Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 135 (2018) 147–154

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Computer Science and Computational Intelligence 2018

Faceture ID: face and hand gesture multi-factor authentication using deep learning

Earl Ryan M. Aleluya[a]*, Celesamae T. Vicente[b]

[a,b]*Mindanao State University – Iligan Institute of Technology, 9200 Iligan City, Philippines*

Abstract

Access control is the act of providing privacy to a resource, and authentication through a single factor is no longer reliable to provide robust protection against unauthorized access. Hence, there is a rapid growth of exploring novel multi-factor authentication (MFA) methods which combine two or more authentication factors– inherence, possession, and knowledge. Despite the increasing use of MFA, to the best of authors' knowledge, none have so far explored the combination of face, one-time password (OTP) and hand gesture in MFA. Thus, this study produces a proof-of-concept of this combination to form a new authentication method (Faceture ID). Furthermore, this study highlights three contributions: i) face verification with single-sample gallery set using pre-trained Deep Convolutional Neural Network, ii) handwriting gesture recognition using Leap Motion controller for tracking motion and Convolutional Neural Network for classification, and lastly, iii) a new MFA method utilizing face, OTP and hand gesture. The experimental results on the face verification show an average false acceptance rate of 1.94% with average genuine acceptance rate of 67.7%, from the self-built database where the facial images are exposed to variations in pose, expression, and occlusion. In addition, the classifier for the handwriting gesture recognition can predict gestures at about 96% for both precision and recall. Furthermore, the proposed MFA provides a novel systematic approach, high accuracy and performance with the intent to contribute on strengthening the security on privacy of resources against identity theft and attacks.

Keywords: multi-factor authentication; deep learning; convolutional neural network; face verification; leap motion; gesture recognition;

* Corresponding author. Tel.: +639-09-729-4834.
E-mail address: earlryan.aleluya@g.msuiit.edu.ph

10.1016/j.procs.2018.08.160

1. Introduction

In the light of security incidents, the single factor authentication (SFA) is no longer reliable to provide robust protection against unauthorized access [1]. A report from Breach Level Index recounts 974 data breaches occur from January to June of the year 2016 where 64% of them constitute to identity attacks [2]. Traditional SFA implemented either password-based security mechanism or biometrics that has security drawbacks. Passwords, the most common authentication factor, has suffered from theft attacks, password cracking and susceptibility to phishing [3]. In the same way, biometrics alone does not guarantee safety from unauthorized users, raises privacy concerns and is convenient only for limited applications since the system becomes very slow for a large number of users [4].

Hence, the two-factor authentication (2FA) has been introduced in order to enhance security in authentication systems by extending the SFA. In 2FA, the user provides dual means of identification to gain access to specific resources: one of which is typically a security code (something the user knows) and the second factor (something the user has) such as a card, to reduce the possibility of online identity theft. Although 2FA may seem promising to be the cure for securing networks and resources, there are many security holes that this type of authentication will not protect against and still vulnerable to man-in-the-middle attacks [5].

With this, there is a growing development of novel multi-factor authentication (MFA) methods to establish secure and authorized communication between a user and server over an insecure channel [6]. This scheme combines authentication techniques belonging to different factors which can be categorized into three groups– inherence (something you are), possession factor (something you have) and knowledge (something you know) [7]. It can be implemented in the following combinations – knowledge and possession, possession and inherence, inherence and knowledge, or all authentication factors. MFA are actively relevant in mobile environment, remote authentication, multi-server environment, wireless sensor networks, cloud computing, web applications, banking and commerce, session initiation protocol, continuous authentication and many more [8].

