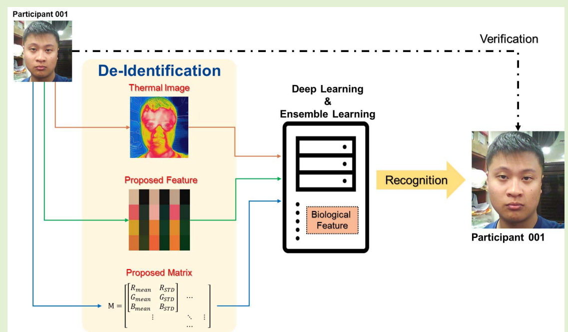
A De-Identification Face Recognition Using Extracted Thermal Features Based on Deep Learning

Chih-Hsueh Lin, *Member, IEEE*, Zhi-Hao Wang, and Gwo-Jia Jong, *Member, IEEE*

***Abstract*—Facial recognition based on visible-light RGB image is a mature technique. With the continuous progress of civilization, the protection against personal privacy and the effectiveness of the system is more stringent. In the past research, the distance from facial features was used as the de-identified biological feature, but this could not solve the problem of using a photo to deceive the system. Therefore, we used thermal image to convert into features by using the proposed feature extraction method, and uses deep learning, random forest, and ensemble learning to build a face prediction model. The proposed feature extraction method cuts the facial image (RGB and thermal image) into 12 and 48 blocks respectively as well as regenerates the feature image and the feature matrix. Based on the experimental results, accuracy of RGB image only with 0.834, feature image with 0.953, and feature matrix with 0.967. The feature images and feature matrices produced by the proposed feature extraction method can achieve better prediction performance. High accuracy of thermal imaging can solve the problem of fake photos and the issue of personal privacy de-identification could be solved.**



***Index Terms*—Face recognition, de-identification, biological feature, thermal features, ensemble learning.**

I. Introduction

In recent years, when science and technology have evolved rapidly, more and more occasions have begun to pay attention to personnel identity issues, especially with national security or high-secret research sites. However, in many cases it still relies on manual identity verification. This usually brings some defects, especially during the inspection process, which is very susceptible to non-objective factors. This means that our manual verification process does not fully adhere to the principles of fairness, impartiality, and openness. Therefore, identification technology is very important and many automatic identification and verification systems have been developed to assist our identification process. Early identification mostly used RFID tags, but this method is easy to be stolen or copied, so biometric technology has received a lot of research in recent years. Most biometric identification systems use fingerprints, voice prints, palm prints, irises, etc. to perform verification, and use various algorithms to find biometrics for identity recognition. There are also a wide range of applications in life, such as using fingerprints to identify user and airport customs officers identify the immigrants through iris, which shows that the biometric identification method is both convenient and reliable. However, the performance of RGB image face recognition system is deeply affected by the ambient illumination, and it can be easily deceived in the form of images, reducing its security. While biometrics are being actively researched, a new issue is being discussed by many scientists or sociologists, personal privacy in research materials, in other words, de-identification of data that is from human.

In state-of-the-art research on face recognition, machine learning and deep learning have been implemented in a large number of applications for face recognition. In particular, convolutional neural networks (CNN), which have particularly outstanding results of image learning, are deployed in most systems. De Freitas Pereira *et al.* [1] proposed a deep convolution neural network method of visual spectral images to extract image features. The research results specifically pointed out that relatively high-level or advanced image features may not exist in the same domain as the original image. As the learning system evolved, Batchuluun *et al.* [2] further

Manuscript received March 12, 2020; revised April 2, 2020; accepted April 2, 2020. Date of publication April 6, 2020; date of current version July 17, 2020. The associate editor coordinating the review of this article and approving it for publication was Prof. Kazuaki Sawada. *(Corresponding author: Zhi-Hao Wang.)*

The authors are with the Department of Electronics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan (e-mail: cslin@nkust.edu.tw; zhwang0401@hotmail.com; gjjong@cc.kuas.edu.tw).

Digital Object Identifier 10.1109/JSEN.2020.2986098

proposed a compound neural network combining CNN and LSTM. Since CNN models are suitable for image feature extraction, LSTM models are suitable for extracting temporal features. Thus, when the two are combined to capture more spatial and temporal features. Although the use of a learning system based on biometric can quickly improve the efficiency and accuracy of face recognition, it still cannot solve the issue of biological security. Thus, George *et al.* [3] and Kotwal *et al.* [4] both proposed presentation attack detection (PAD) to prevent subjects from trying to deceive the system.

In addition to using advanced neural networks of anti-counterfeiting, infrared images and thermal images can also provide quite good solutions. Hermosilla Vigneau *et al.* [5] Analyzed the problems caused by infrared facial image changes of time when used in facial recognition system. The results show that most of the local binary pattern is not time-varying, but this is a significant change when facial images are acquired over time. Jian *et al.* [6] proposed a facial recognition method characterized by facial temperature, which is used to build infrared thermal facial images and perform facial recognition. Kaur and Khanna [7] not only used thermal images to implement face recognition, but also proposed a template protection method based on cancelable biometrics. The results show that not only can generate distinguishable and confidential revocable pseudo-biological identities, but also reduce identity size by 50%.

