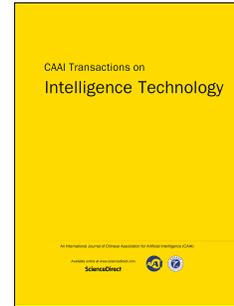


Accepted Manuscript

Face recognition using both visible light image and near-infrared image and a deep network

Kai Guo, Shuai Wu, Yong Xu



PII: S2468-2322(17)30014-8

DOI: [10.1016/j.trit.2017.03.001](https://doi.org/10.1016/j.trit.2017.03.001)

Reference: TRIT 41

To appear in: *CAAI Transactions on Intelligence Technology*

Received Date: 30 January 2017

Accepted Date: 28 March 2017

Please cite this article as: K. Guo, S. Wu, Y. Xu, Face recognition using both visible light image and near-infrared image and a deep network, *CAAI Transactions on Intelligence Technology* (2017), doi: 10.1016/j.trit.2017.03.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Face Recognition Using Both Visible light Image and Near-infrared image and a Deep Network

Kai Guo, Shuai Wu, Yong Xu
 School of Computer Science and Technology
 Harbin Institute of Technology Shenzhen Graduate School
 Shenzhen, China
 guokaiwork@yeah.net

Abstract—In recent years, deep networks have achieved outstanding performance in computer vision, especially in the field of face recognition. In terms of the performance for a face recognition model based on deep network, there are two main closely related factors: 1) the structure of the deep neural network, and 2) the number and quality of training data. In real applications, illumination change is one of the most important factors that significantly affect the performance of face recognition algorithms. As for deep network models, only if there is sufficient training data that has various illumination intensity could they achieve expected performance. However, such kind of training data is hard to collect in the real world. In this paper, focusing on the illumination change challenge, we propose a deep network model which takes both visible light image and near-infrared image into account to perform face recognition. Near-infrared image, as we know, is much less sensitive to illuminations. Visible light face image contains abundant texture information which is very useful for face recognition. Thus, we design an adaptive score fusion strategy which hardly has information loss and the nearest neighbor algorithm to conduct the final classification. The experimental results demonstrate that the model is very effective in real-world scenarios and perform much better in terms of illumination change than other state-of-the-art models. The code and resources of this paper are available at <http://www.yongxu.org/lunwen.html>.

Index Terms—Deep Network, Face recognition, Illumination change, Insufficient training data

I. INTRODUCTION

Biometrics is one of the most important branches of pattern recognition [1-3]. Face recognition is one of the most attractive biometric techniques. Nevertheless, face recognition in real applications is still a challenging task [4]. The main reason is that the face is a non-rigid object, and it often has different appearance owing to various facial expression, different ages, different angles and more importantly, different illumination intensity. In recent years, deep learning has become more and more prevalent in computer vision. AlexNet [5] which is designed by Alex Krizhevsky got the champion of ILSVRC-2012 competition and outperformed the second place nearly 10 percents.

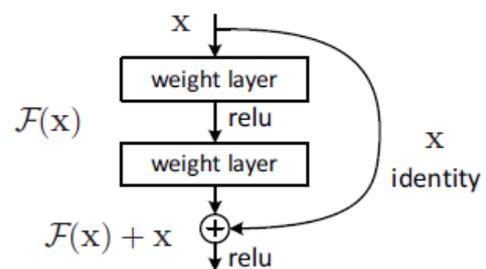


Figure 1: shortcut structure

From then on, researchers realized the powerful analysis ability of deep convolutional network. Simonyan [6] proposed a very deep convolutional network, which is called VGG net, and get the second place of ILSVRC-2014 competition. VGGnet could be divided into several blocks and each block contains several convolutional layers which have identical kernel size and channel number. GoogleNet [7] is specially designed for the ILSVRC-2014 competition by google team. It applies a structure called inception to preserve locally compact connection, while the global structure is sparse. GoogleNet won the first place of the ILSVRC-2014 competition whose top-5 error rate is less than 7%. VGGnet and GoogleNet are very deep and have 19 and 22 layers respectively. On account of the great success of VGG and GoogleNet, researchers began to apply different methods to increase the depth of network. However, a degradation problem is exposed: when the network depth increases, accuracy gets saturated and then degrades rapidly. Such degradation, unexpectedly, is not related with overfitting. In order to solve such problem, He [8] designs a shortcut structure which is illustrated in Figure 1. This structure combines the input x and $F(x)$ as the final output, so $F(x)$ is considered as the residual. The network piled up by this structure is called residual network. Residual network could have the depth of 152 layers and the top-5 error is less than 4% on imagenet database.

