



Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor

SHIJIE RAO, YIDONG HUANG, KAIYU CUI,  AND YALI LI*

¹Department of Electronic Engineering, Tsinghua University, Shuangqing Road, Haidian District, Beijing 100084, China

*Corresponding author: liyali13@tsinghua.edu.cn

Received 5 July 2022; revised 6 September 2022; accepted 24 September 2022; published 8 November 2022

Modern face recognition systems usually combine RGB, depth, and infrared cameras to do face antispoofing, but they are still not robust enough to unknown 3D high-quality mask attack. In our work, we developed a snapshot hyperspectral image sensor based on metasurface nanostructures to obtain the high-precision hyperspectral information of the detected face, and we built a practical antispoofing face recognition system using our new sensor. Experiments show that our sensor can reconstruct the reflectance spectrum of human skin, and this spectral information captured by our sensor can be quite effective and robust to identify spoof faces. We attack our system with several types of spoof faces, and our system reaches 97.98% accuracy in real-world testing scenes. © 2022 Optica Publishing Group under the terms of the Optica

Open Access Publishing Agreement

<https://doi.org/10.1364/OPTICA.469653>

1. INTRODUCTION

Facial recognition is a biometric identity authentication method. Unlike traditional passwords, biometric properties cannot be forgotten or stolen, thus providing better security [1]. But one can still display a printed photo or wear a mask to attack the face recognition system [2]. To defend against these attacks, some systems use technics to acquire 3D images of the face and do 3D face recognition [3,4]. Some systems do video analysis to identify fake faces [5,6]. Some systems use algorithms to do anomaly detection in RGB images and detect fake faces [7]. However, with the development of 3D printing and biomimetic silicone technology, some 3D disguise masks can look so real that even human eyes can be deceived [8]. These 3D masks bring a great challenge to current face recognition systems [9,10]. Moreover, current face antispoofing (FAS) algorithms based on RGB, RGBD, or NIR images usually perform badly on interdata set evaluations [11]. This domain adaptation problem makes the performance of FAS systems deployed in the real world unpredictable. Therefore, besides using these common cameras, some FAS methods try to capture more robust and discriminating features between real and fake faces using advanced sensors, such as SWIR [12], thermal cameras [13], light-field cameras [14], and polarized cameras [15]. But these sensors are either too expensive, too large, or not convenient to use, which makes them less practical to be integrated into real-world face recognition systems [16].

Spectrum analysis is an effective tool to identify different materials. Because of the absorption of hemoglobin in human blood, the reflectance spectrums of human skin [17] have two minimum points, at 545 and 575 nm (see Fig. S1 in Supplement 1), which

is hardly imitable using masks. FAS approaches based on hyperspectral images (HSIs) are most effective and reliable [18,19] and are more robust than approaches based on RGB cameras [20,21]. But traditional hyperspectral cameras rely on optics grating and mechanical scanning systems. They are usually expensive, bulky, and cost a lot of time to capture one HSI. Hence, like the other advanced sensors mentioned above, they are also impractical in real-world use. And there are few public FAS data sets based on HSIs [22]. In recent years, on-chip spectral imaging sensor technology has been developed rapidly. Using silicon metasurfaces and computational imaging technics, an on-chip snapshot HSI sensor can be achieved [23,24], which can capture HSI data at a video rate and may bring hyperspectral sensing into everyday life.

In our work, we adopted the on-chip snapshot HSI technics to build a hyperspectral image sensor and designed a novel FAS system. Compared with existing FAS works, our anti-spoofing face recognition system is based on hyperspectral information. Our work is the first (we believe) to build a practical antispoofing face recognition system using an integrated hyperspectral sensor. It can effectively detect almost all kinds of spoofing attacks using spectrum analysis and is reliable to unknown attacks in the real world. Compared with the traditional hyperspectral imager, our on-chip hyperspectral image sensor is much cheaper, faster, and smaller (see Fig. S2 and Table S1 in Supplement 1). As a traditional hyperspectral camera needs over 100 s to scan one HSI, our sensor only needs a 50-ms snapshot. It can measure the reflectance spectrum of faces at high accuracy and reveal the absorption peaks of hemoglobin. And it can be easily integrated into existing face recognition systems.

