Nguyen, Ngu; Sigg, Stephan

User Authentication based on Personal Image Experiences

Published in:
2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018

DOI:
10.1109/PERCOMW.2018.8480087

Published: 02/10/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Nguyen, N., & Sigg, S. (2018). User Authentication based on Personal Image Experiences. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018* (pp. 336-341). Article 8480087 IEEE. https://doi.org/10.1109/PERCOMW.2018.8480087

User Authentication based on Personal Image Experiences

Ngu Nguyen
Aalto University
Espoo, Finland
le.ngu.nguyen@aalto.fi

Stephan Sigg
Aalto University
Espoo, Finland
stephan.sigg@aalto.fi

***Abstract*—Building upon the concept of collective computing [1], which combines cloud, crowd and shroud technologies, we propose a further application domain for the fourth generation of computing: Usable Security. Combining the three constituent technologies enables novel, stronger and personalized authentication mechanisms. In particular, we combine implicit memory of people (the crowd), obtained from wearable camera devices (the shroud) and supported by edge and cloud facility (the cloud) in order to generate image-based authentication challenges which are transient and personalized .**

I. Introduction

In the seminal 2016 paper [1], Gregory P. Abowd introduced the notion of collective computing. In essence, integrating cloud technology, the human crowd as well as wearable and IoT devices has the potential of creating improved, real-time services that define the fourth generation of computing. Personal navigation is an early example of the tremendous potential that originates from such integration. Further examples are in the health domain, education and commerce (cf. [1]) We argue that usable security is another domain that strongly benefits from collective computing.

Traditional means of authentication do not well combine with the convenience required for todays myriad of IoT and mobile devices. For instance, popular pattern-based authentication schemes are insecure and easily circumvented by shoulder-surfing or smudge-attacks [2]. Biometric information, another common means of authentication, is easily stolen without notice. Personal experience, on the other hand, is a unique piece of information that is continuously refreshed and difficult to access for an adversary. For instance, consider a group of people jointly participating in activities such as walking some route or visiting the same physical location. Those people, even if they share their physical location, experience their surrounding differently and pay attention to different stimuli. We propose an experience-based authentication scheme exploiting the cloud, the crowd and the shroud.

In particular, as depicted in Figure 1, we exploit experiences in image sequences captured by wearable cameras (the shroud) to form always-fresh authentication challenges with respect to individual experiences (the crowd). These image sequences are processed in the network edge or cloud and anonymously analyzed together with sequences obtained from other individuals (the crowd and the cloud) in order to generate login-challenges that exploit the unique (i.e. non-redundant with other people's experiences) experience of an individual. The authentication challenges generated in such a way that they are always fresh, easy to solve for an individual because it reflects her/his unique perceived experiences and therefore challenging to be broken by an adversary that has no access to the unique-experienced patterns.

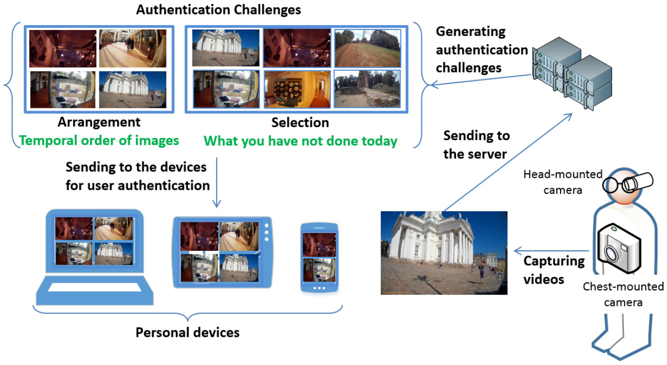

Fig. 1. Collective computing architecture for our image-based user authentication approach

In the rest of the paper, we first review related work on representative authentication mechanisms in Section II. Then, Section III explains techniques used to develop our image-based authentication schemes and Section IV analyses security issues. Next, we describe the experiments and investigate the outcomes in Section V. Finally, the paper concludes with a summary of results and future work.