The less sophisticated MFA system is designed based on the combination of knowledge and possession factors. One-time password (OTP) is a practical authentication service to prevent password phishing attacks [9] which eliminates the need to preset user passwords during the authentication process while leveraging the security through sending OTPs from the server to the mobile devices. This setup combines the static password (knowledge) and the OTP (possession). Another combination is using both inherence and possession factors that requires a biometric data and provides an OTP. In [10], a handwritten signature using the mouse movement (inherence) with the OTP (possession) generated from an open source Google authenticator. Their process eliminates the cost involving digital signature authenticity for remote authentication. Cascading inherence and knowledge factors, which requires a biometric data and a static password, is also one method in MFA. In [11], a behavioral handwritten signature (inherence) with the static password (knowledge) achieves a high security on a cloud-based application that addresses the challenge in modern hand-held devices to ensure the privacy of the users.

Although the combination of only two authentication factors has been well-studied for means of controlling access, a systematic literature introduced various issues and challenges in 2FA algorithms [8] that purports several applications to use all factors to leverage the security of the verification process. One prototype incorporates fingerprint (inherence), tap sequence (knowledge) and near-field communication (NFC) tag to the reader in the smartphone, together with an NFC sticker, keychain, or NFC-enabled wearable device [12]. Various biometric signals are also being actively studied: one recent trend is the use of multimodal data for achieving high reliability [13], yet it generally requires multiple sensors, which result in high developmental costs. This gap was addressed by Hyunsoek Choi and Hyeyoung Park who proposed a novel multimodal biometric system that used two heterogeneous biometric signals obtained from a single vision sensor: facial image and gesture video [14]. However, multimodal biometric system used a parallel structure to capture both the two factors which make the system dependent on the performance of both verification systems.

In the milieu of increasing reliance on MFA, to the best of author's knowledge, none have so far integrated the facial biometrics, OTP, and hand gesture. Thus, this present study introduces "Faceture ID", a proof-of-concept of a new MFA that cascades face verification, OTP messaging and hand gesture recognition. The system uses image classification technique to recognize the face and handwriting gestures. Along addressing the limitation of the multimodal biometric system, instead of using parallel verification system, this study uses authentication factors sequentially. In this way, if the first verification process concludes an unauthorized user, the authentication will not proceed to the next verification process, thereby, preventing longer queuing time. Moreover, this study attends to improve the performance of the face verification system using deep learning method by adapting a pre-trained FaceNet model wherein this model applies Convolutional Neural Network. The proposed system used Leap Motion Controller (LMC) on hand gesture since it is inexpensive and less sophisticated than Kinect sensor, has good detection performance and accurate [15]. The proposed system is designed to be implemented in facilities requiring high security with less population, unlike sensors for mobile phones are designed for portability. In addition, it is introduced to overcome the drawbacks of the traditional authentication system of its reliability and performance.

This paper is further organized as follows: Section 2 is devoted to reviewing the latest approaches to face verification and hand gesture recognition. The proposed method is explained in Section 3, while Section 4 presents the experimental results. Finally, the conclusion is made in Section 5.

2. Related works

This section discusses the existing methods on face recognition and dynamic handwriting gesture recognition.

2.1. Methods of face verification system

Face recognition is a biometric application in which the system either classifies human identity according to the face (face identification) or verifies whether two images belong to the same subject (face verification) [16]. Face recognition can be clustered based on the methods on how discriminative features are obtained and how they are used for classification.

Image-set-to-single-image face recognition (SIFR) is the framework that uses many facial samples per person during training and evaluates a single facial image. Although, image-set-to-image-set face recognition (SSFR) is different from conventional SIFR since there is only one facial image in the probe set. SSFR assumes the training set and the testing set to have multiple sample images for each person [17]. In 2016, the FaceNet garnered a remarkable accuracy of 99.63% in the LFW dataset [18]. With the result, SSFR through deep learning models has become an active research nowadays.

The limitation of the availability of facial images for every person produces a new challenge in face recognition, and this challenge is coined as single sample per person problem (SSPP). Thus, the single-image-to-single-image face recognition (IIFR) uses one sample per person for the training and probing. Unlike IIFR, the single-image-to-image-set face recognition (ISFR) does not ignore the collection of the multiple probe samples. In which most real-world applications, the probe samples usually can be captured on the spot easily while there is usually a single register for each person on the database [17].

