Chapter 13
Mobile User Re-authentication Using Clothing Information

Hoang (Mark) Nguyen, Ajita Rattani and Reza Derakhshani

Abstract Biometric authentication has become a popular alternative to passwords on mobile devices. However, most implementations do not incorporate any mechanisms to ascertain whether the originally authenticated user is still in control of the mobile device. Thus, the user has to re-scan for any subsequent device access, which may lead to biometric scan fatigue. One solution to this problem is to re-authenticate the user via ancillary surrogates of identity that are likely to be stable and unique in the short term and easier to acquire compared to the primary biometric modality, such as opportunistically captured clothing information. The aim of this paper is to investigate such clothing information as a soft biometrics for short-term mobile user re-authentication. To this aim, we propose a novel method for segmentation and matching of clothing ROI from images captured via front-facing camera of mobile devices, without explicitly requiring the face to be present. Experimental investigations on a large-scale mobile dataset show error rates as low as 2.5% using this method.

13.1 Introduction

Mobile devices are playing a significant role in daily life, not only for communications but also for entertainment, e-commerce, and even remote health services. However, mobile phones are misplaced, lost, and stolen more than other computing devices. Therefore, efforts have been directed at the development of biometrically secure

H. (Mark) Nguyen (✉) · R. Derakhshani
Department of Computer Science and Electrical Engineering,
University of Missouri at Kansas City, Kansas City, USA
e-mail: hdnf39@mail.umkc.edu

R. Derakhshani
e-mail: derakhshanir@umkc.edu

A. Rattani
Department of Electrical Engineering and Computer Science,
Wichita State University, Wichita, KS, USA
e-mail: ajita.rattani@wichita.edu

mobile access and transactions. The use of biometric technology in mobile devices is referred to as *mobile biometrics* [9, 16, 17, 24]. Biometrics Research Group, Inc.[1] has predicted that by 2020, mobile biometrics will transition from consumer adoption phase to full maturity, enabling the technology to overtake existing authentication technologies. By 2020, it is estimated that biometrics will be ubiquitous, installed on 100 percent of mobile devices.

Thus, many commercial solutions as well as well as academic studies have been focusing on mobile user authentication via strong primary biometric traits. In particular, modalities based on face [9, 16] and ocular region [14, 15, 17, 20] acquired using selfie images are of interest given that they do not require any specialized sensors.[2] Fingerprint and near-infrared iris captured [4, 25] using dedicated sensors installed in mobile devices have also been used for mobile user authentication.

However, most of these methods focus on entering the user into the authenticated state via the primary biometric but provide no explicit or robust solution to keep the user in that state. In other words, they have no mechanism to determine whether the user authorized after the initial successful authentication is still the same person in control of the device [10]. If the device locks up or logs out after the initial access, the user has to frequently re-scan his or her biometrics using the primary modality to regain access to the device and its services, each time requiring a certain level of cooperation and attention, leading to bad user experience. Alternatively, if a timer is used to extend the initial authenticated state, there is still a risk of illegitimate access to the sensitive information on the device by an intruder if the device was taken from its original user in the meantime. To mitigate this problem, there is a need for short-term, low friction user re-authentication to properly extend the authenticated state after the initial primary biometric scan by the authorized user [2, 10, 23, 26].

Two most important factors for frequent and even continuous user authentication are reliability and usability. Primary biometrics such as face, eye, and finger scans are highly reliable but require a non-negligible active user cooperation for an acceptable scan (e.g., aligning the face or eyes with the camera or placing a clean finger on the fingerprint scanner), reducing their utility for frequent re-authentication. Further, these traits might not be available due to the user's pose. Less cooperative soft biometrics such as gender, skin color, and other face attributes, as well as other modalities like keystrokes and device movement dynamics [12, 23] have gained attention for user re-authentication in the background.

In this work, we investigate the use of clothing information as soft biometrics for short-term mobile user re-authentication. Clothing information has been studied extensively in person re-identification for multi-camera surveillance systems [5–7]. The advantages of using clothing information for mobile user re-authentication are as follows:

---

[1]https://www.biometricupdate.com/201703/special-report-mobile-biometric-applications.

