Pattern Recognition Letters 82 (2016) 154–161

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

DeepIris: Learning pairwise filter bank for heterogeneous iris verification☆

Nianfeng Liu, Man Zhang*, Haiqing Li, Zhenan Sun, Tieniu Tan

Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Available online 31 October 2015

Keywords:
Biometrics
Iris verification
Convolutional neural networks
Deep learning
Iris recognition

ABSTRACT

Heterogeneous iris recognition (HIR) is in great demand for a large-scale identity management system. Iris images acquired in heterogeneous environment have large intra-class variations, such as different resolutions or different sensor optics, etc. Therefore, it is challenging to manually design a robust encoding filter to face the complex intra-class variations of heterogeneous iris images. This paper proposes a deep learning based framework for heterogeneous iris verification, namely DeepIris, which learns relational features to measure the similarity between pairs of iris images based on convolutional neural networks. DeepIris is a novel solution to iris recognition in two main aspects. (1) DeepIris learns a pairwise filter bank to establish the relationship between heterogeneous iris images, where pairs of filters are learned from two heterogeneous sources. (2) Different from two separate steps in terms of handcrafted feature extraction and feature matching in conventional solutions, DeepIris directly learns a nonlinear mapping function between pairs of iris images and their identity supervision with a pairwise filter bank (PFB) from different sources. Thus, the learned pairwise filters can adapt to new sources when given new training data. Extensive experimental results on the Q-FIRE and the CASIA cross sensor datasets demonstrate that EER (Equal Error Rate) of heterogeneous iris verification is reduced by 90% using DeepIris compared to traditional methods.

1. Introduction

Iris recognition [7] has been widely used for personal identification due to the unique, complex, and stable texture patterns in iris. After over decades of research, conventional iris recognition under homogeneous and controlled conditions has been extensively studied. Recently, more attention has been paid on uncontrolled and heterogeneous iris recognition, i.e., matching iris images across different domains. It includes the problems of iris images captured at different resolutions or from different sensors, etc. In real-world applications, iris images are usually captured from different domains. For example, both close-range and long-range iris sensors have been developed for iris image acquisition in recent years, where iris images are represented at different resolutions. Low-resolution iris images captured at a distance or on the move [14] usually lose many texture details, whereas enrolled iris images are captured with high-resolution in controlled conditions. Some iris systems only use iris sensors from one vendor but the system update procedure introduces different versions of iris sensors. Specifically, large-scale identification applications such as national ID systems usually use different sensors from a number of vendors to capture iris images. Some examples of heterogeneous iris images are shown in Fig. 1. It is obvious that there are significant differences at iris texture details among these iris images acquired from different domains. Heterogeneous iris images have large intra-class variations, which challenge the conventional well-performed iris recognition systems. The International Biometric Group[1] concluded that the verification performance of heterogeneous iris recognition degraded significantly. This encourages researchers to propose more effective algorithms for heterogeneous iris verification.

Existing methods for heterogeneous iris verification mainly focus on bridging the gap between heterogeneous iris images, which often use three approaches based on pixel, feature, and code level. Image enhancement is the most intuitive pixel based solutions for heterogeneous iris matching. This kind of method attempts to generate the similar visual appearance from different sources by reconstructing [15] or predicting [17], etc. However, these solutions focus on visual effects rather than the recognition accuracy. Feature

☆ This paper has been recommended for acceptance by Maria Frucci.

* Corresponding author. Tel.: +86 10 8254 4611.
E-mail address: zhangman@nlpr.ia.ac.cn (M. Zhang).

[1] http://www.biometric-scatalog.org/itirt/itirt-FinalReport.pdf

http://dx.doi.org/10.1016/j.patrec.2015.09.016
0167-8655/

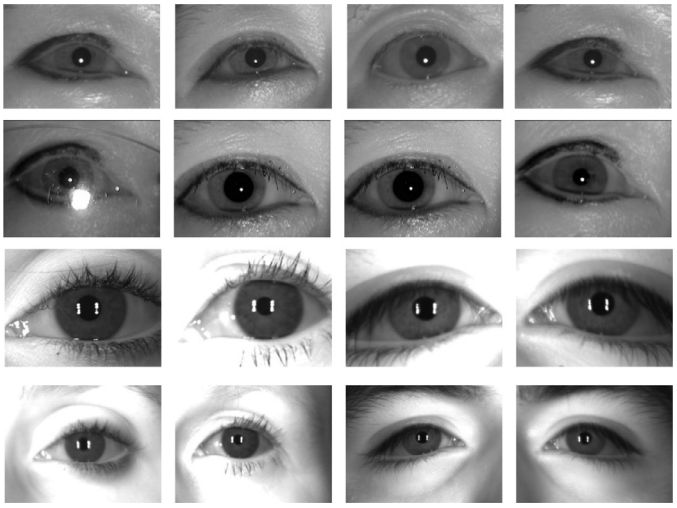

Fig. 1. Iris images captured by different sensors and at different distances. The first and the second row are iris images captured by different sensors. Images in the first row are acquired by IKEMB-110, images in the second rows are acquired by Iris Guard H100. In the first two rows, images of the same columns belong to the same subject. The third and the fourth rows are iris images captured from different distances. Images in the third row are acquired at 5 ft, images in the forth row are acquired at 11 ft. In the last two rows, images of the same columns belong to the same subject.

and code level methods extract the same features from heterogeneous sources at the first step. A code level information fusion [12] is proposed to increase the similarity of iris code from different resolutions. Other literature [19,20], extracts features that are invariant for both kinds of sensors. These mentioned methods extract the features through the same filter template for heterogeneous iris images. In particular, the exactly same filters are not designed for different domains, which could be suboptimal to filter images from different sources. While automatically learned filters taking the sources difference into account which might lead to better verification performance.

