Feature Extraction Efficient for Face Verification Based on Residual Network Architecture

Thananchai Khamket[1] and Olarik Surinta[2](✉)

[1] Applied Informatics Group, Department of Information Technology, Faculty of Informatics,
Mahasarakham University, Mahasarakham, Thailand
thananchai.k@msu.ac.th

[2] Multi-Agent Intelligent Simulation Laboratory (MISL), Department of Information
Technology, Faculty of Informatics, Mahasarakham University, Mahasarakham, Thailand
olarik.s@msu.ac.th

Abstract. Face verification systems have many challenges to address because human images are obtained in extensively variable conditions and in unconstrained environments. Problem occurs when capturing the human face in low light conditions, at low resolution, when occlusions are present, and even different orientations. This paper proposes a face verification system that combines the convolutional neural network and max-margin object detection called MMOD + CNN, for robust face detection and a residual network with 50 layers called ResNet-50 architecture to extract the deep feature from face images. First, we experimented with the face detection method on two face databases, LFW and BioID, to detect human faces from an unconstrained environment. We obtained face detection accuracy $> 99.5\%$ on the LFW and BioID databases. For deep feature extraction, we used the ResNet-50 architecture to extract 2,048 deep features from the human face. Second, we compared the query face image with the face images from the database using the cosine similarity function. Only similarity values higher than 0.85 were considered. Finally, the top-1 accuracy was used to evaluate the face verification. We achieved an accuracy of 100% and 99.46% on IMM frontal face and IMM face databases, respectively.

Keyword: Face verification · Face detection · Facial landmarks · Deep feature · Local descriptor

1 Introduction

Face verification is a sub-system of face recognition systems that can detect, extract features, and verify human identity from frontal faces [1]. Many algorithms and applications had been proposed to address face recognition systems. However, to achieve accurate face recognition many technical gaps need to be improved, such as insufficient training data, low quality of image source, and low light conditions.

Due to the computation time, various well-known computer vision techniques were proposed to challenge face recognition systems, such as scale-invariant feature transform (SIFT) [2] and histogram of oriented gradients (HOG) [3]. These techniques could

P. Chomphuwiset et al. (Eds.): MIWAI 2021, LNAI 12832, pp. 71–80, 2021.
https://doi.org/10.1007/978-3-030-80253-0_7

perform fast computation while training, and test without a graphics processing unit (GPU). Because of the high accuracy of measured face recognition, a deep learning technique [4] was proposed, called convolutional neural networks (CNNs). However, it requires a GPU while training.

A face verification system framework can be categorized into three sections; face detection or facial landmark localization, feature extraction or deep feature, and face verification [5–7].

In particular, a face verification system generally follows a common approach, and different solutions have been proposed for each step of it. These steps can be summarized as:

- **Face Detection**. In this step, face detection and facial landmark localization are applied to find human faces from the various devices, such as CCTV, web camera, video capture, and camera. Face detection mainly deals with finding the whole human face from the image and video. The facial landmark localization is defined as the localization of specific key points on the frontal face, such as eye contours, eyebrow contours, nose, mouth corners, lip, and chin [5, 7].
- **Feature Extraction Techniques**. The local descriptors, such as SIFT and HOG, are applied to extract the robust features (namely handcrafted features) from the face and key points detected as described in the first section. Instead of handcrafted features, CNN architectures have recently been applied to learn from the training images and then create the feature vector [8], called a deep feature. This method creates robust deep features from the whole face.
- **Similarity and Verification**. In this step, first, the similarity function is employed to find the similarity value between query faces and faces from the database. Second, a threshold value is used to verify if two faces are similar or not.

In this paper, we focus on enhancing the face verification system by proposing face detection and feature extraction methods based on convolutional neural networks (CNNs). As for face detection, the max-margin object detection (MMOD) method was designed to detect faces over the pyramid images. To define if that object is the face, we sent the object to the simple CNN model to extract the deep features and classify it as a face or not. The MMOD + CNN method discovered faces (even small faces) in unconstrained environment images. For face verification, we then proposed the ResNet-50 model to extract the deep feature from the specific face location that was extracted using the MMOD + CNN method. The face verification system can work in various face databases.

Paper Outline. The remainder of the paper is structured as follows: Sect. 2 explains and reviews related work in face verification. In Sect. 3, the proposed face verification system is described in detail. In Sect. 4, the experimental results, face databases, evaluation, and discussion are presented. In Sect. 5, the conclusion and suggestions for future work are given.