Face recognition could be considered as a special classification task and the deep network is pretty suitable for face recognition. Deep neural networks have powerful feature extraction ability, and can obtain competitive extractor by using massive training sets. For some public face data sets, such as Labeled Faces in the Wild Home (LFW) [9], the accuracy of deep networks can even reach 99.8%. However, face

recognition is far from perfect in engineering applications. For the LFW database, the face image is relatively easier for recognizing, it only contains part of the scenes, postures. However, in real word applications, as we pointed out earlier, the condition is more complicate, the appearance of face images may suffer from various facial expressions, different angles and varying illuminations. Compared with other factors, illumination change has the greatest influence on face recognition. Figure 2 shows that face images obtained under different environmental illuminations are much different. These differences eventually lead to a negative impact on the face recognition system.



Figure 2: Face images captured by the visible light camera under different illuminations.

For deep learning algorithms in engineering applications, massive training data is necessary. Baidu researchers proposed to demonstrate the importance of enough training data for deep networks [10]. They used mobile phone cameras to collect a number of face images, and used a trained deep learning model to test the verification task, reaching an accuracy of 85% when the false alarm rate is 0.0001. After adding new training data (Chinese celebrities collected from web sites) to train the model they achieved an accuracy of 92.5% when the false alarm rate is still 0.0001. This example tells us that more training data is necessary and helpful for deep networks and can lead to better performance.

If we want to solve the illumination change problem using deep networks, there must be enormous training data that have various illumination intensity. However, it is very tough to collect such kind of dataset in real word. In this paper, we design a deep network system using both visible light image and near-infrared image. Besides conventional recognition tasks of visible light face images, recognition of near-infrared face images also attracts the attention of many researchers. As we know, near-infrared face images are usually less sensitive to illumination change. Compared to near-infrared images, visible light face images could reflect more details of the faces. Therefore, we apply both visible light image and near-infrared image in order to solve the illumination change problem and at the same time, preserve the advantage of visible light images. First of all, we use public visible light face data resources from the internet to train a deep network model, which is referred to as the first model. Then we use a number of near-infrared face images to re-train the obtained deep network model. After re-training is completed, we use the ultimate deep network model as feature extractor of near-infrared face images and refer to it as the second model. After that, we exploit the cosine distance between the test sample and training samples to get the class score and apply an adaptive score fusion strategy and

nearest neighbor algorithm to conduct the final classification.

Here, we need to point out that we use two different devices to obtain face images: a visible light camera and a near-infrared camera. Therefore, our system can simultaneously collect two kinds of face images, near-infrared face images and visible light face images. The first model and second model are respectively applied to visible light face images and near-infrared face images to extract features. Then the cosine distance will be calculated to get the class score. Here the score could be considered as the correlation intensity that between the test sample and training samples. Because these two kinds of images are respectively beneficial to capture light-invariant features and texture features of face, we finally combine the score to conduct the final classification. The score fusion strategy combines the two separate models into a union one. According to the result of the experimental analysis, our model has a good performance in practical application scenes.