[2]A selfie is a self-portrait image of a user captured using the ubiquitous front-facing cameras available in virtually all mobile devices.

- Clothing, as something that one has, and after being temporarily tied to the user identity at the time of the primary biometric scan, is usually unique and stable enough to be used for re-authentication for ensuing several minutes.
- Though clothing, as detailed above, may constitute a temporary visual representation of an individual, it is inherently revocable and unlike other soft biometrics the information stored in the template generally does not compromise user's privacy.
- Clothing ROI is a much larger target compared to the face and eyes, and thus it can be acquired from the front-facing camera while a user is naturally interacting with the target application with no explicit cooperation (except an initial consent to allow the method).

It should be noted though that this method is not applicable to scenarios where people wear uniform clothing, nor when the device camera is not in the general direction of the user's torso. The latter is indeed a benefit, since re-authentication should not happen when the user is not naturally interacting with the app that requested the service. That is also the time window when the OS permissions allow the use of the device cameras.

Our earlier study in [11] consisted of a preliminary investigation on the use of clothing information for mobile user re-authentication. The new contribution of this work over [11] are as follows:

1. A new deep learning-based method for more accurate segmentation of clothing ROI from selfie images that is robust to different user poses, rendering this method much more applicable to everyday mobile use cases.
2. Evaluation of SURF keypoint detectors and patch descriptors for matching clothing ROIs from selfie image pairs, followed by a comparative evaluation of this non-learning-based texture descriptor method with learning-based methods across various scales to better understand the pros and cons of each methodology.

The rest of this paper is organized as follows: Sect. 13.2 describes the existing work related to continuous mobile user authentication. Section 13.3 describes the proposed segmentation and matching methods for clothing-based short-term user re-authentication. Experimental validations of the proposed method are discussed in Sect. 13.4. Conclusions and future work are given in Sect. 13.5.

13.2   Previous Work

In this section, we discuss existing soft biometric methods applicable to the mobile device user re-authentication.

Samangouei et al. [23] proposed facial attributes such as gender, ethnicity, eyeglasses, hair color, skin type, and face shape as an auxiliary authentication method for mobile devices. Binary SVM classifiers were trained for each attribute. The learned classifiers were applied to the selfie image of the user for attribute's extraction. Authentication was done by comparing the extracted attributes with the enrolled attributes of the user.

Zhao et al. [26] investigated the touch-based continuous mobile authentication via proposing a novel Graphic Touch Gesture Feature (GTGF). In this method, touch traces were converted to images for the explicit representation of the touch dynamics. The touch sequences were first segmented and normalized so that traces have a fixed number of sample points. Then, the samples on the normalized traces were converted into shapes and intensity values of the GTGF. User authentication was performed by computing L1-norm between a pair of GTGF images. In [22] a text-based multimodal biometric approach utilizing linguistic analysis, keystroke dynamics, and behavioral profiling was proposed for continuous mobile user authentication.

Crouse et al. [2] proposed an unobtrusive continuous authentication system based on face matching. Performance and accuracy for unconstrained face matching were improved by integrating data from the device accelerometer, gyroscope, and magnetometer to correct camera sensor orientation and hence face image.

Rattani et al. [18, 19] proposed convolutional neural networks for gender and age prediction from ocular images captured using mobile devices for performance enhancement and potential re-authentication. In another work [10], authors exploited the use of eyebrows for short-term mobile user authentication. Eyebrows, being one-sixth of the facial region, is computationally efficient and offers fast throughput for continuous re-authentication in mobile devices. To this aim, the histogram of oriented gradients and GIST descriptors extracted from left and right eyebrow regions were evaluated.

The above studies, though helpful in their given contexts, do not solve the problem of user re-authentication without needing the face to be in view, or they may require user interaction with an additional touch-based modality. To the best of our knowledge, the line of studies starting with [11] was the first attempt on continuous user authentication using clothing information from selfie images in the mobile environment. In that the preliminary study, learning-based methods using local texture descriptors along with support vector machines (SVM) were applied on clothing ROI that was approximated through heuristics.