2. RESULTS

A. Hyperspectral Image Sensor

We designed and fabricated a snapshot hyperspectral image sensor to obtain hyperspectral information about the captured object. The overall architecture of our FAS system is shown in Fig. 1. Resonant metasurface structures can be used to frequency-modulate incident light [25]. By integrating the metasurface structures on top of a CMOS image sensor (CIS), the intensity of the modulated light is measured by CIS. Combined with a spectrum reconstruct algorithm, the one-shot HSI can be realized [23,26]. The HIS device consists of tens of thousands of microspectrometers. Each microspectrometer is achieved by combining metasurface units with CIS. For a single microspectrometer, assuming that the spectrum of the incident light is $F(\lambda)$, N metasurface units of transmission response $H_i(\lambda)$, $i = 1, 2, 3, \dots, N$ are attached to the surface of the CIS. The absorption response of the CIS imaging system is $T(\lambda)$. Then, the signal intensity I_i received by the CIS with the i th metasurface can be described as

$$I_i = \int F(\lambda) H_i(\lambda) T(\lambda) d\lambda. \quad (1)$$

In the discrete form, I_i can be described as

$$I_i = \sum_{\lambda} f(\lambda) b_i(\lambda) t(\lambda), \quad (2)$$

where $f(\lambda)$, $b_i(\lambda)$, $t(\lambda)$ is the discrete sampling of $F(\lambda)$, $H_i(\lambda)$, $T(\lambda)$. $b_i(\lambda)$ and $t(\lambda)$ can be measured through experiment. Let $r_i(\lambda) = b_i(\lambda)t(\lambda)$; then we get

$$I_i = \sum_{\lambda} f(\lambda) r_i(\lambda). \quad (3)$$

For N metasurface units, I_i , $i = 1, 2, 3, \dots, N$ are measured independently. Therefore, $I = (I_1, I_2, \dots, I_N)^T \in R^{N \times 1}$ can be described in the matrix form,

$$I = Rf, \quad (4)$$

where $f = (f(\lambda_1), f(\lambda_2), \dots, f(\lambda_M))^T \in R^{M \times 1}$ and $R = (r_1(\lambda), r_2(\lambda), \dots, r_N(\lambda))^T \in R^{N \times M}$.

In our work, we sampled $F(\lambda)$, $H_i(\lambda)$, $T(\lambda)$ from 450 to 750 nm at an interval of 0.5 nm; thus $M = 601$. And 49

metasurface units arranged in a 7×7 array are used as a single spectrometer; thus, $N = 49$. In order to improve the performance of the metasurface-based hyperspectral image sensor, we chose 49 different metasurface units from the database to optimize the column correlation of matrix R . As N is much smaller than M , in practice, a compressive sensing (CS) algorithm is used to reconstruct f ; we adopted the following convex optimization method to solve the CS problem:

$$\min_y \{ \|RDy - I\|_2^2 + \alpha \|y\|_1 \}, \quad f = Dy, \quad (5)$$

where D is the sparse dictionary trained using a spectrum database [27] and a K-SVD [28] algorithm according to [29]. α is the weights of l_1 -norm regularization.

We adopted the freeform-shaped meta-atoms method [30] to generate and optimize metasurfaces, thus reducing the column correlation of R (more details can be found in Algorithm S1 in Supplement 1). By arranging the microspectrometers in a 2D array, a hyperspectral facial imaging device can be achieved. Combined with the spectrum reconstruction algorithm, the snap-shot hyperspectral image sensor can be realized (see Fig. 2).

First, we tested the performance of our hyperspectral image sensor. We used our sensor to capture and reconstruct the reflectance spectrum of a live face and some spoof faces. The results are shown in Fig. 3. The images are captured under white LED lighting. In Fig. 3, blue curves represent the data acquired from GaiaField Pro-V10 [31] commercial hyperspectral camera. They are considered as the reference for comparison. Orange curves represent the spectrums reconstructed by our sensor. Figure 3 shows that our HSI sensor can measure the reflectance spectrum at a relatively high accuracy. Cosine similarity between the reconstructed spectrum and the reference is adopted to quantify the performance of our sensor. And, on the four test samples, which are a live face [Fig. 3(a)], a paper mask [Fig. 3(b)], a silicone mask [Fig. 3(c)], and raw silicone material [Fig. 3(d)], respectively, the cosine similarities are 99.25%, 99.76%, 99.80%, and 99.68%. Moreover, the commercial HSI camera needs more than 100 s to capture one image, while our sensor only needs a snapshot within 50 ms. Figure 3(a) shows that there are two hemoglobin characteristic absorption peaks, at 545 and 575 nm, in the live face test sample (more quantitative analysis can be found in Fig. S5 in Supplement 1). These two valleys at around 545 and 575 nm are labeled as A1

