DEEP LEARNING BASED OFF-ANGLE IRIS RECOGNITION

Ehsaneddin Jalilian, Georg Wimmer, Andreas Uhl[*]

University of Salzburg
Department of Computer Science
Jakob-Haringer-Straße 2, 5020, Salzburg, Austria

Mahmut Karakaya

Kennesaw State University
Department of Computer Science
Marietta,GA,USA,30060

ABSTRACT

Even with trained operators and cooperative subjects, it is still possible to capture off-angle iris images. Considering the recent demands for stand-off iris biometric systems and the trend towards "on-the-move-acquisition", off-angle iris recognition became a hot topic within the biometrics community. In this work, CNNs trained with the triplet loss function are applied to extract features for iris recognition. To analyze which parts of the eye are most suited for the CNN-based recognition system, experiments are carried out using image data from different parts of the eye (full eye, eye zoomed to iris, iris only, iris normalized, eye without iris). To analyze the impact of different gaze angles on the recognition performance, experiments are applied on: (1) different gaze angles separately, (2) image data with increasing differences in the gaze angles, and (3) corrected off-angle image data. The experiment results show superior performance of the CNN trained with the triplet loss on the iris images with more lateral gaze angles ($\geq 30°$). However, higher differences in the gaze angles between images deteriorate the network performance. Also, the results are about the same for the different parts of the eye and correcting the gaze angle did not really improve the performance of the CNN.

Index Terms— Deep-learning, CNN, Iris recognition, Off-angle iris recognition, Triplet loss

1. INTRODUCTION

Iris recognition is one of the most reliable and popular biometric techniques. Under ideal (constrained) image acquisition conditions, iris recognition achieves precise recognition/identification results with very low false acceptance rates. However, in the case of none-ideal (relaxed) conditions, the eye images contain off-angle related distortions such as corneal refraction, distorted three-dimensional iris textures, missing iris boundaries and non-circular iris shapes in addition to common noises such as specular/light reflections, non-uniform illumination and low contrast. Off-angle iris recognition focuses on addressing these challenges.

This work aims to utilize the power of Convolutional Neural Networks (CNNs) to learn and extract iris features from off-angle eye image data. The proposed CNN approach is trained using the triplet loss function [1]. The key advantage of the proposed approach is that the CNN learns to extract features from the eye for subject recognition instead of applying a direct class assignment. In this way, the CNN can be applied to subjects that were not registered during the training process, contrary to CNNs trained with common loss functions like the SoftMax loss. In our experiments, we analyze the impact of different gaze angles on the recognition performance. We further examine which features and parts of the eye are most suited for the triplet loss based iris recognition by using image data from different parts of the eye (full eye, eye zoomed to iris, iris only, iris normalized, eye without iris). Furthermore, we analyze if correcting the off-angle images by transforming them to the frontal view can compensate for some off-angle distortions. Finally, we compare the recognition performance of the proposed approach against some state-of-the-art off-angle iris recognition algorithms.

2. RELATED WORK

Several different techniques have been proposed to address the off-angle iris recognition problem, *e.g*. [2] [3]. Gangwar *et al*. segmented the eye in the polar space using adaptive filters and then refined the segmentation outcome in the Cartesian space [4]. We further denote this approach as 'Irseg'. Uhl and Wild proposed the WAHET (Weighted adaptive hough and ellipsopolar transforms) algorithm, which uses an adaptive Hough transform at multiple resolutions to estimate the approximate position of the iris center [5]. Jalilian and Uhl utilized fully convolutional encoder-decoder networks to segment the iris pixels in images acquired in a wide set of heterogeneous conditions, including off-angle images [6]. Based on this approach, they proposed an end-to-end off-angle iris recognition pipeline utilizing the Refinenet [7] network in their later work [8]. Using the segmented output, iris features were extracted using quadrature 1-D Gabor wavelets. We further denote this approach as 'Segmentation-CNN'. Previous

[*]This project was partly funded from the FFG KIRAS project AUTFingerATM under grant No. 864785 and the FWF project "Advanced Methods and Applications for Fingervein Recognition" under grant No. P 32201-NBL.

