Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier XXX/XXX

iFace 1.1: A Proof-of-Concept of a Facial Authentication Based Digital ID for Smart Cities

ALAKANANDA MITRA[1], (Graduate Student Member, IEEE), DAN BIGIOI[2], SARAJU P. MOHANTY[1], (Senior Member, IEEE), PETER CORCORAN[2], (Fellow, IEEE), and ELIAS KOUGIANOS[3], (Senior Member, IEEE)

[1]Dept. of Computer Science and Engineering, University of North Texas, USA
[2]College of Science and Engineering, (C3I Research Centre), National University of Ireland, Galway, Ireland, (email: d.bigioi1@nuigalway.ie)
[3]Dept. of Electrical Engineering, University of North Texas, USA

Corresponding author: Dan Bigioi (e-mail: d.bigioi1@nuigalway.ie).

ABSTRACT "Smart Cities" are a viable option to various issues caused by accelerated urban growth. To make smart cities a reality, smart citizens need to be connected to the "Smart City" through a digital ID. A digital ID enables citizens to utilize smart city facilities such as healthcare, transportation, finance, and energy with ease and efficiency. In this paper, we propose a proof-of-concept of a facial authentication-based end-to-end digital ID system for a smart city. Facial authentication systems are prone to various biometric template attacks and cyber security attacks. Our proposed system is designed to detect the first type of attack, especially deepfake and presentation attacks. Users are authenticated each time they use facilities in a smart city. Facial data is stored in the cloud in a lookup table format with an unidentifiable username. The process is very secure as no data leaves the device during authentication. Our proposed solution achieved 97% accuracy in authentication with a False Rejection Ratio of 2% and False Acceptance Ratio of 3%.

INDEX TERMS Smart City, Digital ID, Internet-of-Things (IoT), Deepfake, Presentation Attack, Facial Authentication System, Convolutional Neural Network, Triplet Loss, Lookup Table

I. INTRODUCTION

The world's population is increasing at an unprecedented rate. It is estimated that 70% of the global population will be living in cities by the year 2050 [1]. Such rapid urbanization will create more carbon emissions and pollution which in turn will negatively impact the environment and people's health. It will also create a greater demand for energy, food, and resources. Smart cities have emerged as a resilient and sustainable solution to the problems caused by rapid urbanization. They are envisioned as the future of urbanization, where residents of such cities can benefit from smart transportation, health care, energy, and other seamlessly integrated services which are all connected through the Internet of Things (IoT). Fig. 1 shows how smart city stakeholders are connected.

In the last twenty years, the idea of a smart city has stepped forward due to advancements in hardware and software, growth in information and communication technologies (ICT), and initiatives offered by various tech giants [2].

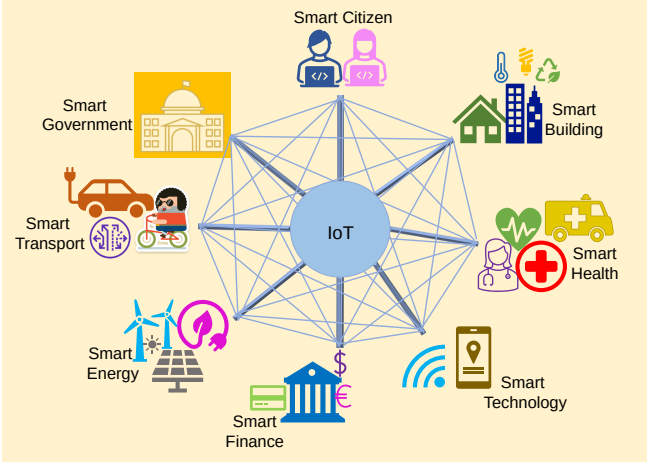


FIGURE 1. Components of Smart City

For a practical implementation of the smart-city digital infrastructure, citizens should have a way to easily connect to the amenities offered by the smart city. A wide range of services, such as smart healthcare, smart government and smart financial products should be accessible to the citizens. Checking bank accounts, ordering products online, using smartphones, driving electric vehicles, locking doors with smart locks, lighting homes with smart lights and paying at electronic parking meters are just some of the ways people are already transforming traditional towns into "smart cities." Therefore, in this paper, we propose a universal digital identity system that can unlock the full potential of a smart city by connecting its citizens to all its amenities in a simple but efficient manner.

The use of digital IDs has already started to become the norm in various parts of the world, e.g., India's *Aadhaar*, a biometric based identification number, the digital ID system in Australia, and the digital ID wallet initiative in the EU. Using digital ID in smart cities is a similar idea.

In recent years, a variety of cutting-edge facial recognition systems, various computing platform-based facial authentication systems, and threat detection approaches have been presented in various publications. Even though there have been many creative ideas, no end-to-end face authentication system that could be used for a digital ID system in smart cities has been offered.

In this paper, we propose a proof-of-concept for a facial authentication-based digital ID system that would enable citizens to effortlessly access smart city services. The proposed system detects both presentation and deepfake attacks with a high success rate. With this system, individuals may easily and effectively use the numerous smart city services that have been made available. This paper is an improved and extended version of the original work presented in [3].

The rest of this paper has been organized as follows: Section II presents digital ID and smart cities in detail, including the need for a digital ID in a smart city and the challenges of implementing the digital ID system. Section. III discusses the novel contributions of our work. Related works in a smart city context are discussed in Section IV. Section V depicts the proposed work. Performance of the face authentication system is presented in Section VI. Section VII winds up the paper with suggestions for future work.

II. DIGITAL ID AND SMART CITY

A. ROLE OF DIGITAL ID IN SMART CITY

In traditional cryptographic systems, secret keys are used to authenticate users. Often, users write down their secret keys, save them somewhere, share them with others, or simply forget them. Sometimes the keys are so simple or tied up with people's life events that they are easy to predict. The authentication system collapses if the secret key is no longer private.

On the other hand, biometrics-based digital IDs are person-dependent and discrete as biometrics represent physiological or behavioral traits of a person. With such an ID, there is no need to preserve the confidentiality of the key because the users themselves are the secret keys.

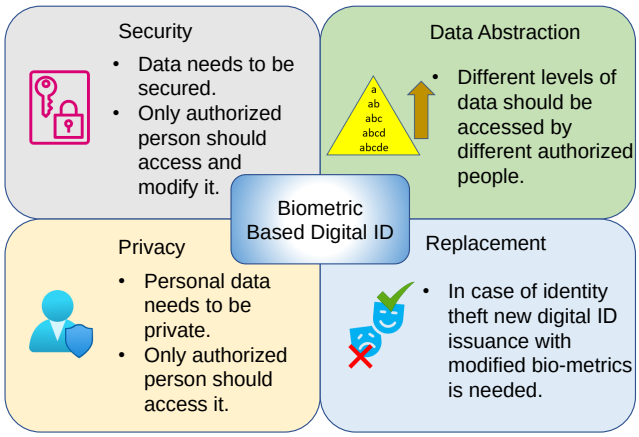


FIGURE 2. Mandatory Requirements for Digital ID in a Smart City.

Activities such as opening a bank account or availing of any age-restricted service require proper user verification. A digital ID can reduce the paperwork typically associated with such activities, giving citizens easy and efficient access to all the services provided by the smart city. The importance of using a digital ID in a smart city is multi-fold:

1) The digital ID system revamps the operational capacity of a city at a granular level.
2) Emergency medical services will be improved by introducing digital ID. If a critically sick person reaches a hospital without any traditional ID, a digital ID can save his lives. The doctor can access the patient's medical history and give proper medical care.
3) During any natural disaster, efficient verification of individuals' identities can be performed.
4) Digital IDs can facilitate improved administrative capacity for amenities around smart cities, e.g., banks, driver's licenses, retail, and transportation.
5) Online education is operating in parallel with the traditional brick and mortar schooling system. A Digital ID will offer a fair system with more flexibility to both students and educators.

B. CHALLENGES OF DIGITAL IDENTIFICATION

Building such a system is challenging in light of multiple considerations, including cyber security attacks, data privacy, security, and inclusion.

1) Vulnerability to Biometrics and Cyber Security Attacks

Biometric-based digital ID systems are prone to various biometric and cyber security attacks. As our proposed digital ID is facial biometric based, biometric attacks are our area of concern.

a: Presentation Attack

The proposed digital ID verification system is reliant on accurate face recognition. Facial recognition software is vulnerable to face spoofing or presentation attacks. The vulnerable points of a digital identity system (DIS) are listed in Table. 1 [4]. The first two attacks on the table are presentation attacks.

b: Deepfake Attack

Another common vulnerability of a digital ID system is the deepfake attack. Deepfakes are AI generated fake images or videos that do not exist and can easily fool human eyes. It is a type of presentation attack [5]. It also poses a serious threat to facial recognition systems [6]. With the rapid progress of deep learning, deepfakes are gaining the potential to fool even the best facial recognition systems.