13.3 Proposed Method

The main steps involved in the proposed method are (a) selfie-pose-invariant clothing ROI segmentation and (b) robust matching of the features extracted from clothing ROI. We evaluated the efficacy of both learning and non-learning methods for the latter. Next, we discuss these steps in detail.

13.3.1 Clothing Segmentation

The segmentation task can be viewed as a pixel-wise labeling where the system differentiates between the pixels of clothing from those of the background. Deep

learning-based segmentation methods have been outperforming traditional methods. It has become common to use convolutional encoder–decoder models for this purpose. The encoder layers extract features from input data while the decoder layers reconstruct the image by way of the feature maps [8]. The model produces a binary mask of the original image size delineating the background from the foreground target object, respectively.

In this work, we used U-Net [21]-based deep learning model for clothing ROI segmentation. U-Net is a convolutional neural network that was originally developed for biomedical image segmentation. The network architecture of U-Net consists of contracting part (encoder) on the left and expansive path (decoder) on the right. The encoder is a repeated application of two 3×3 convolutions, followed by rectified linear units (ReLU), and 2×2 max pooling operation. Similarly, each decoder layer consists of an upsampling using 2×2 up-convolution, a concatenation of corresponding feature maps from the contracting path, and two 3×3 convolutions followed by ReLUs. This was also the first network to introduce skip connections for directly connecting the upsampling and downsampling layers. This allows the network to take the context of the image into account, which could be lost through the convolution operation otherwise. The architecture of the network is designed for parametrization with fewer training images, and it yields more precise segmentations.

For clothing segmentation, we trained the U-Net model with 1000 selfie images collected from the web. The dataset was further augmented by adding Gaussian blur, scaling, and rotation to the original selfie images, along with target binary masks. The training clothing masks were created using MATLAB's built-in "imageLabeler" function.

Figure 13.1 shows the architecture of U-Net for clothing mask generation from selfie images.
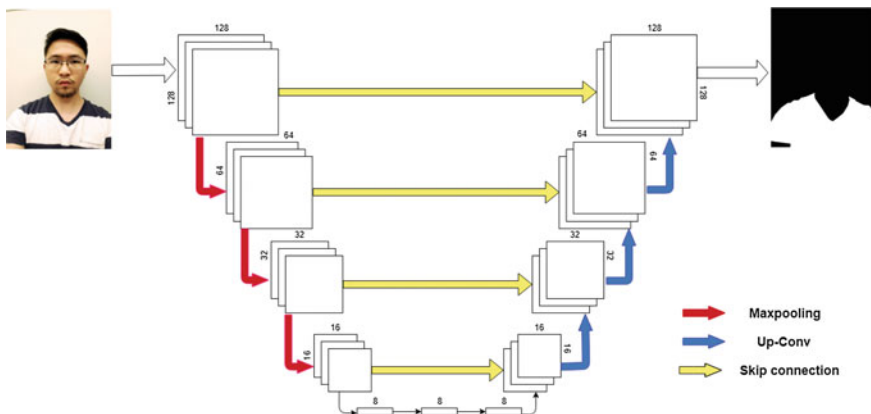

Fig. 13.1 Architecture of U-Net model used for clothing mask generation from selfie images

13.3.2   Clothing Matching

Clothing matching is the process of confirming whether two visual representations are from the same clothes or not. This is done by feature extraction from segmented clothing ROIs and matching them using either learning or non-learning-based methods. Next, we discuss our proposed learning and non-learning methods for the purpose.

