Expert Systems With Applications 172 (2021) 114600

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Detection of spoofing attacks for ear biometrics through image quality assessment and deep learning

İ. Toprak [*], Ö. Toygar [1]

Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey

ARTICLE INFO

Keywords:
Ear biometrics
Spoof detection
Printed photo attack
Image quality measure
Deep learning

ABSTRACT

Ear recognition systems are one of the popular person identification systems. These biometric systems need to be protected against attackers. In this paper, a novel method is proposed to detect spoof attacks within ear recognition systems. The proposed method employs Convolutional Neural Network (CNN) which is based on deep learning and Image Quality Measure (IQM) techniques to detect printed photo attacks against ear recognition systems. Full-reference and no-reference image quality measures are used to extract ear image features. Score-level fusion is used to combine the scores obtained from image quality measures. Finally, decision-level fusion is employed to fuse the decisions obtained from CNN and IQM systems. The final decision is obtained as real or fake image as the output of the whole system. The experiments are conducted on publicly available ear datasets namely, AMI, UBEAR, IITD, USTB set 1 and USTB set 2 and the obtained results are compared with the state-of-the-art methods that are focused on printed photo attacks as well.

1. Introduction

Today, the identification systems based on biometric traits become inevitable for human life. The application of the person identification systems has wide range in the area of the biometric community. Ear recognition systems are recently used in the area of biometrics. Many studies have been done so far to propose novel and robust ear recognition systems. Since ear of the human has rich, reliable, distinctive and invariant features and no cooperation with the user is required, it is a preferred biometric trait to be implemented in the person identification systems (Hassaballah, Alshazly, & Ali, 2019; Alqaralleh & Toygar, 2018; Omara, Feng, Zhang, & Zuo, 2016; Yuan & Mu Chun, 2012; Toygar, Alqaralleh, & Afaneh, 2018). Although biometric trait based person identification systems provide some advantages in human's life, lack of security of these systems can cause horrible outcomes. The intruders may attempt to have access to these systems' accounts by applying illegal methods. In other words, if the intruder tries to have access to someone else's account for specific purposes, this is called spoofing. There are many ways for intruders to do that. A person identification system which is based on any biometric trait can be attacked with printed photo of a biometric trait. The most common attack types are printed photo attack, digital photo attack, replay video attack, mask attack and plastic surgery attack (Kisku & Rakshit, 2017). Consequently, these systems need to be protected against fraudulent attacks. Many anti-spoofing methods which are mostly based on deep learning, texture, motion, image quality and liveness have been proposed by scientists for spoof detection. The details related to these methods are explained in the next section. In this study, a novel and efficient method is proposed to counter printed photo attacks to an identification system which is based on ear biometric trait. In the implementation of the proposed method, deep learning based Convolutional Neural Network method and 5 Image Quality Measures (IQM) are fused to obtain a robust anti-spoofing algorithm. There are some contributions of this study. Firstly, the fusion of CNN and IQM functions for ear anti-spoofing systems is implemented the first time in this study. Secondly, the fusion of 5 image quality metrics for the detection of spoofing attacks for ear biometrics through feature-level, decision-level and score-level fusion strategies are presented. Finally, ear anti-spoofing system results are demonstrated the first time on four ear databases namely, AMI, UBEAR, IITD and USTB in this study. The rest of this paper is organized as follows: the literature review and the details of the methodology are explained in Section 2 and 3, respectively. Section 4 describes the proposed method. Experimental results are demonstrated in Section 5. Finally, the paper is concluded in Section 6.

* Corresponding author.
E-mail addresses: imren.toprak@emu.edu.tr (İ. Toprak), onsen.toygar@emu.edu.tr (Ö. Toygar).
[1] Principal corresponding author

https://doi.org/10.1016/j.eswa.2021.114600
Received 7 August 2019; Received in revised form 10 January 2021; Accepted 10 January 2021
Available online 15 January 2021

2. Literature review

According to the literature, anti-spoofing techniques are based on texture, liveness, motion and image quality (Kisku & Rakshit, 2017). Texture-based methods are implemented in the feature extraction step of an anti-spoofing system. In these type of methods, in order to determine whether the biometric test image is real or fake, the comparison of the test image and the training image is performed by analysing their texture patterns. In the study of Raghavendra and Busch (2015), Multi-Scale Binarized Statistical Image Feature (M-BSIF) extraction method has been applied to describe the textures of the iris images for presentation attack detection system. In Boulkenafet, Komulainen, and Hadid (2016), Local Binary Patterns (LBP), Co-occurrence of Adjacent Local Binary Patterns (CoALBP), Local Phase Quantization (LPQ), Binarized Statistical Image Feature (BSIF) and Scale-Invariant Descriptor (SID) have been implemented on colour images for face spoof detection system.