II. Related Work

Body-mounted camcorders continuously capture first-person-view high-quality videos from the user's perspective [1]. They can be equipped with wireless communication to enable data processing on the cloud and have been integrated in applications increasingly recently. Castro *et al.* [5] extracted an individual's behavioral routines and predicted daily activities from a long-term egocentric photo dataset (the shroud). They applied an ensemble method combining a Convolutional Neural Network and a Random Decision Forest, which could be executed on a remote server (the cloud). Beyond individual behavior, egocentric videos reveal important cues on social

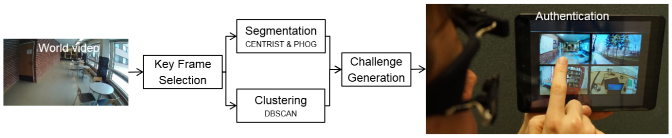

Fig. 2. User authentication based on egocentric videos

interactions. Bambach *et al.* [3] detected and distinguished hands in first-person-view videos (the shroud) to recognize complex interacting activities. Moreover, first-person vision helps us to study which visual parts of the physical world are meaningful and shared across multiple people (the crowd). A clustering algorithm was developed by Yonetani *et al.* [21] to identify these shared visual experiences. The popularity of on-body cameras encourages us to propose a novel authentication mechanism based on autobiographical photo sequences.

Autobiographical authentication is developed on the intersection of users' memories and information recorded with personal on-body devices. Das *et al.* [6] proved its effectiveness through two online questionnaires and one field study on mobile phone usage data. As an authentication scheme based on existing memory, Sun *et al.* [17] proposed to generate passwords from icons of selective installed applications on a user's smartphone (the shroud). To be authenticated, a user is required to discriminate between valid and decoys icons, where the latter are chosen from an app market (the crowd) and personalized for each user. While the previous schemes relied only on data in smartphones, our system provides a two-factor authentication mechanism comprised of an on-body camera and user context (i.e. chronological egocentric images), which can be leveraged to log-in multiple personal devices.

III. User Authentication based on Egocentric Videos

We propose and analyze two schemes to generate graphical passwords in collective computing (*experience-arrangement* and *experience-selection*) and six formats to visualize the authentication challenges.

A. Image-based Password Generation

Our password generation process is shown in Figure 2. First-person-view videos are captured by a wearable camera mounted to an eyeglass frame or the chest (clipped to the pocket) of a user. They are then processed to select key frames which contain memorable events, such as object interaction or remarkable scenes. The remaining images are organized into distinctive events using segmentation and clustering techniques. We introduce two authentication schemes: *experience-arrangement* and *experience-selection*.

Keyframe selection: To select video frames that contain clear information, i.e. containing details that help to recall the associated timeline. We extract key points using the method of Rosten and Drummond [16], which identifies blurred and cluttered images. Formally, suppose that the video consists of k frames and each frame v_i has n_i key points, we select v_i if $n_i \geqslant mean(\{n_i | i \in \mathbb{N}, 1 \leqslant i \leqslant k\})$. For ease of processing on shroud-devices, we reduce the number of images extracted directly from the recorded videos to a fixed ratio (e.g. five images per second).

Segmentation: The $k' \leqslant k$ selected images are then segmented into scenes, based on similarity of visual features. If the difference between two frames is below a certain threshold, they belong to the same segments. Thus, a *segment* describes an identifiable context (i.e. location or activity) describing the experience of a subject. We define the similarity threshold as the mean value of the Euclidean distance between two consecutive (representative) frames in the feature space. Suppose f is the function that extracts the feature vector from each video frame, we define the threshold as $mean(\{d(f(v_i), f(v_{i+1})) | i \in \mathbb{N}, 1 \leqslant i \leqslant (k'-1)\})$ where d is the Euclidean distance function. The preliminary analysis of our experiments showed that this produced better segmentation results than the average pairwise distance between every frame pair. Note that the selected images are not continuous but still belong to distinguishable experiences (e.g. working in front of a computer, walking along a corridor, climbing stairs, etc). When generating passwords, we select at most one representative image from each segment.

