Exploring New Authentication Protocols for Sensitive Data Protection on Smartphones

Chiara Galdi, Michele Nappi, Jean-Luc Dugelay, and Yong Yu

Abstract

Smartphones are increasingly becoming a tool for ubiquitous access to a number of services including but not limited to e-commerce and home banking, and are more and more used for sensitive data storage. If on the one hand this makes the smartphone a powerful tool in our private and professional life, on the other it has brought about a series of new challenging security and privacy threats and raised the need to protect users and their data through new secure authentication protocols. In this article, we illustrate how the security level of a human authentication system increases from traditional systems based on the use of passwords or badges to modern systems based on biometrics. We have moved a step forward by conceiving an authentication protocol based on the combined recognition of human face and smartphone *fingerprint*. Thanks to image processing techniques, both the distinctive characteristics of the face and of the device that captured the face image can be extracted from a single photo or video frame and used for a double check of user identity. The fast-technological development of smartphones, allows performing sophisticated operations on the device itself. In the edge computing perspective, the burden of biometric recognition and source camera identification can be moved on the end user side.

Index Terms
Automatic authentication, multi-modal authentication, face recognition, source digital camera identification.

Introduction

Smartphones are by definition devices able to perform many of the functions of a computer. Their technology has a rapid development that is quickly overcoming the initial limits related to insufficient memory or low computational power. This has widened their use in daily life such as for email checking, messaging, and personal data storage (including private photos and passwords), but also in security-critical tasks, namely home banking operations, use of credit cards or other payment methods for online shopping, and remote access to workstations. The scenario described above has led to two consequences that are addressed in this article:

- The user and their smartphone are inseparable.
- Sensitive data and access to remote services must be protected.

The number of smartphone users worldwide is forecast to reach 2.1 billion in 2020 (from Statista - The portal for statistics, 2017). It is reported that in 2015 about eight-in-ten Americans used to shop online, 51% using a cellphone (source: Survey “Online Shopping and E-Commerce”, by Pew Research Center). In 2016, Kaspersky Security Network (KSN), estimated mobile banking attacks increase of 1.6 times, compared to 2015. Pew Research Center also reports (January 2016) that 28% of smartphone owners do not use a screen lock or other security features in order to access their phone or protect sensitive data stored on it. Finally, Acuity Market Intelligence has published its latest “Biometric Smartphone Update”, which reveals that the number of smartphones

Chiara Galdi and Jean-Luc Dugelay are with the Department of Digital Security, EURECOM, Sophia Antipolis, Biot, 06410 France
e-mail: {chiara.galdi, jean-luc.dugelay}@eurecom.fr

Michele Nappi is with the Department of Computer Science, Università degli Studi di Salerno, Fisciano, SA, 84084 Italy
e-mail: mnappi@unisa.it

Yong Yu is with the School of Computer Science, Shaanxi Normal University, Xi'an, 710062, China
e-mail: yuyong@snnu.edu.cn

incorporating biometrics has doubled since January 2016. These data define the scenario that have given rise to the proposed authentication protocol. On one hand, there is the ever-increasing need of secure authentication procedures and on the other, biometrics are spreading through smartphone applications. One important aspect addressed by the proposed protocol, is the ease of use. For convincing smartphone owners, including the ones that do not use any kind of security feature, it is important to design authentication protocols easy-to-use and as transparent as possible to the user. In the following sections, it is illustrated how the security and ease of use requirements are achieved by the proposed solution.

As mentioned before, the initial smartphone limits related to insufficient memory or low computational power are being overcome by fast-technological development. This allows performing sophisticated operations, including image and video processing on the device itself without requiring more demanding computation to be processed on the server side. In the edge computing perspective [1], the burden of biometric recognition and source camera identification can be moved on the end user side.

Kaspersky Lab Resource Center lists the "Top 7 Mobile Security Threats" on their website. Data leakage, unsecured Wi-Fi, and phishing attacks are part of them. These threats can be faced by a wise user behavior, such as avoid sharing personal information, always check the source, and use unique passwords. The threat addressed here, is the attempt to access the smartphone itself or a remote service by fooling the authentication system. Given the predominant role that the smartphone has assumed in our daily lives it is very unlikely to lend it to someone for a long period or do not immediately realize to have forgotten or lost it. Thus, it is a much more suitable *object* for user authentication compared to badges or keys, something that the user always brings with them. Existing techniques for authentication on smartphones include personal identification number (PIN), numeric password, pattern, and biometrics. The latter have been increasingly adopted on smartphones in recent years, but also often been fooled. The famous hacking of the Apple TouchID fingerprint scanner and then the Samsung Galaxy S8 iris scanner bypassed less than a month after it was shipped to public, have demonstrated the need of new and robust protocols for user authentication. The proposed authentication protocol combines the recognition of the user's smartphone with the recognition of the user based on their face. The user is only required to record a short video clip of their face. From that single clip, both face and device recognition is performed. The system work flow is illustrated in Fig. 1. Besides ensuring a higher level of security than using biometric recognition only, as detailed in the following section, the proposed system has several advantages:

1. Although the system consists of a double recognition, the acquisition process is very easy and fast.
2. The system is more robust to attacks since both the face features and the smartphone signature must be replicated to fool the system.

Given the pervasiveness of technology in our lives, just think of the so-called Internet of things (IoT), this approach can be further applied for fast and secure authentication for any kind of smart object [2].

This article is an extended version of the paper "Secure User Authentication on Smartphones via Sensor and Face Recognition on Short Video Clips"[3], previously presented at the 12th International Conference on Green, Pervasive and Cloud Computing (GPC 2017).

