Auditory perception vs. face based systems for human age estimation in unsupervised environments: from countermeasure to multimodality

Muhammad Ilyas, Amine Nait-Ali

Version of Record: https://www.sciencedirect.com/science/article/pii/S016786552030427X
Manuscript_f5dff029f985eb3a178b027cdddc4c30

Auditory perception vs. face based systems for human age estimation in unsupervised environments: from countermeasure to multimodality

Muhammad Ilyas[a,c], Amine Nait-ali[b,c]

Universite Paris-est creteil Paris[a,b]

ARTICLE INFO

Keywords:
Age estimation
Countermeasures
Forensics
Multimodal biometrics
Unsupervised biometrics
RaS-DeeP.

ABSTRACT

Face-based age estimation systems are commonly considered in biometric applications as well as in other fields such as forensics or healthcare. For security purposes, features extracted from the face can be used to verify or estimate the age of individuals in order to control their access to physical or logical resources. The main problem in using facial biometrics is its sensitivity, to acquisition (e.g. illumination, pose, occlusion, image quality, etc.), to face expression, and especially to potential attacks in unsupervised environments. In this work, we propose a robust modality using both random auditory stimulation and Deep-learning based age estimation, though individual perception (RaS-DeeP): (1) as a countermeasure to prevent attacks on face-based age estimation systems , but also (2) : as a complementary modality in a multimodal biometric system (i.e. face-sound perception) in order to improve the performances of face-based age estimation system. Used as countermeasure, we show that RaS-DeeP provides promising results with an EER value of 4.2%. On the other hand, when considering the multimodal system face-auditory perception, we show that, the performance of face age estimation system is enhanced with an EER of 3.3%. To evaluate the performance of multimodal system in real-time, 71 subjects from different age ranges achieving five repetitions, participated in our experiment.

1. Introduction

The actions to offer a proper security of confidential data and services require a setup of specific protocols to control users' access. Although passwords provide security to some extent, they are in many cases simply guessed or cracked. While complicated passwords are hard to memorise and are usually kept in unsecured means, one password can give access to multiple resources to fraudsters and allow them to manipulate the information as required.
A sophisticated replacement to passwords is provided by biometric systems. Biometrics is a digital representation of human biological and physiological features such those extracted from the iris, face, fingerprint, voice, palmprint, etc. Even though biometrics is considered secure compared to classical authentication techniques, they may be vulnerable to spiteful attacks. In fact, presentation attacks is considered when a subject tries to access secured information illegitimately, either through a physical or a digital attack. For instance, in face biometric system, an attacker can use a 3D mask or cosmetics as a physical attack to look like a genuine subject. A digital attack, on the other hand, can be manipulated by injecting a recorded video or an image of a genuine subject to a biometric system.

In a real case, it has been reported that a young man impersonated himself as an old man and successfully boarded to the plane in a flat cap [8]. This young man used an artificial face mask made of silicon in order to deceive the border authorities. In the same context, the black hat test has been publicly available designed by many manufacturer in laptops to provide information about how to spoof facial based systems [3]. Such cases demonstrate the possibilities of successful spoof attacks on facial recognition based systems in the real-world.

From the state of the art, many face recognition anti-spoofing systems have been proposed to prevent such biometric systems from being attacked [9], but regarding face age estimation attacks, the field is still challenging and many questions are pending, especially in some unsupervised environments where users deals with their own verification system. As an example of an unsupervised environment, one can consider a scenario when users try to access a website/social-media, initially reserved only for a specific category of people whose ages are within a given range (e.g. under 12 years old). When dealing with this kind of environment, users may have limited resources (e.g. laptops/smartphones). This means that advanced anti-spoofing solutions such those requiring 3D face analysis, infrared analysis, etc, may not be employed. Moreover, in an unsupervised environment, users can achieve many attempts of potential attacks (e.g. and adults who aims to appear young or vice-versa). Within this context, two main contributions are highlighted in this paper. In the first contribution, a countermeasure of face age estimation attacks in an unsupervised environment is considered. This is achieved using a random auditory stimulation and a Deep-learning based age estimation system (RaS-DeeP). The concept of Ras-Deep has been introduced in our previous work [4] as a single anti-spoofing modality. Regarding the second contribution, we show that one can improve the performances of existing face age estimation using the same protocol (RaS-DeeP) in a multimodal system. In both cases (i.e. countermeasure and multimodal system), individual perception is evaluated.