13.3.2.1   Learning-Based Method

We define a learning-based method as one where the discriminant (or the similarity metric) is learned via training data. In the proposed learning-based method, tile texture features are used to train an SVM as the learned similarity metric. The trained SVM is then used for re-authentication. Based on the literature features and our various experiments, we found local binary pattern (LBP) [13], histogram of oriented gradient (HOG) [3], and color histogram (CH) to be most effective for this task. LBP is a simple visual descriptor that encodes the differences between the given center pixel with those in its neighborhood. HOG computes the local gradient orientation of the dense grid with local contrast normalization. LBP and HOG both operate on gray-scale images. CH generates color information from the histogram of R, G, and B channels. All features are extracted by dividing clothing ROI into 2×3 non-overlapping tiles at four different image scales ($1\times$, $0.5\times$, $0.25\times$ and $0.125\times$), an arrangement that was experimentally determined to be most effective. All these LBP, HOG, and CH feature vectors are then concatenated into a single vector as shown in Fig. 13.2 and used for training and testing the SVMs. We experimentally determined linear SVMs to provide the best generalization.
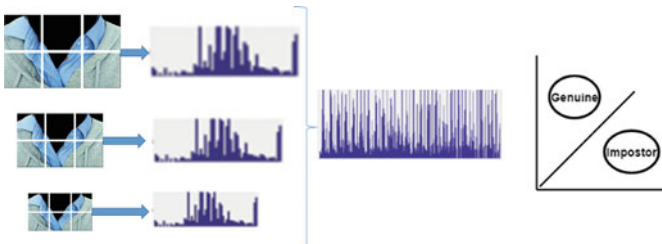

Fig. 13.2   Features extracted from a clothing ROI that is divided into 2×3 blocks at three different scales. All the extracted features from the different scales are concatenated into a single vector prior to classification
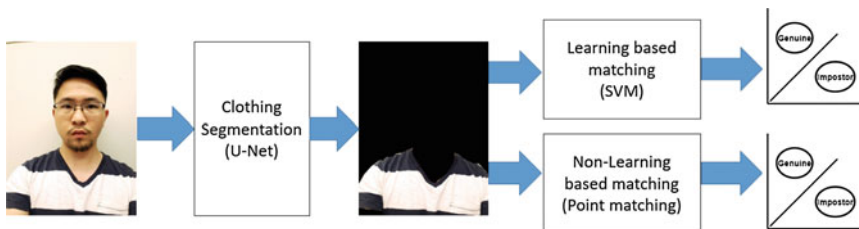
13.3.2.2 Non-learning-Based Method

We define a non-learning-based method as one where the discriminant is a pre-defined distance metric, such as Euclidean or Manhattan distance. In our non-learning-based method, we used the venerable speeded up robust features (SURF) [1]. SURF has been proven to be one of the best local feature detectors and descriptors for object recognition and image classification. In order to detect interest points, it uses Hessian matrix with the approximation of Gaussian smoothing. Similar to the scale-invariant feature transform (SIFT), interest points are calculated at different scales of the image pyramid. The descriptors around each interest point are computed using the first-order Haar wavelet responses which represent the intensity distribution of pixels within a block. The match score is computed as the number of matched SURF points between enrollment and verification clothing ROIs using the sum of absolute differences (Manhattan distance), experimentally deemed to be the best for this use case. Figure 13.3 shows the matching of SURF descriptors from clothing pairs coming from same (genuine) and different (impostor) clothing ROIs.

The obvious advantage of learning-based method is its higher accuracy over non-learning methods given its data-driven similarity metric. However, non-learning methods are usually computationally more efficient, do not require an extensive training process, and being more generic they may generalize better over certain unseen datasets. Figure 13.4 shows the overall proposed system.
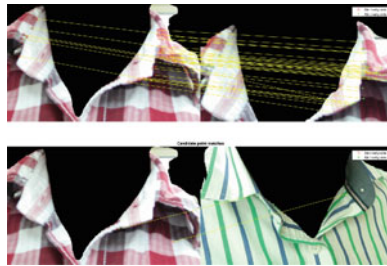
Fig. 13.3 SURF point matching between a pair of similar (genuine) clothing ROIs (top) and different (impostor) clothing (below)

Fig. 13.4 Overview of the short-term user re-authentication system based on clothing information. The main steps are clothing segmentation using U-Net followed by matching using proposed learning or non-learning-based methods

13.4 Experimental Validation

13.4.1 Dataset and Protocol

The dataset used in this work is a subset of full face mobile dataset used to generate VISOB dataset [17]. VISOB dataset was collected by acquiring full face selfie images from around 550 healthy adults using front-facing cameras of mobile devices. The subset of the dataset consisting of about 240,000 selfie images from 293 subjects using an OPPO N1 cellular phone. Out of the whole subset, the pre-trained segmentation algorithm detected masks for about 85,000 of images containing enough clothing information. Approximately half of these images were used for training and testing. Both the sets were further subdivided based on lighting conditions at the time of capture: daylight and indoor office lighting, for experimental analysis of system performance across different lighting conditions. Equal error rate (EER), area under the ROC curve (AUC), and precision and recall were used as performance metrics in our analysis.