Some research has also pointed out that thermal images combined with features of specific locations on the face can effectively improve recognition performance. Past research has been able to accurately give the feature positions and relative proportions of faces [8], [9]. Gündüz Arslan *et al.* [10] Analyzed and compared the prevalence of three different face types of young Turkish people based on the facial index, and evaluated the longitudinal and size of each type. Sun and Zheng [11] developed an automatic heterogeneous face alignment method that does not rely on eye-center positioning. Hardin *et al.* [12] analyzed the images, researchers can gather information about metabolism and vascular activity to identify physiological abnormal changes. Jain *et al.* [13] Discovered the most sensitive areas on the human face, which can monitor changes in facial temperature and get 83% accuracy to predict a lie / truth response. Li *et al.* [14] proposed a novel non-invasive infrared thermal imaging framework that can estimate the thermal comfort level of the occupants by measuring the skin temperature collected from different facial areas using a low-cost thermal camera. The results show that ears, nose and cheeks are the most representative of a person's thermal comfort with an average accuracy of 85%.

In addition, it has been proved in some research that thermal images combined with RGB images can improve the recognition performance. Cao *et al.* [15] proposed a data-enhanced joint learning (DA-JL) method. This method interconverts cross-modal differences by incorporating synthetic images into the learning process. An end-to-end framework was proposed by Wang *et al.* [16], which consists of a generative network and a detector network to translate thermal facial images into visible. Hermosilla *et al.* [17] proposed a new face recognition system based on fusion of thermal and visible descriptors. Wang *et al.* [18] proposed a new method that used visible and thermal images as features of face recognition and classified them by support vector machines. Peng *et al.* [19] designed a couple representation similarity measure (CRSM) to measure the similarity between the obtained graphical representations. Extensive experiments have been performed in multiple HFR scenes, including viewing sketches, forensic sketches, near-infrared images, and thermal infrared images.

Despite all these efforts, limitation still exists. Basically, all of these researches were done on RGB image or thermal image based without further feature extraction. Therefore, their benefits cannot be validated. Additionally, infrared image cannot use for long-term recognition, which is noticeable when the face images have been acquired with a time lapse. In facial recognition, using only RGB images or thermal images cannot achieve excellent recognition performance. Based on the shortcomings of these research, we propose a face recognition system based on deep learning and using RGB images and thermal images as input datasets. In the image processing part, a novel feature extraction method is also proposed to extract biological features in thermal images to enhance the prediction results of the model.

The main contributions in this research are as follows:

- Proposed an image recognition method with de-identification function using feature images and feature matrix.
- Proposed a novel feature extraction method based on RGB images and thermal images, which are image and matrix.
- Establishing an ensemble learning-based face recognition, and RGB images, thermal images, feature images, and feature matrices is input dataset as well as output a prediction of participant.

The rest of this paper organization is briefed as follows: Section 2 presents the method related to deep learning and facial image feature extraction. Section 3 delivers the results related to the preprocessing results and prediction of the deep learning framework. Discussion is presented in Section 4. Finally, Section 5 concludes the research.

II. Methods

A. Proposed Architecture

The proposed architecture of this research as shown in Fig. 1. The aim of this research is developing a thermography-based face recognition method to improve prediction accuracy and de-identify the raw facial image. The core of this architecture is a deep learning framework that is based on the convolution neural networks, and support vector machine. The dataset used in this research is divided into three parts, namely training set, validation set and test set. According to the type of input data, the three prediction models are trained for different purposes, the first one is for the raw RGB image and thermal image, the second one is for the feature image which is generate from proposed feature extraction method, the third is the feature matrix. In this architecture, 9 models will be trained, Predication results of each model will list the

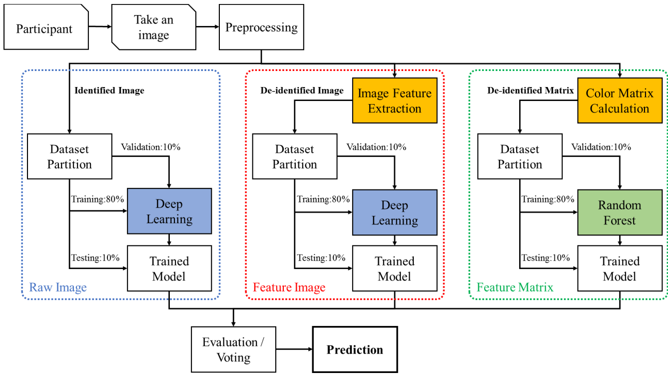


Fig. 1. Diagram of proposed architecture.

