Automated Classification of Mislabeled Near-Infrared Left and Right Iris Images using Convolutional Neural Networks

Yixin Du, Thirimachos Bourlai and Jeremy Dawson
West Virginia University
LCSEE, 395 Evansdale Drive, Morgantown, WV 26506-6070, U.S.A.
yidu@mix.wvu.edu, thirimachos.bourlai@mail.wvu.edu and jeremy.dawson@mail.wvu.edu

Abstract

In this paper, we propose a Convolutional Neural Network (CNN) with unified architecture (no need to re-design it for each unique iris database used) that operates well in a diverse set of iris databases. The CNN is designed to automatically recognize mislabeled left and right iris images by iris recognition system operators, and thus, extend the capabilities of a conventional iris recognition system. Our proposed approach is composed of three steps. First, for each iris database used as input, a CNN is trained using part of the database. Second, an empirical parameter optimization study is conducted so that classification performance is acceptable. Finally, the proposed classifier is tested on the remaining images of the same database used for training. The performance of the proposed network is evaluated on small- and large-scale iris databases, including the NIST's Iris Challenge Evaluation (ICE), LG ICAM 4000 iris, the CASIA Lamp, and the Pupil Light Reflex (PLR) databases. Experimental results show that independent of the databases used or whether the classification performance is tested on either a left- or right-side dataset, our approach results in a classification performance ranging from 97.5 to 100%.

1. Introduction

Iris recognition has been used for reliable personal identification for more than 20 years [7]. A typical iris recognition system is composed of a set of algorithmic steps following data collection and management, including preprocessing of iris images, feature extraction and matching [13]. Performance results using such systems is very high due to the fact that iris is one of the most unique and stable biometric modalities. The main reasons are two-fold: (i) iris area structure remains unchanged over time [9] and (ii) the large variability of color, texture and shape of irises among individuals results in a very low false matching rate [4]. While the iris modality is very reliable, one of the major drawbacks of an iris recognition system can be the involvement of biometric operators that may mislabel raw left for right iris images and vice versa during or after data collection. Such errors can affect the reliability of iris matchers [2]. In Figure 1, we illustrate the process of generating iris datasets, including the collection of iris images, iris labeling (left vs. right) and, finally, collection post-processing that automatically determines whether the collected iris images are from the right or left side of a subject's face. Motivated by the challenges associated with the problem, including the fact that most iris images do not always have the eyebrow present (it can be either absent, partially occluded, or fully present), we propose a reliable classifier to identify the mislabeled iris images and, thus, eliminate potential drops in iris matching performance.

1.1. Motivation

Related to the aforementioned problem, the most relevant works in the literature are discussed below. First, Li and Savvides [12] used a single classification attribute by comparing the locations of pupil and iris centers. If the pupil center is located on the left hand side of the iris center, it is considered to be a right eye image and if otherwise a left eye image. Authors reported that they achieved a 93.9% and 92.6% accuracy for the left and right eye datasets respectively, when using the NIST ICE database [14]. One potential drawback of their proposed system is that the algorithm depends on the accuracy of the iris/pupil segmentation step. Iris segmentation is not always accurate enough due to the case of iris occlusion from eyelids/eyelashes. In addition, there are other noise factors that can negatively affect iris segmentation accuracy including reflections and eye blinking.

Bhat and Savvides [3] evaluated the performance of Active Shape Models (ASM) on the problem of left vs. right eye classification. The model is reported to be trained on 100 left and 100 right eye images of the NIST ICE database [14]. For each training image, landmarks across the top and