In this study, we propose a general deep learning model named DeepIris to solve the problem of heterogeneous iris matching. DeepIris aims to learn the relational filters between pairs of iris images based on convolutional neural networks (CNNs). The relational features, obtained by the pairwise filters, are able to measure the similarities of local regions between pairs of iris images. DeepIris automatically learns a nonlinear mapping function between the input heterogeneous iris image pairs and the output labels based on CNNs. Our approach is different from conventional methods that apply two separate steps in terms of iris feature extraction and feature matching, where the supervision is fully exploited for feature learning. The contributions of this model are threefold.

(1) This study is the first attempt to solve iris verification using deep learning method. Experimental results demonstrate that the EER (Equal Error Rate) of heterogeneous iris verification is reduced by 90% compared to traditional methods on two iris image databases Q-Fire and CASIA.

(2) The relational filters are learned by taking the domain difference into account. In contrast to traditional methods, heterogeneous sources are processed by the source-specific filters, i.e., the peculiarity of sources are taken into consideration.

(3) The proposed model is an end-to-end solution for heterogeneous iris matching, where relational filters adaptive to heterogeneous sources are automatically learned rather than hand-crafted.

The remainder of this paper is organized as follows. The previous work is reviewed in Section 2. The proposed model is described in Section 3. The experimental results and analysis are in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

Cross-resolution iris verification aims to match low-resolution probe iris images with high-resolution gallery ones captured at enrollment stage. Compared to high-resolution iris images, low-resolution ones usually lack a great amount of texture information. To improve the matching accuracy between low-resolution and high-resolution iris images, a straightforward way is to super resolve low-resolution iris images to obtain its corresponding high-resolution ones, namely super-resolution. For example, the quality of low-resolution iris images are improved by prediction [17] or weighted averaging [15]. However, these methods focus on enhancing iris images for better visual effects, but not improving the verification accuracy between iris images at different resolutions.

Since the pixel level reconstruction are unstable, [16] and [12] propose to operate at a higher level which can achieve higher accuracy. In particular, [16] fuses multiple Gabor filtered low-resolution iris images at the feature level and [12] uses a set of binary feature codes of low-resolution iris images. But these methods may still be problematic when there are not enough low-resolution iris images for fusion or coding.

Recently, the accuracies of low-resolution matching are improved by metric learning [2,3,21]. For example, a novel Mahalanobis distance learning algorithm [13] is proposed where a metric distance is learned between low-resolution and high-resolution iris images in code level by scaling and rotating the feature space. The learned metric distance exhibits strong generalization ability to separate cross-resolution data. This kind of methods ignore source difference and utilize only single handcrafted ordinal filter, i.e., filtering the cross-resolution iris images with exactly the same filter template, which may lead to lower accuracy. In addition, they modify the provided feature space without feature selection, where the final performance could be heavily affected by some noisy features.

In addition to cross-resolution matching, another closely related problem is cross-sensor iris matching. Iris images captured by different sensors follow different distributions [4] which can easily degrades the matching accuracy [1,5,6]. The first cross-sensor iris dataset is built and studied in [5,6]. They conclude that the overall accuracy is affected by the selected algorithms and the cameras. Authors in [1] propose a framework that first predicts the source camera and then applies selective enhancement algorithms to achieve better performance. The benefits of these methods are that each source is selectively preprocessed. Afterwards, feature-level methods are introduced to select effective cross-sensor features. For example, a coupled dictionary [19] are learned for feature extraction and a large margin feature selection method [20] is provided for cross-sensor iris verification. However, these feature-level methods are often limited by the filter pool even though it is over complete. In addition, all the feature selection methods filter the cross-sensor iris images with the same template and ignore the influence of different sensors. Moreover, computational cost at the training stage is intractable [19,20]. Even if a large training dataset is given, they could not take the full advantage of the large scale data.

3. Methods

The goal of iris verification is to decide whether the input pairs of iris images belong to the same subject or not, which is actually a binary classification task. To deal with this problem, we proposed a novel framework namely DeepIris, whose architecture is shown in Fig. 2. In the following, we will detail this framework in terms of the preprocessing of iris images, the construction of pairwise filter layer, the algorithm implementation and the score level fusion.

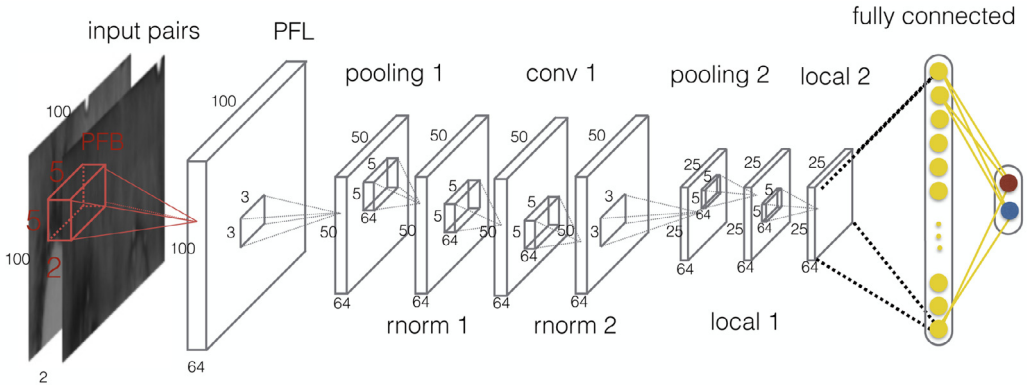

Fig. 2. The architecture of the proposed method. The rnorm is denoted as normalization layer. The pooling is denoted as max-pooling layer. The neuron numbers are marked beside layers.

