The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W9-2025
ISPRS Intl. Workshop “Photogrammetric and computer vision techniques for environmental and infraStructure monitoring, Biometrics and Biomedicine”
PSBB25 , 9–11 June 2025, Moscow, Russia

Iris image key points extraction based on handcrafted features in neural network

Elizaveta Andreeva[1], Elena Pavelyeva[2]

[1] Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 119991, Russia, Moscow, Leninskie Gory, MSU – elizabeth01@bk.ru
[2] Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 119991, Russia, Moscow, Leninskie Gory, MSU – pavelyeva@cs.msu.ru

Keywords: Key Points, Iris Recognition, Biometrics, Image Processing, Deep learning, Computer Vision.

Abstract

In this paper a neural network method for iris image key points detection based on handcrafted features using the Key.Net architecture is proposed. Due to the use of the handcrafted features in CNN, the proposed method combines the robustness of classical key points detection methods and high accuracy of neural networks. Additional Hermite-based convolutional filters are integrated into the network to improve keypoint localization. A synthetic dataset is generated from normalized iris images using geometric and photometric transformations. Matching of iris image key points is performed using HardNet descriptors, followed by geometric filtering and confidence-based ranking. Experimental evaluation demonstrates the robustness of the proposed method to the presence of eyelids and eyelashes without using any segmentation masks. The proposed approach achieves an Equal Error Rate (EER) value of 0.096% on the CASIA-IrisV4-Interval database. These results show the potential for the combination of handcrafted filtering with deep learning for accurate and interpretable iris recognition.

1. Introduction

Human iris identification is considered one of the most reliable approaches in biometrics. At the same time, the data collection process is quite simple and does not require special equipment, which makes iris biometrics promising for research.

Different iris recognition methods are based on both classical mathematical and neural network solutions. Classical mathematical methods are mostly based on handcrafted features that are focused on edge distribution (Alonso-Fernandez et al., 2009) or that are derived from Gabor (Daugman, 2009), Fourier (Vijaya Kumar et al., 2003), Hermite (Pavelyeva, 2013) or wavelet transforms (Ma et al., 2004). Furthermore, some approaches such as Difference of Gaussians (Lowe, 2004), Harris-Laplace and Hessian-Affine (Mikolajczyk and Schmid, 2004) use combinations of image derivatives to compute image feature maps, that is similar to the operations performed in the layers of trained convolutional neural networks (CNNs). Neural network methods are based on deep learning models for feature extraction, such as ResNet (Boyd et al., 2019), VGG (Minaee et al., 2016), DenseNet (Hafner et al., 2021), some methods also use attention mechanisms (Luo et al., 2021). Complex-valued networks (Nguyen et al., 2022) can extract both phase and amplitude information for improved feature quality. FeatNet (Zhao and Kumar, 2019) is designed specifically for iris recognition and performs detection, segmentation, and recognition in a unified pipeline using spatially corresponding features. Additionally, Capsule Networks have been explored for their ability to model spatial hierarchies in iris patterns (Kuhifayegh and Rajabi, 2025). Some deep learning methods do not require classical iris image normalization or precise image segmentation. For example, DeepIrisNet2 uses spatial transformer layers and dual CNN pipelines to extract robust features under non-ideal conditions (Gangwar et al., 2019).

One of the approaches to iris recognition feature extraction involves iris image key points detection. Often the key points and their descriptors are obtained using the scale-invariant feature transform (SIFT) algorithm. In (Quinn et al., 2021) the authors analyze the distribution, stability, and distinctiveness of key points and demonstrate that local feature-based matching can be effective for images captured at visible wavelengths. In (Rathgeb et al., 2019) the binary SIFT-based feature vectors from the key points descriptors are obtained. In (Alvarez-Betancourt and Garcia-Silvente, 2016) three detectors are used to identify distinctive key points: Harris-Laplace, Hessian-Laplace and Fast-Hessian, and three information sources of SIFT features are combined at matching score level. To construct the iris image key points descriptors, the phase congruency information can be used, since phase congruency is invariant to lighting and contrast (Protsenko and Pavelyeva, 2019). Key points descriptors can also be constructed using fractional phase congruency, calculated by the fractional wavelet transform (Protsenko and Pavelyeva, 2023).

This paper proposes a neural network method for detecting iris image key points based on handcrafted features using the Key.-Net architecture. Due to the use of the handcrafted features in CNN, the proposed method takes advantage of both classical key points detection methods and neural network approaches: robustness combined with high accuracy.

