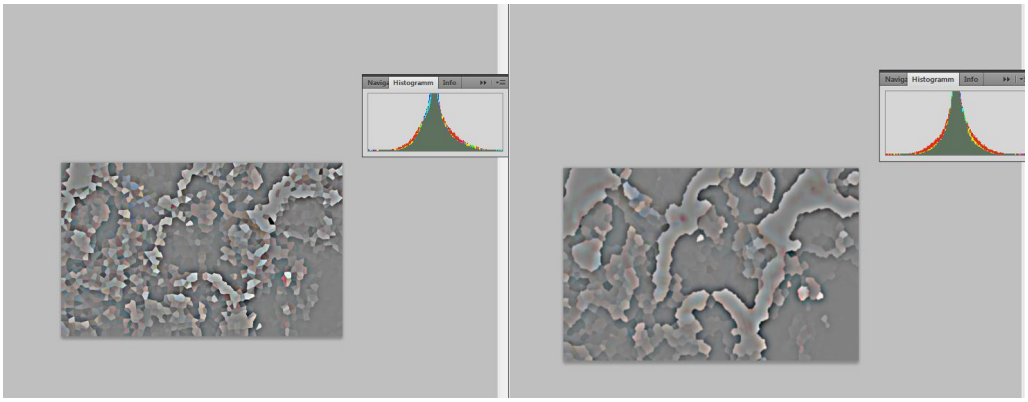

Figure 2.8: Example of high-pass filtering and histogram comparison for a privacy sensitive image, once filtered with Oil Paint and once filtered with Crystallize, both at strength medium

3 Finding Image Content to Obfuscate

Protecting the user's privacy during mobile photo browsing, in this report, describes handing the user as much control as possible over who is allowed to see which of her personal photographs and who is not. As complete control is hardly achievable, it is important to know how severe the intrusion into privacy is, depending on the content of the photograph. That is why an online survey on privacy sensitive image content was conducted among 36 participants to highlight the kinds of images the user definitely needs control over, in order to then test the obfuscation performance of the filters for those pictures in particular.

3.1 Methods

The participants rated the privacy sensitivity of eight different photo subjects on a slider with 20 invisible steps. All topics needed to be rated twice, once, assuming that oneself is displayed and once, assuming that someone else is displayed. In both cases participants were told to assume that the pictures are stored on their personal smartphone. The topics were:

- Person consuming alcohol
- Person consuming drugs
- Person during the exchange of affection
- Person making an unattractive face
- Person smoking a cigarette
- Person being drunk
- Person being naked
- Young child under the age of 6 (subject was only rated for pictures displaying someone else)

Example images can be found in attachment A.

3.2 Results

Of the 36 participants taking part in the online survey, 21 were female and 15 were male. All were between the ages of 20 and 34 and all but two had either matriculation standard or a university degree. There are only two types of image content where almost all participants agreed that those are highly sensitive: nudity and drug use (see figure 3.1). In comparison to other topics (for instance "Person smoking a cigarette", where the ratings were quite diverse) those results show that most users agree that some topics definitely pose a threat to one's privacy. Interestingly, the only topic where the results for assuming the participant herself or someone else is displayed differed noticeably was smoking (see figure 3.2). This indicates that several people do not want everybody to know that they smoke, but do not care about pictures of other people smoking cigarettes.

The results of the online survey show that there are certain kinds of photo subjects that are generally regarded as more sensitive or more critical to the personal privacy, than others. In order to properly evaluate the performance of the selected distortion filters for privacy protection, it is important that the filters are tested on such critical photographs. As a result, a combination of insensitive pictures (displaying random content like food, buildings and people) and sensitive pictures (displaying either nudity or drug use) are utilized in the user study.

