Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems | School of Information Systems

4-2018

Empirical study of face authentication systems under OSNFD attacks

Yan LI
Singapore Management University, yan.li.2009@smu.edu.sg

Yingjiu LI
Singapore Management University, yjli@smu.edu.sg

XU, KE
Singapore Management University, kexu.2013@phdis.smu.edu.sg

Qiang YAN
Singapore Management University, qiang.yan.2008@smu.edu.sg

Robert H. DENG
Singapore Management University, robertdeng@smu.edu.sg

DOI: https://doi.org/10.1109/TDSC.2016.2550459

Citation

LI, Yan; Yingjiu LI; XU, KE; YAN, Qiang; and DENG, Robert H.. Empirical study of face authentication systems under OSNFD attacks. (2018). *IEEE Transactions on Dependable and Secure Computing*. 15, (2), 231-245. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/3339

Empirical Study of Face Authentication Systems under OSNFD Attacks

Yan Li, Yingjiu Li, Ke Xu, Qiang Yan, Robert H. Deng

Abstract—Face authentication has been widely available on smartphones, tablets, and laptops. As numerous personal images are published in online social networks (OSNs), OSN-based facial disclosure (OSNFD) creates significant threat against face authentication. We make the first attempt to quantitatively measure OSNFD threat to real-world face authentication systems on smartphones, tablets, and laptops. Our results show that the percentage of vulnerable users that are subject to spoofing attacks is high, which is about 64% for laptop users, and 93% smartphone/tablet users. We investigate liveness detection methods in the real-world face authentication systems against OSNFD threat. We discover that under protection of liveness detection, the percentage of vulnerable images is 18.8%, but the percentage of vulnerable users is as high as 73.3%. This evidence suggests that the current face authentication systems are not strong enough under OSNFD attacks. Finally, we develop a risk estimation tool based on logistic regression, and analyze the impacts of key attributes of facial images on the OSNFD risk. Our statistical analysis reveals that the most influential attributes of facial images are image resolution, facial makeup, occluded eyes, and illumination. This tool can be used to evaluate OSNFD risk for OSN images to increase users' awareness of OSNFD.

Index Terms—Face authentication, online social networks, OSN-based facial disclosure, liveness detection

1 Introduction

Face authentication systems have been widely available on various consumer-level computing devices such as smartphones, tablets, and laptops which have built-in camera capability. Popular face authentication systems include Face Unlock [15], Facelock Pro [12], and Visidon [46] on smartphones/tablets, and Veriface [30], Luxand Blink [32], and FastAccess [47] on laptops. Face authentication requires zero memory efforts from users and usually generates higher entropy than legacy password [36]. Thus face authentication systems provide attractive alternatives of legacy passwords. Previously, the major obstacle for an adversary to compromise face authentication is the physical proximity required to capture a victim's facial images. However, this is no longer necessary as the emergence of online social networks (OSNs).

OSNs provide a platform for facilitating social interactions. Numerous personal data including personal images are published in OSNs such as Facebook and Google+ at every moment. For example, 350 million images are published by users on Facebook every day [48]. It is very likely that these images contain facial images where the users' faces can be clearly seen. These facial images could become an abundant resource for potential attackers to exploit, which introduces the threat of OSN-based facial disclosure (OSNFD). OSNFD affects the strength of face authentication as OSNFD can disclose facial images and compromise face authentication in a large scale.

To understand the threat of OSNFD against face authentication, we collect users' facial images published in OSNs and build a dataset containing important image attributes that are common in real-life photos but rarely used in prior controlled study on face authentication [8], [20]. Using these images, we simulate spoofing attacks against typical real-world face authentication systems which are designed for smartphones, tablets, and laptops. Since all target systems [12], [15], [30], [32], [46], [47] are closed-source with no programmable testing interfaces, enormous efforts are made for image collection and testing.

We make the first attempt to provide a quantitative measurement on the threat of OSNFD against typical face authentication systems in use. Our study reveals that the percentage of vulnerable users that are subject to OSNFD attacks is high, though the percentage of vulnerable images which can be used for OSNFD attacks is moderate. The percentage of vulnerable users is 77% on average. Our results are different for the systems on smartphones/tablets and on laptops. Further investigation shows that the quality of images is a more important factor affecting the success rate of spoofing attacks as compared to quantity.

In order to mitigate the OSNFD attacks, various liveness detection mechanisms are designed to distinguish between legitimate face biometrics and forged face biometrics. We examine the effectiveness of the liveness detection mechanisms available in the target face authentication systems. Our results show that when liveness detection is in use, the percentage of vulnerable images becomes low, which is 18.8% on average. However, the percentage of vulnerable users is still high, which is 73.3% on average. All these findings show that the current face authentication systems are not strong enough under OSNFD attacks.

- *Yan Li, Yingjiu Li, Ke Xu, Qiang Yan, and Robert H. Deng are with School of Information Systems, Singapore Management University. E-mail: {yan.li.2009, yjli, kexu.2013, qiang.yan.2008, robertdeng}@smu.edu.sg*

We develop a risk estimation tool to evaluate how likely certain facial images can be used in effective OSNFD attacks. Logistic regression is used to extract key attributes of facial images affecting the success rate of OSNFD spoofing attacks. Our statistical analysis shows that the success rate is significantly affected by image resolution, occlusion of eye, occlusion of mouth, blurred image, facial makeup, dim lighting condition, and illumination in general. We further investigate the statistical significance of these attributes for different face authentication systems with different security settings and on different platforms among different users. The proposed risk estimation tool achieves a precision of 81%, a recall of 83%, and an F1 score of 82% on average. It can help users evaluate the risk of uploading their images to OSNs, thus increasing their awareness of OSNFD threat. We further discuss the costs and implications of mitigating the threat of OSNFD spoofing attacks.

This paper extends a prior work [31] in that a comprehensive analysis on the effectiveness of liveness detection mechanisms and a detailed statistical analysis of characteristics of OSNFD are provided.

2 Background

2.1 Face Authentication System and Related Work

Face authentication is a biometrics-based user authentication mechanism, which verifies a user's identity by using information extracted from the user's facial features. As illustrated in Figure 1, a typical face authentication system uses a camera to capture the user's facial image/video as input, and then verifies it with enrolled biometric information for the claimed identity. The objective of a face authentication system is to recognize a user as long as the input is collected from the legitimate user, while rejecting the inputs from all other users.

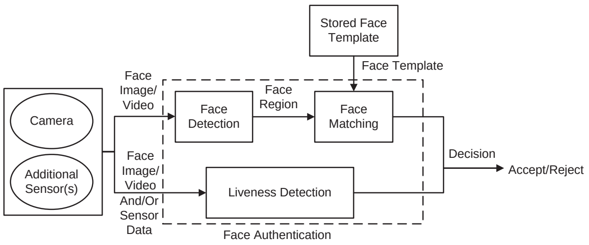


Fig. 1. Work flow of a typical face authentication system

Two key modules are involved in this verification process. The first module is the face detection module, which identifies the face region and removes irrelevant information of an image. The processed image is then passed to the next module named face matching. This module computes a similarity score for the input image based on an enrolled face template containing key features which can be used to distinguish a user from other users and imposters. Different algorithms may be used for these two modules, but all face authentication systems generally have these two modules and follow this work flow. In the end, a face authentication system outputs the final decision (i.e. accepting or rejecting a claim) according to whether or not the similarity score is higher than a matching threshold. This threshold is carefully chosen so as to achieve a proper balance between false rejection rate and false acceptance rate.

It is well-known that face authentication is subject to spoofing attacks. An attacker may compromise an authentication system by displaying some images or videos of a legitimate user in hard copies or on screen [4], [6], [13]. Liveness detection is a major countermeasure designed and deployed to mitigate the risk of spoofing attacks.

We summarize the closely related work in terms of face recognition and liveness detection in this section.

2.1.1 Related Work on Face Recognition

For face recognition, holistic approaches and local landmark based approaches have been studied before [1], [50]. The holistic approaches, such as PCA-based algorithms and LDA-based algorithms, use the whole face region as input. Local landmark based approaches extract local facial landmarks such as eyes, nose, mouth, etc and feed the locations and statistics of these local facial landmarks into a structure classifier.

Face authentication is an important application of face recognition. Trewin et al. [44] show that face authentication is faster, and it causes lower interruption of user memory recall in a comparison to other authentication solutions which base on voice, gesture, and typical password entry. Another advantage of face authentication is that it provides stronger defense against the repudiation threat than token based authentication and password based authentication [36]. Besides face authentication, face identification is another application of face recognition, which compares an input facial image with multiple registered users and identifies the user in the input image. Face identification may cause privacy leakage in OSNs due to identifiable personal images published in OSNs [19]. Compared to their work, we focus on the impact of publishing personal images in OSNs to the effectiveness of face authentication systems under OSNFD attacks.

2.1.2 Related Work on Liveness Detection

Liveness detection is designed to distinguish between legitimate input face biometrics from live users and forged face biometrics. Liveness detection methods can be categorized according to liveness indicators, including texture pattern, motion of 3D face, real-time response, and multimodal [7], [9], [24].