The SSPP problem is quite common in practical face-related applications. Several IIFR and ISFR methods present approaches to learn discriminative information from the single image in the training set. Thus, this study aims to develop a face verification system on top a pre-trained deep learning model (which is an SSFR) to address the SSPP problem by obtaining a high-level representation of the single image in the gallery set.

2.2. Methods of handwriting gesture recognition

Due to the advantages of convenience and naturalness, hand gestures have been widely used for mid-air interaction between the machine and the human such as in virtual reality and intelligent robot control. The difficulty in recognizing dynamic hand gestures lies on what sensor to use for detection and what model for classification [19].

Data glove-based and visual-based methods are the two approaches to acquire hand motion. Data gloves are wearable devices embedded with gyroscopes, accelerometers and other physical sensors to obtain 3D information. On the other hand, Microsoft Kinect and Leap Motion controller are visual-based hardware which rely on the captured depth images to extract the positions of the hands, joints, and fingers [20].

The challenge in dynamic handwriting gesture recognition is not only on the classification but must also address the problems in detection, tracking and feature extraction. The Leap Motion controller (Figure 1) detects and tracks the 3D coordinates of the hands' frame by frame directly. This simplified tracking motivates this study to use LMC for detection. The 3D information of the hand is collected to form a time series. However, according to [20], time series prediction is unstable and limited. Thus, this study represents each dynamic gesture as an image containing the trajectory.

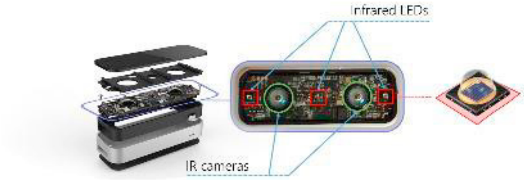

Fig. 1. Components of the Leap Motion controller.

The proposed system used LMC since the performance of the said hardware, which is not just a portable device, is remarkable as evaluated by several researchers. In 2013, Weichert, Bachmann, Rudak, and Fisseler stated that the localization precision of the sensor is within 0.2-millimeter deviation [21] . In 2014, the experiments from Guna, Jakus, Pogacnik, Tomazic, and Sodnik concluded that the sensor is reliable and accurate for tracking points [22]. Lastly, Bachmann, Weichert, and Rinkenauer pointed out that the sensor has 7.2% error rate for detecting hands and fingers, making it plausible to be a pointing device [23]. Since the sensor is well-evaluated, there is a recent effort from researchers in this field to establish a robust recognizer of dynamic hand gestures using the sensor [15 20]. In terms of its usage, there are many motion sensing devices available in the marketplace. The Leap Motion controller was chosen for this project because of its accuracy and low price. Unlike the Kinect, which is a full body sensing device, the Leap Motion controller specifically captures the movements of a human hand, albeit using similar IR camera technology [15].

3. System architecture

The proposed "Faceture ID" has the system architecture as designed below in Figure 2. It has two stages – *registration* and *authentication*. The *registration* stage refers to the task of populating the database with the authorized person's name, phone number, and facial image. Meanwhile, the *authentication* stage refers to the login process and it composes of sequential verification. First, the facial image of the person is captured using a camera and then being compared to the samples inside the database. Where there is a match from the templates, the authentication process continues. Otherwise, the person is rejected accordingly. Afterward, the system generates an OTP and sends it to the person's mobile phone. Once received, the person writes the OTP on mid-air within the coverage of the Leap Motion controller. The person will be granted access when the handwriting gestures are correctly supplied.

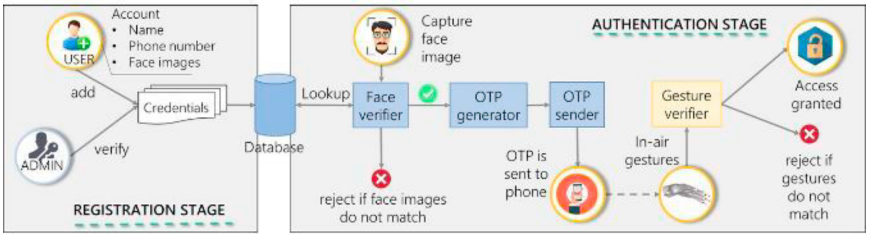

Fig. 2. System architecture of the proposed "Faceture ID".