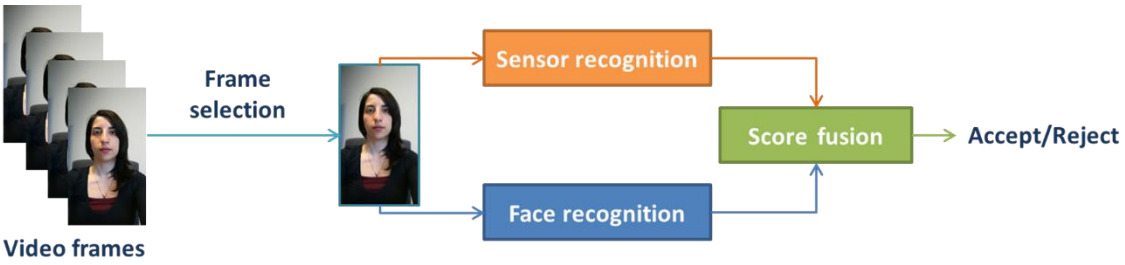

FIGURE 1 PROPOSED AUTHENTICATION SYSTEM WORK FLOW.

User Authentication

Authentication can be performed based on one or a combination of the following items [4]:

- Something the user knows (e.g., password, personal identification number (PIN), secret answer, pattern).
- Something the user has (e.g., smart card, ID card, security token, software token, smartphone).
- Something the user is or does (e.g. fingerprint, face, iris, gait).

The last are known as biometrics. As a premise, it is worth considering that passwords can be forgotten or snatched by malicious people, physical objects such as badges and ID documents can be lost or stolen. Biometrics can hardly be stolen and the process of falsification is much more complicated (e.g. plastic surgery). The most recent biometric recognition systems also embed mechanisms to recognize live biometrics (liveness detection) and fakes (presentation attack detection). If we consider all possible combinations of the three factors of authentication, we obtain the ranking, from lowest to highest security, illustrated in Figure 2 [4].

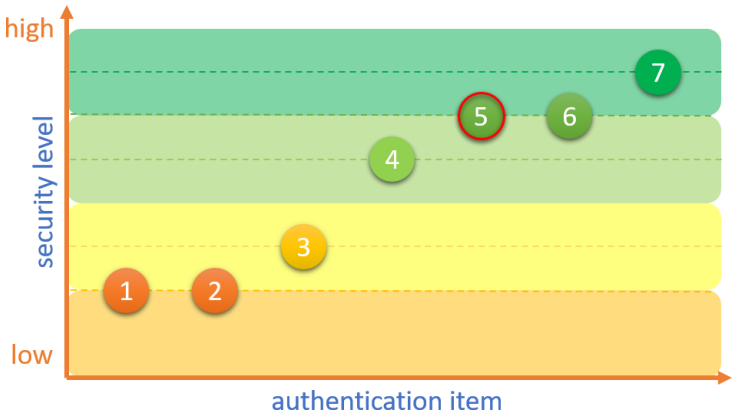

FIGURE 2 AUTHENTICATION SYSTEMS SECURITY LEVELS: (1) SOMETHING THE USER KNOWS; (2) SOMETHING THE USER HAS; (3) SOMETHING THE USER KNOWS + SOMETHING THE USER HAS; (4) SOMETHING THE USER IS OR DOES; (5) SOMETHING THE USER HAS + SOMETHING THE USER IS OR DOES; (6) SOMETHING THE USER KNOWS + SOMETHING THE USER IS OR DOES; (7) SOMETHING THE USER KNOWS + SOMETHING THE USER HAS + SOMETHING THE USER IS OR DOES.

Figure 2 plots the relative degrees of security. As mentioned before, the proposed system is of the type “Something the user has (smartphone) + something the user is (face)”, and assures a higher security level compared to the use of biometrics only.

Smartphone Recognition

The smartphone is identified based on the distinctive pattern, also called *camera fingerprint* or *camera signature*, left by its digital camera on the captured photos. That is why this technique is

also referred as *source digital camera identification*. Each imaging sensor has a noise pattern originated from imperfections during the manufacturing process and different sensitivity of pixels to light due to the inhomogeneity of silicon wafers of which the sensor is composed [5]. Even sensors of the same model can be distinguished by analyzing the sensor pattern noise (hereinafter SPN). The technique to extract and compare the SPN from an image has been first presented by Lukas et al. in [6] and further improved by Li in [5].

An image can be represented by its frequencies in the so-called frequency domain. Low frequencies correspond to homogeneous image regions while high frequencies describe the image details including edges but also the sensor noise. The SPN of a sensor is obtained by applying a de-noising filter in the wavelet domain to isolate the frequencies associated with the sensor noise. However, since both noise and scene details are located in high frequencies, it is observed that the SPN can be affected by the image content [5]. Li's approach, namely the enhanced sensor pattern noise (ESPN), is based on the idea that strong SPN components are more likely to be originated from the scene details and thus must be suppressed, while weak components should be enhanced. Figure 3 illustrates how the SPN is still contaminated by picture details (i.e. edges) while the ESPN is less affected.