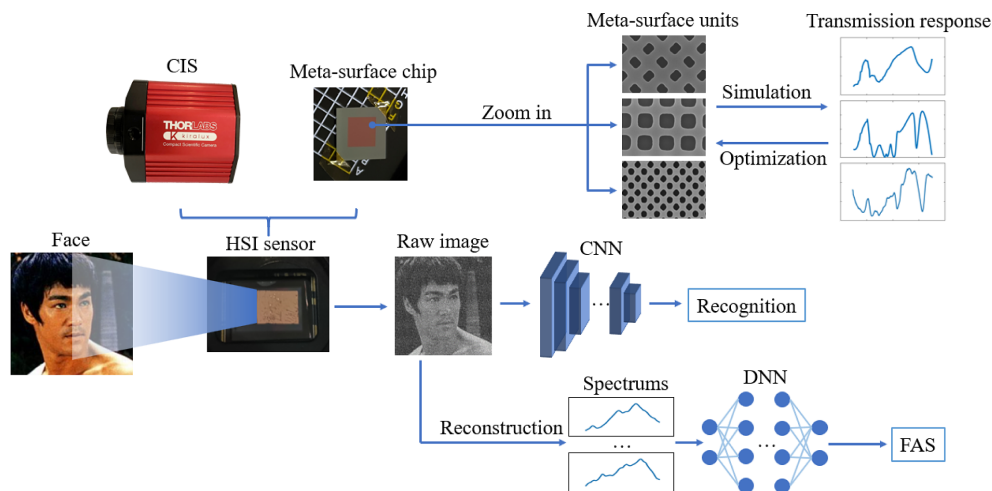


Fig. 1. Overall architecture of our proposed method.

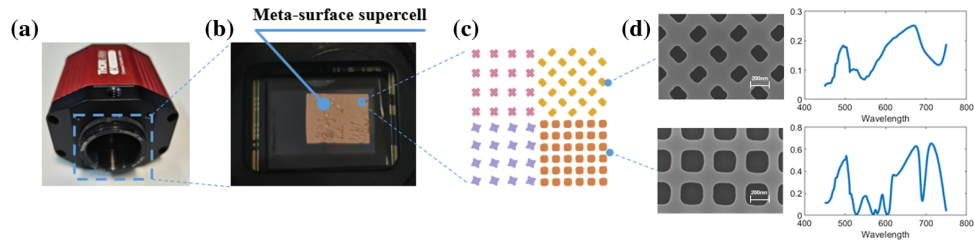


Fig. 2. Structure of the designed HSI image sensor. (a) Fabricated HSI image sensor; (b) a metasurface supercell is attached to the surface of the CIS. (c) A microspectrometer consists of N different metasurface units, and the metasurface supercell consists of more than 40,000 microspectrometers. (d) The transmission responses of different metasurface units are designed to have a minimum correlation.

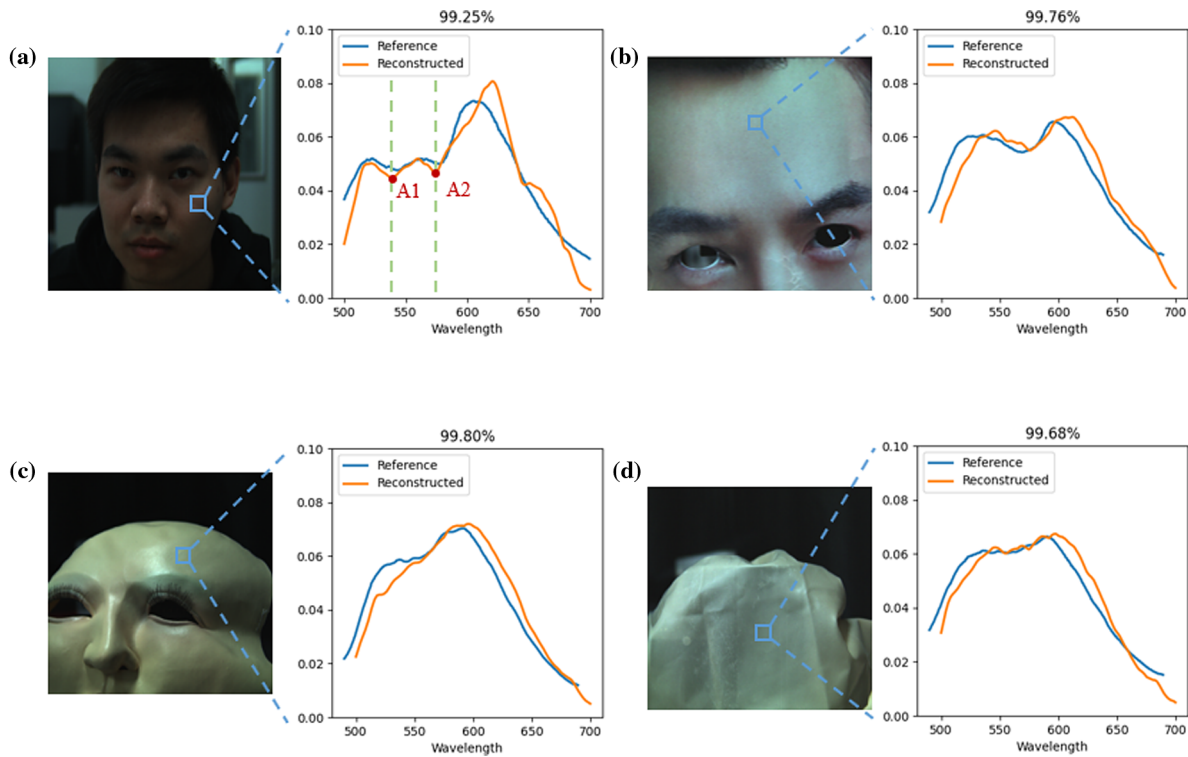


Fig. 3. Spectrum measure results of our sensor. (a) Live face; (b) paper mask; (c) silicone mask; (d) raw silicone material.