CNN based approaches for iris recognition used off-the-shelf CNN features [9], capsule network architectures [10] or iris information from both eyes [11]. Ahmed and Fuller used the triplet loss for iris recognition using segmented iris images [12]. Zhao and Kumar proposed the Extended Triplet Loss (ETL) function, which is specifically developed to incorporate the bit-shifting and non-iris masking in normalized iris images [13]. In [14], a CNN trained with the SoftMax loss was applied for off-angle iris recognition using basically the same image data as in this work. The experimental setup was to analyze the classification performance (ROC curve) of the CNN for different parts of the eye using training data from either all gaze angles or only the frontal view. Because of the SoftMax loss training, the CNN recognition system in [14] can be applied only to subjects that were used for the training of the CNN. So, compared to [14] the main differences are that we use a different CNN training, different experiments and additionally we employ gaze angle correction.

3. METHODOLOGY

In this work we aim to address four main questions:

Q1: Are different gaze angles easier or harder for iris recognition systems? To find out if iris images with extreme gaze angles are harder to recognize than iris images that are less off-angle or entirely not off-angle (frontal view), the EER is computed separately for the images of 11 different gaze angles ($-50°, -40°, \ldots, +40°, +50°$). That means for each gaze angle, only similarity scores between images of the considered gaze angle are computed for the EER.

Q2: How tolerant are iris recognition systems to off-angle iris data? To find out the impact of differences in the gaze angle between images on the results of recognition systems, we compute the EER using only similarity scores between images with a maximum gaze angle difference of θ with $\theta \in \{0°, 10°, 20°, 30°, 40°\}$. For example if $\theta = 30°$, that means we employ all similarity scores between image pairs with a maximum gaze angle difference of $30°$ (*e.g.* $+50°$ and $+20°$ or $-20°$ and $-40°$ but not for image pairs with gaze angles of *e.g.* $-30°$ and $+10°$).

Q3: Which parts of the eye work best for the triplet loss based CNNs? Contrary to classical eye recognition systems that segment and normalize the iris, CNNs may also use information from other parts of the eye than only the iris, like *e.g.* eye shape, eye brows, eye wrinkles, skin texture and so on. In order to find out which parts of the eye can be used for subject recognition with the triplet loss CNN, we carry out experiments using image data from different parts of the eye. We use the following image data in our experiments: (1) full eye images, (2) images zoomed to the iris, (3) images with only the iris, (4) images where the iris is removed and (5) images of the normalized iris. Fig. 1 shows exemplary images.

Q4: Does gaze angle correction improve the results? We aim to find out if it is beneficial to correct the image gaze angles by bringing them back to the frontal view.

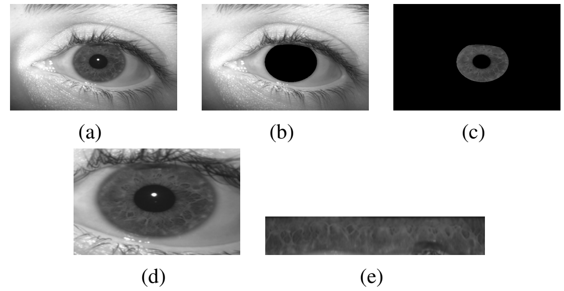

Fig. 1: Exemplary images of: Full eye (1a), No iris (1b), Iris only (1c), Zoomed iris (1d), and Iris Normalized (1e).

3.1. Off-angle iris recognition using triplet loss CNNs

The triplet loss [1] requires three input images at once (a so called triplet), where two images belong to the same class (the so called Anchor image and a sample image from the same class, further denoted as Positive) and the third image belongs to a different class (further denoted as Negative). The triplet loss trains the network to minimize the distance between the Anchor and the Positive and maximize the distance between the Anchor and the Negative. The triplet loss using the squared Euclidean distance is defined as follows:

$$L_{(A,P,N)} = \max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0),$$

where A is the Anchor, P the Positive and N the Negative. α is a margin that is enforced between the positive and negative pairs and is set to $\alpha = 1$. $f(I)$ is an embedding (the CNN output) of an input image I. Summarized, this means the CNN is trained to create an embedding $f(I)$, such that the Euclidean distances between embeddings of images from the same class (eye) is small, whereas the Euclidean distance between embeddings of any pairs of images from different eyes is large. To specifically train the CNN to tolerate differences in the gaze angle between images, we allow differences in the gaze angle between the Anchor and Positive and the Anchor and Negative with a maximum of $10°$. It turned out that using only small differences in the gaze angle ($10°$) is enough to make the CNN more off-angle robust and works better than using higher gaze angle differences between the three input images. For CNN training we employ the hard triplet selection [1] (only those triplets are chosen for training that actively contribute to improving the model). As net architecture we use the Squeeze-Net [1] (SqNet) [15]. SqNet is a small neural networks that is specifically created to have few parameters and only small memory requirements. The size of the CNN's last convolutional filter layer is adapted so that a 256-dimensional output vector (embedding) is produced. The implementation of the network was realized in PyTorch. The CNNs are trained for 400 epochs with the ADAM optimizer, starting with a learning rate of 0.001 . One major problem with CNN classifiers (using common loss functions like the SoftMax loss) utilized for biometric recognition is that the CNNs are only able to identify those subjects which have