Clustering: We cluster images using the similarity of feature vectors. We use *Density-based spatial clustering of applications with noise* (DBSCAN) [9], which groups together feature points that are in a neighborhood. The algorithm requires two parameters: the distance threshold to group similar images and the minimum number of images in each cluster. The similarity threshold of images is determined by analysing the difference of consecutive frames. We use the same threshold value for both segmentation and clustering. The second parameter allows us to discard *noise* images, which pop up abruptly and are *non-informative*. These images are blurred, dark or do not contain interesting objects (almost empty). They do not provide useful information to assist the user's memory.

Challenge generation: We investigate two authentication schemes from egocentric videos: *experience-arrangement* and *experience-selection*. The former requires users to organize multiple experiences in their chronological order while the latter demands them to pick experiences that satisfy a time-based criteria. The system can vary the number of experiences in each password to satisfy different authentication purposes, such as instant log-in or fallback authentication. In addition, multiple challenges can be displayed sequentially to strengthen the security.

Experience-arrangement: In this scheme, users exploit their memory to arrange a set of experienced images in the right chronological sequence. To generate this kind of passwords, video frames which are in the same cluster but appear in different segments are removed from the candidate image set. This is because they are repetitive scenes or activities, which confuse the users when they try to guess the occurrence order. In Section V-A, we evaluated the user effort of this scheme in an object-interaction scenario.

Experience-selection: In this scheme, the system displays a grid of experienced images and the user is required to select the ones that satisfy a criteria, such as *"What have you not done today?"*. To do that, we supply the clustering algorithm with videos from consecutive days. After being segmented and clustered, images co-appearing on both days are discarded. We do not remove repetitive scenes occurring in the same day because they do not influence the user's choices. This type of passwords can be extended to diverse temporal scales. For example, the question *"What have you done after 11:00am today?"* separates events based on the recording time.

B. Image Features

To characterize global and local attributes of video frames, a feature vector is required to represent the whole scene appearance and the characteristics of local sub-regions. We concatenate the Census Transform Histogram (CENTRIST) [20] and the pyramid of histograms of orientation gradients (PHOG) [4]. These visual descriptors have been successfully applied in scene classification [20] and object-based image categorization [4].

We then perform Principle Component Analysis in order to reduce the number of dimensions, which benefits the segmentation and clustering stage in terms of computational load. After that, each feature vector which is utilized in segmentation and clustering contains only 100 entries.

C. Password Formats

In order to further improve performance and convenience in timeline-based login challenges, we investigated the use of image-based localization approaches [8]. The combination of both systems enables a new class of localization challenges based on historical location or experience and can also foster the use of clearer, pre-prepared images to improve the user experience. In particular, after receiving the location, our system then provides a representative, clear image of that location instead of the raw image recorded by the egocentric camera. In addition, authentication challenges based on location or context are possible.

We consider the following types of authentication challenges as depicted in Figure 3: (1) raw egocentric image, (2) clear images based on the location of the user, (3) textual description of the location, (4) textual description of the context, (5) textual description of context and location, and (6) combination of (2) and (5). These authentication challenges are evaluated for their mental load in section V-C.

IV. Security Analysis

Our proposed approach facilitates a two-factor authentication mechanism composed of a wearable camera C and chronological image sequences. To log-in a personal device D, attackers must obtain both C and D. If the possession time is not enough to generate a new image-based challenge, adversaries need to reconstruct image timelines of users. The experiment in Section V-A has shown that the effort of informed attackers is significantly higher than that of legitimate users. This allows us to detect an attack. If the attacker, however, succeeds in wearing C for a sufficient duration to form new passwords, our scheme can not guard against the attack. In such case, we suggest to leverage gait characteristics [11] or head movements [14] for legitimate user identification.

Scheme		**Entry time (s)**
Experience-arrangement	*Object-interaction*	9.77
	Daily condition	9.79
Experience-selection	*Daily routine*	4.67
	Unfamiliar environments	3.10
Passfaces [10]		18.25
PassApp [17]		7.27

TABLE I
Entry time (seconds) of our approach comparing with representative image-based authentication schemes

In order to handle privacy issues related to wearable cameras, there have been solutions to filter or modify captured images. For example, object recognition and scene classification were applied to identify sensitive areas [18] and computer screens [13]. Video frames can also be degraded [15] before they are displayed. In this case, legitimate users are still able to recognize the images because they have observed the scenes [7]. In addition, there exist methods to protect private images processed on untrusted cloud servers, such as the one presented by Wang *et al.* [19].

