2023 International Conference on Artificial Intelligence of Things and Systems (AIoTSys) | 979-8-3503-1227-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIoTSys58602.2023.00042

Smart Minion: A Low-cost Serverless Multimodal Access Control System Based on Face Recognition and Gesture Recognition

Zijie Chen, Yiming Gao, and Junrui Liang
School of Information Science and Technology,
ShanghaiTech University, Shanghai 201210, China
Email: {chenzj1, gaoym, liangjr}@shanghaitech.edu.cn

***Abstract*—With the advance of the Internet of Things (IoT) and artificial intelligence (AI) technology, more and more applications appear for mobile computing, such as face recognition for access control. Yet, most artificial intelligence of things (AIoT) products in our homes are based on cloud service application programming interfaces (APIs). All the original data are sent wirelessly to the server for cloud computing, which may lead to privacy issues and high energy consumption. In this study, we propose a novel low-cost serverless access control system with multimodal inferring capability. Within the system, practical on-device face recognition and cost-efficient dynamic gesture recognition are implemented for liveness detection. We utilize the MobileNetV2 model for transfer learning the task-specific face data and a random forest model for one-dimensional gesture recognition. Both tiny machine learning (TinyML) models are successfully deployed on a low-cost microcontroller (MCU). The multimodal inferring perform pretty well. After using the sequential inferring mechanism, the robustness of the system is further reinforced. A low-cost prototype is assembled for field tests. The compact all-in-one product based on the proposed system with a cute minion theme has also been designed. The energy consumption of different MCU operations is measured in detail. The system provides a valuable reference for realizing pervasive smart sensing and, thus, contributes to the bright future of AIoT.**

***Index Terms*—Face recognition, gesture recognition, access control, transfer learning, tiny machine learning, AIoT.**

I. INTRODUCTION

Nowadays, the integration of machine learning (ML) and the Internet of Things (IoT), called artificial intelligence of things (AIoT), has kept attracting increasing interest from both academia and industry, due to the wide deployment and the relatively low energy consumption of IoT devices [1], [2]. The end-edge-cloud orchestrated system enables the endpoint IoT device to fulfill artificial intelligence (AI)-based tasks that were hardly imagined before, such as autonomous driving and smart home [3]. To achieve identity verification and access control, face recognition is one of the most widely used perceiving techniques in several AIoT systems [4]. With advanced deep learning (DL) models, excellent performance for face recognition can be realized [5] for their powerful end-to-end learning capacity.

Yet, many AIoT devices rely on resource-rich servers due to the limited computing and storage ability of IoT devices. All raw data sensed by the end devices are transmitted to the server for analysis, then the result of the machine learning model is sent back to the device for continuous application workflow [6]. Sending end data from two separate locations in this centralized architecture causes large energy consumption and undesired data privacy issues, especially for the battery-powered sensors in highly private places. Recently, researchers have been paying attention to providing AI functionalities for the edge devices [7], [8], such as smartphones, FPGA(s), Raspberry Pi(s), and personal computers, whose relatively high cost and energy consumption, compared with microcontrollers (MCUs) [9], remain a barrier to the genuine ubiquitous AIoT.

TinyML, a booming branch of state-of-the-art ML techniques, enables the low-cost MCU to run on-device ML models at a milliwatt-level power consumption without real-time support of large servers [10]. TinyML-based AIoT applications have been increasingly proposed, such as gesture recognition [11] and face recognition [12]. The essential techniques of TinyML are to simplify and compress the ML model and deploy it on the end device where data originates [13]. It significantly reduces potential data security problems. Moreover, end-side inferring minimizes the data transmitting cost, which is highly beneficial for systems where the wireless transmitter dominates the power consumption.

For real-world practical face recognition, protective mechanisms are required, such as liveness detection and spoof detection [4], enhancing the anti-deception ability of the face recognition system. The compressed model embedded in the MCU can barely support the security functions mentioned above. This may be why there is a vacancy for commercial products with on-device access control based on face recognition. Methods for practical face recognition have been proposed based on action imitation and multimodal sensors [14]. Yet, while using these typical techniques, extra relatively high cost and power consumption are demanded, which restrains the applications based on low-cost battery-powered devices.

