IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 January 2023, accepted 31 January 2023, date of publication 6 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3242654

RESEARCH ARTICLE

CloudNet: A LiDAR-Based Face Anti-Spoofing Model That Is Robust Against Light Variation

YONGRAE KIM[1,2], HYUNMIN GWAK[2,3], JAEHOON OH[3], MINHO KANG[4], JINKYU KIM[1], HYUN KWON[5], AND SUNGHWAN KIM[3]

[1]Department of Computer Science and Engineering, Korea University, Seoul 02841, South Korea
[2]Mustree Company Ltd., Seoul 05029, South Korea
[3]Department of Applied Statistics, Konkuk University, Seoul 05029, South Korea
[4]Department of Physics, Korea University, Seoul 02841, South Korea
[5]Department of Artificial Intelligence and Data Science, Korea Military Academy, Seoul 01819, South Korea

Corresponding author: Sunghwan Kim (shkim1213@konkuk.ac.kr) and Hyun Kwon (hkwon.cs@gmail.com)

This work was supported in part by the National Research Foundation of Korea (NRF) through the Ministry of Education, Science and Technology under Grant NRF 2020R1C1C1A01005229 and Grant NRF 2021R1A4A5032622.

ABSTRACT Face anti-spoofing (FAS) is a technology that protects face recognition systems from presentation attacks. The current challenge faced by FAS studies is the difficulty in creating a generalized light variation model. This is because face data are sensitive to light domain. FAS models using only red green blue (RGB) images suffer from poor performance when the training and test datasets have different light variations. To overcome this problem, this study focuses on light detection and ranging (LiDAR) sensors. LiDAR is a time-of-flight depth sensor that is included in the latest mobile devices. It is negligibly affected by light and provides 3D coordinate and depth information of the target. Thus, a model that is resistant to light variations and exhibiting excellent performance can be created. For the experiment, datasets collected with a LiDAR camera are built and CloudNet architectures for RGB, point clouds, and depth are designed. Three protocols are used to confirm the performance of the model according to variations in the light domain. Experimental results indicate that for protocols 2 and 3, CloudNet error rates increase by 0.1340 and 0.1528, whereas the error rates of the RGB model increase by 0.3951 and 0.4111, respectively, as compared with protocol 1. These results demonstrate that the LiDAR-based FAS model with CloudNet has a more generalized performance compared with the RGB model.

INDEX TERMS Deep learning, face anti-spoofing, LiDAR, point cloud.

I. INTRODUCTION

Face recognition systems are widely used in various applications owing to their convenience and excellent performance. However, this technology is vulnerable to presentation attacks, such as print, replay, and 3D masks. In particular, 2D printers and mobile devices can easily generate print and replay attacks. Advances in scanners and 3D printers have enabled the production of high-quality 3D masks. Now, obtaining images of certain people's faces through the Internet is easy; consequently, sophisticated spoofs can be created for malicious purposes. Therefore, numerous studies have been conducted to improve the face anti-spoofing (FAS) model.

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

Traditionally, FAS has adopted handcrafted methods, such as eye blinking [1] or gaze tracking [2]. Owing to the rapid developments in deep learning technology, end-to-end deep learning-based FAS models have also been studied extensively [3], [4], [5]. Several of these studies focused on commercial red green blue (RGB) cameras as it is an excellent solution that considers both the performance and cost [6], [7], [8]. However, some industries, such as mobile payments, require a secure model with lower errors, even if the costs are higher. Therefore, numerous studies have been recently conducted to further improve the performance of FAS models using advanced sensors [9], [10], [11], [12], [13]. Advanced sensors include near-infrared (NIR), short-wavelength infrared (SWIR), depth sensor, thermal, light field, and polarization cameras. In practice, these sensors perform excellently at detecting presentation attacks. The

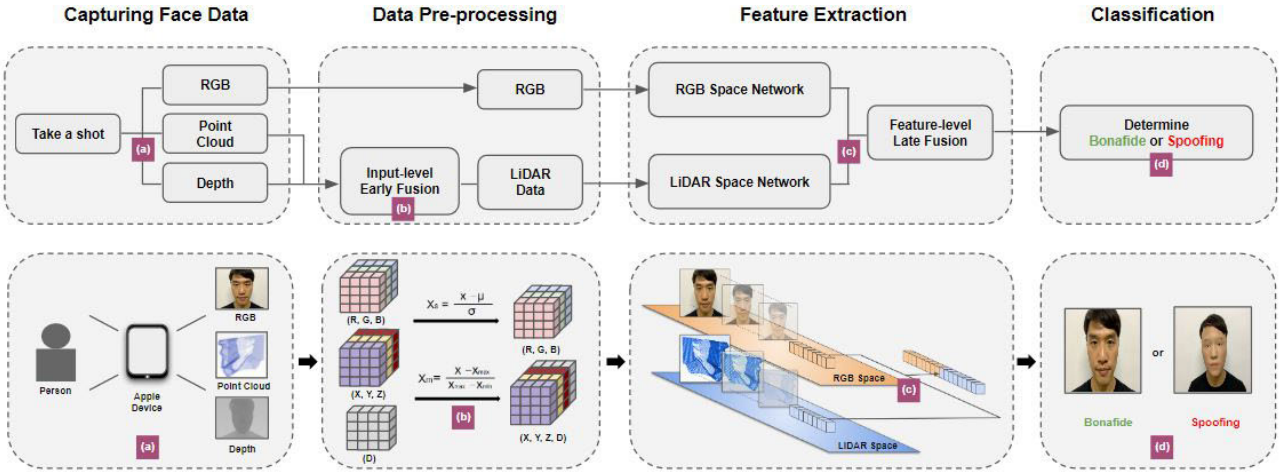


FIGURE 1. Proposed framework of FAS using LiDAR sensor. (a) Detecting a face with LiDAR application to obtain RGB, point cloud, and depth image. Depth is created based on the point cloud. (b) During data pre-processing, point cloud and depth are combined through early fusion. (c) Features of RGB and LiDAR data are extracted by separate networks and late fusion is performed. (d) FAS model determines whether the photographed face is bonafide or spoofing.

light detection and ranging (LiDAR) sensors also have this advantages. LiDAR-based multi-modal FAS model uses 3D spatial and depth information as well as color information and it leads excellent performance.

LiDAR sensors have another advantage in that they provide FAS models that are robust against light variations [14], [15]. One of the challenges that FAS studies face is creating a generalized model for environments, such as light and background [16], [17], [18], [19], [20]. In particular, face data are significantly affected by the intensity of light [17]. This implies that when a FAS model is delivered in an actual service, the face data obtained from different illuminations may be misidentified. This can have catastrophic consequences for financial services, such as mobile payments. This problem can be overcome by collecting training data from numerous environments. However, face data is difficult to collect because of the nature of biometric data, and collecting such data in various environments is even more challenging. To solve this problem, this study focused on LiDAR sensors. LiDAR sensors measure the distance by calculating the round-trip delay of the light signal emitted by the laser to the target. This provides 3D spatial coordinates of the points that make up the target and are called point clouds. Compared with RGB data, whose values fluctuate with light variation, LiDAR point clouds are negligibly affected by light. Therefore, using LiDAR sensors can reduce the impact of light on the model performance. Finally, LiDAR sensors have recently been integrated into mobile devices, making it convenient for creating multi-modal models without the need for additional hardware. This is beneficial as it allows for real-world mobile applications such as those that use both RGB cameras and LiDAR sensors.