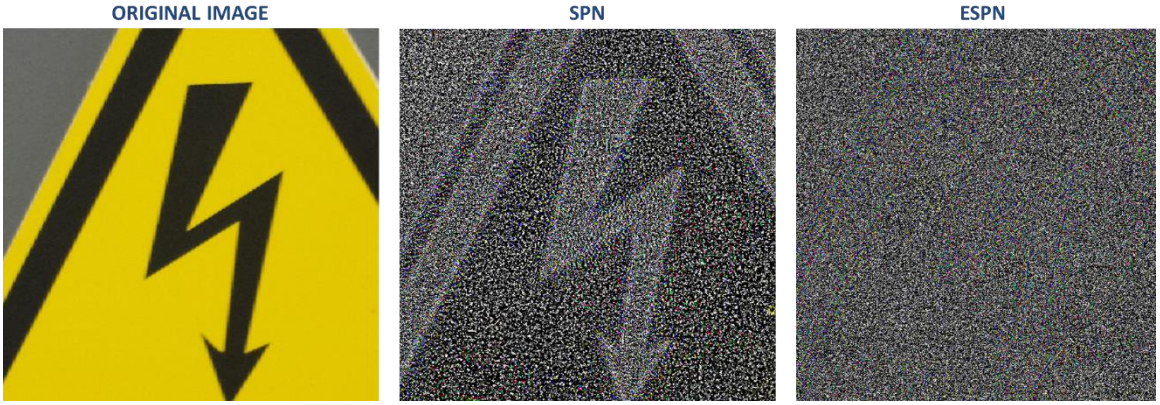

FIGURE 3 COMPARISON BETWEEN THE SPN (MIDDLE) AND THE ESPN (RIGHT) EXTRACTED FROM A PICTURE (LEFT).

Sensor Pattern Noise from Videos

It is known that videos are strongly compressed. The SPN comparison achieves optimal performances on still images but source digital camera identification from videos is much more challenging. The sensor noise pattern is strongly impacted by video compression, and it is demonstrated that SPN performance drastically drops [7]. The identification rate can be improved by selecting from the video only the I-frames for SPN estimation. A compressed video is made up of three kinds of frames: the *intra-coded picture* (I-frame), the *predicted picture* (P-frame), and the *bidirectional predictive picture* (B-frame). I-frames are the least compressible. An I-frame is a complete image, like a JPG image file. P and B frames hold only part of the image information (the part that changes between frames, e.g. moving objects). Thus, part of the SPN is lost. P-frames hold only the changes in the image from the previous frame. For example, in a scene where a person moves across a stationary background, only the person's movements are encoded. B-frames only store differences between the current frame and both the preceding and following frames. Therefore, the SPN is best preserved in I-frames. In the H.264/MPEG-4 AVC standard, the granularity is brought down to the "slice level." A slice is a spatially distinct region of a frame that is encoded separately from any other region in the same frame. I-slices, P-slices, and B-slices take the place of I, P, and B frames [8]. In [9] it is shown how performance improves by selecting only I-frames, or a weighted combination of I-frames and P-frames. In [10], Chen et al. propose a

technique for determining whether two video clips came from the same camcorder by mean of the Maximum Likelihood Estimator for estimating the SPN, and of normalized cross-correlation for SPN comparison. Other factors can affect the SPN, for example video stabilization and the additional video compression operated by some website when uploading a video [11]. The latter is a major issue since videos with criminal content are often posted on line on social networks or web platforms for video sharing and the additional compression steps make more difficult to associate the video to the source digital camera.

Face Recognition

Face recognition, and biometric recognition in general, consists in compactly representing the features of the face. This representation is also known as biometric template. The method we adopted is based on the histogram of oriented gradients (HOG) [12]. The idea behind this technique is that object appearance and shape can be represented by the distribution of local intensity gradients (i.e. a directional change in the intensity or color) or edge directions, even without precise knowledge of the corresponding gradient or edge positions. The resulting HOG descriptors are then used as input of a conventional support vector machine (SVM) based classifier.

Experimental Results

The proposed protocol, of which the workflow is illustrated in Fig. 1, requires in input a short video clip depicting the user face. A single I-frame is selected, it can be chosen according to many criteria such us image quality in terms of focusing. The frame is processed by two modules that can work independently, namely the face recognition module and the source digital camera identifier. Each module provides a score indicating how likely is that the input image comes from the authorized user. In the following sections, the performance of single modules and of their fusion are presented. This section also describes the employed dataset. Performances are assessed in terms of equal error rate (EER), recognition rate (RR), cumulative match score curve (CMS), receiver operating characteristic curve (ROC), and area under ROC curve (AUC).

Database

The experiments could be carried out thanks to the publicly available database for Source Camera REcognition on Smartphones, namely the SOCRatES database. SOCRatES is currently made up of about 6.200 images and 670 videos captured with 67 different smartphones of 13 different brands and about 40 different models. This database has been particularly designed for developing and benchmarking of source digital camera identification techniques. SOCRatES is freely available to other researchers for scientific purposes[1].

Source camera recognition performance evaluation

When the ESPN is extracted from an image, it is compared with the Reference Sensor Pattern Noise (RSPN) of a digital camera. For each sensor, the RSPN has been estimated by averaging the SPN from the I-frames of 9 out of 10 videos (the 10th video has been used as test sample).

A video contains several I-frames and in this first experiment we have tested which I-frame yields to best performances. As mentioned before, SOCRatES contains short videos of about 2-5 seconds. In these short clips, only few I-frames are present. The RSPN has been computed over N frames (N = 36 if the videos are long enough to have 4 I-frames each) extracted from 9 out of 10 clips. The

[1] http://socrates.eurecom.fr/

10th video has been used as test sample and its ESPN has been extracted using Li's technique on its I-frames. In the graphs illustrated in Fig. 4, a comparative performance evaluation shows that using the first I-frame of the test clip, leads to better recognition performances compared to the use of the second or third I-frame. Performances have been assessed on a pool of 630 videos recorded by 63 different smartphones. The performance values are summarized in the following:

- First I-frame: EER = 0.3013; RR = 0.3492; AUC = 0.7717
- Second I-frame: EER = 0.3322; RR = 0.2857; AUC = 0.7339
- Third I-frame: EER = 0.3360; RR = 0.2903; AUC = 0.7330

Compared to the performances reported in [7] obtained on the same dataset but without I-frame selection, an improvement of around 7% of the rate of correct classification has been obtained.