[1] pytorch.org/vision/0.8/_modules/torchvision/models/squeezenet.html

been used for the training of the neural network. If new subjects are added in a biometric application system, then the nets need to be trained again since the new subjects can only be classified as one of the subjects that were used for training. This problem can be avoided using the triplet loss function. During training, data augmentation is applied by resizing the input images to a size of 234×234 followed by extracting patches of size 224×224 at a random position of the resized image (± 5 pixels in each direction).

CNN training is the same for all experiments. For CNN training, we use image data across all gaze angles ($-50°$ to $+50°$) to improve the off-angle robustness of the CNNs and also to have more available training data. The only limitation is that the gaze angle difference of the three images of a triplet must not exceed $10°$. The details of the employed image data in the experiments are specified in Section 3.3. Distances between feature vectors of images of the evaluation data are measured using the Euclidean distance d. To transform the Euclidean distance to a similarity metric, the Euclidean distances are inverted ($d \rightarrow 1/d$) and normalized so that the resulting similarity values range from zero to one.

3.2. Gaze angle correction

The main issue with off-angle iris images is the related distortions such as: 3D structural changes, missing iris boundaries, and perspective and refraction distortions, which erode and deform the geometric profile of the iris. So, correcting the images (bringing them back to frontal view) may help to correct these distortions and improve the performance of iris recognition systems. To investigate this, we additionally extend our experiments to gaze angle corrected image data. For gaze angle correction, we apply the approach presented in [8]. The gaze angles are determined using various measures of the eye that are unique for different gaze angles and then the images are re-projected to the frontal view.

3.3. Experimental Framework

Dataset: For our experiments we used 4400 iris images captured from 40 subjects of an off-angle iris database [16]. The iris images are captured by two near-infrared sensitive IDS-UI-3240ML-NIR cameras. Images at $0°$ gaze angle were captured by a frontal fixed camera, and off-angle images were captured by a frontal moving camera rotating horizontally from $-50° \; to + 50°$ in angle with a $10°$ step-size. Each camera captured 10 (gray scale) iris images per stop, giving 10 frontal and 100 off-angle iris images per subject.

Comparison methods: To compare our proposed approach with the state-of-the-art, the experiments are carried out with three methods from previous publications on the same database (using the Full eye images). The IrisSeg [4] algorithm and the WAHET [5] algorithm are two classical methods, and the Segmentation-CNN is a deep learning based method [8]. The technical details of these algorithms have already been explained in Section 2. We used the implementations of these methods as provided by the Iris Toolkit (USIT) [2] from the University of Salzburg.

Training approach: We employ 2-fold cross validation to train and evaluate the CNNs. For this, we divide the whole database into two equal parts (20 subjects per fold). In the first fold, one part is used as training data and the other one as evaluation data. In the second fold, the roles are switched.

Recognition metrics: To quantify the recognition performance the Equal Error Rate (EER) is calculated. Genuine and imposter scores are computed for all possible pairs of images of the evaluation data. The only limitations are made by the two different experiments, where either only similarity scores between images of the same gaze angle are taken into account (Q1), or only similarity scores between images with a certain maximum gaze angle difference (Q2). For the triplet loss CNN, there are different mappings of the images to the CNN output feature space for the CNNs of each fold, and therefore the EER has to be computed for each fold separately using only similarity scores between images of the evaluation data. Thus, we report mean EER over both folds.

4. RESULTS

In this section we present the results of our experiments to answer the four questions raised in Section 3. First, we present the results of the experiment to measure the impact of the gaze angles on the recognition performance (Q1). For this, the EER is computed separately for the images of each of the 11 different gaze angles. Fig. 2 presents the results across the different gaze angles for the triplet loss CNN using image data from different parts of the eye. As we can observe, more extreme gaze angles do not worsen the results compared to lower gaze angles. Furthermore, it does not really matter which parts of the eye are used, the results are always about the same. Yet interestingly, it seems that the results are slightly better when removing the iris information ('No iris') than for keeping it ('Full eye'). Fig. 4 shows the results obtained with and without gaze angle correction for the triplet loss CNN approach (image data zoomed to the iris) along with the results of the comparison methods. For the results of the triplet loss CNN, we can observe that correcting the gaze angles improves the results very slightly, but not consistently across all gaze angles. At more extreme gaze angles the triplet loss CNN achieves the best results. The comparison approaches perform worse at more extreme gaze angles. At lower gaze angles, Segmentation-CNN applied to the uncorrected image data achieves the best results.