To achieve robust on-device face recognition in resource-constrained devices, in this study, we propose a low-cost multimodal system where gesture recognition based on only one infrared sensor is utilized to complement face recognition toward practical real-time access control. Two TinyML models are simultaneously deployed in the MCU for multimodal inferring. It is hoped the system can promote the development of AIoT applications and provide inspiration for future pervasive

979-8-3503-1227-0/23/$31.00 ©2023 IEEE
DOI 10.1109/AIoTSys58602.2023.00042

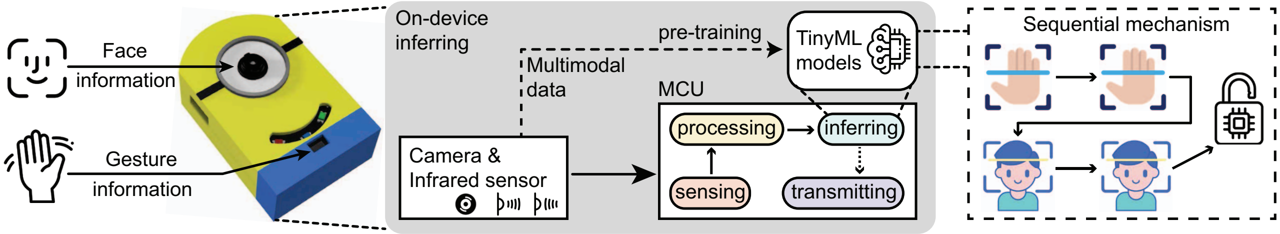


Fig. 1. Architecture of the proposed offline multimodal access control system based on face and gesture recognition with an emphasis on the low-cost characteristic.

sensing and ubiquitous AI studies.

II. Related Work

TinyML is a burgeoning field at the intersection of embedded systems and machine learning [15]. It brings more intelligence to the ultra-low power and always-on endpoint devices like MCU to infer real-world data with ML algorithms. On-device inferring relieves the omnipresent dependency on remote servers, benefiting in low-power, low-cost, low-latency, and privacy-preserving characteristics. It revolutionizes real-time data processing and unleashes many AIoT applications [16]. In [17], an energy-aware TinyML system is proposed for real-time face recognition on battery-free IoT devices powered by solar energy. Additionally, the authors in [18] introduce a static hand gesture recognition system based on an 8×8 infrared array sensor with endpoint intelligence.

However, for robust access control, the face recognition process not only includes face detection, image pre-processing, feature extraction, and feature matching but also comprises anti-deception techniques such as liveness detection [19]. Today's common methods for liveness detection on commercial products are based on action imitation or multimodal sensors. They record and upload the video, in which the user cooperates to complete a series of actions, to the server or utilize multimodal sensors, such as the binocular camera, for liveness analysis, which may somehow increase the energy consumption and privacy issues or the cost of the hardware setting.

As far as the authors know, a robust serverless access control system based on multimodal face and gesture recognition, with an emphasis on ultra-low cost characteristics, remains unexplored.

III. System Overview

Fig. 1 illustrates the architecture of the proposed access control system. The end device is shaped like a cute compact minion, embedded with a low-power MCU with IoT capability, a monocular camera, and an infrared distance sensor, which is installed on a door or any other places that require controlling and restricting access to resources, areas or information in real-time. The camera and infrared sensor are utilized to capture the human face and gesture information, respectively.

For TinyML development and deployment, the multimodal data is first collected and preprocessed for pre-training ML models. Note that the preprocessing part should be lightweight enough since the MCU would operate the same process during real-world applications. In addition, the lower the sampling rate, the lower the energy consumption for the data acquisition in a fixed time interval. After training, the model is compressed with skills like parameter quantization and pruning [13]. For fast deploying, it is popular to choose an inference frame designed for running mobile machine learning, such as the TensorFlow Lite [20]. Subsequently, two compressed models for gesture and face recognition are converted into documents suitable for embedded systems.

The MCU executes intelligent evaluation when AI models are deployed. Face recognition is complemented by gesture recognition for practical on-device identity authentication. The immediate real-world data is sensed and preprocessed. After on-device inferring, human identification is recognized. Toward more robust and lower power, we adopt a sequential mechanism for data inferring, i.e., only after two continuous correct gestures are recognized, the face recognition is opened, and only after two sequential face recognition is passed would the system open the access. For occasions with relatively low security, only gestures or faces would be sampled, lowering the latency of the authentication process. The MCU can be designed to send an alarm to the server when a suspicious condition is detected. Other than this alarm, all MCU operations can work offline.