Face swapping through Face Swapping Generative Adversarial Network (FSGAN) [7] eases the creation of deep fakes as there is no need to train the FSGAN for hours with source and target images. It means deepfakes can spread more quickly and easily than ever before, as people with a basic knowledge of this technology are now capable of creating them.

c: Indirect Attack

In the Table. 1, attacks from row 3 to row 8 are indirect attacks. These attacks target the cyber security system directly rather than biometric attacks, as they target databases, or channels, or even the device itself. In this paper, indirect attacks have not been addressed as they fall outside the scope of this work.

2) Privacy and Data Security

Fig. 2 depicts the conditions of a biometric-based digital ID. Two major concerns are the security and privacy of information. Data should be secured and obtained only by the titled person. Data abstraction should also be followed, where people with different levels of authorization access different levels of data. As digital ID is related to someone's biometric data, storing this data safely is another important factor. Hence, the design approach needs to be secure-and-private-by-design [8].

3) Exclusion

Ideally, everyone in the smart city should be enrolled in the digital ID system. However, there will be scenarios where, e.g., people may not want to be enrolled in the system; people may not be capable of providing the required biometrics due to physical disabilities; or they may not be digitally aware. In those scenarios, alternative traditional identification, e.g., a paper ID, is required.

4) Data Privacy Regulation Policy

Data privacy regulation policies are different across the globe. Europe has the General Data Protection Regulation (GDPR), whereas North America has different regulations like HIPAA, FERPA, COPPA, and FCRA for different types of consumer data. There are many more such regulations worldwide, and any physical digital ID system for a city would have to ensure that it adheres to all rules and regulations surrounding consumer data privacy. As the approach presented in this paper is simply a proof of concept for a facial authentication based digital ID system, the application of data privacy regulations for the hypothetical smart city falls outside the scope of this paper.

TABLE 1. Vulnerable points in DIS.

Serial No.	Type of Attacks	Real World Scenarios
1	Present fraud face biometrics at the camera.	Using a 3D face mask, 2D photo, a video clip of the attacked face.
2	Submit saved digital photo of face instead of using camera.	Resubmit earlier photo.
3	Trojan Attack during feature extraction.	Selection of predefined features by hacker.
4	Alter feature set after extraction.	When face matching is done at a different place than feature extraction, change of some packets in TCP/IP stack remotely.
5	Attack the matcher.	Matcher shows intruder defined scores.
6	Alter the database.	Any entry of the database can be changed.
7	Attack the channel carrying data from the database.	Intercept the channel and alter the data before it reaches matcher.
8	Change the final score.	Hacker can change the final result failing the FRS.

III. NOVEL CONTRIBUTIONS OF THE CURRENT PAPER

Here, the novelty of our work is discussed:

- The user registration process is remotely and securely performed on the edge device. Facial embeddings are extracted using the user's device in an offline application. These embeddings contain no identifiable information and are private by design as they are just a series of numbers generated by a neural network. However, they are encrypted for extra protection, and sent to be stored on a remote cloud server operated by the smart city.
- The authentication process is also done on the edge device. The photograph, face embedding extraction, and authentication of the user are all done on the edge. No facial data is sent to the cloud during authentication. It makes the process resilient to various indirect attacks.
- Due to its fast and easy deepfake generation technique, FSGAN deepfakes pose a serious threat to people. Our system is capable of detecting deepfakes created by FSGAN.
- The proposed system can detect the presentation attack. A check is performed each time a user tries to access any facility within the smart city by taking a picture during the time of access. This ensures the correct user is present and decreases the risk of presentation

attack. No data is stored anywhere in the system during authentication. It is performed as users come and go. With this feature, the risk of presentation attack on this kind of system goes down.
- No facial biometric data leaves the edge platform during the authentication phase. No data is stored during authentication. The only data that is stored is the reference facial embeddings extracted during enrollment. These are used to retrain the model on the cloud, and it never leaves the cloud storage at any time.

IV. RELATED WORKS

In the past few years, research on computer vision and pattern recognition has been boosted by significant advancements in deep learning techniques, new ways of thinking about parallel computing, and monumental developments in hardware. Because facial authentication is at the heart of our digital ID system, this section discusses some of the most recent work on the three main features of the proposed system: the FR system, face features-based authentication, and morphing attack detection (MAD).

A. FR SYSTEMS:

In the early days of facial recognition, a holistic approach [9] was used, such as image projection on low-dimensional space [10], Laplacianface [11], and sparse representation [12]. In the early 2000s, more local features-based face detectors [13], [14] were introduced. However, from 2012 onward high-accuracy deep learning-based techniques have been predominant.

In 2014, DeepFace [15] transformed the direction of facial recognition techniques. An accuracy of 97.35%, close to human accuracy, was obtained with this 9-layer deep neural model on the LFW dataset [16].

Deep neural network-based techniques have used diverse architectures, different loss functions, and various image processing techniques. In the same year as DeepFace [15], another paper [17] performed face verification using high-level features and deep ConvNets. Face features from different face parts helped the model to achieve a higher accuracy of 97.45% on the LFW dataset [16]. Both works used a cross-entropy-based softmax loss. However, the later versions [18] and [19] of DeepID used a Euclidean distance-based loss named contrastive loss. Here, absolute distances between image pairs are calculated.

Another face recognition model is FaceNet [20], where a new loss, triplet loss, was used for feature learning and clustering. The method had an accuracy of 99.63%. Triplet loss is another Euclidean distance-based loss where the relative difference in distances between matching pairs is considered. Another important facial recognition system was proposed in [21]. Here, marginal loss was proposed for deep face recognition with a comparable accuracy of 99.48% on the benchmark LFW dataset. It minimizes the intra-class variation and maximizes the inter-class distances simultaneously.

A pose invariant facial recognition technique was proposed using Disentangled Representation Learning GAN in [22]. This FR system has been evaluated for various illumination and angular positions of the face. Another competitive FR system is *CosFace* [23]. Large margin cosine loss (LMCL) has been introduced by redefining *Softmax loss* as a *cosine loss*. One of the state-of-the-art FR systems, *ArcFace* is presented in [24]. An Additive Angular Margin Loss has been proposed for face recognition. It is a highly accurate system. It achieved the highest accuracy of all the discussed FR systems on the LFW dataset.

After studying several of the aforementioned state-of-the-art FR systems, FaceNet [20] has been chosen as the FR system in the proposed digital ID. FaceNet [20] offers high accuracy as well as ease of application on the edge platform.

B. FACIAL FEATURES BASED AUTHENTICATION SYSTEMS:

In the last few years, various cloud-based, cloud-edge-based, and IoT mobile device-based face authentication systems have been researched in literature.

A detailed survey has been made for face verification and authentication for IoT mobile devices in [25]. In [26], a face verification system for a mobile device, from face registration to face verification, was proposed using light normalization and information fusion. However, no security measures were undertaken. Another work [27] proposed biometric-based security for IoT infrastructure using pairing-based cryptography.

A face verification technique for mobile devices is presented in [28]. The Viola-Jones detector has been used for face detection and subspace metrics for authentication. It has a low error rate. But no security measures were implemented. Another facial feature based active authentication technique for mobile phones has been proposed in [29]. A short video is used as the input of the face verification system. The detection rate showed high accuracy when authentication and enrollment were done from the same session videos. It is not suitable for real-time use where different session data needs to be compared. Similarly, a face authentication system for mobile devices is implemented in [30]. Here, face detection has been performed with Haar-like features and the AdaBoost algorithm. Face authentication, on the other hand, has been done with a local binary pattern.

Another mobile-friendly deep learning based face detector has been proposed in [31]. Various illuminations and extreme poses were considered. Without CUDA, mobile GPUs have been used to implement deep neural network models. [32] presents a fingerprint and face template based method. The face verification is SVM based, whereas the fingerprint verification is minutiae based. According to the authors, the "Secure sketch" cryptography and geometric translation make the method forgery-free. An enhanced biometric capsule-based authentication method was proposed in [33]. MTCNN [34] has been used for face detection and FaceNet [20] has been used for face feature extraction. A deep learning-based

face verification method has been proposed in [35] for an IoT-cloud setting. The face verification part is done by a tree-based cloud model. The edge part is optional for processing and filtering images.