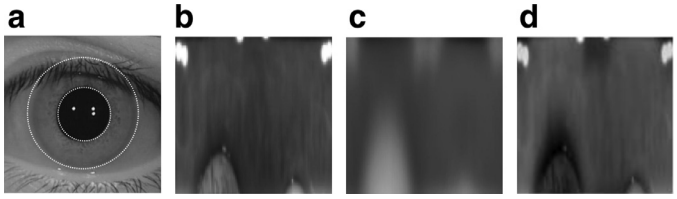

Fig. 3. Iris image preprocessing. (a) Localized iris images. (b) Normalized iris image. (c) The average gray image which is computed in the training data. (d) The preprocessed iris image which is subtracted by the average gray image at each pixels.

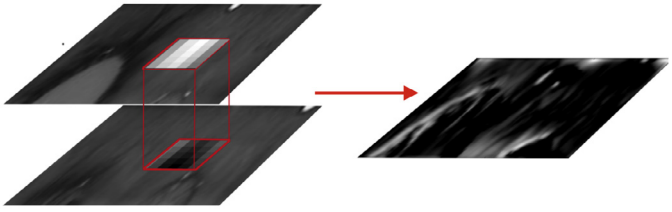

Fig. 4. The pipeline of Pairwise Filter Layer. A pair of heterogeneous images are fed into the layer. After filtered by the learned pairwise filters, the two feature maps are summarized into the similarity map. The similarity map demonstrate the similarity between the input pairs of iris images which is depicted in Fig. 6.

3.1. Iris image preprocessing

As shown in Fig. 3, the preprocessing of iris images include the following steps. (1) The iris texture regions are segmented from original iris images. (2) The annular region of the iris is transformed to unwrapped or rectangular iris images, which is usually called iris normalization. We then closely follow the preprocessing method in [9], and downsample the normalized iris images to 100×100. The downsampling process can prevent overfitting in some extent, because the original iris image are large, which increases the parameters greatly and leads to easier overfitting. (3) After that, every pixel is subtracted by the mean gray value which is computed using the whole training dataset. This step aims to rescale all the data to have zero energy. Experimental results show that different preprocessing strategies can influence the performance significantly.

3.2. Pairwise filter layer

When dealing with the heterogeneous iris images, existing methods usually extract features from different sources with the same filter template. This might be suboptimal because the feature extraction does not consider the difference of distribution between different sources. In fact, designing robust filters for heterogeneous iris verification is rarely investigated. Because it is difficult to manually design robust filters to deal with the complex variations, which results from cross-resolution, cross-sensor, illumination, head pose, gaze etc. So we decide to adaptively learn source-specific filters based on convolutional neural networks.

The pairwise filter layer takes into pairs of iris images as input, and a couple of filters in the pairwise filter layer is shown in Fig. 4. The pairwise filters convolve a pair of input images and summarize their feature maps to generate the similarity map. While in the context of the architecture of DeepIris in Fig. 2, the input is two gray normalized iris images, which are stacked into two-channel instead of spliced into one map. So the convolutional operation of pairwise

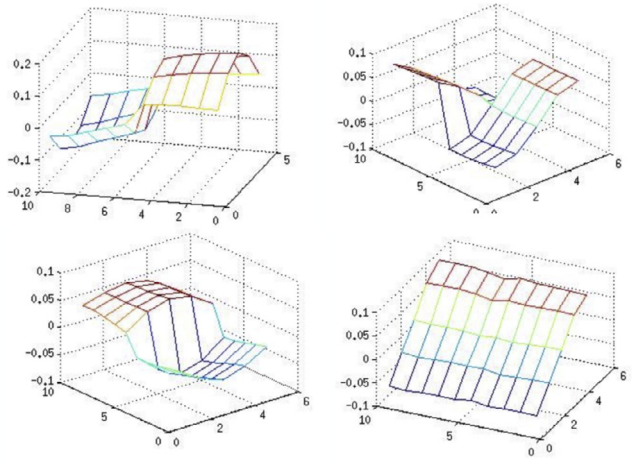

Fig. 5. Some learned pairwise filters. Each 10×5 image indicates a pair filter. The left 5×5 and the right 5×5 are the two filters in one filter pair.

filter layer is expressed as

$$Y_j = \sum_{i=1}^{2} X_i * W_{i,j} + B_j \tag{1}$$

where (X_1, X_2) denotes the stacked input image pair. Y_j is the j-th similarity map of pair (X_1, X_2). $W_{1,j}$ and $W_{2,j}$ are the j-th pairwise filters for pair (X_1, X_2), respectively. B_j is the bias of the j-th similarity map for pair (X_1, X_2). $*$ denotes convolutional operation. The ReLU nonlinearity $max(0, x)$ is used for the hidden neurons, which has better fitting ability than the sigmoid function as argued in [11].

