Personal Identification Based on Deep Learning Technique Using Facial Images for Intelligent Surveillance Systems

Van-Huy Pham, Diem-Phuc Tran, and Van-Dung Hoang

Abstract—**Identity recognition is a very important task in intelligent surveillance systems. Today, identity recognition systems have achieved high accuracy and widely used in specific application areas such as recognition system based on retina imaging in immigration inspection, civil security and citizen management. In these systems, human is required to be submissive for data acquisition to identify themselves. However, the automated monitoring systems are required to be active for information retrieval and human is passively monitored in this situation. In this kind of approach, human recognition is still a challenging task for the overall system performance. This study proposes a solution for human identification based on the human face recognition in images extracted from conventional cameras at a low resolution and quality. Our proposed approach for human identification is based on a deep learning method for feature extraction and classification for human identification using a similarity estimation. This approach was evaluated on some standard databases which are available online and also on our own collected dataset. The results from the comparison to the state of the art approach illustrate that our proposed approach achieves high accuracy and is suitable for practical applications.**

Index Terms—**Personal identification, face image, feature extraction.**

I. INTRODUCTION

In recent years, with the rapid development of science and technology, monitoring systems based on artificial intelligence techniques have been studied and applied in many fields of security monitoring, medicine application, education, intelligence transportation systems, etc. [1]-[4]. One of the most interesting orientations is focusing on how to improve the efficiency and feasibility of monitoring solutions, ensuring security in the specialized situation or public applications [5]. In narrow specialized applications, that systems could achieve high precision, such as human identification systems using retinal image, DNA parsing, fingerprint image, information types from other biometrics. However, in public surveillance applications, there are many limitations for human identification, for example, identifying person's name in crowds at public airports or docks, train stations.,. Thus, it is possible to classify the systems in some different approaches: (1) systems which require an obligation

to nominate person in order for active data receiving; (2) systems which would not require an obligation to nominate person for the task. The data should be received within a passive manner in applications such as security monitoring systems, detecting crime in the crowds and so on. In the first kind of the applications, data of human retina, fingerprint can only be received when human is required to provide information for the systems. However, there are some surveillance systems which are required to be actively and fully automated in the task of data acquisition, meanwhile recognition systems based on retinal image, fingerprints could not be done. In this kind of applications, a human recognition system based on face, or human walking style information is applicable. Like human beings, the human brain can identify the name of a person if it has enough information about this person. In this study, we propose an approach with expecting to be able to identify a large number of human identities through their facial images by providing an enough amount of image data for training a machine learning model for human recognition and identification task.

II. RELATED WORKS

Deep learning techniques have been widely studied by many scientists and have achieved high accuracy in many applications and research. Deep learning techniques have been used extensively in various areas such as object classification, product defecting monitoring, fraud detection, handwriting recognition, voice recognition, medical diagnosis, and so on. There are some studies, which related to person identification, which is primarily driven by the use of special biometric data. There are some groups researching on the topic of identification based on analyzing the iris image, retinal, blood vessel data such as [6]-[9]. Face biometric quality assessment can be used for human recognition [10]. Authors presented a biometric quality assessment method for facial image processing. In that method, a classifier is trained for simultaneously predicting categories of facial image and degree of the degradation. Experimental results on the CASIA, FLW, YouTube dataset conducted that the proposed approach the effectiveness of the proposed method. Another author group presented an approach for multimodal biometric personal identification and verification [9]. The method is a cascaded multimodal biometric system with combination of fingerprint and iris recognition. The key ideas for fingerprint iris recognition are minutiae extraction of fingerprint information and encoding the Log Gabor filtering. The experiments on some dataset demonstrated the superior performance of that approach compared to unimodal systems. Another method also used iris for biometric identification presented in a pattern [7]. Meanwhile authors in [6] presented

a technique for online iris image compression and personal identification, which called iris image compressor and identifier. That method converts the iris image of eye in the form of Laplace–Beltrami Spectra, then Laplace–Beltrami spectra is converted into the form of Strakos matrix. The matrices of Eigen values are calculated for identifying a person. The proposed method was experimented on a hundred iris images of CASIA database. The experimental results illustrated that the method is robustness, efficient and economically feasible. Patwari [8] presented an algorithm using retinal blood vessels bifurcation for personal identification. Their method focuses on detection of retina blood vessels and measurement bifurcation points of blood vessels. The proposed method was performed on real data of 300 images which collected and analyzed by doctors. That algorithm achieved high accuracy score, which is suitable for real application.