In this study, the so-called CloudNet, a LiDAR-based FAS model, is proposed. As shown in Fig. 1, CloudNet determines the liveness of a face using RGB images, point clouds, and depth images obtained from a LiDAR-based camera. CloudNet consists of a RGB space and LiDAR space networks to learn the separate weights for RGB, point clouds, and depth data. The architecture of CloudNet is a binary classifier based on Resnet34. This is because recent multi-modal FAS methods have adopted Resnet [21], [22], VGG [23] and so on as backbone for image classification tasks. To verify the model performance, a dataset collected by the LiDAR sensor was required. Because no public face dataset has been built with LiDAR sensors, in this study, the LiDAR dataset for FAS (LDFAS) was built using an Apple iPad equipped with LiDAR sensors. Three protocols were used to confirm the superiority of the model according to light variation. In protocol 1, the training and test sets had the same light domains. Protocols 2 and 3 constructed these sets with different light domains. The RGB model and CloudNet had error rates of 0.0667 and 0 for protocol 1, 0.4618 and 0.1340 for protocol 2, and 0.4778 and 0.1528 for protocol 3, respectively. CloudNet increased the errors by 0.1340 and 0.1528, whereas the RGB model increased the errors by 0.3951 and 0.4111. This demonstrates that CloudNet with LiDAR sensors is a more generalized model than the RGB model. In addition, we investigated the trade-offs caused by using the LiDAR sensor. The details of the costs will be discussed in the experimental results and discussion sections in Section V. The contributions of this study can be summarized as follows.

- A method to create a generalized model for the light domain was devised using a LiDAR sensor.
- The LDFAS, which contains point clouds and depth using LiDAR sensors, was built.
- CloudNet was designed to efficiently train point clouds using a LiDAR sensor.

The remainder of this paper is organized as follows. Section II discusses related work. Section III describes the dataset built herein (LDFAS). Section IV explains the proposed method in detail. Section V covers the experimental setup, evaluation metrics, experimental results, and

ablation studies; additionally, it discusses the results. Finally, Section VI concludes the study.

II. RELATED WORK

A. MULTI-MODAL FACE ANTI-SPOOFING

Deep learning-based FAS study can be divided into two categories depending on the sensor used [24]. The first category utilizes only a commercial RGB camera [25], [26]. As previously mentioned, using RGB cameras is a excellent way for creating low-cost and high-performance FAS models. However, in certain high-security scenarios, such as face payment and vault entrance, a extremely low rate of false acceptance is required. As a result, a second category which uses a special sensor, along with or without a commercial RGB camera has been introduced. These specialized sensors, including NIR sensors [9], [27], [28], SWIR sensors [9], [28], depth sensors [9], [10], [27], [28], thermal sensors [9], [11], [28], light-field cameras [12], and four-way polarization cameras [13], increase the accuracy of FAS models. SWIR sensors are known to effectively protect against 3D mask attacks caused by moisture on real faces [9]. Reference [27] has shown through ablation study to reduce the error rate of FAS models through the addition of depth and IR sensors. Thermal sensors effectively block attacks based on the fact that the average temperature of the human face is 36–37°C [11]. Additionally, a light-field camera and four-directional polarization sensor improve the FAS model performance [13]. This study belongs to the second case. Herein, a LiDAR sensor, which is a time-of-flight-based depth sensor, was used.

B. MULTI-MODAL FUSION

Multi-modal fusion is a method of combining data collected from different modalities to achieve more accurate results [29]. It is widely used in various fields from affective computing [30] to autonomous driving [31]. Recent studies have demonstrated that by using a combination of visual, vocal, and textual data, it is possible to more accurately identify psychological patterns from multiple perspectives [30], [32]. In the field of autonomous driving, multi-modal fusion has also been used [29], [33]. RGB images provide rich visual information, but are sensitive to light variation. Point clouds do not affect by light but have limitations in terms of resolution. Autonomous driving study fuses RGB images and point clouds together to use the data complementarily to overcome their own limitations [29]. Currently, multi-modal models often use one of three methods for combining data: early fusion, middle fusion, and late fusion. Early fusion combines data at the pre-processing stage, middle fusion is used during the feature extraction phase, and late fusion combines the output from multiple models to produce the final result [29].

C. LiDAR SENSOR AND POINT CLOUD

The LiDAR sensor measures the distance by calculating the round-trip delay time when the light signal emitted from the laser reaches the target [34]. It has been used as an observation technology for precise atmospheric analysis and global environmental observation via mounting on aircraft and satellites, and as a important technology for laser scanners and 3D imaging cameras in autonomous driving. Recently, mobile applications that use LiDAR for face recognition and clothes measurement had also been studied [35], [36]. The sensor generates point cloud data, which is a 3D representation of the target. Point cloud can be learned in deep learning models via three approaches [37]. The first is to project a point cloud onto a 2D plane and then learn the features using conventional 2D convolutional neural networks (CNNs). The second is a Voxel-based learning method that learns using a 3D space-based 3D CNN called Voxel. Finally, the third is learning pixel-by-pixel. The first method was used in the present study. As is well-known, the feature distribution of LiDAR images changes drastically at different image locations despite the similarities between regular RGB and LiDAR images [38]. Recently, some methods have been devised for deep learning models to effectively learn from point cloud data. Typically, SqueezeSegV3 adapts the SAC block [38], whereas FPS-Net uses the MRF-RDB block [39]. To solve this problem, a separate network for each dataset was designed herein. More information is provided in Section IV.

III. LiDAR DATASET FOR FACE ANTI-SPOOFING (LDFAS)

The LDFAS was built to develop a LiDAR-based FAS model. The Dataset is composed of 8,640 face data collected from 36 Koreans. (2880 images in each of RGB, point cloud, and depth configurations). This section describes the LiDAR application, data collection procedure, comparision with multi-modal based public datasets, and evaluation protocols. Examples of these datasets are presented in Fig. 2.

A. LiDAR APPLICATION

An artkit-based mobile app that simultaneously generated RGB, point cloud, and depth data was used [35]. This camera application also provides information on how to map all points in the point cloud to a specific pixel of the RGB image. Depth images are derived from point clouds. A 3D point cloud had 45,192 points and depth had a resolution of 256 $\times$ 192, and RGB was generated at 1440 $\times$ 1080 pixels. The RGB image and the point cloud were each captured with the 12 MP Wide Camera and the TOF 3D LiDAR scanner, respectively.