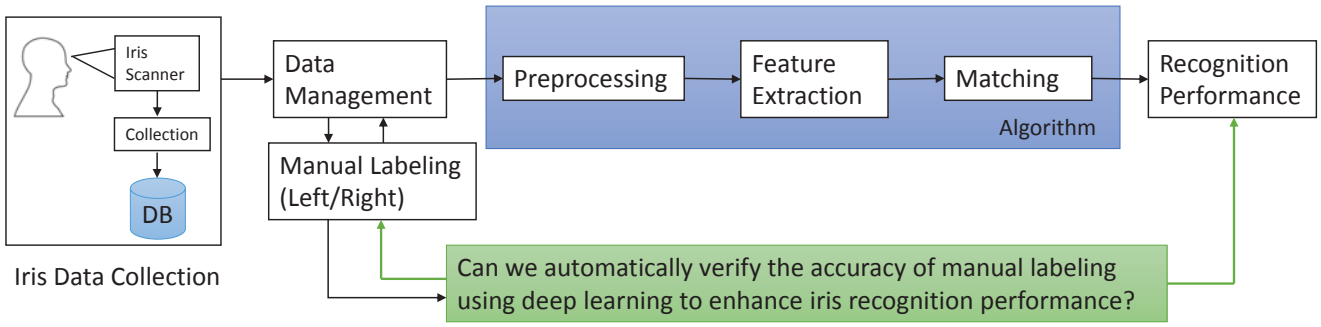

Figure 1: An overview of an iris recognition system after data collection is performed. First, iris images are collected by using an iris scanner, followed by a data management process where images are manually labeled into left and right eye folders. In practical applications there is potential risk of mislabeling the left and right iris images. The green box highlights our contribution to the recognition process by automatically verifying the accuracy of manual labeling using deep learning.

lower eyelid are manually labeled to describe the iris shape. Once the eye is localized using an eye detector, ASM is applied to iteratively search for the desired landmarks around the initial estimate using gray-level statistics. For testing, a relative likelihood ratio is computed to serve as the classification threshold. The authors reported that they reached 62.5% and 67.7% accuracy on the NIST ICE left and right eye datasets respectively. Abiantun and Savvides [2] developed a tear-duct detector using several techniques including boosted Haar features, support vector machines, PCA and LDA. The training is performed by cropping a 55×55 window on the tear-duct that is the positive class while the remaining part of the image is the negative class. The highest accuracy on the NIST ICE dataset is reported to be 69%. The major assumption in the aforementioned two papers is that the input iris images are complete, and no case of occlusion exist. However, in the iris database the authors used, there are several iris images on which either the tear-duct or the lateral angle (eye corner) are partially or fully occluded (see Figure 2). The problem is that there are approximatively 30% iris images in the NIST ICE dataset that the tear-duct is either partially or fully occluded [2]. This issue can significantly affect the classification performance of an iris recognition system, the reliability of which depends on whether the system can automatically determine whether an input iris image is coming from the left or right eye. The question is what algorithmic approach can be designed and developed to efficiently solve this problem.

Such an approach can be based on Convolutional Neural Networks (CNNs) that have lately become very popular due to the fact that CNNs can effectively learn novel feature representations and easily adapt to different computer vision problems. CNNs have been successfully used in many areas such as face recognition [18], visual document analysis [5][15], and image super-resolution [8]. A typical CNN architecture includes multiple stages, and each stage is composed of three layers: a convolution layer, a non-linearity layer, and a feature pooling layer [11].

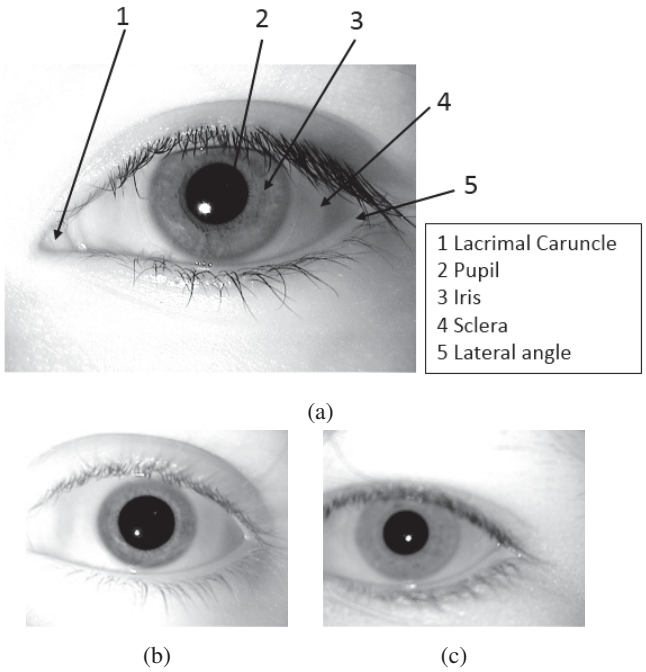

Figure 2: (a) Anatomy of the iris region when lacrimal caruncle (tear-duct) and lateral angle are fully present. (b) Tear-duct partially occluded. (c) Tear-duct fully occluded on a left-eye image. Occlusions such as (b) and (c) bring challenges to left and right iris image classification as an iris recognition system can hardly detect the tear-duct [2][3].