The majority of the aforesaid facial authentication systems lack security measures. However, for any facial authentication system, security and privacy are the two critical criteria that need to be fulfilled. These two factors have been prioritized in the design of the proposed digital ID.

C. ATTACK DETECTION:

In this subsection, we discuss papers addressing various attacks on facial recognition systems. User liveness has been addressed in a majority of the papers.

For face authentication, an acoustic sensor based liveness detection method has been proposed in [36] with an accuracy of 96.02% and a false alarm rate of 3.97%. It uses the unevenness of the stereo structure of a real face to check the liveness of the user. Another liveness detection method has been proposed in [37] using photoplethysmograms of two simultaneous videos of the face and fingertips.

Some papers also focus on deepfake detection systems. In [38] and [39] deepfake videos have been detected. These systems, which are based on convolutional networks, have a high level of accuracy. Dynamic lip movement analysis has been done in [40] to detect deepfake attacks. In [41], the potential threat of face swapping to electronic Know Your Customer (eKYC) has been discussed, and a detection system has been proposed. Another IoT-friendly deepfake detection method has been described in [42]. The LightGBM classifier has been used to classify the images based on features from the Gray Level Co-occurrence Matrix (GLCM). In [43], an anti-spoofing facial recognition system has been proposed. COTS RFID tag array has been used to extract biometric features of the face and 3D geometry. 95.7% success rate is achieved with 4.4% EER. Most of the papers in this area also focus on a specific attack, but not in the context of facial authentication systems.

The discussion above shows that in the last few years, a number of state-of-the-art FR systems, different cloud/edge-cloud/mobile based facial authentication methods, and different attack detection techniques have been developed. Regardless of all the state-of-the-art methods, no end-to-end facial authentication system that can be used for implementing a digital ID system for smart cities has been proposed. Our objective is to propose a proof-of-concept of a viable but simple facial authentication based digital ID system for smart cities with high success rates in detecting attacks and authenticating smart city citizens.

V. iFACE 1.1 : PROPOSED FACIAL AUTHENTICATION BASED DIGITAL ID SYSTEM OF SMART CITIES

iFace 1.1 is the proposed digital ID system for smart cities. Once facial authentication is performed, residents will be able to use different city services.

A. SYSTEM LEVEL ARCHITECTURE

iFace 1.1 is presented as a four-layered architecture which is spread between edge and cloud platforms, as shown in Fig.3.

1) Layer-1: The input layer of the authentication system is denoted as layer-1. At this level, smart city residents capture photos using the cameras attached to their personal devices. The photo is then sent for authentication into layer-2. Users also provide their username.
2) Layer-2: Layer-2 is the edge computing platform. The photo, along with the username from the previous layer, works as the input for this layer. No data is transmitted at this stage over the open communication channels as both the layers are located on the same device. Hence, no encryption of facial data is required at this point. Layer-2 handles the data processing and computing parts of iFace 1.1.
3) Layer-3: The cloud computing platform is the third layer. Long-range communication technologies, e.g., *4G* and *LTE* connect the edge devices to this layer. During registration, biometric data is transmitted to this layer after being encrypted at layer-2. Encrypted biometric data and usernames are stored in this layer.
4) Layer-4: The last layer consists of smart city stakeholders like various smart city facilities. Residents can avail different amenities after they are authenticated with the iFace 1.1 system.

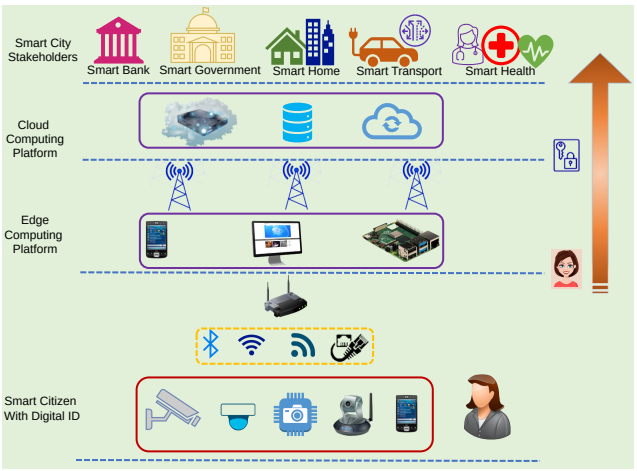


FIGURE 3. End-to-End System Level Framework of iFace 1.1

B. SYSTEM OVERVIEW

The aim of the paper is to propose a proof-of-concept of a working end-to-end digital ID system that can be implemented in a smart city. The digital ID system works in two phases:

- Registration/Enrollment Phase
- Authentication Phase

1) Enrollment Phase

The registration process of a new applicant is shown in Fig. 4. In this phase, an existing government-issued ID is checked to avoid any forgery, and then a unique username is provided to the new applicant.

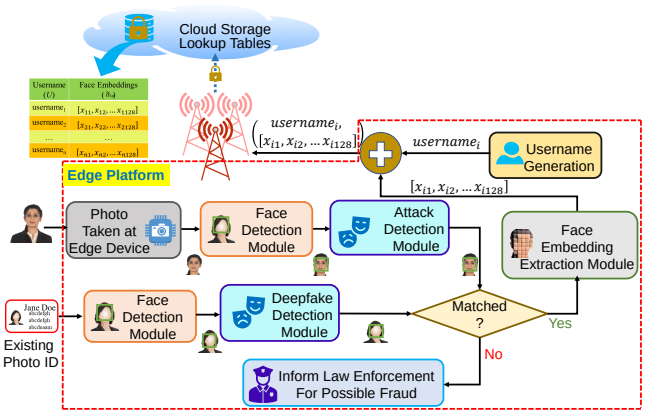


FIGURE 4. New User Registration

An edge device camera takes a neutral frontal face (NFF) image. Next, the face is detected in the image by the face detection (FD) module. The photo is then checked for deepfake and presentation attacks in the Attack Detection (AD) module. If the AD module does not verify the photo, law enforcement authorities are notified of possible fraud. For legitimate applicants, a facial embedding is extracted from the detected face in the photo. The embedding, along with the username, is added to a *lookup table* on the cloud server of the smart city. Data encryption may be added at this step.

2) Authentication Phase

Authentication is done on the edge, as in Fig. 5. The user provides their username, and a photo is captured by the edge device. Once the face is detected in the FD module, it is checked for any possible attack in the AD module. Next, the facial embedding is extracted, and the system predicts the username associated with the face embedding. If the predicted username is matched with the input username, access to the facility is granted; otherwise, it is denied.

C. SYSTEM MODULES

The system pipeline consists of various modules. Each module serves a specific purpose.

1) Username Generation (UG) Module

The UG module is an important module where unique and user-specific usernames are generated during registration. Fig.6 shows a sample username generation module workflow. During registration, the applicant provides their name. A unique username is generated as per the sample workflow.

During enrollment, the generated username and facial embedding extracted from the photo are sent to the cloud data storage and added to the *lookup table*. Data encryption should be used in each step to make the process more secure. Fig. 7 shows a sample *lookup table* before encryption. The stored data is used to train the classifier from time to time.