B. DATA COLLECTION PROCEDURE

The participants were instructed to sit in front of the camera and look towards the sensor. LDFAS dataset are divided into three subsets: indoor, outdoor and indoor (dark). Table 2 shows the explanation of LDFAS's subsets. During the indoor subset, the participants were positioned 70-90cm away from the camera while bonafide and 3D mask were photographed. The lighting was maintained between 170-180 lux. The participants were also asked to slightly rotate their heads and 20 images were taken without video. The outdoor subset was collected during the day and data was collected at different

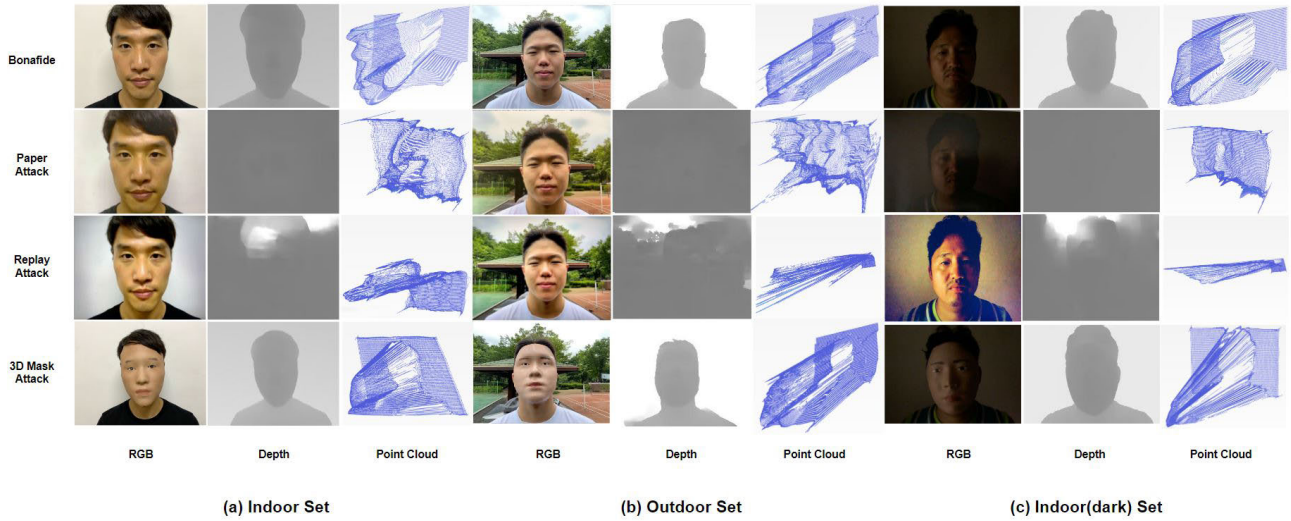


FIGURE 2. Examples from the LDFAS. Face data of three persons taken from LDFAS. The samples belong to indoor, outdoor, and indoor(dark) subsets. The point clouds are constructed using an online 3D viewer (https://www.creators3d.com/online-viewer).

TABLE 1. Summary of statistics of the LDFAS and comparison to multi-modal based public datasets. # of Bonafides and # of Spoofings means the number of images.

Public Dataset	# of Sub.	# of Bonafides	# of Spoofings	Modality	Attack Type	Year
CeFA [27]	1607	6300	27900	RGB, Depth, IR	Print, Replay, 3D Mask	2020
HQ-WMCA [9]	51	555	2349	RGB, Depth, NIR, SWIR, Thermal	Print, Replay, 3D Mask, Glasses, Wigs, Tatoo	2020
PADISI-Face [28]	360	1105	924	RGB, Depth, NIR, SWIR, Thermal	Print, Replay, 3D Mask, Glasses, Funny eye	2021
LDFAS	36	720	2160	RGB, LiDAR	Print, Replay, 3D Mask	2022

locations and at different times. The distance between the camera and the participants was also adjusted randomly. The indoor (dark) subset was collected in a dark indoor environment with varying degrees of darkness. The distance between the camera and the participants was also adjusted randomly and the participants were asked to slightly rotate their heads, 20 images were taken without video.

TABLE 2. Table of Explanation for LDFAS subsets.

Subset	Light	Background	Components
Indoor	constant	constant	RGB, Point Cloud, Depth
Outdoor	variable(bright)	variable	RGB, Point Cloud, Depth
Indoor(dark)	variable(dark)	variable	RGB, Point Cloud, Depth

Print attack and replay attack were made with bonafide. All print attacks and replay attacks in the three subsets were photographed under the same lighting conditions at the same location and at the same time, respectively. An interesting point is that, as shown in Fig. 2, replay attacks taken by the LiDAR sensor did not appear as a completely flat surface. This phenomenon occurred because the light shot from the LiDAR sensor was reflected on the surface of the device reproducing the replay attack. Therefore, it was important to collect all replay attacks in the LDFAS under the same conditions. Print attacks were made with a laser printer and replay attacks were made by playing on an Apple's iPad device. The 3D mask was made of thermoplastic polyurethane (TPU) and the Cubicon 3DP 320C Single Plus 3D Printer. The numbers of bonafides and attacks are listed in Table 3.

TABLE 3. Table that shows the number of data in subsets.

Subset	Subjects	Bonafides	Papers	Replays	3D Masks
Indoor	12	240	240	240	240
Outdoor	12	240	240	240	240
Indoor(dark)	12	240	240	240	240

C. COMPARISION WITH MULTI-MODAL BASED PUBLIC DATASETS

Recently, several public datasets have been built for study on multi-modal based FAS in recent years. In comparison to other datasets, Table 1 shows the novel aspect of LDFAS. To the best of our knowledge, the most recent datasets are CeFA, HQ-WMCA, and PADISI-Face [9], [27], [28]. CeFA is a large dataset with 1,607 participants and includes presentation attacks in the form of impersonation using RGB camera, depth, and IR sensors [27]. HQ-WMCA also is a large dataset and includes presentation attacks in the form of impersonation and obfuscation, it was constructed using RGB, depth, NIR, SWIR, and thermal sensors [9]. PADISI-Face is also a large dataset with 360 participants, it is composed of various modalities similar to HQ-WMCA and includes presentation attacks in the form of impersonation and obfuscation [28]. The main difference in the LDFAS dataset that we have built is the use of a new modality called LiDAR. The LiDAR sensor generates point cloud data, which