IV. On-device Face Recognition

Compared with password, card, and fingerprint unlocking, face recognition proves a touchless method with far more convenience, which has become increasingly prevalent for intelligent access control systems. However, high-cost device requirements and potential privacy leakage issues in network communications may restrain the smart homes development, which makes a low-cost offline face recognition solution worthy of research.

A. Data Acquisition

To validate the proposed system, we deployed a serverless face recognition demo in our lab. An MCU (ESP32, Espressif Inc.) and an off-the-shelf camera model (OV2640, OmniVision Inc.) are utilized for data acquisition and real-world face recognition. To simulate the access control scenario, the end

Fig. 2. Data illustration. A small part of the dataset of human face grayscale images from different capturing perspectives with 96×96 resolution.

	S1	S2	S3	S4	S5	S6	S7	Nobody
S1	100%	0%	0%	0%	0%	0%	0%	0%
S2	0%	100%	0%	0%	0%	0%	0%	0%
S3	0%	0%	100%	0%	0%	0%	0%	0%
S4	0%	0%	0%	100%	0%	0%	0%	0%
S5	0%	0%	3.2%	0%	96.8%	0%	0%	0%
S6	0%	0%	0%	0%	0%	100%	0%	0%
S7	0%	0%	0%	0%	0%	0%	100%	0%
Nobody	0%	0%	0%	0%	0%	0%	0%	100%
F1 Score	1.00	1.00	0.97	1.00	0.98	1.00	1.00	1.00

Fig. 3. Confusion matrix of the TinyML model based on MobileNetV2 for eight-class face recognition on seven subjects.

device was kept relatively constant in position during sampling. We chose an office with more color changes as the background for testing the system's robustness in different environments. The image is captured and transmitted to the PC through WiFi for data preprocessing and model training. The image preprocessing part simply includes grayscale and normalization to decrease the on-device data processing delay. Labeled human face images from seven different subjects were collected, where the data of each subject includes diversity in time (day, night), face angle, face distance from the camera, etc. Therefore, a dataset of more than 2500 grayscale images with 96×96 resolution is acquired, a small portion of which is shown in Fig. 2.

B. Transfer Learning

Instead of training a robust convolutional neural network (CNN) for face recognition from scratch, we plan to adopt MobileNet [8] for transfer learning. An ML technique where a pre-trained model with a large-scale dataset is utilized as a starting point for solving a new related task. It accelerates the learning process and improves performance with limited task-specific data [21]. MobileNet, a well-known efficient model for mobile and embedded vision applications, has already captured general features and patterns that are applicable to various visual recognition tasks. Depthwise separable convolution, a combination of depthwise and pointwise convolution, is the key technique for its lightweight.

We chose MobileNetV2 [22], which is more powerful and lightweight than MobileNetV1. The hyperparameter, width multiplier, is set to be 0.05 for extremely reducing the number of model parameters and computational complexity. Except for the input layer, final layer, and output layer, other layers of the model are frozen. The neuron number and dropout rate of the final layer are 0.1 and 8. The neuron numbers of the input and output layers are 9216 and 8. With the help of Edge Impulse [23], and TensorFlow Lite [20], we trained and compressed the model, whose confusion matrix is shown in Fig. 3. The size of the compressed model is only 158 KB.

In the field test, each object can be recognized locally with a confidence rate higher than 90% in less than 1 second. When a stranger appears, the confidence rate will be lower than 70%. The end device can also accurately discern the unknown person with a simple threshold setting. Together with the sequential judgment mechanism, the system performs more robustly. However, the system sometimes misjudges when using deception, such as pictures, for stress tests. So, we continue to empower the system with low-cost dynamic gesture recognition to improve security.