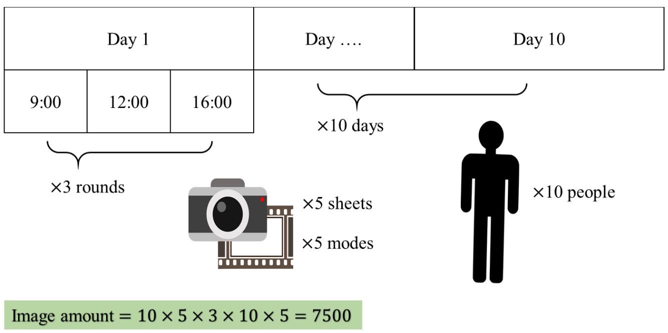


Fig. 2. Configuration of image data collecting.

probability of Top 5 subjects. Finally, the majority vote process is used to determine the final prediction result. A detailed description of each part of the proposed architecture will be described in the following sections.

B. Thermal Imaging and Experimental Configuration

The FLIR One is a mobile thermal camera that is used to take a thermal image on this research. There are two cameras on the FLIR One, the first with normal RGB imaging and the second with far infrared imaging. Since the temperature distribution characteristics of each materials are different, conventional RGB images cannot be used. Instead, thermal images and RGB images should be viewed simultaneously for better detection results.

In order to adapt to changing applications, FLIR One offers seven different color scales to highlight different temperature feature, which are seven modes including Contrast, Rainbow, Iron, Gray (White Hot), Lava, Arctic and Wheel. Under normal conditions, the temperature of the face does not change drastically, and the temperature is continuously distributed. Therefore, the selected thermal image display mode must be able to observe small thermal changes. The four modes, Rainbow, Iron, Gray, and Wheel, have such characteristics, which are suitable for scenes with small thermal changes and low contrast. Thus, in this research, the thermal images generated under these four modes are compared for the face recognition effect.

To make the collected images have the same conditions, Fig. 2 shows the experimental configuration of the image capture schedule. The image output contains RGB images

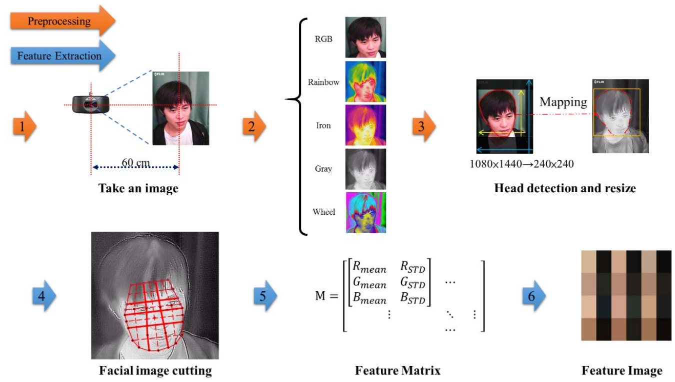


Fig. 3. The schematic diagram of model dataset acquisition and data preprocessing.

and thermal images of 4 color modes., and the face must be in the center of the image as well as distance between FLIR One and the participant is 60 cm. The chronological collection was performed on 10 participants. **Participants do not need to face the camera completely, they can look to the left or right, as long as the camera can capture at least 50% of the full-face image. They are 6 males (24 years old to 26) and 4 females (22 years old to 24).** One participant will respectively take 5 images at 09:00, 12:00, and 16:00 **in indoor** on each day. **And the brightness of the lighting is not an important issue because thermal imaging is used.** During the 10-day collection, a total of 7,500 images were stored in the dataset from participants. These 7,500 images will be used to train, validate and test prediction models of face recognition.

C. Image Preprocessing and Feature Extraction

The primary preprocessing in this research is image resize and image cutting, and feature extraction will produce two features, which are feature image and feature matrix. Fig. 3 shows the schematic diagram of preprocessing and feature extraction.

First, after acquiring one image, four other thermal images will be generated. In other words, each photographing program will have one RGB image and four thermal images, which are generated by Rainbow, Iron, Gray, and Wheel modes. Since the raw image size is up to 1080 × 1440, considering the input size limitation of the deep learning model, all images are cut to leave only the head and resize to 240 × 240. The Landmark function is used to detect and label facial feature points based on resized image, as shown in Fig. 4. The feature points output by Landmark do not completely in line with our design, thus, the feature points are further expanded based on the detected feature points.

Since the features of organs cannot be clearly recognized in thermal images, the Landmark function is not possible to achieve feature marking. In this research, the Landmark function will be implemented on the RGB image to obtain the coordinates, and synchronously mapped to the other four thermal images. At the 4th step in Fig. 3, the facial thermal image can be cut into 24 or 48 blocks after connecting the coordinates into a grid. From the 5th to 6th step, the RGB

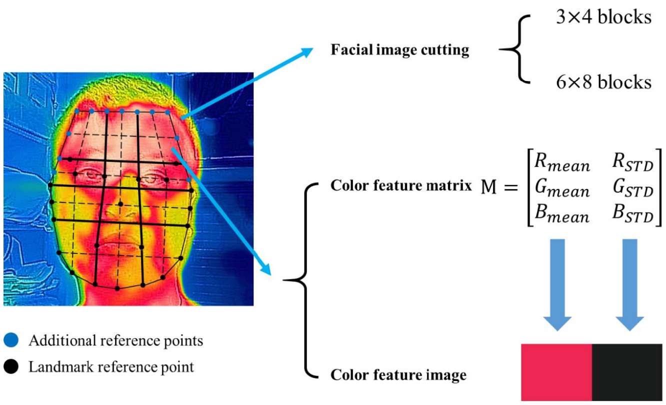


Fig. 4. The cutting grid and feature extraction of facial image.