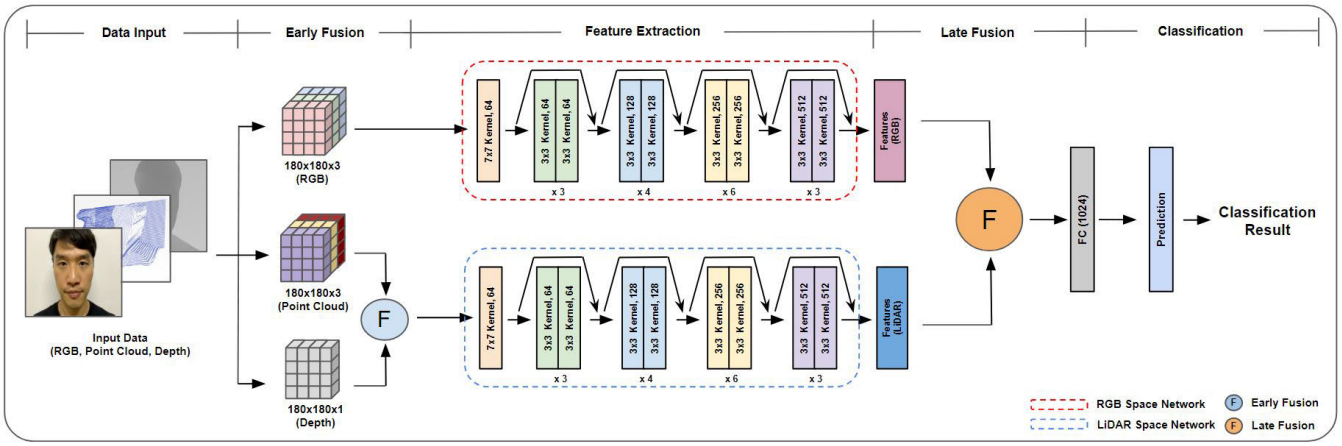


FIGURE 3. Architecture of CloudNet.

is why we constructed a dataset composed of RGB, point cloud, and depth map.

D. EVALUATION PROTOCOLS

The goal of this study was to develop a generalized FAS model considering light variations. Three protocols were designed for this purpose. Protocol 1 corresponds to when the learning and test datasets are in the same light conditions. By contrast, protocols 2 and 3 used different light conditions. The indoor, outdoor, and indoor (dark) sets were tested while training only the indoor sets. Details of each protocol are listed in Table 4.

TABLE 4. Description of evaluation protocols.

Protocol	Train Set	Test Set	# of Train Data	# of Test Data
1	Indoor	Indoor	720	240
2	Indoor	Outdoor	720	960
3	Indoor	Indoor(dark)	720	960

IV. PROPOSED METHOD

In this section, the CloudNet architecture is explained. The structure is composed of a RGB space and LiDAR space networks. Each network extracts facial features from the RGB and LiDAR data (point cloud and depth). CloudNet performs both early fusion and late fusion to classify bonafide and spoofing images. Herein, binary cross-entropy was used as the loss function. The architecture of the model is shown in Fig. 3.

A. ARCHITECTURE

The input data for CloudNet are RGB, point clouds, and depth. CloudNet performs two fusion operations. The first one is an early fusion of point cloud and depth. The second one is a late fusion of the RGB space network and the LiDAR space network. The fusion operation is represented as follows.

$$F_{fusion} = [I_1,\ I_2] \tag{1}$$

where F represents the fusion operation. Accordingly, the entire CloudNet network can be described as follows.

$$Y = \sigma([N_{rgb}(I_{rgb}),\ N_{lidar}([I_{pc},\ I_d])]) \tag{2}$$

where I_{rgb}, I_{pc}, and I_d represent the RGB, point cloud, and depth, respectively; N_{rgb} and N_{lidar} represent the RGB and LiDAR networks, respectively; and σ denotes the sigmoid function, which is a non-linear activation function [40]. Herein, both the RGB and LiDAR networks were implemented using Resnet34, which is a CNN-based network exhibiting outstanding performance in image classification [41]. The difference between the networks used in previous studies and Resnet34, which was used herein, is that Resnet34 does not have a fully connected layer. After the first early fusion, the input images passes through the Resnet34-based inner networks for the second late fusion. When the late fusion operation is completed, they go to the fully connected layer and finally pass to the activation function.

The CloudNet consists of two networks owing to the characteristic of point cloud. The feature distribution of LiDAR images differs significantly from that of RGB images. As shown in Fig. 4, the feature distribution of RGB was confirmed to be different from that of the point cloud and depth images. A CNN applies the same weight matrix to all channels of the input image. Therefore, herein, a model that learns features from RGB and LiDAR data separately was designed.

B. LOSS FUNCTION

The FAS model is a binary classification method that classifies input images as bonafide or spoofing. Herein, a binary cross-entropy loss function was introduced to train the proposed network; this function was also used in [42], [43], and [44]. The loss function can be described as follows.

$$BCELoss = -(y\log(p) + (1-y)\log(1-p)) \tag{3}$$

where y is the ground truth value and p is the predicted value.

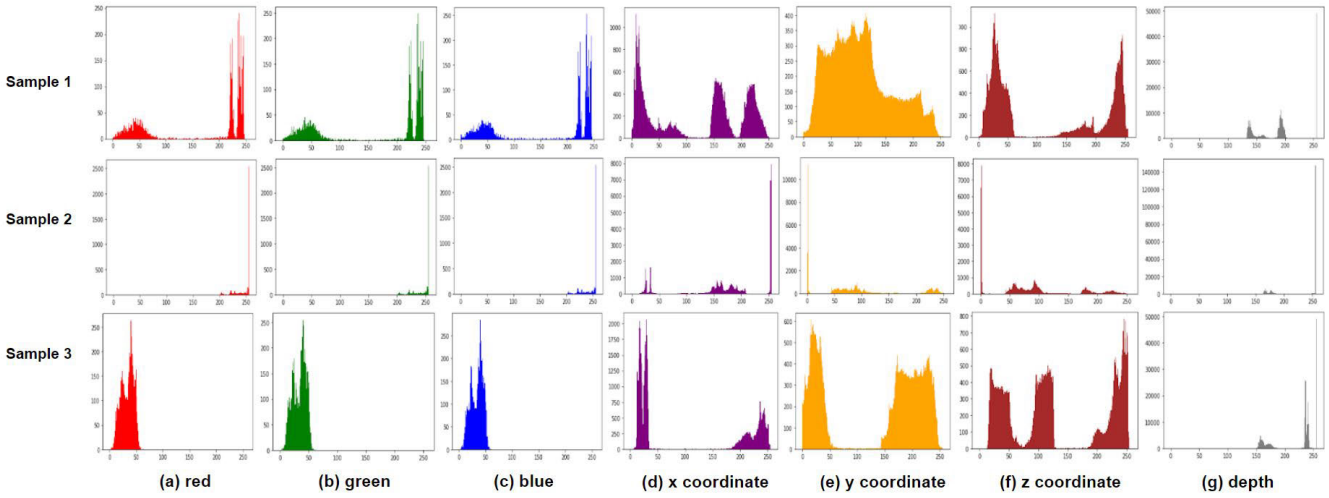


FIGURE 4. Distribution of RGB, point cloud, and depth of three random data of indoor protocol. Histograms of (a) red, (b) green, and (c) blue contributions of an RGB image. Histograms of (d) x-, (e) y-, and (f) z-coordinates of the point cloud. (g) Histogram of the depth image. The point cloud and depth exhibit very different distributions compared with the RGB.

V. NUMERICAL EXPERIMENTS

In this section, the experimental setup, evaluation metrics, experimental results, which are conducted followed by three protocols presented in Section III, ablation study, and discussion are presented.