The texture pattern based liveness detection approaches detect specific texture patterns for fake facial images due to printing process or properties of digital screen. Maatta et al. propose to detect liveness by extracting local binary patterns from a single image [33]. IDIAP team takes local binary patterns from each video frame and builds a global histogram for liveness detection [7]. The above approaches usually require very diverse dataset of paper and printing texture patterns [24]. Akhtar et al. analyze the quality of input biometric images based on their local features and global structures for detecting the spoofing attacks [3]. Patel et al. propose to detect the spoofing attacks using face videos based on the analysis of moiré pattern which often appears on digital screens [38]. However, the moiré pattern on digital screen can be reduced or eliminated by resizing or rotating photos or applying mathematical filters [41]. The effectiveness of these approaches could be affected if the attack is performed using a photo/video displayed on a screen with high display resolution.

The motion based liveness detection approaches assume that a real face is a 3D object which moves differently from 2D fake faces. These approaches are usually associated with optical flow analysis because different patterns of optical flow represent differences between the movement of 3D faces and 2D faces [24]. Bao et al. analyzes optical flows generated from a holistic 3D face for liveness detection [5]. Kollreider et al. analyzes the optical flows based on the detection of ears, nose, and mouth as these facial landmarks generate different optical flow patterns [25]. Tirunagari et al. propose to use dynamic mode decomposition to analyze the movements of eyes, lip, and head and the local binary pattern from input videos [43]. However, the above approaches usually require high-quality input videos with high frame rates and ideal lighting which may be difficult to achieve in practice.

The real-time response based liveness detection approaches require interactions with users in real time, such as eye blink and head rotation. Pan et al. propose a solution which requires users to blink their eyes for liveness detection [37]. However, these approaches can be bypassed with multiple images or videos which contain the required liveness traits [39]. In order to mitigate this threat, some liveness detection approaches based on abrupt changes of motions and motion continuity [11], [35] are proposed if sufficient images of intermediate stages are available. Considering the case of eye blink, an eye blink is an activity containing rapid closing and opening of eyelid [37]. An eye blink usually lasts for 0.1-0.4 seconds [21]. Thus the motion of closing/opening eyes approximately takes 0.05-0.2 seconds. However, the front-facing cameras on existing mobile devices can capture 1-4 frames at most for this motion, which may not be sufficient. This situation would change if high-speed cameras become popular in the future.

Multimodal based liveness detection approaches take face biometrics and other biometrics together, such as fingerprint and iris for liveness detection [13]. Fingerprint refers to the flowing pattern of ridges and furrows located on the tip of a finger while iris consists of a random structure of minutiae or points of detail. The multimodal based approaches rely on the fusion of face biometrics and these other biometrics for liveness detection [6], [13]. However, these approaches require additional hardware or must be used in specific environment.

Some liveness detection approaches combine multiple liveness indicators in order to defend against the spoofing attacks [7], [9]. AMILAB team proposes to detect liveness based on texture pattern, motion of 3D face, and real-time response together [7]. CASIA team proposes to combine texture pattern and motion of 3D face in liveness detection [9]. They analyze the texture by multi-scale local binary patterns and the 3D face by dense optical flows. These approaches can be affected by the quality of input images/videos and illumination.

The summary of these liveness detection methods is presented in Table 1. While these liveness detection methods can be used to thwart image and video spoofing attacks to a certain degree, we focus on the impact of OSNFD attacks to typical face authentication systems in use, and develop a risk estimation tool to increase users' awareness before they publish their personal images in OSNs.

TABLE 1
Summary of the existing liveness detection methods.

Types of liveness detection	Liveness detection method	Features	Detection of attacks
Texture pattern	Maatta et al. [33]	Local binary patterns	Photo spoof
	IDIAP team [7]	Local binary patterns, global histogram	Photo spoof
	Akhtar et al. [3]	Local features, global structures	Photo spoof Video spoof
	Patel et al. [38]	Moiré patterns	Video spoof
Motion	Bao et al. [5]	Optical flow field	Photo spoof
	Kollreider [7]	Optical flow patterns	Photo spoof
	Tirunagari et al. [43]	Face dynamics, local binary pattern	Photo spoof Video spoof
Real-time response	Pan et al. [37]	Eyeblink	Photo spoof
	NG et al. [35]	Eyeblink	Photo spoof
Multimodal	Galbally et al. [13]	Face, fingerprint	Photo spoof Video spoof
Combination	AMILAB [7]	Texture, face movement, eye blink	Photo spoof
	CASIA [9]	Texture, face movement	Photo spoof Video spoof

2.2 OSN-based Facial Disclosure and Threat Model

The OSN-based facial disclosure (OSNFD) addresses the issue when users' face biometrics is involuntarily disclosed by sharing personal images in OSNs. These disclosed face biometrics would raise security risks against face authentication systems.

The impact of the spoofing attacks was believed to be limited due to the requirement that an adversary had to be physically close to a victim in order to collect the required information. Therefore, it is generally considered sufficiently secure as an authentication factor for common access protection [6].

However, this belief may be questionable since OSNFD becomes a common phenomenon. OSNFD supplies an adversary with abundant facial images to exploit and makes large-scale identity theft possible for those who use face authentication. Our work investigates the OSNFD threat and quantitatively measures its impacts. We consider OSNFD-based attacks where an adversary attempts to forge a valid input from image resources disclosed from OSNFD so as to pass face authentication. Our study focuses on image-based attacks unless explicitly mentioned.

The OSNFD threat may be mitigated with liveness detection technologies, which rely on extra information sources or heuristic algorithms to distinguish a live user from a captured image/video. All the existing sophisticated liveness detection technologies associate with considerable costs, which will be explained later in Section 5.1. This may explain that only weak liveness detection technologies are currently deployed on the face authentication systems designed for consumer-level computing devices [25], [37]. For example, eye blinking detection is a common heuristic used by many face authentication systems [15], [37], [46] including Google's Face Unlock; however, it can be easily bypassed using two facial images as demonstrated in [39]. Similar tricks can also apply to other weak liveness detection mechanisms such as head rotation detection. The detailed evaluation on the effectiveness of the liveness detection mechanisms will be presented in Section 3.2.4. Even worse is that the existing liveness detection mechanisms are disabled by default in most popular face authentication systems [15], [30], [46], as they may have negative impacts on accessibility.

3 User Study and Empirical Results

In order to quantitatively measure the impacts of OSNFD, we conduct a user study to collect real personal images that have been shared in OSNs. The collected images are used to test against real-world face authentication systems chosen from the most popular face authentication products in terms of user base [16], [45]. Among these face authentication systems, Face Unlock [15], Facelock Pro [12], and Visidon [46] are designed for the platforms on smartphones/tablets while Veriface [30], Luxand Blink [32], and FastAccess [47] are designed for the platforms on laptops or desktops. This section describes the detailed process of data collection and the results of our empirical analysis. We use the following classifications in our discussion.

First, we classify the security settings of a face authentication system into *low* and *high*. Most of face authentication products [12], [15], [30], [32], [46], [47] provide very limited choices on security settings that generally affect the recognition threshold used in the face matching module. For example, Google's Face Unlock [15] does not provide any option for users to adjust its security strength. Most of our tested products [12], [30], [32], [46], [47] only have two options for users, labeled as "high accessibility" (i.e. low security) and "high security". Only Lenovo's Veriface [30] provides a scrollbar for users to adjust its security strength from the lowest to the highest. Therefore, we use "low" to indicate that a target system enforces the weakest security protection, and use "high" to indicate the strongest security protection achievable to the system.

Second, we classify face authentication systems into *mobile* and *traditional*. A system is labeled as mobile if it is used for smartphones or tablets, while a traditional system is used for laptops or desktops. A mobile system is usually more tolerant to varied environments, as it should be accessible no matter where a user uses the device. Laptops is considered as traditional as it is not expected to be used from anywhere at any time like what users expect smartphones and tablets.

Third, we classify users into different groups according to the pattern of their sharing behaviors. As observed in our study, it is quite common that a user tends to upload edited images where facial landmarks are significant changed to create better visual appeal. Therefore, it is also an important factor that needs to be considered.

These classifications represent three major factors that affect the effectiveness of OSNFD-based attacks, which are security settings, target platforms, and user behaviors, respectively. We use them as controlled parameters to evaluate the severity of OSNFD, and more sophisticated statistical analysis will be given in the next section to identify the key attributes that can be used to mitigate the OSNFD threat.

3.1 User Study and Data Collection

74 participants are involved in our user study, including 36 males and 38 females with an age range between 19 and 35. Most of these participants are university students. Each participant is paid with 10 dollars as a compensation. The study consists of three parts, all conducted in a quiet room. In the first part, we ask each participant to select and download 20 *facial* images published within the last 12 months in popular OSNs such as Facebook, Google+, Instagram, and etc. The downloaded facial images are used for spoofing attacks to the face authentication systems in our test. A facial image is defined as an image where a participant's face can be seen. A participant's face may not be perfect due to many negative effects such as blur, occlusion (e.g. covered by a sunglasses), head rotation. Such negative effects are examined in our study.