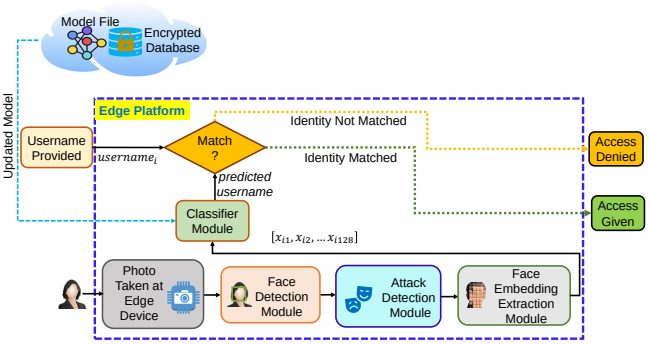


FIGURE 5. Authentication at Edge

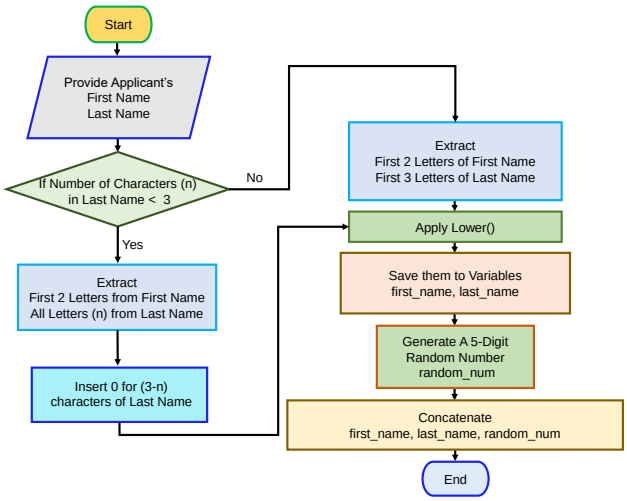


FIGURE 6. Username Generation Module Workflow

Lookup Table	
Username	**Face Embedding**
abcde02846	$[x_{11}, x_{12}, \ldots x_{1128}]$
fghij96784	$[x_{21}, x_{22}, \ldots x_{2128}]$
…	…
klmn069001	$[x_{n1}, x_{n2}, \ldots x_{n128}]$

FIGURE 7. Sample Lookup Table

2) Face Detection (FD) Module

Accurate face detection is the first step of face authentication. There are mainly two types of state-of-the-art face detection methods: deep neural network based and handcrafted features based or machine learning-based. Deep neural network

based face detectors, e.g., Multitask Cascaded CNN [44], RetinaFace [45], Fast RCNN [46], Faster RCNN [47], Mask RCNN [48], YOLO [49], and SSD [50], have emerged as successful face detectors. They are more accurate and robust. However, the majority of them are heavy and poorly suited for deployment on an embedded device.

In our work, a lightweight, fast, and efficient model is required. Hence, we chose a machine learning-based face detector over a deep neural network-based one. For security and data integrity purposes, the face detector should not work under any face occlusion. Face detection from the frontal face photo is required. Hence, the Viola-Jones Haar Cascade face detector [51] has been chosen as the face detection module from [14], [51]–[53]. It detects the face from the photo. It is tiny in size ($\sim$ 1MB) so a good fit for an IoT environment where resources are limited. It does not work under conditions where the face is occluded, which is a mandatory condition of the system to avoid any fraudulent activity. The Dlib HoG [54] face detector also works with mostly frontal faces. But, the extracted face by Dlib HoG mostly excludes the forehead of the detected face, as shown in Fig. 8. It is not desired for face authentication purposes.

Therefore, a frontal face image is necessary each time for security reasons. But for practical implementation, an alternative state-of-the-art suitable method can be used.

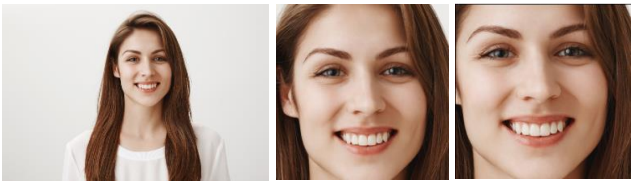


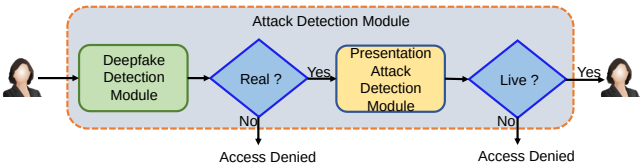
FIGURE 8. Extracted Faces using Haar Cascade (Middle) and Dlib HoG (Right Most) Face Detectors from Original Face (Left Most) (Photo Courtesy: Microsoft Power Point)

3) Attack Detection (AD) Module

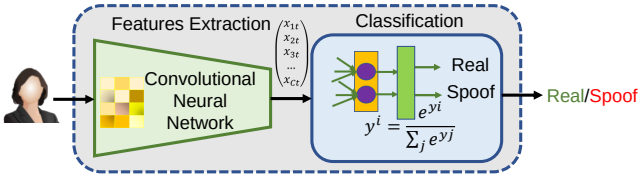
The attack detection module is the next module in the digital ID system pipeline. It checks for various presentation attacks. A person with ill intention can gain access to any facility in a smart city by fooling the facial recognition systems. A spoof image or non-living object can be presented to the camera of the edge device instead of a live person. To detect such an attack, the neutral frontal face photo of the user, taken through the edge device camera, is passed through the AD module as in Fig. 9(a). The AD module comprises of two subsystems—the Deepfake Detection (DD) and Presentation Attack Detection (PAD) modules.

a: Deepfake Detection (DD) Module

The deepfake detection (DD) module is used to detect if there is a deepfake attack on the system. Ideally, the module should be trained to detect any type of deepfake. However, for the purposes of this proof-of-concept, the model has been trained to only detect deepfakes, generated by face swapping



(a) System Level Diagram of Attack Detection Module



(b) Detection Process

FIGURE 9. Attack Detection Module

techniques. More research on deepfakes is needed to propose a generalized deepfake detection module.

We follow the procedure shown in Fig. 9(b) from [55] to detect the deepfakes. As MobileNetV2 [56] is suitable for an edge platform, pre-trained MobileNetV2 is fine tuned to extract the distinguishing features of the images. The last layer of the MobileNetV2 structure is replaced by a *fully connected layer* with *Softmax* activation and 2 nodes, and is used as a classifier as [38], [39]. *Binary Cross Entropy* loss has been used for training.

The DeepfakeTIMIT (DFTIMIT) [6], [57] dataset for fake images and VidTIMIT [58] for real images have been used for training and validation of the DD module, as shown in Table. 2. The dataset has been split into 80% train and 20% validation.

TABLE 2. Datasets for Deepfake Detection Module

Dataset	Image Type	No. of Images	Remarks
VidTIMIT	Real	34,004	10 videos taken at 3 settings for 32 subjects have been used.
DeepfakeTIMIT (HQ)	Fake	33,988	Total 620 videos of 32 subjects have been used.
DeepfakeTIMIT (LQ)	Fake	34,025	Image extraction rate 25 fps.

The classifier layer was trained for 10 epochs, and then the end-to-end model was trained for 15 epochs. The best model is evaluated as per the validation accuracy. The same and cross-dataset evaluation has been performed.

Table. 3 presents the accuracy obtained for the Deepfake Detection module. Perfect accuracy is achieved when the model is trained and evaluated on the DF-TIMIT (LQ) and VidTIMIT datasets. But in a real scenario, we counter with high-quality fraud. When the testing data is high quality and close to reality, we get a more realistic accuracy of

94.83%, which means that our system can find 94.83% of face-swapped deepfake images.

TABLE 3. Accuracy of Deepfake Detection Module for Different Evaluation Scenarios

Training Dataset	Testing Dataset	Accuracy (%)
DF-TIMIT (LQ)	DF-TIMIT (LQ)	100.00
DF-TIMIT (HQ)	DF-TIMIT (HQ)	94.83
DF-TIMIT (HQ)	DF-TIMIT (LQ)	96.91

* For real images → VidTIMIT dataset.

b: Presentation Attack Detection (PAD) Module

To detect presentation attacks, the same pipeline shown in Fig. 9(b) is followed. Here we use EfficientNet B0 [59] as the feature extractor as it shows better results. A *GlobalAveragePooling* layer, followed by a *dense* layer of 2 nodes and a *Softmax* activation function, has been used as the classifier. As presentation attacks have been approached as a binary classification problem, *Binary Cross Entropy* loss has been used. Transfer learning is also used here to shorten training time and improve accuracy.

The Replay Attack dataset [60] has been used to train and evaluate the Presentation Attack Detection module. The dataset is an imbalanced one. For our work, we partially used the spoof part of the dataset to make a balanced one. The dataset details used in our work have been stated in Table 4. The number of frames extracted from test videos is fewer than that from training videos, as the duration of the test videos is shorter than the training videos in the dataset. Frames have been extracted using ffmpeg [61].

TABLE 4. Dataset for Presentation Attack Detection Module

Dataset	Type	Image Type	No. of Images	Remarks
Replay Attack	Train	Real	899	60 original train videos
		Spoof	891	120 spoof videos
	Test	Real	80	80 original test videos
		Spoof	191	200 spoof test videos

Here, the initial number of epochs for the classifier training is kept to 5, and end-to-end model training to 10. No data augmentation has been done. The model has been evaluated with the test section of Replay Attack [60] dataset.

Our system can detect 93.0% of presentation attacks. The performance of the module is evaluated through *confusion matrix* in Fig. 10. *Precision*, *recall*, *F1-score*, and *accuracy* have been calculated in Table. 5.