A. EXPERIMENTAL SETUP

This study argues that a LiDAR sensor can provide an FAS model that is robust against light variation. Additionally, CloudNet is suggested suitable for training with the RGB, point cloud, and depth images. To support this argument, three models, namely Resnet34 with RGB, Resnet34 with three shots, and CloudNet with three shots (referred to as RGB, LiDAR, and CloudNet models, respectively), were employed herein. The models were trained and tested according to the three protocols mentioned in Section III. RGB, point cloud, and depth were all resized to 180×180. In training stage, we used the Adam optimizer and set the learning rate to 1e-3. The batch size was 4 on single 2080Ti GPU. We trained models with maximum 1000 epochs. He Initialization was used as the weight initialization method. All codes were implemented with pytorch.

B. EVALUATION METRICS

To evaluate the performance of CloudNet, first, the bonafide presentation classification error rate (BPCER), attack presentation classification error rate (APCER), and average classification error rate (ACER) were used as the evaluation metrics. These metrics were proposed in ISO/IEC 30107-3:2017 for performance assessment of presentation attack detection mechanisms [45]. BPCER is the proportion of bonafides incorrectly rejected as an attack. APCER is the percentage of attacks incorrectly accepted as bonafides. ACER is the average of BPCER and APCER. Additionally, a receiver operating characteristic (ROC) curve was used. To quantitatively compare the ROC curves, the area under curve (AUC) values of the graphs were determined.

C. EXPERIMENTAL RESULTS

Table 5 reports the models' BPCER, APCER, and ACER values under protocol 1. The ACER values for the RGB, LiDAR, and CloudNet models were 0.0667, 0.025, and 0, respectively. CloudNet performed the best, followed by the LiDAR and RGB models. The experimental results indicate that when the test set is in the same light domain as the training set, using point cloud and depth, although subtle, improves the performance of the FAS model. Further, CloudNet allows learning point cloud and depth images more effectively.

TABLE 5. BPCER, APCER and ACER under protocol 1.

Model	Input	BPCER	APCER	ACER
Resnet34 [41]	RGB	0.1333	0	0.0667
Resnet34 [41]	RGB, Point Cloud, Depth	0.05	0	0.025
CloudNet	RGB, Point Cloud, Depth	0	0	0

Table 6 reports the models' BPCER, APCER, and ACER values under protocol 2. The ACER for the RGB model was 0.4618. The error rate of the model increased by 0.3951 compared with protocol 1. By contrast, the ACER values for the LiDAR and CloudNet models were 0.1958 and 0.1340, respectively. This corresponds to an increase of 0.1708 and 0.1340, respectively. CloudNet had the smallest increase in errors, followed by the LiDAR and RGB models. This increase in the error rate shows how generalized the model is with respect to light variation.

Table 7 reports the models' BPCER, APCER, and ACER values under protocol 3. Compared with protocol 1, the ACER values of the models increased by 0.4111, 0.3340, and 0.1528, respectively. Similar to the experimental results obtained under protocol 2, CloudNet exhibited the smallest ACER growth, followed by the LiDAR and RGB models.

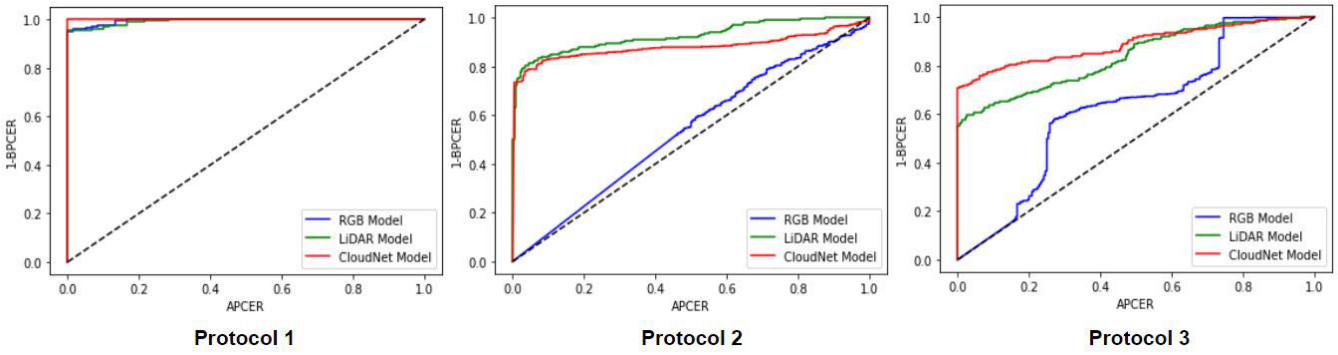


FIGURE 5. ROC curves of RGB model, LiDAR model, and CloudNet for protocols 1, 2, and 3, respectively. The AUC values are reported in Table 8.

TABLE 6. BPCER, APCER and ACER under protocol 2.

Model	Input	BPCER	APCER	ACER
Resnet34 [41]	RGB	0.4069	0.5167	0.4618
Resnet34 [41]	RGB, Point Cloud, Depth	0.3875	0.0042	0.1958
CloudNet	RGB, Point Cloud, Depth	0.2347	0.0333	0.1340

TABLE 7. BPCER, APCER and ACER under protocol 3.

Model	Input	BPCER	APCER	ACER
Resnet34 [41]	RGB	0.7889	0.1667	0.4778
Resnet34 [41]	RGB, Point Cloud, Depth	0.7181	0	0.3590
CloudNet	RGB, Point Cloud, Depth	0.3056	0	0.1528

Furthermore, the performance of the models was compared based on AUC values. First, ROC curves, shown in Fig. 5, were plotted. The AUC values were measured to quantitatively compare the ROC curves. Table 8 reports the AUC values of the three models. For protocol 1, the RGB, LiDAR, and CloudNet models had AUCs of 0.9956, 0.9931, and 1.0, respectively. For protocol 2, the AUC values decreased by 0.4594, 0.0661, and 0.1198, respectively. For protocol 3, they decreased by 0.3705, 0.1552, and 0.1094, respectively. This reduction in AUC values also supports the argument that LiDAR data render FAS models robust against light variation. The AUC value under protocol 3 also demonstrates that CloudNet is a better model than the LiDAR model. However, under protocol 2, the LiDAR model had a higher AUC value than that of CloudNet.

TABLE 8. AUC values under three protocols.

Model	Input	Protoc.1	Protoc.2	Protoc.3
Resnet34 [41]	RGB	0.9956	0.5362	0.6251
Resnet34 [41]	RGB, Point Cloud, Depth	0.9931	0.9270	0.8379
CloudNet	RGB, Point Cloud, Depth	1.000	0.8802	0.8906

Finally, we investigated the trade-offs caused by using the LiDAR sensor. Table 9 shows the number of model parameters, latency and Multiply-Adds (MAdds). The delay time was calculated by running the program that tests 100 data 100 times and taking its average and standard deviation. The latency of the LiDAR model was 2% higher than the RGB model. The number of MAdds for the parameters increased by 0.01M, 0.1G. On the other hand, CloudNet's latency increased by 16% compared to the RGB model and the number of parameters and MAdds also increased by almost double.