In the second part, we capture each participant's facial images with 35 controlled head poses and 5 typical facial expressions using a Canon EOS 60D (18.0-megapixel D-SLR CMOS camera). The resulting images are 5184×3456 in size with the inner pupil distance of the subjects typically exceeding 400 pixels. 35 controlled head poses are specified by both horizontal and vertical rotations. Rotation angles are represented as (rot_H, rot_V), where rot_H is the angle of horizontal rotation while rot_V is the angle of vertical rotation. The value range of rot_H includes $0°$, $10°$ to left/right, $20°$ to left/right, and $30°$ to left/right, while the value range of rot_V includes $0°$, $10°$ to up/down, and $20°$ to up/down. We choose these boundary values according to the common restriction of existing face authentication systems [1], which indicates that a participant cannot pass user authentication if rot_H exceeds $30°$ or rot_V exceeds $20°$ degrees. In our test, the captured images with head rotation are used to examine the impact of head poses to the face authentication systems by displaying these images to the camera on an LCD screen. They also serve as the ground truth for labeling the head poses of downloaded facial images.

Each participant's facial images are captured with not only controlled head poses, but also typical facial expressions, including neutral expression, smile without showing teeth, smile showing teeth, closed eyes, and open mouth. The images with facial expressions are used to investigate the impact of facial expressions to the target face authentication systems. During image capturing, a continuous lighting system is used to eliminate the shadow on participants' faces.

We use a helmet equipped with a gyroscope to control the head rotation of participants. The use of gyroscope achieves a low measuring error of less than 1 degree for measuring head rotation in all cases [34]. For each head pose, we ask participants to face to the DSLR camera and help them adjust their heads to the frontal position in the way similar to [20]. Then the participants rotate their heads to the required angles with help of the gyroscope. The gyroscope generates real-time rotation angles and broadcasts them via WiFi. This rotation information is received and displayed on an iPad screen, and shown to the participants. Then, we ask the participants to hold their head poses while we remove their helmet gently without causing any movement of their heads during the helmet removal. After that, the images of each head pose are captured immediately.

In the final part, each participant is asked to fill in a questionnaire for collecting the participant's attitudes towards the usage of face authentication systems and OSNs.

3.2 Empirical Results

Based on collected images, we inspect the realistic threat of OSNFD against the latest versions of popular real-world face authentication systems. Our experiment procedure is similar to prior work [8], [28], which is described as follows: Each participant enrolls his/her frontal images into all tested face authentication systems in a quiet room with normal lighting. During the enrollment, each of the face authentication systems enrolls participant's faces via a built-in camera. Note that except Facelock Pro, the enrollment processes of the tested face authentication systems are automatic and similar which do not require any participants' interference. The enrollment process by Facelock Pro differs from the other face authentication systems in that users need to click on a button in order to trigger image capturing by Facelock Pro. After enrollment, we use each participant's own OSN images to test whether they can be used to log in a target face authentication system for his/her own account. The authentication processes of these tested systems are automatic and essentially the same. The participant's OSN images are displayed on an LCD screen with a resolution of 1600×900 pixels. The result on whether a target system can be spoofed by an OSN image is recorded for each target system and for each image.

In our user study, we are interested in *vulnerable images* and *vulnerable users*. A vulnerable image, denoted by $VulImage$, is defined as a facial image which is wrongly accepted as a genuine user by a face authentication system during user authentication and therefore enables an adversary to circumvent the face authentication system. The examples of vulnerable/non-vulnerable images are shown in Figure 2. A vulnerable user, denoted by $VulUser$, is a user enrolled in a face authentication system who has at least one vulnerable image published in OSNs.

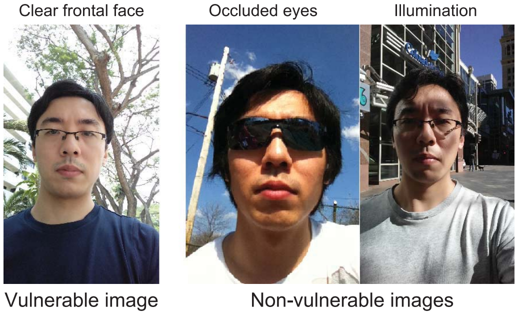


Fig. 2. Examples of vulnerable/non-vulnerable images

Table 2 shows that all tested face authentication systems are vulnerable to OSNFD in general. On average, 39% of OSN images and 77% of participants are vulnerable. Among these face authentication systems, Visidon is more vulnerable at its low security level, for which 68% of the images and 97% of the participants are vulnerable. Note that Google's Face Unlock comes as a built-in feature in all Android-based systems whose versions are higher than 4.0 [15]; 45% of the OSN images and 86% of the participants are vulnerable in this case.

TABLE 2
Overall percentage of $VulImage$ and $VulUser$

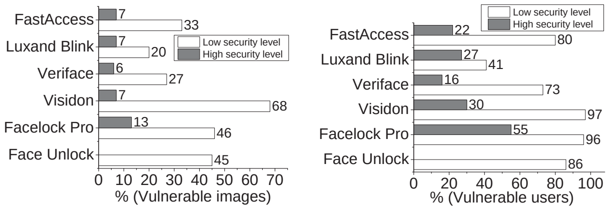
	$VulImage\%$	$VulUser\%$
Face Unlock	45%	86%
Facelock Pro	46%	96%
Visidon	68%	97%
Veriface	27%	73%
Luxand Blink	20%	41%
FastAccess	33%	80%
Average	39%	77%

Although the percentage of vulnerable images is moderate in Table 2, the quantity of vulnerable images is very large due to the huge amount of images in OSNs. The large amount of vulnerable images existing in OSNs create an online arsenal for potential attacks. Since users usually share their personal images with their friends in OSNs, most of them tend to publish their images in which their faces can be clearly viewed. Consequently, the percentage of vulnerable users is high as observed in our study.

In the following, we analyze the security settings, target platforms, and user behaviors to the effectiveness of OSNFD attacks in terms of vulnerable images and vulnerable users. We also evaluate the effectiveness of the liveness detection mechanisms available in the target face authentication systems which can be used to mitigate OSNFD attacks to a certain degree.

3.2.1 Impacts of Security Settings

Security settings specify the security strength of a face authentication system against potential attacks. As previously explained, most of face authentication products [12], [15], [30], [32], [46], [47] provide very limited choices on security level. So we focus our analysis on lowest and highest security level that can be provided by each system, which are denoted as low security and high security, respectively. Since there is only one security level in Face Unlock and the observed security strength of Face Unlock is comparable to the other systems in low security level, we classify its security level as low. As expected, Figure 3 shows that the face authentication systems in low security level are facing more severe OSNFD threat than those in high security level. On average, 40% of the images and 79% of the participants are vulnerable for the face authentication systems in low security level while 8% of the images and 30% of the participants are vulnerable for the face authentication systems in high security level.



(a)

(b)

Fig. 3. Percentage of $VulImage$ and $VulUser$ in different security levels

The change of security settings generally affects the recognition threshold in the face matching module. As the security level is raised, the recognition threshold becomes higher which imposes more restrictions for matching between login facial image and pre-stored facial image. Therefore the face authentication imposes more rigid restrictions on the login facial image. The major restrictions observed in our study are head pose and lighting condition.

For head pose, we use *acceptable head pose range* to measure the tolerance of a face authentication system on head pose variations. It describes the head rotation range of head poses with which at least 50% of the participants successfully log in the face authentication systems. In these tests, we use the images collected with controlled head poses as test inputs (i.e. login images) for the systems where the participants' frontal face images are enrolled as described at the beginning of Section 3.2. The results show that the systems in low security are more tolerant for the variations of head poses than the systems in high security by about $10°$.

For lighting condition, we further classify it into different types of illumination and low lighting [14], [26], [50]. The face authentication systems in low security level are observed to have higher tolerance for variation of lighting conditions than the systems in high security level. In our study, illumination is observed in 27% (394 out of 1440) of the OSN images while low lighting is observed in 18% (266 out of 1440) of the OSN images. On average, 81% of the OSN images with illumination and 79% of the OSN images with low lighting cannot be used to log in the face authentication systems in low security level while 96% of the OSN images with illumination and 94% of the OSN images with low lighting cannot be used to log in the systems in high security level.

On the other hand, a face authentication system in low security level has higher tolerance for varied login environments, which is necessary for the system to be usable in the complex environments. Clear tradeoffs between security and accessibility are observed on our tested systems at different security settings. An increase in security strength inevitably decreases the accessibility. We conduct a follow-up experiment to collect quantitative evidence for the impact of these tradeoffs.

20 participants are invited to this follow-up study. The participants' facial images have been enrolled in the face authentication systems during our user study as described in Section 3.2. To mimic different login environments, the experiments are conducted between 2pm-4pm in a sunny day at four fixed indoor/outdoor locations, including 1) a meeting room with a normal lighting condition, 2) a meeting room with a dim lighting condition, 3) outdoor ground in the sunshine, and 4) shelter of a building. This setting simulates a situation when a user registers to an authentication system in one place, but uses it in many other places. The participants are asked to login to each face authentication system without activating any liveness detection. In this experiment, no OSN images are used; only live legitimate users attempt to login. Each participant has at most three chances for each login before we record it as a false rejection.

Table 3 shows the false rejection rates of the face authentication systems at the low security level are lower than those at the high security level. The highest false rejection rate observed is 85% for Veriface at its high security level. This will cause a significant concern on the accessibility. From our questionnaire on user perception, 70% of the participants think it is important to successfully log in their smartphones, tablets, or laptops at the time they want to use. If the face authentication system is not always functional, 67% of the participants give up using the system which causes the serious accessibility problem to their devices. This may also explain why the popular face authentication systems always use low security level by default.