II. RELATED WORKS

Face recognition has been a prevalent research for many years. There are many classical algorithms for face recognition. Here we can simply classify these methods into three categories: 1) Local feature method, subspace-based methods and more sparse representation methods. Local feature methods are mainly proposed to solve varying facial expression problem. As we know, the face is a non-rigid object, and the change of facial expressions and other factors will lead to changes in facial features. But researchers have found that some local features of face images do not change severely. Therefore, local features have been exploited for face recognition. For example, Gabor [11], LBP [12], SIFT [13] features all show promising performance in face recognition. 2) The basic idea of subspace-based methods is to exploit a transform factor to map the high dimensional face image to a low-dimensional feature space. PCA [14] is one of the most classical subspace-based methods. On the basis of PCA, 2DPCA [15] is proposed to improve the robustness of PCA. LPP (Local Preserve Projection) [16] is an unsupervised method and could make the feature vector preserve the relationship of original neighbors. LDA [17] applies Fisher discriminate function to guarantee that in the feature space, samples within a class will have close relationship and different classes have weak correlation. So the property of the feature space is dependent on the methods that formulate the transform factor. The advantages of the subspace-based algorithms are that on one hand they can remove redundant information of the original images, and on the other hand they can reduce the time complexity because the dimension of feature space is much less than that of the original image space. 3) Sparse representation has attracted much attention of researchers in the field of signal processing, image processing, computer vision, and pattern recognition. Before deep networks become famous, sparse representation has been considered as one of the most effective methods for face recognition. SRC (sparse representation based classification) [18] first assumes that the test sample can be sufficiently represented by the training samples. Specifically, SRC exploits the linear combination of training samples to represent the test sample and

computes sparse representation coefficients of the linear representation system, and then calculates the reconstruction residuals of each class employing the sparse representation coefficients and training samples. The test sample will be classified as a member of the class which leads to the minimum reconstruction residual. Xu [19] proposed a two-step sparse representation based classification method. Xu firstly chooses the k nearest neighbors in the first step, then uses these neighbors to represent the test sample in the second step. Deng [20] proposed an extended sparse representation method (ESRM) for improving the robustness of SRC by eliminating the variations in face recognition, such as disguise, occlusion, and expression. The quintessence of sparse representation methods is the sparsity. It means most coefficients of the training samples are 0, which is advantageous for preserving useful training samples.

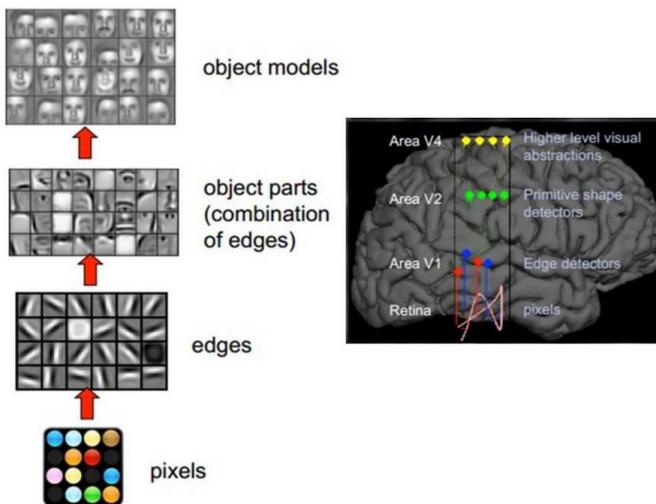


Figure 3: Visual mechanism of human brain.

Compared with the conventional face recognition methods, face recognition models based on deep network can always achieve better performance. As we know, deep networks can simulate the human brain's thinking process. The memory and conception for objects in our brain is not stored in a single place, instead, it is distributed stored in a vast network of neurons. This is consistent with the deep network whose forward calculation is the process of abstraction. Figure 3 illustrates the comparison of face recognition based on deep networks and our brain. Here we need to point out that the widely used deep learning method often adopts the weight-sharing network structure, so it can greatly reduce the complexity of the network architecture and the number of weights.

At present, some academic and commercial institutions have designed different deep networks for face recognition, such as FaceNet [21] (Google), VGGNet [22] (oxford research group),

Table 1: Results with weak illumination change

Different methods	Number of registered samples		
	3	5	7
PCA	0.8799	0.8882	0.8907

LPP	0.9227	0.9267	0.9321
LDA	0.9640	0.9840	0.9862
CNN	0.9901	0.9927	0.9933

Table 2: Results with strong illumination change

Different methods	Number of registered samples		
	3	5	7
PCA	0.6261	0.6197	0.6157
LPP	0.6480	0.6515	0.6675
LDA	0.7644	0.8060	0.8416
CNN	0.8005	0.8035	0.8104