13.4.2 Results

In this section, we present and discuss the result of proposed clothing segmentation and matching using learning and non-learning-based methods.

13.4.2.1 Clothing Segmentation

In order to evaluate the segmentation accuracy, we used precision and recall metrics given in Eqs. 13.1 and 13.2, respectively. In these equations, S is the segmentation mask obtained by U-Net model, and R is the ground truth label mask. Precision is the fraction of pixels that are segmented correctly over the total pixels in clothing mask generated by U-Net. Recall is the fraction of pixels that are segmented correctly over the total pixels in the ground truth label mask. Using the above equations, we obtained precision and recall of 94.73 and 94.03%, respectively. The high precision and recall rates suggest the efficacy of the proposed method for clothing ROI segmentation. Figure 13.5 shows the examples of segmented clothes and clothing masks.

$$Precision = \frac{S \cap R}{|S|} \tag{13.1}$$

$$Recall = \frac{S \cap R}{|R|} \tag{13.2}$$

Fig. 13.5 Example of **a** original selfie images, **b** segmented clothes, **c** and the corresponding masks obtained by our U-Net model segmentation. The eye regions have been masked in order to preserve the privacy of the participants
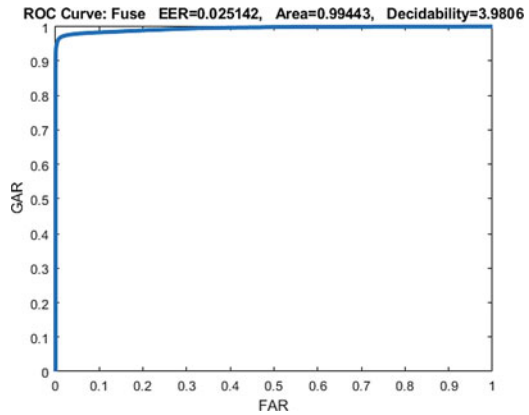
13.4.2.2   Learning-Based Clothing Matching

Table 13.1 shows the performance of learning-based method for clothing matching in terms of EER and AUC across same and different lighting conditions. Recall that learning-based method consists of feature level fusion of LBP, HOG, and CH feature vectors for SVM training and classification. Understandably, a very low error rate is obtained when training and testing sets are acquired under the same lighting conditions. The least EER of 2.5% was obtained when training and testing sets were acquired using indoor office lighting condition. However, the EER increased when lighting conditions were varied. The EER increased to 10.7% when the training images were acquired in office lighting conditions and test images came from daylight captures. Similarly, EER increased to 13.9% when the training images were acquired under daylight conditions and test images came from indoor office lighting conditions. This suggests that the method is sensitive to illumination variations. Figures 13.6 and 13.7 show ROC curves of the learning-based method across same and different lighting conditions.
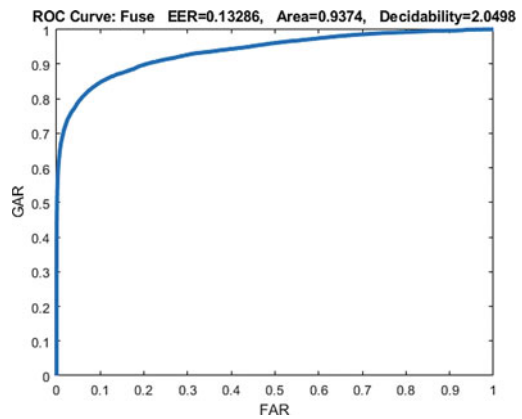
Table 13.1 AUCs and EERs of learning-based method with same and different lighting conditions

Train	Test	AUC	EER (%)
Office light	Office light	0.994	2.5
Daylight	Daylight	0.992	3.5
Office light	Daylight	0.954	10.7
Daylight	Office light	0.937	13.9