Scientists state that the use of motion of the user, such as head movement, lips movement, eye blinking and expression changes, has contribution in the solution of an anti-spoofing problem. In that approach, motion of a person is tracked to detect fraudulent attempts to the identification system. In the study of Edmunds and Caplier (2018), a novel countermeasure method has been proposed to detect spoof attacks to a face recognition system. In that method, Conditional Local Neural Fields (CLNF) algorithm is applied for face tracking in the low-level of the proposed method. Further, Fisher vectors are used to describe the motions in the mid-level of the proposed method. According to that study, motion-based methods can be used as an extra countermeasure in anti-spoofing systems. Besides, liveness detection is another way to cope up with the problem of spoofing. In order to detect liveness of the biometric trait, hardware-based and software-based tecniques are implemented in the literature. Hardware-based techniques can be implemented in the sensor-level of an anti-spoofing system to measure the sweat of the fingerprint, facial thermogram, blood pressure or reflection of eye by integrating extra sensing devices. On the other hand, in the study of Gragnaniello et al. (2015), a liveness detection algorithm has been implemented by using LBP for print attack of iris images on mobile devices. Further, in the study of Singh and Arora (2017), eye blinking and lip movement have been considered to detect liveness of the facial images by implementing morphological operations. In biometrics community, image quality assessment technique is another popular way to counter spoof attacks. The images are classified as genuine or impostor image by taking the quality difference between them into consideration. In the study of Galbally, Marcel, and Fierrez (2014), 25 image quality measures are computed and used for spoof detection of attacks based on iris, face and fingerprint biometric traits. Moreover, image quality based, motion-based and texture-based techniques are combined to obtain robust algorithms for spoof detection in some studies such as Feng et al. (2016) and Farmanbar and Toygar (2017). CNN-based deep learning methods are widely used methods for anti-spoofing algorithms for different kinds of biometric traits such as face, iris, fingerprint, fingervein and even speech. The authors proposed face anti-spoofing method which is based on fusion of image quality and motion cues with the approach of CNN in Feng et al. (2016). Next, in the study (Kim, Park, Song, & Yang, 2016), a deep learning based method for fake fingerprint detection has been proposed. Further, in the paper (Czajka, Bowyer, Krumdick, & VidalMata, 2017), the authors have compared their two studies. The first study is based on handcrafted features that are classified by Support Vector Machine (SVM). The second study is based on learned features using CNN. Consequently, CNN-based approach performed better results compared to the first study. Additionally, authors have proposed finger vein presentation attack detection algorithm which is based on deep CNN in the paper (Raghavendra, Venkatesh, Raja, & Busch, 2017). Moreover, Dinkel, Qian, and Yu (2018) proposed speech spoof detection method based on deep Neural Networks (NN) on the dataset of BTAS2016 and overcome other methods.

3. Methodology

In this study, CNN-based deep learning and Image Quality Assessment methods are employed to propose a robust and efficient ear anti-spoofing method. The employed methods are explained below.

3.1. Image quality assessment

Image Quality Assessment (IQA) is applied to measure the quality of the image in the presence of noise, blur, contrast, change of illumination or any other distortion. In the literature, there are two types of Image Quality Measures (IQM) namely, Full-Reference (FR) IQM and No-Reference (NR) IQM. Conceptually, FR IQM functions are used to compare the original image and its distorted version. In contrast, NR IQM functions are used to evaluate the quality of the original image without comparison. In this study, Peak Signal to Noise Ratio (PSNR), Gradient Phase Error (GPE), Structural Similarity Index (SSIM) and Reduced Reference Entropic Difference (RRED) are employed as FR IQM functions. Furthermore, High-Low Frequency Index (HLFI) is employed as NR IQM function. The mathematical formulas of PSNR, GPE and HLFI are described below (Galbally et al., 2014). The details of the SSIM and RRED functions are explained in Wang, Bovik, Sheikh, and Simoncelli (2004) and Soundararajan and Bovik (2012), respectively. PSNR is calculated as follows:

$$PSNR\left(I,\widehat{I}\right)=10\times log\left(\frac{max\left(I^2\right)}{MSE\left(I,\widehat{I}\right)}\right) \tag{1}$$

where *MSE* is the Mean Squared Error (MSE) and it is computed as follows:

$$MSE\left(I,\widehat{I}\right)=\frac{1}{N\times M}\times\sum_{n=1}^{N}\sum_{n=1}^{M}\left(I_{i,j}-\widehat{I}_{i,j}\right)^2 \tag{2}$$

where I is the original image, $\widehat{I}$ is the distorted image and $(N\times M)$ is the size of the image.

The formula for GPE is as follows:

$$GPE\left(I,\widehat{I}\right)=\frac{1}{N\times M}\sum_{n=1}^{N}\sum_{n=1}^{M}|arg\left(G_{i,j}\right)-arg\left(\widehat{G}_{i,j}\right)|^2 \tag{3}$$

where $(G_{i,j})$ and $(\widehat{G}_{i,j})$ represent gradient maps of I and $\widehat{I}$ in the x and y directions, respectively.