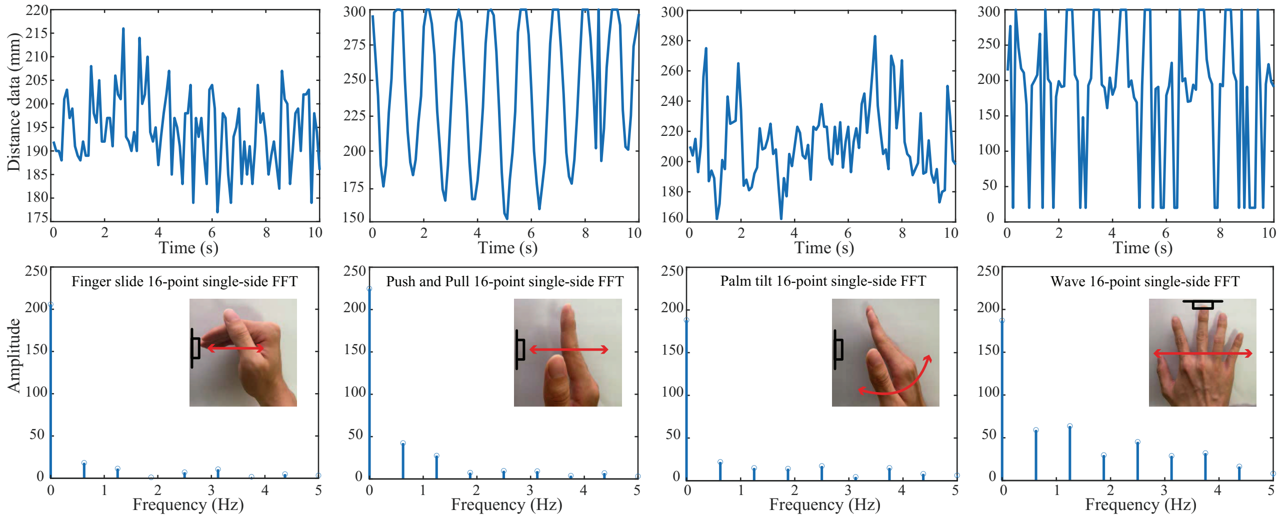
V. ON-DEVICE GESTURE RECOGNITION

Gesture recognition is one of the hottest topics for information verification in recent years, benefiting more convenient human-computer interaction applications. Touchless gesture recognition solutions are mainly based on radio frequency, camera, or infrared rays. To realize ultra-low cost demand, the proposed system adopts an infrared distance sensor with a built-in transmitter and receiver unit for one-dimensional dynamic gesture recognition.

A. Data Acquisition

We utilized an MCU (ESP32, Espressif Inc.) and an infrared model with a sensing chip (VL53L0, STM Inc.) based on time-of-flight for data acquisition and real-world gesture recognition. Data is transmitted to the PC through the serial port. In addition, the end device was kept relatively constant in position during sampling to simulate the access control scenario. As we know, a low sampling rate is usually adopted for battery-powered sensors to save energy during data sampling. Thus, in this study, we use the data sampled at an ultra-low rate of 10 Hz for dynamic gesture classification.

A dataset of over 2000 seconds of four different gestures and background noise signals is collected. The gesture patterns include finger slide, push and pull, palm tilt, and wave, whose



(a) (b) (c) (d)

Fig. 4. Sampled data illustration. (a)-(d) 10-second segment data sampled at 10 Hz of four respective gestures, finger slide, push and pull, palm tilt, wave, and their corresponding unipolar frequency-domain values of 16 consecutive points randomly chosen.

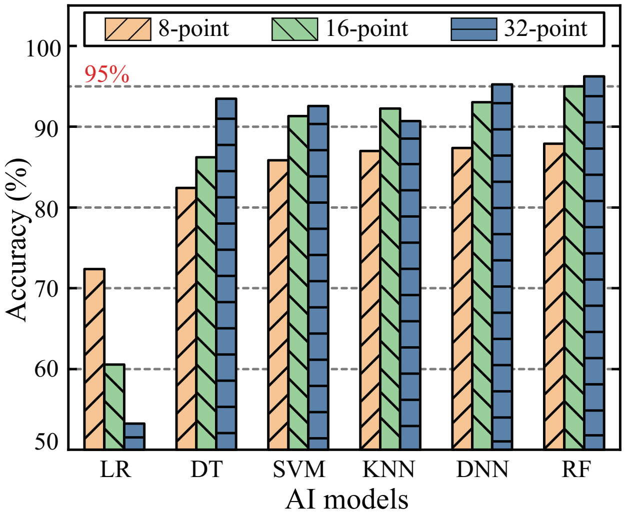


Fig. 5. Performance of six AI models under different data lengths for 5-class gesture recognition.

data are collected at different times and distances, respectively. To eliminate the background noise and outlier data, we set the maximum distance value to 300 mm, which matches the scenario of being near the sensor for identity verification.