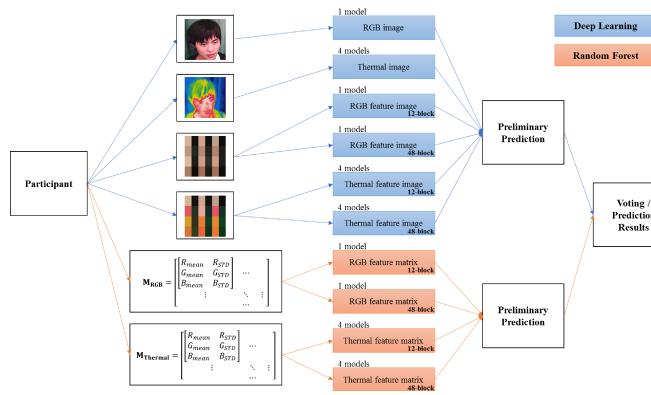


Fig. 5. Diagram of model arrangement and input dataset.

color mean value and standard deviation (STD) the value of each block in the grid were calculated to build a new matrix from color parameter, which is called feature matrix in this research. Another feature image is created based on color parameter from feature matrix. After preprocessing and feature extraction, a resized image, a feature matrix, and a feature image were obtained, and these are the input-data of the prediction model.

D. Ensemble Learning of Model and Datasets

The purpose of this research is to verify the effectiveness of thermal imaging to improve face recognition. De-identification and thermal features are the topics of concern in this process. In order to protect the privacy of participants (such as name), all images were desensitized before using, and participants are named using code names (for example, the first name is ID-01). These RGB images and thermal images are processed to generate more biological features to enhance recognition performance. A total of 7,500 images consisting of 5 categories and each with 1,500 images, which were collected from 10 participants. The dataset in each category is randomly divided into three parts: 1200 (80%) for training, 150 (10%) used for validation, and 150 (10%) for testing. After preprocessing and feature extraction, each category will indirectly generate 1,500 feature images and 1,500 feature matrices for other recognition sub models, as shown in Fig. 5. In other words, the proposed architecture consists of 25 sub-models, and the final result is generated by ensemble learning based on voting. The training set was used to train our model while the validation set was used to estimate how well our model had been trained. The test set was used to test the effectiveness of the research results.

E. Deep Learning and Random Forest

In the implementation of deep learning, we choose a universal deep learning development platform, the NVIDIA Deep Learning GPU Training System (DIGITS). The DIGITS simplify common deep learning tasks such as managing data on multi-GPU systems, designing and training neural networks. DIGITS is completely interactive so that user can focus on designing and training networks, rather than programming and debugging. DIGITS is also highly compatible, and it can run many different deep learning frameworks, such as CAFFE, Torch, and TensorFlow. CAFFE is a deep learning framework made with expression, speed, and modularity in mind as well as contained many classic deep neural network (DNNs) models such as GoogLeNet, AlexNet, LeNet, and VGGNet. Among these models, GoogLeNet has excellent performance in image applications in classic CNN. Considering the convenience of implementation, DIGITS is selected as the platform to implement CAFFE framework, and train GoogLeNet as the core model to observe the image prediction results in the proposed architecture.

Random Forest is the ensemble of many random trees, also known as a kind of Ensemble Learning. This kind of learning architecture is a combination of multiple weak learners to build a strong and stable model, and this model is less likely to be biased or overfit. In recent years, random forests have been very popular, and they have been used in many machine learning applications. It has strong performance and predictive ability, and is easy to use. In addition, random forests can also observe the importance of each feature. In this study, the characteristics of random forests are used for the classification of feature matrices for face recognition of participants.

Our face recognition is based on thermal facial image and its feature to enhance performance of recognition. In this proposed architecture mainly contained two steps: 1) feature image and feature matrix (biological feature) are obtained from the raw image based on the proposed feature extraction method, 2) train deep learning models to validate the predictive performance of different input dataset. In this research, we used deep learning techniques to automatically perform face recognition task. Specially, a fully convolutional neural network inspired from GoogLeNet were employed to predictive/classify participants from RGB image, thermal image, and feature image as well as the Random Forest is using for feature matrix. Each type of input data will have an independent sub-model for prediction, and finally determine the identity of the participant based on the voting results of the sub-model.

III. Results

A. Proposed Feature Extraction of Facial Image

In image processing issues, image size affects not only resolution but also processing speed. In view of this, all

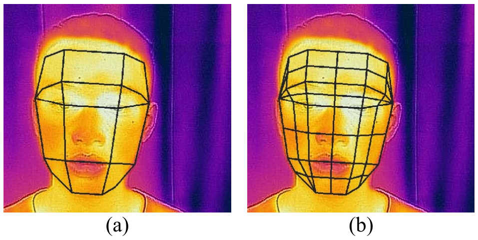


Fig. 6. Face image cutting results based on landmark. (a) 12 blocks, (b) 48 blocks.