TABLE 5. Classification Report of Presentation Attack Module -Trained and Tested on Replay Attack Dataset

Test images	Precision (%)	Recall (%)	F1-score
191 Spoof	100.0	90.0	94.0
80 Real	80.0	100.0	89.0
Macro Average	90.0	95.0	92.0
Weighted Average	94.0	93.0	93.0
Total 271	**Accuracy (%)**	93.0	

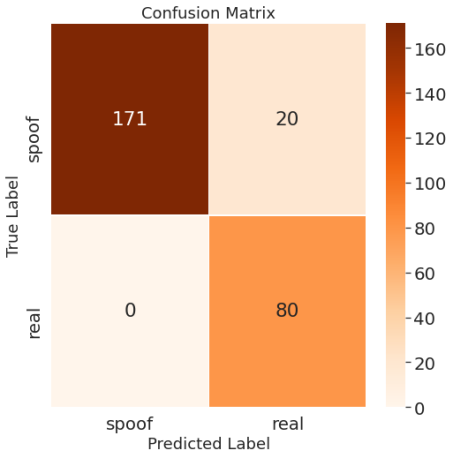


FIGURE 10. Confusion Matrix for Presentation Attack Module

4) Face Features Extraction (FFE) Module

One of the major modules of the system is the feature extraction module, as it extracts the facial features from an image. We wanted to select a simple but highly accurate method as the Face Features Extraction (FFE) module. Hence, we followed the process as in FaceNet [20] instead of other new state-of-the-art methods to extract facial features. In image classification, state-of-the-art accuracy has been obtained in Google's EfficientNets. The model size and computational complexity are notably low. Hence, instead of using the original deep learning network of FaceNet, we use EfficientNet B0 as the feature extractor [62].

Fig. 11 shows the workflow diagram of the FFE module at training. We use EfficientNet B0 [59], pretrained on ImageNet [63] as the backbone feature extractor, and connect a *GlobalAveragePooling* layer followed by a *dense* layer of 128 nodes without any activation function. *L2 normalization* is used to extract face embedding as in [20]. The module extracts facial features from the face image. These features are expressed in terms of a 128-dimensional feature vector.

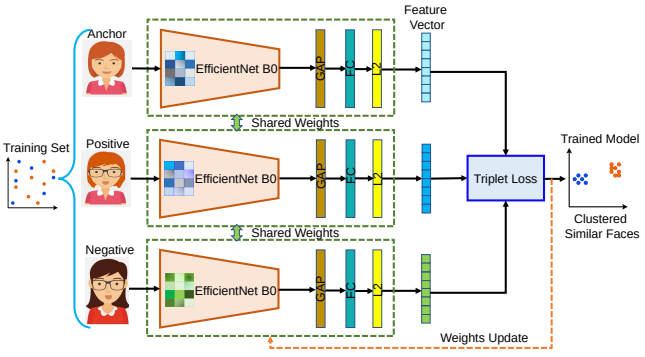


FIGURE 11. Training of Face Features Extraction Module

The Triplet Loss function [20] has been used to train the feature extraction module. During training, the network

learns to calculate the optimum *Euclidean* distance among images through embedding.

Once the feature extractor has been trained, the trained model is used to get face embeddings from the test image.

To train the face features extraction module, we downloaded images of the main characters of the American sitcom 'Friends' using the Bing Search API and created a customized dataset. The dataset details are mentioned in Table. 6. We used images without any occlusion in the face.

TABLE 6. Customized Dataset for Face Features Extraction Module and Classification Module

Dataset	Character Names	No. of Images
Main 6 Characters of American sitcom 'Friends'	Chandler	45
	Joey	50
	Monica	47
	Phoebe	43
	Rachel	36
	Ross	30

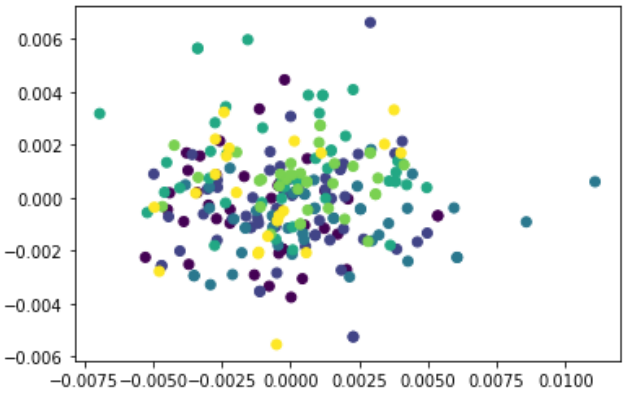
This module is trained for 100 epochs. A set of triplets, i.e. anchor image, positive image, and negative image, are generated before training. To generate the triplets, the below procedure has been followed:

1) For our dataset, we have 6 classes (6 characters). 2 classes are chosen randomly.
2) 2 images are chosen from one class, and 1 image is selected from the other class. This process is also random. From 2 images, one is chosen randomly as the anchor image and the other as the positive image, whereas the single image from the other class is chosen as the negative image.
3) For each anchor image, we chose 10 positive and 10 negative random images. So for our 247 images, 24, 700 combinations of triplets were generated.

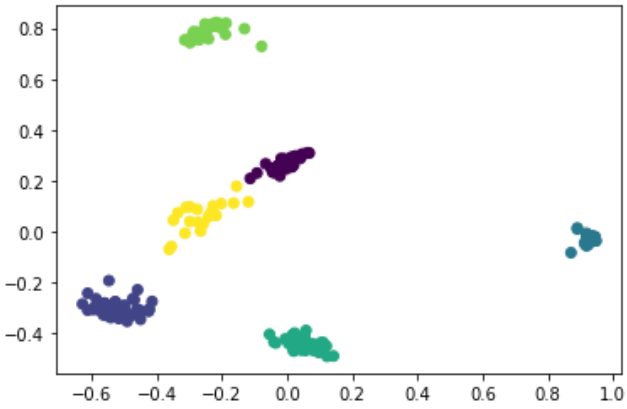
Fig. 12 shows the Principal Component Analysis (PCA) plot of the embeddings of our training dataset for the FFE module. Here, 2 principal components are used to reduce the dimensionality of the embedding vectors. Fig. 12(a) shows the embeddings before training the module, and Fig. 12(b) shows the same embeddings, clustered after the training.

5) Classifier

Once the FFE module extracts the features, a classifier is used to predict the identity of the image. In our work, a k-Nearest Neighbor (kNN) classifier with 3 neighbors, an auto-tree algorithm, and 30 leaves is used to make the prediction. The distance metric of the classifier is chosen as *Euclidean*. We train the classifier with the face embedding extracted from the trained FFE module. It authenticates or denies a face by measuring the distance to k nearest neighbors and finally decides by taking a majority voting. Fig.13 shows the classifier's training. The classifier predicts the username corresponding to the face embedding.



(a) Before Training



(b) After Training

FIGURE 12. Embedding Plots of Six Main Characters of American Sitcom 'Friends'

Embeddings from the FFE module are used as the training data for classification. The dataset details are mentioned in Table. 6.

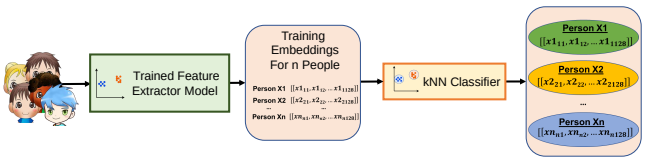


FIGURE 13. Classifier Training

Embeddings from the trained FFE module are used as the input of the classifier. So, the number of training images of the classifier is equal to the number of original images i.e. 247.

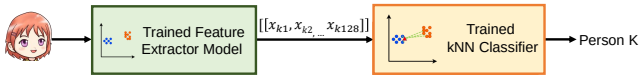
VI. PERFORMANCE OF THE PROPOSED DIGITAL ID FACE AUTHENTICATION SYSTEM

In this section, the performance of iFace 1.1 is described, along with a comparison to existing works.

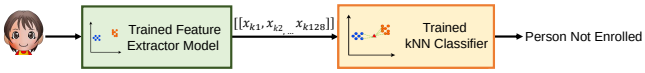
The user is authenticated in an intuitive way by comparing the input username with the predicted username from the

face embeddings extracted during authentication on the smart device. This authentication is conducted on the edge.

Fig. 14 shows the authentication process. If the person is already enrolled in the system and the predicted username matches with the input username, the classifier authenticates as in Fig. 14(a). The person gets access to the facility. But if the person is not enrolled in the system or if the predicted username does not match with the input username, the system does not give access to the facility to the person as in Fig. 14(b).