Fig. 13.6 ROC of learning-based method for clothing matching when the training and test images are all acquired under indoor office lighting conditions

Fig. 13.7 ROC of learning-based method when the training and test images are acquired under daylight and office lighting conditions, respectively

13.4.2.3   Non-learning-Based Clothing Matching

Table 13.2 shows the performance of non-learning-based SURF matcher. Again, it can be seen that lower EERs are obtained when the pair of selfie images were captured under the same lighting conditions. EERs of 11.9 and 13.9% were obtained when images were acquired under the same office lighting or daylight conditions,

Table 13.2 AUCs and EERs of non-learning method using same and different lighting conditions

Train	Test	AUC	EER (%)
Office light	Office light	0.943	11.9
Daylight	Daylight	0.929	13.8
Office light	Daylight	0.888	19.7
Daylight	Office light	0.875	18.9

Fig. 13.8 ROC of the non-learning method when the training and test images are acquired under office lighting condition

ROC Curve: SURF   EER=0.11896,   Area=0.94292,   Decidability=1.5986
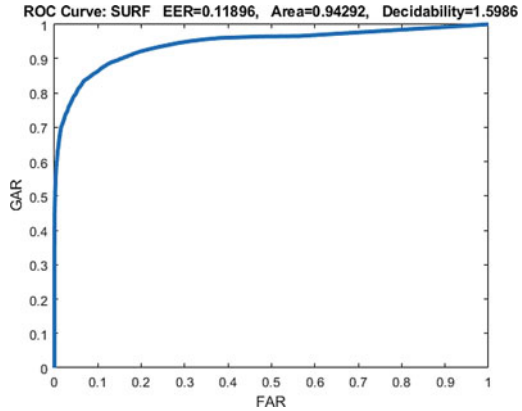
Fig. 13.9 ROC of the non-learning method when the training and testing images are acquired under office and daylight conditions, respectively

ROC Curve: SURF   EER=0.19669,   Area=0.88761,   Decidability=1.1525
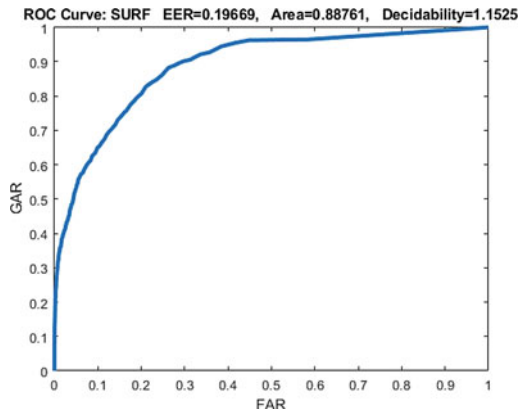
respectively. However, the performance drops for training and testing across different lighting conditions. 18.9 and 19.7% EERs were obtained when training and testing images were acquired under mixed office light and daylight conditions.

Figures 13.8 and 13.9 show the ROCs for non-learning clothing matching under same and different lighting conditions, respectively.

13.5   Conclusion and Future Work

In this paper, we showed the utility of partial clothing information, seen on the user's upper torso during uncooperative, free form interaction with a mobile device with front-facing cameras, for short-term re-authentication. We treat such clothing information as a soft identifier (something that user has and does not change in short term) if and when tied to a strong identifier such as a primary biometric that enters the user into the authenticated state. Here we show that, using our proposed clothing

segmentation and matching methods, one can obtain acceptable error rates to keep the user authenticated if he/she returns to a previously (biometrically) authorized device after a short period of time, without needing extra explicit biometric scans, for better user experience. The obtained error rates for matching clothing information are quite low when the verification clothing images are captured under similar lighting conditions that were used for training (2.5 and 11.9% EERs for learning and non-learning-based matching methods, respectively). However, the error rates increase across different lighting conditions. As a part of future work, a large-scale retraining and evaluation of the proposed methods will be conducted on other available mobile datasets. The proposed methods can be made more resilient to varying lighting conditions by including lighting variability into larger training sets, utilizing lighting-equalizing preprocessing, and by employing more resilient matching. More specifically, deep learning-based methods will be developed for matching clothing ROIs. Further, an adaptive fusion of clothing information with other available soft biometrics traits, such as the presence of eyeglasses, skin color, and gender, will be investigated for further performance enhancements.