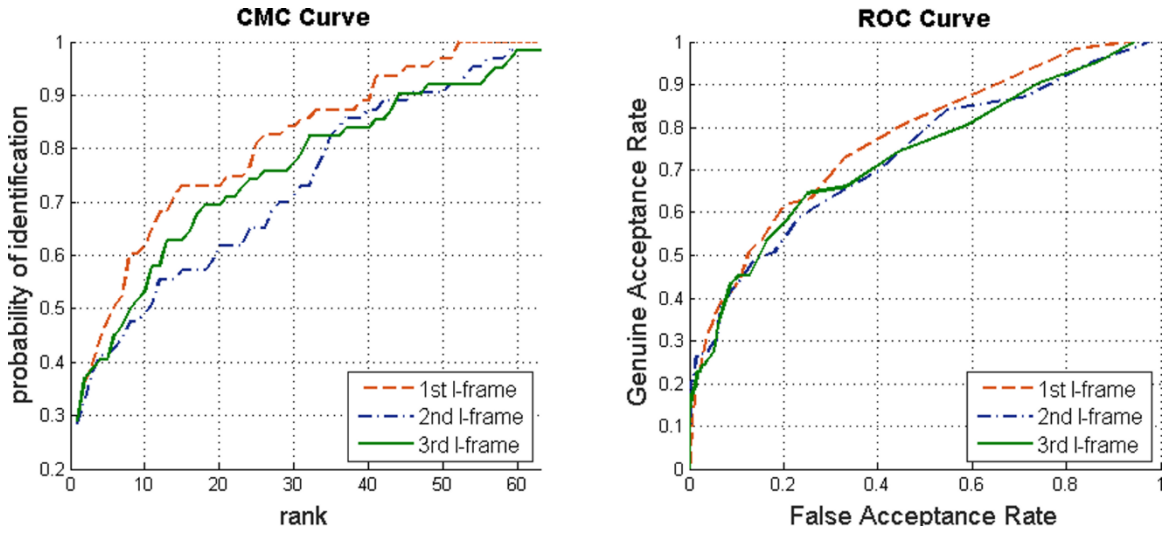

FIGURE 4 SENSOR RECOGNITION PERFORMANCES COMPARISON WHEN USING THE 1ST, 2ND, OR 3RD I-FRAME FOR SPN EXTRACTION.

Face recognition performance evaluation

For what concerns face recognition, 10 face pictures for each user have been collected. Eight out of 10 images have been used to train a SVM on the extracted HOG features. One picture, randomly selected from the 2 remaining ones, has been used as test sample. The performances obtained are: EER = 0.07; RR = 0.80; AUC = 0.97.

Score fusion performance evaluation

The videos collected in the SOCRatES database do not portray face images. In order to simulate system performances, face pictures collected with the same devices that recorded the videos have been used. A total of 59 pairs device-face have been then defined. A sample is thus genuine if the combination device-person is enrolled in the system.

Fusion is performed at score level. The modules contribute equally to the final score since they represent two different entities, that is the smartphone and the user, and have the same importance in the computation of the final score. Alternatively, a voting procedure could be used for the final accept/reject decision.

The system performances after fusion are: EER = 0.06; RR = 0.83; AUC = 0.97.

Fig. 5 reports the performances graph for the aforementioned experiments.

It is worth mentioning that even if the source camera identifier gives relatively small contribution to the fused authentication score, the contribution in terms of security of the system is great. Let us recall that the system performs a double check of user identity, based of two distinct and

independent entities. Fooling the system is much more complicated since both the face and camera fingerprint must be replicated. Compared to systems based on biometrics only, the ease of use is the same, since the camera recognition is completely transparent to the user, but the security assured is much higher.

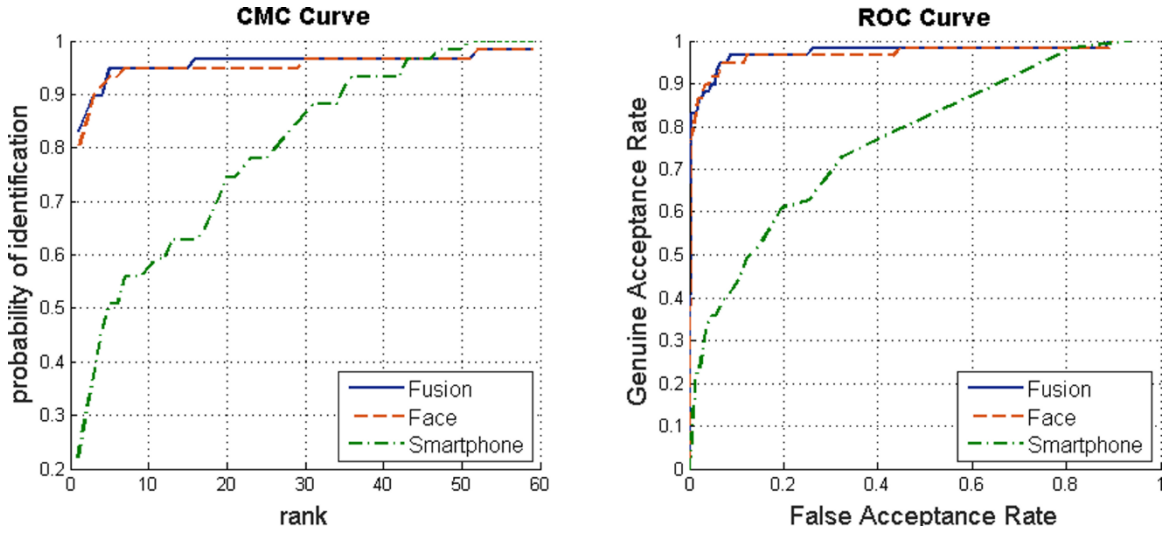

FIGURE 5 SINGLE FEATURES AND FUSION PERFORMANCE COMPARISON.