DeepFace [23] (Facebook) and DeepID [24] (CUHK group). FaceNet exploits very deep networks to perform face recognition. It uses nearly 8 million images of 2 million people and applies the triple loss strategy to train the network. DeepFace model applies a network trained by 4 million images. Here we need to point out that face recognition in DeepFace are a two-step process. It firstly exploits deep network to extract the face feature and then performs classification. Moreover, DeepFace applies an integrated neural network to conduct face alignment in the preprocessing stage. DeepID is on the basis of DeepFace, and it uses plenty of patches of face images to train different deep networks and finally combines the output of these networks as the feature of face image. In the classification, DeepID applies Joint Bayesian classifier in order to make the classification more robust.

However, these models heavily depend on the quality and quantity of the training data. For illumination change problems, as we pointed out earlier, there is not enough training data with varying illuminations intensity. So the deep network model will not perform well. Table 1 illustrates the performance of different methods with weak illumination change, while Table 2 illustrates that with strong illumination change. Comparing these two tables, we can conclude that both conventional methods and deep network are not competent for varying illuminations problem. Therefore, we design a model which could make use of near-infrared images to handle varying illuminations because near-infrared images are less sensitive to illumination change.

III. NETWORK ARCHITECTURE AND MODEL TRAINING

A. Network architecture

Table 3 shows the reported performance of different models on the LWF database. In this paper, we choose VGG Net [22] as the deep learning network architecture. It is because that in the engineering applications, VGG Net can achieve reasonable balance between the accuracy and time efficiency. DeepID applies a multi-model feature fusion algorithm, which uses 200 deep network models to extract features of a face image. As we know, the forward processes of 200 deep networks are very time consuming because there are enormous convolution operations, so DeepID is really not suitable for engineering applications and

we do not choose it. The complete network architecture of VGG is shown in Figure 6.

Table 3: Accuracy of different deep model

Deep model	Net	ACC
Fisher Vector	-	93.1
DeepFace	3	97.35
DeepID	200	99.47
FaceNet	1	98.87
VGG_Face	1	98.95

We can see that except for the fully connected layer, there are 5 blocks in the VGG Net model. Each block consists of several convolutional layers followed by additional nonlinear operations, such as ReLU operation and local response normalization. The kernel size and channel number in one block are identical. The last layer of each block is always a pooling layer. So the feature map size of convolutional layers in one block is smaller than that of the previous block. Focusing on the first four blocks, we can find that the channel number is doubled after a pooling layer. More channels will effectively compensate for the information loss caused by the pooling layer. After 5 blocks, there are 3 fully connected layers. The convolution operation is mainly used to extract local features of face images, while the full join operation can extract global features of the whole face image. Moreover, the main purpose of using the ReLU operation is to eliminate overfitting and perform the local normalization operation on the convoluted features. Network-specific parameters are shown in Table 4.

B. Training and Databases

The training process is composed of two phrases that result in two separate models to handle different kinds of face images. Firstly, we apply CelebFaces Attributes Dataset (CelebA) of the Chinese University of Hong Kong to train the VGG net and refer to the training results as the first model. Next, we preserve the parameter of the first model and fine-tune the model using the CASIA NIR dataset and PolyU NIR Face Database. The training result is referred to as the second model. The CelebA dataset [25] contains 202599 face images captured from 10177 identities, and contains rich posture and background variations, Figure 4 illustrates some examples of CelebA dataset. The first model is aimed at extracting features from visible light image. The CASIA NIR database contains 2490 NIR face images of 197 peoples and the PolyU NIR Face Database includes 35000 NIR Face images of 350 peoples. More information is shown in [26-27]. The second model is effective in handling near-infrared images owing to the fine-tuning step.

In order to verify whether the proposed face recognition method which exploits both visible light image and near-infrared image is effective to illumination changing problems, we need a set of visible light and near-infrared face images to perform verification experiments. At present, most public databases contain only visible light or near-infrared face images, and only a few public databases contain both of them. This paper uses the HIT LAB2 face dataset [28] and SunWin Face database to verify our method. These two databases both

contain near-infrared and visible light face images under changeable conditions as Figure 5 shows.



