Towards Secure Personal Device Unlock using Stereo Camera Pan Shots

Rainhard D. Findling and Rene Mayrhofer

Department for Mobile Computing, Upper Austria University of Applied Sciences,
Softwarepark 11, 4232 Hagenberg, Austria,
rainhard.findling@fh-hagenberg.at,
rene.mayrhofer@fh-hagenberg.at,
WWW home page: http://www.fh-ooe.at/en/

Abstract. Personal mobile devices hold sensitive data and can be used to access services with associated cost. For security reasons, most mobile platforms hence implement automatic device locking after a period of inactivity. Unlocking them using approaches like PIN, password or an unlock pattern is both problematic in terms of usability and potentially insecure, as it is prone to the shoulder surfing attack: an attacker watching the display during user authentication. Therefore, *face unlock* – using biometric face information for authentication – was developed as a more secure as well as more usable personal device unlock. Unfortunately, when using frontal face information only, authentication can still be circumvented by a photo attack: presenting a photo/video of the authorized person to the camera. We propose a variant of face unlock which is harder to circumvent by using all face information that is available during a 180° pan shot around the user's head. Based on stereo vision, 2D and range images of the user's head are recorded and classified along with sensor data of the device movement. We evaluate different classifiers for both grayscale 2D and range images and present our current results based on a new stereo vision face database.

1 Introduction

Personal devices like smartphones hold important and private user data in the regular case, like access to emails/private messages, contacts, calendars and location information. Hence, current smartphone platforms provide lock mechanisms as entering PIN/password or drawing an unlock pattern, with which the device has to be unlocked before usage. Unfortunately, the majority of current unlocking approaches like entering a PIN/password or drawing an unlock pattern are prone to the shoulder surfing attack [15,18] (an attacker watching the display while the user authenticates) and other attacks, such as the smudge attack for unlock patterns [2,22] (an attacker analyzing the smudges remaining on the display after unlocking). An attacker could consequently perform a replay attack to unlock another person's personal device.

Face unlock (face based authentication) is an approach to unlock a personal device, which aims at a) increasing usability, as the user does not have to remember an authentication secret and the authentication process can potentially be done faster, and b) being more robust against the shoulder surfing attack, as replay attacks are more complicated for face-based authentication. Unfortunately, using frontal face information only for face unlock, the unlock mechanism can be circumvented by photo attacks [1,11,19] (presenting a photo/a short video of the user with sufficient quality to the personal device's camera). For many regular users, such data can be grabbed from social networks. Therefore, using frontal face information only for a face unlock approach cannot be considered secure enough for a personal device holding important and private data.

We propose a variant of face unlock which is more robust and more secure against photo attacks than using frontal face information only. Our approach is based on our previous work [3] and uses all face information available from a 180° stereo camera pan shot around the user's head, which the user can take by simple panning the mobile phone with one arm in a half circle around the head. Using the mobile device's stereo camera, we record stereo images – a left/right pair of grayscale images – from multiple perspectives of the user's head. We record images instead of a video stream as they are usually of higher quality. For obtaining range images – grayscale images, in which the brightness represents the distance from the camera to the object – out of the recorded pair of stereo images, a stereo to range algorithm is applied. Then the grayscale and range images both get used for authentication. To circumvent a system based on this approach, grayscale and range images of multiple perspectives around the user's head would be needed. Therefore, an attacker will no longer be able to simply grab the attack data from a social network site, but is required to use a more complex data source – like a 3D model or a previously recorded stereo vision pan shot of the user's head.

2 Face Unlock based on Stereo Vision Pan Shot

For our current stereo vision pan shot face unlock (see figure 1), we first record grayscale stereo images of the user's head at multiple angles, along with gyroscope sensor data for each pair of images. For a pan shot of 180°, we record nine such image pairs (one pair for about each 22.5°). Using stereo to range algorithms, a range image can be derived from each stereo camera image pair. We use block matching stereo correspondence algorithm implemented in OpenCV [7] as our stereo to range approach – which currently delivers unsatisfying results: the resulting range image has large areas not covered with range information (displayed as white areas). Therefore, the further evaluation of our approach is done on the basis of precalculated range images taken out of our face database, as described in section 3.