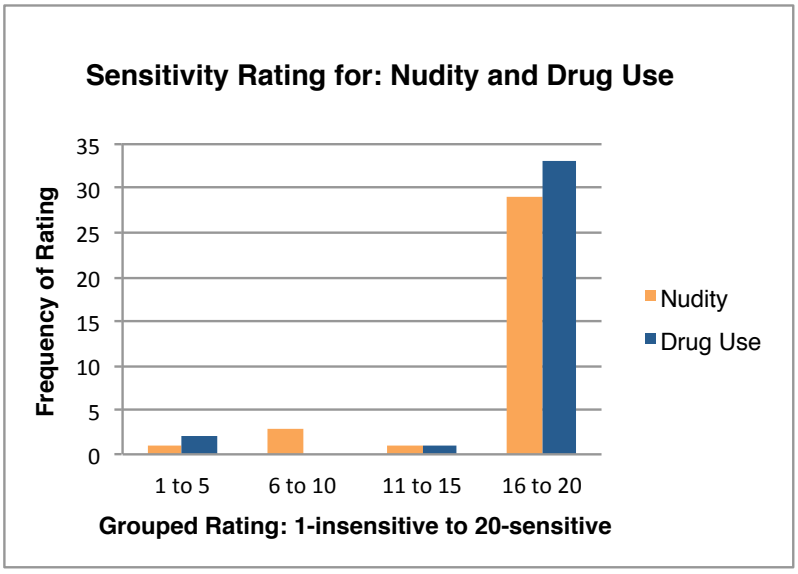

Figure 3.1: Histogram of the consistent ratings of privacy sensitivity by 36 participants for the photo subjects nudity and drug use assuming that the participant is the displayed person. The results for assuming another person is displayed are similar

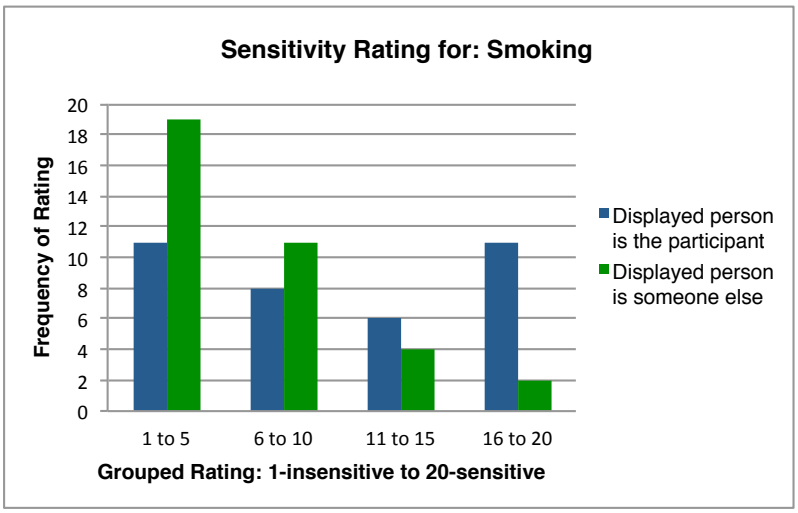

Figure 3.2: Histogram of the diverse ratings of privacy sensitivity by 36 participants for the photo subject: Person smoking a cigarette

4 User Study: Filter Evaluation

The goal of the user study is to evaluate the performance of the three selected filter types at the three levels of strengths, to find out, which of them are suited for implementation in a privacy enhanced photo gallery for smartphones. A quantitative experiment was conducted with 24 participants, who are familiar with browsing photos on a smartphone and were recruited using social media, mailing lists and word-of-mouth advertising.

4.1 Methods

4.1.1 Prototype

For this study a first prototype of a filtered privacy enhanced photo gallery was implemented on an Android device (Samsung Galaxy S3, see figure 4.1, left). The prototype application displays a one-page photo grid of 12 photo thumbnails (standard square crop) per page. As mentioned before, the photos used for the study consist of privacy insensitive and sensitive pictures, the sensitive pictures showing either nudity or drug use. In each run the participant's task is to select all images with privacy sensitive content in the one-page grid ("Mark all pictures showing nudity or drug use") as fast as possible. There are always either four or eight sensitive pictures to select. The participants select the pictures by touch and are also able to correct their selection by deselecting or reselecting images. The prototype indicates selection by lowering the alpha value of the image (thumbnail on black background becomes darker). The right image of figure 4.1 shows an example for one of those gallery grids with a Crystallize filter. During the experiment, the prototype records all inputs from the user with a timestamp and a reference to the image and its attributes (privacy sensitivity, familiar or not, position in grid, etc.). In addition to this activity log, a summary of the most relevant test data for analysis and the collected image data for all photographs in the current grid are both created as csv files within the prototype. To avoid difficulties, the prototype was supplied with the study design and developed in a way that it automatically generates the correct order of conditions for the current user id and randomizes the arrangement of the photographs within the grid.