Figure 4: Examples of CelebA database

Brief introduction of the SunWin Face and HIT LAB2 data sets are as follows:

The SunWin Face database contains 4000 face images from 100 identities. It has two parts: 1) 2000 visible light pictures from the 100 identities. For each person, 10 pictures are collected under normal light, the other 10 pictures per person are captured under abnormal light. 2) 2000 near-infrared pictures from the 100 identities. For each person, 10 pictures are also obtained under normal light and the other 10 pictures are captured under abnormal light. The collected database contains different facial expressions, lights and other changes. A visible light camera and a near-infrared camera were used to collect data at the same time.

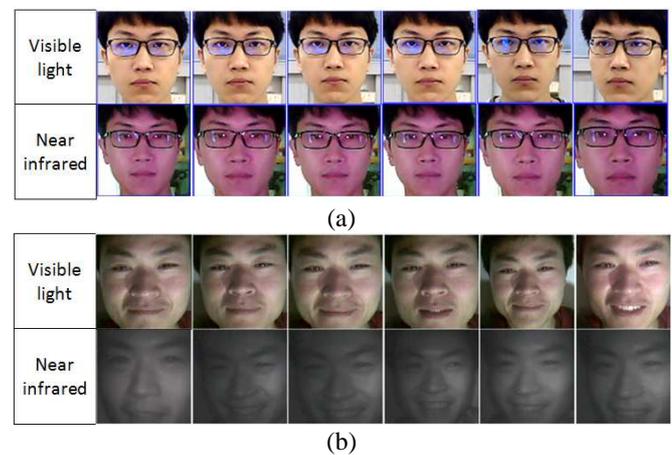


Figure 5: Examples of (a) Sun Win (b) HIT LAB2

The HITSZ Lab2 dataset was collected and issued by Harbin Institute of Technology. The database contains a total of 2000 face images from 50 volunteers. The image size is 200X200. These images were collected under the following different lighting conditions: (a) natural light (b) natural light + left light (c) natural light + right light (d) natural light + left and right side lights. The image also contains significant posture or facial expression changes.

C. Score-fusion

There are three fusion strategies: pixel-based fusion, feature level fusion and score-level-based fusion. Simultaneously using visible light and near-infrared face images can make the extracted face feature more comprehensive. Empirically, for face recognition, the score based fusion strategy is better than the feature based fusion strategy. The reason is that the feature-based fusion will cause information loss. The strategy based on score fusion avoids this defect, so the score fusion strategy can always obtain better experimental results [29]. Therefore, this paper applies the score fusion strategy to conduct the final classification.

In our face recognition system, the first model and the second model will respectively process the visible light image and near-infrared image and extract feature from both images. Then we apply the cosine distance to calculate the score of both features between test sample and training samples, Here the score could be considered as the correlation intensity that between the test sample and training sample. After that, we use the weighted combination strategy to perform score fusion [30], as shown in (1).

$$F_i = \alpha V_i + \beta N_i \quad (1)$$

where F_i denotes the fusion score, V_i denotes the score of the visible image and N_i denotes the score of the near-infrared image V_i and N_i are processed using the normalization algorithm, as shown in (2),

$$Score = \frac{Score - \min}{\max} \quad (2)$$

The normalized processing is necessary because different models of the face image have different score ranges and direct score fusion is meaningless α and β are the weight parameters of V_i and N_i with the following conditions $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$. Here we can just empirically

choose the value of α and β , but the performance will be severely affected if the value is not suitable. So, in this paper, we choose an adaptive way for determining the weights as shown in (3).

$$\alpha = \frac{V_i}{V_i + N_i}, \beta = \frac{N_i}{V_i + N_i} \quad (3)$$

As presented in (3), the adaptive strategy will assign a higher weight to the higher score and a lower weight to the lower one. After score fusion, we apply the nearest neighbor algorithm to perform classification and the test sample is labeled as the corresponding training sample with the maximum find score. In (1), a higher score contains more information and acts as the main component, while the lower one acts as a secondary factor. When N_i is much greater than V_i , the illumination usually changes much. Under this condition, if visible light face image are captured with dark illumination, it is reasonable to regard the features of the visible image as secondary factors, and the near-infrared light feature as the primary factor. Figure 6 shows the score fusion strategy in detail.