Other approaches focus on different kinds of information for biometric measurement for human identification, such as palatal, skin texture in [11]-[13]. Gibelli and *et al.* in [11] presented method based on 3D superimposition techniques for constructing 3D models of palatal reggae to personal identification. The research provides a uniqueness assessment of palatal rugae based on their anatomical 3D conformations. Experimental results showed that the method achieved high precision and applicable. A personal recognition method based on hand back skin texture is proposed in [12]. In that paper, authors investigated problem of the hand back skin texture pattern for personal identification and gender classification. The model was trained and classified the hand back skin texture patterns in good result scores. Evaluation results demonstrated an efficient for assist human identification and gender classification. Meanwhile, Yang and *et al.* in [14] proposed an approach based on fusion at the feature level of finger vein and finger dorsal texture. The method was compared the state-of-the-art unimodal biometrics on the performance. Compared results on established dataset illustrated that their approach achieved higher identification accuracy and lower equal error rates to other methods.

In different approach, our contribution presents an identity recognition approach based on deep learning techniques using facial image data, which is collected from surveillance camera with normally quality of image.

III. PROBLEM STATEMENT AND SYSTEM OVERVIEW

Human identification plays an important role in many applications such as monitoring and human management systems. This research focuses on identity recognition based on face information which is extracted from surveillance cameras. According to previous discussion, identity recognition uses special information of a person such as retinal, irises, fingerprints, and other types of biometrics often provide high accuracy and have been applied in practical systems. However, there are some types of applications that cannot coerce human to provide their information. In this situation, the surveillance systems have to be active for collecting the data and human is to be passive in the process of systems.

This article presents some results on identity recognition

task studied by the use of artificial intelligence and image processing techniques for face detection and identification. The general scheme of the identity recognition system is shown in the diagram in Fig. 1.
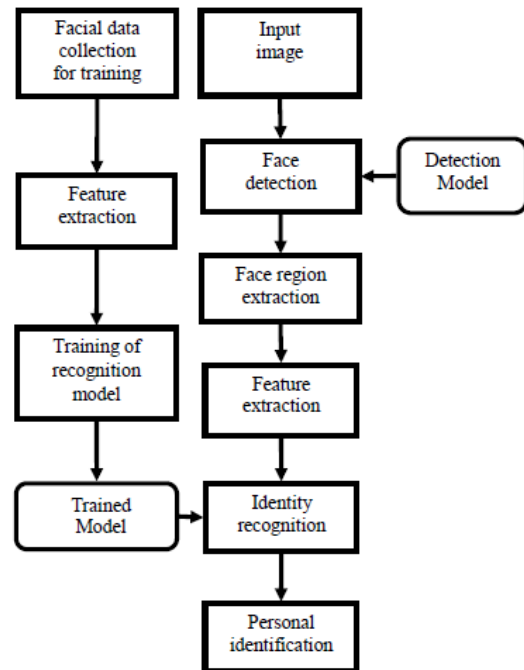

Fig. 1. General flowchart of the identity recognition system.

The proposed solution consists of two tasks: (1) detecting positions of faces from the input image acquired from surveillance cameras; (2) identity recognition of the detected face regions from the first task. In this paper, we focus on how to improve the accuracy of human identity recognition in the second task. The task of facial detection was done by the proposed method for real time face detection by Viola[15]. In the detection task, a facial model was trained on a face dataset using the Haar-like feature and AdaBoost classifier, resulting in a detection model with detection rate of 98%s. We provide a recognition solution based on the tasks of feature extraction and identity recognition using deep learning approach. There is a number of approaches for this problem with the use of shallow learning techniques based on the traditional feature extraction method. In traditional approaches, there are many methods for feature extraction such as HOG, SHIFT, LBP, Haar-wave... The feature extraction results are then fed as inputs for classifiers trained by SVM, boosting, decision tree, GA, etc. In our approach, deep learning technique is used for both the objectives of feature extraction and classification task.

IV. ARCHITECTURE PROPOSAL

A. Deep Learning Architecture Construction

In general, there are some existing pre-trained models which have already trained based on huge datasets such as AlexNet, Googlenet, VGG, RCNN, ... that can be used for transfer learning on our dataset. However, in our approach, using of pre-trained models are inappropriate because the input size of the old models is unsuitable to the collected images. Additionally, the trained parameters do not significantly support for improving accuracy of the identity

recognition system. There are some ways for constructing the deep learning network for object recognition. In this research, we proposed a deep architecture based on a serial convolutional neural network.