HLFI metrics is calculated as follows:

$$HLFI\left(I\right)=\frac{\sum_{i=1}^{i_l}\sum_{j=1}^{j_l}\left|F_{i,j}\right|-\sum_{i=i_h+1}^{N}\sum_{j=j_h+1}^{M}\left|F_{i,j}\right|}{\sum_{i=1}^{N}\sum_{1}^{M}\left|F_{i,j}\right|} \tag{4}$$

where $(F_{i,j})$ represents Fourier transform of I.

3.2. CNN-based deep learning

Generally, Neural Network (NN) systems consist of neurons that are placed in the input layer, output layer and hidden layers. These neurons are connected in a specific way in order to communicate with each other. Each neuron in the network has a specific weight. Input neurons take the signal from the environment and combine it with its weight. Afterwards, the computation result is transmitted to the subsequent layer's neurons. Finally, output layer's neurons convey the computation result to the environment (LeCun, Bengio, & Hinton, 2015).

Convolutional Neural Network (CNN) is a widely used supervised deep learning model which was developed by LeCun, Bottou, Bengio, and Haffner (1998). The first component of this model consist of

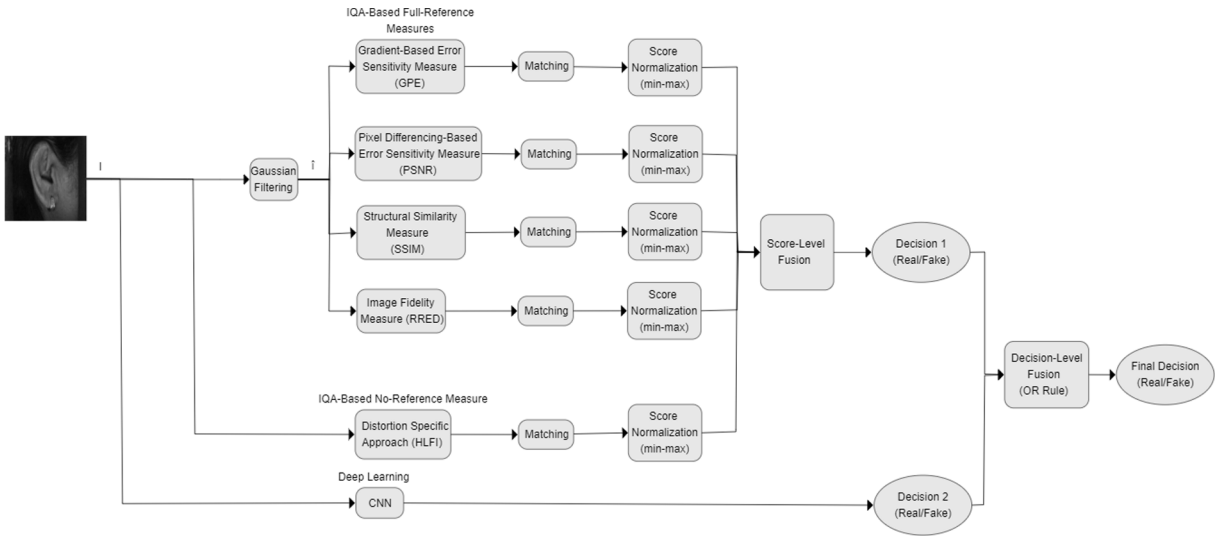

Fig. 1. General block diagram of the proposed ear anti-spoofing method.

convolutional layers. In order to extract feature representation of an input image, convolutional operation is applied by using convolution filters in that layer. The input image is searched to detect different visual elements by convolving it with learned multiple filters. Among these filters are vertical filter, horizontal filter or diagonal filter where each filter extracts different features. The convolved results are called as feature maps. The number of obtained feature maps is equal to the number of convolutional filters used. In order to transmit computed output values of neurons of current layer to the next layer, non-linear function is needed to detect non-linear features. The most commonly used non-linear function in CNN architectures is Rectified Linear Unit (ReLU) because it works better in terms of speed (Wang & Chen, 2020; Agarap, 2018). ReLU puts zero instead of negative values which represent black. Implementing the first convolutional layer provides to obtain low-level features such as edges, colour, texture, gradient orientation, etc. of an image. Additionally, if more convolutional layers are added into implementation, high-level features will be obtained as well. This approach will led network to learn the input image deeper. Consequently, we can say that using multiple convolutional layers is advantageous.