2 Related Work

Some approaches to face verification have focused on using computer vision techniques to generate the handcrafted feature. Recently, deep learning techniques have been used to train and create a deep feature. Furthermore, the facial landmarks localization technique is proposed to find the specific frontal face locations. To learn the achieving face verification systems, we briefly explain the research related to computer vision techniques, deep learning, and facial landmarks localization techniques.

Khunthi et al. [9] proposed a face verification system. Their system included two steps; face detection and face encoding. The face detection step found that the histogram of oriented gradients combined with the support vector machine (HOG + SVM) obtained an accuracy of 99.60% on the BioID dataset. In the face encoding step, the ResNet-50 architecture was used to extract the feature vector. The ResNet-50 architecture achieved 100% accuracy on the BioID and FERET datasets. It also provided a high accuracy of 99.60% on the ColorFERET dataset.

For the facial landmarks localization techniques, Kazemi and Sullivan [10] proposed a novel technique to estimate the location of facial landmarks using gradient boosting for learning an ensemble of regression trees. This technique could detect 194 facial landmarks in a millisecond. Khan [6] proposed a framework that detected only 49 facial landmarks from eyes (12 marks), eyebrows (10 marks), nose (9 marks), and lips (18 marks). Furthermore, Amato et al. [7] compared the effectiveness between facial landmarks features and deep feature or feature extraction for verifying faces. For facial landmarks features, it returned 68 key points located on the face. Then, the distance values between the 68 key points and the center key point were computed (68-point feature) and used as the feature vector. In addition, the 68 distance values were divided by the maximum distance value [8], called the pairs feature. The experimental results showed that the pairs feature provided better results than the 68-point feature. For deep feature, they used the VGGFace network to extract the deep feature. As a result, the deep feature outperformed the facial landmarks features.

For the deep feature, Taigman et al. [11] proposed a convolutional neural network to learn from a very large-scale labeled face dataset collected online. In their training CNN network, the 3D-aligned face with three channels, including red, green, and blue, were used as the input of the network. This network returned the deep feature of 4,096 dimensions. Subsequently, Srisuk and Ongkittikul [12] invented face recognition with weighted CNN architecture. In this method, the face components were extracted from the face image. The face components were then used as the input of the CNN network to extract the deep feature that represents the whole face. In addition, Parkhi et al. [13] introduced the VGGFace network based on the VGG network. The output of the first two fully connected (FC) layers was 4,096 dimensions. The last FC layer created 2,622 or 1,024 dimensions that depended on the loss functions.

Furthermore, Hof et al. [14] designed a device that included a mmWave radar sensor to capture human faces. Firstly, the mmWave radar sensor captured the energy reflected from the face at each distance from 8, 16, and 24 cm. Secondly, 6,000 real numbers were used as the input of the autoencoder network. Stochastic gradient descent (SGD) was used when training the network to reduce the reconstruction error. Finally, the mean square error (MSE) was proposed to find the reconstruction error between input and output faces reconstructed from the autoencoder network. The experimental results showed a correlation between different captures of the same face when tested on the 200 faces of different people.

As can be seen from the above, both well-known handcrafted and the deep feature have been accepted by researchers. We will present our proposed face verification system in the following section.

3 The Proposed Face Verification System

This paper presents a framework for accurate detection and verification of the human face. The proposed framework is shown in Fig. 1 and described in the following section.

3.1 Ace Detection Using MMOD + CNN

This section provides the concepts of max-margin objection detection (MMOD) combined with a convolutional neural network (CNN) that were applied to face detection.

King [15] proposed max-margin object detection to detect objects in images. In this method, to deal with small faces, small sliding windows with 50 × 50 pixels were slid through the image pyramid (see Fig. 2). This method skipped unnecessary sliding windows by considering only sliding windows that had a window scoring value larger than the max-margin value. Therefore, the sliding windows were sent to the simple CNN to extract the deep feature.

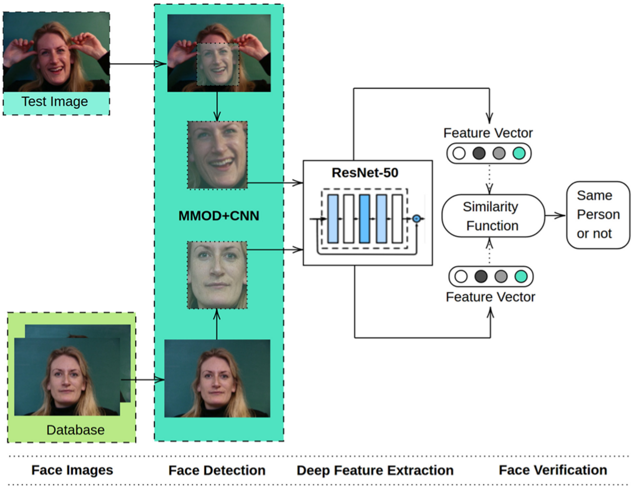


Fig. 1. Framework of the proposed face verification system.