and A2 in Fig. 3(a). The reflectance spectrum curve is W-shaped at around 520–600 nm, which is the unique feature of human skin reflectance spectrum. And the reconstructed spectrums of spoof faces [Figs. 3(b)–3(d)] do not have this specific characteristic, which indicates that our HSI sensor is capable of doing antispoofing face recognition. More results about the reconstructed spectrums of spoof and live faces under different lighting conditions can be found in Fig. S3 in Supplement 1.

B. Face Antispoofing

Our hyperspectral face image camera can measure the spectrum of the visible band (450 ~ 750 nm). Considering that the characteristics of human skin reflectance spectrum are at about 545 and 575 nm, we used the spectral range of 500 ~ 650 nm to do FAS. And we adopted the mean-blur method to denoise, thus reducing the positive distortion of the reconstructed spectrums. That is, to get the spectrum of a certain key point, the spectrums of nine nearby pixels are reconstructed and averaged. For each spectrum sample, the vector is normalized to a mean of 0 and a variance of 1. We designed a FAS classifier based on transformer [32] (see Fig. 4)

and used our own hyperspectral FAS data set (see Algorithm S2 in Supplement 1) to train and evaluate the classifier. The detailed specification of our network architecture and training procedure can be found in Algorithm S3 in Supplement 1.

Figure 5 shows two samples from the testing set. The upper one is a live face and the lower one is a screen-displayed face. From the images taken by RGB camera [Figs. 5(a) and 5(b)], live and spoof faces are hard to distinguish. However, using hyperspectral information, the spoof face can be easily detected [Figs. 5(g) and 5(h)]. Spectral analysis makes antispoofing live skin segmentation possible. After training, we evaluated the FAS classifier on the testing set; the receiver operating characteristic (ROC) curve and confusion matrix are shown in Fig. 6. We employed several metrics to evaluate our model, such as accuracy, area under curve (AUC), false rejection rate (FRR), and false acceptance rate (FAR). AUC is the area under the ROC curve. It is one of the common metrics used to evaluate the performance of a binary classifier. We regard live faces as positives and spoof faces as negatives. Then FRR represents the proportion of live faces misclassified as spoof faces, and FAR represents the proportion of spoof faces misclassified as live faces. Our method reaches 98.75% AUC and 95.42% accuracy.

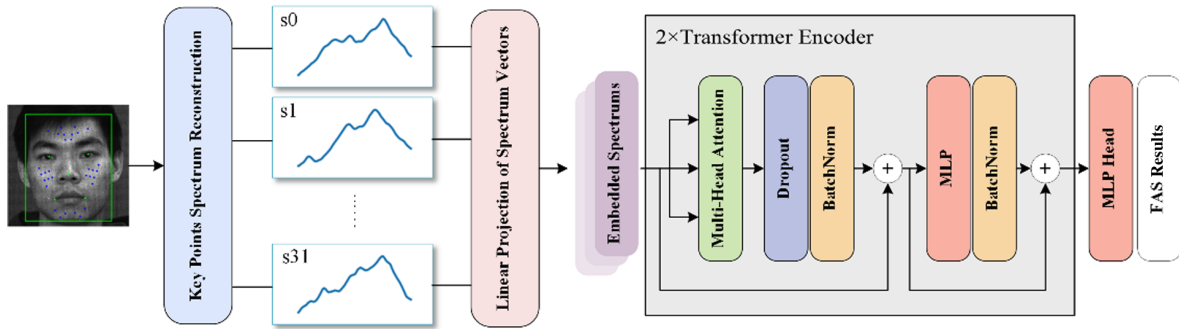


Fig. 4. Architecture of the FAS classifier. For each input, there are 32 spectrum sample vectors. By adopting the self-attention mechanism using Transformer Encoder, not only are the features of each spectrum sample analyzed respectively, but also the cross-correlations of 32 samples are taken into account. Finally, the output vector by Transformer Encoder is sent to a multilayer perceptron (MLP) to get the final FAS results. More detailed information about the network can be found in Algorithm S3.