The article is organized as follows: In section 2, several

*Amine Nait-ali
ORCID(s):

anti-spoofing attacks for face based systems are briefly discussed. In section 3, the proposed method for human age estimation is presented. Afterwards, the performance of RaS-Deep as a countermeasure for Deep EXpectation (DEX) is discussed in section 4. In sections 6, the performances of multimodal system (Face, auditory perception) is described in details. Finally, the work is concluded in section 7 along with some ideas for future research.

2. Related Work

A strong facial based age estimation systems are vulnerable to spoof attacks such as using photographs, 3D masks, videos. Photographs can be easily found in any social media network, 3D masks can be developed by using several existing special materials, video's can be created in any events and could be utilized later for spoofing purposes.

The anti-spoofing system using facial features have four different categories [12]: user behavior modeling, relying on extra devices, data driven characterization methods and relying on user corporation. Users's behaviour modeling is related to the behaviour of a subject, such as eye-blinking, movement of the face and different portion of the head. an additional use of hardware such as infrared cameras or images, thermal camera's , 2D camera's, and 3D camera's are related to the methods rely on extra devices for detection of spoof attacks. The data driven characterization methods are artificial cleaning and enhancement of the quality of the images used for the age estimation process.

Hardware based anti-spoofing systems are more likely considered as accurate compared to software based systems, because specialized scanners can catch or highlight directly some intrinsic variations among genuine and synthetic 3D structure faces and other spectral characteristics [10]. The software based solution for detection of spoof-attacks are divided into two categories (active and passive approaches). Furthermore, user engagement can be used very well for facial spoof attacks since we humans are social, while the picture or video attacks can not satisfy the random criteria for intervention. With random facial movement, the identification or measurement of facial 3D structure is very hard. Active software based strategies can make generalizations across various acquisitions and attacks scenarios due to higher time period and system variability for verification.

Passive software-based techniques are more favorable especially for facial anti-spoofing systems compared to countermeasures because they are fast, simple and less invasive. Because of the growing number of public databases for benchmarks, several passive software-based methods for facial anti-spoofing have been introduced. Passive techniques analyze several facial characteristics such as texture , frequency content, and quality , or motion cues (blinking eyes), mouth movements, and facial expression changes, or even color variation due to blood circulation (pulse), to discriminate face artifacts from genuine ones. The results shown by passive software-based techniques are promising in publicly available databases while for the databases collected under unknown conditions, this can affect the performances of the proposed techniques.

Initially, the performances of Convolutional Neural Networks (CNNs) were promising for persons identification and human age estimation using facial characteristics trained through publicly available databases. On the other hand, several methods of fusion have been suggested for a more general countermeasure for a particular form of attack such as linear fusion of frame and video analysis or feature level fusion.

Considering the challenges in the field of biometrics for human age estimation, we are presenting a countermeasure system with another biometric modality. By combining the physical and behavioral characteristics, biometric systems will be able to enhance the performances in the near future. RaS-DeeP is successfully demonstrated for the first time with promising results. In this paper, the RaS-DeeP is used as an anti-spoofing system for automatic face based human age estimation system.

3. Methods and materials

3.1. Face based human age estimation using Deep Expectation (DEX CNN)

For age estimation using facial feature, we used DEX (Deep EXpectation) from the state-of-art [11]. In this section the pipeline of the proposed approach is briefly explained.

Face alignment Many datasets utilized for this work do not show centered frontal faces for DEX technique, but rather faces and the detection occurs and the faces are aligned for training and testing. Instead of deploying complex approaches for precise detection of landmarks, facial characteristics, image wrapping and face alignment for an image several simple and easy approaches also exist [15].

3.1.1. Architecture

In our experiment, Deep Expectation [11] has been considered. CNN has been trained on IMDB-WIKI and ImageNet images using the same model. While we deployed Pinellas County Sheriff's Office (PCSO) for fine-tuning the CNN to adapt the content of face to estimate human age. Furthermore, for evaluation purpose, CNN has been tuned on the training portion of the each datasets. The performance can be enhanced through fine-tuning which allows the CNN to choose the particularities, the bias of each dataset, and the distribution.

To quantify the results $\epsilon - error$ is considered, there is no proper background of this approach where only a bunch of observers can vote. The standard deviation is considered for the predicted age of the subject voted by the observer. The maximum votes decides the label of prediction, the mean age μ and stand deviation σ calculate the error as:

$$\epsilon = 1 - e^{\frac{-z-\mu}{2\sigma^2}} \tag{1}$$

The final value of $\epsilon - error$ is calculated by taking the average value of errors. The value of $\epsilon - error$ is between 1 and 0 (wrong prediction is 1 while correct prediction is 0).