2. Key.Net Neural Network Application

In our work the Key.Net neural network architecture (Barroso-Laguna and Mikolajczyk, 2022) is used (Figure 7). The choice of Key.Net as the keypoint detector is motivated by its hybrid architecture and stability on limited datasets. It combines both handcrafted and learned CNN filters to highlight objects at different scale levels. Due to this, the model maintains a balance between quality and speed of operation and can obtain more consistent localization of iris image key points compared to pure learning-based detectors.

Key.Net model has three levels of scale for the input image,

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-7-2025

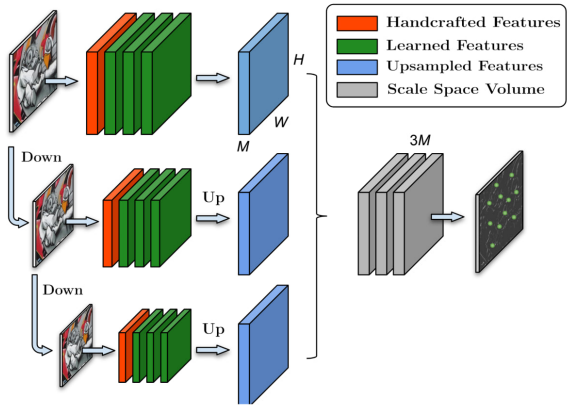

Figure 1. Key.Net architecture.

each with reduced discretization by a downsampling factor of 1.5. Handcrafted filters generate feature maps that are concatenated and processed through a set of learnable filters. The idea behind this multiscale approach is that the model processes images with different resolutions and generates a response map, where each pixel in the image is assigned an estimate of how likely it is a key point.

3. Dataset Preparation and Synthetic Pair Generation

The Key.Net training takes place on the basis of a synthetic dataset generated from normalized iris images. These images are obtained by preprocessing of the CASIA-IrisV4-Interval database images, that contains 2654 iris images from 249 individuals (Figure 2). The dataset is split into ~70% for the training set (1858 images, 158 individuals), ~10% for the validation set (261 images, 38 individuals), and ~20% for the test set (535 images, 53 individuals). Each class corresponds to a unique eye of a specific individual.

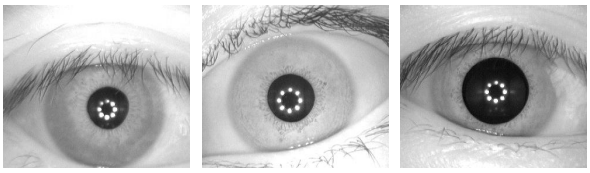

Figure 2. Some examples of iris images from the CASIA-IrisV4-Interval database.

Each image is normalized to a fixed size of 64×512 pixels using iris boundary segmentation and unwrapping (Figure 3). In this work only $3/5$ part of the height of the initial normalized image (closer to the iris pupil) is taken as the normalized iris image, since the areas around iris pupil have more sharp iris texture in the taken iris image database.

From the normalized dataset, we create synthetic training pairs by extracting small patches and applying random geometric transformations: scale $[0.5, 1.5]$, skew $[-0.3, 0.3]$, and rotation $[-5°, 5°]$ (Figure 4). These known transformations define ground truth mappings between pairs of patches, allowing us to obtain the fully supervised training of the detector. Areas without meaningful texture are discarded based on the response of handcrafted filters falling below a predefined threshold. To

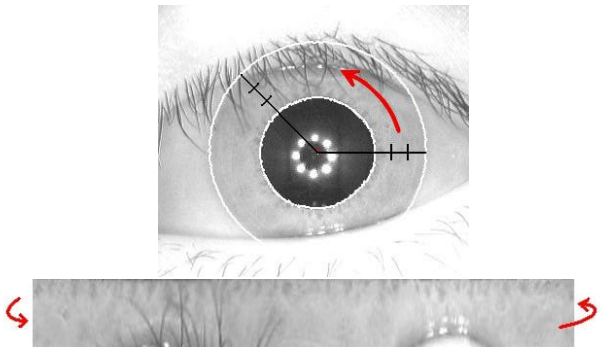

Figure 3. The iris image and iris normalization.

simulate real-world variability, additional photometric augmentations are applied to one image in each pair, including random changes in contrast and brightness.

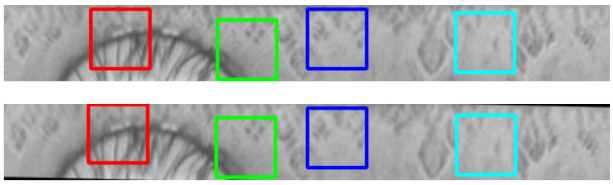

Figure 4. Synthetic dataset generation.