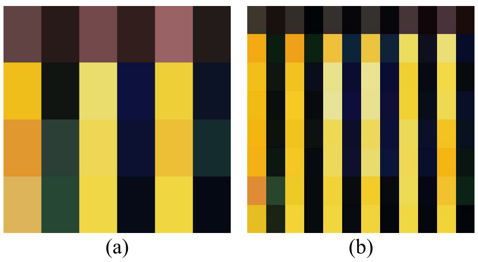


Fig. 7. Feature image based on block of facial image. (a) 12 blocks, (b) 48 blocks.

the images will cut the extra images to leave the head and resize to 240 × 240 to retain important image information before doing the main work. In the proposed feature extraction method, several blocks are cut from the facial image based on Landmark and the color parameters in the blocks are calculated to generate a new matrix. We cut the face image into 12 (3 × 4) and 48 (6 × 8) blocks (as shown in Fig. 6), and calculated the mean value and STD value of the RGB color in each block, so each block is a 1 × 6 color parameter matrix. This new facial feature matrix will also become a new feature image (as shown in Fig. 7), and a block will generate two color patches, which are generated by the mean and STD values, respectively.

B. Image-Based Model Training and Validation

This section describes the results of modeling, validation and testing of deep learning models for images. Each image mode has 1500 data. According to the architecture shown in Fig. 1, 1200 data are used for modeling, 150 are used for validation, and the remaining 150 are used for testing. There are 15 image models (see Fig. 5) for different types of images. In the complete procedure, the history of model is based on the accuracy and loss with each epoch, and the accuracy of the model is calculated using the confusion matrix. The confusion matrix, which clearly lists the prediction results of each participant. Based on this result, the prediction accuracy of each participant can be calculated. In order to facilitate the comparison of model performance, we average the accuracy of 10 participants to represent the accuracy of each prediction model. This table of confusion matrix is used in training, validation, and testing to calculate the accuracy of predictions.

Fig. 8 shows the modeling history of one of the 15 models. This graph simultaneously records the accuracy and loss of training and validation for model tuning. In order to obtain fair training results, we ensure that all 15 models have achieved

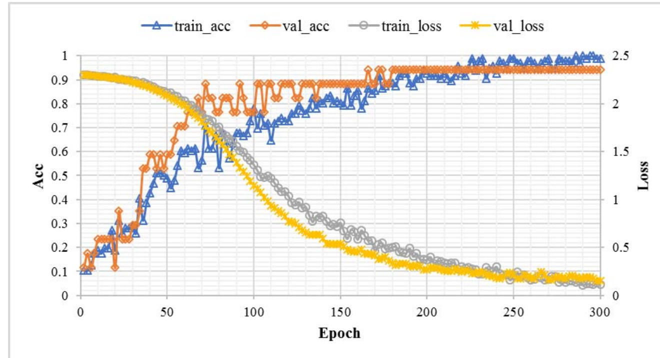


Fig. 8. The history curve of accuracy and loss for thermal image (Wheel mode).

TABLE I
CONFUSION MATRIX OF IMAGE PREDICTION BASED ON DEEP LEARNING IN TESTING (RGB IMAGE)

Label \ Predict	ID001	ID002	ID003	ID004	ID005	ID006	ID007	ID008	ID009	ID010
ID001	**97**	2	1	1	3	4	4	3	2	3
ID002	4	**95**	0	1	2	1	2	3	9	3
ID003	1	2	**94**	0	0	3	1	3	7	9
ID004	1	0	0	**98**	0	1	9	5	4	2
ID005	0	2	1	6	**99**	1	2	3	5	1
ID006	1	2	1	3	2	**101**	4	2	1	3
ID007	2	4	2	2	2	3	**100**	2	2	1
ID008	4	0	2	6	2	0	1	**96**	6	3
ID009	0	2	0	1	1	1	1	1	**110**	3
ID010	2	1	2	1	4	3	2	4	2	**99**

convergence results. The confusion matrix is used to calculate the accuracy rate, as shown in Table I, record the RDB image test results and get an accuracy rate of 0.824. Test results for all image modes will be filled in Table II. It can be observed from the comparison table that all image modes get an accuracy of 1 during the training. During the validation, the image can still reach an accuracy of 1. In the 12-block, the RGB image only has an accuracy of 0.412. The worst accuracy in the thermal image has the Iron mode with 0.706. In the 48-block, the RGB image also gets the lowest accuracy rate. The worst of the thermal images is still the Iron mode, but the accuracy rate has been increased from 0.706 to 0.882. In the testing, the RGB image can only achieve an accuracy of 0.824 when tested using the image, and the prediction effect cannot be clearly seen using the feature image. Among the thermal images, the Iron mode is still the worst, and the other three modes can achieve a maximum accuracy of 0.947 in a 48-block thermal feature image.