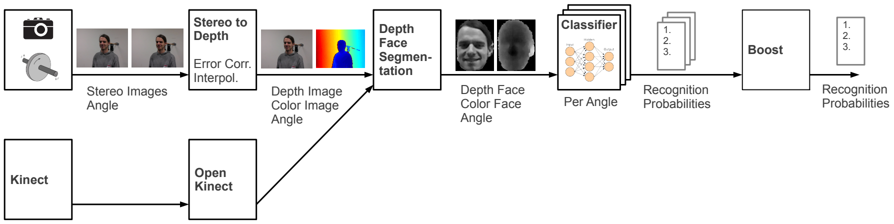


Fig. 1: Overview of the stereo vision pan shot face unlock system.

For processing face images, the face-related information of the image should get cropped first. In our approach, we use range based face segmentation – searching and cutting out faces in a range image – as described in section 4. Based on segmented grayscale and range face images and their device-angle information at the time of recording, face recognition is performed. For grayscale and range faces and for different perspectives, we use different classifiers. As classifiers, we use Support Vector Machines [12] and Neural Networks [10] for face recognition, as described in section 5. Each classifier estimates a probability of recognizing a certain person. To combine the probabilities of the classifiers for different angles and grayscale and range images, boosting those classifiers – using approaches like AdaBoost [4] or LogitBoost [5] – may be an appropriate option, and will be in the focus of our future research.

3 Evaluation Data

As we are not aware of a face database which contains face images in a form as they would be required for testing a pan shot face recognition approach, we created the Hagenberg pan shot face database 2013. It contains 20 pan shot image sets of 30 people with realistic indoor lightning conditions, recorded from 9 different perspectives around the user's head (see figure 2). For each perspective and person each pan shot image set contains:

- One high quality colored DSLR image.
- One pair of colored stereo images, recorded with an up to date mobile device with stereo camera.
- One color and one range image, recorded with a Microsoft Kinect and the OpenKinect framework.

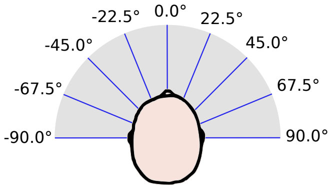


Fig. 2: Angles at which the pan shot images have been recorded.

The details of the Hagenberg pan shot face database 2013, including a recording setup description will be described in a future work. The evaluation of our current stereo vision pan shot face unlock approach is done based on the color and range Kinect images, contained in the face database (see figure 3).

4 Range Face Segmentation

Face recognition should be performed on basis of the grayscale and range input images. To only pass face related data to the classifiers, the face has to get extracted from the image first. One approach to extract a face from an image is to perform grayscale based face detection, such as the well known approach of

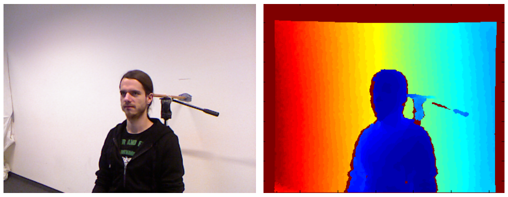


Fig. 3: OpenKinect color and range images, used as input for the system evaluation.

Viola and Jones [21] with Lienhart and Maydt [8]. In our previous work [3], this approach resulted a) in a notable amount of false positives/negatives, specially for the profile perspective [14], which causes the classifiers to already learn wrong data, and b) in having not face-related information (background) around the corners and borders of the extracted area.

Hence, many different approaches for more precise face segmentations have been proposed, such as [9,13,16,17]. For our current pan shot face recognition, we rely on a simple, but for our needs yet effective range-template based face segmentation:

1. A coarse person segmentation removes those parts of the image, which have a bigger distance to the camera than a predefined threshold value.
2. The human face gets searched in the range image, using an "average human face range template" in combination with a sliding window approach. For each of the nine perspectives there exists on such average human face range template (see figure 4).
3. Finally, for the best fit of the template in the image, the known area of face in the template gets cut out for both the grayscale and the range input image. This results in both one segmented grayscale and range face image (see figure 5).