The deep network architecture includes 25 layers, as shown in Table I, which consists of several network layers such as input image layer, convolution layer, rectified linear unit (ReLu) layer, standard cross normalization layer, max-pooling layer, and fully connected layer. The network model maps the input image into the serial hierarchical analysis. The input layer of the network is processed on image with 192×192×3 size.

TABLE I: THE DEEP NETWORK ARCHITECTURE CONSISTS OF 23 HIDDEN LAYERS AND THE INPUT LAYER AND THE FINAL CLASSIFICATION LAYER

TT	Layer name	Description
1	Image Input	192x192x3 image
2	Convolution	64 7x7x3 convolutions, stride [1 1]
3	ReLU	Rectify linear unit
4	Normalization	Cross channel normalization
5	Max Pooling	3x3 max pooling, stride [2 2]
6	Convolution	64 7x7x64 convolutions, stride [1 1]
7	ReLU	Rectified linear unit
8	Max Pooling	3x3 max pooling, stride [2 2]
9	Convolution	64 7x7x64 convolutions, stride [1 1]
10	ReLU	Rectified linear unit
11	Max Pooling	2x2 max pooling, stride [1 1]
12	Convolution	96 7x7x64 convolutions, stride [1 1]
13	ReLU	Rectified linear unit
14	Max Pooling	2x2 max pooling, stride [1 1]
15	Convolution	96 7x7x96 convolutions, stride [1 1]
16	ReLU	Rectified linear unit
17	Normalization	Cross channel normalization
18	Max Pooling	3x3 max pooling, stride [2 2]
19	FulConnection	1024 fully connected layer
20	ReLU	Rectified linear unit
21	FulConnection	512 fully connected layer
22	ReLU	Rectified linear unit
23	FulConnection	n fully connected layer
24	Softmax	Softmax
25	Classification	Crossentropyex n classes

B. Data Augmentation for Training

One of the important problems in deep learning is the required number of samples, which is large enough for training a deep learning model for appropriate applications. Therefore, it is important to augment data for small datasets. Recently, there are some data augmentation method [16]-[21] that were commonly used to augment data to avoid the problem of data overfitting in deep learning. The underlying concept of image augmentation is that the rotation, flipping, and other deformations can be applied to enhance data without changing and loosing characteristics and properties of data. In our experiments, we have applied several kinds of

data augmentation as follows.

Color normalization and balance: the facial images were collected in different resources, illumination conditions and from different types of devices. Therefore, it can help to improve the system accuracy by normalization of the colors of the images before using them for the training to improve performance of classification system [22].

Geometry transformations: affine transformations such as shearing, distorting and scaling randomly warp stroke data for image classification [23]. Thus, affine transformations are very well suited to augment data for improving the overall performance and mitigating overfitting of the training task, especially, in the case of deep learning. Images from the training set are affine transformed for producing a larger dataset used for training. We used various techniques for resampling such as rotation, stretching, shearing: random rotation with angle between -10° and 10°, random shearing with angle between -10° and 10°; random shearing stretching with stretch factor between 5° and 15°; flipping is also applied in this situation because almost facial shapes are mirror. All images are normalized by converting all pixels into [-2.0, 2.0] range to create normalized data.

Fig. 2. Example of augmentation images: 30 outputs from an original input.

TABLE II: DATA AUGMENTATION FOR IMPROVING ACCURACY

Dataset	#Class	#Samp	Original set	Augmentation	
			ACC	#Samp	ACC
ATT-faces	40	240	25.00	6000	88.13
Our dataset	27	1673	91.89	18140	96.14

The accuracy result (ACC) of the recognition system when training on the original dataset and the data augmentation have shown that training on the augmented dataset has significantly improving the accuracy rate of recognition system. Especially, in the case of the ATT dataset, with only 10 samples for each class, the model is insufficient the problem of under-fitting for training deep learning models. It only results in the accuracy of 25%. However, the trained model on augmented dataset have reached higher accuracy.

V. EXPERIMENTAL RESULTS

We evaluated the proposed method in comparison with the

traditional approach on some datasets as a basis for the evaluation. The comparisons were conducted using: (i) the traditional method based on the use of HOG-feature descriptor and SVM-recognition; (ii) a CNN model which were trained on the original collected dataset; (iii) a CNN model which were trained on the augmented dataset. Details of the number of samples used for the training of (ii), (iii) and the accuracy results are shown in Table II.