C. Matrix-Based Model Training and Validation

In this research, Random Forest is using Dlib (Open Source Library) on Anaconda. For the feature matrices of the five

TABLE II
ACCURACY COMPARISON TABLE OF DEEP LEARNING RESULTS IN DIFFERENT CUTTING METHODS AND IMAGE MODES

	Train			Validation			Testing		
	Image	12-block	48-block	Image	12-block	48-block	Image	12-block	48-block
RGB	1	1	1	1	0.412	0.647	0.824	0.352	0.617
Rainbow	1	1	1	1	1	1	0.882	0.882	0.947
Iron	1	1	1	1	0.706	0.882	0.706	0.681	0.647
Gray	1	1	1	1	0.765	1	0.882	0.663	0.945
Wheel	1	1	1	1	0.824	1	0.941	0.657	0.941

TABLE III
CONFUSION MATRIX OF IMAGE PREDICTION BASED ON RANDOM FOREST IN TRAINING (RGB FEATURE, 12 BLOCKS)

Predict / Label	ID001	ID002	ID003	ID004	ID005	ID006	ID007	ID008	ID009	ID010
ID001	**120**	0	0	0	0	0	0	0	0	0
ID002	0	**117**	0	1	0	1	1	0	0	0
ID003	0	1	**119**	0	0	0	0	0	0	0
ID004	0	0	0	**119**	0	1	0	0	0	0
ID005	0	0	1	0	**115**	1	1	2	0	0
ID006	0	0	0	1	0	**119**	0	0	0	0
ID007	0	0	0	0	0	0	**119**	0	1	0
ID008	0	0	0	0	0	0	1	**119**	0	0
ID009	0	0	0	0	0	1	1	0	**118**	0
ID010	0	0	0	0	0	0	0	0	1	**119**

images, there are 1500 data each, 80% for training, 10% for validation, and the rest for testing. The input format of the data is a two-dimensional matrix [1500, 6 $*$ 12] and [1500, 6 $*$ 48], where 6 represents the mean value and STD value of the RGB parameters, and 12 and 48 respectively represent the number of cut squares of the facial image. In the model's training parameters, the estimator is set to 1500, the criterion is set to gini, and sampling with replacement when building the tree.

Table III shows the confusion matrix of the training results of the RGB feature matrix with 12-block. There are 1200 data in this matrix for training. We then used 10% of the data for model validation and the remaining 10% for testing. Table IV is used to record the accuracy of all modes. The results show that when using the feature matrix for prediction, the accuracy of the RGB mode (12-block) is only 0.335hile the lowest accuracy of the thermal image is still as high as 0.823. Among them, Iron and Gray reached a test accuracy of 0.941 on 12 blocks and 48 blocks.

D. Ensemble Learning Model Testing

This section described the results of ensemble learning. Since this study proposes 25 sub-models for face recognition

TABLE IV
ACCURACY COMPARISON TABLE OF RANDOM FOREST RESULTS IN DIFFERENT CUTTING METHODS AND IMAGE MODES

	Train		Validation		Testing	
	12-block	48-block	12-block	48-block	12-block	48-block
RGB	0.987	1	0.667	0.764	0.335	0.647
Rainbow	1	1	1	1	0.882	0.882
Iron	1	1	1	1	0.941	0.941
Gray	1	1	1	1	0.941	0.941
Wheel	1	1	1	1	0.823	0.941

TABLE V
ACCURACY COMPARISON TABLE OF RANDOM FOREST RESULTS IN DIFFERENT CUTTING METHODS AND IMAGE MODES

	Raw Image	**Feature Image (12)**	**Feature Image (48)**	**Feature Matrix (12)**	**Feature Matrix (48)**
RGB	0.834	0.413	0.647	0.393	0.653
Rainbow	0.887	0.893	0.947	0.873	0.907
Iron	0.673	0.707	0.873	0.927	0.947
Gray	0.887	0.767	0.953	0.953	0.967
Wheel	0.947	0.813	0.927	0.813	0.953
※Test Round:150 Accuracy of Ensemble Learning:1					

for different input data, it is quite suitable to use the architecture of ensemble learning to vote for comprehensive prediction results. We used 10% of the data to vote, so a total of 150 rounds were executed. The voting results are based on the majority results predicted by the 25 sub-models as the final prediction results, and in addition to the conditions for a majority decision, the number of votes as the final result must exceed half the total number of votes. The prediction accuracy of each last sub-model is also recorded in Table V.

Under the ensemble learning architecture, the majority answer is correct every time, so the accuracy rate is 1. When the test results in Table II and Table III are compared with Table V, it can be observed that the results of the model self-test and the results of ensemble learning are quite close. According to such results, it can be shown that each sub-model has a stable prediction ability that has converged, and the same response trend is also observed.

IV. DISCUSSION

In this research, we proposed a new facial image extraction method for face recognition. The proposed method cuts the facial image (RGB and thermal image) into 12 and 48 blocks respectively, and then reconstructs it into a feature image and a feature matrix. The images of 10 participants participated in the training, validation and testing of the model, and finally performed the final prediction using ensemble learning.

First, RGB images and thermal images are used to build deep learning prediction models. The test results are recorded in Table II. From the table, we can observe the following results. (1) When the input data type is image (RGB and thermal), a good accuracy rate can be obtained. The Wheel mode (0.941) works best and the Iron mode (0.706) works the worst. (2) When the input data type is a feature image, the prediction effect of 48-block is better than 12-block. (3) The prediction performance of RGB feature images is significantly lower than that of thermal feature images. (4) Compared with other image modes, the Rainbow mode of thermal image can achieve high accuracy with image, 12-block and 48-block.