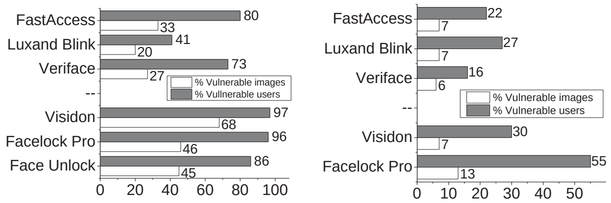
TABLE 3
Significant increase in false rejection rates when using high security level settings. The increments of false rejection rates are more significant for traditional platform-based systems (the last three systems).

System	Security level	Room+ normal lighting	Room+ dim lighting	Outdoor ground	Shelter
Face Unlock	N/A	0%	5%	10%	0%
Facelock Pro	Low	0%	10%	10%	0%
	High	0%	45%	60%	25%
Visidon	Low	0%	5%	5%	0%
	High	5%	55%	65%	50%
Veriface	Low	0%	25%	35%	20%
	High	10%	60%	85%	60%
Luxand Blink	Low	0%	30%	50%	45%
	High	5%	55%	70%	55%
FastAccess	Low	0%	15%	30%	15%
	High	5%	55%	65%	55%

3.2.2 Impacts of Target Platforms

The target platform of a face authentication system imposes the platform-specific requirements on both security and usability. In our tested systems, Face Unlock, Facelock Pro, and Visidon are targeting for mobile platform, while Veriface, Luxand Blink, and FastAccess are targeting for traditional platform.

Figure 4 shows that the OSNFD threat for mobile platform is generally more severe than the OSNFD threat for traditional platform. On average, in low security level, 53% of the images and 93% of the participants are vulnerable for the face authentication systems on mobile platform while 27% of the images and 64% of the participants are vulnerable for the systems on traditional platform. In high security level, 10% of the images and 43% of the participants are vulnerable for the face authentication systems on mobile platform while 7% of the images and 22% of the participants are vulnerable for the face authentication systems on traditional platform.



(a) Low security level (b) High security level

Fig. 4. Difference in $VulImage$ and $VulUser$ between systems targeting for mobile platform and traditional platform.

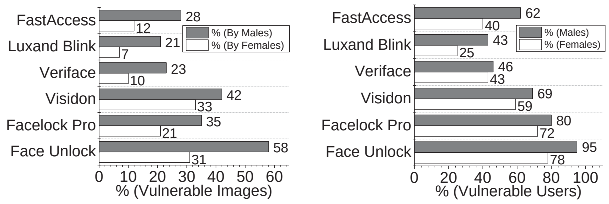
These results clearly show the difference caused by platform-specific requirements. Compared to a traditional system, a mobile system is usually designed to be more robust and more tolerant to varied environments such as outdoor environment in order to meet accessibility expectation by users. Meanwhile it leads to the more severe OSNFD threat for mobile platform based systems. This difference is confirmed by the results of our questionnaire, which shows that 91% of the participants believe that it is important to log in smartphones or tablets in both indoor and outdoor environment while only 36% of the participants think it is important to log in laptops in both indoor and outdoor environment.

This difference is also revealed in our tests on head pose and lighting condition. Our results show the systems targeting for mobile platform have higher tolerance for variations of the head poses than the systems targeting for traditional platform by about 10°.

Our tests on lighting conditions further show the face authentication systems targeting for mobile platform are more tolerant to variations of the lighting conditions. In our study, 81% of the OSN images with illumination and 77% of the OSN images with low lighting cannot be used to log in the face authentication systems targeting for mobile platform, while these rates increase to 96% for the images with illumination and 96% for the images with low lighting on traditional platform.

3.2.3 Impacts of User Behaviors

The difference in user behavior is another major factor influencing the quality of shared images that decides whether these images can be eventually used for successful OSNFD-based attacks. Our study reveals that the participants who publish more facial images in OSNs are not necessarily more vulnerable than those who publish less facial images in OSNs. In fact, the OSNFD threat is more severe among the participants who publish facial images with higher quality in OSNs.



(a) (b)

Fig. 5. Difference in $VulImage$ and $VulUser$ between females and males

To illustrate the impact of user behaviors, we use the different sharing behaviors and the different OSNFD threat between females and males as example. In our study, female participants are reported to publish facial images in OSNs more frequently than male participants in general. On average, each of the female participants publishes 65 facial images per year while each of the male participants publishes 34 facial images per year. However, the OSNFD threat for the females is less severe than that for the males, as shown in Figure 5.

This can be explained by the lower quality of the OSN images published by the females. We find that the female participants are more likely to publish blurred images, edited images, or images with their makeup. The blur, edit, and makeup can degrade the quality of an image and therefore lead to the difficulty in face recognition [10], [23]. In our study, 12% of the OSN images suffer from these negative effects. Among these low quality images, 61% are published by the females while only 39% of the images are published by the males. All of these blurred, makeup, or edited images fail to pass at least one face authentication system.

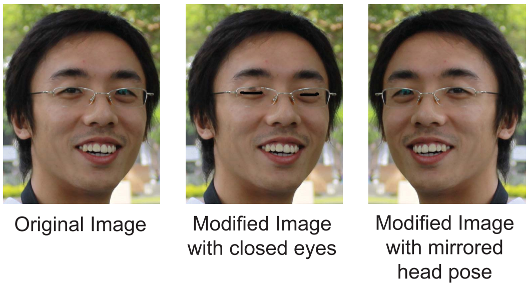
3.2.4 Effectiveness of Liveness Detection

Liveness detection can be used to mitigate OSNFD attacks. The purpose of liveness detection is to distinguish between a real face and a fake face. Among the six tested face authentication systems, liveness detection mechanisms are available on three of them. In particular, Face Unlock and Visidon's liveness detection detects eye blink while Veriface's liveness detection detects head rotation[1]. We examine

[1] At the moment of activating liveness detection, Face Unlock and Visidon show users messages requiring the users to blink their eyes while Veriface shows users messages requiring the users to rotate their heads.

the effectiveness of these liveness detection mechanisms against OSNFD attacks.

We randomly choose 20 subjects (including 10 females and 10 males) and their OSN images to test the available liveness detection mechanisms. In order to imitate the required facial motions including eye blink and head rotation, we use Photoshop software to modify these images as shown in Figure 6. In particular, for imitating eye blink, we firstly replace the eyes in each original image with two black lines to imitate the closed eyes. Then the original image with open eyes and the modified image with closed eyes are displayed on an LCD screen alternatively and quickly so as to imitate eye blinks [39]. To imitate head rotation, we firstly flip the face in each original image containing horizontal head rotation ranging between 10° and 20° and generate a face image with mirrored head pose. Then the original image and the modified image with a mirrored head pose are displayed on an LCD screen alternatively and quickly so as to imitate the head rotation as required by Veriface and other head rotation based liveness detection.

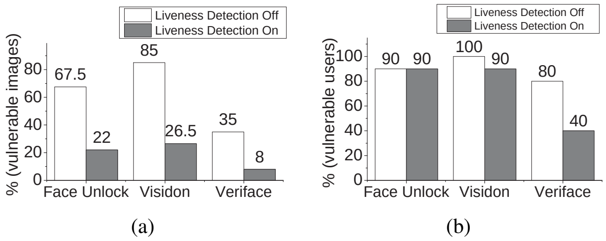

Original Image

Modified Image with closed eyes

Modified Image with mirrored head pose

Fig. 6. Modified sample facial images

Figure 7 shows that the liveness detection mechanisms which we test can still be bypassed due to OSNFD attacks. When the liveness detection mechanisms are turned off, on average, 62.5% of images and 90% of users are vulnerable. After the liveness detection mechanisms are turned on, 18.8% of images and 73.3% of users are vulnerable on average.



(a)

(b)

Fig. 7. Percentage of $VulImage$ and $VulUser$ for three face authentication systems where liveness detection is turned on/off

The detection of eye blink and head rotation relies on accurate detection of facial landmarks, including eyes, mouth, nose, and ears, which requires high-quality login facial images with high resolutions and good lighting conditions [34], [50]. This reduces significantly the percentage of vulnerable images when the liveness detection mechanisms are turned on. However, since most images are high quality, the decrease in the percentage of vulnerable users is moderate, which is still high even if the liveness detection mechanisms are turned on.

In our user study, the eye blink based liveness detection on Face Unlock and Visidon is more vulnerable to OSNFD attacks than the head rotation based liveness detection on Veriface. With liveness detection, on average, 24.3% of images and 90% of users are vulnerable for Face Unlock and Visidon while 8% of images and 40% of users are vulnerable for Veriface. Compared to the eye blink based liveness detection, the head rotation based liveness detection put more restrictions on the quality of login facial images because the detection of head rotation usually requires an accurate localization of multiple facial landmarks including eyes, nose, mouth, and ears. It is thus more difficult for attackers to discover and manipulate appropriate facial images to bypass the liveness detection on Veriface as compared to Face Unlock and Visidon. We also examine Visidon and VeriFace in high security level with liveness detection on. In particular, the percentages of the vulnerable images for Visidon and VeriFace decrease to 15% and 2%, respectively, while the percentages of the vulnerable users for Visidon and VeriFace decrease to 50% and 25%, respectively.