(a) Person In System



(b) Person Not In System

FIGURE 14. Authentication Process

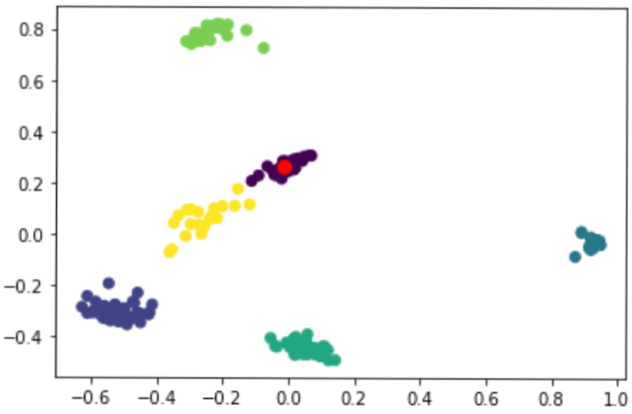
During authentication, data collection, processing, feature extraction, and prediction are all done on the edge platform. As user verification is conducted each time the user accesses the facilities, the odds of impersonating a person are very low. This provision makes the system robust. Encryption during data transfer and storage further secures the system.

When a face needs to be authenticated in the system, the embedding is extracted from the image through the trained FFE module. Two such scenarios are plotted in PCA plots, as in Fig. 15 along with training data embeddings. If the person is already enrolled in the system, the PCA plot is as in Fig. 15(a) and if the user is not yet enrolled in the system, Fig. 15(b) depicts that scenario. 171 spoofs have been correctly detected among 191 spoof images.

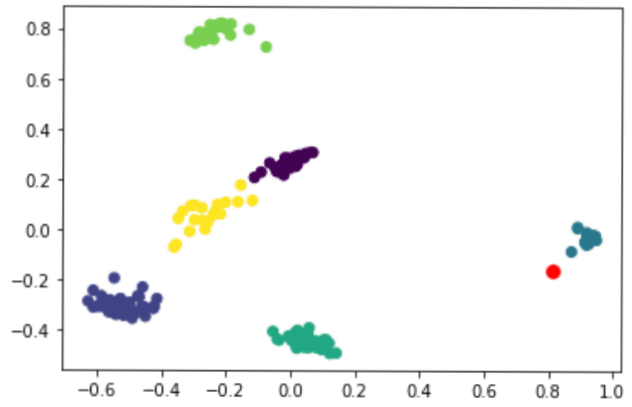
The system has been evaluated with 50 unseen frontal images of 6 enrolled people and 500 frontal images of the CelebA [64] dataset. The first set is also used as the authorized people to measure the false rejection ratio (FRR) and the second set as the intruders to measure the false acceptance ratio (FAR).

The performance of the facial recognition system of the proposed digital ID is shown in Table. 7 and Fig. 16. FRR is calculated with 50 images of enrolled people. One of the 50 images was falsely rejected, generating an FRR of 2%. We evaluate the system with 500 images from the CelebA [64] dataset. 15 images were falsely accepted. FAR is calculated to 3%.

The evaluated images are not all frontal faces. The foreheads or cheeks of some faces are occluded with hair. Those images are the cause of the false acceptance ratio. There is a variation in age for certain 'Friends' characters in our training and testing dataset. This was intentionally chosen to show the effect of the aging of a person in his/her digital ID.



(a) For Enrolled Person



(b) For Not Enrolled Person

FIGURE 15. Embedding Plot of Sample Person in the Clustered Training Dataset

TABLE 7. Performance of the Proposed Digital ID System

Dataset	Type	No. of Test Images	No. of Authentication		
			Correct	False Acceptance	False Rejection
Own Data	Enrolled User	50	49	0	1
CelebA	Not Enrolled User	500	485	15	0

Various metrics like *precision*, *recall*, *F1-score*, *accuracy*, *FRR*, and *FAR* have been calculated in Table. 8 to evaluate the digital ID system.

Table. 9 compares the proposed iFace 1.1 with the existing face authentication systems. The methods that are applicable to mobile devices or on the edge platform are stated there. It is clear from Table. 9 that different metrics have been used to evaluate the performance of these systems. Most of the systems perform face authentication without providing any security measures. They are mainly standalone facial authentication or verification systems. On the other hand, our paper presents an end-to-end facial authentication-based digital ID system that can detect both presentation attacks

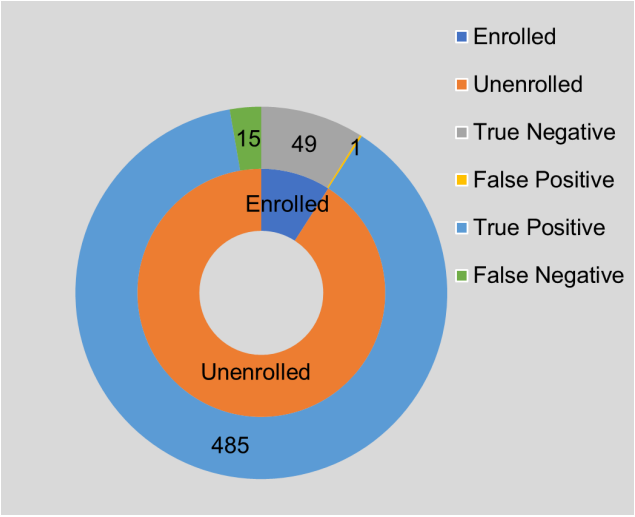


FIGURE 16. Digital ID Performance

TABLE 8. Performance Metrics of Facial Authentication System

Test images	Precision (%)	Recall (%)	F1-score
500 Impostors	100.0	70.0	98.0
50 Users	77.0	98.0	86.0
Macro Average	88.0	97.0	92.0
Weighted Average	98.0	97.0	97.0
Total 550	**Accuracy (%)**	97.0	
	FRR	2.0	
	FAR	3.0	

and deepfake attacks.

TABLE 9. Performance Metrics of Facial Authentication System

Papers	Performance Metrics	Face Authentication	Attacks Resiliency
Tao et al. [26]	EER 2%	Yes	No
Tao et al. [28]	EER 1.2%	Yes	No
Hadid et al. [30]	Acc. varies with image size	Yes	No
Sarkar et al. [31]	Acc. 88%	No	No
Masud et al. [35]	Acc. $>$ 95%	Yes	No
Mitra et al. [3]	FRR 2.77%	Yes	Yes
Current Paper	**Acc. 97%, FRR 2%**	**Yes**	**Yes**

EER $\rightarrow$ Equal Error Rate. Acc. $\rightarrow$ Accuracy.
FRR $\rightarrow$ False Rejection Ratio.

VII. CONCLUSION AND FUTURE WORK

In this article, a proof-of-concept of an end-to-end facial authentication-based digital ID system for smart cities has been proposed. Several things have been accomplished here:

- Our system is capable of detecting the various intruder attacks mentioned in Sections II-B1a and II-B1b. The system is resilient to deepmorph deepfake attacks, mentioned in Section II-B1b. It shows high accuracy even with high quality face swapped GAN generated images.
- It can detect presentation attacks, mentioned in Section II-B1a, with an accuracy of 93%.
- The false acceptance ratio and false rejection ratio of the system are fairly low.
- The face authentication system has an accuracy of 97%. As facial authentication has been done at the edge, the risk of security compromise is reduced.
- No photos are stored anywhere in the system. Face features are stored in the cloud in terms of a numerical value. Hence, it is not possible to reverse engineer the photos from these numbers, which makes the process secure.
- Biometric data is stored without any identifiable information about the user, eliminating the data privacy regulation issues.

Currently, the deepfake detection module detects the face-swapped images. However, a comprehensive deepfake detection module for other sources of deepfakes will be added. The PAD module upgrade is another area where more experiments will be performed. Our proposed system has demonstrated promise. However, more study and experimentation are needed before smart city deployment.