Fig. 5 shows four learned pairwise filters. The two filters in one filter pair are depicted together for better illustration. The two pictures in the first column indicate that the pairwise filters process the

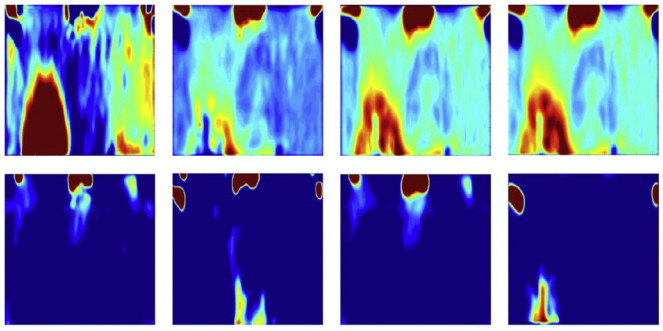

Fig. 6. The feature maps processed by the pairwise filters. The first row shows inter-subject similarity map and the second row shows intra-subject similarity map. The color of the pixel in the similarity maps indicate the similarity between the local region of iris images pairs. The deeper color (dark blue) indicate the lower response. The images are enhanced for better visual effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

local regions of the images pairs in a different way. For the second column, the pairwise filters are more likely to adopt a different filter at each map and then summarize the two maps. Fig. 6 shows the similarity maps after processed by the pairwise filters. The similarity maps show the similarities at each pixel between the local region of the iris image pairs. The lower response (the darker color in the picture) indicates higher similarity. The first row shows the inter-subject similarity maps and the second row shows the intra-subject similarity maps. As we known, the inter-subject similarity maps have strong response to the PFB (pairwise filter bank), which results in more textures and lighter color in visualization. On the other hand, when the input pairs belong to intra-class subjects, the similarity maps have lower response (the deeper color, dark blue) which demonstrates the greater similarity between the two input images.

3.3. Architecture

The model is composed of nine layers including one pairwise filter layer, one convolutional layer, two pooling layers, two normalization layers, two local layers and one full connection layer. The size of each layer is depicted in Fig. 2. The input pairwise iris images are normalized and reshaped to 100 × 100 gray scale images as shown in Fig. 3. The input pairwise gray scale images are prepared as a two-channel matrix to pass through a set of pairwise filters in the first convolutional layer. There are 64 pairs of filters which have a size of 5 × 5 in this convolutional layer. A two-pixel zero-padding is apply to the border, so the activation after the first layer is still 100 × 100. Overlap spatial pooling is carried out by pooling a 3 × 3 region with a stride of 2, so this layer down samples the image into half size of 50 × 50. The normalization layer conducts a cross-map normalization in each unit. Experimental results show that the performance can be increased with two additional normalization layers in the network. The second stage is similar to the first one excepting for that the normalization layer and the pooling layer are switched. Because the normalized iris images have structure relations to some extent, i.e., like the occluded eyelids usually lie in the same area, local layers seem to be capable to capture more local information. Experiments also show that the accuracy improves by adding the local layers. We employ two local layers following the second stage, each of which includes 64 filters of the size 3 × 3 with one-pixel padding. The final full connections receive 25 × 25 × 64 neurons and output two predictions, corresponding to the two given labels, i.e., 0 is the intra-class pair and 1 is the inter-class pair. The whole framework is shown in Fig. 9, where a list of architecture is designed with different configurations and implementations.

All the convolutional layers, local layers and the full connection layers use the Rectified Linear Unit [11] as the activation function. Firstly, we train the network with a large learning rate of 0.001 and reduce by a factor of 10 when the classification error rate on the evaluation set stops decreasing. We empirically find that even when the training classification error decreases to almost zero, reducing the learning rate can still help to lower the classification error on the test set.

3.4. Score level fusion

There exists rotation differences between iris images due to head motion. Generally, the iris recognition algorithms translate the normalized iris images by a few pixels to search for the best match. In this work, the normalized iris images are translated within one pixel to correct the rotation variance. Given a pair of iris images, they both are normalized to 100 × 100 in pixels as shown in Fig. 3, denoted as X_1 and X_2, respectively. Then each image is translated by one pixel to the left and to the right. Therefore three copies of X_1 (i.e., X_{11}, X_{12}, X_{13}) are generated. At the same time, if the orders are taken into consideration, there are two kinds of pairs which are X_1–X_2 and X_2–X_1. Finally, six pairs are generated for every input of two images: X_{11}–X_2, X_{12}–X_2, X_{13}–X_2, X_2–X_{11}, X_2–X_{12}, X_2–X_{13}. The total six pairs are fed into the model and six similarity scores of the input pairs are reported. The score fusion will decide the best alignment of the verified two iris images. In this way, the proposed method can correct the rotation variance. Three fusion methods are compared during the experiments, which are maximum fusion, minimum fusion and average fusion. In the experiment part, score level fusion methods increase the performance to a large extent, which demonstrates its effectiveness.

4. Experiments

To evaluate the effectiveness, we compare our model with other state-of-the-art methods on two datasets: the Q-FIRE [10] dataset and the CASIA cross sensor iris dataset. The description of the two datasets are listed in Section 4.1. The sizes of the training and testing set are given in Section 4.2. Some factors affecting the performance are analyzed in Section 4.3. The comparisons with the state-of-the-art methods are shown in Sections 4.4 and 4.5. The equal error rate (EER) and receiver operating characteristic (ROC) curve are used as quantitatively measure in the heterogeneous iris verification.

4.1. Description of the datasets

Q-FIRE dataset: Iris images in Q-FIRE [10] are captured in different distances at 5, 7, and 11 ft by the same sensor. So the resolution of the radius of iris rings results in high resolution (about 300 pixels), medium resolution (about 200 pixels) and low resolution (about 100 pixels). The first 160 subjects are selected to construct the cross-resolution dataset. The 5 ft and 11 ft iris images with high illumination are selected as high-resolution and low-resolution iris images. Some mislocalized iris images are dropped because of serious blur and occlusion. The total number of images in the dataset is 3123 (high-resolution iris images) and 2902 (low-resolution iris images) for the first 160 selected subjects.