V. Experiments and Analysis

We conducted experiments with various scenarios to evaluate the user effort on our proposed schemes. The settings included object-interaction in an office room, daily activities, routine, and unfamiliar environments. The user effort in terms of mean entry time is summarized in Table I. In addition, we evaluated an attacking scenario with informed adversaries to assess the security of the approach.

Experience-arrangement and *experience-selection* passwords are both appropriate for touch screens. Thus, we implemented web applications that supported touch-based interaction to reduce the user effort. We proposed six formats to enhance the usability of the authentication challenges and their comparison was based on the NASA Task Load Index [1].

A. Experience-Arrangement Authentication

We evaluated the *experience-arrangement* authentication scheme in a usual workplace with basic furniture and office equipment. The scenario was challenging because all activities occurred at similar location. Thus, our system relied only on object appearance and interaction to form authentication challenges. The room and its layout were unfamiliar with all seven subjects (four females, two wearing glasses) participating in the experiment. Each subject was asked to perform at least five object-interaction experiences in any order, and later solve passwords generated from the recorded egocentric videos. The suggested experiences include: reading a book, working on a laptop, writing on paper, writing on a board, viewing a poster,

[1] http://www.nasatlx.com/

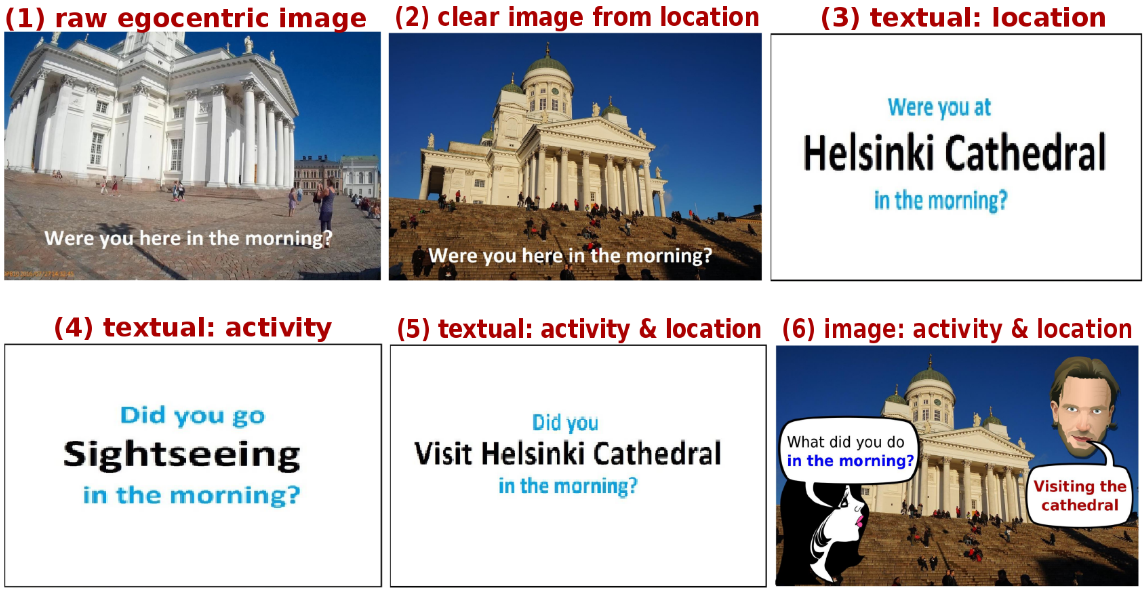

Fig. 3. Enhanced password formats to improve the usability

talking to a person, using a smartphone, unboxing an item, playing a board game, and using a paper-cutter. The objects are put on the desks (laptop, paper, gameboard, and paper-cutter), hung on the wall (poster and board), or in the subject's pocket (smartphone). The person who communicates with the subject is in the same room.

In this scenario, we used the Pupil headset [2], which has a camera attached to an eyeglass frame. We developed a web-based application with HTML and JavaScript. It allows touch-based interaction from mobile devices as well as access from desktop computers.