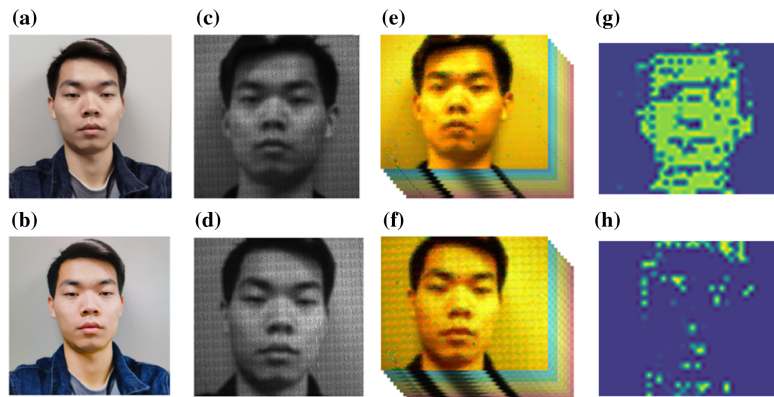


Fig. 5. Which one is real? (a), (b) Images taken by an RGB camera; (c), (d) output of raw images of our sensor; (e), (f) reconstructed HSIs; (g), (h) pixel-level liveness detection results by FAS classifier. Bright yellow pixels indicate the area where live skin is detected.

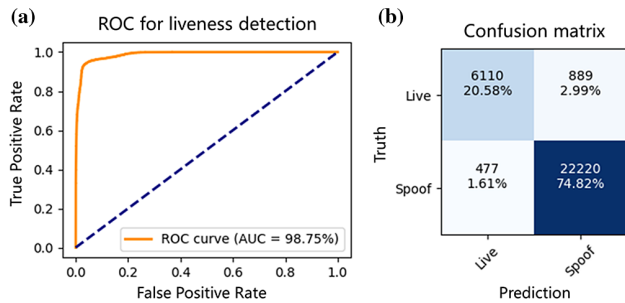


Fig. 6. Testing results of FAS classifier. (a) ROC curve of the testing results. Our classifier achieves an AUC of 98.75%, which indicates that the classifier performs well on the FAS task. (b) Confusion matrix of the testing results; the accuracy of our classifier is 95.42%. Only 2.1% of the spoof faces are misclassified to the live face.

The FRR is 12.70%, and the FAR is only 2.10%, which means that only 2.10% of the spoof faces are misclassified as live faces. Further analysis shows that the performance of the spectrum measurement has a greater impact on the results than the FAS classification algorithm. Measurement noise may distort the reconstructed spectrum, which results in a relatively high FRR. In addition, we utilized a randomly points sampling method to generate the data set, which resulted in a certain number of noisy samples and reduced FRR. By improving the fabrication of the image sensor and the spectrum reconstruction algorithm, the FAS performance can be further improved.

C. Study of Cross-Domain Evaluation

In the intradomain evaluation, both the training data and the testing data are taken under sunlight and LED light. The results are shown as Exp. 1 in Table 1. Then, to test the robustness of our method, we did some studies of cross-domain evaluation. As the spectrum of the light source may have an effect on the reflection spectrum measurement, we trained and evaluated our FAS method on HSIs taken under different light sources. In Exp. 2, we trained the network on the data taken under LED light and tested the network on the data taken under sunlight. As data taken under sunlight were completely unseen to the network during the training procedure, the testing accuracy and AUC dropped a little. These results show that the light source does have a certain effect on FAS performance, especially on FRR. If the classifier is trained and evaluated on HSIs taken under different light sources, the FRR could be relatively high. In Exp. 3, both the training data and testing data are obtained only under sunlight. Therefore, the network is trained and evaluated specially for a sunlight situation. In this experiment, the accuracy and AUC both improved slightly,

Table 1. Cross-Domain Evaluation Results

	Training Data		Testing Data		Acc. (%)	AUC (%)	FRR (%)	FAR (%)
	Sun	LED	Sun	LED				
Exp. 1	✓	✓	✓	✓	95.42	98.75	12.70	2.10
Exp. 2		✓	✓		88.95	92.77	24.21	5.77
Exp. 3	✓		✓		96.37	98.65	6.07	2.64

and the FRR dropped dramatically. The results demonstrate that if the classifier is trained and evaluated on HSIs taken under the same light source, the FRR would be much lower. Therefore, if the light conditions of the deployment site are known, we can train the network focused on the data taken under the certain light source, and the performance of our FAS method can be further improved.

D. Antispoofing Face Recognition System

To build a practical real-time antispoofing face recognition system, we only use spectrums from 32 face key points to do FAS. On the one hand, it is unnecessary to use spectral information of every pixel to do FAS. On the other hand, more key points mean more spectrums need to be reconstructed, which increases the computation and data storage cost. Therefore, we used only 32 face key points to build an antispoofing face recognition demo. Certainly, in real-world applications, the number of key points can be chosen appropriately. More key points bring better FAS performance, thus providing better security for the face recognition system.