Learning Model	O/P neuron	Pre-traning with IMDB-WIKI			
		Uniform		Balance	
		MAE	e-error	MAE	e-error
SVR on Conv 5_3		4.57	0.41		
SVR on fc6		3.69	0.32		
SVR on fc7		3.67	0.32		
Regression	1	3.65	0.31		
Classification	10	4.24	0.38	3.91	0.30
Classification	50	3.56	0.29	3.51	0.30
Classification	101	3.52	0.30		
Expected Value	10	3.55	0.31	3.50	0.29
Expected Value	50	3.34	0.29	3.31	0.28
Expected Value	101	3.25	0.28		

Table 1
The number of out neuron varies accordingly while conv5 3 (100,352 dim) is the last convolutional layer. fc6 (4,096 dim) and fc7 (4,096 dim) are last fully connected layers

3.1.2. Output layer and expected value

For each object class, the pre-trained CNN model (VGG-16) with output layer of 1000- softmax normalization neurons is used for classification task. Meanwhile, human age estimation is not a classification task. The last layer is replaced with only 1 output for regression and Euclidean loss function has been deployed. Considering the higher error,it is not possible to train a CNN directly for regression problems. Its also relatively hard for the CNN to converge the unstable prediction with a larger gradient. The mean value is calculated for all the training samples in the proposed age range Y_a. For our experiment, we considered a uniform distribution of training samples. The training set is organized according to age range to provide enough data as needed for finer training. The CNN is trained in the same manner as for classification tasks. The value expected for softmax-normalized output probabilities of $|A|$ number of neurons:

$$Z(B) = \sum_{i=1}^{|Y|} t_i \tag{2}$$

Where B = 1,2,.....$|A|$ is a $|A|$ dimensional output layer and for output probability of neuron i the value of softmax-normalization is $B_i \in B$.

3.1.3. Deployment details

In table 1, the performance of Support Vector Regression (SVR) is presented with Radial Basis Function (RBF) kernel, from pooling layer (Conv 5 3), penultimate fc6 and final (fc7) fully connected layer of the deep model with and without pretraining on IMDB-WIKI database. The specialized layers provided higher accuracy as compared to pooling layer which is projected for age estimation. While pooling provides better performance using conventional methods.

DEX for age estimation is vulnerable to spoof attacks as it is using a single image for human age estimation as shown in figure 1. The baseline performance under licit scenario for human age has FAR value of 5.3%, while for zero-effort imposters it became more higher than 19.7% and under spoof attack its nearly 38.9%. Considering the vulnerability of the proposed system for human age estimation using facial, we presented an anti-spoofing system which is successfully demonstrated for human age estimation using auditory perception based [4]

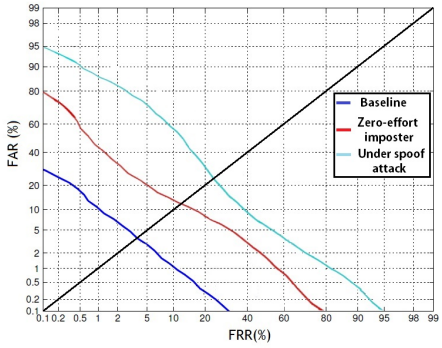

Figure 1: Performance evaluation of the baseline system, using zero effort imposter and under spoof attacks

3.2. Human age estimation using auditory perception and RaS-DeeP

3.3. Age estimation

The protocol for human age estimation based on auditory perception responses is briefly described in Figure 2. The auditory system of a subject is stimulated through a dynamic frequency (frequency changes with a constant proportion) sound, and as a result the audible frequencies have been saved in database. An artificial intelligence based system is estimating the age according to registered frequencies in the database.

3.3.1. Protocol of stimulation

According to the following model, a dynamic frequency sound is generated to stimulate the auditory system:

$$A(t) = A_0 \sin(2\pi\phi(t)t) \tag{3}$$

Where

$$\phi(t) = \alpha t + \phi_1 \tag{4}$$

$A(t) \in \mathbb{R}$, $A_0 \in [0, A_{\max}]$, where $A_0 \in \mathbb{R}^+$ stands for sound amplitude, $\phi(t) \in [f_{\min}, f_{\max}]$, where $f_{\min}, f_{\max} \in \mathbb{R}^+$ and $t, \phi_1, \alpha \in \mathbb{R}^+$ stands respectively for, time, the initial frequency, and the speed of frequency.