The following subsections discuss the architecture of the individual systems – the face recognition, OTP messaging and gesture recognition.

3.1. Face verification system

Each image in the gallery set and probe set undergo same pre-processing to output high-level representation and are discussed as follows:

The *down-sampling* is the step to scale down the resolution of the original captured image which decreases the processing power for future image manipulation. The *face detection* is the step wherein the faces are localized using the OpenFace "dlib" frontal face detector API. The OpenFace library is an open-source implementation of FaceNet [24]. In the *landmark detection* step, there are 68 face landmarks per detected face that describe the shape of the face. Lastly, each face is represented by a vector with 128 features in the *representation* step and it contains the ratio of distances, areas, and angles of the localized landmarks.

When verifying a person, the vector describing the probe image must be similar to the ones saved in the gallery set (database). The similarity between two vectors is computed through magnitude (Euclidean distance) and direction (cosine similarity). Mathematically, the formulas are:

$$distance = \|F_{gallery}.F_{probe}\| = \sqrt{\sum_{i=1}^{128}(f_i^{gallery} - f_i^{probe})^2} \tag{1}$$

and,

$$similarity(F_{gallery}, F_{probe}) = \cos\theta = \frac{F_{gallery} \,.\, F_{probe}}{\|F_{gallery}\|.\,\|F_{probe}\|} \tag{2}$$

3.2. Instant messaging application

Once the system verifies the authorized person, a one-time password is sent to his phone number. The password is composed of six random digits which are generated randomly from the range between 100000 and 999999. The system application connects to the Twilio cloud service to request for sending an outbound SMS. Twilio is a cloud service especially for communication purposes such as chat, call, and SMS.

3.3. Gesture recognition system

Once the OTP is received, the person handwrites the password to perform mid-air interaction using Leap Motion controller. The system architecture is described as follows:

The *pre-processing* module accepts a handwriting gesture which is represented as a series of image frames that are captured from the Leap Motion controller. The application service of the sensor provides directly the 3D information of the index finger and the z-coordinate of this finger is neglected in the process. The *thresholding* module determines the start and end of the trajectory using the speed at the tip of the index finger. When the speed exceeds 10 mm/s, the finger is considered moving, otherwise, it is the end of the gesture. The *conversion* module aims to convert the trajectory of the finger into an image-like canvas using the OpenCV library. This module extracts the region-of-interest and resizes the canvas into a resolution of 28x28 matrix. In *classification* module, a self-constructed convolutional neural network with 7 layers including the input and output layers. The hidden layers consist of two combinations of convolution and pooling layers, plus a fully connected layer.

The digit is classified one at a time. If the digits match the OTP, the person is accepted and is granted with access.

4. Experimental results

This section enumerates the evaluation of individual systems and the fully-integrated system. The experiment is conducted in a Lenovo U31-70 computer with Intel Core i7 CPU and 8 GBs of RAM.

4.1. Evaluation on proposed face verification system

Based on the system architecture of the face verification, there is a need to choose threshold values of Euclidean distance and cosine similarity, and in this work, 0.60 and 0.93 are selected respectively. The values are obtained based on heuristic observation on the different combinations of the two parameters.

Twenty university students composing of 10 males and 10 females are invited to participate in the gathering of facial images. A "Nikon" SLR camera is used to record videos from the participants. Each participant moved his head to different pose in a sequence as follows – front, upward, downward, left, right, counter-clockwise and clockwise directions. For various expression, the participants expressed different emotions like happy with a close and open smile, sad, excited, angry and shocked. Lastly, for occlusion, each participant wears one accessory at a time, either eyeglass, sunglass or a cap.