After each subject finished the suggested activities, we asked her to solve a sequence of four experience-based passwords. A new password was generated whenever the user formed a wrong sequence. When the subject succeeded in submitting the right chronological order of images, the number of attempts and entry time were recorded. Our participants spent on average 9.77 seconds with 1.21 attempts. The duration is comparable with other graphical password schemes such as PassApp [17] 7.27 seconds and Passfaces [10] 18.25 seconds. Remarkably, our proposed scheme exposes a new password if the previous attempt results in a wrong temporal sequence of images, which results in higher security level. Furthermore, for each user, we selected experiences from an ever-changing personalized image set, instead of a static portfolio.

Attacking the object-interaction scenario: We conducted an attacking simulation in which the adversaries had good knowledge on the environment (including locations of the furniture and suggested activities) and tried to solve others' authentication challenges. These attackers reported that they leveraged two strategies to obtain the correct answers: (1) leveraging their knowledge on the suggested activities and the furniture arrangement to form a candidate image order, and (2) fixing an arrangement for every graphical password without paying attention to the images. If they could not determine any answer, they tended to choose a random one. The mean time spent on a single challenge was 58.17 seconds with average 14.90 attempts to success, which was significantly higher than the effort of legitimate users. Hence, thresholds can be set on the user's effort (e.g. limits on entry time and number of attempts) to issue less convenient but more secure user verification methods or to lock the personal devices permanently.

Daily condition assessment: Four subjects (two females, two wearing glasses) agreed to continue the experiment in daily condition, i.e. wearing the camera in two consecutive days. For their comfort, we utilized a Transcend DrivePro^{TM} Body 10 camera [3] which could be clipped to clothes or backpack straps. Whenever possible, the camera wearer started to continuously record videos with the consideration of technical, legal, and social regulations. Recored videos contained visual data of users' daily activities and navigation. The system generated a new authentication challenge whenever the user attempted to log-in. It changed authentication challenges (i.e. new images) if the user constructed a wrong arrangement. The mean entry time was 9.79 seconds and the mean number of attempts was 1.87. The user effort was thus similar to that of the object-interaction scenario.

B. *Experience-Selection Authentication*

To evaluate *experience-selection* passwords, we analyzed two scenarios in which the subjects wore the cameras for several days and then solved the authentication challenges. Another web application was implemented that displayed a

[2] https://pupil-labs.com/

[3] https://www.transcend-info.com/Products/No-704

grid of photos as an authentication challenge. The images were extracted from videos of two consecutive days, namely d_{i-1} and d_i. Each subject was required to pick the valid images which belonged to events happening on d_{i-1} but not on d_i. Our system randomized both the total number of images n and the quantity of valid photos k and did not reveal the latter parameter to users. In a password containing n images (i.e. password length is n), there would be $1 \leq k \leq (n-1)$ valid ones. The user may select and deselect an image multiple times until achieving the correct configuration. In the application, clicking on an image alternated its state. We recorded the number of clicks instead of the number of attempts. The entry time is calculated from when the password fully appears on the screen until when the proper selection is obtained.

Daily routine: Four subjects wore the Transcend DrivePro$^{\text{TM}}$ Body 10 camera in two consecutive days. The recorded videos captured routine activities of the subjects, such as working, traveling (on foot, bicycle, and bus), etc. The *experience-selection* authentication challenges asked them to pick the images that occurred in one day but not in the other. We randomized the password length n from two to eight and did not show the number of valid images to the users. The mean entry time for all configurations is 4.67 seconds and the mean number of clicks is 2.92. If only eight-image passwords are considered, these values are 5.66 seconds and 3.68 clicks, respectively.

Discovering unfamiliar environments: The same camera was worn by a subject over a period of three weeks. The videos mostly contain scenes of home, workplace, and navigation (both on foot and on public transport). The subject even captured videos of a trip to Stockholm (Sweden), Wadern (Germany), and Brussels (Belgium), where the subject visited for the first time. The videos varied in activities, light condition, weather, and location. The subject performed the log-in actions multiple times during the experiment with dynamically-generated graphical passwords. Over various password lengths and number of valid images, the user consumed 3.10 seconds to click 5.14 times on the images on average. In the experiment, two-image passwords were solved on average in 2.2 seconds with 1.44 clicks. It makes this password length suitable for quickly unlocking a personal device. With the most secure configuration (i.e. eight-image passwords), the users solved them on average in 4.78 seconds with 7.44 clicks. In this case, unfamiliar environments can provide more cues to trigger users' memory than routine activities and scenes.