CASIA cross sensor dataset: To evaluate the feasibility of the proposed method, two acquisition systems are adopted to build a new iris database. It includes two close-up sensors (IK, short for IKEMB-110; IG, short for IrisGuard H100) and one long-range iris recognition system (LRI). The LRI captures iris images at a distance of 3–5 m. The IG sensor and the LRI sensor are selected to construct the cross-sensor dataset. They both contain 350 subjects, each subject has 10 images per eye. This dataset has not been published yet.

Table 1
Details of the Q-FIRE dataset used in experiments.

	Number of classes	Number of images
05 feet-train	100	1680
05 feet-test	60	911
11 feet-train	100	1568
11 feet-test	60	966

Table 2
Details of the CASIA cross sensor dataset used in experiments.

	number of classes	number of images
train	300	6000
test	50	1000

Table 3
Comparisons of the strength of downsampling on the Q-FIRE dataset.

Architecture A	EER (%)
100*100	0.15
100*128	0.69
100*256	4.08

Table 4
Comparisons of different data normalization methods on the Q-FIRE dataset.

Architecture A	EER (%)
Simple rescaling	0.15
Per-example mean subtraction	0.96
Without subtraction	4.15

4.2. Experimental settings

Subject-disjoint training and testing are conducted on two datasets. For the Q-FIRE dataset, the first 100 subjects are selected for training and the rest 66 subjects are used for testing. Detailed information of the Q-FIRE dataset is listed in Table 1. The training set of the CASIA cross sensor dataset is created by selecting the first 300 subjects from each sensor and the rest 50 subjects are left for open-set testing. The compared methods use the first 50 subjects for training, and the last 50 subjects for testing (Tables 2 and 3).

During training, our framework attempts to learn filter bank from pairs of heterogeneous iris images which are coming from different sources (e.g., high-resolution vs. low-resolution or sensor A vs. sensor B). The intra-subject heterogeneous pairs are labeled to 1 and the inter-subject heterogenous pairs are labeled to 0. In this way, the verification task is handled as a binary classification problem.

In the test phase, each iris image is organized as pairs with the whole rest images. The total number of testing pairs on the Q-FIRE and CASIA datasets are about 10 millions and 2 millions, respectively.

4.3. Evaluation of training and test strategies

In this section, a list of experiments are conducted on the Q-FIRE dataset. Firstly, a set of different deep architectures are designed to evaluate the proposed model, among which the selected three architectures are shown in Fig. 9. Then the best performing architecture is used to study other factors of variation which may affect the performance of our model, e.g., the depth of the architecture, data preprocessing, dropout technique and normalization layer. The details are shown as follows (Table 4).

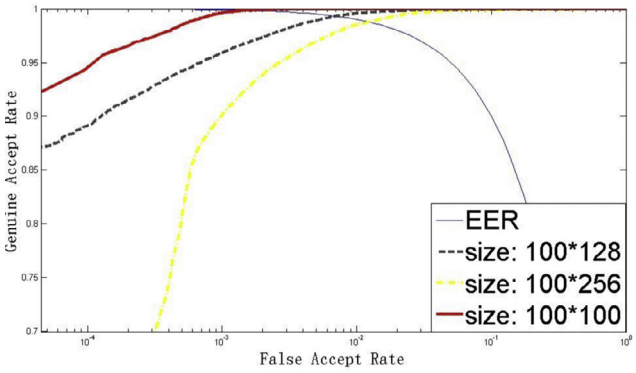

Fig. 7. The ROC curves of the model with different strength of downsampling on the Q-FIRE dataset.

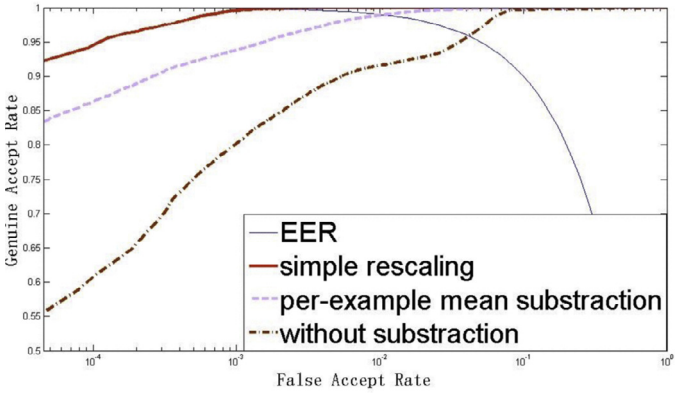

Fig. 8. The ROC curves of the model with different data normalization methods on the Q-FIRE dataset.

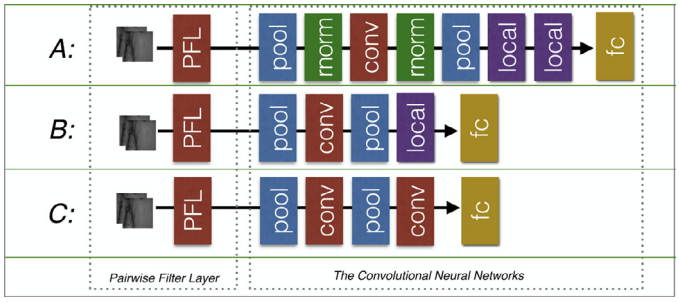

Fig. 9. Different architectures of CNN to classify the feature map obtained by the pairwise filter layer. There are three architectures constructed which are named as A, B and C.