As is shown in Fig. 1, face recognition and face detection are done on the output raw gray-scale image of the sensor directly. And face key points are located using face alignment results (see Algorithm S4 in Supplement 1). Then the spectrums of 32 key points are reconstructed and sent to the FAS classifier. Finally, the face recognition and antispoofing results are output by the system. We deployed our sensor in the real world and recorded videos of live and spoof faces under a variety of illuminations, including sunlight and LED light. A total of 1637 valid frames of 20 different

identities were recorded, of which 428 were of live faces and 1209 were of spoof faces. Our system got 100% accuracy on face recognition and 97.98% accuracy on face antispoofing, and we reached 0% and 7.71% on FAR and FRR. Some testing results are shown in Fig. 7.

3. DISCUSSION

FAS has been a challenging task for all the current face recognition systems. In our work, we proposed a novel FAS method in the combination of software and hardware. A snapshot hyperspectral image sensor is designed and fabricated to get the hyperspectral information of the captured face. Face detection and recognition are done on the raw output image of our sensor. Then, the spectrums of the face key points are reconstructed, and FAS is done by analyzing the spectrum information. Compared with existing work, our method introduced hyperspectral information to the face recognition system using only a snapshot, which is more robust and reliable than other unknown spoofing attacks. And our sensor is based on CMOS-compatible metasurface units. Fabricating the metasurface units on the surface of CIS is all we need to build such a hyperspectral sensor. It is cheaper and faster than traditional hyperspectral cameras. It can be integrated into any existing face recognition devices or mobile phones very easily. And by combining our sensor and algorithm with an RGB camera, a depth camera, or other cameras, we can even build a stronger antispoofing face recognition system. It is very practical in

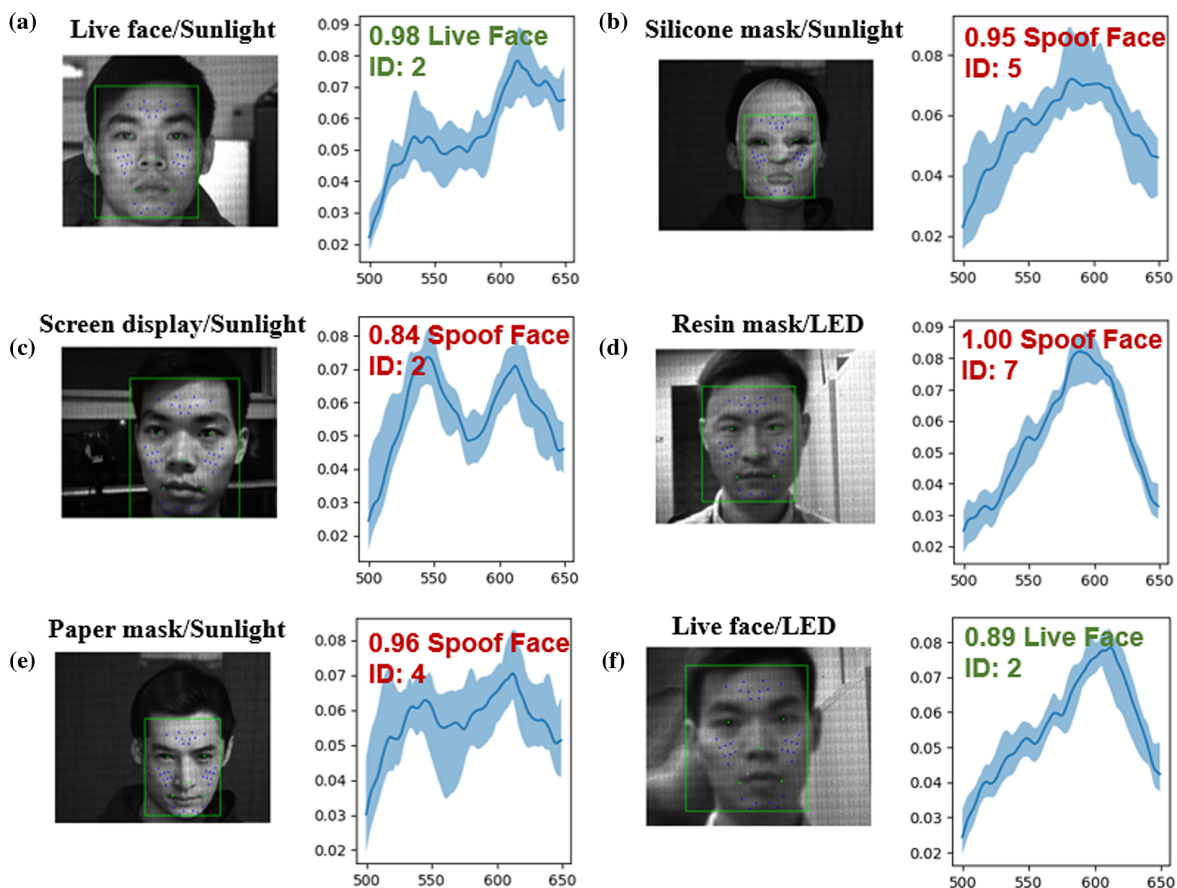


Fig. 7. Real-world antispoofing recognition results. (a) Results for a live face captured under sunlight; (b) results for a silicone mask captured under sunlight; (c) results for a face displayed on the screen; (d) results for a high-quality 3D resin mask captured under LED light; (e) results for a paper mask captured under sunlight; (f) results for a live face captured under LED light.