The current face segmentation results are not fully accurate, as some minor areas of the faces are missing, and some not face-related information is still included in the extracted faces. Therefore, improving the range based face segmentation for our stereo vision pan shot face unlock will be focus of our future research. Still, the quality of our current face segmentation results is good enough for the results to get processed by classifiers, as described in the next section.

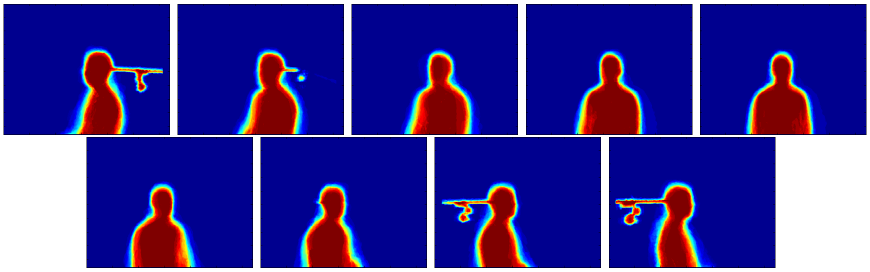
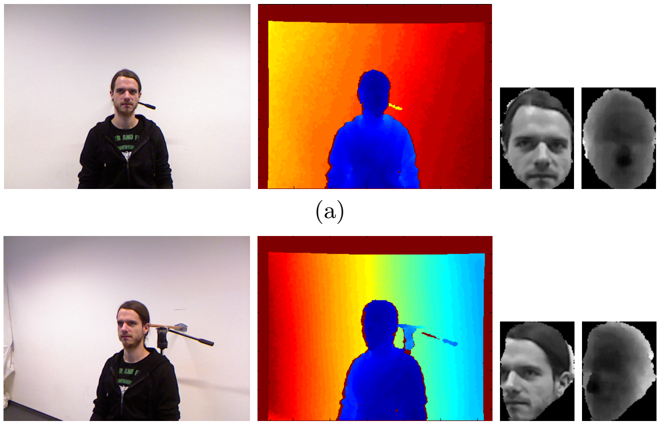


Fig. 4: Average human range templates for all nine perspectives.



(a)

(b)

Fig. 5: a) Frontal and b) side grayscale and range input images, and their corresponding range based segmented faces.

5 Face Recognition

Based on the obtained grayscale and range faces from face segmentation, along with the device rotation angle at the time of recording, face recognition is performed. We use different classifiers for grayscale and range images, and for each of the nine perspectives.In previous work [3], we used Eigenfaces for recognition [20] as a baseline for face recognition – which resulted in a final person recognition rate of 55.8%. In our current approach, we therefore use more promising approaches like Support Vector Machines (SVM) [12] and Neural Networks [10] for face recognition. To find a well adjusted parameter configuration for our recognition, we perform a search for the number of neurones in the hidden layer of the feed forward neural network, and a grid search for the correspondent SVM parameters, as suggested by Hsu et. al. [6].

5.1 Training and Test Procedure

The recognition is done as binary classification. For each of the 30 subjects, the face images of the test subject, recorded at a certain angle, represent the positive class, and the images of all other people of the same angle get assigned to the negative class. This leads to the positive class being $\frac{1}{29}$ of the negative class size. The classifiers get trained with 60% of the data of each the positive and negative class (train set). The remaining 40% of the data (test set) are explicitly used to measure the performance of the best classifiers in the end, see section 5.2.

Neural Networks: training for the feed forward neural networks (FFNN) is done as follows: for each angle, subject and classifier, 10 neural networks get trained with the correspondent part of the rtyain set. 30% of the data get used for training the neural networks, 12% percent to perform cross validation to stop the training, and the remaining 18% to evaluate the 10 generated neural networks against each other. The network with the best performance gets evaluated using the correspondent part of the train set then.