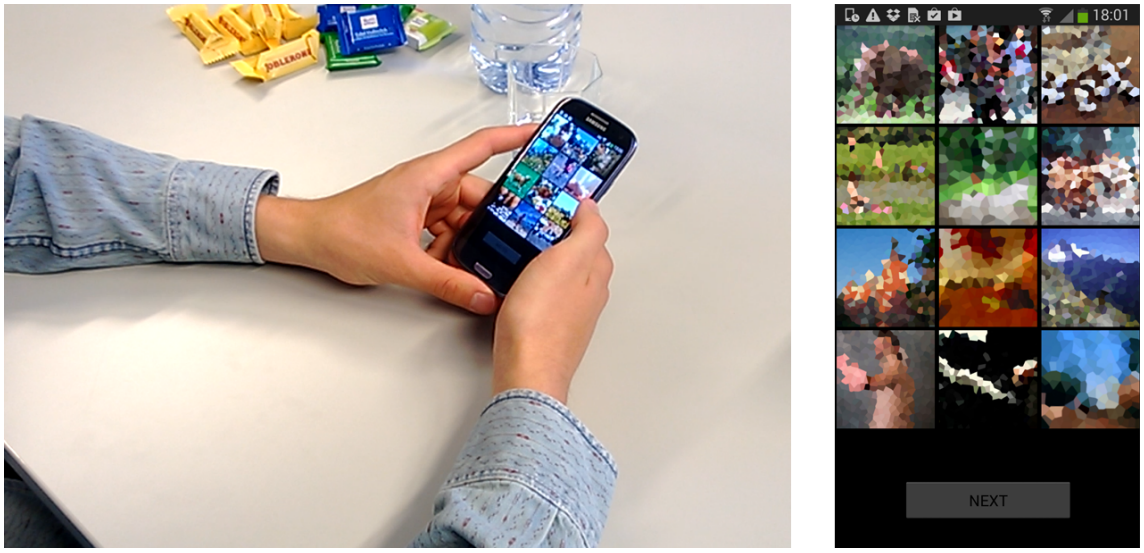

Figure 4.1: Left: User study participant performing a selection task in the prototype gallery. Right: Gallery grid with privacy sensitive and insensitive images and an applied Crystallize filter (high)

4.1.2 Photographs

In order to test all conditions with all participants and simultaneously minimize learning effects, a total of 144 images were used in the experiment. For the insensitive images a broad mixture of random motives like buildings, nature, animals, people, food, etc. were selected with the aim to include both pictures with higher and lower detail. Moreover, some decoy images (like a skin colored dog or a person in a bathing suit) that might look like pictures displaying nudity were included. Concerning the drug use pictures, it was important to select images with an obvious subject, that all participants can easily understand. To avoid misunderstandings, the participants were instructed not to treat alcohol as a drug, but to regard all pictures of cigarette-like objects as well as all kinds of pills as drugs. Choosing the photos displaying nudity, it was important not to include too many bold images, that solely display skin color. Instead, many images with context (for instance a nudist during hiking) were used.

4.1.3 Variables

The i*ndependent variables* in this experiment are:

- Type of image filter (Oil Paint, Crystallize, Pixelate)
- Strength of image filter (none, low, medium, high)
- Role (attacker, user)
- Kind of privacy sensitive content (nudity, drug use)

As *dependent variables* the time the participants needed to complete the task was measured as well as the amount of errors they made. Selecting a picture with no privacy sensitive content is considered a false positive and not selecting a picture, although it has privacy sensitive content (it shows nudity or drug use) is considered a false negative. The independent variable role is evenly distributed among the grid of 12 images, meaning that in each run six images were familiar to the user and six images were unknown or new to her. The kind of privacy sensitive content (nudity or

drug use) is also distributed evenly among the four or eight privacy sensitive pictures per grid. An exemplary assembly of the 12 images within the grid is shown in figure 4.3.

The dependent variable errors in combination with the time the participants need to make a selection gives information on how well the participants were able to recognize the known images as well as how difficult it was for them to recognize the content of the unknown pictures. An ideal filter would produce a high error rate for the previously unknown images (attacker role) and a low error rate for the familiar images (user role) while not causing the participants to need noticeably large amounts of time to complete the selection task.