In our experiment, the sound is generated over a duration T. A user is required to communicate in real time with the system such that subjects should respond in real-time when they stop hearing or start hearing the dynamic frequency. The experiment is based on two tests:

- Test-one: the sound is generated from lower frequency $f_{\min}$ to higher frequency $f_{\max}$ and subjects should respond while unable to perceive sound.

- Test-two: the second test starts automatically. In this case, the sound is generated from a higher frequency $f_{\max}$ to lower frequency $f_{\min}$ and subjects should respond while start perceiving a sound.

The resultant frequencies are saved in a database (first test frequency and second test frequency) for further processing to estimate human age using the random forest classifier.

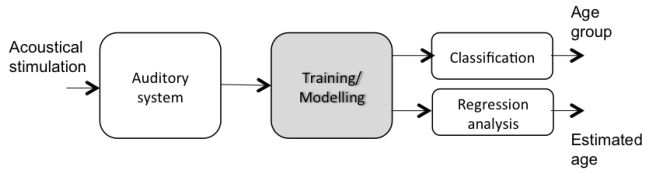

Figure 2: Flow diagram of the proposed for human age estimation and classification

3.3.2. Performance and Vulnerabilities to spoofing

The system based on auditory perception responses for human age estimation and classification is vulnerable to spoof attacks such that an adult user misleads the system by ending the experiment with a high-frequency in the first test, and respond after some seconds as the second test of experiment. Similar scenario can be considered when a young user impersonates an adult . In order to overcome the problem spoof-attacks, an anti-spoofing system is designed for the system based on auditory perception [4].

In Figure 3, the performance of the proposed system for human estimation is shown under licit scenario (i.e. baseline with no spoof attacks)and under spoof scenario (i.e. baseline with spoof attacks). The performance of a biometric system can be evaluated through the value of Equal Error Rate (EER). Low EER value indicates describes high accuracy system. The EER value under licit scenario is about 2% and under spoof scenario, the EER value is raised to 60%, which provides the evidence of the vulnerability of the proposed age estimation system based on auditory perception.

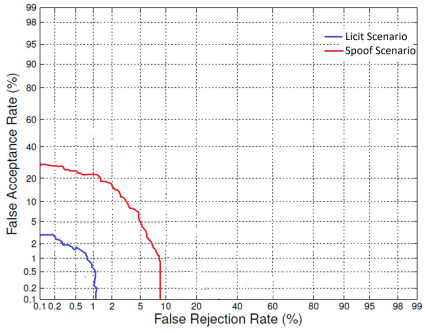

Figure 3: Performance evaluation of auditory perception based system under Licit scenario vs spoof scenario

3.4. RaS-DeeP for age verification

RaS-Deep is successfully demonstrated in previous work [4]. Hence, we are presenting RaS-Deep for age verification according to Algorithm 3.4.

A subject requires to take the test for age estimation, using face based system as shown in Figure 6. A camera captures a photo and uses it as an input while artificial intelligence based system estimates the age accordingly. Our proposed anti-spoofing system uses the estimated age as an input for age verification. Ten random sounds are generated with different pre-defined frequencies, following a uniform distribution. The intensity of sound generated by the anti-spoofing has hearable and unhearable frequency sounds according to the input age of the subject. To enhance the performance of the system, some of the hearable frequencies are randomly repeated to ensure that the test subject provides legitimate feedback. According to our standard database, the minimum and the maximum value of hearable and unhearable sound frequencies for each age is assigned. The hearable and unhearable frequencies are unpredictable for the spoof attacker among the generated sounds.

The response of the subject for each frequency sound has a specific value $1/u$ in the final score, where u is the number of sound frequencies, randomly generated as shown in Algorithm 3.4. If the final score is greater than the decision threshold τ, the subject will be considered genuine and verify the input age.

The decision threshold τ have been selected according to the standards given by:

$$\tau^*_{EER} = arg.min|FAR(\tau, D_{dv}) - FRR(\tau, D_{dv})| \quad (5)$$

The decision threshold τ is the balanced value (EER) between False Acceptance Rate (FAR) and False Rejected Rate (FRR). FAR is the ratio of the number of false acceptance divided by the number of verification attempts while FRR is the ratio of the number of false rejection divided by the number of verification attempts. The performance of the proposed algorithm can be demonstrated as a function of decision threshold τ_{EER}. A test subject is required to take the test for age estimation using auditory perception. As the auditory perception based system estimated the age, our proposed anti-spoofing system would verify the age of the test subject. The system is designed to generate ten random frequencies of sound according to our standard database by taking the estimated age of the subject as an input. Among these ten sound frequencies, some of them are audible and some of them are inaudible for the test subject. To make it more secure against spoofing attacks, some of the audible frequencies are repeated to ensure that the test subject provides the same feedback. According to our previous study, the minimum and the maximum values of audible and inaudible sound frequencies for each age are assigned from a reference database. One can note that it is difficult for a spoof attacker to guess the audible and inaudible sound frequencies in the set of generated sound frequencies. Every