TABLE 9. Latency, params and MAdds of models.

Model	Input	Latency(s)		Params(M)	MAdds(G)
		Mean	Std		
Resnet34 [41]	RGB	4.06	0.03	21.29	2.57
Resnet34 [41]	RGB, Point Cloud, Depth	4.16	0.24	21.30	2.67
CloudNet	RGB, Point Cloud, Depth	4.70	0.14	42.57	5.16

D. ABLATION STUDY

In addition, ablation studies were performed. Additional experiments were performed using the point cloud and depth data. For the multi-modal models, the approach of early fusion, late fusion, and hybrid fusion was applied. Experiments in Section IV were conducted for the cases of: point cloud only, depth only, RGB and point cloud combined, and RGB and depth combined. The experimental results are listed in Table 10. According to the results of the point cloud and depth experiments conducted under protocol 1, these data are not suitable for performing FAS operations on their own compared to RGB model. Unlike the RGB model with an error rate of 0.0667, the point cloud and depth models exhibited high error rates of 0.3028 and 0.2750, respectively. Essentially, if the learning and test datasets are in the same domain, RGB provides stronger discrimination compared with point cloud or depth. However, the experimental results obtained under protocols 2 and 3 suggest that the point cloud and depth are negligibly affected by light variations. This is an obvious advantage that RGB does not have. Next, the model performance was investigated using RGB and point cloud, and RGB and depth. The experimental results indicated that models built using RGB and depth performed better than those constructed using RGB and point cloud. Furthermore, additional ablation studies confirmed that training RGB and LiDAR data separately was effective. All models, including those using RGB and point cloud, RGB and depth, RGB, point cloud, and depth, demonstrated better performance when using a late fusion approach instead of an early fusion approach. Additionally, when training RGB, point cloud, and

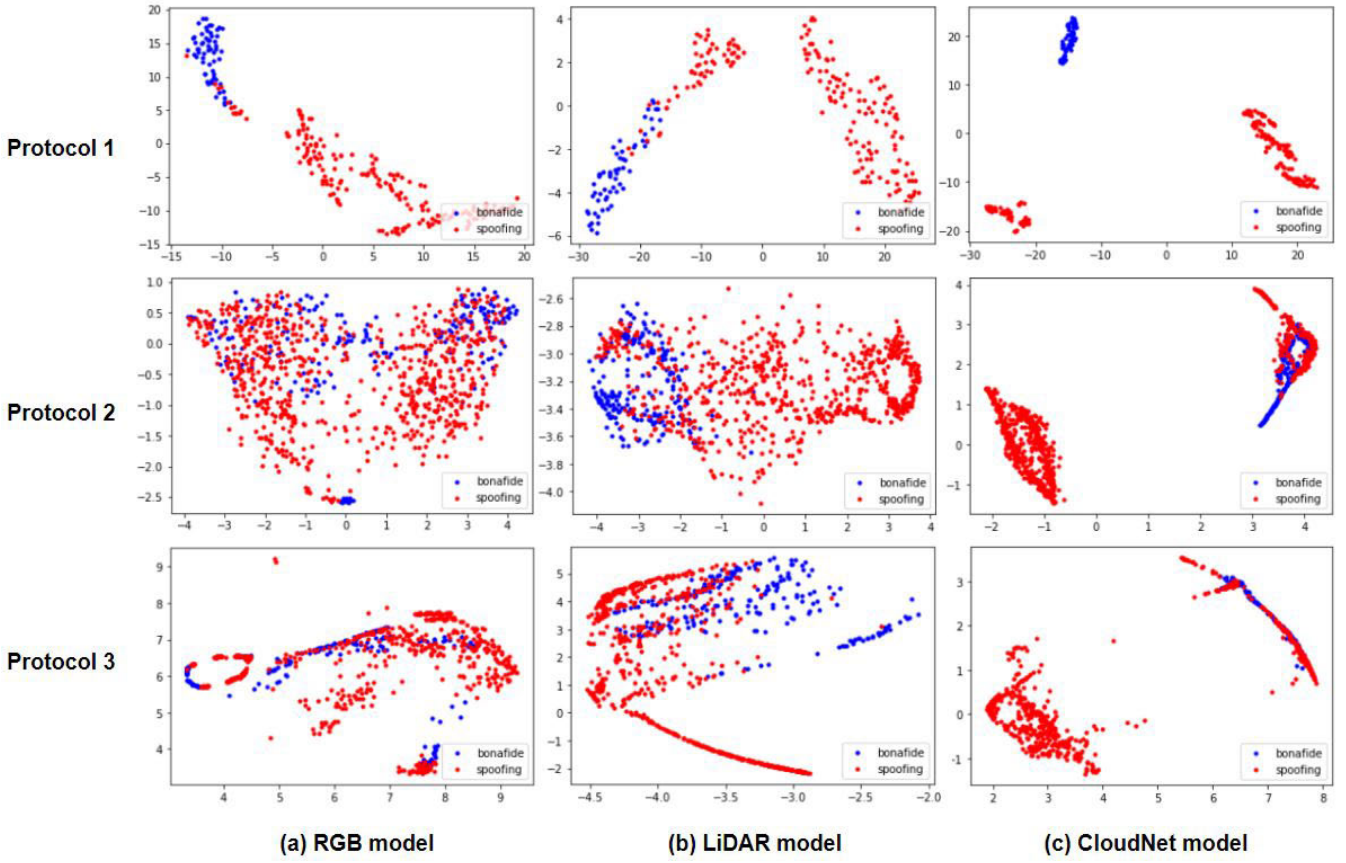


FIGURE 6. Feature distribution diagrams of RGB, LiDAR, and CloudNet models expressed using T-distributed stochastic neighbor embedding (T-SNE) technique. Blue dots represent bonafides and red dots represent spoofing attacks. Plot results obtained using (a) Resnet34 with RGB. (b) Resnet34 with RGB, point cloud and depth, and (c) CloudNet with RGB, point cloud and depth.

TABLE 10. BPCER, APCER and ACER including ablation study under three protocols.