To evaluate the system, two scenarios are experimented – *condition positive* and *condition negative*. In a *condition positive* scenario, the probe images from the recorded videos are cross-referenced to the images of authorized people in the gallery set. While in a *condition negative* scenario, the system is assumed to be scrutinized by unauthorized people through cross-referencing the images from the videos of a specific person to the gallery set which omits the existence of his sample image. Table 1 summarizes the classification reports on the evaluation in *condition positive* scenario. Having average precisions above 90% is an advantage to the system since it means that there is a minimal chance that an authorized person is misclassified as someone else in the dataset. Moreover, the system predicts above 77% accuracy even in different pose, various expressions, and even wearing an eyeglass. However, the accuracy lowers when a person wears sunglass or a cap because some landmarks are not visible.

Table 1. Classification report for face verification on condition positive scenario.

Challenge	Images per person	Average precision	Average recall	Average f-score
Various pose	2,522	0.99	0.86	0.92
Various expression	901	0.98	0.84	0.90
Occlusion from eyeglass	648	0.96	0.77	0.85
Occlusion from sunglass	648	0.91	0.35	0.50
Occlusion from cap	657	0.94	0.55	0.64

The two scenarios, when combined, produce a binary confusion matrix. It demonstrates the overall verification performance based on the total number of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP). The genuine acceptance rate (GAR), false rejection rate (FRR), genuine rejection rate (GRR) and false acceptance rate (FAR) are calculated from values of TP, FN, TN, and FP. Table 2 summarizes the system performance on two different scenarios, while Figure 3 visualizes them in a binary confusion matrix. The chance that the system accepts falsely an unauthorized person is below 2%. Though FAR is minimal, this issue is still addressed through the proposed MFA method.

Table 2. Performance summary of the proposed face verification system.

Challenge	Condition positive				Condition negative			
	TP	*FN*	*GAR*	*FRR*	*TN*	*FP*	*GRR*	*FAR*
Various pose	43548	6892	0.8634	0.1366	50415	25	0.9995	0.0005
Various expression	15172	2848	0.8420	0.1580	17385	635	0.9648	0.0352
Occlusion from eyeglass	10039	2921	0.7746	0.2254	12695	265	0.9796	0.0205
Occlusion from sunglass	4542	8418	0.3505	0.6495	12676	284	0.9781	0.0219
Occlusion from cap	7273	5867	0.5535	0.4465	12892	248	0.9811	0.0189
Total	80574	26946	3.384	1.616	106063	1457	4.903	0.0969
Average	16115	5389	**0.677**	**0.323**	21213	291	**0.981**	**0.0194**

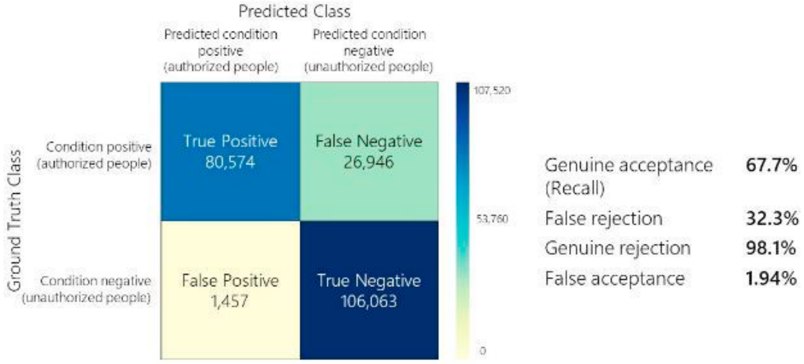

Fig. 3. Binary confusion matrix for the face verification system.

4.2. Evaluation on Leap Motion-based gesture recognition

To produce a good classifier, this study constructs a dataset containing 500 images per digit. The 400 images for each digit are used for training while the rest are used for the evaluation. Figure 4 shows the system performance as visualized in a confusion matrix and Table 3 summarizes the report. Based on table, the classifier garners 96% for the average values of precision, recall and f-scores. This result concludes that the proposed handwriting gesture recognition using Leap Motion controller for detection is a feasible method. cap.