feedback for each generated sound frequency has a value of $1/b$ to calculate the final score, where b is the total number of randomly generated sound frequencies, as shown in Algorithm 3.4. The final score must be greater than a decision threshold τ to prove that the subject is genuine and verify the input age. More details about the algorithm is given in Algorithm 3.4.

[**RaS-Deep for age verification**] [3.4]

procedure OUTPUT: AGE VERIFICATION
    x=*actual-age*,
    y= *nbr-freq*,
    z= *nbr-randomn-hearable-freq*,
    b=*nbr-repetition-hearable-freq*,
    Input: $x, y, z, b, e(\tau_{EER})$
    $x \leftarrow$ *insert()*
    F3, *F4* = [*min-f(x)*, *max-f(x)*]
    F1, *F2* = [*min-f(x)*, *max-f(x)*]
    TAB : *rand-f(y)*
    for *i=1*:*y* **do**
        TAB[i]=0
    end for
    hearable-indices [ ] $\leftarrow z$
    for *i=1*:*length(hearable-indices)* **do**
        TAB[hearable-indices(i)] = (F1,F2)
    end for
    for *i=1*:*b)* **do**
        $b \leftarrow$ *rand[ ]*
        b = TAB[position]
    end for
    for *i=1*:*y)* **do**
        if *TAB[i]* $\neq 0$
        TAB[i] = (F3,F4)
        endif
    end for
    for *i=1*:*y* **do**
        play sound(<TAB[i])
        user feedback[i] $\leftarrow$ *ask-user-feedback* >
    end for
    nbr-correct-anwers $\leftarrow$ *verify-feedback(user-feedback[])*
    if *(nbr-correct-answers*$\geq \tau_{EER}$*)*: *verified-age* $\leftarrow 1$
    else
    verified-age $\leftarrow 0$
    end

4. Performance Evaluation of RaS-Deep as a countermeasure

4.1. Datasets collection

Three datasets are collected as mentioned in Table 2 under two scenarios, while 1140 male subjects and 470 females subjects participated in our experiment under spoof scenario and licit scenario by using the algorithm 3.4 :

-The **development dataset** decides the optimized point of decision threshold for better performance on an operational value. With both male and female subjects 450 trails are conducted.

- The **anti-spoofing dataset** is used to assess the vulnerability of the Ras-Deep with the proposed threshold τ value. While 570 trails were conducted for anti-spoofing dataset.

- The **Countermeasure dataset** is a two-class dataset, to identify genuine and spoofed trials. Although 630 trials have been conducted by all the subjects (both male and female) in the age range of 12 to 51 years. The test set is utilized to determine the efficiency for a specific threshold computed through the development set. The test set is equally distributed into enrolment and probe subsets.

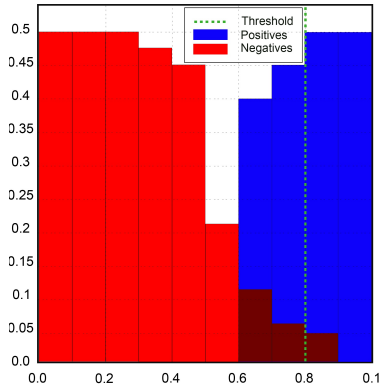

Figure 4: Score distribution of genuine users and zero-effort imposters and the decision threshold τ is presented with a dashed line.

4.2. Optimization of decision threshold value

To optimize the decision threshold, the development data set is used. The biometrics system's baseline performance is tested under licit scenario with zero-effort impostor and genuine trails. The decision threshold τ at 80% shows the minimum misclassification of impostor trials as genuine [4]. Efficiency of the proposed biometric system can be demonstrated with the Detection Error Trade-off (DET) profile as shown in Figure 3. Under the licit scenario the EER value 3.3% represent by blue line while baseline under spoof attack is represented by red line, where the FAR increases exponentially.

The predetermined threshold is utilized to estimate the efficiency of the system using a test set D_{test} (FAR,FRR, and Half Total Error Rate (HTER) of D_{test}) [4]. Equal Error Rates (EER), defined in the ROC or DET curve as a point where the FAR is equal to the FRR, are the performances of the system while at 0.5 yields, representing the Half Total Error Rate (HTER).