The next component of a CNN model is called pooling layer which aims to reduce the computation of data and select important features (Phan, Hertel, Maass, & Mertins, 2016; Zhou et al., 2016). The function of this layer is subsampling the output feature representation of convolutional layer to reduce the dimensionality of the feature map. One of the most commonly used pooling layers is Max pooling which selects maximum value from subdivided feature map. Afterwards, obtained features are flattened to construct one dimensional feature vector and this feature vector is fed to fully-connected layers. The principle of fully-connected layer is like traditional NN model. It consists of input layer, output layer and hidden layers. In a fully-connected layer, each neuron of current layer is connected to all neurons of the next layer. In this context, the obtained feature vector will be received by neurons of input layer and pass through all hidden layers. Finally, the output of the last hidden layer will be the input to output layer to be classified as a specific class. In the output layer, Softmax classifier is applied for the classification of an image. The pattern of input image is learned by the convolutional network and the classification provides the output. In that network, every iteration of training is achieved by applying back-propagation algorithm. In order to obtain the best classification rate, the network model will be trained over a number of epochs (Gu et al., 2018). In the CNN model, memorizing training data problem which is also known as overfitting occurs in the case of small number of image data. This problem can be prevented by using regularization techniques where dropout is one of these techniques (Wu & Gu, 2015). The working principle of Dropout is that some neurons are dropped in every iteration of NN model. Therefore, the network is not dependent to specific neurons (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). Moreover, in order to optimize CNN model, some techniques are available. One of these techniques is Batch Normalization (BN) that is applied to normalize the inputs of each layer to overcome internal covariant shift problem (Ioffe & Szegedy, 2015).

4. Proposed method

Fusion techniques become compulsory to combine multiple results into a common one in the solution of computer vision applications. In the literature, feature-level fusion, score-level fusion and decision-level fusion are the most commonly applied fusion techniques (Ross & Jain, 2003). Firstly, in the feature-level fusion technique, extracted multiple features are fused in order to obtain common feature vector for further process. Besides, score-level fusion technique is applied into obtained scores after matching step of the method. In this technique, the obtained scores from multiple algorithms are normalized to bring them into a common scale so that they can be ready to fuse. There are several types of normalization techniques namely, min–max, tanh and z-score. Finally, when multiple decisions are obtained after decision module, decision-level fusion either with OR rule, majority voting or weighted majority voting technique is applied to have the final decision. In this study, feature-level fusion technique has been applied to combine the extracted features from 5 IQM functions. Furthermore, decision-level fusion with OR rule and majority voting technique and score-level fusion with aforementioned normalization techniques have been applied.

Firstly, IQM functions are employed to extract distinctive features from test image in the proposed method. In the first step, the original test image (I) is filtered to obtain the smoothed image (Î) by using Gaussian 3×3 kernel filter which has 0.5 as σ value. This approach expects that the loss of quality delivered by Gaussian filtering varies between genuine and fake biometric test images (Galbally et al., 2014). Afterwards, (I) and (Î) are used as an input to FR IQM functions. In the meanwhile, (I) is used as an input to NR IQM function. In the matching step, Manhattan distance is calculated between test image and all training images according to the obtained results in the previous step. In the next step, min–max score normalization is applied to adjust scores obtained in different scale to a common scale. Score-level fusion is employed as a final step to obtain final decision (real\fake). The block

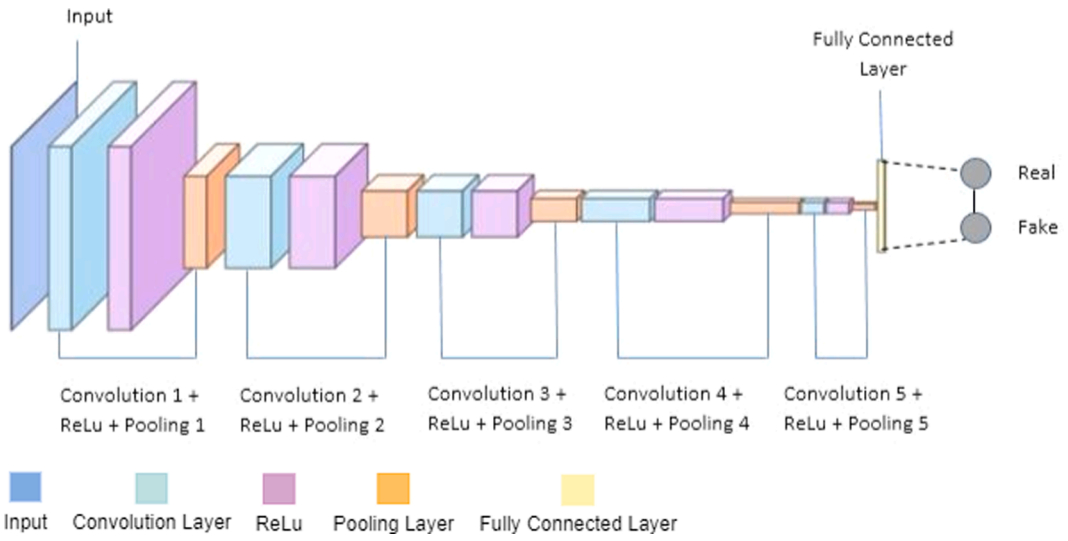

Fig. 2. Architecture of the CNN part of the proposed ear anti-spoofing method.