Image preprocessing. To study the effect of the downsampling in the data preprocessing, we use different sizes of the input pairs for training and compare their performance. Fig. 7 shows that the downsampling is helpful to improve the performance. The reason is that the bigger size leads to easier overfitting when training data is limited. In addition, the down-sampling of the input images can greatly minimize the number of parameters which potentially accelerates the learning procedure.

We also compare the results of two preprocessing methods: simple rescaling and per-example mean subtraction. Simple rescaling subtracts the mean gray value of the whole dataset from each image, while per-example mean subtraction subtracts the mean gray value of the image itself. Fig. 8 shows the ROC curve of the model with different data normalization methods. It is obvious that the performance degrades greatly if the data subtraction is not performed. It demonstrates the necessity of the preprocessing procedure. We can also find

Table 5
Comparisons of different architecture on the Q-FIRE dataset. Details of architecture A, B and C are shown in Fig. 9.

Architecture	EER (%)
Architecture A	0.15
Architecture B	0.39
Architecture C	0.72

Table 6
Comparisons of different implementations in architecture A. The second column is the EER on the Q-FIRE dataset.

Arch	EER (%)
A	0.15
A-without dropout	0.64
A-unnormalized	0.69

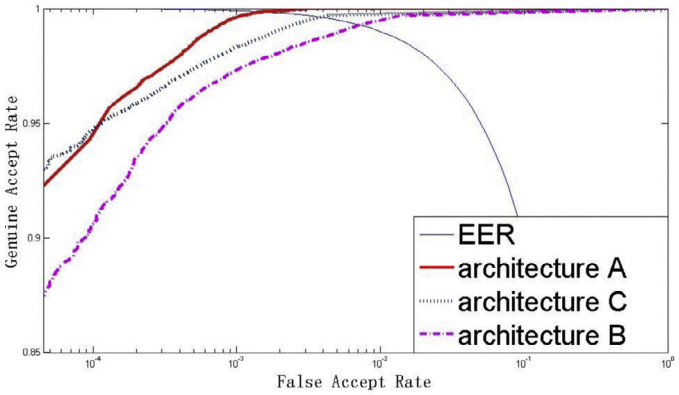

Fig. 10. The ROC curves of different implementations in architecture A on the Q-FIRE dataset.

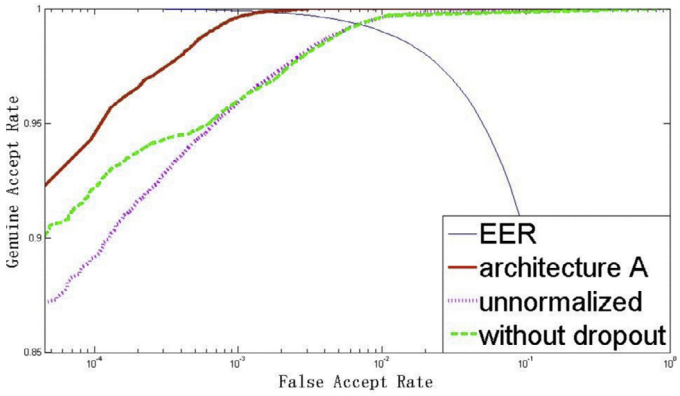

Fig. 11. The ROC curves of different implementations of architecture A on the Q-FIRE dataset.

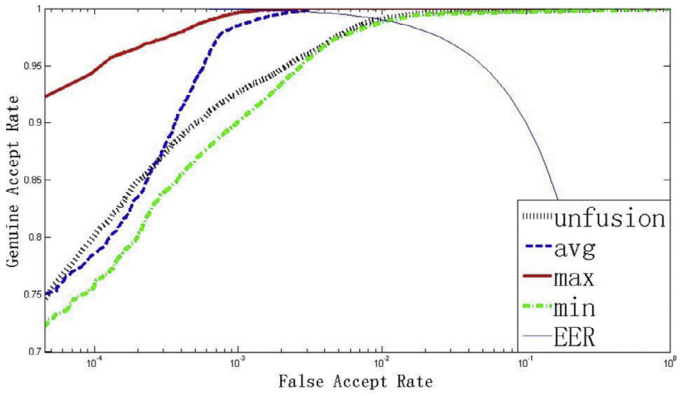

Fig. 12. The ROC curves of three score level fusion methods on the Q-FIRE dataset.

that simple rescaling performs better than per-example subtraction in some extent.

Architecture depth. Considering that the depth of the architecture determines the capability of the model, we compare three networks from shallow to deep in the following. Fig. 9 shows the details of the compared three architectures. From Table 5 and Fig. 10, we can conclude as follows. (1) Architecture A is the best performing architecture, which demonstrates that the increasing depth of the model is beneficial to the accuracy improvement. In fact, the deeper network can greatly increase the fitting and generalization abilities. (2) Architecture B performs better than architecture C at the same depth. Notice that the only difference between architecture B and architecture C is the second last layer, where Architecture B uses a local layer rather than a convolutional layer. It demonstrates that the local layer can be more helpful, because local layers captures the local relations between the pairwise iris images.

Dropout technique. Dropout technique is recently introduced in [11] for neural networks. The core idea of dropout is to reduce the co-adaptation of neurons to learn robust features. Generally, we realize dropout by dropping each neuron with a probability of 0.5. Those neurons which are dropped out make no contribution to both forward pass and backward propagation. In every training epoch, the neural network will sample a different architecture. In the test phase, all the neurons are used but the output is multiplied by 0.5 [11]. We perform a comparative experiment with the best architecture A, and find that when the dropout technique is removed, the EER will increase from 0.15% to 0.64%. The dropout technique is important to the performance of deep learning based iris recognition (Table 6).