5. RaS-DeeP as countermeasure for DEX

Anti-spoofing systems, countermeasures, several combinations of countermeasure approaches and algorithms are attracting the researchers. It provides an assistance to researcher to introduce new ideas to build novel countermeasures. While last development in robustness to particular field, fused countermeasures allowed an adjustable anti-spoofing structure by which recently developing vulnerabilities can be immediately covered with flexible fusion approaches or

Datasets	Development dataset		Anti-Spoofing dataset		Countermeasure Dataset	
Gender	Male	Female	Male	Female	Male	Female
Licit Scenario	230	100	250	110	380	110
Spoofing Scenario	70	50	160	60	100	40
Total trials	450		570		630	

Table 2
Datasets Collected for different scenarios

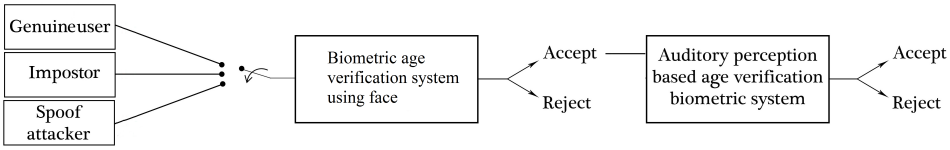

Figure 5: Spoofing countermeasure integrated with biometric system

the extension of modern countermeasures. The fusion of two biometric modalities, facial features and auditory perception approaches are used for human age estimation as a countermeasure. As shown in Figure 5, countermeasure sub-systems is alternatively integrated, with the biometric system. Countermeasure is consist of two steps:

- Human age verification using facial features
- Human age verification using auditory perception responses

To demonstrate the performance of the proposed face based human age estimation system, we used Diverse Fake Face Dataset (DFFD) [2] to train deep learning model. The DFFD dataset consists of several publicly accessible data sets and photographs which are extracted / exploited using methods open to the public. Through integrating numerous outlets for actual photos, we are able to use different resolution and picture consistency with both natural and synthetic / manipulated pictures. The proliferation of face recognition, biometric systems activation, and social networking is a major opportunity for malicious actors to incorporate fake or distorted photographs to distribute incorrect facts for someone's reputations to harm. It helped by the quality management in the synthesis and manipulation of realistic image by the techniques of generative adversarial network. The DFFD dataset incorporates several forms of fraudulent in one set of database such as facial attributes manipulation, entire face synthesis, face expression swap, and face identity swap.

We considered 50% of the data into training, 5% for validation and for testing 45% , while an subject in one partition does not exist in the other partition such that the testing and training subject are different [14].

Experimental setup for countermeasures: The proposed experiment for countermeasure consists of three steps:

- The subject is required to use the face based system for age estimation as shown in figure 6
- The subject will again verify his age by using RaS-Deep using the approach of algorithm 3.4.

All the results are saved in a database in order to compare the final results. If the required responses are equal or greater than the decision threshold τ, the user is genuine and vice versa.

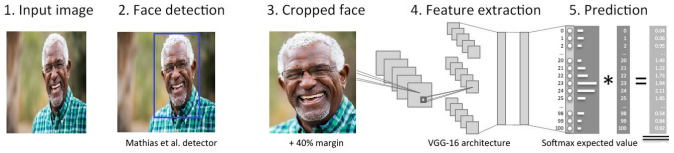

Figure 6: Pipeline of DEX architecture [11]

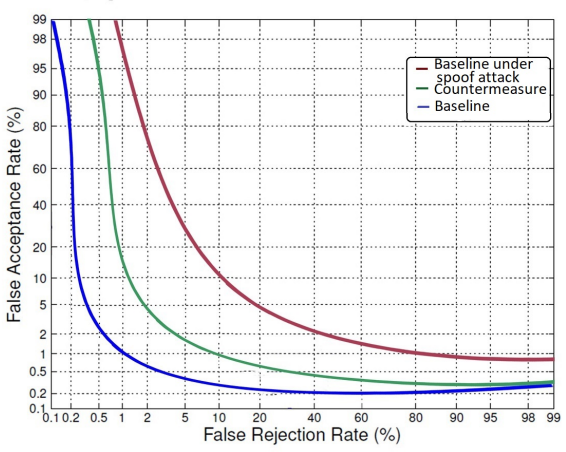

Figure 7: The Detection Error Trade-off (DET) profiles are obtained for baseline (red line), (green line) demonstrates the efficiency of RaS-Deep constrained to spoof attacks, and (blue line) demonstrates the efficiency of the countermeasure biometric system under spoof attack