1.2. Contributions

In this paper, we propose a Convolutional Neural Network (CNN) with unified architecture that operates well

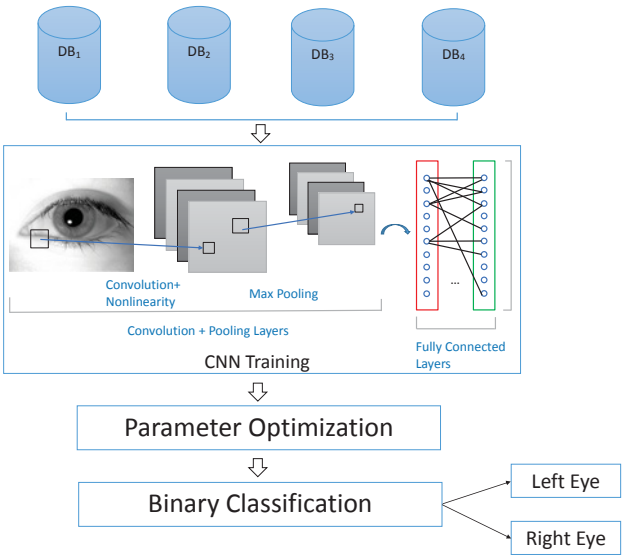

Figure 3: Overview of the proposed method for classification of left- and right-eye iris images using four different iris databases. CNN training/validation, parameter optimization and testing have been conducted on each of the four datasets separately.

when tested on four iris datasets with/without occlusion to automatically recognize mislabeled left and right iris images. There are three steps in our proposed method as shown in Figure 3. First, for each iris database used as input, a CNN is trained using part of the database. Second, an empirical parameter optimization study is conducted so that classification performance is acceptable. Finally, the proposed classifier is tested on the remaining images of the same database used for training. The advantages of our proposed approach are that (i) it is very fast (milliseconds are needed to classify each iris image), (ii) it is not limited in the type or size of the dataset used, (iii) it is not limited on the partial or full occlusion of the tear-duct or lateral angle, and (iv) it is a unified method, i.e., no need to redesign its architecture to be applied on alternative datasets. The accuracy of our proposed method ranges from 97.5 to 100%.

1.3. Paper Organization

Section 2 introduces the proposed method. In Section 3, we compare our results with others, and we show the network's performance on large scale datasets. Section 4 concludes the paper.

2. Methodology

The concept of Convolutional Neural Network (CNN) is inspired by the biological visual cortex as discussed in Hubel and Wiesel's work [10]. The visual cortex is the part of the brain that processes visual information. It contains cells acting as filters over the input space. Connections between neurons of adjacent layers are established by a CNN in which each pair of input/output has an associated weight. The model is an assembly of inter-connected nodes with weights and biases, and the output node is the inner product of each node and its weight. The output node can then be compared against the threshold for decision making. In our work we apply a CNN and successfully overcome challenges in left vs. right eye image classification, including cases of occlusion of the tear-duct.

2.1. Neural Networks

A Neural Network is composed of three layers: an input layer, a hidden layer, and an output layer. The units in the hidden/output layers are called neurons. Neurons of two adjacent layers are connected by weights. Each neuron receives signals from the inputs, then produces and transmits the resulting signal. A feed-forward neural network is composed of a series of functions [16]

$$f(x) = f_K(...f_2(f_1(x, w_1), w_2)..., w_K). \quad (1)$$

x denotes the original input, and w_k is the weight associated with step k. The above function takes the output of previous step and feeds it as the input for the current step. $w = (w_1, ..., w_k)$ are original weights randomly generated, and eventually learned from the data. In our problem, x represents the 2-D iris image to be classified, while all the remaining x_k are intermediate feature maps generated using the linear filter w_k. Figure 4 shows the convolution results after applying a Gaussian 12-random-filter bank toward a sample image.