Score normalization is a necessary step when combining different recognition modules. The algorithms employed can generate scores that are different in terms of distribution and numerical range. In the past, several different methods of score normalization have been proposed, addressing different issues that can emerge during the fusion process. In our experiments, we tested five normalization techniques, namely: Max-Min, Z-score, Median/MAD, TanH, and Sigmoidal. Please refer to the following article for reference: [13]. As illustrated in Fig. 6, Median/MAD is outperformed by all other techniques with Z-score and TanH performing slightly better (they both achieved a slightly lower EER) than the others.

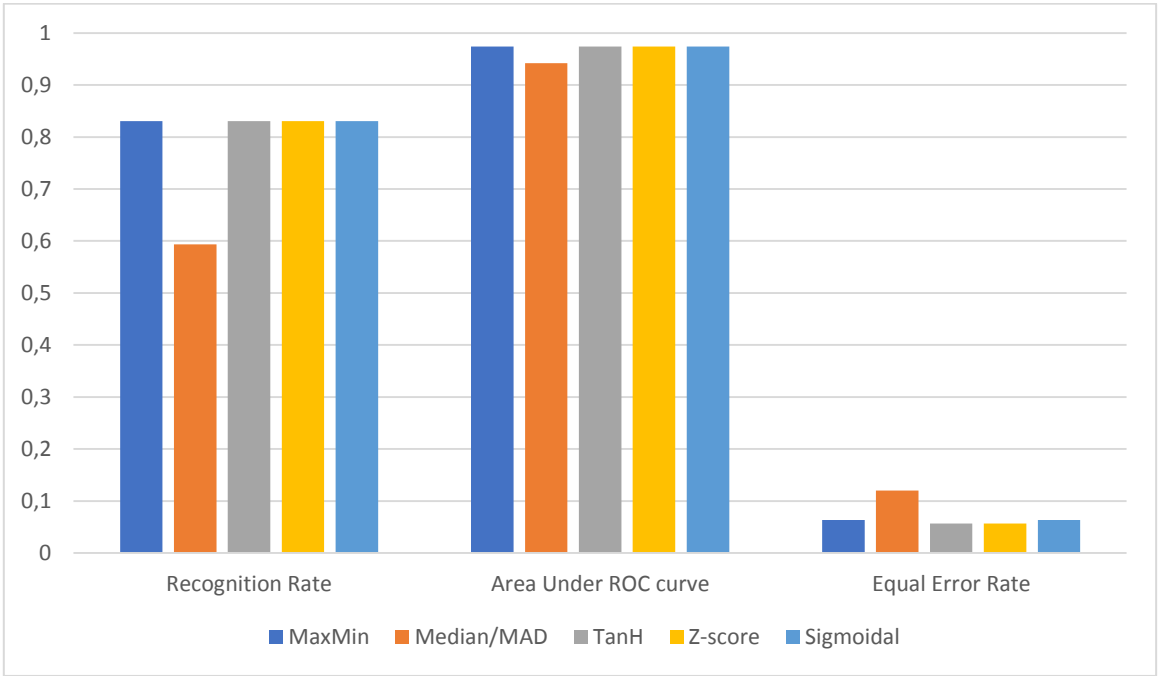

FIGURE 6 COMPARISON OF DIFFERENT TECHNIQUES FOR SCORE NORMALIZATION

The average computational time for face recognition on smartphone has been evaluated on a Samsung Galaxy S6 smartphone: face detection plus face template creation takes around 400 milliseconds; while template matching takes around 30 milliseconds. Concerning the smartphone recognition module, the heaviest operation in terms of computational time is the reference SPN computation. However, this operation is performed only once, when enrolling the smartphone. On the other side, the SPN comparison is the fastest operation since it consists in simple correlation. Further details about the computational time will be made available with the development of a complete system prototype.

Conclusions

The need of protecting smartphone users' sensitive data and access to remote services led us to conceive an innovative authentication protocol. To the best of our knowledge, this is the first work proposing the combination of source camera identification and face recognition for real-time user authentication from videos. The authors have previously presented a system combining iris and sensor recognition from still images in [14]. Here, the use of videos as input data presents a considerable challenge since the SPN is significantly affected by strong video compression. However, by simply selecting I-frames from a set of short video clips, a rate of correct classification equal to 77% is obtained by the source camera recognition module and a rate of 97% is achieved by the combination with the face module.

When dealing with biometric recognition, a question arises about privacy protection. How to protect sensitive data, such as the face picture, used for authentication? The solution mostly adopted is to never store the original picture/biometric sample, but only circulate its compact representation, namely the *template*. In addition, the template must never be externally visible decrypted.

The proposed protocol assures a more secure authentication by combining different authentication items, namely the user's face and their smartphone. Also, the acquisition process is simple and fast. Further improvement of the proposed solution, include but are not limited to the combination with other biometric traits, such as voice, or the integration of presentation attack or liveness detectors [15].