5.1. Comparative study

We compared our method to the state of the art methods as shown in table 5.1. For Long Short-Term Memory

Method	EER(%)	HTER
Fine-tuned VGG-Face [7]	7.33	
LSTM-CNN [16]	8.11	11.93
Boulkenafet et al [1]	6.20	7.44
Siddique et al [13]	5.37	
Yang et al [17]	5.99	
Fusion (DEX+RasDeep)	3.33	

(LSTM), CNN the temporal characteristics are used [16], the holistic characteristics are extracted for classification and CNN is used for feature extraction and SVM is used for classification after applying PCA to the response of the last layer. Some state-of-the-art approaches, along with a classifier, use handcrafted features.

As shown in table5.1, the performance of our proposed system outperforms others by comparing the EER values. It proves that the fusion of both facial features and auditory perception responses system (RAS-Deep) contains more discriminative details.

We also tested other architectures from the state of the art with the proposed database. Actually, we compared our method performances with other state of the art methods which analyze the distortions in spoof images, concatenated representation and color moment. While it is mentioned in a very few articles. In comparison to [17, 13], our proposed system achieved 3.33% while other systems gave lower accuracy for the proposed database.

5.1.1. Evaluation protocol

The experimental setup of the proposed anti-spoofing countermeasure is shown in Figure 5, while Figure 7 provides the Detection Error Trade-off (DET) curve with different profiles under spoof attack. A clear view of efficiency's of different systems have been demonstrated by calculating the EER value. To generate the DET profiles, three different kind of system configurations are involved for evaluation. The first configuration (blue profile) demonstrates the efficiency of the countermeasure biometric system (under spoof attack). While the second configuration (green profile) demonstrates the efficiency of RaS-Deep constrained to spoof attacks. The third configuration (red profile) demonstrates the baseline under spoof attack.

Whatever the modality used for integration, the proposed system achieved promising results to secure biometric systems from spoof attackers. The EER value is 4.2% for the countermeasure under spoof-attack.

6. Multimodal age estimation

For human age estimation, a unimodal biometric system using auditory perception is already designed showing promising results [5] as compared to face based age estimation system [11]. We demonstrated several applications hlusing human age estimation based on auditory perception responses [6]. In this article, the compatibility of the auditory perception based system for human age estimation is demonstrated in integration with face based system.

Years	12-20	21-35	36-50	51-65
Male	9	8	11	10
Female	8	6	8	11
Total number of subjects	17	14	19	21

Table 3
Dataset for testing in real-time

We proposed a multimodal age estimation system to enhance the performance of the existing biometric systems for human age estimation. We created a fusion of two biometric modalities such as face and auditory perception responses as shown in figure 8.

An image is provided to face based human age estimation system as an input, and the deep learning trained model estimates the age accordingly while an auditory perception based system for human age estimation is also executed in parallel to estimate the age of the subject through auditory perception responses ($F_{min} = 20hz$, $F_{max} = 20,000hz$, $t = 20sec$). Both the outputs (estimated age using face and estimated age using auditory perception) are then fused together to provide the final results of estimated age of the proposed subject.

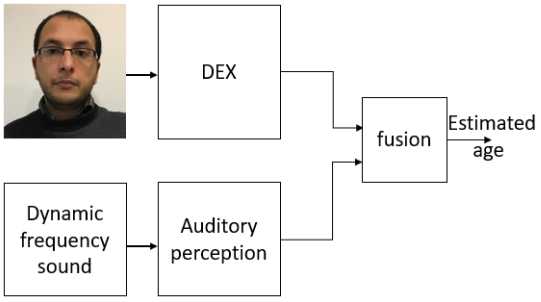

Figure 8: Diagram of multimodal approach for human age estimation

6.1. Performance evaluation

The trained facial feature based human age estimation is tested in real time with 71 subjects from different age range (12-67 years) with five repetitions, while the same subject conducted the test for human age estimation using auditory perception responses as shown in table 3.

In figure 9, the comparative evaluation is presented. The facial features based human age estimation system shows promising results with an EER value of 5.43% represented by the blue line, while auditory perception based system achieved EER value of 2.6% represented by the black line.

Furthermore, the fusion of face and auditory perception based system for human estimation enhances the performance of the facial based age estimation, as shown in figure 9 . The fusion based system achieved the EER value of 3.3%, it results in an overall improvement complementariness of fusion further reduces the chances of error that is compared favorably to the state-of-the-art.