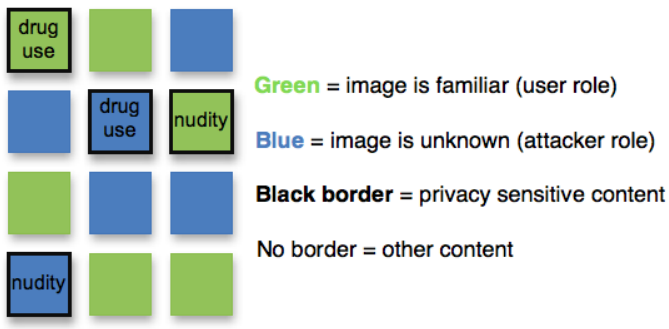

Figure 4.2: Exemplary assembly of 12 images within one gallery grid with the variables role and kind of privacy sensitive content evenly distributed among them

4.1.4 Procedure

In a repeated measures design, each participant was testing all combinations of the independent variables filter and strength leading to a total of 12 runs for each participant. A replicated 12x12 Latin square design was used to counterbalance the variables filter and strength for 24 participants and minimize learning effects. The allocation of whether there were four or eight pictures to select is even - meaning that each condition gets allocated twelve times eight and twelve times four and that each participant needs to select six times four pictures and six times eight pictures. Before the actual experiment, every participant was taken through a test run, to try out interaction with the gallery and avoid insecurities or problems during the real tests.

4.1.5 Photo Questionnaire

In those cases where the independent variable role is user, the images need to be familiar to the participant meaning that she has seen them before. As the filters should be tested especially for pictures with privacy sensitive content and all participant should be testing with the same images, it was not possible to use personal photographs of the participants. That is why all participants were sent an online image questionnaire, one week prior to the experiment, in the form of a set of pictures they clicked through and looked at. The pictures were accessible via a unique URL for each participant, inside an online survey tool, that ensured that everybody had looked through all the pictures exactly once and was also shown the cropped and resized thumbnail version of each image within this questionnaire. This was done to simulate a realistic use case: Assuming that smartphone owners regularly take pictures with their phones and assuming that they see the pictures once in detail and once in the thumbnail gallery, during photographing - ordinarily about one week passes until they need the picture again (e.g. to show it to friends) and search for it in the gallery.

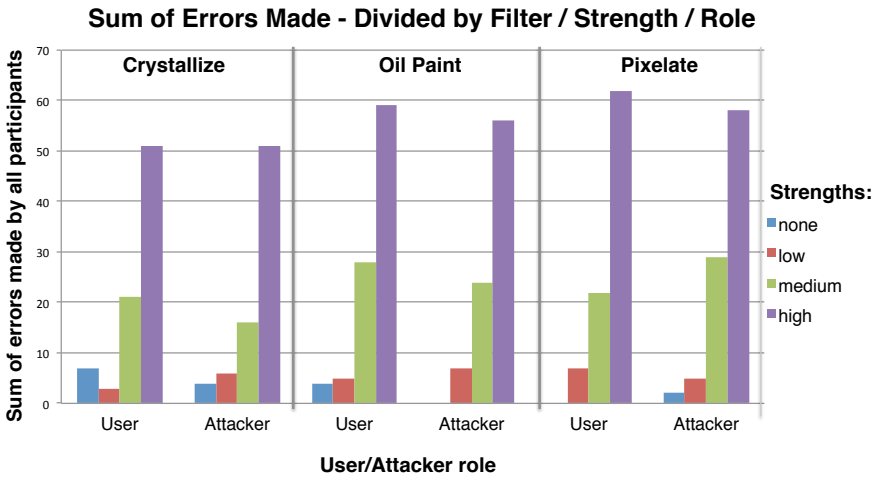

Figure 4.3: Sums of errors made by all participants during the selection task separated by filter, strength and role. The strengths medium and high show an increase of errors. There are no obvious differences between the roles user and attacker

4.2 Results

Eight of the 24 participants of the experiment were female and sixteen male. They were on average 25 years old (all between the ages of 21 and 33) and all of them have matriculation standard, eleven of them have a university degree. Thirteen of the participants reported to be nearsighted and one of them farsighted, but only seven of those wear glasses at all times and hence also wore them during the experiment. All but two participants own a smartphone for at least one year, but everyone stated to be familiar with using a smartphone and browsing photos on a smartphone. The participants successfully worked through the photo questionnaire one week prior to the study. They were instructed to look at each of the 84 survey pages for at least three seconds and complied with this request by looking at the pages for 6.85 seconds on average.