Normalization layer. The local response normalization layer is helpful to generalization capability [11]. Architecture A with normalization achieves a better EER 0.15% than 0.69% of A-unnormalized model, which removes all the normalization layers (Fig. 11).

Score level fusion. The conventional iris recognition algorithms translate the normalized iris images a few pixels to search the best match. Similarly, we design a score level fusion method to find the best match of the input pairs, i.e., two testing iris images can generate six pairs as stated in previous sections. From Figs. 13 and 12 we can conclude that maximum score fusion improves the performance. Especially when the false accept rate is significantly low, the maximum fusion can ensure the low false reject rate.

4.4. Cross resolution iris verification

The analysis above validates the effectiveness of DeepIris. The best architecture and its best configurations are selected to compare with the state-of-the-art methods on the Q-FIRE dataset. In the comparison, the experimental results of four methods are reported. The first one is denoted as ITML [8] which learns a globality metric. The second one is denoted as LMNN [18] which learns the metric by both considering the locality and globality. The third one is MDML [13] which learns a Mahalanobis distance metric on code level features. The proposed convolutional neural network based method is denoted as Proposed. Fig. 14 compares the ROC curves of the four methods, and their corresponding EERs are shown in Table 7. We can see that the proposed method achieves the best performance. Not only the EER is only 1/10 of other methods, but also the genuine accept rate (GAR) is as high as 95% when the false accept rate (FAR) is significantly low to 10^{-4}.

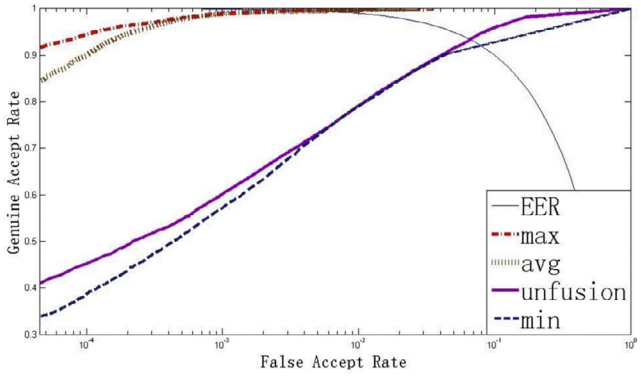

Fig. 13. The ROC curves of three score level fusion methods on the CASIA dataset.

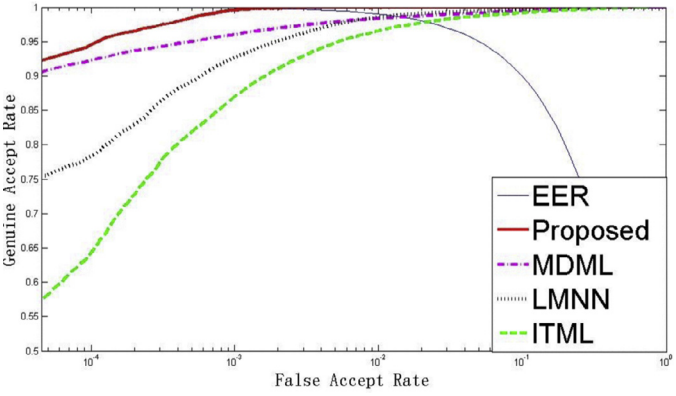

Fig. 14. Comparisons of the proposed algorithm with the existing metric learning methods on the Q-FIRE dataset. ITML [8], LMNN [18] and MDML [13].

Table 7
Performance of the compared methods on the Q-FIRE dataset.

Methods	EER (%)
ITML [8]	2.35
LMNN [18]	1.73
MDML [13]	1.67
Proposed	0.15

Table 8
Performance of the compared methods on the CASIA dataset.

Methods	EER (%)
CLP [19]	5.81
Proposed	0.31

4.5. Cross sensor iris verification

To further verify the performance of our DeepIris, we have carried out the experiments on the CASIA cross sensor dataset. The state-of-the-art coupled linear programming (CLP) [19] method is selected for comparison. The ROC curve is drawn in Fig. 15 and the EER is shown in Table 8. Our method achieves 0.31% EER, outperforming the state-of-the-art CLP method with 5.50% improvement. It is clear that the proposed method outperforms the CLP, due to its automatically learned source-specific filters.

4.6. Computational cost

It takes about 0.7 ms for DeepIris to process one pair of images on a GTX Titan GPU. This is extremely fast for iris feature extraction and iris feature matching compared with traditional methods. Therefore, the proposed DeepIris method is suitable for real-word iris verification systems with the advantages of high accuracy and low complexity.

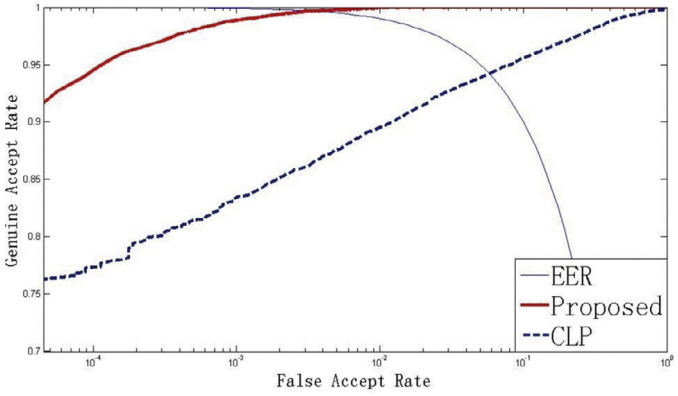

Fig. 15. Comparisons of the proposed algorithm with CLP [19] on the CASIA cross sensor dataset.