Before the model training process, the data preprocessing part is required. Fast Fourier transform (FFT) is a popular feature extraction method that converts the signal from its original domain to a representation in the frequency domain. We perform a single-side FFT to the original data. Then, we add the average value of the spectral amplitude in each set of data. After normalization, the data is ready for model training. Fig. 4 illustrated the 10-second segment data sampled at 10 Hz of four respective gestures and their corresponding unipolar frequency-domain values of 16 consecutive points randomly chosen. Note that the gesture data are not periodic, and the average amplitude varies depending on the time or other factors. Since the dynamic gestures are generally not high-frequency, the amplitude of the low-frequency region is relatively high.

B. Machine Learning Analysis

We apply six well-known supervised machine learning classification algorithms, including Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Deep Neural Network (DNN), on the 5-class preprocessed gesture data for testing their performance in classification.

For comparability, traditional ML models are based on the default settings of the sklearn library, and the results are based on 5-fold cross validation. LR's regularization parameter and penalty are 1.0 and L2; DT's max depth, minimal samples split, and criterion are none, 2, and gini; SVM's regularization parameter and kernel function are 1.0 and rbf; KNN's neighbor number and distance metric are 5 and minkowski; RF's tree number, max depth, minimal samples split, and criterion are 13, none, 2, and gini, whose model size is 247 KB. DNN is set to 3 hidden layers, and the neuron distribution is (32, 16, 16); the activation function is ReLU; the output function is Softmax; the optimizer is adam; the loss function is cross entropy function; the batch size is 16; the learning rate is 0.1.

The number of sampled time-domain data is required to be an exponent of 2 for FFT transform; thus, we test the system performance under data lengths of 8, 16, and 32 points. Fig. 5 shows the performance of different models under different data lengths for 5-class gesture recognition. Except for the LR algorithm, all models reach a high accuracy even when the

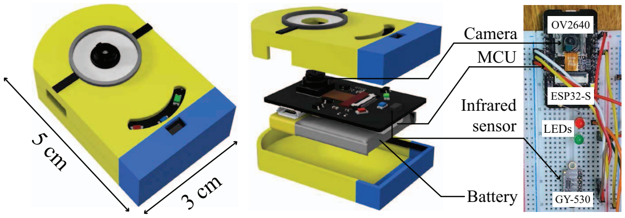


Fig. 6. Prototyped all-in-one compact prototype with a cute minion theme.

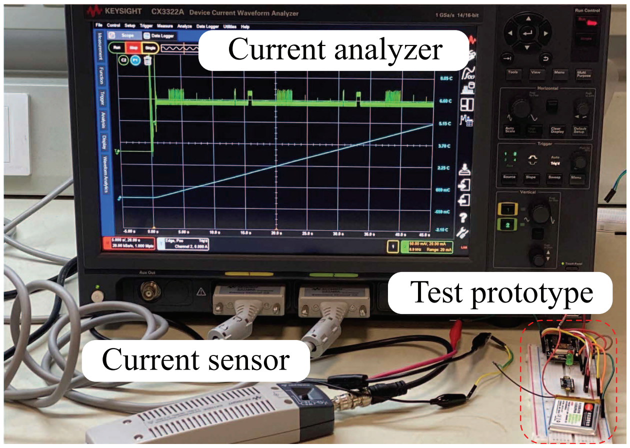


Fig. 7. Experimental setup for energy consumption measurement.

TABLE I
MEASURED ENERGY CONSUMPTION OF SPECIFIC OPERATIONS.

Atomic operation	**Processed data**	**Duration**	**Energy**
Initialization	-	800 ms	264.1 mJ
I^{2}C sampling & normalization	float16×1	50 ms	21.1 mJ
FFT & averaging	float16×16	0.8 ms	495.0 μJ
Inferring (RF)	float16×10	0.8 ms	462.0 μJ
Image capturing	int8×320×240×3	0.4 ms	165.0 μJ
Data preprocessing	int8×320×240×3	120 ms	52.8 mJ
Inferring (MobileNetV2)	float16×96×96×1	640 ms	287.1 mJ

data length is only 8. Therefore, AI methods can satisfy high performance in identifying poorly sampled one-dimensional gesture signals. 16-point seems the most appropriate sampling mode for balancing the sampling time and accuracy tradeoff. Since the sampling rate is 10 Hz, the corresponding window size is 1.6 seconds. In addition, RF achieves the highest accuracy, up to 94.9% in the 16-point case, so we choose RF for TinyML deployment and energy measurement.