4.2.1 Errors

Each time the participants selected a photograph with privacy insensitive content (false positive) or failed to select a photograph with sensitive content (false negative) an error was recorded. The total amount of errors made by the participants (false positives and false negatives) needs to be looked at separately for the roles user (familiar photographs) and attacker (previously unknown photographs) to find out whether the familiarity of the previously seen images had an influence on the participants' ability to recognize the content of the pictures. Figure 4.4 shows a histogram of the sum of errors made by all participants. It appears that the strengths medium and high produce an increase of mistakes. There are no visible or distinct differences between the roles user and attacker. The figure also gives reason to believe that the strength low performed rather poorly at obfuscating image content. For both amounts of error counts (user and attacker) the Kolmogorov-Smirnov test as well as the Shapiro-Wilk test are significant ($p < .05$) for all filters and strengths (see attachment B, table A.1). The values of the dependent variable "errors" are not normally distributed. Consequently, non-paramteric tests (Wilcoxon test and Friedman test) are used to analyze the data. Comparing the amount of errors participants made as users and as attackers with a Wilcoxon test shows no significant differences between the two for any of the filter-strength combinations (see attachment B, table A.2). Thus, it appears that simulating the user perspective by showing the participants the original photographs prior to the experiment did not have a visible effect on their ability to recognize the filtered pictures. As this experiment did not succeed in

Filter	Chi-Square (3)	Asymp. Sig.
Oil Paint	49.912	.000
Crystallize	50.321	.000
Pixelate	55.031	.000

Table 4.1: Significance and Chi-Square values of the Friedman test comparing the amount of errors made at the different levels of strength (high, medium, low, none) separated by filter. There are significant ($p < .05$) differences. This indicates that for all three filters, some levels of strength performed significantly better in obfuscating the images content, hence leading to the participants making more mistakes

simulating an authentic user perspective, only the results for the attacker role are valid and will be included in further tests. A Friedman test reveals that for all three filter types (Crystallize, Oil, Pixelate) the amount of errors made at the different levels of strength (none, low, medium, high) differs significantly ($p < .05$) (see table 4.1). To analyze these differences, a follow-up Wilcoxon test was conducted. Thus the significance level is corrected to $p < .025$. For the Crystallize and the Pixelate filter, the amount of errors made at filter strengths high and medium is significantly ($p < .025$) higher than at the baseline filter strength none. However the amount of errors made at the filter strength low is not significantly different to the baseline. Looking at the Oil filter, it appears that not only the amount of errors made at strength high and medium, but also at strength low significantly ($p < .025$) differs from the baseline none. Table 4.2 shows all significance values. It should be noted that total error count for the Oil Paint baseline (none) is constant and equals zero. The settings Pixelate-none as well as Crystallize-none all produced at least some errors. As all filter types at strength none show the unfiltered original image, this difference among the three baselines cannot be caused by the filter type itself and is most likely random. It might however be an explanation for the significant differences of Oil Paint filter none to Oil Paint filter low.

Filter	none-high T-TestStatistic (p-value)	none-medium T-TestStatistic (p-value)	none-low T-TestStatistic (p-value)
Oil Paint	0.00 (.000)	0.00 (.001)	0.00 (.008)
Crystallize	0.00 (.000)	12.00 (.007)	2.50 (.317)
Pixelate	0.00 (.000)	0.00 (.001)	2.00 (.257)

Table 4.2: Significance values and test statistics of follow-up Wilcoxon tests to analyze differences between levels of filter strength. There are significant ($p < .025$) differences between the strengths high and none as well as medium and none for all filters. The Oil Paint filter also shows a significant difference between low and none

4.2.2 Time

The participants needed 15.46 ± 7.61 seconds on average to find and select the sensitive images in one gallery. During the experiment it became obvious that the participants used different strategies to solve the selection task. Some scanned the gallery and all images first before then quickly making their selection. Others looked at each image individually and selected a chosen image right away. There were also mixtures of those two strategies. This leads to the problem that the times

that pass before or between the selections vary greatly and are by no means comparable. Thus, the only timespan that can be analyzed and to compare the different filters and strengths is the total time (from the moment the gallery grid is loaded, until the participant confirms her selection by clicking on the "ready" button at the top of the prototype gallery). Although Friedman's test shows significant ($p < .05$, Chi-Square = 9.250) differences between the total time needed by participants between the three different filters at strength high the follow-up Wilcoxon test with corrected significance ($p < .025$) can not confirm any significant differences between the three filter types. Apart from this, no statistically significant differences were found between any of the filter types or strengths. Presumably, the time needed to recognize image content increases with filter strength, but at the same time, as the amount of recognizable content decreases, so does the time needed to complete the selection (because fewer images are selected).

4.2.3 Content

The participants' task was to select all privacy sensitive pictures that either showed nudity or drug us. Looking at the amount of errors made in the attacker role, a Wilcoxon test reveals that the content of the pictures had an influence on the amount of mistakes at two filter-strength combinations: Oil Paint at strength high and Pixelate at strength medium. Here, the amount of false negatives of drug use images is significantly ($p < .05$) higher than the amount of false negatives of images showing nudity (see attachment B, table A.3). It appears that at those two filter settings the participants (in the attacker role) had less difficulties recognizing photos displaying nudity than photos displaying drug use.