The comparison results between the traditional approach and deep learning model trained on augmented dataset showed that the recognition system based on CNN with data augmentation is more accurate than shallow machine learning. In the following subsections, details of the evaluated results on some different datasets will be presented.
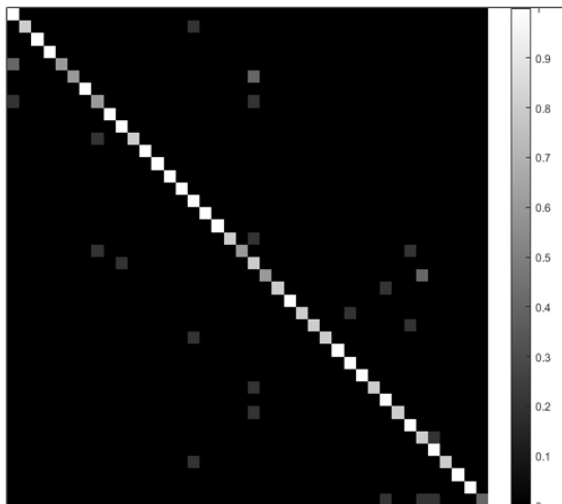

Fig 3. Some facial examples of ATT-faces dataset.

Fig. 4. Confusion matrix of recognition rate on ATT-faces.

A. Evaluation on AT&T Dataset

The ATT_faces dataset was created by the AT&T Laboratories of Cambridge University [25]. This dataset consists of 400 face images of forty people, each person sampled with 10 different images, which stored in individual folders for each person. The images were taken at different times, changing light, facial expressions (eyes open/aiming, smiling/ not smiling) and wearing glasses or without glasses. All images are taken with a dark homogeneous background for objects in an upright, frontal position (with tolerance for some movement, a litter bit changes). Images with PGM format and gray-scale value with 256 gray levels per pixel the

uniform image resolution with 92 × 112 pixels. Some examples of facial pattern are shown in Fig. 3.

In this experiment, we partitioned dataset into 60% samples for training and 40% samples for evaluation. Different to our dataset, in ATT dataset, facial samples contain of the entire face and head of a person. A comparison on experimental results are shown in Fig. 4 and Table III. The results illustrated that our CNN model using training set, which is augmented to 25 times, is significant improvement to the traditional approach using HOG for feature description and SVM for classification.

TABLE III: COMPARISON RESULTS OF HOG+SVM AND CNN ON ATT-FACES DATASET

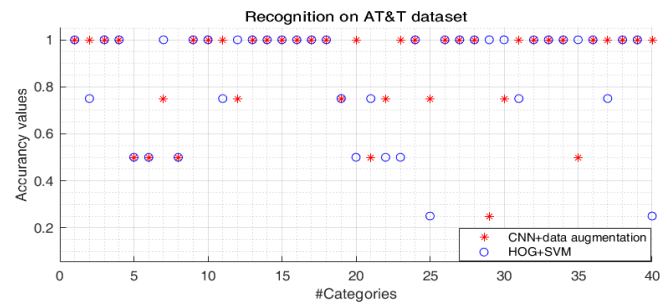| Methodology	Training samples	Test samples	Accuracy
HOG+SVM	240	160	85,00%
Our CNN	6000	160	88.13%

Fig. 5. Recognition results on each person of ATT-faces data.

B. Evaluation Results on Our Collected Dataset

Fig. 6. Some facial examples of our collected dataset.

The experimental dataset was collected by our team, which is called DStudent dataset. The data samples were taken by a low-resolution camera which mounted in a class of a high school. Facial images were captured under different lighting conditions. Then sample images were cropped and normalized so that each image contains the entire head with a

full face. The DStudent dataset contains 2,787 faces of 27 different students. Each student was sampled from 100 to 110 facial samples. Some face patterns are shown in Fig. 6.

In our experiment, we used about 60% samples of each student for training and about 40% sample of each student for evaluation. Experimental results, which are shown in Table IV. It illustrates that our CNN model trained on our dataset, which is augmented to 25 times, is significantly improved in compared with the traditional approach using HOG feature description and SVM for classification.