diagram of the proposed method is illustrated in Fig. 1. Moreover, deep learning based method CNN is employed to discriminate test image either as real or fake. The architecture of CNN part is illustrated in Fig. 2. CNN part includes 5 convolutional layers with size of 3×3 filter map to extract features and non-linear ReLu activation function to detect non-linear features. The number of output filters in the first, second, third, fourth and fifth convolution layers are 32, 64, 64, 96 and 32, respectively. After each convolutional layer, max pooling layer with 2×2 pooling size is applied. Further, Batch Normalization (BN) follows each max pooling layer. Before flattening of feature maps, Dropout with 0.2 probability is applied to overcome overfitting problem. Next, concatenated feature vector is used as an input layer to feed the Neural Network (NN). Epoch number is adjusted to 250 to train the NN. In the network, Softmax classifer is used to classify input ear image as real or fake. After implementation of aforementioned methods separately, two decisions are obtained. Therefore, in order to obtain final decision among obtained decisions, decision-level fusion with OR rule is applied.

5. Experimental results

In order to evaluate the performance of the ear anti-spoofing systems, several experiments have been conducted. Firstly, every method has been implemented separately to observe the performance individually. Afterwards, fusion techniques have been applied to increase the performance of the individual methods. The details of the conducted experiments are explained below.

The aforementioned experiments have been conducted on 5 different datasets namely, UBEAR, AMI, IITD, USTB set 1 and set2 (Raposo, Hoyle, Peixinho, & Proença, 2011; Gonzalez et al., YYYY; Kumar & Wu, 2012; Mu & Yuan, YYYY). As a summary, left and right ears of 50 subjects are selected randomly which makes 100 ear images in total from UBEAR and AMI databases to construct our experimental datasets. The selected images have the size of 236×159 and 702×492 for UBEAR and AMI datasets, respectively. There are right ear of 124 subjects with the size of 204×272 in the IITD dataset that is used for our experiments. Finally, the first and the second dataset of USTB database are used for constructing our experimental datasets. In this context, right ear of 60 subjects with the size of 150×80 and right ear of 76 subjects with the size of 400×300 are used for set 1 and set 2, respectively. The spoof databases for aforementioned databases are constructed by us because there is no available spoof database for ear biometric trait in the biometric community. In this context, the original ear images are printed on A4 paper by using Olivetti d-colour mf223 printer device that has resolution of 1800×600. Next, each printed ear image is captured by Iphone 6s camera that is 12 megapixel. Captured ear images are resized for all databases separately and stored in spoof databases. Afterwards, train and test sets are constructed for all aforementioned databases. Half of the ear images which are real in one database and corresponding fake ear images are stored in the train set. The rest of the ear images of that database and corresponding fake ear images are stored in test set of that database. Consequently, 5 train sets and 5 test sets are constructed for aforementioned databases. The obtained results are shown in terms of False Fake Rate (FFR), False Genuine Rate (FGR) and Half Total Error Rate (HTER). FFR represents the number of real images that are classified as impostor and FGR represents the number of fake images that are classified as genuine. Additionally, HTER is computed as (FFR + FGR)/2.

5.1. Preliminary experiments

Implementations of deep learning method which is CNN and image quality measures namely PSNR, GPE, SSIM, HLFI and RRED have been performed separately. The results for each method on AMI, UBEAR and IITD databases for ear anti-spoofing are shown in Table 1 and the error

Table 1
Effects of deep learning and image quality assessment methods for ear anti-spoofing on AMI, UBEAR and IITD databases (Results are in percentages).

	AMI Database			UBEAR Database			IITD Database		
Method Name	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
CNN	0.0	1.0	0.5	69.0	0.0	34.5	1.0	1.0	1.0
PSNR	42.0	26.0	34.0	35.0	39.0	37.0	13.0	30.0	21.5
GPE	10.0	4.0	7.0	45.0	27.0	36.0	2.0	10.0	6.0
SSIM	4.0	11.0	7.5	46.0	39.0	42.5	0.0	39.0	19.5
RRED	24.0	28.0	26.0	58.0	54.0	56.0	34.0	31.0	32.5
HLFI	43.0	40.5	41.5	52.0	43.0	47.5	2.0	50.0	26.0

Table 2
Effects of deep learning and image quality assessment methods for ear anti-spoofing on USTB set 1 and set 2 datasets (Results are in percentages).