4.2.4 Feedback

A lot of valuable feedback could be gathered from the participants of this study. All of them agreed that the topic of privacy protection during mobile photo browsing is important and there are few to none satisfying solutions to-date. A large majority of the participants indicated great interest in a privacy enhanced photo gallery and stated that they could imagine to use such a gallery on their personal smartphone. However, it was also evident that the usability of and the control over a private gallery are of utmost importance. At all times the owner of the gallery wants to be in complete control over whether his images are filtered or not and possibly even to which extent they are filtered. Not being able to recognize and thus use one's personal images anymore would obviously be a criterion of refusal for all participants. Some participants noted that they would not want to arouse suspicion from their friends and family by using such an obfuscated gallery. In other words, they believe that a filtered, distorted gallery could actually increase the interest in the contents of the gallery. However, if the protection works well enough, an increased interest does not necessarily mean an increased risk. Although such concerns need to be addressed, one should not abandon means of protecting one's privacy, just because they might raise suspicion from others.

5 Discussion

Viewing through the total of 84 pictures did not have the desired effect that the used photos were still implicitly familiar to the participants one week later during the experiment. It is possible that this is due to the large amount of images that had to be used or due to the time that passed between viewing the pictures and having to recognize them (one week). The way the user perspective was simulated in this experiment had some weaknesses and resulted in only the attacker role data samples being analyzed further. Nevertheless this only means, that the familiarity of photos could not be simulated correctly. It does not confound the general concept or the research questions.

The results of this first study confirm that all three selected filters (Oil Paint, Crystallize, Pixelate) can obfuscate the content of the gallery images to an extent that caused the participants to make more mistakes during the selection task. While the filter strengths medium and high for all filters produced distinct increases in the amount of errors, the strength low performed rather poorly. In terms of errors made, none of the three filter types clearly outperforms the others. In summary, it appears that the selection process for the set of filters came up with three comparable and suitable filter types that (at the strengths medium and high) can all be used for protecting gallery images on a smartphone.

The equally well performing filter types can not be distinguished by the time participants needed to find and select photographs in a gallery. It appears that obfuscating gallery images does not have a huge effect on the time needed to scan a gallery (be it the time until an image is found or the time until someone decides he is not able to recognize anymore content). Due to the already mentioned different methods of searching and selecting images it was not possible to gather further insights on time differences among different images or between the attacker and user role (on this, also refer to the paragraph on limitations).

Two filter settings (Oil Paint at strength high and Pixelate at strength medium) lead to the result that participants performed better at recognizing pictures displaying nudity than they did at recognizing pictures displaying drug use. Inspite of selecting the pictures showing nudity carefully to not be too bold, the rather big portions of skin color within those images still seem to make recognition a little easier, even when the picture is unfamiliar. Although the Crystallize filter also preserves those color proportions within the image it did not show a significant difference. It could be that the polygon shaped regions confused the participants - because they are not used to this form of distortion - resulting in them being less confident about selecting a picture following a speculation.

6 Limitations

Despite all the interesting learnings from this user study, there are also some limitations that need to be considered. The issue of not succeeding in simulating familiarity of photos correctly lead to the loss of insights on the user perspective. Nevertheless by neglecting the user role data, it was possible to analyze the results for the attacker role without corruption. The results of the time needed by the participants are the only exception. Due to the widely varying methods of searching and selecting images (and with that widely varying differences between certain interaction timestamps) it is not possible to regard the isolated time values of the attacker role. One can only look at the total time it took participants to select user and attacker role images. Considering that there were not significant differences between the amount of errors of the two roles it is, however, safe to assume that the user role elements do not strongly distort the results. As every photo gallery is somewhat unique, it was attempted to make a diverse selection of images that is representative for the average smartphone gallery. Still, the results of this experiment might not apply to each and every composition of smartphone galleries there are. Especially galleries that are highly homogeneous could change the user's or the attacker's ability to recognize distorted photographs.

7 Summary

In the process of the user study, to evaluate the selected filters, a lot of valuable insights into the matter were generated and used to refine the methods for further research. For future studies, we need to consider that users and attacker have different main interests and thus their performances need to be tested differently. When the user or owner of a gallery looks for an image, she has a specific image in mind and needs to find and recognize this specific image. The attacker's

main interest, on the other hand, is to recognize content and subjects of pictures. The design of the second study takes this into account by separating the tested tasks for the user and attacker role.