real-world deployment. Furthermore, using the same strategy, an anti-spoofing device can also be designed for other biometric-based identity authentication systems, such as fingerprint recognition and palm-print recognition.

The system we implemented takes full advantage of the snapshot miniature HSI technology, which brings real-time hyperspectral sensing into small devices and mobile devices. Furthermore, using CMOS fabrication, on-chip integrated metasurface units, and CIS, we can achieve mass production. Noniterative algorithms such as deep neural networks can also be adopted to boost the spectrum reconstruction procedure. Our solution is expected to be widely used in real-world applications.

4. MATERIALS AND METHODS

A. Device Fabrication

The nanostructure of the designed metasurface is fabricated on a 220 nm thick silicon-on-insulator (SOI) using electron beam lithography (EBL) and inductively coupled plasma (ICP). Then buffered hydrofluoric is used to corrode the silicon-dioxide layer under the SOI. Finally, the metasurface is transferred and attached to the surface of the CIS using polydimethylsiloxane (PMDS). We used Thorlabs CS235MU monochrome camera as our CIS.

B. Face Detection and Recognition

At least two steps are needed for a regular face recognition system: face detection and face recognition. Using deep-learning algorithms, there are several efficient and reliable approaches to do face detection [33,34] and face recognition [35,36]. As face alignment is necessary for the following FAS procedure, we adopted multi-task convolutional neural network (MTCNN) [37] to do face detection and face alignment. MTCNN is a lightweight convolutional neural network (CNN) with three stages. It can do joint face detection and alignment at high computational efficiency. As for face recognition, we trained ResNet [38] with ArcFace [39] loss to embed face images. The face detection and recognition networks are first trained on WiderFace [40] and MS1M [41] data sets and then fine-tuned on the images captured by our sensor. Finally, the trained algorithms are performed on the raw gray-scale images are produced by our image sensor.

Funding. China State Key Development Program in 14th Five-Year (2021YFF0602103, 2021YFF0602102).

Acknowledgment. The authors thank Tianjin H-Chip Technology Group Corporation and Innovation Center of Advanced Optoelectronic Chip and Institute for Electronics and Information Technology in Tianjin, Tsinghua University for their support during SEM and ICP etching. We thank Prof. Shengjin Wang for suggestions on writing the manuscript, Dr. Sheng Xu for helping with acquiring the image data, and Yue Zou for helping with English language editing.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