To find more accurate faces, we can adjust more sub-image layers because the image pyramid is a sequence of an image defined as sub-image layers. The image sequence is obtained through the scale of the down-sampling. The bottom sub-image layer is the original image, and the size of the next layer is calculated according to the specific function.

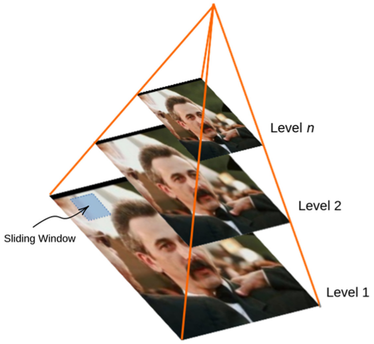


Fig. 2. Illustration of the image pyramid.

3.2 Deep Feature Extraction Using ResNet-50 Architecture

Deep feature extraction is the deep learning method proposed to extract the robust feature with the CNN architectures [16–18]. In this paper, the pre-trained ResNet-50 [19] model was applied to train a set of face images and then extracted the deep features from the layer before the fully connected layer. We then used the deep feature of the query face image to compare with other face images from the databases.

The purpose of the ResNet architecture was quite different from typical CNN architecture designed with a feedforward network without skipping any convolution layers. The ResNet architecture added the shortcut connections that allow skipping the next building block only when the number of feature maps was equal. Then, the output of the current building block was added to the outputs of the stacked layer.

4 Experiments

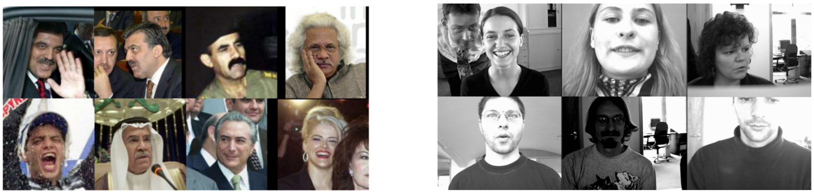
In this section, we report some experiments to evaluate the effectiveness of our approach on several face image databases; LFW, BioID, and IMM. We applied the face detection accuracy value for face detection to evaluate the face detection algorithms described in Eq. 1. For face verification, when verifying the person between the query face and faces in the database, we considered only the similarity value when the similarity value (see Eq. 2) was larger than 0.85. Furthermore, we used only Top-1 accuracy to verify if it was the same person.

4.1 Face Databases

Labeled Faces in the Wild (LFW). Huang et al. [20] provided the LFW face database to study face recognition in unconstrained environments. The LFW dataset consists of 13,233 images of 5,749 persons with the size of 250 × 250 pixels. We used this dataset only to evaluate the face detection algorithm. The LFW face database is shown in Fig. 3(a).

BioID Face Database. This dataset contains 1,521 grayscale images of 23 persons with a pixel resolution of 384 × 288 pixels [21]. We performed both face detection and face verification on the BioID dataset. The BioID face database is shown in Fig. 3(b).

IMM Face Database. In 2003, Stegmann et al. [22] proposed the IMM face database. This database is divided into two datasets; IMM face and IMM frontal face. First, the IMM face dataset contains 240 images of 40 different humans with a pixel resolution of 640 × 480 pixels and is stored in a color image. The humans were not allowed to wear glasses. This dataset aimed to define the shape model with the 58 facial landmarks. Second, the IMM frontal face dataset comprises 120 images of 12 persons. However, we applied the IMM face database to experiment with the proposed face verification system. The IMM face database is shown in Fig. 4.



(a) (b)

Fig. 3. Sample images from (a) LFW and (b) BioID datasets.

(a) (b)

Fig. 4. Example images from IMM face database. (a) IMM face and (b) IMM frontal face.

4.2 Evaluation Metrics

Face Detection Accuracy. The accuracy of face detection depends on two factors; detected and error faces [9]. The equation can be calculated as:

$$FDA = \frac{(p - n) * 100}{N} \tag{1}$$

where p is the number of positive faces detected after using the face detection method, n is the number of negative faces, and N is the number of face images.