C. Enhanced Password Formats

To evaluate the complexity and workload of distinct login formats (cf. Figure 3), we utilized the NASA-TLX, which is comprised of 6 sub-scales representing somewhat independent variables: Mental, Physical, and Temporal Demands, Frustration, Effort, and Performance. It is assumed that some combinations of these dimensions likely represent the workload experienced by most people performing most tasks. The dimensions were selected after an extensive analysis of the primary factors that define workload for different people performing a variety of activities. Variables are rated for each task within a 100-points range with 5-point steps. In order to achieve subject-independent rating, in a second step, subject-dependent individual weighting is created by subjects rating variables for their pairwise perceived importance of the given task. The number of times each is chosen as being more important than others gives the weighted score, which is then multiplied by the scale score of the variable and divided by 15 to get a workload score from 0 to 100 (the overall task load index).

Chal.	TLX	Mental	Temp.	Perf.	Effort	Frustr.
(1)	37.58	7.96	6.8	6.0	4.05	12.71
(2)	4.33	0.78	0.44	1.36	1.18	0.4
(3)	28.62	8.67	7.24	4.82	4.69	3.2
(4)	16.85	4.24	5.02	2.36	3.27	1.96
(5)	19.93	3.89	5.27	3.38	3.31	4.09
(6)	24.11	5.24	3.47	3.24	4.04	8.11

TABLE II

Results of the Task Load Index for six distinct password formats. The table presents the weighted scores from the test. The physical dimension has been left out since it is not meaningful for this task

In particular, we compared six distinct login designs, each exploiting distinct pre-processing of the images to present. To be comparable and in order to focus on the impact of the way of presentation rather than the complexity of the challenge, we reduced the examples to the simplest case (cf. figure 3). The different authentication challenges are (1) the plain image as recorded by the egocentric camera, (2) a clean, representative of the scene provided by a server storing crowdsourced images, (3) a textual description of the location where the picture was taken, (4) a textual description of the activity conducted, (5) a textual description combining location and context, and (6) a dialogue-style combination of (2) and (5).

Ten participants aged 26-40 years with 70% female subjects participated. Tests were taken individually for all participants and in all cases, the interviewer has been present during the complete duration of the test. Participants initially have been introduced to the concept of authentication challenges based on egocentric vision and were then made familiar with the six different challenges considered (printed on paper). For each of the six formats, the participants were presented to respective formats and then they were asked to complete the test. For this, we exploited the online-version of the NASA-TLX test.

Results are presented in Table II. The TLX score is lowest for the format (2) and highest for (1). This shows that non-processed images directly taken from the egocentric camera require higher user effort to interpret. In addition, it was interesting to observe that all textual presentations scored lower TLX scores than when the unprocessed image was chosen in the format (1). This result is represented also by the individual weighted scores such as frustration which is again highest for (1) and lowest for (2).

D. Image Memorability Classification

The subjects participating in our experiments reported that they had been confused with some images in the authentication challenges; especially, when images were so generic that they could not help the users to recall the right chronological order, even though they are clear. This motivates us to design and implement an image classification algorithm to discard generic images, such as the ones that contain corridors or stairs.

In order to train the classifier, the subjects picked particularly confused and memorable images manually. As a result, they selected 192 confused images and 234 memorable images. These images were fed to an online evaluation service [4], which was developed from crowdsourced photos, to assess their memorability [12]. The mean memorability score of the former was 0.51 while that of the latter was 0.76 (1 is the most memorable). We then built a Support Vector Machine classifier using LibSVM [5] with CENTRIST [20] and PHOG [4] features.

We randomly split the above image sets into training and testing subsets. The classifier parameters were optimized with cross-validation. Then, the optimal model was applied on the testing subsets. We repeated this process ten times. The average correct classification rate and F-measure were 92% and 0.93, respectively. It shows that the technique is useful for filtering unmemorable images.