- J. A. Martins, R. L. Lam, J. Rodrigues, and J. H. du Buf, "Expression-invariant face recognition using a biological disparity energy model," *Neurocomputing* **297**, 82–93 (2018).
- N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE Trans. Inf. Forensics Secur.* **9**, 1084–1097 (2014).
- Y. Cai, Y. Lei, M. Yang, Z. You, and S. Shan, "A fast and robust 3D face recognition approach based on deeply learned face representation," *Neurocomputing* **363**, 375–397 (2019).
- S. Z. Gilani and A. Mian, "Learning from millions of 3D scans for large-scale 3D face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1896–1905.
- G. Pan, Z. Wu, and L. Sun, *Recent Advances in Face Recognition, Liveness Detection for Face Recognition* (IntechOpen, 2008), pp. 109–124.
- H. Jee, S. Jung, and J. Yoo, "Liveness detection for embedded face recognition system," *Int. J. Comput. Inf. Eng.* **2**, 2142–2145 (2008).
- Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "Face anti-spoofing with human material perception," in *European Conference on Computer Vision—ECCV* (Springer, 2020), pp. 557–575.
- S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3D mask face anti-spoofing database with real world variations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2016), pp. 100–106.
- S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks," in *IEEE 9th International Conference on Biometrics Theory, Applications and Systems* (IEEE, 2018), pp. 1–7.
- S. Jia, G. Guo, and Z. Xu, "A survey on 3D mask presentation attack detection and countermeasures," *Pattern Recogn.* **98**, 107032 (2020).
- A. Sabaghi, M. Oghbaie, K. Hashemifard, and M. Akbari, "Deep learning meets liveness detection: recent advancements and challenges," arXiv:2112.14796 (2021).
- 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 Behavior Identity Sci.* **2**, 399–409 (2020).
- J. Seo and I. Chung, "Face liveness detection using thermal face-CNN with external knowledge," *Symmetry* **11**, 360 (2019).
- M. Liu, H. Fu, Y. Wei, Y. A. U. Rehman, L. Po, and W. L. Lo, "Light field-based face liveness detection with convolutional neural networks," *J. Electron. Imaging* **28**, 013003 (2019).
- Y. Tian, K. Zhang, L. Wang, and Z. Sun, "Face anti-spoofing by learning polarization cues in a real-world scenario," in *4th International Conference on Advances in Image Processing* (2020), pp. 129–137.
- Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: a survey," arXiv:2106.14948 (2021).
- A. Currà, R. Gasbarrone, A. Cardillo, C. Trompetto, F. Fattapposta, F. Pierelli, P. Missori, G. Bonifazi, and S. Serranti, "Near-infrared spectroscopy as a tool for in vivo analysis of human muscles," *Sci. Rep.* **9**, 8623 (2019).
- M. Ardabilian, A. Zine, and S. Li, "Multi-, hyper-spectral biometrics modalities," in *Hidden Biometrics* (Springer, 2020), pp. 127–153.
- Y. Liu, M. Zheng, and Q. Li, "Face liveness verification based on hyperspectrum analysis," in *31st International Conference on Advanced Information Networking and Applications Workshops* (IEEE, 2017), pp. 612–615.
- M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatio-spectral information fusion and PLS regression," *IEEE Trans. Image Process.* **24**, 1127–1137 (2015).
- Z. Xie, J. Niu, L. Yi, and G. Lu, "Regularization and attention feature distillation base on light CNN for hyperspectral face recognition," *Multimedia Tools Appl.* **81**, 19151–19167 (2021).
- S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalera, H. J. Escalante, and S. Z. Li, "Casia-surf: a large-scale multi-modal benchmark for face anti-spoofing," *IEEE Trans. Biometrics Behavior Identity Sci.* **2**, 182–193 (2020).
- J. Xiong, X. Cai, K. Cui, Y. Huang, J. Yang, H. Zhu, Z. Zheng, S. Xu, Y. He, and F. Liu, "One-shot ultraspectral imaging with reconfigurable metasurfaces," arXiv:2005.02689 (2020).
- Z. Wang, S. Yi, A. Chen, M. Zhou, T. S. Luk, A. James, J. Nogan, W. Ross, G. Joe, and A. Shahsafi, "Single-shot on-chip spectral sensors based on photonic crystal slabs," *Nat. Commun.* **10**, 1020 (2019).
- F. Yesilkoy, E. R. Arvelo, Y. Jahani, M. Liu, A. Titti, V. Cevher, Y. Kivshar, and H. Altug, "Ultrasensitive hyperspectral imaging and biodetection enabled by dielectric metasurfaces," *Nat. Photonics* **13**, 390–396 (2019).

26. J. Xiong, X. Cai, K. Cui, Y. Huang, J. Yang, H. Zhu, W. Li, B. Hong, S. Rao, and Z. Zheng, "Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces," *Optica* **9**, 461–468 (2022).
27. U. Utzinger, "Spectra database hosted at the University of Arizona," <http://spectra.arizona.edu/>.
28. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).
29. S. Zhang, Y. Dong, H. Fu, S.-L. Huang, and L. Zhang, "A spectral reconstruction algorithm of miniature spectrometer based on sparse optimization and dictionary learning," *Sensors* **18**, 644 (2018).
30. J. Yang, K. Cui, X. Cai, J. Xiong, H. Zhu, S. Rao, S. Xu, Y. Huang, F. Liu, and X. Feng, "Ultraspectral imaging based on metasurfaces with freeform shaped meta-atoms," *Laser Photon. Rev.* **16**, 2100663 (2022).
31. Dualix, "GaiaField-pro hyperspectral imaging camera," <https://www.dualix.com.cn/en/Goods/desc/id/123/aid/999.html>.
32. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (2017), pp. 5999–6009.
33. P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-stream neural networks for tampered face detection," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, 2017), pp. 1831–1839.
34. A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," *Artif. Intell. Rev.* **52**, 927–948 (2019).
35. I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: a survey," in *31st SIBGRAPI Conference on Graphics, Patterns and Images* (IEEE, 2018), pp. 471–478.
36. M. Lal, K. Kumar, R. H. Arain, A. Maitlo, S. A. Ruk, and H. Shaikh, "Study of face recognition techniques: a survey," *Int. J. Adv. Comput. Sci. Appl.* **9**, 42–49 (2018).
37. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.* **23**, 1499–1503 (2016).
38. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
39. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 4690–4699.
40. S. Yang, P. Luo, C. Loy, and X. Tang, "Wider face: a face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 5525–5533.
41. Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-Celeb-1M: a dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision* (Springer, 2016), pp. 87–102.