Another step of CNN is called max-pooling. It is a non-linear down-sampling method to partition the input image into a set of rectangles with a maximum value in each sub-region. Max-pooling can reduce the system's original high dimensionality by eliminating non-maximal values for the purpose of increasing computational speed. The feed-forward neural network processes the convolution and max-pooling steps in a recursive manner such as three convolutions followed by one max-pooling and, finally, the process is repeated until some certain thresholds are met.

Back-propagation is a widely used algorithm for training CNN. The network is initialized with some random weights $w = (w_1, ..., w_k)$ and biases. Then, the error signals are passed backwards to update the weights and minimize the mean squared error between the prediction and actual target value:

$$E = \sum_i [Y_i - f(X_i, w_i)]^2 \quad (2)$$

These modifications toward weights are made from the output layer through each hidden layer, and eventually down to the first hidden layer.

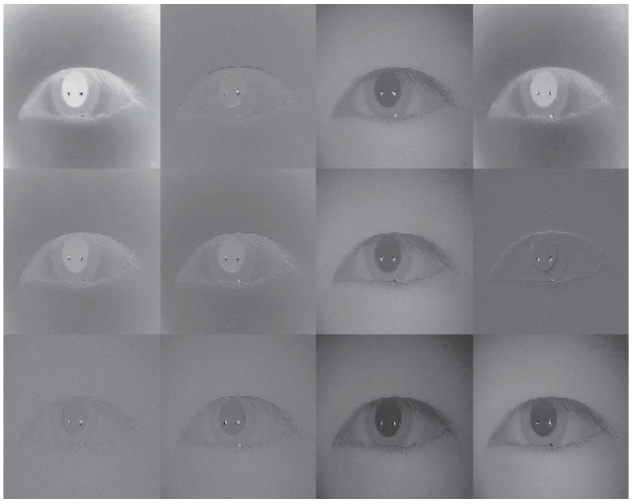

Figure 4: Convolutions using a 12-random-filter bank on a sample iris image.

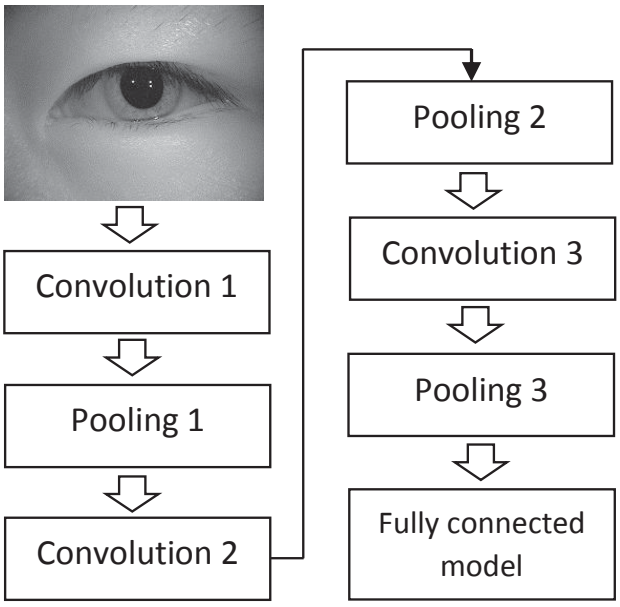

Figure 5: Architecture of the proposed neural network. Convolution and max-pooling are performed in a recursive manner three times.

The following subsections will introduce the architecture of the network and parameter optimization that optimizes the network's performance on the datasets.

2.2. Proposed Approach

The neural network architecture we used is based on Vedaldi et al. [16] that has the following structure: The first convolution layer has a 20-random-Gaussian-filter bank with 5×5 window size. The stride of the first layer is set to 1. The second layer is a 2×2 max pooling layer to extract important information from the feature map. Then, we repeat the convolution and max pooling layers 2 more times. An overview of the architecture of the network proposed is shown in Figure 5.

2.3. Parameter Optimization

In the pre-processing step, all images are down-scaled to 32×32. The color images are converted to grey scale. For each database, we selected 50% images for training/validation and the rest for testing. There are four parameters that may affect classification accuracy in the proposed network: batch size (p_1), epoch number (p_2) that determines the maximum number of iterations, momentum (p_3) that updates the current weight by adding a fraction of the previous one, and learning rate (p_4) that controls the changes of the size of weight and bias [17]. We empirically ran a set of experiments on each database, where first, the coarse and then a fine tuning process is performed. The purpose of the coarse is to evaluate the potential range of each parameter that produces reasonable classification accuracy, while fine tuning scrutinizes the combinations of the four parameters within a potential range so that an acceptable parameter set can be selected.