References

1. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: Leonardis A, Bischof H, Pinz A (eds) Computer Vision - ECCV 2006. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 404–417
2. Crouse D, Han H, Chandra D, Barbello B, Jain AK (2015) Continuous authentication of mobile user: fusion of face image and inertial measurement unit data. In: International conference on biometrics (ICB), pp 135–142
3. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 886–893
4. Gao M, Hu X, Cao B, Li D (2014) Fingerprint sensors in mobile devices. In: 2014 9th IEEE conference on industrial electronics and applications, pp 1437–1440
5. Hermans A, Beyer L, Leibe B (2017) In defense of the triplet loss for person re-identification. CoRR, abs/1703.07737
6. Jaha ES, Nixon MS (2014) Soft biometrics for subject identification using clothing attributes. In: IEEE international joint conference on biometrics, pp 1–6
7. Jaha ES, Nixon MS (2016) From clothing to identity: Manual and automatic soft biometrics. IEEE Trans Inf Forensics Secur 11(10):2377–2390
8. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 3431–3440
9. Marsico MD, Galdi C, Nappi M, Riccio D (2014) Firme: face and iris recognition for mobile engagement. Image Vis Comput 32(12):1161–1172
10. Mohammad AS, Rattani A, Derakhshani R (2018) Short-term user authentication using eyebrows biometric for smartphone devices. In: IEEE computer science and electronic engineering conference
11. Nguyen H, Sai R, Li Z, Derakhshani R (2018) User re-identification using clothing information for smartphones (accepted). In: IEEE International symposium on technologies for homeland security (HST), pp 1–6
12. Niinuma K, Park U, Jain AK (2010) Soft biometric traits for continuous user authentication [dec 10 771–780]. IEEE Trans Inf Forensics Secur 6(4):771–780

13. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
14. Rattani A, Derakhshani R (2017) On fine-tuning convolutional neural networks for smartphone based ocular recognition. In: 2017 IEEE international joint conference on biometrics (IJCB), pp 762–767
15. Rattani A, Derakhshani R (2017) Online co-training in mobile ocular biometric recognition. In: 2017 IEEE international symposium on technologies for homeland security (HST), pp 1–5
16. Rattani A, Derakhshani R (2018) A survey of mobile face biometrics. Comput Electr Eng 72:39–52
17. Rattani A, Derakhshani R, Saripalle SK, Gottemukkula V (2016) Icip 2016 competition on mobile ocular biometric recognition. In: 2016 IEEE international conference on image processing (ICIP), pp 320–324
18. Rattani A, Reddy N, Derakhshani R (2017) Convolutional neural network for age classification from smart-phone based ocular images. In: 2017 IEEE international joint conference on biometrics (IJCB), pp 756–761
19. Rattani A, Reddy N, Derakhshani R (2018) Convolutional neural networks for gender prediction from smartphone-based ocular images. IET Biom 7:423–430(7)
20. Reddy N, Rattani A, Derakhshani R (2018) Ocularnet: deep patch-based ocular biometric recognition. In: 2018 IEEE international symposium on technologies for homeland security (HST), pp 1–6
21. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597
22. Saevanee H, Clarke N, Furnell S, Biscione V (2015) Continuous user authentication using multi-modal biometrics. Comput Secur 53:234–246
23. Samangouei P, Patel VM, Chellappa R (2015) Attribute-based continuous user authentication on mobile devices. In: 2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS), pp 1–8
24. Tao Q, Veldhuis R (2006) Biometric authentication for a mobile personal device. Third annual international conference on mobile and ubiquitous systems: networking services., San Jose CA, pp 1–3
25. Thavalengal S, Bigioi P, Corcoran P (2015) Iris authentication in handheld devices - considerations for constraint-free acquisition. IEEE Trans Consum Electron 61(2):245–253
26. Zhao X, Feng T, Shi W (2013) Continuous mobile authentication using a novel graphic touch gesture feature. In: 2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS), pp 1–6