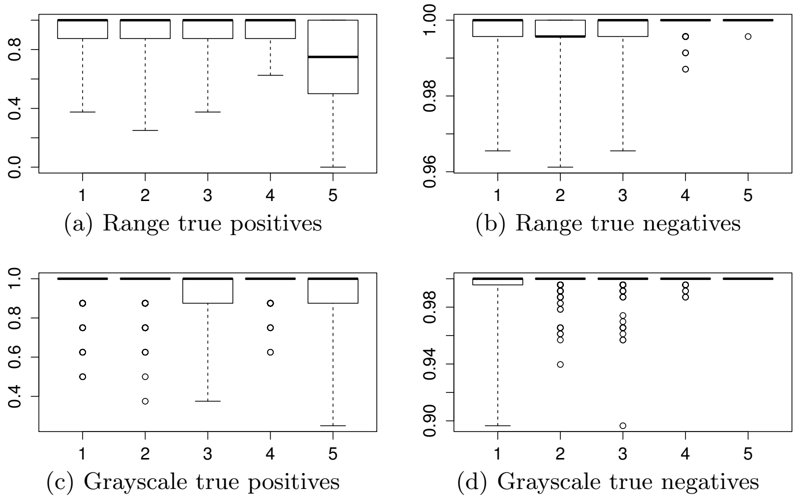
Support Vector Machines: training for the Support Vector Machines is done as follows: for each angle, subject and classifier, one support vector machine gets trained using the correspondent part of the train set, and evaluated on the correspondent part of the test set then.

5.2 Current Recognition Results

The classification results of the best performing support vector machine with linear kernel and radial kernel, and best performing neural networks (see table 1) are shown in the tables for range and grayscale classification results. The corresponding boxplot provides an overview of true positive and true negative classification results for both range and grayscale faces for all perspectives combined (see figure 6).

Nr.	Classifier	Neurons	Kernel	Cost	Gamma
1	FFNN	10	–	–	–
2	FFNN	17	–	–	–
3	FFNN	25	–	–	–
4	SVM	–	Linear	1	–
5	SVM	–	Radial	1	0.01

Table 1: Classifier parametrization.



(a) Range true positives

(b) Range true negatives

(c) Grayscale true positives

(d) Grayscale true negatives

Fig. 6: Face recognition results, all perspectives combined: for range a) true positives and b) true negatives, and grayscale c) true positives and d) true negatives.

The results show clearly: the range based face recognition performs slightly worse over the grayscale face recognition. For all classifiers – except of two – the median is 1, but for the mean and first quartile, a clear distinction to the favor of grayscale face recognition is visible: the first quartile of true positive rate for range classifiers is at maximum 87.5%, compared to at least the same value for grayscale classifiers. As the positive class is much smaller in size than the negative class, true negative results are better overall. Again, grayscale performs slightly better than range: the first quartile goes down to 99.57% for three range classifiers, while going down to the same value for one grayscale classifier only. The best performing classifier (SVM with linear kernel) has a mean true positive rate of 93.89% for range faces, which is slightly lower than the mean of 96.85% for grayscale faces. The true negative rate of 99.95% is again slightly lower than the true negative rate of 99.97%. Still, the overall recognition rate obtained by this classifier indicates that range based pan shot face recognition is possible and can be combined with grayscale face recognition results for further usage.

We therefore argue that using additional range faces for pan shot based face unlock will be a feasible approach – even if our range face recognition results are slightly worse over the grayscale recognition results. The slightly worse range recognition rate will be offset by the increased effort, which an attacker will have to accept in order to obtain the additional range data of the user's face.