2.4. Datasets

There are four datasets used for the evaluation of the proposed network: NIST ICE 2005, CASIA Lamp, LG, and PLR.

- The NIST ICE 2005 is the dataset used for Iris Challenge Evaluation conducted by the National Institute of Standards and Technology. The data was collected by the Computer Vision Research Lab at the University of Notre Dame using an LG 2200 iris imaging system [14]. There are 1,528 and 1,425 left and right eye images respectively. Figure 6 shows some example images in the NIST ICE dataset.

- The CASIA iris image database was collected by the Chinese Academy of Science Institute of Automation for iris recognition [1]. We specifically selected the CASIA Lamp for our experiment because it simply shows one eye profile excluding other objects such as glasses in the CASIA Thousand. Collected by a handheld iris sensor produced by OKI, the CASIA Lamp has 411 subjects, each of which has 20 images for both the left and the right eye.

- The LG dataset is a private dataset collected using LG ICAM 4000. The sensor was mounted on a steel frame and placed on an adjustable height table in order to facilitate data capture from individuals of varying height. Data was obtained from 1099 participants over a pe-

Database	Left eye	Right eye	Occlusion
NIST ICE	1,528	1,425	Yes
LG	10,980	10,980	Yes
CASIA Lamp	8,131	8,080	No
PLR	62,579	63,355	No

Table 1: Number of left and right eye images and occlusion information in the databases applied in this work.

(a) Left eye images

(b) Right eye images

Figure 6: Example iris images in the NIST ICE database.

riod of 16 months. Five sessions of left/right iris images were collected from each participant during one visit. The LG sensor captures two images per eye from both a left and right illuminator (10 images per participant), resulting in a dataset consisting of 10,980 left and 10,980 right iris images.

- The Pupil Light Reflex (PLR) iris database was collected by Crihalmeanu and Ross [6] at West Virginia University including 54 subjects with different iris colors, ethnicities and age. They collected videos of ocular regions under varying illumination conditions including using ramp and flash light stimuli. We selected all the frames of each subject in the database for evaluation.

Table 1 compares the databases used in this work. There are around 11,000 and 8,100 images for both left and right eye in the LG and CASIA Lamp respectively. We included around 125,000 iris images in total from the PLR dataset due to the large amount of frames of each subject available. Occlusions exist in the NIST ICE and LG datasets, and there are no occlusions in the CASIA Lamp or PLR datasets. Such a diverse set of iris databases (small or large, with or without occlusion) provides us the opportunity to fully test our network's performance.

Method	% Accuracy (left)	% Accuracy (right)
Segmentation [12]	93.9	92.6
ASM [3]	62.5	67.7
Tear-duct detection [2]	69.9	
CNN	98.9	98.5
CNN (flipped)	97.5	98.6

Table 2: Comparison of classification accuracy on the NIST ICE database.

3. Experiments

Initially the network only produced 71.9% and 69.1% classification accuracy of left and right eyes in the NIST ICE dataset. Then, in the coarse tuning step, we found that the epoch number is not a significant factor that affects performance, and therefore we kept it fixed at 80. The meaningful range of the remaining three parameters are as follows: $p_2 \in [14,\ 20]$, $p_3 \in [0.85,\ 0.92]$, $p_4 \in [0.0005,\ 0.0012]$. The fine tuning was conducted within the above range, with two replications per parameter setting and 512 observations in total. For each replication, we randomly selected half of the images in both the left and right eye folders for training/validation and used the rest for testing. The classification accuracy for each class was recorded. We observed that the highest accuracy can be achieved with $p_1 = 80$, $p_2 = 18$, $p_3 = 0.92$, $p_4 = 0.0012$.

We compared our result with three other referenced works on the NIST ICE dataset and illustrated in Table 2. For better evaluation of the proposed network architecture, we flipped the images and performed the experiment again. We achieved 98.9% and 98.5% classification accuracy respectively on the original images of the NIST ICE dataset, and 97.5% and 98.6% on the flipped images.