We further evaluate the accessibility of the face authentication systems with liveness detection activated by a follow-up study. 20 participants are asked to login to the three face authentication systems with liveness detection on. The results of our study show that the false rejection rates on the systems with their liveness detection on are higher than those with their liveness detection off, especially for the outdoor environments where mobile devices and laptops are usually used. The highest false rejection rate observed is 95% for Veriface. This is because the systems with their liveness detection on impose more restrictions on the quality of login images and the environments of login process. The activation of liveness detection leads to lower accessibility to end users.

TABLE 4
The false rejection rates of three face authentication systems with liveness detection on.

System	Security level	Room+ normal lighting	Room+ dim lighting	Outdoor ground	Shelter
Face Unlock	N/A	0%	5%	20%	0%
Visidon	Low	0%	5%	15%	0%
	High	5%	65%	80%	60%
Veriface	Low	0%	40%	55%	40%
	High	20%	80%	95%	75%

4 Risk Estimation

Although the OSNFD threat is significant as shown in the previous section, we observe the effectiveness of OSNFD-based attacks may be significantly reduced by manipulating certain attributes of facial images. In this section, we examine these key attributes via statistical analysis in terms of the three major factors including security settings, target platforms, and user behaviors which affect the effectiveness of OSNFD-based attacks. These key attributes are used to develop an estimation tool for end users to calculate the risk of their shared images.

4.1 Key Attributes

From the theoretical perspective, there are still many challenges for face recognition algorithms. These challenges also become key attributes that limit the effectiveness of spoofing attacks. The common attributes addressed in the prior study [1] include head pose, lighting condition, facial expression, facial occlusion, and image resolution. Beside these traditional attributes, we also observe blur, facial makeup, and editing (using Photoshop-like software) as the extra key attributes which often appear in the real world images shared in OSNs, though they are usually not considered in the controlled settings of traditional study on face authentication. We describe the details of these key attributes as follows.

Head pose is a prominent challenge to face recognition. The performance of face recognition algorithms in face authentication can be significantly affected if the head pose in a login image and the head pose in the pre-stored facial image are different [50]. The affecting variations of a head pose mainly include two out-of-plane rotations, namely horizontal rotation and vertical rotation [34].

Lighting condition is another prominent challenge in the realm of face recognition. The variation of lighting conditions mainly includes illumination and low lighting [14], [26], [50]. The illumination is mainly caused when direct light shoots on the 3D structure of a face and strong shadows can be casted which diminish facial features [14], [50]. The illumination can be classified into side illumination and top/bottom illumination [14]. Low lighting is another negative lighting condition, which usually happens when a facial image is taken in dim environment or with extreme bright background. The low lighting may diminish facial features since the luminance in face region is too low for face recognition algorithms to recognize [26].

Facial expression such as smile, surprise, etc, can change face geometry and therefore affect the performance of face recognition algorithms [1]. The common facial expressions include neutral expression, smile without showing teeth, smile showing teeth, closed eyes, open mouth, and other expressions.

Facial occlusion often happens in real world due to additional accessories on face, such as sunglasses, scarf, hands on face, etc. The occlusion can result in the failure of face appearance representation or imprecise facial feature searching and localization, and therefore have negative influence on the performance of face recognition algorithms. The common facial occlusions include forehead occlusion, eyebrow occlusion, eye occlusion, cheek occlusion, and mouth occlusion [1].

The **resolution** of an image can affect accuracy of facial landmark localization and therefore influence the performance of face recognition algorithms. As the resolution of face images decreases, the performance of the face recognition algorithms drops [50].

The **blur** in a facial image causes difficulty in accurate localization of edges of facial region and facial landmarks (i.e. eyes, nose, mouth, etc) by face recognition algorithms and therefore harms the performance of the algorithms.

Facial **makeup** can substantially change the appearance of a face and facial landmarks, such as the alternations of perceived facial shape, nose shape, location of eyebrows, etc. These alternations by the facial makeup, especially by non-permanent facial makeup, challenge face recognition significantly [10].

The **editing** of an image introduces noise pixels and changes the appearance of the face in the image [2], [10]. Face recognition algorithms can be affected by these noises and appearance changes due to the edited image.

The labeling of the collected OSN dataset is performed by three researchers with the help of some automatic tools. For each OSN image, the face region in the image is firstly extracted using popular face detection software Picasa [17]. Then the resolution of the face region, which is a positive number, is automatically calculated by a program we developed, named ResolutionCalculator. The head poses in the image, which vary between 0 and 90, are estimated with typical head pose estimation algorithms including POSIT and LGBP [34]. The estimated head poses are also manually validated by comparing the OSN image and the participant's images with controlled head poses which are captured in our user study. The rest of the attributes in the image are manually labeled with binary values "yes/no" (i.e. 1 or 0) by three researchers in a majority-vote manner. In particular, we manually label the attributes related to lighting conditions according to the shadow and histogram of face region. The attributes related to facial expressions, facial occlusions, blur, makeup, and edit are manually labeled by comparing the OSN image with the images captured earlier in our user study.

All these attributes significantly degrade the image quality and therefore lead to the failure of OSNFD-based attacks. They are used as input parameters to build our risk estimation tool in the next section.

4.2 Risk Estimation Model

We use binomial logistic regression [22] to model the impact of the key attributes introduced in the previous subsection. The notions of these attributes are defined in Table 5. Then the key attributes of each image can be represented by an input parameter vector, denoted as $V = (rot_H, rot_V, ill_{sd}, ill_{tb}, dm, bg, FEx_n, FEx_s, FEx_{st}, FEx_{ce}, FEx_m, FEx_{ot}, Occ_{fh}, Occ_{eb}, Occ_{eye}, Occ_{chk}, Occ_{mh}, res, blur, mk, ed)$. For the output, we assign an OSN image to either a positive class or a negative class. The positive class means the image can be used to pass the login of a specific face authentication system, otherwise the image will be in the negative class.

TABLE 5
Parameters related to the key attributes

Attribute	Parameter	Notation
Head pose	Horizontal rotation	rot_H
	Vertical rotation	rot_V
Lighting condition	Side illumination	ill_{sd}
	Top/bottom illumination	ill_{tb}
	Dimness	dm
	Bright background	bg
Facial expression	Neutral	FEx_n
	Smile without showing teeth	FEx_s
	Smile showing teeth	FEx_{st}
	Closed eyes	FEx_{ce}
	Open mouth	FEx_m
	Other expressions	FEx_{ot}
Facial occlusion	Occluded forehead	Occ_{fh}
	Occluded eyebrow	Occ_{eb}
	Occluded eye	Occ_{eye}
	Occluded cheek	Occ_{chk}
	Occluded mouth	Occ_{mh}
Resolution	Resolution	res
Blur	Blur	$blur$
Facial makeup	Makeup	mk
Edit	Edit	ed

Binomial logistic regression is a classic probabilistic classification model [22], which accepts multiple predictor variables as inputs, and predicts the outcome for a dependent variable which has only two possible types, such as "positive" vs "negative". Thus it is a proper tool to calculate the probability of an image assigned to the positive class based on the key attributes extracted from an OSN image. Given a parameter vector V_i of a facial image i and a face authentication system in a security level, the regression function is

$$\ln\left(pr_i/(1 - Pr_i)\right) = \beta_0 + \beta_1 v_1 + \cdots + \beta_m v_m \qquad (1)$$

where pr_i is the probability that an image i is assigned to the positive class, v is a parameter in V_i, and β is a regression coefficient. The risk score of the facial image i is the value of pr_i. The facial image i is assigned to the positive class if $pr_i \geq 0.5$. Otherwise, i is assigned to the negative class. The correctness of these assignments is verified with the ground truth data collected from the previous empirical analysis.

For each combination of face authentication system and its security level, we examine the model fitting of binomial logistic regression and the significance of the parameters by using the real world OSN images and run binomial logistic regression on SAS software [42]. The detailed statistical test results are reported in Appendix A. The likelihood ratio test and wald statistic [22] for all the face authentication systems are smaller than 0.0001.

Our statistical analysis shows the most influential attributes are resolution res (*p-value* $p < 0.0001$), occluded eye Occ_{eye} ($p = 0.0255$), occluded mouth Occ_{mh} ($p = 0.0223$), makeup mk ($p = 0.0094$), blur $blur$ ($p = 0.0283$), dimness dm ($p = 0.0413$), and illumination ill_{sd} ($p = 0.0469$). Resolution res has positive impact on the risk of OSNFD. It is because higher resolution contributes to more accurate facial landmark localization and results in better performance of face recognition and increases the risk of OSNFD. The occluded eye Occ_{eye}, occluded mouth Occ_{mh}, makeup mk, blur $blur$, dimness dm, and illumination ill_{sd} have negative impact and lower the risk of OSNFD. In particular, the occluded eye and occluded mouth leads to decrease in the performance of face recognition algorithms, as accurate localization of eyes and mouth is important for the alignment process in all major face recognition algorithms [1]. Makeup can significantly change the appearance of the face and the facial landmarks and therefore lowers the performance of face recognition. The blur, dimness, and illumination are prominent attributes which cause difficulty in face recognition since it diminishes facial features and leads to difficulty in localization of these facial features.