Second, we present the results of the experiment to measure the off-angle robustness of the approaches (Q2). For the computation of the EER, we employed all similarity scores between image pairs with gaze angle differences that are up to $0°$, $10°$, $20°$, $30°$ and $40°$. Fig. 3 presents the triplet loss

[2] http://www.wavelab.at/sources/USIT

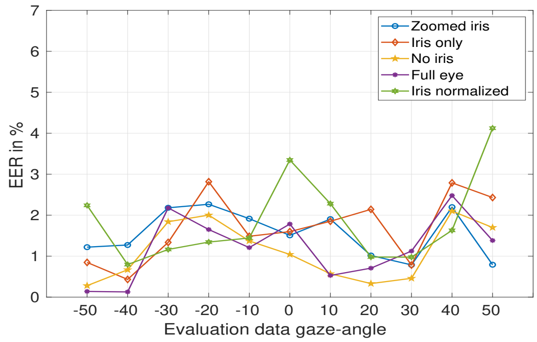

Fig. 2: Recognition results (EER in%) of the triplet loss CNN network. Only similarity scores between images of the same gaze angle are employed for EER computation

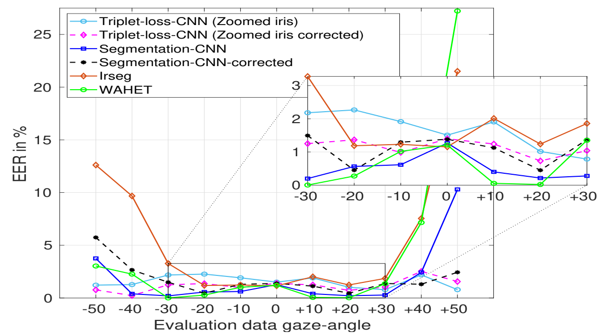

Fig. 4: Recognition results (EER in%) of the triplet loss CNN (with and without gaze angle correction) along with the comparison methods evaluated on the different gaze angle groups

CNN results for image data from different parts of the eye. As we can observe, the results are quite similar with respect to the used eye parts. Using the 'Zoomed iris' image data performs best and using 'Iris only' performs worst.

Fig. 5 shows the results obtained with and without gaze angle correction for the triplet loss CNN approach (image data zoomed to the iris) along with the results of the comparison methods. Interestingly, the triplet loss CNN performs slightly better on less off-angle images (up to 20°) using the corrected image data (Zoomed Iris corrected), while for the more extreme off-angle images (30° and 40°) slightly better results are achieved using the uncorrected data (Zoomed iris). So, we may conclude that the CNN is already off-angle robust due to its training with off-angle image data. The correction process actually deteriorates the results on higher gaze angles, where the correction process is the most difficult. The best results in this experiment are achieved using Segmentation-CNN on the corrected data. Thus, correcting the off-angle data can effectively improve the recognition performance for images that are acquired at different gaze angles, but not for our proposed method. Segmentation-CNN using uncorrected image data performs similar to the proposed approach and the two other methods perform worst.

5. CONCLUSION

In this work, CNNs trained with the triplet loss were used for iris recognition. The experiments were specifically designed to answer 4 questions (Q1-Q4) related to the use of off-angle iris image data. We showed that unlike the other comparison methods, the results of the proposed CNN approach did not decrease at stronger gaze angles and maintained an EER of around 2% across all gaze angles, making it a better choice when dealing with more extreme off-angle iris images ($\geq 30°$). The experiments yet showed that higher differences in the gaze angles between images deteriorate the results of the proposed CNN approach ($EER \approx 2\%$ at 0° difference and $EER \approx 8\%$ at 40° difference), but to a lesser extent than most of the comparison methods. In this case, the Segmentation-CNN combined with gaze angle correction proved to be a better choice. We also showed that it is not so important which parts of the eye images are used for subject recognition, as eventually the results remain similar. An interesting outcome of the experiments was that removing the iris information ('No iris') often gave better results than keeping it ('Full eye'). This, together with the fact that using image data from only the iris ('Iris only') often performed worst in the experiments, indicates that iris information is less suitable for subject recognition than information from other parts of the eye, at least for the proposed CNN approach. Finally, we showed that correcting the gaze angle did not really improve the CNN results, neither for the experiments on separate gaze angles nor for the experiments to test the off-angle robustness. However, the Segmentation-CNN method did clearly benefit from using rotation corrected data.