IV. EXPERIMENT

Table 5 shows the accuracy of some face recognition models on the LWF and YTF database. The LFW Database contains more than 13,000 face images collected from the internet. Each face has been assigned with a single class label. YouTube Face Database YTF [31] contains 3,425 videos of 1,595 different people. All the videos were downloaded from YouTube website. Each person has 2.15 videos on average. The shortest video has 48 frames, while the longest video has 6,070 frames. The average length of a video is 181.3 frames.

From the experimental results in Table 5 we can see that the face recognition algorithm based on the deep network is

Table 4: The parameter of VGGNet network architecture

Layer	Type	Name	Filt dim	Numfilt
0	input	data	-	-
1	conv	conv11	3	64
2	conv	conv12	64	64
3	pool	pool1	-	-
4	conv	conv21	64	128
5	conv	con22	128	128
6	pool	Pool2	-	-
7	conv	Conv31	128	256
8	conv	Conv32	256	256
9	conv	Conv33	256	256
10	pool	Pool3	-	-
11	conv	Conv41	256	512
12	conv	Conv42	512	512
13	conv	Conv43	512	512
14	pool	Pool4	-	-

15	conv	Conv51	512	512
16	conv	Conv52	512	512
17	conv	Conv53	512	512
18	pool	Pool5	-	-
19	Full_connection	Fc1	512	4096
20	Full connection	Fc2	4096	4096
21	Full connection	Fc3	4096	10177
22	softmax	softmax	-	-

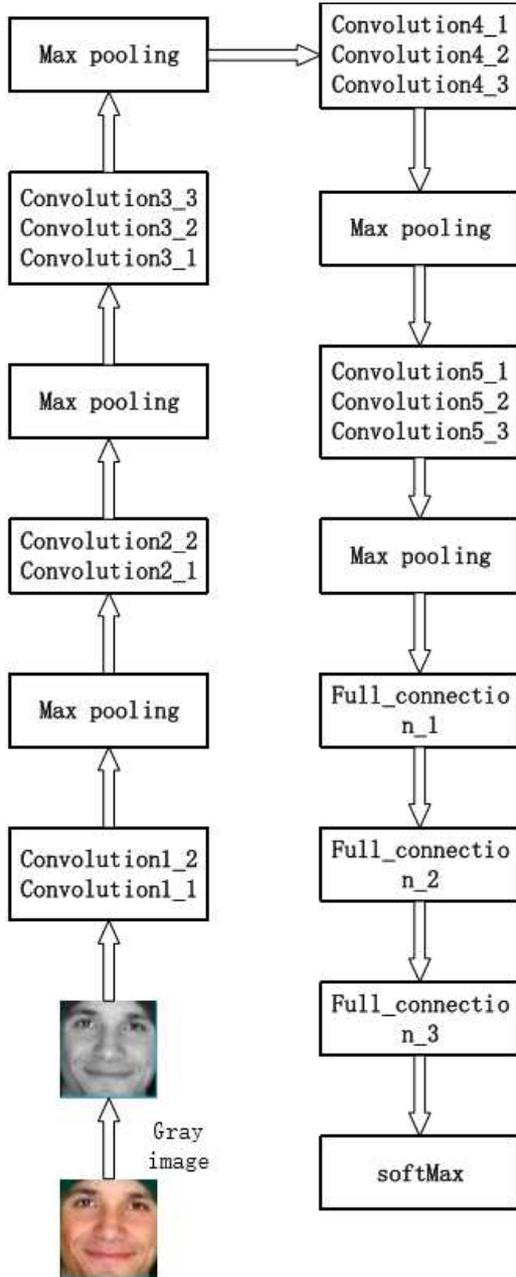


Figure 6: Deep VGG Net structure

significantly better than the traditional methods. On the LFW database, the VGG Net model can reach an accuracy of 98.95%. For some commonly used face recognition algorithms, such as LBP, the accuracy is only 85.17%, far below than that of the

deep network model. For the verification set YTF [31], we can get the similarity conclusion. However, the performance of both Local Binary Pattern (LBP) [32] method and Fisher Vector Faces (FVF) [33-34] method decline greatly comparing to the LFW database. The reason is that the images in the YTF database are taken from the video, which are more complex in attitudes, expressions and other factors.