The parameters related to other attributes, including head pose and facial expression, are generally not statistically significant. Among the collected OSN images, the variations of head pose and facial expression are limited since users are usually cooperative when these images are captured and tend to publish the images from which they are easily recognized. As observed in our study, the head poses in most OSN images are within the acceptable head pose ranges of the face authentication systems, which causes the insignificance due to lack of samples with extreme head pose. On the other hand, facial expressions observed in most OSN images are only mild-mannered expressions including neutral expression, smile without showing teeth, smile showing teeth, closed eyes, open mouth. These common expressions do not have significant impact as they have been well handled in current face recognition algorithms [1]. Other extreme facial expressions, such as making faces, do significantly affect the face recognition, but they are observed in only 5% of the OSN images.

In the following subsection, we further analyze the detailed impacts of the key attributes from the three major perspectives including security settings, target platforms, and user behaviors which can affect the effectiveness of OSNFD-based attacks as shown in 3.2.

4.2.1 Detailed Impacts of the Key Attributes

The security settings of the tested face authentication systems can be configured to either low security level or high security level. As shown in Table 9 and Table 10 in Appendix A, for evaluating the risk of OSNFD on face authentication systems at the low security level, the most influential attributes are resolution res ($p < 0.0001$), occluded eye Occ_{eye} ($p = 0.0014$), occluded mouth Occ_{mh} ($p = 0.0079$), makeup mk ($p = 0.0068$), blur $blur$ ($p = 0.0283$), dimness dm ($p = 0.0248$), and illumination ill_{sd} ($p = 0.0469$). For evaluating the risk of OSNFD on face authentication systems at the high security level, the most influential attribute is resolution res ($p < 0.0001$).

From the low security level to the high security level, the authentication systems raise the recognition threshold and impose more rigid restrictions on the quality of login facial images. In our study, for the systems at their low security level, the average resolution of vulnerable images is 92481 pixels. However, for the systems at their high security level, the average resolution of vulnerable images increases to 111552 pixels. The resolution res makes a significant contribution to the risk of OSNFD.

The face authentication systems are classified into mobile platforms and traditional platforms. Our statistical analysis shows that resolution res ($p < 0.0001$), head pose rot_V ($p = 0.0029$), occluded eye Occ_{eye} ($p = 0.0223$), occluded mouth Occ_{mh} ($p = 0.0223$), makeup mk ($p = 0.0094$), blur $blur$ ($p = 0.0012$), and illumination ill_{sd} ($p = 0.0239$) are significant attributes for evaluating the risk of OSNFD on mobile platforms while resolution res ($p < 0.0001$), blur $blur$ ($p = 0.0283$), dimness dm ($p = 0.0413$), and illumination ill_{sd} ($p = 0.0469$) are significant attributes for evaluating the risk of OSNFD on traditional platforms.

Compared to the face authentication systems on mobile platforms, the systems on traditional platforms impose more rigid restrictions on the quality of login face images such as higher image resolution and better lighting conditions. Our results show that the average resolution of vulnerable images for mobile platforms is 89633 pixels while the average resolution of vulnerable images for traditional platforms increases to 110747 pixels. Besides blur $blur$ and illumination ill_{sd}, the dimness dm makes a significant contribution for evaluating the risk of OSNFD on traditional platforms because the existence of dimness leads to a low face recognition rate.

The quality of the facial images published by female participants is generally lower, as it is presented in Section 3.2.3. According to Table 8 in Appendix A, makeup mk ($p = 0.0002$), resolution res ($p < 0.0001$), occluded eye Occ_{eye} ($p = 0.0014$), and head pose rot_V ($p = 0.0008$) are the most influential attributes for evaluating the risk of OSNFD for female participants while resolution res ($p < 0.0001$) is most influential attribute for evaluating the risk of OSNFD for male participants.

Female participants are more likely to publish their facial images with makeup, non-frontal head poses, low resolutions, and occluded facial landmarks. These attributes can degrade the quality of images and lead to a low face recognition rate. Thus, makeup mk, resolution res, occluded eye Occ_{eye}, and head pose rot_V make a significant contribution to the evaluation of the risk of OSNFD for females. Compared to the female participants, the risk of OSNFD for the facial images published by male participants is mainly affected by image resolution res.

4.3 Model Evaluation

To evaluate the performance of the proposed risk estimation tool, we use a subject-disjoint cross-validation method. In each round, for each of the face authentication systems at a specific security level, we randomly choose 80% of the OSN images to train the model and use the risk estimation tool to automatically classify the rest of the images. The selection of the images for training set and evaluation set is subject-disjoint. The above process is repeated by 10 rounds. The performance is measured by standard classification evaluation metrics, including precision, recall, and F1 score [40].

Precision is defined as the percentage of the true positive images among the images assigned to the positive class by the risk estimation tool, which can be calculated by $tp/(tp + fp)$ where tp is the number of true positive images and fp is the number of false positive images. Recall is defined as the percentage of the true positive images detected by the risk estimation tool among the positive images in ground truth, which can be calculated by $tp/(tp+fn)$ where tp is the number of true positive images and fn is the number of false negative images. F1 score considers both the precision and the recall, which can be calculated by $\mathsf{F1} = 2\times\mathsf{precision}\times\mathsf{recall}/(\mathsf{precision}+\mathsf{recall})$.

Table 6 shows the performance evaluation metrics of the risk estimation tool. On average, the risk estimation tool achieves a precision of 81%, a recall of 83%, and an F1 score of 82%. The performance evaluation indicates that the risk estimation tool detects most of the vulnerable images which can lead to successful OSNFD-based attacks if these images are published in OSNs.

TABLE 6
Effectiveness of our risk estimation tool

System	Security level	Precision	Recall	F1 score
Face Unlock	N/A	73%	77%	75%
Facelock Pro	Low	70%	69%	69%
	High	81%	75%	78%
Visidon	Low	79%	90%	84%
	High	86%	92%	89%
Veriface	Low	79%	68%	73%
	High	90%	98%	94%
Luxand Blink	Low	84%	87%	85%
	High	87%	90%	88%
FastAccess	Low	77%	67%	72%
	High	89%	95%	92%
Average	N/A	81%	83%	82%

5 Discussion

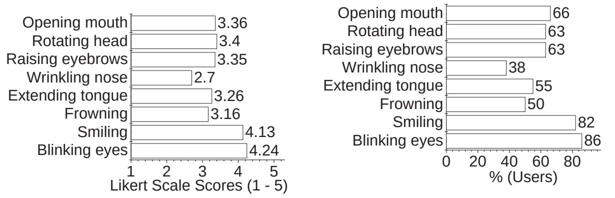
5.1 Costs of Liveness Detection

The liveness detection mechanisms deployed on popular face authentication systems include eye-blinking and head rotation detection. The advantages of such liveness detection include no additional hardware support, moderate quality of input images, and relatively low usability cost, which are important to consumer-level products as they are price-sensitive and accessibility-first.

Various liveness detection mechanisms have been proposed to enhance face authentication, including blinking eyes, rotating head, smiling, frowning, extending tongue, wrinkling nose, raising eyebrows, and opening mouth [18]. To evaluate the usability of such liveness detection mechanisms, we conduct an online survey among all 74 participants in our user study. The online survey includes a number of questions on a 5-points Likert scale to collect the participants' preference and perception on these mechanisms related to the facial motions and facial expressions.

In the Likert Scale, 5 means most comfortable, while 1 means most uncomfortable from a user's point of view.

Figure 8(a) shows that most users prefer blinking eyes and smiling in liveness detection while wrinkling noses, frowning, and extending tongue have a lower usability. On the other hand, the facial motions and facial expressions with higher Likert scores can be easily manipulated or they are more likely to appear in OSN images. Our study shows that blinking eyes can be easily manipulated by modifying facial images and smiling is observed in 65.7% of OSN images. Therefore, the liveness detection based on them is more vulnerable to the OSNFD attacks although more than 80% of users like to use them, as shown in Figure 8(b). In contrast, the less preferable facial expressions, including wrinkling noses, frowning, and extending tongue, are observed in only 4.7% of OSN images. The liveness detection based on these facial expressions can mitigate OSNFD attacks more effectively, because it is more difficult for adversaries to obtain suitable OSN images which contain different facial expressions and similar conditions so that they can be stitched together to simulate facial motions. Unfortunately, only 47.7% of the users would like to conduct such facial expressions in our user study.



(a) 5-point Likert scale scores, ranging from 1 (most uncomfortable) to 5 (most comfortable), for the facial motions and facial expressions

(b) Percentage of users choosing to activate liveness detection based on the facial motions and facial expressions

Fig. 8. Usability of facial motions and facial expressions commonly used in liveness detection

TABLE 7
Costs associated with existing liveness detection for face authentication, where * indicates that a significant cost is incurred for end users or device manufacturers.

Types of liveness detection	Examples of liveness detection methods	Image quality	Additional hardware	Usability cost
Real-time response	Eye blinking [15], [46], head rotation [30], facial expression	Low/Middle	No	Middle/High*
Motion of 3D face	Optical flow from holistic face [5], optical flow lines [25]	High*	No	Low
Texture pattern	LBP based texture analysis [7], [33]	High*	No	Low
Multimodal	Face and fingerprint or iris [13]	Middle/High*	Yes*	Middle/High*

The liveness detection methods based on facial motions and facial expressions require real-time response during face authentication. Besides the liveness detection methods in this category, several sophisticated liveness detection techniques, including texture pattern, motion of 3D face, and multimodal techniques, have been proposed for face authentication [13], [24]. However, all of them are associated with considerable costs as shown in Table 7 [24], [37]. Their costs include requiring additional hardware, high-quality images, ideal environments, and high user collaborations. They may not be suitable for existing consumer-level face authentication systems and need to be further improved.