5. Conclusions

In this paper we have proposed a deep learning based method to solve heterogeneous iris verification based on convolutional neural networks. The automatically learned pairwise filters and the deep representations are learned jointly from raw pixels instead of using handcrafted features. The core contribution of this paper is automatically learning the relational features to express similarities for heterogeneous iris images from raw pixels. The pairs of filters are learned for heterogeneous sources, respectively. Thus the difference between sources are taken into account for better performance. Experimental results show that the proposed method achieves promising and competitive performance for both cross-resolution and cross-sensor iris verification. The method is more advantageous after training on big iris data, which is currently available with the fast and wide deployments of iris recognition systems. Although DeepIris model needs more time for training (a few hours in our experiments), the learned model can process a pair of iris images for identity verification extremely fast.

Acknowledgments

This work is supported by National Basic Research Program of China (Grant no. 2012CB316300), National Natural Science Foundation of China (Grant no. 61273272), and Beijing Baidu Netcom Science Technology Co., Ltd.

References

[1] S. Arora, M. Vatsa, R. Singh, A. Jain, On iris camera interoperability, in: Proceedings of International Conference on Biometrics: Theory, Applications and Systems, 2012, doi:10.1109/BTAS.2012.6374599.
[2] X. Ben, W. Meng, R. Yan, K. Wang, An improved biometrics technique based on metric learning approach, Neurocomputing 97 (2012) 44–51.
[3] S. Biswas, K.W. Bowyer, P.J. Flynn, Multidimensional scaling for matching low-resolution face images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 2019–2030, doi:10.1109/TPAMI.2011.278.
[4] K.W. Bowyer, S.E. Baker, A. Hentz, K. Hollingsworth, T. Peters, P.J. Flynn, Factors that degrade the match distribution in iris biometrics, Identity Inf. Soc. 2 (3) (2009) 327–343.
[5] R. Connaughton, A. Sgroi, K. Bowyer, P. Flynn, A cross-sensor evaluation of three commercial iris cameras for iris biometrics, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011, doi:10.1109/CVPRW.2011.5981814.
[6] R. Connaughton, A. Sgroi, K. Bowyer, P.J. Flynn, A multialgorithm analysis of three iris biometric sensors, IEEE Trans. Inf. Forensics Secur. 7 (3) (2012) 919–931.

[7] J. Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Trans. Pattern Anal. Mach. Intell. 15 (11) (1993) 1148–1161, doi:10.1109/34.244676.
[8] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of International Conference on Machine Learning, ACM, 2007.
[9] Z. He, T. Tan, Z. Sun, X. Qiu, Toward accurate and fast iris segmentation for iris biometrics, IEEE Trans. Pattern Anal Mach. Intell. 31 (9) (2009) 1670–1684, doi:10.1109/TPAMI.2008.183.
[10] P. Johnson, P. Lopez-Meyer, N. Sazonova, F. Hua, S. Schuckers, Quality in face and iris research ensemble (q-fire), in: Proceedings of International Conference on Biometrics: Theory Applications and Systems, IEEE, 2010.
[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of Advances in Neural Information Processing Systems, 2012.
[12] J. Liu, Z. Sun, T. Tan, Code-level information fusion of low-resolution iris image sequences for personal identification at a distance, in: Proceedings of International Conference on Biometrics: Theory, Applications and Systems, 2013, doi:10.1109/BTAS.2013.6712692.
[13] J. Liu, Z. Sun, T. Tan, Distance metric learning for recognizing low-resolution iris images, Neurocomputing (2014).
[14] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. LoIacono, S. Mangru, M. Tinker, T.M. Zappia, W.Y. Zhao, Iris on the move: acquisition of images for iris recognition in less constrained environments, Proc. IEEE 94 (11) (2006) 1936–1947.
[15] K. Nguyen, C. Fookes, S. Sridharan, S. Denman, Quality-driven super-resolution for less constrained iris recognition at a distance and on the move, IEEE Trans. Inf. Forensics Secur. 6 (4) (2011) 1248–1258, doi:10.1109/TIFS.2011.2159597.
[16] K. Nguyen, S. Sridharan, S. Denman, C. Fookes, Feature-domain super-resolution framework for gabor-based face and iris recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, doi:10.1109/CVPR.2012.6247984.
[17] K.Y. Shin, K.R. Park, B.J. Kang, S.J. Park, Super-resolution method based on multiple multi-layer perceptrons for iris recognition, in: Proceedings of International Conference on Ubiquitous Information Technologies Applications, 2009, doi:10.1109/ICUT.2009.5405701.
[18] K.Q. Weinberger, J. Blitzer, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, in: Proceedings of Advances in Neural Information Processing Systems, 2005, pp. 1473–1480.
[19] L. Xiao, Z. Sun, R. He, T. Tan, Coupled feature selection for cross-sensor iris recognition, in: Proceedings of International Conference on Biometrics: Theory, Applications and Systems, 2013, doi:10.1109/BTAS.2013.6712752.
[20] L. Xiao, Z. Sun, R. He, T. Tan, Margin based feature selection for cross-sensor iris recognition via linear programming, in: Proceedings of Asian Conference on Pattern Recognition, 2013, doi:10.1109/ACPR.2013.34.
[21] E.P. Xing, M.I. Jordan, S. Russell, A.Y. Ng, Distance metric learning with application to clustering with side-information, in: Proceedings of Advances in Neural information Processing Systems, 2002, pp. 505–512.