Second, a feature matrix is generated after the image is cut and recreated. This matrix is used to predict random data by using random forest. The prediction results are shown in Table IV. We can observe that the worst prediction performance after the thermal image is converted into the thermal feature matrix has an accuracy of 0.882, which is a fairly high accuracy rate relative to the RGB feature matrix. Regardless of the 12-block or 48-block cutting method of the RGB feature matrix, a reliable accuracy rate cannot be obtained. In other words, the RGB image cannot use the proposed method to enhance the prediction performance. On the contrary, the thermal feature matrix has achieved good results, especially Iron mode (0.941) and Gray mode (0.941).

From the prediction results of deep learning, random forest, and ensemble learning, we can find the following interesting phenomena. RGB images are only suitable for prediction using images. If RGB is converted into feature images or feature matrices, the visible features on the face will be lost. At this time, photos can be disguised as humans to deceive the recognition system, and personal privacy issues will also be raised. Three modes of thermal image, Rainbow mode, Gray mode and Wheel mode are suitable for image and feature image for face prediction, and the prediction performance is higher than RGB image. Although the Iron model is not suitable for prediction with thermal images, it has an amazing effect on the prediction of the thermal feature matrix. When the RGB face image is converted into a thermal image, it is the first de-identification process. When it is converted into a thermal feature image or a thermal feature matrix, it is the second de-recognition process. For human faces, in addition to visible features that can be used as a basis for identification, invisible thermal features are important nonforgeable biological features. Therefore, after de-identification, the system not only can still recognize, but also can obtain a higher prediction ability than the original image. However, each thermal image mode has its own suitable temperature distribution. Therefore, different types of input data can be mutually enhanced through the ensemble learning architecture to reduce he possibility of misidentification.

V. Conclusions

In this research, we proposed a new image extraction method for face recognition, which are used deep learning, random forest, and ensemble learning to recognize human. The results show that facial thermal images and facial RGB images can also achieve a high prediction effect, and the feature images and feature matrices produced by the proposed method can achieve better prediction performance. Based on our experimental results, we proposed the following suggestions. (1) The thermal image can be de-identified by converting the feature image and the feature matrix. (2) The FLIR-based thermal image capture system should consider using Rainbow mode, Gray mode, and Wheel mode as the thermal image and produce feature image, and Iron mode for produce the thermal feature matrix. (3) Non-FLIR-based systems use thermal feature images and thermal feature matrices to enhance the recognition of the original thermal image. (4) Although deep learning and random forests already have high accuracy, using the ensemble learning integration sub-model can enhance prediction performance.

References

[1] T. de Freitas Pereira, A. Anjos, and S. Marcel, "Heterogeneous face recognition using domain specific units," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1803–1816, Jul. 2019, doi: 10.1109/TIFS.2018.2885284.

[2] G. Batchuluun, H. S. Yoon, J. K. Kang, and K. R. Park, "Gait-based human identification by combining shallow convolutional neural network-stacked long short-term memory and deep convolutional neural network," *IEEE Access*, vol. 6, pp. 63164–63186, 2018, doi: 10.1109/ACCESS.2018.2876890.

[3] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 42–55, 2020.

[4] K. Kotwal, S. Bhattacharjee, and S. Marcel, "Multispectral deep embeddings as a countermeasure to custom silicone mask presentation attacks," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 1, no. 4, pp. 238–251, Oct. 2019.

[5] G. Hermosilla Vigneau, J. L. Verdugo, G. Farias Castro, F. Pizarro and E. Vera, "Thermal face recognition under temporal variation conditions," *IEEE Access*, vol. 5, pp. 9663–9672, 2017, doi: 10.1109/ACCESS.2017.2704296.

[6] B. Jian, C. Chen, M. Huang and H. Yau, "Emotion-specific facial activation maps based on infrared thermal image sequences," *IEEE Access*, vol. 7, pp. 48046–48052, 2019, doi: 10.1109/ACCESS.2019.2908819.

[7] H. Kaur and P. Khanna, "Random distance method for generating unimodal and multimodal cancelable biometric features," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 709–719, Mar. 2019, doi: 10.1109/TIFS.2018.2855669.

[8] R. M. A. Kiekens, A. M. Kuijpers-Jagtman, M. A. van 't Hof, B. E. van 't Hof, and J. C. Maltha, "Putative golden proportions as predictors of facial esthetics in adolescents," *Amer. J. Orthodontics Dentofacial Orthopedics*, vol. 134, no. 4, pp. 480–483, Oct. 2008.

[9] M. Peter Prendergast, "Facial -proportions," in *Proc. Adv. Surgical Facial Rejuvenation*, pp. 15–22, Jul. 2011.

[10] S. G. Arslan, C. Genç, B. Odabaş, and J. D. Kama, "Comparison of facial proportions and anthropometric norms among turkish young adults with different face types," *Aesthetic Plastic Surgery*, vol. 32, no. 2, pp. 234–242, Mar. 2008.