6 Conclusion / Future Work

We presented a variant of face unlock which uses stereo vision and 180° pan shots around the user's head, and therefore, is harder to circumvent by a photo attack than with using grayscale and frontal face information only. Based on range images from our stereo vision face database, we perform range-template-based face segmentation to find and cut out faces in the range, and the corresponding grayscale images. Using these faces, we evaluate grayscale and range face recognition capabilities of different Support Vector Machines and Feed Forward Neural Networks. For the best performing classifier, we achieve a mean true positive rate of 93.89% for range, and of 96.85% for grayscale face recognition. For the same classifier, the true negative rate is 99.95% for range, compared to 99.97% for grayscale face recognition. These results indicate that range based face recognition can be used along with grayscale face recognition in a pan shot face unlock scenario. This will increase the amount of required data – and therefore the effort an attacker will have to accept to obtain this data – in order to successfully circumvent a grayscale and range pan shot face unlock system.

Our future research will focus on improving the range template based face segmentation, as on combining the results obtained from grayscale and range classifiers for different perspectives to scalar values, using approaches such as classifier boosting.

7 Acknowledgments

This work has been carried out by u'smile, the Josef Ressel Centre for User-Friendly Secure Mobile Environments. We gratefully acknowledge funding and support by the Christian Doppler Gesellschaft, A1 Telekom Austria AG, Drei-Banken-EDV GmbH, LG Nexera Business Solutions AG, and NXP Semiconductors Austria GmbH. We further want to thank Fabian Wenny for helping at the face segmentation implementation and Christan Gabler for helping at the creation of the Hagenberg stereo vision pan shot face database 2013.

References

1. A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7, 2011.
2. A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX conference on offensive technologies*, pages 1–7, Berkeley, CA, USA, 2010.
3. R. Findling and R. Mayrhofer. Towards face unlock: On the difficulty of reliably detecting faces on mobile phones. In *Proc. MoMM 2012: 10th International Conference on Advances in Mobile Computing and Multimedia*, pages 275–280. ACM, December 2012.
4. Y. Freund and R. E. Schapire. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting. In P. M. B. Vitányi, editor, *Second European Conference on Computational Learning Theory*, pages 23–37, 1995.
5. J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:2000, 1998.
6. C. W. Hsu, C. C. Chang, and C. J. Lin. *A practical guide to support vector classification.* Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2003.
7. K. Konolige. Small vision systems: Hardware and implementation. In Y. Shirai and S. Hirose, editors, *Robotics Research*, pages 203–212. Springer London, 1998.
8. R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing 2002*, pages 900–903, 2002.
9. I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comput. Vision*, 60(2):135–164, Nov. 2004.
10. T. M. Mitchell. *Machine Learning.* McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
11. G. Pan, Z. Wu, and L. Sun. *Liveness Detection for Face Recognition*, chapter Recent Advances in Face Recognition, page 236. InTech, 2008.
12. P. J. Phillips. Support vector machines applied to face recognition. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Neural Information Processing Systems*, volume 10, pages 803–809, 1998.
13. C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, Aug. 2004.
14. M. C. Santana, O. Déniz-Suárez, L. Antón-Canalís, and J. Lorenzo-Navarro. Face and facial feature detection evaluation - performance evaluation of public domain haar detectors for face and facial feature detection. In A. Ranchordas and H. Arajo, editors, *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.*, volume 2, pages 167–172, 2008.
15. F. Schaub, R. Deyhle, and M. Weber. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 13:1–13:10, New York, NY, USA, 2012. ACM.
16. M. Segundo, L. Silva, O. Bellon, and C. Queirolo. Automatic face segmentation and facial landmark detection in range images. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 40(5):1319 –1330, oct. 2010.
17. H. Seibert. Efficient segmentation of 3d face reconstructions. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pages 31–34, July 2012.
18. F. Tari, A. A. Ozok, and S. H. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on Usable privacy and security*, SOUPS '06, pages 56–66, New York, NY, USA, 2006. ACM.
19. R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, and F. Roli. Fusion of multiple clues for photo-attack detection in face recognition systems. In *International Joint Conference on Biometrics*, pages 1–6, Oct. 2011.
20. M. Turk and A. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1):71–86, Jan. 1991.
21. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:511–518, 2001.
22. E. von Zezschwitz, A. Koslow, A. De Luca, and H. Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 277–286, New York, NY, USA, 2013. ACM.