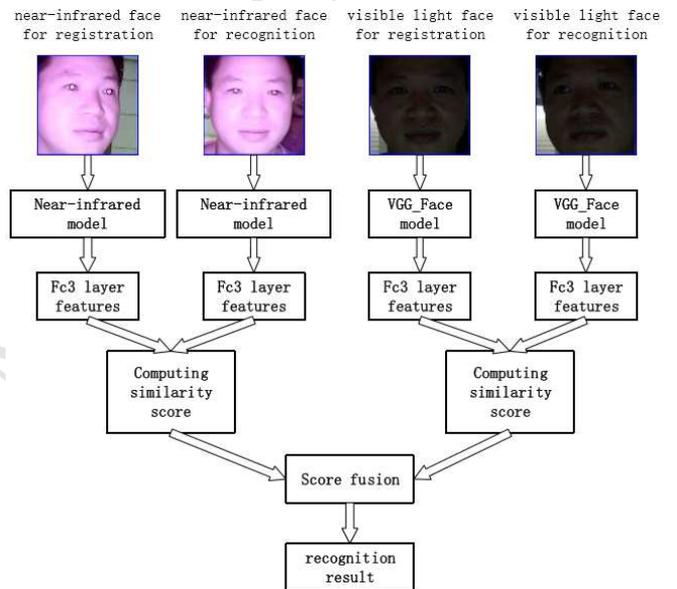


Figure 7: Score fusion process

However, we can see that the accuracy of VGG Net on the YTF data set is only 1% less than the LFW database, this means the robustness of the deep network model to these changes is very powerful. Therefore, we choose VGG Net as our basic model.

However, deep network is not perfect in real word applications. It is sensitive to illumination change because collecting vase training data that has varying illuminations is very difficult. Table 6 shows the comparing results of VGG Net between strong illumination change and weak illumination change. We can see that the accuracy of VGG Net significantly decline if the illumination change is strong on both Sun Win database and HIT database.

In order to solve this problem, we design a model that applies the near-infrared image as the auxiliary image to eliminate the effect of illumination change. We use HIT LAB2 and SunWin Face database to test the performance of our model. In the test phase, five face images of a person were randomly selected as training data, then the VGGNet model was used to extract the features and cosine distance is used as a measure of similarity of

face. The results are shown in Tables 7 and 8.

Table 5: Accuracy on different algorithm

algorithm	LWF	YTW
LBP [32]	98.73	83.89
FVF [33], [34]	97.56	91.31
VGG Net	98.95	97.3

Table 7 is the experimental results of the HIT LAB2 database. We can see clearly that when the lighting change is weak, the face recognition accuracy of VGGNet can reach 98.74%. But when the illumination changes drastically, the accuracy declines to 89.80%. In the case of little change of illumination, the accuracy of our method based on deep learning and score fusion can reach 99.56%. In terms of our model without score fusion, although the accuracy is slightly lower than VGG Net in weak light change condition, our model is significantly higher than VGG Net under the condition of strong illumination change. When we apply score fusion, our model can outperform VGG Net under both illumination conditions.

Table 6: Accuracy of VGG Net

	Weak light change	strong light change
Sun Win	99.01	80.34
HIT LAB2	99.31	89.83

We further test our method via the Sunwin Face database. The experimental results are shown in Table 8. The conclusion is similar to that on HIT LAB2. Our model can achieve a better recognition result for drastic illumination change. Moreover, according to Table 7 and Table 8, we can also demonstrate the significance of score fusion. On both database, our model applying the score fusion could achieve better performance than that without score fusion. Score fusion strategy could make full use of feature from both visible light image and near-infrared image, so it is reasonable for better performance.