	USTB Set 1 Database			USTB Set 2 Database		
Method Name	FFR	FGR	HTER	FFR	FGR	HTER
CNN	3.0	6.0	4.5	1.0	0.0	0.5
PSNR	18.0	9.0	13.5	3.0	0.0	1.5
GPE	14.0	8.0	11.0	7.0	7.0	7.0
SSIM	17.0	15.0	16.0	2.0	0.0	1.0
RRED	21.0	15.0	18.0	1.0	3.0	2.0
HLFI	16.0	14.0	15.0	0.0	0.0	0.0

rates on USTB datasets are demonstrated in Table 2. As shown in Table 1 and Table 2, CNN performs better results with HTER values of 0.5, 34.5, 1.0 and 4.5 for datasets of AMI, UBEAR, IITD and USTB set 1, respectively. On the other hand, the minimum HTER value which is 0.0 is obtained by HLFI for USTB set 2. According to the comparison of IQM functions, GPE achieves better performance with HTER values of 7.0, 36.0, 6.0 and 11.0 for AMI, UBEAR, IITD and USTB set 1 datasets. On the other hand, HLFI performs better result with HTER value of 0.0 for USTB set 2 dataset. Since the obtained results are not consistent for all datasets, fusion techniques have been applied to obtain more accurate results.

In order to determine which fusion technique works better for the combination of the results of 5 IQM functions, Score-Level Fusion (SLF) with min–max, tanh and z-score normalization, Feature-Level Fusion (FLF) and Decision-Level Fusion (DLF) with Majority Voting techniques have been applied. According to the results shown in Table 3 and Table 4, Score-Level Fusion with min–max and z-score normalizations perform same results with HTER value of 0.0 on datasets of AMI, UBEAR, USTB set 1 and set 2. Meanwhile, DLF performs 0.0 HTER value for all datasets except UBEAR dataset. According to the results, Score-Level Fusion with min–max normalization technique is preferred because of its computation simplicity.

In order to have consistent and robust results for all datasets, fusion technique of IQA and CNN is proposed for ear anti-spoofing problem in this paper. Results for implementation of the proposed method on AMI, UBEAR, IITD, USTB set 1 and set 2 databases are shown in Table 5. As it is shown, HTER value of 0.0 has been obtained on AMI, UBEAR, IITD, USTB set 1 and set 2, respectively. As a result, our proposed method achieves the best performances for all datasets.

The execution times of the proposed method including both training and test times (model application time) in minutes for AMI, UBEAR, IITD, USTB Set 1 and USTB Set 2 databases are 316.65, 322.35, 23.41, 9.57 and 14.03, respectively.

5.2. Comparison with the state-of-the-art

The comparison results which have been made with our proposed method and the other biometrics anti-spoofing methods are presented in Table 6. State-of-the-art systems use CNN for countering against spoof attacks and involve face, fingerprint and iris biometric traits. In the first study (Wang, Nian, Li, Meng, & Wang, 2017), the method is proposed for face anti-spoofing and the obtained error rates are 1.2 and 2.3 for their private database and CASIA, respectively. The second study (Rehman, Po, & Liu, 2018) proposes an anti-spoofing method for face biometric trait and the obtained error rates are 19.12 and 8.39 for CASIA-FASD and Replay-Attack databases, respectively. In the third study (Li et al., 2018), the proposed anti-spoofing method is for face biometric trait and the obtained error rates are 1.4, 1.2, 0.0 and 7.0 for CASIA-FASD, Replay-Attack, MSU and Rose-Youtu databases, respectively. Further, a novel method is implemented for fingerprint biometric trait (Yuan et al., 2018) and 3.7 and 6.45 error rates are obtained for LivDet2013 and LivDet2011 databases, respectively. Lastly, iris anti-spoofing CNN-based method is proposed in Kuehlkamp, Pinto, Rocha, Bowyer, and Czajka (2018) and 3.28, 0.68, 9.45 and 14.92 error rates are obtained on Notre Dame, Warsaw, Clarkson and IITD + WVU databases, respectively. The comparison of the results of our method and the results of the state-of-the-art methods shows that the proposed method in this study achieves the best perfomances for all datasets with zero error rates.

6. Conclusion

In this paper, a novel ear anti-spoofing method is implemented to detect printed photo attacks by combining CNN and 5 IQM functions. Four of these IQMs are full-reference and one of them is no-reference. According to the preliminary experiments, results are not consistent for all datasets whenever deep learning and IQM methods are employed separately. Therefore, the fusion of CNN and IQM is proposed in this study. Comparison of our proposed method with the state-of-the-art CNN-based methods shows that our method achieves the best performances for all datasets used in the experiments. As a future work,

Table 4
Decision-level and score-level fusion results (in percentages) of IQMs on USTB set 1 and set 2 datasets.