TABLE IV: COMPARISON RESULTS ON OUR DATASET

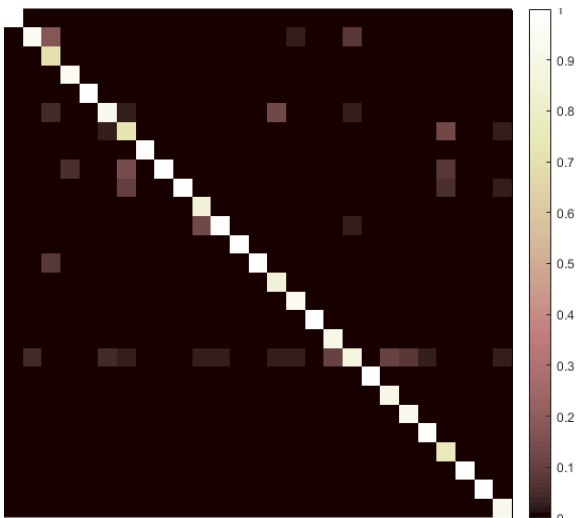| Methodology	Training samples	Test samples	Accuracy
HOG+SVM	1673	1114	92.18%
Our CNN	18140	1114	96.14%

Fig. 7. Confusion matrix of recognition rate on our dataset.

Fig. 8. Recognition results on each person of our dataset.

VI. CONCLUSION

Presently, there are some proposed solutions for identity recognition problem. In some specialized applications, the system achieved high accuracy for real application. However, public surveillance systems require general information for recognition of human identity still a challenging task due to some different reasons. In this study, we proposed an approach based on deep learning for recognition of identity human. Experimental results showed that the recognition system based on our proposed deep learning technique for human identification using facial images reached high accuracy, which can be applied in practical applications. Our proposed method has been studied on the basis of empirical

data, comparing of several machine learning methods to provide an effective solution for people identification problem. Our contribution consists of conducting evaluations to help the selection of an effective method for human identification problem and providing resulted details of each person identity with high accuracy using deep learning based on convolutional neural network. However, the system was only evaluated on a dataset with a small the number of people, from 27 to 84 categorical people. In future, we are improving the recognition model by experimenting on big dataset, so that the system can be applied to real application in intelligent monitoring systems.

REFERENCES

[1] Hironori Hattori, Namhoon Lee, Vishnu Naresh Boddeti, Fares Beainy, Kris M Kitani, Takeo Kanade, "Synthesizing a Scene-Specific Pedestrian Detector and Pose Estimator for Static Video Surveillance", *International Journal of Computer Vision*, pp. 1-18, 2018.
[2] Van-Dung Hoang, Kang-Hyun Jo, "Path planning for autonomous vehicle based on heuristic searching using online images," *Vietnam Journal of Computer Science*, Springer Berlin Heidelberg, 2014, pp. 1-12.
[3] Vishwanath A Sindagi, Vishal M Patel, "A survey of recent advances in cnn-based single image crowd counting and density estimation", *Pattern Recognition Letters*, vol. 107, pp. 3-16, 2018.
[4] Diem-Phuc Tran, Van-Dung Hoang, Tri-Cong Pham, Chi-Mai Luong, "Pedestrian activity prediction based on semantic segmentation and hybrid of machines", *Journal of Computer Science and Cybernetics*, vol. 34(2), pp. 113-125, 2018.
[5] Van-Dung Hoang, Van-Dat Dang, Tien-Thanh Nguyen, Diem-Phuc Tran, "A solution based on combination of RFID tags and facial recognition for monitoring systems," in 5th NAFOSTED Conference on Information and Computer Science (NICS), pp. 386-390, 2018.
[6] Kamta Nath Mishra, "An Efficient Technique for Online Iris Image Compression and Personal Identification," in Proceedings of International Conference on Recent Advancement on Computer and Communication, pp. 335-343, 2018.
[7] Keith J Hanna, "System and method for iris data acquisition for biometric identification," Google Patents, 2018.
[8] Manjiri B Patwari, Ramesh R Manza, Yogesh M Rajput, Manoj Saswade, Neha Deshpande, "Personal identification algorithm based on retinal blood vessels bifurcation," in 2014 International Conference on Intelligent Computing Applications (ICICA), pp. 203-207, 2014.
[9] Mohamed Elhoseny, Ahmed Elkhateb, Ahmed Sahlol, Aboul Ella Hassanien, "Multimodal biometric personal identification and verification," *Advances in Soft Computing and Machine Learning in Image Processing*, pp. 249-276: Springer, 2018.
[10] Jun Yu, Kejia Sun, Fei Gao, Suguo Zhu, "Face biometric quality assessment via light CNN", *Pattern Recognition Letters*, vol. 107, pp. 25-32, 2018.
[11] Daniele Gibelli, Danilo De Angelis, Valentina Pucciarelli, Francesco Riboli, Virgilio F Ferrario, Claudia Dolci, Chiarella Sforza, Cristina Cattaneo, "Application of 3D models of palatal rugae to personal identification: hints at identification from 3D-3D superimposition techniques", *International journal of legal medicine*, vol. 132(4), pp. 1241-1245, 2018.
[12] David Zhang, Guangming Lu, Lei Zhang, "Hand Back Skin Texture for Personal Identification," *Advanced Biometrics*, pp. 213-233: Springer, 2018.
[13] Kelly Sauerwein, Tiffany B Saul, Dawnie Wolfe Steadman, Chris B Boehnen, "The effect of decomposition on the efficacy of biometrics for positive identification", *Journal of forensic sciences*, vol. 62(6), pp. 1599-1602, 2017.
[14] Wenming Yang, Xiaola Huang, Fei Zhou, Qingmin Liao, "Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion", *Information sciences*, vol. 268, pp. 20-32, 2014.
[15] Paul Viola, Michael J. Jones, "Robust Real-Time Face Detection", *Intenational Journal of Compute Vision*, vol. 57(2), pp. 137-154, 2004.
[16] Kazuhisa Matsunaga, Akira Hamada, Akane Minagawa, Hiroshi Koga, "Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble", *ArXiv e-prints*, 2017].
[17] Iván González-Díaz, "Incorporating the Knowledge of Dermatologists to Convolutional Neural Networks for the Diagnosis of Skin Lesions", *ArXiv e-prints*, 2017].