However, more powerful front-facing cameras with higher speed and resolution and various sensors are emerging in mobile devices, which open up new possibilities for reliable and usable liveness detection including both hardware-based liveness detection and multi-biometrics.

5.2 Implications of Our Findings

Face authentication does provide an attractive alternative of user authentication for its non-intrusive and zero-memory procedure. However, the appearance of OSNFD brings a significant threat to question the practicality of face authentication as a usable authentication factor. Nowadays, a huge amount of personal facial images/videos have been published in OSNs that can be accessible to potential adversaries without the previously required physical proximity. Therefore, face biometrics can now be disclosed in large scale and acquired by adversaries remotely. Face biometrics are no longer secrets only owned by the users and can be disclosed to anyone who has access to victim's personal images shared in OSNs.

Raising the security level of face authentication systems could mitigate the OSNFD threat by scarifying the accessibility, which leads to the inconvenience for legitimate users. Liveness detection is another major countermeasure to mitigate the spoofing attack against the face authentication systems. Unfortunately, existing liveness detection techniques available on consumer-level computing devices can be easily circumvented by one or two facial images as presented in Section 3.2.4. More reliable liveness detection like multi-modal mechanisms usually relies on using additional authentication factor (e.g. another biometrics such as voice and fingerprint). This introduces another liveness detection problem for the additional authentication factor, which may not be reliable. For example, voice and fingerprint can also be spoofed. Even worse, more serious privacy concerns will rise if a system requires to collect many biometrics information from a user [49], which may eventually cause the rejection of the liveness detection mechanism.

The current face authentication systems are not strong enough to thwart OSNFD attacks. The existing liveness detection techniques are either too weak to defend against OSNFD attacks or not suitable to be deployed on consumer-level devices. More reliable and usable liveness detection is needed to mitigate the threats.

5.3 Limitations

Ecological validity is a challenge to any user study. Like most prior research [19], [38], our study only recruits students in university. These participants are more active in using consumer-level computing devices and sharing images in OSNs. Thus the evaluation of the OSNFD may vary with other populations.

In the user study design, it is still a challenge to collect facial images with precisely controlled head poses [34]. Like the prior head pose data sets [20], [29], the accuracy of the head poses in our data set may be affected by the poor ability of the participant to accurately direct his/her head, the unconscious movement of human beings and limit of resources. In another experiment of examining the false rejection rates of the face authentication systems, we choose 4 locations to mimic different login environments in daily life. Since it is impossible for all the participants to do the tests at the same time and at the same physical positions, the background of image inputs captured by the camera may change.

Another challenge in our study is to accurately estimate parameters [27] such as head pose, illumination, and makeup in our collected OSN dataset. Since the accuracy of automatic labeling tools is limited [1], we manually label the OSN images with the help of automatic tools.

It is also possible to further improve our risk estimation tool. To our best knowledge, our work is the first attempt to semi-automatically detect the vulnerable images that can be used to attack face authentication. Our current risk estimation tool can serve as a baseline for future improvement by refining the key parameters and the statistical model. It is also valuable to incorporate automatic high accuracy labeling for those hard-to-label attributes like illumination and facial makeup, once the ongoing research [10], [14], [34] resolves these challenges.

6 Conclusion

In this paper, we investigated the threat of OSN-based facial disclosure (OSNFD) against some real-world face authentication systems. Our results show that these face authentication systems are vulnerable to OSNFD attacks. We analyzed the vulnerability of typical face authentication systems against OSNFD attacks in terms of security settings, target platforms and user behavior. We further develop a risk estimation tool to help users evaluate the risks of publishing their personal images in OSNs. We also examined the existing liveness detection methods in the presence of OSNFD attacks and showed that the effectiveness and usability of the existing liveness detection are not sufficient and need to be improved.

7 Acknowledgment

The work is supported by the Singapore National Research Foundation under NCR Award Number NRF2014NCR-NCR001-012. The study of Yingjiu Li on this subject is also funded through a research grant C220/MSS13C004 from MOE's AcRF Tier 1 funding support through Singapore Management University.

References

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, 2007.

[2] M. Abdel-Mottaleb and M. H. Mahoor. Assessment of blurring and facial expression effects on facial image recognition. In *Advances in Biometrics*, pages 12–18, 2005.

[3] Z. Akhtar, C. Micheloni, C. Piciarelli, and G. L. Foresti. Mobio_livdet: Mobile biometric liveness detection. In *AVSS*, pages 187–192, 2014.

[4] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *IJCB*, pages 1–7, 2011.

[5] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In *IASP 2009*, pages 233–236. IEEE, 2009.

[6] B. Biggio, Z. Akhtar, G. Fumera, G. Marcialis, and F. Roli. Security evaluation of biometric authentication systems under real spoofing attacks. *Biometrics, IET*, 1:11–24, 2012.

[7] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, et al. Competition on counter measures to 2d facial spoofing attacks. In *IJCB*, pages 1–6. IEEE, 2011.

[8] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *BIOSIG*, pages 1–7, 2012.

[9] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, et al. The 2nd competition on counter measures to 2d face spoofing attacks. In *ICB 2013*, pages 1–6. IEEE, 2013.

[10] A. Dantcheva, C. Chen, and A. Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *BTAS*, pages 391–398, 2012.

[11] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay. Moving face spoofing detection via 3d projective invariants. In *ICB*, pages 73–78, 2012.

[12] Facelock.mobi. http://www.facelock.mobi/facelock-for-apps.

[13] J. Galbally, S. Marcel, and J. Fierrez. Biometric anti-spoofing methods: A survey in face recognition. *Access*, 2:1530–1552, 2014.

[14] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *TPAMI*, 23(6):643–660, 2001.

[15] Google. http://www.android.com/about/ice-cream-sandwich/.

[16] Google. https://play.google.com/store/apps?hl=en.

[17] Google. https://picasa.google.com/.

[18] Google. http://www.google.com/patents/US8457367.

[19] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *WPES*, pages 71–80, 2007.

[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *IVC*, 28(5):807–813, 2010.

[21] Harvard. http://bionumbers.hms.harvard.edu//bionumber.aspx ?id=100706&ver=1.

[22] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant. *Applied logistic regression*. Wiley. com, 2013.

[23] F. Hua, P. Johnson, N. Sazonova, P. Lopez-Meyer, and S. Schuckers. Impact of out-of-focus blur on face recognition performance based on modular transfer function. In *ICB*, pages 85–90, 2012.

[24] O. Kahm and N. Damer. 2d face liveness detection: An overview. In *BIOSIG*, pages 1–12, 2012.

[25] K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. *IVC*, 27(3):233–244, 2009.

[26] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognitionła review. *CVIU*, 97(1):103–135, 2005.

[27] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, pages 365–372, 2009.

[28] I. Lab. https://www.idiap.ch/dataset/replayattack.

[29] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *TPAMI*, 27(5):684–698, 2005.

[30] Lenovo. http://en.wikipedia.org/wiki/VeriFace.

[31] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng. Understanding osn-based facial disclosure against face authentication systems. In *ASIACCS*, pages 413–424, 2014.

[32] Luxand. http://www.luxand.com/.

[33] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *IJCB 2011*, pages 1–7. IEEE, 2011.

TABLE 9
The statistical test results for the risk of OSNFD on different face authentication systems at low security level.