	USTB Set 1 Database			USTB Set 2 Database		
Method Name	FFR	FGR	HTER	FFR	FGR	HTER
DLF(PSNR,GPE,SSIM,RRED, HLFI)	0.0	0.0	0.0	0.0	0.0	0.0
FLF(PSNR,GPE,SSIM,RRED, HLFI)	11.0	6.0	8.5	3.0	0.0	1.5
SLF-tanh(PSNR,GPE,SSIM, RRED,HLFI)	0.0	0.0	0.0	0.0	0.0	0.0
SLF-min–max(PSNR,GPE,SSIM, RRED,HLFI)	0.0	0.0	0.0	0.0	0.0	0.0
SLF-z-score(PSNR,GPE,SSIM, RRED,HLFI)	0.0	0.0	0.0	0.0	0.0	0.0

Table 5
Proposed method (IQA + CNN) results (in percentages) for ear anti-spoofing.

	Proposed Method (IQA + CNN)		
Database Name	FFR	FGR	HTER
AMI Database	0.0	0.0	0.0
UBEAR Database	0.0	0.0	0.0
IITD Database	0.0	0.0	0.0
USTB Database Set1	0.0	0.0	0.0
USTB Database Set2	0.0	0.0	0.0

Table 3
Decision-level and score-level fusion results (in percentages) of IQMs on AMI, UBEAR and IITD databases.

	AMI Database			UBEAR Database			IITD Database		
Method Name	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
DLF(PSNR,GPE,SSIM,RRED,HLFI)	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0
FLF(PSNR,GPE,SSIM,RRED,HLFI)	10.0	12.0	11.0	30.0	27.0	28.5	1.0	32.0	16.5
SLF-tanh(PSNR,GPE,SSIM,RRED,HLFI)	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
SLF-min–max(PSNR,GPE,SSIM,RRED,HLFI)	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0
SLF-z-score(PSNR,GPE,SSIM,RRED,HLFI)	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0

Table 6
Comparison of the proposed ear anti-spoofing method with the state-of-the-art biometrics anti-spoofing methods using CNN (Results are in percentages).

Reference	Methods Used	Biometric Trait	Databases Used	#Subject	Size of Image	#Epoch	HTER
Wang et al. (2017)	CNN	Face	Collected by them	20	256*256	–	1.2
			CASIA	50			2.3
Rehman et al. (2018)	CNN	Face	CASIA-FASD	50	96*96	500	19.12
			Replay-Attack	50			8.39
Li et al. (2018)	CNN	Face	CASIA-FASD	50	128*128	50	1.4
			Replay-Attack	50			1.2
			MSU	35			0.0
			Rose-Youtu	20			7.0
Yuan et al. (2018)	CNN	Fingerprint	LivDet2013	–	200*200	100	3.7
			LivDet2011	–			6.45
Kuehlkamp et al. (2018)	BSIF + CNN	Iris	Notre Dame	–	260*260	–	3.28
			Warsaw	–			0.68
			Clarkson	–			9.45
			IITD + WVU	–			14.92
Proposed Method	IQA + CNN	Ear	AMI	100	256*256	250	0.0
			UBEAR	100			0.0
			IITD	62			0.0
			USTB Set 1	30			0.0
			USTB Set 2	38			0.0

different types of attacks such as replay video attack and digital photo attack can be investigated on ear biometrics. Further, ear anti-spoofing systems can be integrated to face anti-spoofing systems to develop a more robust anti-spoofing system.

CRediT authorship contribution statement

İ. **Toprak:** Methodology, Software, Investigation, Writing - original draft, Visualization. Ö. **Toygar:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv:1803.08375.

Alqaralleh, E., & Toygar, O. (2018). Ear recognition based on fusion of ear and tragus under different challenges. *International Journal of Pattern Recognition, 32*, 1856009–1856019.

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2016). Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security, 11*, 1818–1830.

Czajka, A., Bowyer, K. W., Krumdick, M., & VidalMata, R. G. (2017). Recognition of image-orientation-based iris spoofing. *IEEE Transactions on Information Forensics and Security, 12*, 2184–2196.

Dinkel, H., Qian, Y., & Yu, K. (2018). Investigating raw wave deep neural networks for end-to-end speaker spoofing detection. *The IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26*, 2002–2014.

Edmunds, T., & Caplier, A. (2018). Motion-based countermeasure against photo and video spoofing attacks in face recognition. *The Journal of Visual Communication and Image Representation, 50*, 314–332.

Farmanbar, M., & Toygar, O. (2017). Spoof detection on face and palmprint biometrics. *The Signal and Image Processing, 11*, 1253–1260.

Feng, L., Po, L., Li, Y., Xu, X., Yuan, F., Cheung, T. C. H., & Cheung, K. W. (2016). Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *The Journal of Visual Communication and Image Representation, 38*, 451–460.

Galbally, J., Marcel, S., & Fierrez, J. (2014). Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing, 23*, 710–724.

Gonzalez, E., Alvarez, L. & Mazorra, L. Ami ear database http://www.ctim.es/research_works/ami_ear_database.