[18] Afonso Menegola, Michel Fornaciali, Ramon Pires, Sandra Avila, Eduardo Valle, "Towards Automated Melanoma Screening: Exploring Transfer Learning Schemes", *ArXiv e-prints*, 2016].

[19] Afonso Menegola, Michel Fornaciali, Ramon Pires, Flávia Vasques Bittencourt, Sandra Avila, Eduardo Valle, "Knowledge Transfer for Melanoma Screening with Deep Learning", *ArXiv e-prints*, 2017].

[20] Afonso Menegola, Julia Tavares, Michel Fornaciali, Lin Tzy Li, Sandra Avila, Eduardo Valle, "RECOD Titans at ISIC Challenge 2017", *ArXiv e-prints*, 2017].

[21] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", *Nature,* vol. 542(7639), pp. 115-118, Feb 02, 2017.

[22] Tri-Cong Pham, Chi-Mai Luong, Muriel Visani, Van-Dung Hoang, "Deep CNN and Data Augmentation for Skin Lesion Classification," in Asian Conference on Intelligent Information and Database Systems, pp. 573-582, 2018.

[23] Van-Dung Hoang, My-Ha Le, Truc Thanh Tran, Van-Huy Pham, "Improving Traffic Signs Recognition Based Region Proposal and Deep Neural Networks," in Asian Conference on Intelligent Information and Database Systems, pp. 604-613, 2018.

[24] Ajmal Mian, "Illumination invariant recognition and 3D reconstruction of faces using desktop optics", *Optics express,* vol. 19(8), pp. 7491-7506, 2011.

[25] Johan Stephen Simeon Ballot, "Face recognition using Hidden Markov Models," Stellenbosch: University of Stellenbosch, 2005.

Van-Huy Pham received the Ph.D in Computer Science from Ulsan University, South Korea, in 2015, and M.S. degree in Computer Science from University of Sciences, Ho Chi Minh City, Vietnam in 2007. Since 2015, he has been a lecturer and researcher at Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam. His main research interests include artificial intelligence, image processing, computer vision.

Diem-Phuc Tran is a Ph.D student in Computer Science from Duy Tan University, Da Nang City, Vietnam. Now, he is working in Information and Communications department of Quang Binh province. His main research interests include artificial intelligence, image processing, computer vision, auto vehicle.

Van-Dung Hoang received his Ph.D. degree from University of Ulsan, Korea, in 2015, and , and M.S. degree in Computer Science from Hanoi National University of Education, Vietnam in 2007. He has been serving as a senior lecturer in Quang Binh University, Vietnam, since 2002. His research interests covers in a wide area, which includes pattern recognition, machine learning, computer vision, medical image processing, vision based robotics and intelligent systems, comminication network.