Cosine Similarity for Face Verification. The cosine similarity measurement ($cos(\theta)$) [23, 24] between the feature vector of face a and face b can be defined as:

$$cos(\theta) = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}} \tag{2}$$

where a_i and b_i are elements of vector a and b. The meaning of the similarity value of one is the same vector, while zero is the opposite.

Top-1 Accuracy. The result must be precisely the correct answer. In our study, the actual class matches with the highest similarity value calculated by the similarity function.

4.3 Evaluation

In order to evaluate the robustness of the face verification system, we divided the evaluation method into two steps; face detection and face verification.

For face detection, we made comparisons with three well-known face detection techniques; MMOD + CNN, Haar-Cascade, and HOG + SVM. The comparative results are shown in Table 1.

Table 1. Performance comparisons of face detection algorithms on LFW and BioID databases.

Face detection algorithms	Face detection accuracy (%) on face databases	
	LFW	BioID
Haar-Cascade	93.05	93.29
HOG + SVM	99.43	98.88
MMOD + CNN	**99.91**	**99.54**

From Table 1, the experimental results showed that the MMOD + CNN technique performed slightly better than the HOG + SVM method on both LFW and BioID face databases. A high accuracy above 99.5% was obtained from the MMOD + CNN method. Therefore, it can be seen that both the MMOD + CNN and HOG + SVM methods yielded an accuracy above 99% on the LFW face database, which was designed to detect faces in unconstrained environments.

To evaluate the face verification, we experimented with the effect of feature extraction techniques. Three feature extraction techniques; SIFT, HOG, and ResNet-50, were performed. Additionally, we extracted robust features, including SIFT and HOG features from 68-landmark localization (LL) and dense grid area (DG). The ResNet-50 extracted deep features from the face area. The experimental results are shown in Table 2.

We examined the face verification techniques on the IMM face databases. The IMM face databases were divided into training and test sets with a ratio of 60:40. Each of these face databases were split into two more face databases; IMM frontal face and IMM face. The IMM frontal face database contained only the frontal face images, while the IMM face database contained various face orientations and different light conditions.

Table 2 shows that ResNet-50 extracted robust 2,048 deep features and outperformed other local descriptor methods (SIFT and HOG). The ResNet-50 achieved 100% on the IMM frontal face database and 99.46% on the IMM face database. Consequently, the experimental result obtained an accuracy of 100% on the IMM frontal face database when extracting the local features from the 68-landmarks. Also, while extracting the local features with HOG and SIFT methods from the dense grid area obtained an accuracy of 100% and 96.2%, respectively. The high accuracy was achieved because the IMM frontal face contained only frontal faces. The results showed that extracting the local features with HOG and SIFT methods from the dense grid area performed better than extracting the local descriptors from the landmark localization on the IMM face database.

Table 2. Performance comparison of face verification with different feature extraction methods on IMM face database.

Feature extraction methods	Number of features	IMM frontal face	IMM face
LL + SIFT	8,704	100	90.81 ± 0.17
LL + HOG	544	100	94.35 ± 0.11
DG + SIFT	14,080	96.2 ± 0.23	95.79 ± 0.21
DG + HOG	880	100	98.82 ± 0.15
ResNet-50	2,048	100	99.46 ± 0.51

4.4 Discussion

In this section, we discuss the selection of the face verification system. Two main processes (face detection and face verification) are considered.

The MMOD + CNN face detection method can detect a small face because the small sliding windows with 50 × 50 pixels were sliding through the pyramid of images. Also, the max-margin value was used to consider the window scoring value from the overlap windows and rejected that window if it had a low score. Subsequently, we sent each window to a simple CNN architecture to create robust deep feature. It is quite fast because it created the deep feature from the window with high scoring.

A ResNet-50 architecture can be performed to extract the robust deep feature even when applied to extract from the unconstrained face, such as the face in different orientations, emotions, and light conditions (see Fig. 3(a)). It provided a high similarity value when comparing the query face and faces in the database were compared. By extracting robust deep feature with the ResNet-50 architecture, we can excluded face landmark localization and dense grid processes. We then selected the ResNet-50 architecture and set it into the face verification process.

5 Conclusion

In this research, we proposed a face verification system, including face detection and face verification. In face detection, we proposed to use the convolutional neural network and max-margin object detection, namely the MMOD + CNN method, to detect faces in unconstrained environments. It can detect a normal face, a small face, and even a part of a face. The results show that the MMOD + CNN method provided a high detection accuracy of more than 99.5% on the LFW and BioID databases. We then assigned the detected faces to the ResNet-50 model, the convolutional neural network (CNN) architecture, to extract the 2,048 deep features. We evaluated the ResNet-50 on the IMM face databases, including IMM frontal face and IMM face. The experimental results showed that the ResNet-50 model obtained 100% accuracy on the IMM frontal face database and 99.46% on the IMM face database. Additionally, the local descriptors (histogram of oriented gradients: HOG and scale-invariant feature transform: SIFT) that extracted the local features from the dense grid method outperformed the local descriptors that extracted the features from the face landmark localization method.