Table 3. Classification report of the proposed gesture recognition.

Digit	Precision	Recall	F-score
0	0.98	1	0.99
1	0.99	0.96	0.97
2	0.92	0.95	0.94
3	0.90	0.90	0.90
4	0.95	0.94	0.94
5	0.99	0.99	0.99
6	0.95	0.98	0.97
7	1	0.98	0.99
8	0.96	0.92	0.94
9	0.95	0.97	0.96
Average	0.96	0.96	0.96

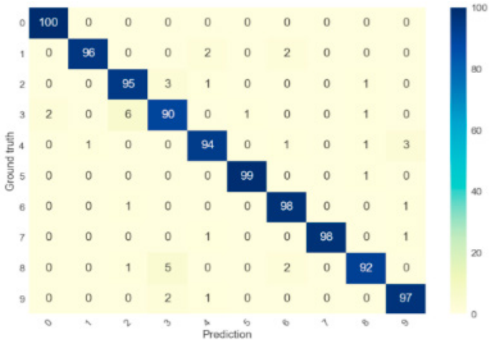

Fig. 4. Confusion matrix for the handwriting gesture classifier.

4.3. Evaluation on the proposed method

Since the proposed method of this study is a sequential verification process, it is fundamental to examine the performance of the standalone face verification system and the proposed MFA. As can be seen in Figure 5, the proposed "Faceture ID" is superior to the face verification system which has a minimal chance to be infiltrated by an unauthorized person. This effect is emphasized through requiring additional authentication factors. Every authorized person has a phone number that is saved inside the database for the messaging of OTP, and in result, the unauthorized person cannot bypass the latter authentication process. Thus, the proposed MFA gains a zero false acceptance rate, resulting in the improvement of DET curves as shown in a solid curve.

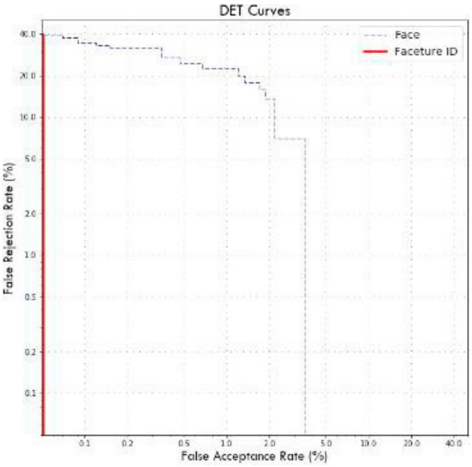

Fig. 5. Detection Error Tradeoff (DET) curves of the face verification and the proposed method.

5. Conclusion and future work

The present study introduces "Faceture ID", a proof-of-concept of a new novel MFA that cascades face verification, OTP and hand gesture recognition. The experimental results on face verification show more than 77% accuracies (average recalls) on instances where images are exposed to variations in pose, expression, and occlusion from eyeglass. The chance that an unauthorized person may gain an access is below 2%, and this outcomess makes the system a recommendation for implementing practical face-related applications such as surveillance, monitoring, and public safety. The quantitative evaluation of the developed gesture recognition presents 96% for both precision and recall. In other words, the classifier for the handwriting gestures shows a feasible method using the projection of trajectory-based gestures. However, the performance of the system is limited to the capability of the Leap Motion controller to obtain 3D coordinates of the joints in each hand.

Further, this proposed study can be a baseline research for future endeavors in this field of work which have similar methodologies. Though, the proposed framework can be further improved by implementing a behavioral gesture recognition which can distinguish the unique strokes of the user, thereby strengthening the security.

Acknowledgments

The paper would like to acknowledge the School of Graduate Studies and the Mindanao State University - Iligan Institute of Technology. Special acknowledgment to the Department of Science and Technology – Engineering Research and Development for Technology (DOST-ERDT).