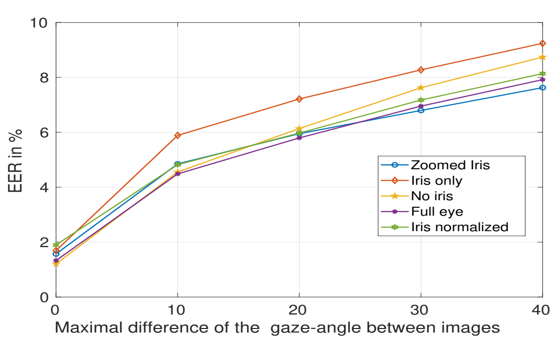

Fig. 3: Recognition results (EER in%) of the triplet loss CNN network using only similarity scores between images with maximal gaze angle differences between 0° and 40°

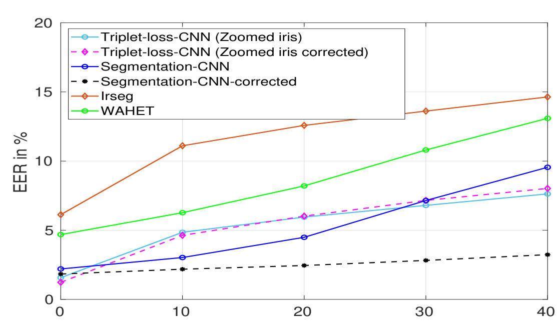

Fig. 5: Recognition results (EER in%) for the triplet loss CNN approach (with and without gaze angle correction) and the comparison approaches for an increasing maximum gaze angle difference between images

6. REFERENCES

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[2] John Daugman, "Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1927–1935, 2006.

[3] Samir Shah and Arun Ross, "Iris segmentation using geodesic active contours," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 824–836, 2009.

[4] Aabhishek Gangwar, Akanksha Joshi, Ashutosh Singh, Fernando Alonso-Fernandez, and Josef Bigunün, "Irisseg: A fast and robust iris segmentation framework for non-ideal iris images," *2016 International Conference on Biometrics (ICB)*, pp. 1–8, 2016.

[5] Andreas Uhl and Peter Wild, "Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation," in *2012 5th IAPR international conference on biometrics (ICB)*. IEEE, 2012, pp. 283–290.

[6] Ehsaneddin Jalilian and Andreas Uhl, "Iris segmentation using fully convolutional encoder–decoder networks," in *Deep Learning for Biometrics*, Ajay Kumar Bir Bhanu, Ed., chapter 6, pp. 133–155. Springer, (ZG) Switzerland, 2017.

[7] Guosheng Lin, Milan Anton, Shen Chunhua, and Ian Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5168–5177.

[8] Ehsaneddin Jalilian, Mahmut Karakaya, and Andreas Uhl, "Cnn-based off-angle iris segmentation and recognition," *IET Biometrics*, pp. 1–18, 2021.

[9] Kien Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan, "Iris recognition with off-the-shelf cnn features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.

[10] Tianming Zhao, Yuanning Liu, Guang Huo, and Xiaodong Zhu, "A deep learning iris recognition method based on capsule network architecture," *IEEE Access*, vol. 7, pp. 49691–49701, 2019.

[11] Alaa S. Al-Waisy, Rami S.R. Qahwaji, Stanley S. Ipson, Shumoos Al-Fahdawi, and Tarek A.M. Nagem, "A multi-biometric iris recognition system based on a deep learning approach," *Pattern Analysis and Applications*, vol. 21, pp. 783–802, 2018.

[12] Sohaib Ahmad and Benjamin Fuller, "Thirdeye: Triplet based iris recognition without normalization," *CoRR*, vol. abs/1907.06147, 2019.

[13] Zijing Zhao and Ajay Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3829–3838.

[14] Mahmut Karakaya, "Deep learning frameworks for off-angle iris recognition," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems*. IEEE, 2018, pp. 1–8.

[15] Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and $<$1mb model size," *CoRR*, vol. abs/1602.07360, 2016.

[16] Mahmut Karakaya, Del Barstow, Hector Santos-Villalobos, and Joseph Thompson, "Limbus impact on off-angle iris degradation," in *International Conference on Biometrics (ICB)*, 2013, pp. 1–6.