Gragnaniello, D., Sansone, C., & Verdoliva, L. (2015). Iris liveness detection for mobile devices based on local descriptors. *Pattern Recognition Letters, 57*, 81–87.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition, 77*, 354–377.

Hassaballah, M., Alshazly, H. A., & Ali, A. A. (2019). Ear recognition using local binary patterns: A comparative experimental study. *Expression Systems and Applications, 118*, 182–200.

Ioffe, S. & Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv, 1–11.

Kim, S., Park, B., Song, B. S., & Yang, S. (2016). Deep belief network based statistical feature learning for fingerprint liveness detection. *Pattern Recognition Letters, 77*, 58–65.

Kisku, D. R., & Rakshit, R. D. (2017). Face spoofing and counter-spoofing: A survey of state-of-the-art algorithms. *Transactions on Machine Learning and Artificial Intelligence, 5*, 31–73.

Kuehlkamp, A., Pinto, A., Rocha, A., Bowyer, K. W., & Czajka, A. (2018). Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection. *IEEE Transactions on Information Forensics and Security, 14*, 1419–1431.

Kumar, A., & Wu, C. (2012). Automated ear identification using ear imaging. *Pattern Recognition, 45*, 956–968.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*, 2278–2324.

Li, H., He, P., Wang, S., Rocha, A., Jiang, X., & Kot, A. C. (2018). Learning generalized deep feature representation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 13*, 2639–2652.

Mu, Z. & Yuan, L. Ear recognition laboratory at ustb http://www1.ustb.edu.cn/resb/en/index.htm.

Omara, I., Feng, L., Zhang, H., & Zuo, W. (2016). A novel geometric feature extraction method for ear recognition. *Expression Systems and Applications, 65*, 127–135.

Phan, H., Hertel, L., Maass, M. & Mertins, A. (2016). Robust audio event recognition with 1-max pooling convolutional neural networks. arXiv:1604.06338.

Raghavendra, R., & Busch, C. (2015). Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. *IEEE Transactions on Information Forensics and Security, 10*, 703–715.

Raghavendra, R., Venkatesh, S., Raja, K. B., & Busch, C. (2017). Transferable deep convolutional neural network features for fingervein presentation attack detection. *Workshop on Biometrics and Forensic (IWBF)*, 1–5.

Raposo, R., Hoyle, E., Peixinho, A., & Proença, H. (2011). Ubear: A dataset of ear images captured on-the-move in uncontrolled conditions. *The IEEE International Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, 84–90.

Rehman, Y. A. U., Po, L. M., & Liu, M. (2018). Livenet: Improving features generalization for face liveness detection using convolution neural networks. *Expression Systems and Applications, 108*, 159–169.

Ross, A., & Jain, A. (2003). Information fusion in biometrics. *Pattern Recognition Letters, 24*, 2115–2125.

Singh, M., & Arora, A. S. (2017). A robust anti-spoofing technique for face liveness detection with morphological operations. *Optics, 139*, 347–354.

Soundararajan, R., & Bovik, A. C. (2012). Rred indices: Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing, 21*, 517–526.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Proceed. IEEE, 15*, 1929–1958.

Toygar, O., Alqaralleh, E., & Afaneh, A. (2018). Symmetric ear and profile face fusion for identical twins and nontwins recognition. *Sig. Image Vid. Proc., 12*, 1157–1164.

Wang, S., & Chen, Y. (2020). Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. *Multimed Tools Appl, 79*, 15117–15133.

Wang, Y., Nian, F., Li, T., Meng, Z., & Wang, K. (2017). Robust face anti-spoofing with depth information. *J. Vis. Commun. Imag. Rep., 49*, 332–337.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*, 600–612.
Wu, H., & Gu, X. (2015). Towards dropout training for convolutional neural networks. *Neural Networks, 71*, 1–10.
Yuan, C., Xia, Z., Jiang, L., Cao, Y., Wu, Q. J., & Sun, X. (2018). Fingerprint liveness detection using an improved cnn with image scale equalization. *IEEE Access, 7*, 26953–26966.
Yuan, L., & Mu Chun, Z. (2012). Ear recognition based on local information fusion. *Pattern Recognition Letters, 33*, 182–190.
Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H. & Xu, B. (2016). Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. arXiv: 1611.06639.

İmren Toprak received B.S., M.Sc. and Ph.D. degrees in Computer Engineering Department of Eastern Mediterranean University in 2012, 2015 and 2020, respectively. Her research interests include biometrics and computer vision.

Önsen Toygar received her B.S., M.S. and Ph.D. degrees in 1997, 1999 and 2004 respectively from Computer Engineering Department of Eastern Mediterranean University, Northern Cyprus. Since September 2004, she worked in Computer Engineering Department of Eastern Mediterranean University. She is currently an Associate Professor in the department and served as the Vice Chair of the department between September 2011 and January 2013. Her current research interests are in the area of biometrics, computer vision, image processing and digital forensics.