References

1. Nag K, Roy A, Dasgupta D. An Adaptive Approach Towards the Selection of Multi-Factor Authentication. In 2015 IEEE Symposium Series on Computational Intelligence; 2015 December. p. 463-472.
2. Ba Z, Ren K. Addressing smartphone-based multi-factor authentication via hardware-rooted technologies. In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS); 2017. p. 1910-1914.
3. Uluagac AS, Liu W, Beyah R. A multi-factor re-authentication framework with user privacy. In 2014 IEEE Conference on Communications and Network Security; 2014. p. 504-505.
4. Prasad KS, Varanasi A, Kumar UV. Two Factor Authentication System using Intervened password and Color Pattern. International Journal of Scientific & Engineering Research. 2015 June; 6(6).
5. Akram S, Misbahuddin , Mohammed G. A usable and secure two-factor authentication scheme. International Security Journal: A Global Perspective. 2012 January; 21(4).
6. Khan S, Akbar MA, Shahzad F, Farooq M, Khan Z. Secure biometric template generation for multi-factor authentication. Pattern Recognition. 2015 February; 48(2).
7. Bauckman DT, Johnson NP, Robertson DJ, inventors; Inc. ZS, assignee. Multi-factor authentication. U.S. patent 8,984,605. 2015 March 17.
8. Velásquez I, Caro , Rodríguez A. Authentication Schemes and Methods: a Systematic Literature Review. Information and Software Technology. 2018 September; 94: p. 30-37.
9. Huang CY, Ma SP, Chen KT. Using one-time passwords to prevent password phishing attacks. Journal of Network and Computer Applications. 2011 July; 34(4).
10. Hema D, Bhanumathi S. Mouse behaviour based multi-factor authentication using neural networks. In 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT); 2016. p. 1-8.
11. Khan S, Akbar MA. Multi-factor authentication on cloud. In 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA); 2015. p. 1-7.
12. Yohan A, Lo NW, Lie HR. Dynamic multi-factor authentication for smartphone. In 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC); 2016. p. 1-6.
13. Ross A, Jain AK. Multimodal biometrics: an overview. In 2004 12th European Signal Processing Conference; 2004. p. 1221-1224.
14. Choi H, Park H. A multimodal user authentication system using faces and gestures. BioMed research international. 2015.
15. McCartney R, Yuan J, Bischof HP. Gesture recognition with the leap motion controller. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV); 2015: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). p. 3.
16. Xu X, Le HA, Dou P, Wu Y, Kakadiaris IA. Evaluation of a 3D-aided pose invariant 2D face recognition system. In 2017 IEEE International Joint Conference on Biometrics (IJCB); 2017. p. 446-455.
17. Shang K, Huang ZH, Liu W, Li ZM. A single gallery-based face recognition using extended joint sparse representation. Applied Mathematics and Computation. 2018 March; 320.
18. Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 815-823.
19. John V, Umetsu M, Boyali A, Mita M, Imanishi M, Sanma N, et al. Real-time hand posture and gesture-based touchless automotive user interface using deep learning. In 2017 IEEE Intelligent Vehicles Symposium (IV); 2017. p. 869-874.
20. Hu JT, Fan CX, Ming Y. Trajectory image based dynamic gesture recognition with convolutional neural networks. In 2015 15th International Conference on Control, Automation and Systems (ICCAS); 2015. p. 1885-1889.
21. Weichert F, Bachmann D, Rudak , Fisseler D. Analysis of the accuracy and robustness of the leap motion controller. Sensors. 2013 May; 13(5).
22. Guna J, Jakus G, Pogačnik , Tomažič , Sodnik. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors. 2014 February; 14(2).
23. Bachmann D, Weichert F, Rinkenauer G. Evaluation of the leap motion controller as a new contact-free pointing device. Sensors. 2014 December; 15(1).
24. Amos B, Ludwiczuk B, Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science. 2016 June.