Acknowledgements. This work has been funded by the Faculty of Informatics, Mahasarakham University, Thailand [grant number it2-05/2560].

References

1. Bah, S.M., Ming, F.: An improved face recognition algorithm and its application in attendance management system. Array **5**, 100014 (2020). https://doi.org/10.1016/j.array.2019.100014
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). https://doi.org/10.1023/b:visi.0000029664.99615.94
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2005). https://doi.org/10.1109/cvpr.2005.177
4. Prana, K.B., Manikandan, J.: Design and evaluation of a real-time face recognition system using convolutional neural networks. Procedia Comput. Sci. **171**, 1651–1659 (2020). https://doi.org/10.1016/j.procs.2020.04.177
5. Ouanan, H., Ouanan, M., Aksasse, B.: Facial landmark localization: past, present and future. In: 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), pp. 487–492 (2016). https://doi.org/10.1109/cist.2016.7805097
6. Khan, F.: Facial expression recognition using facial landmark detection and feature extraction via neural networks, pp. 1–7. arXiv https://arxiv.org/abs/1812.04510 (2018)
7. Amato, G., Falchi, F., Gennaro, C., Vairo, C.: A comparison of face verification with facial landmarks and deep features. In: 8th International Conference on Advances in Multimedia (MMEDIA), pp. 1–6 (2018)
8. Rassadin, A., Gruzdev, A., Savchenko, A.: Group-level emotion recognition using transfer learning from face identification. In: 19th ACM International Conference on Multimodal Interaction (ICMI), pp. 544–548 (2017). https://doi.org/10.1145/3136755.3143007
9. Khunthi, S., Saichua, P., Surinta, O.: Effective face verification systems based on the histogram of oriented gradients and deep learning techniques. In: 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), pp. 215–219 (2019). https://doi.org/10.1109/isai-nlp48611.2019.9045237

10. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1867–1874 (2014). https://doi.org/10.1109/cvpr.2014.241
11. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701–1708 (2014). https://doi.org/10.1109/cvpr.2014.220
12. Srisuk, S., Ongkittikul, S.: Robust face recognition based on weighted DeepFace. In: International Electrical Engineering Congress (iEECON), pp. 1–4 (2017). https://doi.org/10.1109/ieecon.2017.8075885
13. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: The British Machine Vision Conference, pp. 1–12 (2015). https://doi.org/10.5244/c.29.41
14. Hof, E., Sanderovich, A., Salama, M., Hemo, E.: Face verification using mmWave radar sensor. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 320–324 (2020). https://doi.org/10.1109/icaiic48513.2020.9065010
15. King, D.E.: Max-margin object detection, pp. 1–8. arXiv https://arxiv.org/abs/1502.00046 (2015)
16. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. **54**(10), 6232–6251 (2016). https://doi.org/10.1109/tgrs.2016.2584107
17. Mahmood, A., Bennamoun, M., An, S., Sohel, F.: Resfeats: residual network based features for image classification. In: IEEE International Conference on Image Processing (ICIP), pp. 1597–1601 (2017). https://doi.org/10.1109/icip.2017.8296551
18. Özyurt, F.: Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. J. Supercomput. **76**(11), 8413–8431 (2019). https://doi.org/10.1007/s11227-019-03106-y
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/cvpr.2016.90
20. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, pp. 1–14 (2008)
21. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust face detection using the Hausdorff distance. In: Audio- and Video-Based Biometric Person Authentication, pp. 90–95 (2001). https://doi.org/10.1007/3-540-45344-x_14
22. Stegmann, M.B., Ersbøll, B.K., Larsen, R.: FAME – a Flexible appearance modelling environment. IEEE Trans. on Med. Imaging **22**(10), 1319–1331 (2003). https://doi.org/10.1109/TMI.2003.817780
23. Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010. LNCS, vol. 6493, pp. 709–720. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19309-5_55
24. Zheng, L., Idrissi, K., Garcia, C., Duffner, S., Baskurt, A.: Triangular similarity metric learning for face verification. In: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–7 (2015). https://doi.org/10.1109/fg.2015.7163085