VI. EXPERIMENTAL EVALUATION

To comprehensively validate the proposed system, we successfully deployed the two TinyML models in an MCU and empowered it with multimodal inferring capability. A prototype is simply assembled and used for the field test. Subsequently, measurements of energy consumption are also carried out in detail.

A. Prototype

The system can be based on most 32-bit Cortex-M or Xtensa LX6 MCUs. We have assembled all the required hardware on a breadboard as an original prototype for testing, which comprises an MCU (ESP32-S, Espressif Inc.), a monocular camera (OV2640, OmniVision Inc.), an infrared distance sensor (GY-530) with a sensing chip (VL53L0, STM Inc.), and some LEDs for status indication. The entire cost for the prototype is less than US$ 7, which would be much lower after mass production. In addition, we have designed an all-in-one compact prototype with a cute minion theme, whose eye and logo on the pant are the camera and infrared sensor, respectively, as shown in Fig. 6.

B. Energy Measurement

The whole process for the proposed offline access control system can be divided into four types of states: sleeping, sampling, data preprocessing, and inferring. During the normal working process, the real-world dynamic human gesture information is first sensed in 1.6 seconds under a 10 Hz sampling rate. To further reduce energy consumption, the MCU would enter light sleep mode during every interval between samplings. Then, the original data is processed and inferred on the device for gesture recognition. After recognizing the target gesture twice in a row, the system enters face recognition mode. The camera first captures a color image in JPEG format with 320×240 resolution. Subsequently, the image is converted to RGB888 format and scaled to 96×96 resolution. Before inferring, the image is converted to grayscale to decrease the delay. After recognizing the target face twice in a row, the system finishes one-round access control.

Fig. 7 shows the experimental setup for measuring the energy consumption of the system operations introduced above. The infrared sensor is sampled using the I^{2}C bus embedded in the MCU, which is connected with a supporting current sensor in series and powered by a 3.7 V lithium battery. Then, the current is measured by a CX3322A current analyzer (Keysight Inc.) connected across the current sensor.

The energy consumption statistics of a round of on-device gesture recognition and face recognition are listed in Table I. The duration and required energy of TinyML inferring for face recognition are much higher than gesture recognition due to the difference in computational complexity. The average power for the whole process is at the mW level. The duration is less than 5 seconds when using the sequential inferring mechanism. The system can be more energy-efficient by using the ultra-low-power system on chip (SoC) or energy-aware programming technology [24].

Most of the time, the MCU would enter sleep mode to save energy, where the current is at μA level. A button awakens it and works intermittently. For the anomaly warning function, the system can actively transmit a message to the homeowner through WiFi, GSM, or NB-IoT protocols if the face recognition or the gesture recognition cannot be passed for a long time, simply by adding an additional communication peripheral. Or we can simply set the maximum time for gesture and face recognition to 5 seconds, respectively, beyond which the system will delay for several minutes as a penalty.

C. Discussion

In the stress test, we tried to disturb the system with different unknown gestures and human faces. The result proves the robustness of the system with the sequential mechanism. In real-life applications, the system can be built quickly by conveniently collecting the user's face data with the accompanying App and sending it to the cloud server for model training and compressing, and finally using Over-The-Air (OTA) for end-side deployment.

Only one infrared sensor with an ultra-low sampling rate brings a relatively long sensing period for system stability. To enhance the interactive experience and performance, the sensor number and sampling rate can be increased in less energy-constrained applications. The proposed system focuses on customized privacy-preserving scenarios with a relatively small dataset, such as smart home and smart driving, whose robustness has been validated. Due to the space limitation, the variations in the dataset, model size, feature extraction method, and system performance brought about by recognizing much more complex and diverse face and gesture information will be analyzed in detail in our future work.

VII. CONCLUSION

This paper introduced a multimodal access control system based on face recognition and gesture recognition with an emphasis on low-cost demand. With advanced TinyML techniques, the system worked totally offline, maximally protecting the users' privacy. A transfer learning technique based on the well-known MobileNetV2 model was leveraged in this study to reach high-accuracy human face recognition. The performance of six well-known ML models under one-dimensional gesture data sampled at a 10 Hz rate were compared in detail. RF with 13 trees achieved accuracy up to 95% for 5-class gesture recognition given only 16-point original data. After successfully deploying two TinyML models in one MCU, the system is validated in field tests with an assembled low-cost prototype ($<$ \$7). The results proved the system's robustness with a designed sequential inferring mechanism. Comprehensive energy measurements were carried out to validate the feasibility of the battery-powered solution. Such a design provided valuable inspiration for future ubiquitous AIoT applications and studies.