Attributes		Face Unlock	Facelock Pro	Visidon	Veriface	Luxand Blink	FastAccess
rot_H	*P value*	0.8207	0.6517	0.6068	0.441	0.1371	0.9559
	Coefficient	-0.00098	-0.00172	-0.00201	-0.00407	-0.0128	-0.00027
rot_V	*P value*	0.0005★	0.0002★	0.7292	0.8102	0.1242	0.0011★
	Coefficient	-0.0313	-0.0299	-0.00283	-0.00246	-0.0246	-0.0315
ill_{sd}	*P value*	<.0001★	0.0239★	0.0206★	0.0469★	0.0414★	0.034★
	Coefficient	-1.5931	-0.7479	-0.6632	-0.7501	-2.2443	-0.6005
ill_{tb}	*P value*	0.2152	0.5554	0.337	0.6778	0.746	0.7218
	Coefficient	-0.5369	-0.2072	-0.2962	-0.2357	-0.4123	-0.195
dm	*P value*	0.2765	0.0248★	0.0035★	0.0004★	0.0049★	0.0228★
	Coefficient	-0.2414	-0.4608	-0.5629	-1.2419	-2.2378	-0.6126
bg	*P value*	0.5851	0.6232	0.1426	0.2427	0.2585	0.5456
	Coefficient	-0.2138	-0.1686	-0.4182	-0.638	-1.3363	-0.3025
FEx_n	*P value*	0.4701	0.5031	0.9539	0.4619	0.6762	0.4608
	Coefficient	1.1708	12.1979	1.7251	1.7988	1.8263	2.4745
FEx_s	*P value*	0.6091	0.6978	0.9533	0.8321	0.9083	0.7664
	Coefficient	1.3664	0.6781	12.3712	0.4969	-0.496	0.7354
FEx_{st}	*P value*	0.6674	0.9167	0.9547	0.8156	0.697	0.996
	Coefficient	1.1483	0.1826	12.002	-0.5469	-1.6799	0.0125
FEx_{ce}	*P value*	0.9982	0.7453	0.9623	0.9883	0.9932	0.9067
	Coefficient	0.00669	-0.7071	9.9899	-13.4194	-13.7513	0.3331
FEx_m	*P value*	0.9982	0.7453	0.9623	0.9883	0.9932	0.9067
	Coefficient	0.00669	-0.7071	9.9899	-13.4194	-13.7513	0.3331
FEx_{ot}	*P value*	0.6135	0.6329	0.9608	0.7905	0.7797	0.7205
	Coefficient	-1.3496	-0.8336	10.3791	-0.6231	-1.2065	-0.886
Occ_{fh}	*P value*	0.2785	0.3658	0.0801	0.0137★	0.1161	0.0885
	Coefficient	-1.2366	0.2072	-0.4292	-0.7965	-0.8285	-0.6359
Occ_{eb}	*P value*	0.6332	0.17	0.2477	0.6318	0.0579	0.2146
	Coefficient	-0.1374	-0.3652	-0.3076	-0.1757	-1.3736	-0.4163
Occ_{eye}	*P value*	<.0001★	0.0014★	<.0001★	0.0835	0.1596	0.0009★
	Coefficient	-2.3839	-0.8268	-1.2979	-0.668	-1.5658	-1.4977
Occ_{chk}	*P value*	0.0853	0.0124★	0.0355★	0.1415	0.5469	0.0909
	Coefficient	-0.5953	-0.7223	-0.5271	-0.5805	0.378	-0.6681
Occ_{mh}	*P value*	0.0005★	0.0001★	0.0007★	0.0079★	0.9721	0.0046★
	Coefficient	-1.8545	-2.1559	-1.2322	-2.1112	-15.1083	-1.9149
res	*P value*	<.0001★	<.0001★	<.0001★	<.0001★	<.0001★	<.0001★
	Coefficient	1.9227	2.1419	3.2119	4.7961	8.2094	4.474
$blur$	*P value*	0.5405	<.0001★	0.0003★	0.0283★	0.0078★	0.0072★
	Coefficient	-0.2191	-1.6006	-1.0922	-1.1189	-2.4965	-1.3147
mk	*P value*	<.0001★	<.0001★	<.0001★	0.0068★	0.9713	0.0001★
	Coefficient	-3.5659	-1.6768	-2.0062	-1.3347	-13.4519	-2.7998
ed	*P value*	0.4278	0.3935	0.2544	0.7024	0.0483★	0.8062
	Coefficient	-0.3167	-0.3378	-0.4549	-0.215	-1.9747	-0.1177

TABLE 10
The statistical test results for the risk of OSNFD on different face authentication systems at high security level.

Attributes		Facelock Pro	Visidon	Veriface	Luxand Blink	FastAccess
rot_H	*P value*	0.3614	0.242	0.0511	0.883	0.082
	Coefficient	-0.00539	-0.013	-0.0296	-0.00147	-0.0213
rot_V	*P value*	0.0029★	0.3448	0.4049	0.5589	0.1232
	Coefficient	-0.0352	-0.0198	-0.0312	-0.0116	-0.0397
ill_{sd}	*P value*	0.0974	0.7467	0.679	0.4299	0.8434
	Coefficient	-1.0475	-0.7183	-4.3193	-1.0409	-0.4704
ill_{tb}	*P value*	0.066	0.3217	0.917	0.0272★	0.6541
	Coefficient	-1.222	-2.2165	-1.0495	-2.8132	-3.7415
dm	*P value*	0.3431	0.0599	0.351	0.9508	0.0413★
	Coefficient	-0.3239	-1.6103	-5.8915	-13.8053	-2.3548
bg	*P value*	0.8877	0.6957	0.9844	0.9672	0.8458
	Coefficient	-0.1001	-5.4439	-0.1927	-11.3403	-0.4639
FEx_n	*P value*	0.657	0.9657	0.9896	0.828	0.9678
	Coefficient	6.5169	8.1925	3.2357	6.7898	-6.7898
FEx_s	*P value*	0.8312	0.9785	0.9953	0.9662	0.9819
	Coefficient	1.188	4.0849	3.7357	0.6321	3.8179
FEx_{st}	*P value*	0.9907	0.9874	0.9983	0.9731	0.992
	Coefficient	-0.0653	2.3982	1.3745	-0.502	1.6839
FEx_{ce}	*P value*	0.9879	0.9879	0.9987	0.994	0.9912
	Coefficient	-12.9673	-2.551	1.0544	-11.515	-1.943
FEx_m	*P value*	0.9879	0.9879	0.9987	0.994	0.9912
	Coefficient	-12.9673	-2.551	1.0544	-11.515	-1.943
FEx_{ot}	*P value*	0.8219	0.9796	0.9955	0.9546	0.9877
	Coefficient	-1.2547	-3.8644	3.5507	-0.8491	2.5989
Occ_{fh}	*P value*	0.089	0.1873	0.0003★	0.0919	0.0706
	Coefficient	1.222	-4.0753	-5.6302	-2.28	-4.1719
Occ_{eb}	*P value*	0.1085	0.1116	0.0262★	0.7679	0.6558
	Coefficient	-0.8694	-1.7841	-3.6723	-0.2618	-0.6768
Occ_{eye}	*P value*	0.0223★	0.578	0.5021	0.0255★	0.585
	Coefficient	-1.6262	-6.2644	-4.1741	-1.4811	-3.6529
Occ_{chk}	*P value*	0.0611	0.3064	0.2473	0.4126	0.2548
	Coefficient	-1.1122	-1.3555	-7.6076	-0.8082	-1.7596
Occ_{mh}	*P value*	0.0223★	0.1787	0.4788	0.9708	0.524
	Coefficient	-2.5118	-1.8801	-8.1448	-14.8264	-7.7025
res	*P value*	<.0001★	<.0001★	<.0001★	<.0001★	<.0001★
	Coefficient	4.1225	10.8057	18.9514	8.512	12.0517
$blur$	*P value*	0.0012★	0.6068	0.3789	0.9698	0.4244
	Coefficient	-2.6587	-9.975	-9.4925	-15.2418	-8.9533
mk	*P value*	0.0094★	0.697	0.7208	0.9717	0.6202
	Coefficient	-2.6839	-6.6644	-3.3963	-12.5103	-5.1342
ed	*P value*	0.0461★	0.2398	0.5003	0.0287★	0.5336
	Coefficient	-1.5032	-1.5236	-10.3199	-2.9023	-9.4999

TABLE 8
The significant attributes for the risk of OSNFD among females/males.

Attributes		**Female**	**Male**
rot_V	*P value*	0.0008★	0.3212
	Coefficient	-0.0393	-0.0256
Occ_{eye}	*P value*	0.0014★	0.9952
	Coefficient	-0.8417	-0.209
res	*P value*	<.0001★	<.0001★
	Coefficient	1.9447	9.2784
mk	*P value*	0.0002★	0.7873
	Coefficient	-0.9086	-7.0665

[34] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *TPAMI*, 31(4):607–626, 2009.

[35] E.-S. Ng and A.-S. Chia. Face verification using temporal affective cues. In *ICPR*, pages 1249–1252, 2012.

[36] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[37] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *ICCV*, pages 1–8, 2007.

[38] K. Patel, H. Han, A. K. Jain, and G. Ott. Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks. In *ICB*, pages 98–105, 2015.

[39] J. Rice. http://www.androidpolice.com/2012/08/03/android-jelly-beans-face-unlock-liveness-check-circumvented-with-simple-photo-editing/.

[40] C. J. V. Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 1979.

[41] S. Saravanakumar et al. Removal of moiré pattern noise in images using median and gaussian filter. *IJSETR*, 2(2):pp–380, 2013.

[42] SAS. http://www.sas.com/.

[43] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. Ho. Detection of face spoofing using visual dynamics. *TIFS*, 10(4):762–777, 2015.

[44] S. Trewin, C. Swart, L. Koved, J. Martino, K. Singh, and S. Ben-David. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In *ACSAC*, pages 159–168, 2012.

[45] VagueWare.com. http://www.vagueware.com/top-globally-popular-face-recognition-software/.

[46] Visidon. http://www.visidon.fi/en/Home.

[47] S. Vision. http://www.sensiblevision.com/en-us/home.aspx.

[48] K. Wagner. http://mashable.com/2013/09/16/facebook-photo-uploads/.

[49] J. D. Woodward. Biometrics: Privacy's foe or privacy's friend? *Proceedings of the IEEE*, 85(9):1480–1492, 1997.

[50] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *CSUR*, 35(4):399–458, 2003.

APPENDIX A
STATISTICAL TEST RESULTS

In this section, we provide the detailed results of statistical tests. The results include *P values* and estimated *coefficients* of the key attributes. The statistically significant results are marked with ★.

Yan Li is currently a research fellow at Singapore Management University.

Yingjiu Li is currently an Associate Professor in the School of Information Systems at Singapore Management University.

Ke Xu is currently a Ph.D. student in Information Systems at Singapore Management University.

Qiang Yan is currently a privacy engineer in Google.

Robert H. Deng is currently a Professor in the School of Information Systems at Singapore Management University.