Model	Input	Fusion Approch	Protocol1 (Indoor)			Protocol2 (Outdoor)			Protocol3 (Indoor-dark)		
			BPCER	APCER	ACER	BPCER	APCER	ACER	BPCER	APCER	ACER
Resnet34 [41]	RGB	-	0.1333	0	0.0667	0.4069	0.5167	0.4618	0.7889	0.1667	0.4778
Resnet34 [41]	Point Cloud	-	0.1556	0.4500	0.3028	0.1847	0.3792	0.2819	0.1542	0.5625	0.3583
Resnet34 [41]	Depth	-	0.1000	0.4500	0.2750	0.2292	0.3208	0.2750	0.1403	0.2417	0.1910
Resnet34 [41]	RGB, Point Cloud	Early Fusion	0.0388	0.0500	0.0444	0.0888	0.7625	0.4257	0.6208	0.1625	0.3917
Resnet34 [41]	RGB, Point Cloud	Late Fusion	0.0111	0	0.0055	0.1861	0.2042	0.1951	0.2847	0.0583	0.1715
Resnet34 [41]	RGB, Depth	Early Fusion	0	0	0	0.2792	0.3917	0.3354	0.1667	0.5722	0.3694
Resnet34 [41]	RGB, Depth	Late Fusion	0	0	0	0.1542	0.1417	0.1479	0.3083	0	0.1542
Resnet34 [41]	RGB, Point Cloud, Depth	Early Fusion	0.0500	0	0.0250	0.3875	0.0042	0.1958	0.7181	0	0.3590
Resnet34 [41]	RGB, Point Cloud, Depth	Late Fusion	0	0	0	0.1514	0.1792	0.1653	0.3167	0.0042	0.1605
Resnet34 [41] (CloudNet)	RGB, Point Cloud, Depth	Hybrid Fusion (Early + Late)	0	0	0	0.2347	0.0333	0.1340	0.3056	0	0.1528

depth together, the use of both early and late fusion, such as in the CloudNet model, resulted in better performance than using only late fusion.

Lastly, the extracted features were visualized to determine how well the proposed model classifies bonafide and spoofing images. The T-distributed stochastic neighbor embedding (T-SNE) technique was used to transform high-dimensional features extracted by deep learning models into 2D features [46]. This technique was applied to the models we experimented with in section IV. Fig. 6 shows the feature distributions of the models expressed by the T-SNE.

E. DISCUSSION

Through the experiments, we have found that the performance of RGB model is severely poor when tested on datasets with domain shift in light. Compared to protocol1 where the light domains of the training and test sets were the same, the performance of the RGB model greatly decreased in protocols 2 and 3 where the light domains were different. On the other hand, LiDAR sensors have been found to improve the performance of FAS models and make them more robust to light changes, as confirmed by experimental results and an ablation study. In addition, CloudNet could

further improve the performance of the LiDAR model. This suggests that optimizing the way LiDAR data is trained can improve the FAS model. Meanwhile, it is necessary to note that there is a trade-off involved. The LiDAR model's computational cost, measured in the number of model parameters and MAdds, is similar to that of the RGB model, with only a small difference of 0.01M and 0.1G. However, CloudNet's computational cost is twice that of the RGB model. This suggests that the increase in cost is primarily due to the CloudNet structure, rather than the use of LiDAR sensor. Therefore, it is considered a future study to reduce the cost of the CloudNet model while maintaining performance.

VI. CONCLUSION

In this study, an FAS model that uses a LiDAR sensor with an RGB camera was proposed. LiDAR provides 3D coordinate and depth information and has the advantage of robustness to light variation. Herein, the LDFAS was constructed to verify the superiority of the model. LDFAS consists of three subsets with different light variations. Based on this, with three different protocols were chosen for experimenting: 1) the same light domain, 2) brighter light domain, and 3) darker light domain, compared with the training set. Additionally, CloudNet was designed to learn separate weights for the RGB and LiDAR data (point cloud and depth). The experimental results revealed that using a LiDAR sensor provides robustness to light variation compared with the RGB model. In addition, CloudNet performed better than RGB and LiDAR models. However, the current CloudNet model also had the drawback of being heavier than a regular LiDAR model. This means that there is a possibility for LIDAR-based FAS models to improve. The task of studying a better LiDAR-based FAS model through model lightweighting will be left as a future study task.

ACKNOWLEDGMENT

(Yongrae Kim and Hyunmin Gwak are co-first authors.)

REFERENCES

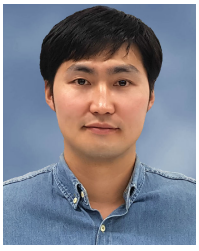
[1] J.-W. Li, "Eye blink detection based on multiple Gabor response waves," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 5, Jul. 2008, pp. 2852–2856.
[2] J. Bigun, H. Fronthaler, and K. Kollreider, "Assuring liveness in biometric identity authentication by real-time face tracking," in *Proc. IEEE Int. Conf. Comput. Intell. Homeland Secur. Pers. Saf. (CIHSPS)*, Jul. 2004, pp. 104–111.
[3] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 389–398.
[4] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 389–398.
[5] Z. Yu, Y. Qin, X. Li, Z. Wang, C. Zhao, Z. Lei, and G. Zhao, "Multi-modal face anti-spoofing based on central difference networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 650–651.
[6] G. B. de Souza, J. P. Papa, and A. N. Marana, "On the learning of deep local features for robust face spoofing detection," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 258–265.
[7] B. Chen, W. Yang, and S. Wang, "Face anti-spoofing by fusing high and low frequency features for advanced generalization capability," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Aug. 2020, pp. 199–204.
[8] Y. Zuo, W. Gao, and J. Wang, "Face liveness detection algorithm based on livenesslight network," in *Proc. Int. Conf. High Perform. Big Data Intell. Syst. (HPBD&IS)*, May 2020, pp. 1–5.
[9] G. Heusch, A. George, D. Geissbühler, Z. Mostaani, and S. Marcel, "Deep models and shortwave infrared information to detect face presentation attacks," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 4, pp. 399–409, Jul. 2020.
[10] J. Connell, N. Ratha, J. Gentile, and R. Bolle, "Fake iris detection using structured light," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8692–8696.
[11] J. Seo and I.-J. Chung, "Face liveness detection using thermal face-CNN with external knowledge," *Symmetry*, vol. 11, no. 3, p. 360, Mar. 2019.
[12] M. Liu, H. Fu, Y. Wei, Y. A. U. Rehman, L.-M. Po, and W. L. Lo, "Light field-based face liveness detection with convolutional neural networks," *J. Electron. Imag.*, vol. 28, no. 1, 2019, Art. no. 013003.
[13] Y. Tian, K. Zhang, L. Wang, and Z. Sun, "Face anti-spoofing by learning polarization cues in a real-world scenario," in *Proc. 4th Int. Conf. Adv. Image Process.*, Nov. 2020, pp. 129–137.
[14] B. Wu, A. Wan, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1887–1893.
[15] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, "SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 4376–4382.
[16] T. Cai, F. Chen, W. Liu, X. Xie, and Z. Liu, "Face anti-spoofing via conditional adversarial domain generalization," *J. Ambient Intell. Hum. Computing.*, vol. 13, pp. 1–14, 2022.
[17] L. Zhang and C. Zhang, "A MrPPG method for face anti-spoofing," in *Proc. IEEE Asia–Pacific Conf. Image Process., Electron. Comput. (IPEC)*, Apr. 2022, pp. 823–827.
[18] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing," *Pattern Recognit.*, vol. 115, Jul. 2021, Art. no. 107888.
[19] L. Birla, P. Gupta, and S. Kumar, "SUNRISE: Improving 3D mask face anti-spoofing for short videos using pre-emptive split and merge," *IEEE Trans. Depend. Secure Comput.*, early access, Apr. 19, 2022, doi: 10.1109/TDSC.2022.3168345.
[20] J.-D. Lin, H.-H. Lin, J. Dy, J.-C. Chen, M. Tanveer, I. Razzak, and K.-L. Hua, "Lightweight face anti-spoofing network for telehealth applications," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 1987–1996, May 2022.
[21] A. Liu, Z. Tan, J. Wan, Y. Liang, Z. Lei, G. Guo, and S. Z. Li, "Face anti-spoofing via adversarial cross-modality translation," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2759–2772, 2021.
[22] F. Jiang, P. Liu, X. Shao, and X. Zhou, "Face anti-spoofing with generated near-infrared images," *Multimedia Tools Appl.*, vol. 79, nos. 29–30, pp. 21299–21323, Aug. 2020.
[23] S. Jia, X. Li, C. Hu, G. Guo, and Z. Xu, "3D face anti-spoofing with factorized bilinear coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 4031–4045, Oct. 2021.
[24] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: A survey," 2021, *arXiv:2106.14948*.
[25] S. Saha, W. Xu, M. Kanakis, S. Georgoulis, Y. Chen, D. P. Paudel, and L. Van Gool, "Domain agnostic feature learning for image and video based face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 802–803.
[26] R. Shao, X. Lan, and P. C. Yuen, "Regularized fine-grained meta face anti-spoofing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11974–11981.
[27] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, "CASIA-SURF CeFA: A benchmark for multi-modal cross-ethnicity face anti-spoofing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1179–1187.
[28] M. Rostami, L. Spinoulas, M. Hussein, J. Mathai, and W. Abd-Almageed, "Detection and continual learning of novel face presentation attacks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14851–14860.
[29] Z. Zou, X. Zhang, H. Liu, Z. Li, A. Hussain, and J. Li, "A novel multimodal fusion network based on a joint-coding model for lane line segmentation," *Inf. Fusion*, vol. 80, pp. 167–178, Apr. 2022.
[30] Z. Lian, B. Liu, and J. Tao, "SMIN: Semi-supervised multi-modal interaction network for conversational emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Jan. 7, 2022, doi: 10.1109/TAFFC.2022.3141237.