The method has been tested on a large database of videos collected with 63 different smartphones, namely the SOCRatES database.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2017YFB0802003, 2017YFB0802004), National Cryptography Development Fund during the 13th Five-year Plan Period (MMJJ20170216) and the Fundamental Research Funds for the Central Universities (GK201702004). The work is also partially supported by the EU Horizon 2020 research and innovation programme under the projects PROTECT (grant agreement 700259) and IDENTITY (grant agreement 690907).

References

[1] B. P. Rimal, D. P. Van, and M. Maier, "Mobile Edge Computing Empowered Fiber-Wireless Access Networks in the 5G Era," in IEEE Communications Magazine, vol. 55, no. 2, February 2017, pp. 192-200. doi: 10.1109/MCOM.2017.1600156CM

[2] M. Shahzad and M. P. Singh, "Continuous Authentication and Authorization for the Internet of Things," in IEEE Internet Computing, vol. 21, no. 2, Mar.-Apr. 2017, pp. 86-90. doi: 10.1109/MIC.2017.33

[3] C. Galdi, M. Nappi, and J.-L. Dugelay, "Secure User Authentication on Smartphones via Sensor and Face Recognition on Short Video Clips," in International Conference on Green, Pervasive, and Cloud Computing, May 2017, pp. 15-22.

[4] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," in Proceedings of the IEEE, vol. 91, no. 12, Dec 2003, pp. 2021-2040. doi: 10.1109/JPROC.2003.819611

[5] C. T. Li, “Source camera identification using enhanced sensor pattern noise,” in IEEE Trans-actions on Information Forensics and Security 5(2), 2010, pp. 280-287.
[6] J. Lukas, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise,” in IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, June 2006, pp. 205-214.
[7] C. Galdi, F. Hartung, and J.-L. Dugelay, “Videos versus still images: Asymmetric sensor pattern noise comparison on mobile phones,” in Electronic Imaging, Media Watermarking, Security, and Forensics 2017, 2017, pp. 100-103(4).
[8] G. J. Sullivan and T. Wiegand, "Video Compression - From Concepts to the H.264/AVC Standard," in Proceedings of the IEEE, vol. 93, no. 1, Jan. 2005, pp. 18-31.
[9] W. H. Chuang, H. Su, and M. Wu, “Exploring compression effects for improved source camera identification using strongly compressed video,” in 18th IEEE International Conference on Image Processing, Brussels, 2011, pp. 1953-1956. doi: 10.1109/ICIP.2011.6115855
[10] M. Chen, J. J. Fridrich, M. Goljan, and J. Lukas, “Source digital camcorder identification using sensor photo response non-uniformity,” in Electronic Imaging 2007, SPIE Proceedings Vol. 6505, Security, Steganography, and Watermarking of Multimedia Contents IX, 2007, pp. 65051G-65051G.
[11] W. Van Houten and Z. Geradts, “Using sensor noise to identify low resolution compressed videos from YouTube,” in International Workshop on Computational Forensics, 2009, pp. 104-115.
[12] N. Dala and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, June 2005, pp. 886-893.
[13] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” in Pattern Recognition, Volume 38, Issue 12, December 2005, pp. 2270-2285. doi: 10.1016/j.patcog.2005.01.012
[14] C. Galdi, M. Nappi, and J.-L. Dugelay, “Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity,” in Pattern Recognition Letters, Volume 82, Part 2, 15 October 2016, pp. 144-153. doi: 10.1016/j.patrec.2015.09.009
[15] M. De Marsico, C. Galdi, M. Nappi, and D. Riccio, “Firme: Face and iris recognition for mobile engagement,” in Image and Vision Computing, 32(12), 2014, pp. 1161-1172.

Biographies

Chiara Galdi (chiara.galdi@eurecom.fr) received her joint supervision Ph.D. in Economy from the University of Salerno and in Signal and Image processing from Telecom ParisTech in 2016. She is now a postdoctoral fellow at EURECOM. She has been awarded with the Ph.D Thesis Award 2015-2016 assigned by the IEEE Italy Section Biometrics Council Chapter. In September 2016, she obtained the European Biometrics Industry Award for her work titled "Combining Iris and Sensor Recognition on Mobile Phones".

Michele Nappi [SM] (mnappi@unisa.it) received the Ph.D. degree in Applied Mathematics and Computer Science from the University of Padova, Italy, in 1997. He is currently an associate professor at the University of Salerno. He is Senior Member of IEEE, IAPR Member and team leader of the Biometric and Image Processing Lab (BIPLAB). He received several international awards for scientific and research activities and is the President of the Italian Chapter of the IEEE Biometrics Council.

Jean-Luc Dugelay [F] (jean-luc.dugelay@eurecom.fr) obtained his Ph.D. from the University of Rennes in 1992. His thesis work on 3DTV was undertaken at Orange Labs. He then joined EURECOM Sophia Antipolis where he is now Professor in Imaging Security. He is a fellow member of IEEE. He is a o-author of several conference articles that received an award. He is the founding Editor-in-Chief of the EURASIP Journal on Image and Video Processing (JIVP).

Yong Yu (yuyong@snnu.edu.cn) is currently a Professor of Shaanxi Normal University, China. He holds the prestigious one hundred talent Professorship of Shaanxi Province as well. His research interests are digital signatures and secure cloud computing.