VI. Conclusion

We have presented a novel authentication mechanism based on collective computing concepts, exploiting implicit knowledge. The authentication scheme produces always fresh authentication challenges from egocentric video timelines and, in contrast to traditional password, pattern or biometric authentication, is robust against shoulder surfing, smudge attacks or theft of biometric or credential information. The authentication scheme exploits an on-body camera as a security token that can generate image-based authentication challenges through edge- or cloud-supported processing. We have evaluated the approach in various real-life scenarios and with a number of different authentication schemes. Authentication was shown to require about 3-10 seconds, which is similar to other context-based authentication approaches. We evaluated attacking scenarios with informed adversaries to assess the security of our approach.

References

[1] Gregory D. Abowd. Beyond weiser: From ubiquitous to collective computing. *IEEE Computer*, 2016.

[2] Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. Smudge attacks on smartphone touch screens. In *USENIX Conference on Offensive Technologies*, 2010.

[3] S. Bambach, S. Lee, D. J. Crandall, and C. Yu. Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1949–1957, Dec 2015.

[4] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *ACM International Conference on Image and Video Retrieval*, 2007.

[5] Daniel Castro, Steven Hickson, Vinay Bettadapura, Edison Thomaz, Gregory Abowd, Henrik Christensen, and Irfan Essa. Predicting daily activities from egocentric images using deep learning. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 75–82, 2015.

[6] Sauvik Das, Eiji Hayashi, and Jason I. Hong. Exploring capturable everyday memory for autobiographical authentication. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2013.

[7] Tamara Denning, Kevin Bowers, Marten van Dijk, and Ari Juels. Exploring implicit memory for painless password recovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2615–2618, 2011.

[8] Jiang Dong, Yu Xiao, Zhonghong Ou, Yong Cui, and Antti Yla-Jaaski. Indoor tracking using crowdsourced maps. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2016.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1996.

[10] Katherine M. Everitt, Tanya Bragin, James Fogarty, and Tadayoshi Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *SIGCHI Conference on Human Factors in Computing Systems*, 2009.

[11] Yedid Hoshen and Shmuel Peleg. An egocentric look at video photographer identity. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2016.

[12] Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 145–152, 2011.

[13] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. Enhancing lifelogging privacy by detecting screens. In *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2016.

[14] Sugang Li, Ashwin Ashok, Chenren Xu, Yanyong Zhang, Marco Gruteser, Janne Lindqvist, Marco Gruteser, and Narayan Mandayam. Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In *IEEE Conference on Pervasive Computing and Communications (PerCom)*, 2016.

[15] Bernhard Rinner and Thomas Winkler. Privacy-protecting smart cameras. In *Proceedings of the International Conference on Distributed Smart Cameras*, ICDSC '14, pages 40:1–40:5, New York, NY, USA, 2014. ACM.

[16] Edward Rosten and Tom Drummond. Fusing points and lines for high performance tracking. In *IEEE International Conference on Computer Vision*, 2005.

[17] Huiping Sun, Ke Wang, Xu Li, Nan Qin, and Zhong Chen. Passapp: My app is my password! In *International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2015.

[18] Robert Templeman, Mohammed Korayem, David Crandall, and Apu Kapadia. PlaceAvoider: Steering first-person cameras away from sensitive spaces. In *Proceedings of The 21st Annual Network and Distributed System Security Symposium (NDSS)*, February 2014.

[19] Qian Wang, Shengshan Hu, Jingjun Wang, and Kui Ren. Secure surfing: Privacy-preserving speeded-up robust feature extractor. In *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pages 700–710, Los Alamitos, CA, USA, 2016. IEEE Computer Society.

[20] Jianxin Wu and Jim M. Rehg. Centrist: A visual descriptor for scene categorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011.

[21] Ryo Yonetani, Kris M. Kitani, and Yoichi Sato. *Visual Motif Discovery via First-Person Vision*, pages 187–203. Springer International Publishing, Cham, 2016.

[4] http://memorability.csail.mit.edu/demo.html

[5] https://www.csie.ntu.edu.tw/ cjlin/libsvm/