Table 7: accuracy on HIT LAB2

	Weak light change	strong light change
VGG Net	98.74	89.80
No score-fusion	97.96	95.13
Score fusion	99.56	95.31

Table 8: Accuracy on Sun Win

	Weak light change	strong light change
VGG Net	99.26	80.34
No score-fusion	98.14	91.62
Score fusion	99.89	93.98

I. CONCLUSION

In this paper, we proposed a CNN-based model which could

apply both visible light image and near-infrared image to perform face recognition. Besides, we also design an adaptive score fusion strategy which is significantly helpful to improve the performance. Compared with the traditional deep learning algorithm, our proposed method can construct a robust face feature extraction model. In practical it is robust to illumination variation. We validate our model via several data sets. The experimental results show that the new model achieves better performance.

ACKNOWLEDGMENT

This study was supported by the Technology Innovation Project of Shenzhen (No.CXZZ20130318162826126). This research was also supported in part by Shenzhen IOT key technology and application systems integration engineering laboratory.

REFERENCES

- [1] C. Feher, Y. Elovici, Robert Moskovitch, Lior Rokach, Alon Schlar, User identity verification via mouse dynamics, *Information Sciences* 201 (2012) 19–36.
- [2] G.F. Lu, Y. Wang, Feature extraction using a fast null space based linear discriminant analysis algorithms, *Information Sciences* 193 (2012) 72–80.
- [3] H.J. Li, J.S. Zhang, Z.T. Zhang, Generating cancelable palmprint templates via coupled nonlinear dynamic filters and multiple orientation palmcodes, *Information Sciences* 180 (2010) 3876–3893.
- [4] Xu Y, Zhu Q, Fan Z, et al. Using the idea of the sparse representation to perform coarse-to-fine face recognition[J]. *Information Sciences*, 2013, 238(7):138-148.
- [5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2):2012.
- [6] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. *Computer Science*, 2014.
- [7] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// *Computer Vision and Pattern Recognition*. IEEE, 2014:1-9.
- [8] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2016:770-778.
- [9] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Month*, 2008.
- [10] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting ultimate accuracy: Face recognition via deep embedding," 2015.
- [11] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 11, no. 4, pp. 467–76, 2002.
- [12] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 19, no. 6, pp. 1635–50, 2010.
- [13] C. Geng and X. Jiang, "Face recognition using sift features," in *IEEE International Conference on Image Processing*, 2009, pp. 3277–3280.
- [14] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews Computational Statistics*, vol. 2, no. 4, p. 433459, 2010.
- [15] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "An approach for directly extracting features from matrix data and its application in face recognition," *Neurocomputing*, vol. 71, no. 10, pp. 1857–1865, 2008.
- [16] He X, Niyogi P. Locality Preserving Projections (LPP)[J]. *Advances in Neural Information Processing Systems*, 2002, 16(1):186-197.
- [17] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," *Chicago*, vol. 3, no. 6, pp. 27–33, 2013.
- [18] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

- [19] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1255–1262, Sep. 2011.
- [20] J. W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.
- [21] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," pp. 815–823, 2015.
- [22] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in British Machine Vision Conference, 2015.
- [23] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
- [24] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," pp. 1891–1898, 2014.
- [25] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," pp. 3730–3738, 2015.
- [26] S. Z. Li, R. F. Chu, S. C. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–39, 2007.
- [27] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to polyu near-infrared facedatabase," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2337–2344, 2010.
- [28] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "Bimodal biometrics based on a representation and recognition approach," *Optical Engineering*, vol. 50, no. 3, pp. 037 202–037 202, 2011.
- [29] A. Ross, A. K. Jain, and J. Z. Qian, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.
- [30] Y. Xu and Y. Lu, "Adaptive weighted fusion: A novel fusion approach for image classification," *Neurocomputing*, vol. 168, pp. 566–574, 2015.
- [31] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," vol. 42, no. 7, pp. 529–534, 2011.
- [32] . Taigman, L. Wolf, and T. Hassner, "Multiple one-shots for utilizing class label information." in British Machine Vision Conference, BMVC 2009, London, UK, September 7-10, 2009. Proceedings, 2009.
- [33] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in British Machine Vision Conference, 2013, pp. 8.1–8.11.
- [34] O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman, "A compact and discriminative face track descriptor," in Computer Vision and Pattern Recognition, 2014, pp. 1693–1700.