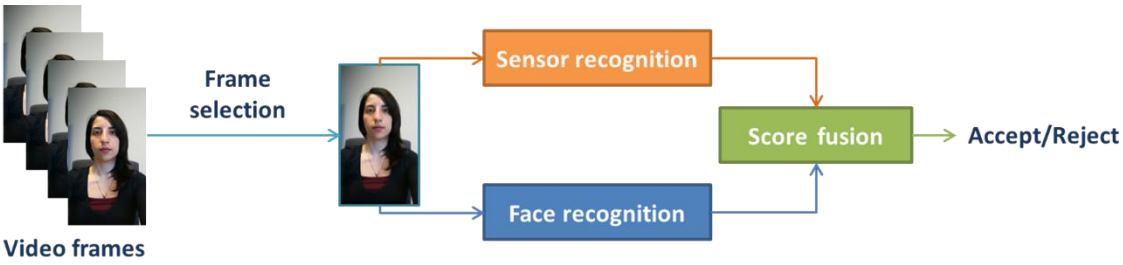

FIGURE 1 PROPOSED AUTHENTICATION SYSTEM WORK FLOW.

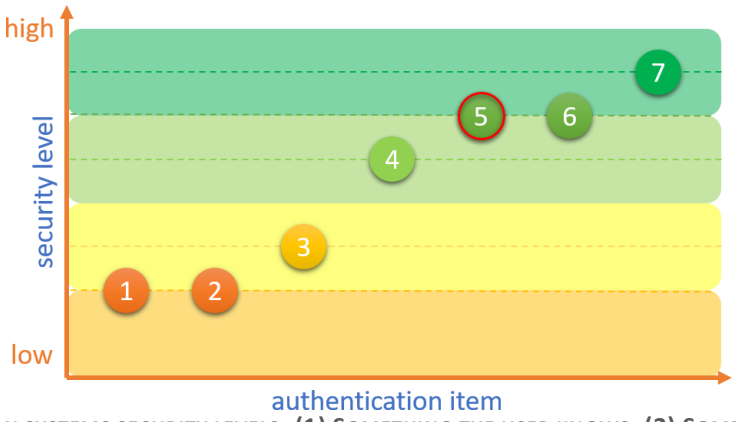

FIGURE 2 AUTHENTICATION SYSTEMS SECURITY LEVELS: (1) SOMETHING THE USER KNOWS; (2) SOMETHING THE USER HAS; (3) SOMETHING THE USER KNOWS + SOMETHING THE USER HAS; (4) SOMETHING THE USER IS OR DOES; (5) SOMETHING THE USER HAS + SOMETHING THE USER IS OR DOES; (6) SOMETHING THE USER KNOWS + SOMETHING THE USER IS OR DOES; (7) SOMETHING THE USER KNOWS + SOMETHING THE USER HAS + SOMETHING THE USER IS OR DOES.

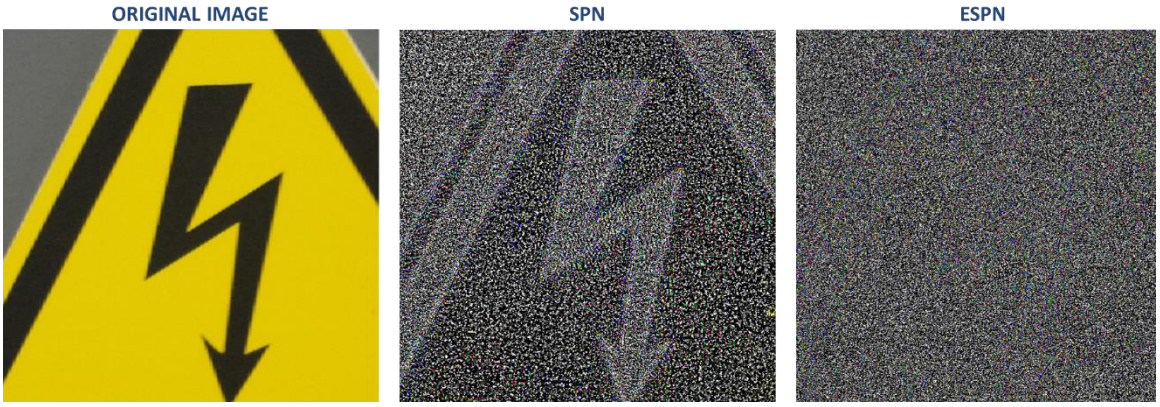

FIGURE 3 COMPARISON BETWEEN THE SPN (MIDDLE) AND THE ESPN (RIGHT) EXTRACTED FROM A PICTURE (LEFT).

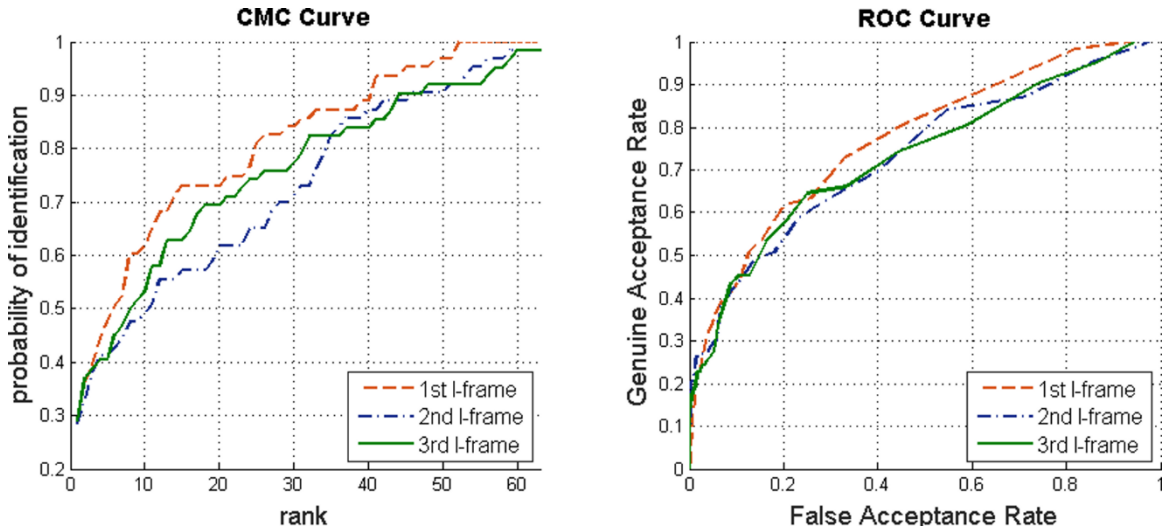

FIGURE 4 SENSOR RECOGNITION PERFORMANCES COMPARISON WHEN USING THE 1ST, 2ND, OR 3RD I-FRAME FOR SPN EXTRACTION.