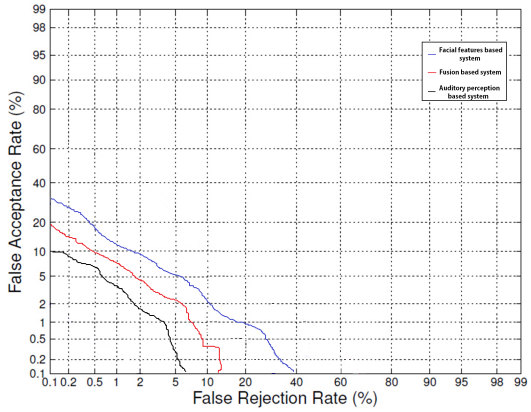

Figure 9: Performance evaluation of multimodal biometric age estimation system

7. Conclusion

In this paper, two main contributions have been highlighted. The first contribution consists of preventing face-based age estimation systems from being attacked. This has been achieved through Ras-DeeP, as a countermeasure based on a random auditory stimulation's. In this study, we found that Ras-Deep provided promising results against spoof-attacks as EER value of 4.2% is achieved. As a second contribution, when the auditory stimulation is considered with face age estimation (DEX), in a multimodal biometric system, better accuracy has been obtained. In particular, EER decreases from 5.43% (using face age estimation), to 3.3% when considering the multimodal system. Obviously, Ras-DeeP is still an emerging approach, but we believe that by optimizing a certain number of parameters, such as the stimulation mode, considering different environment effects, one can enhance the performance of the Ras-Deep. Probably, it will be a challenge for future works.

References

[1] Boulkenafet, Z., Komulainen, J., Hadid, A., 2015. Face anti-spoofing based on color texture analysis, in: 2015 IEEE international conference on image processing (ICIP), IEEE. pp. 2636–2640.

[2] Dang, H., Accessed: 15-08-2020). DFFD: Diverse Fake Face Dataset. URL: http://cvlab.cse.msu.edu/dffd-diverse-fake-face-dataset.html.

[3] Duc, N.M., Minh, B., 2009. Your face is not your password, in: Black Hat Conference.

[4] Ilyas, M., Othmani, A., Fournier, R., Nait-ali, A., 2019. Auditory perception based anti-spoofing system for human age verification. Electronics 8, 1313.

[5] Ilyas, M., Othmani, A., Nait-ali, A., 2020a. Auditory perception based system for age classification and estimation using dynamic frequency sound. MULTIMEDIA TOOLS AND APPLICATIONS .

[6] Ilyas, M., Othmani, A., Nait-ali, A., 2020b. Computer-aided prediction of hearing loss based on auditory perception. MULTIMEDIA TOOLS AND APPLICATIONS .

[7] Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A., 2016. An original face anti-spoofing approach using partial convolutional neural network, in: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), IEEE. pp. 1–6.

[8] Mail, D., January 2015. Accessed: 15-02-2020). Real face spoofing case. URL: http://www.dailymail.co.uk/news/article-1326885/Man-boards-plane-disguised-old-man-arrested-arrival-Canada.html.

[9] Nikisins, O., George, A., Marcel, S., 2019. Domain adaptation in multi-channel autoencoder based features for robust face anti-spoofing, in: 2019 International Conference on Biometrics (ICB), IEEE. pp. 1–8.

[10] Raghavendra, R., Raja, K.B., Busch, C., 2015. Presentation attack detection for face recognition using light field camera. IEEE Transactions on Image Processing 24, 1060–1075.

[11] Rothe, R., Timofte, R., Van Gool, L., 2018. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision 126, 144–157.

[12] Schwartz, W.R., Rocha, A., Pedrini, H., 2011. Face spoofing detection through partial least squares and low-level descriptors, in: 2011 International Joint Conference on Biometrics (IJCB), IEEE. pp. 1–8.

[13] Siddiqui, T.A., Bharadwaj, S., Dhamecha, T.I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N., 2016. Face anti-spoofing with multifeature videolet aggregation, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE. pp. 1035–1040.

[14] Stehouwer, J., Dang, H., Liu, F., Liu, X., Jain, A., 2019. On the detection of digital face manipulation. arXiv preprint arXiv:1910.01717 .

[15] Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. Deepface: Closing the gap to human-level performance in face verification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701–1708.

[16] Xu, Z., Li, S., Deng, W., 2015. Learning temporal features using lstm-cnn architecture for face anti-spoofing, in: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), IEEE. pp. 141–145.

[17] Yang, J., Lei, Z., Li, S.Z., 2014. Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601 .