We were also interested in whether the proposed network will have the same performance on large-scale datasets. CASIA Lamp, LG and PLR databases are all subject-based: all collected images of each subject are put in one single folder. We first ran the experiment with overlapping of one subject's iris images in training and testing as a baseline. Table 3 shows the results. The classification accuracy is very promising (above 99%).

Finally, the proposed network was evaluated in the same manner but without overlapping. For example, we randomly selected all images from 27 subjects in the PLR dataset for training and the images from the rest 27 subjects for testing. The result is shown in Table 4. We noticed a minor degradation in accuracy as compared with results generated with overlapping. Overall, the proposed network produces very desirable results.

Dataset	% Accuracy (left)	% Accuracy (right)
LG	99.8	99.8
LG (flipped)	99.6	99.9
CASIA	99.6	99.3
CASIA (flipped)	99.8	99.6
PLR	99.9	99.6
PLR (flipped)	99.9	99.8

Table 3: Classification accuracy with overlapping of one subject's iris images in testing.

Dataset	% Accuracy (left)	% Accuracy (right)
LG	99.6	99.8
LG (flipped)	99	99.5
CASIA	98.6	99.2
CASIA (flipped)	98.8	98.8
PLR	98.9	100
PLR (flipped)	99.9	98.1

Table 4: Classification accuracy without overlapping of one subject's iris images in testing.

4. Conclusion

We have shown our proposed Convolutional Neural Network (CNN) system designed and developed to automatically classify left vs. right iris images with classification performance ranging from 97.5 and up to 100%. The proposed method has the following advantages. First, it is not limited in the type or size of the dataset used. Second, it is not limited on the occlusion of the tear-duct or lateral angle. Third, it is a unified architecture, i.e., it can be applied on alternative datasets without redesigning its architecture. With such capability, the majority of the potentially mislabeled iris images by iris recognition system operators can be reliably detected. Another advantage is that our training and testing processes are fully-automated, when compared with previous works using ASM or AdaBoost that require manual intervention. Thus, our approach saves design and computational time, especially when dealing with large-scale datasets. Our initial results are very promising, and we are planning to investigate whether we can train a general robust CNN classifier with multiple datasets. The expectation is to avoid training and make our approach more efficient.

References

[1] Casia iris image database. http://biometrics.idealtest.org/.

[2] R. Abiantun and M. Savvides. Tear-duct detector for identifying left versus right iris images. In *the 37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–4, Oct 2008.

[3] S. Bhat and M. Savvides. Evaluating active shape models for eye-shape classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5228–5231, 2008.

[4] M. J. Burge and K. Bowyer. *Handbook of iris recognition*. Springer Science & Business Media, 2013.

[5] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber. Convolutional neural network committees for handwritten character classification. In *the IEEE International Conference on Document Analysis and Recognition (ICDAR)*, pages 1135–1139, 2011.

[6] S. Crihalmeanu and A. Ross. Pupil light reflex in the context of iris recognition–data acquisition setup. *West Virginia University, Tech. Rep*, 2014.

[7] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, Jan 2004.

[8] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014*, pages 184–199. Springer, 2014.

[9] P. Ghosh and M. Rajashekharababu. Authentication using iris recognition with parallel approach. *International Journal of Computer Science and Network Security (IJCSNS)*, 13(5):87–93, 2013.

[10] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968.

[11] Y. LeCun, K. Kavukcuoglu, C. Farabet, et al. Convolutional networks and applications in vision. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 253–256, 2010.

[12] Y.-H. Li and M. Savvides. *Automatically Identifying Mislabelled Iris data: Case of Left vs. Right Irises*. Springer US, 2009.

[13] D. M. Monro, S. Rakshit, and D. Zhang. Dct-based iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):586–595, 2007.

[14] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):831–846, 2010.

[15] P. Y. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the 7th IEEE International Conference on Document Analysis and Recognition*, pages 958–963, 2003.

[16] A. Vedaldi and K. Lenc. Matconvnet – convolutional neural networks for matlab. In *Proceeding of the ACM Int. Conf. on Multimedia*, pages 689–692, 2015.

[17] Wikibooks. Artificial neural networks — wikibooks, the free textbook project, 2015.

[18] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 113–120, 2013.