[11] L. Sun and Z. Zheng, "Thermal-to-visible face alignment on edge map," *IEEE Access*, vol. 5, pp. 11215–11227, 2017, doi: 10.1109/ACCESS.2017.2712159.

[12] W. Hardin and C. Editor, "Thermal imaging to diagnose disease," in *Proc. Vision Online*, Aug. 2018, pp. 1–8.

[13] U. Jain, B. Tan, and Q. Li, "Concealed knowledge identification using facial thermal imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1677–1680.

[14] D. Li, C. C. Menassa, and V. R. Kamat, "Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography," *Energy Buildings*, vol. 176, pp. 246–261, Oct. 2018.

[15] B. Cao, N. Wang, J. Li, and X. Gao, "Data augmentation-based joint learning for heterogeneous face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1731–1743, Jun. 2019.

[16] Z. Wang, Z. Chen, and F. Wu, "Thermal to visible facial image translation using generative adversarial networks," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1161–1165, Aug. 2018.

[17] G. Hermosilla, "Particle swarm optimization for the fusion of thermal and visible descriptors in face recognition systems," *IEEE Access*, vol. 6, pp. 42800–42811, 2018, doi: 10.1109/ACCESS.2018.2850281.

[18] S. Wang, B. Pan, H. Chen and Q. Ji, "Thermal augmented expression recognition," *IEEE Trans. Cybern.*, vol. 48, no. 7, pp. 2203–2214, Jul. 2018, doi: 10.1109/TCYB.2017.2786309.

[19] C. Peng, X. Gao, N. Wang and J. Li, "Graphical representation for heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 301–312, Feb. 2017, doi: 10.1109/TPAMI.2016.2542816.

[20] M. M. Khan, R. D. Ward and M. Ingleby, "Toward use of facial thermal features in dynamic assessment of affect and arousal level," *IEEE Trans. Affective Comput.*, vol. 8, no. 3, pp. 412–425, Jul./Sep. 2017, doi: 10.1109/TAFFC.2016.2535291.

[21] B. A. Rajoub and R. Zwiggelaar, "Thermal facial analysis for deception detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 1015–1023, Jun. 2014, doi: 10.1109/TIFS.2014.2317309.

[22] C. A. Corneanu, M. O. Simón, J. F. Cohn and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016, doi: 10.1109/TPAMI.2016.2515606.

[23] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard and D. Merhof, "A thermal infrared face database with facial landmarks and emotion labels," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1389–1401, May 2019, doi: 10.1109/TIM.2018.2884364.

[24] I. A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis and N. Murtuza, "Multimodal face recognition: Combination of geometry with physiological information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)* vol. 2, San Diego, CA, USA, May 2005, pp. 1022–1029, doi: 10.1109/CVPR.2005.241.

[25] A. M. Guzman, "Thermal imaging as a biometrics approach to facial signature authentication," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 214–222, Jan. 2013, doi: 10.1109/TITB.2012.2207729.

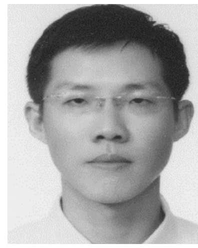
[26] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 613–626, Apr. 2007.

[27] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun. 2013.

[28] M. O. Simón *et al.*, "Improved RGB-D-T based face recognition," *IET Biometrics*, vol. 5, no. 4, pp. 297–303, 12 2016.

[29] B. R. Nhan and T. Chau, "Classifying affective states using thermal infrared imaging of the human face," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 979–987, Apr. 2010.

[30] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis, "Imaging facial signs of neurophysiological responses," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 477–484, Feb. 2009.



Chih-Hsueh Lin (Member, IEEE) was born in Kaohsiung, Taiwan, in 1977. He received the Ph.D. degree in computer science and engineering from National Sun Yat-sen University, Kaohsiung, in 2006.

From 2009 to 2018, he was an Assistant Professor at the Department of Computer and Communication, Shu-Te University, Kaohsiung. Since 2018, he has been an Assistant Professor with the Department of Electronics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung. His research interests include machine learning, information security, biomedical signal processing, and data mining.

Zhi-Hao Wang received the M.S. degree in electronics engineering from the National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, in 2017, where he is currently pursuing the Ph.D. degree. His interests are in renewable energy applications, energy management and control systems of electric vehicles, electromechanical integration, wireless sensing networks, the IoT service applications, and embedded systems.

Gwo-Jia Jong (Member, IEEE) received the B.S. degree in electronics engineering from the National Taiwan University of Science and Technology, Taiwan, in 1984, and the Ph.D. degree in electrical engineering from Paisley University, U.K., in 2000. In 2002, he joined the Faculty of National Kaohsiung University of Science and Technology, Taiwan, where he is currently an Associate Professor with the Department of Electronics Engineering. In 2011, he was also the Director of the Development Center of United Electronics Communication Technology (CECT), Ministry of Education, Taiwan. His research interests include electronics engineering, communication engineering, wireless sensor networks (WSNs), digital communication systems, and system integration.