As for future work, a more efficient digital ID system can be achieved by addressing the following areas:

- A re-enrollment of facial features is needed to accommodate the age-related changes if the person significantly ages.
- Data deletion processes need to be included in the system after someone's death.
- A data update option needs to be there if there is any facial change due to any accident or cosmetic surgery. The provision for deletion of data and re-enrollment of the user with a new user id should be incorporated in the case of identity theft.
- The modules and the facial authentication system need to be upgraded to state-of-the-art systems.
- A large dataset of human faces is required with images generated by various GANs to obtain a generalized deepfake detection module. People of different demographics, races, colors, genders, and ages should be included in the systems. People with glasses, piercings, head coverings, hearing aids, and braces should also be included in the training dataset. The hardware should support this lengthy and resource-intensive training.
- The existing PAD module detects spoofs from the face photo. The PAD module must be tested with different head poses [65], lighting conditions [66], inside, outside, day, and night settings.
- It can be improved by using challenge response techniques with random instructions for motion, e.g., head and eye movement, opening and closing the mouth, or reading aloud any random sentences.

- Voice verification can also be added to enhance one more level of verification.
- The system should be capable of authenticating people with facial occlusions such as sunglasses, masks, and any facial piercings.
- It will be exciting to see if the identical twin scenario can be addressed.

To the authors' best knowledge, iFace 1.1 is the first viable end-to-end proof-of-concept that addresses the main challenges of smart-city digital ID. We hope to see more research in this direction and finally an implemented iFace system in a smart city.

ACKNOWLEDGMENT

A short conference version of this work was presented at [3]

REFERENCES

[1] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The Internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016, doi: 10.1109/MCE.2016.2556879.

[2] "Report: Cisco and IBM leaders in the smart cities technology market," November 21, 2014, Accessed: February 24, 2022. [Online]. Available: https://www.smartcitiescouncil.com/article/report-cisco-and-ibm-leaders-smart-cities-technology-market

[3] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "iFace: A Deepfake Resilient Digital Identification Framework for Smart Cities," in *Proc. IEEE International Symposium on Smart Electronic Systems (iSES)*, 2021, pp. 361–366, doi: 10.1109/iSES52644.2021.00090.

[4] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001, doi: 10.1147/sj.403.0614.

[5] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally, "Introduction to Presentation Attack Detection in Face Biometrics and Recent Advances," 2021. [Online]. Available: https://arxiv.org/abs/2111.11794

[6] P. Korshunov and S. Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection," 2018. [Online]. Available: https://arxiv.org/abs/1812.08685

[7] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment," 2019. [Online]. Available: https://arxiv.org/abs/1908.05932

[8] S. P. Mohanty, "Security and Privacy by Design is Key in the Internet of Everything (IoE) Era," *IEEE Consumer Electronics Magazine*, vol. 9, no. 2, pp. 4–5, 2020, doi: 10.1109/MCE.2019.2954959.

[9] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021, doi: https://doi.org/10.1016/j.neucom.2020.10.081.

[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997, doi: 10.1109/34.598228.

[11] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 3, pp. 328–340, 2005, doi: 10.1109/TPAMI.2005.55.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2008, doi: 10.1109/TPAMI.2008.79.

[13] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002, doi: 10.1109/TIP.2002.999679.

[14] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006, doi: 10.1109/TPAMI.2006.244.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *Proc. IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708, doi: 10.1109/CVPR.2014.220.

[16] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[17] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," in *Proc. IEEE conference on computer vision and pattern recognition*, 2014, pp. 1891–1898, doi: 10.1109/CVPR.2014.244.

[18] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," 2014, doi: 10.48550/ARXIV.1406.4773. [Online]. Available: https://arxiv.org/abs/1406.4773

[19] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," 2014, doi: 10.48550/ARXIV.1412.1265. [Online]. Available: https://arxiv.org/abs/1412.1265

[20] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[21] J. Deng, Y. Zhou, and S. Zafeiriou, "Marginal Loss for Deep Face Recognition," in *Proc. IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 60–68, doi: 10.1109/CVPRW.2017.251.

[22] L. Tran, X. Yin, and X. Liu, "Disentangled Representation Learning GAN for Pose-Invariant Face Recognition," in *Proc. IEEE conference on computer vision and pattern recognition*, 2017, pp. 1415–1424, doi: 10.1109/CVPR.2017.141.

[23] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large Margin Cosine Loss for Deep Face Recognition," in *Proc. IEEE conference on computer vision and pattern recognition*, 2018, pp. 5265–5274, doi: 10.1109/CVPR.2018.00552.

[24] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive Angular Margin Loss for Deep Face Recognition," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699, doi: 10.1109/CVPR.2019.00482.

[25] M. A. Ferrag, L. Maglaras, and A. Derhab, "Authentication and authorization for mobile IoT devices using biofeatures: Recent advances and future trends," *Security and Communication Networks*, vol. 2019, p. 20 pages, 2019, doi: https://doi.org/10.1155/2019/5452870.

[26] Q. Tao and R. Veldhuis, "Biometric Authentication System on Mobile Personal Devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 763–773, 2010, doi: 10.1109/TIM.2009.2037873.

[27] M. S. Hossain, G. Muhammad, S. M. M. Rahman, W. Abdul, A. Alelaiwi, and A. Alamri, "Toward end-to-end biomet rics-based security for IoT infrastructure," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 44–51, 2016, doi: 10.1109/MWC.2016.7721741.

[28] Q. Tao and R. N. Veldhuis, "Biometric Authentication for a Mobile Personal Device," in *Proc. 3rd Annual International Conference on Mobile and Ubiquitous Systems - Workshops*, 2006, pp. 1–3, doi: 10.1109/MOBIQW.2006.361741.

[29] M. E. Fathy, V. M. Patel, and R. Chellappa, "Face-based active authentication on mobile devices," in *Proc. IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2015, pp. 1687–1691, doi: 10.1109/ICASSP.2015.7178258.

[30] A. Hadid, J. Y. Heikkila, O. Silven, and M. Pietikainen, "Face and Eye Detection for Person Authentication in Mobile Phones," in *Proc. First ACM/IEEE International Conference on Distributed Smart Cameras*, 2007, pp. 101–108, doi: 10.1109/ICDSC.2007.4357512.

[31] S. Sarkar, V. M. Patel, and R. Chellappa, "Deep feature-based face detection on mobile devices," in *Proce. IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2016, pp. 1–8, doi: 10.1109/ISBA.2016.7477230.

[32] Y. Sutcu, Q. Li, and N. Memon, "Secure biometric templates from fingerprint-face features," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6, doi: 10.1109/CVPR.2007.383385.

[33] T. Phillips, X. Zou, F. Li, and N. Li, "Enhancing Biometric-Capsule-Based Authentication and Facial Recognition via Deep Learning," in *Proc. 24th ACM Symposium on Access Control Models and Technologies*, ser. SACMAT '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 141–146, doi: 10.1145/3322431.3325417.

[34] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multi-task cascaded convolutional networks," *CoRR*, vol. abs/1604.02878, 2016.

[35] M. Masud, G. Muhammad, H. Alhumyani, S. S. Alshamrani, O. Cheikhrouhou, S. Ibrahim, and M. S. Hossain, "Deep learning-based intelligent face recognition in IoT-cloud environment," *Computer Communications*, vol. 152, pp. 215–222, 2020, doi: https://doi.org/10.1016/j.comcom.2020.01.050.

[36] H. Chen, W. Wang, J. Zhang, and Q. Zhang, "EchoFace: Acoustic Sensor-Based Media Attack Detection for Face Authentication," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2152–2159, 2020, doi: 10.1109/JIOT.2019.2959203.

[37] Y. Chen, J. Sun, X. Jin, T. Li, R. Zhang, and Y. Zhang, "Your face your heart: Secure mobile face authentication with photoplethysmograms," in *Proc. IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9, doi: 10.1109/INFOCOM.2017.8057220.

[38] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "A Novel Machine Learning based Method for Deepfake Video Detection in Social Media," in *Proc. IEEE International Symposium on Smart Electronic Systems (iSES)*, 2020, pp. 91–96, doi: 10.1109/iSES50453.2020.00031.

[39] Alakananda Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "A Machine Learning Based Approach for Deepfake Detection in Social Media Through Key Video Frame Extraction," *SN Comput. Sci.*, vol. 2, no. 2, p. 98, 2021, doi: 10.1007/s42979-021-00495-x.

[40] C.-Z. Yang, J. Ma, S. Wang, and A. W.-C. Liew, "Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1841–1854, 2021, doi: 10.1109/TIFS.2020.3045937.

[41] T.-L. Do, M.-K. Tran, H. H. Nguyen, and M.-T. Tran, "Potential Threat of Face Swapping to eKYC with Face Registration and Augmented Solution with Deepfake Detection," in *Future Data and Security Engineering*, T. K. Dang, J. Küng, T. M. Chung, and M. Takizawa, Eds. Cham: Springer International Publishing, 2021, pp. 293–307, doi: https://doi.org/10.1007/978-3-030-91387-8_19.