Furthermore, the results of the first study reveal that merely showing the participants several original images, one week prior to the study, was not sufficient to create the effect of familiarity that is needed for the concept to work. Research on memorability of stimuli suggests that creating pictures enhances the memorability of these pictures (this is called generation effect) [11]. The act of taking a photo probably also leads to emotions (about the location, the situation, etc.) being connected with that photograph and thus making recollection of it easier. Consequently, to create a realistic use case experiment, the participants need to provide their own personal smartphone photographs. This is the only way to be sure that a bad performance on recognition or recall is not caused by those missing emotions or the missing generation effect.

In summary, the filter selection and evaluation process successfully brought out a set of three suitable filters to use for further evaluation of the concept of a privacy enhanced filtered gallery.

References

[1] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

[2] V. Bruce, Z. Henderson, C. Newman, and A. M. Burton. Matching identities of familiar and unfamiliar faces caught on cctv images. *Journal of Experimental Psychology: Applied*, 7(3):207, 2001.

[3] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10(3):243–248, 1999.

[4] J. Demanet, K. Dhont, L. Notebaert, S. Pattyn, and A. Vandierendonck. Pixelating familiar people in the media: Should masking be taken at face value? *Psychologica belgica*, 47(4):261–276, 2007.

[5] T. Denning, K. Bowers, M. van Dijk, and A. Juels. Exploring implicit memory for painless password recovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2615–2618, New York, NY, USA, 2011. ACM.

[6] R. L. Gregory. Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358):1121–1127, 1997.

[7] A. Harada, T. Isarida, T. Mizuno, and M. Nishigaki. A user authentication system using schema of visual memory. In *Proceedings of the Second International Conference on Biologically Inspired Approaches to Advanced Information Technology*, BioADIT'06, pages 338–345, Berlin, Heidelberg, 2006. Springer-Verlag.

[8] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig. Use your illusion: Secure authentication usable anywhere. In *Proceedings of the 4th Symposium on Usable Privacy and Security*, SOUPS '08, pages 35–45, New York, NY, USA, 2008. ACM.

[9] E. Hayashi, J. Hong, and N. Christin. Security through a different kind of obscurity: evaluating distortion in graphical authentication schemes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2055–2064. ACM, 2011.

[10] G. Holzmann. *Beyond photography: the digital darkroom*. Prentice Hall software series. Prentice-Hall, 1988.

[11] H. Kinjo and J. G. Snodgrass. Does the generation effect occur for pictures? *The American journal of psychology*, 2000.

[12] P. Korshunov, C. Araimo, F. De Simone, C. Velardo, J.-L. Dugelay, and T. Ebrahimi. Subjective study of privacy filters in video surveillance. In *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, pages 378–382. Ieee, 2012.

[13] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

[14] F. Tripp. A privacy-respectful photo browsing (gallery) app for smartphones. Bachelors thesis, Ludwig-Maximilians-Universitaet Muenchen, 2013.

[15] Z. Wang, J. Jing, and L. Li. Time evolving graphical password for securing mobile devices. In *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, pages 347–352. ACM, 2013.

[16] J. Zhang, V. L. Patel, T. R. Johnson, and E. H. Shortliffe. A cognitive taxonomy of medical errors. *J. of Biomedical Informatics*, 37(3):193–204, June 2004.

A Additional Figures and Tables

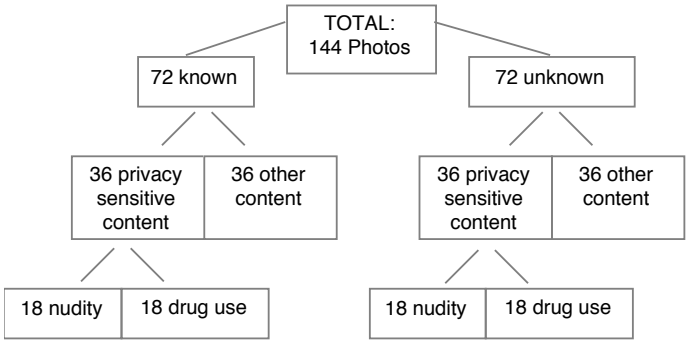

Figure A.1: Explains how the 144 photographs consist of different types of images to achieve an even distribution of the dependent variables role and kind of sensitive content