[31] Z. Huang, X. Mo, and C. Lv, "Multi-modal motion prediction with transformer-based neural network for autonomous driving," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 2605–2611.
[32] M. A. Uddin, J. B. Joolee, and K.-A. Sohn, "Deep multi-modal network based automated depression severity estimation," *IEEE Trans. Affect. Comput.*, early access, Jun. 1, 2022, doi: 10.1109/TAFFC.2022.3179478.
[33] J. Schlosser, C. K. Chow, and Z. Kira, "Fusing LIDAR and images for pedestrian detection using convolutional neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 2198–2205.
[34] R. Roriz, J. Cabral, and T. Gomes, "Automotive LiDAR technology: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6282–6297, Jul. 2022.
[35] K. Ko, H. Gwak, N. Thoummala, H. Kwon, and S. Kim, "SqueezeFace: Integrative face recognition methods with LiDAR sensors," *J. Sensors*, vol. 2021, pp. 1–8, Sep. 2021.
[36] S. Kim, H. Moon, J. Oh, Y. Lee, H. Kwon, and S. Kim, "Automatic measurements of garment sizes using computer vision deep learning models and point cloud data," *Appl. Sci.*, vol. 12, no. 10, p. 5286, May 2022.
[37] B. Yang, Y. Cheng, Z. Jin, X. Ji, and W. Xu, "Generating 3D adversarial point clouds under the principle of LiDARs," 2022.
[38] C. Xu, B. Wu, Z. Wang, W. Zhan, P. Vajda, K. Keutzer, and M. Tomizuka, "SqueezeSegV3: Spatially-adaptive convolution for efficient point-cloud segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 1–19.
[39] A. Xiao, X. Yang, S. Lu, D. Guan, and J. Huang, "FPS-net: A convolutional fusion network for large-scale LiDAR point cloud segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 237–249, Jun. 2021.
[40] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018, *arXiv:1811.03378*.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[42] W. Liu, X. Wei, T. Lei, X. Wang, H. Meng, and A. K. Nandi, "Data-fusion-based two-stage cascade framework for multimodality face anti-spoofing," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 2, pp. 672–683, Jun. 2022.
[43] G. Te, W. Hu, and Z. Guo, "Exploring hypergraph representation on face anti-spoofing beyond 2D attacks," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2020, pp. 1–6.
[44] Q. Yang, X. Zhu, J.-K. Fwu, Y. Ye, G. You, and Y. Zhu, "PipeNet: Selective modal pipeline of fusion network for multi-modal face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 644–645.
[45] *Information Technology Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, Standard ISO/IEC DIS 30107-3:2017, Geneva, Switzerland, 2017.
[46] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.

YONGRAE KIM received the B.S. degree from the Department of Computer Engineering, Handong University, South Korea, in 2017. He is currently pursuing the master's degree with the Department of Computer Science and Engineering, Korea University. His research interests include machine learning and computer engineering.

HYUNMIN GWAK received the B.S. degree from the Department of Statistics, Keimyung University, South Korea, in 2017. He is currently pursuing the master's degree with the Department of Applied Statistics, Konkuk University. His research interests include machine learning and computer vision.

JAEHOON OH received the B.S. degree from the Department of Applied Statistics, Purdue University, West Lafayette, IN, USA, in 2021. He is currently pursuing the master's degree with the Department of Applied Statistics, Konkuk University. His research interests include risk allocation and reinforcement learning.



MINHO KANG is a member of the CMS Group, LHC. He has been researching big data-based research and development with the Department of Physics, Korea University, since 2006, and continues to upgrade Phase-2 Mun on the next generation RPC development for RE3/4-1 chambers, in 2017.

JINKYU KIM received the B.S. degree in electrical engineering and the M.S. degree in electrical and computer engineering from Korea University, South Korea, and the Ph.D. degree in computer science from the University of California at Berkeley, Berkeley, CA, USA, in 2019. He is currently a Professor with the Department of Computer Science, Korea University. His research interests include deep learning for self-driving vehicles, representation learning, explainable AI (XAI), and advisable AI.

HYUN KWON received the B.S. degree in mathematics from the Korea Military Academy, South Korea, in 2010, and the M.S. and Ph.D. degrees from the School of Computing, Korea Advanced Institute of Science and Technology (KAIST), in 2015 and 2020, respectively. He is currently an Assistant Professor with the Korea Military Academy. His research interests include information security, computer security, and intrusion tolerant systems.

SUNGHWAN KIM received the B.A. and M.S. degrees from the Department of Statistics, Korea University, and the Ph.D. degree in biostatistics from the University of Pittsburgh, Pittsburgh, PA, USA, in 2015. He is currently working as an Assistant Professor with the Department of Applied Statistics, Konkuk University. His research interests include deep learning-based models to address the domain problems in the context of vision analysis and omic-data integration.