[42] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "EasyDeep: An IoT Friendly Robust Detection Method for GAN Generated Deepfake Images in Social Media," in *Internet of Things. Technology and Applications*, L. M. Camarinha-Matos, G. Heijenk, S. Katkoori, and L. Strous, Eds. Cham: Springer International Publishing, 2022, pp. 217–236, doi: https://doi.org/10.1007/978-3-030-96466-5_14.

[43] W. Xu, J. Liu, S. Zhang, Y. Zheng, F. Lin, J. Han, F. Xiao, and K. Ren, "RFace: Anti-Spoofing Facial Authentication Using COTS RFID," in *Proc. IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10, doi: 10.1109/INFOCOM42981.2021.9488737.

[44] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[45] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-Shot Multi-Level Face Localisation in the Wild," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5203–5212, doi: 10.1109/CVPR42600.2020.00525.

[46] R. Girshick, "Fast R-CNN," 2015, doi: 10.48550/ARXIV.1504.08083. [Online]. Available: https://arxiv.org/abs/1504.08083

[47] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015.

[48] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017, doi: 10.48550/ARXIV.1703.06870. [Online]. Available: https://arxiv.org/abs/1703.06870

[49] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[50] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot Multibox Detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37, doi: https://doi.org/10.1007/978-3-319-46448-0_2.

[51] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I, doi: 10.1109/CVPR.2001.990517.

[52] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004, doi: https://doi.org/10.1023/B:VISI.0000013087.49260.fb.

[53] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[54] D. E. King, "Dlib-Ml: A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, p. 1755–1758, Dec 2009.

[55] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "Detection of Deep-Morphed Deepfake Images to Make Robust Automatic Facial Recognition Systems," in *Proc. 19th OITS International Conference on Information Technology (OCIT)*, 2021, pp. 149–154, doi: 10.1109/OCIT53463.2021.00039.

[56] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[57] P. Korshunov and S. Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection," Idiap, Idiap-RR Idiap-RR-18-2018, 12 2018, http://publications.idiap.ch/attachments/reports/2018/Korshunov_Idiap-RR-18-2018.pdf.

[58] C. Sanderson and B. Lovell, "Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference," Lecture Notes in Computer Science (LNCS), Vol. 5558, pp. 199–208, 2009, doi: https://doi.org/10.1007/978-3-642-01793-3_21.

[59] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019. [Online]. Available: https://arxiv.org/abs/1905.11946

[60] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. the International Conference of Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1–7.

[61] S. Tomar, "Converting video formats with FFmpeg," *Linux Journal*, vol. 2006, no. 146, p. 10, 2006.

[62] S. Gupta and M. Tan, "EfficientNet-EdgeTPU: Creating Accelerator-Optimized Neural Networks with AutoML ," https://ai.googleblog.com/2019/08/efficientnet-edgetpu-creating.html, Accessed: January 30, 2022.

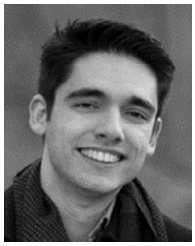
[63] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[64] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," in *Proc. International Conference on Computer Vision*, December 2015, doi: 10.1109/ICCV.2015.425.

[65] S. Basak, P. Corcoran, F. Khan, R. Mcdonnell, and M. Schukat, "Learning 3D Head Pose From Synthetic Data: A Semi-Supervised Approach," *IEEE Access*, vol. 9, pp. 37 557–37 573, 2021, doi: 10.1109/ACCESS.2021.3063884.

[66] V. Varkarakis, W. Yao, and P. Corcoran, "Towards End-to-End Neural Face Authentication in the Wild - Quantifying and Compensating for Directional Lighting Effects," *CoRR*, vol. abs/2104.03854, 2021. [Online]. Available: https://arxiv.org/abs/2104.03854

ALAKANANDA MITRA (Student Member, IEEE) received a Bachelor of Science (Honors) in Physics from Presidency College, University of Calcutta in 2001 and a Bachelor of Technology and Master of Technology in Radiophysics and Electronics from Institute of Radiophysics and Electronics, University of Calcutta in 2004 and 2006 respectively. Currently, she is a doctoral student in the research group at Smart Electronics Systems Laboratory (SESL) in the Department of Computer Science and Engineering at University of North Texas, Denton, USA. She has been awarded the Outstanding Early Stage Doctoral Student, 2022 award by the Department of Computer Science and Engineering, University of North Texas. Along with her course work, she also works as a Teaching Assistant in the department. She has worked as a Project Linked Personnel at Advanced Computing and Microelectronics Unit in Indian Statistical Institute from 2006 to 2007. Her research interests include artificial intelligence, machine learning, deep learning, edge AI, and application of AI/ML approaches in multi-media forensics, smart agriculture, and smart healthcare.



DAN BIGIOI graduated from the National University of Ireland Galway in 2020 with a bachelor's degree in Electronic and Computer Engineering. Upon graduating, he worked as a research assistant at NUIG studying the text to speech and speaker recognition methods under the DAVID (Data-Center Audio/Visual Intelligence on-Device) project. Currently, he is working on his Ph.D. at NUIG, sponsored by D-REAL, the SFI Centre for Research Training in Digitally Enhanced Reality. His research involves studying and implementing novel deep learning-based techniques for Automatic Speech Dubbing and discovering new ways to process multi-modal audio/visual data.

SARAJU P. MOHANTY (Senior Member, IEEE) received the bachelor's degree (Honors) in electrical engineering from the Orissa University of Agriculture and Technology, Bhubaneswar, in 1995, the master's degree in Systems Science and Automation from the Indian Institute of Science, Bengaluru, in 1999, and the Ph.D. degree in Computer Science and Engineering from the University of South Florida, Tampa, in 2003. He is a Professor with the University of North Texas. His research is in "Smart Electronic Systems" which has been funded by National Science Foundations (NSF), Semiconductor Research Corporation (SRC), U.S. Air Force, IUSSTF, and Mission Innovation. He has authored 450 research articles, 4 books, and invented 8 granted/pending patents. His Google Scholar h-index is 45 and i10-index is 185 with 9,400 citations. He is a recipient of 14 best paper awards, Fulbright Specialist Award in 2020, IEEE Consumer Electronics Society Outstanding Service Award in 2020, the IEEE-CS-TCVLSI Distinguished Leadership Award in 2018, and the PROSE Award for Best Textbook in Physical Sciences and Mathematics category in 2016. He is widely credited as the designer for the first digital watermarking chip in 2004 and the first low-power digital watermarking chip in 2006. He has delivered 15 keynotes and served on 13 panels at various International Conferences. He has been the Editor-in-Chief of the IEEE Consumer Electronics Magazine during 2016-2021 and serves on the editorial board of multiple journals/transactions.

PETER CORCORAN (Fellow, IEEE) currently holds the Personal Chair in electronic engineering with the College of Science and Engineering, National University of Ireland Galway (NUIG). He was the Co-Founder of several start-up companies, notably FotoNation (currently the Imaging Division, Xperi Corporation). He has more than 600 cited technical publications and patents, more than 120 peer-reviewed journal articles, 160 international conference papers, and a co-inventor on more than 300 granted U.S. patents. He is an IEEE Fellow recognized for his contributions to digital camera technologies, notably in-camera red-eye correction and facial detection. He is also a member of the IEEE Consumer Technology Society for more than 25 years and the Founding Editor of IEEE Consumer Electronics Magazine

ELIAS KOUGIANOS (Senior Member, IEEE) received a BSEE from the University of Patras, Greece in 1985 and an MSEE in 1987, an MS in Physics in 1988 and a Ph.D. in EE in 1997, all from Louisiana State University. From 1988 through 1998 he was with Texas Instruments, Inc., in Houston and Dallas, TX. In 1998 he joined Avant! Corp. (now Synopsys) in Phoenix, AZ as a Senior Applications engineer and in 2000 he joined Cadence Design Systems, Inc., in Dallas, TX as a Senior Architect in Analog/Mixed-Signal Custom IC design. He has been at UNT since 2004. He is a Professor in the Department of Electrical Engineering, at the University of North Texas (UNT), Denton, TX. His research interests are in the area of Analog/Mixed-Signal/RF IC design and simulation and in the development of VLSI architectures for multimedia applications. He is an author of over 140 peer-reviewed journal and conference publications.