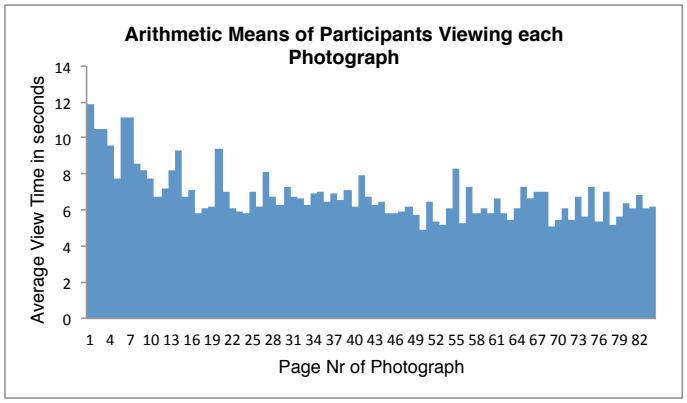

Figure A.2: Arithmetic means of participants view times of each page displaying an image that is later utilized in the study as a familiar image. Participants were instructed to look at each image for at least three seconds and looked at most of the pages for 6.85 seconds on average. Naturally the first few pages were watched a little longer until the participants got used to the survey tool and knew what to do

Filter	Strength	Role	Kolmogorov-Smirnov	Shapiro-Wilk
Oil Paint	high	attacker	.021	.023
Oil Paint	high	user	.007	.001
Oil Paint	medium	attacker	.004	.001
Oil Paint	medium	user	.000	.001
Oil Paint	low	attacker	.000	.000
Oil Paint	low	user	.000	.000
Oil Paint	none	attacker	.000	.000
Oil Paint	none	user	.000	.000
Crystallize	high	attacker	.000	.001
Crystallize	high	user	.004	.003
Crystallize	medium	attacker	.000	.000
Crystallize	medium	user	.000	.001
Crystallize	low	attacker	.000	.000
Crystallize	low	user	.000	.000
Crystallize	none	attacker	.000	.000
Crystallize	none	user	.000	.000
Pixelate	high	attacker	.002	.005
Pixelate	high	user	.000	.007
Pixelate	medium	attacker	.008	.004
Pixelate	medium	user	.000	.000
Pixelate	low	attacker	.000	.000
Pixelate	low	user	.000	.000
Pixelate	none	attacker	.000	.000
Pixelate	none	user	.000	.000

Table A.1: Significance values of both Kolmogorov-Smirnov and Shapiro-Wilk tests for the amount of errors made separated by role (user/attacker), filter and strength. Both tests are significant ($p < .05$) for all of those values. This means that the values of the dependent variable "errors" are not normally distributed

Filter	Strength	T-Test Statistic	Asymp. Sig. (2-tailed) user-attacker
Oil Paint	high	56.00	.508
Oil Paint	medium	74.50	.921
Oil Paint	low	13.50	.480
Oil Paint	none	0.00	.059
Crystallize	high	54.00	.726
Crystallize	medium	35.00	.264
Crystallize	low	13.50	.480
Crystallize	none	5.00	.480
Pixelate	high	23.00	.627
Pixelate	medium	92.00	.403
Pixelate	low	18.00	.564
Pixelate	none	0.00	.157

Table A.2: Significance values and test statistics of Wilcoxon test comparing the amount of errors made in the user and attacker role (familiar and unknown pictures) separated by filter and strength. There are no significant ($p < .05$) differences between the two error scores. This means that showing the photographs to the participants once, one week before the experiment, was not sufficient to create a familiarity effect that improves the ability to recognize the distorted images

Filter	Strength	T-Test Statistic	Asymp. Sig. (2-tailed) nudity-drug use
Oil Paint	high	5.50	.011
Oil Paint	medium	30.00	.763
Oil Paint	low	8.00	.257
Oil Paint	none	0.00	1.000
Crystallize	high	18.00	.564
Crystallize	medium	32.50	.317
Crystallize	low	6.00	.655
Crystallize	none	2.00	.564
Pixelate	high	3.50	.102
Pixelate	medium	9.00	.026
Pixelate	low	1.50	1.000
Pixelate	none	0.00	.157

Table A.3: Significance values and test statistic of Wilcoxon test comparing the amount of false negative errors made by the participants for pictures showing nudity compared to pictures showing drug use. Oil Paint at strength high and Pixelate at strength medium show significant ($p < .05$) differences between the two. This means that for those filter settings the kind of privacy sensitive content had an influence on the participants' ability to recognize it