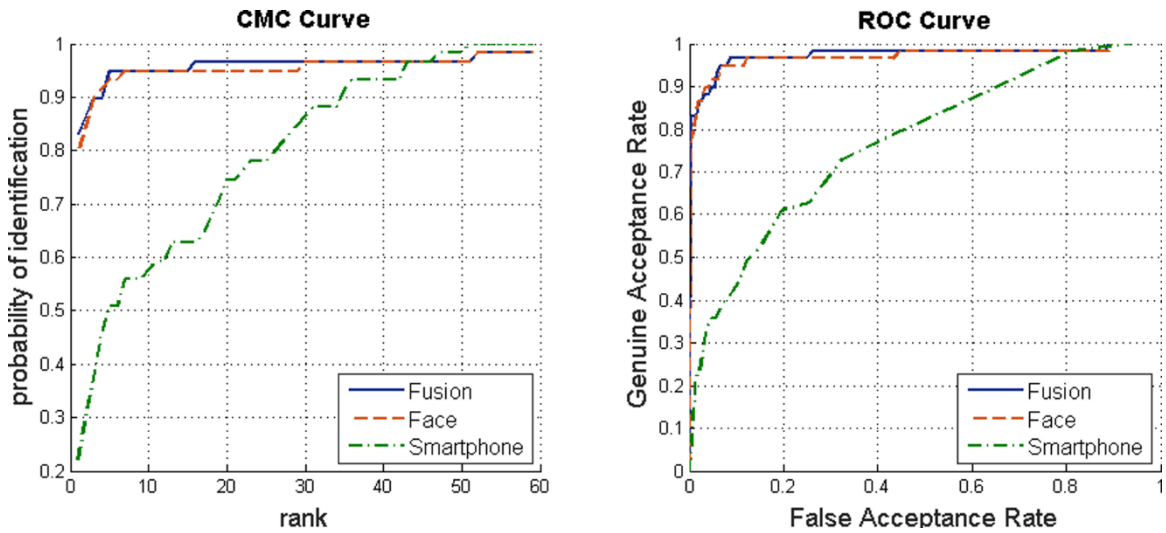

FIGURE 5 SINGLE FEATURES AND FUSION PERFORMANCE COMPARISON.

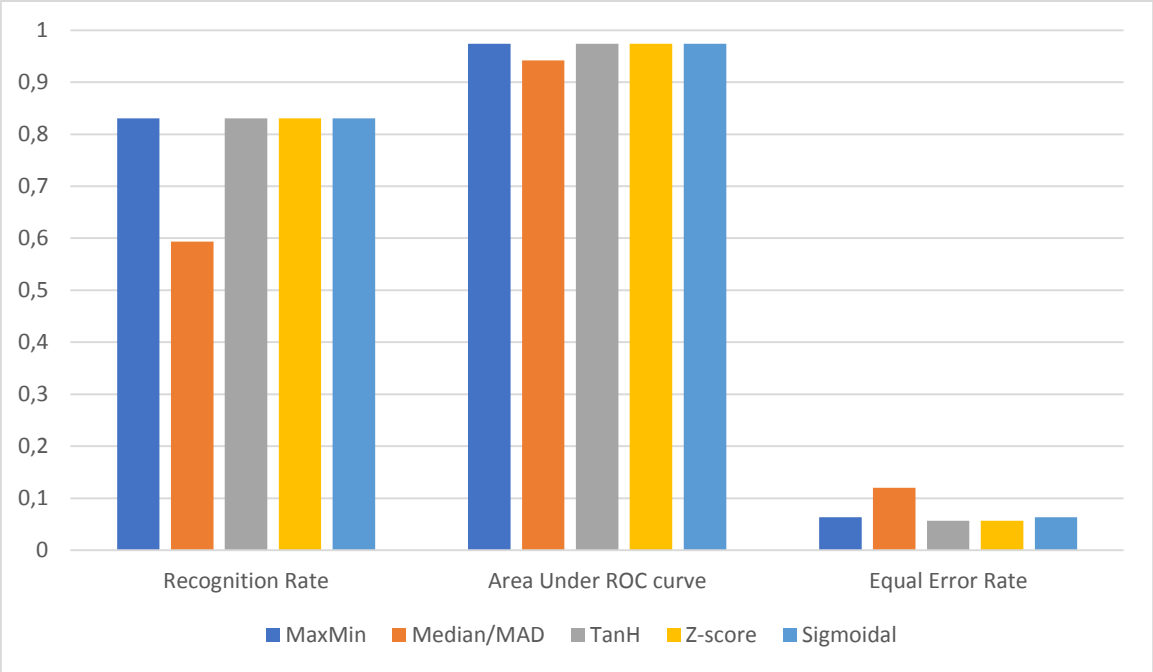

FIGURE 6 COMPARISON OF DIFFERENT TECHNIQUES FOR SCORE NORMALIZATION