REFERENCES

[1] H. Ren, D. Anicic, and T. A. Runkler, "Tinyol: Tinyml with online-learning on microcontrollers," in *2021 international joint conference on neural networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[2] C.-J. Chen, Y.-Y. Huang, Y.-S. Li, C.-Y. Chang, and Y.-M. Huang, "An aiot based smart agricultural system for pests detection," *IEEE Access*, vol. 8, pp. 180 750–180 761, 2020.

[3] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of internet of things for smart home: Challenges and solutions," *Journal of cleaner production*, vol. 140, pp. 1454–1464, 2017.

[4] J. Zhang and D. Tao, "Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2020.

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

[6] M. Monfort Grau, "Tinyml: From basic to advanced applications," B.S. thesis, Universitat Politècnica de Catalunya, 2021.

[7] S. Duan, D. Wang, J. Ren, F. Lyu, Y. Zhang, H. Wu, and X. Shen, "Distributed artificial intelligence empowered by end-edge-cloud computing: A survey," *IEEE Communications Surveys & Tutorials*, 2022.

[8] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9] J. Lin, W.-M. Chen, Y. Lin, C. Gan, S. Han *et al.*, "Mcunet: Tiny deep learning on iot devices," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 711–11 722, 2020.

[10] R. Sanchez-Iborra and A. F. Skarmeta, "Tinyml-enabled frugal smart objects: Challenges and opportunities," *IEEE Circuits and Systems Magazine*, vol. 20, no. 3, pp. 4–18, 2020.

[11] S. Bian and P. Lukowicz, "Capacitive sensing based on-board hand gesture recognition with tinyml," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 4–5.

[12] M. Giordano and M. Magno, "A battery-free long-range wireless smart camera for face recognition," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 594–595.

[13] Y. Zhang, W. Ding, and C. Liu, "Summary of convolutional neural network compression technology," in *2019 IEEE International Conference on Unmanned Systems (ICUS)*, 2019, pp. 480–483.

[14] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–37, 2017.

[15] V. J. Reddi, B. Plancher, S. Kennedy, L. Moroney, P. Warden, A. Agarwal, C. Banbury, M. Banzi, M. Bennett, B. Brown *et al.*, "Widening access to applied machine learning with tinyml," *arXiv preprint arXiv:2106.04008*, 2021.

[16] C. R. Banbury, V. J. Reddi, M. Lam, W. Fu, A. Fazel, J. Holleman, X. Huang, R. Hurtado, D. Kanter, A. Lokhmotov *et al.*, "Benchmarking tinyml systems: Challenges and direction," *arXiv preprint arXiv:2003.04821*, 2020.

[17] A. Sabovic, M. Aernouts, D. Subotic, J. Fontaine, E. De Poorter, and J. Famaey, "Towards energy-aware tinyml on battery-less iot devices," *Internet of Things*, vol. 22, p. 100736, 2023.

[18] W. Dai and L. Zhou, "Tinyml-enabled static hand gesture recognition system based on an ultra-low resolution infrared array sensor and a low-cost ai chip," in *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*. IEEE, 2022, pp. 804–809.

[19] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE access*, vol. 8, pp. 139 110–139 120, 2020.

[20] R. David, J. Duke, A. Jain, V. Janapa Reddi, N. Jeffries, J. Li, N. Kreeger, I. Nappier, M. Natraj, T. Wang *et al.*, "Tensorflow lite micro: Embedded machine learning for tinyml systems," *Proceedings of Machine Learning and Systems*, vol. 3, pp. 800–811, 2021.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[23] S. Hymel, C. Banbury, D. Situnayake, A. Elium, C. Ward, M. Kelcey, M. Baaijens, M. Majchrzycki, J. Plunkett, D. Tischler, A. Grande, L. Moreau, D. Maslov, A. Beavis, J. Jongboom, and V. J. Reddi, "Edge impulse: An mlops platform for tiny machine learning," 2023.

[24] X. Li, H. Tang, G. Hu, B. Zhao, and J. Liang, "ViPSN-pluck: A transient-motion-powered motion detector," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3372–3382, 2021.