In Key.Net the different combinations of first and second image derivatives are used to create the handcrafted filters. A learned block consists of the convolutional layer with 8 filters, batch normalization layer and ReLU activation function. In addition, we incorporate the convolutions with Hermite transform filters to the handcrafted features to expand the representation of features. The Hermite transform functions show promising results for iris image key points detection (Pavelyeva, 2013), so these functions are integrated to our architecture.

The Hermite transform functions are given by:

$$\varphi_n^{(\sigma)}(x) = \frac{1}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma\sqrt{\pi}} e^{-(x/\sigma)^2} H_n\left(\frac{x}{\sigma}\right), \quad n = 0, 1, 2, \ldots, \tag{1}$$

where σ — scale parameter.

Hermite polynomials $H_n(x)$ are defined by recursion as follows:

$$\begin{aligned} H_0(x) &= 1; \\ H_1(x) &= 2x; \\ H_n(x) &= 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x). \end{aligned} \tag{2}$$

All Hermite transform functions for $n > 0$ have zero mean value and are Gaussian function derivatives up to a constant factor. Two-dimensional Hermite transform functions (Figure 5) can be represented as

$$\varphi_{m,n}^{(\sigma_x,\sigma_y)}(x, y) = \varphi_m^{(\sigma_x)}(x)\varphi_n^{(\sigma_y)}(y). \tag{3}$$

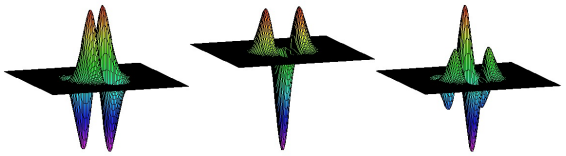

Figure 5. 2D Hermite functions: $\varphi_{1,1}$ (left), $\varphi_{2,0}$ (center), and $\varphi_{2,1}$ (right).

In Hermite transform the convolution of the image with the Hermite transform function is applied. The Hermite transform kernel configurations used in this article to extend the handcrafted features are as follows (Figure 6):

$$\varphi_{2,0}^{(3,3)}(x,y), \varphi_{0,2}^{(3,3)}(x,y), \varphi_{1,1}^{(3,5)}(x,y). \quad (4)$$

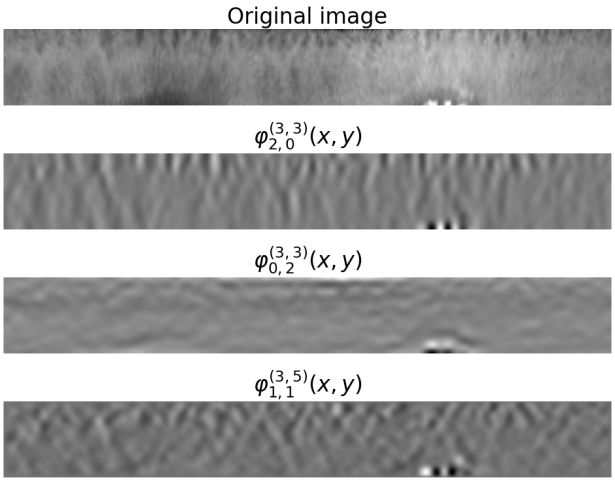

Figure 6. The normalized iris image and the results of applying 2D Hermite filters with 2D Hermite transform functions.

To identify key points across various scales, a Multi-scale Index Proposal (M-SIP) layer is employed. It is based on the Index Proposal (IP) layer, a differentiable module designed to extract key point coordinates from local maxima in a response map. The IP layer divides the response map into fixed-size windows and applies a spatial softmax operation within each window to estimate the position of a dominant key point as a weighted average of response scores. Given images I_a and I_b, and the ground truth homography $H_{b,a}$, which defines the geometric relationship between the image pairs, the loss $\mathcal{L}$ is based on the squared difference between the points extracted by IP layer and the actual maximum coordinates (NMS) in the corresponding windows of I_a and I_b:

$$\mathcal{L}_{IP}(I_a, I_b, H_{a,b}, N) = \sum_i \alpha_i \| [x_i, y_i]_a^T - H_{b,a}[\hat{x}_i, \hat{y}_i]_b^T \|^2,$$
$$\alpha_i = \mathcal{R}_a(x_i, y_i)_a + \mathcal{R}_b(\hat{x}_i, \hat{y}_i)_b, \quad (5)$$

where $[x_i, y_i]_a^T$ - coordinates of the i-th key point in I_a;
$[\hat{x}_i, \hat{y}_i]_b^T$ - corresponding key point coordinates in I_b;
$\mathcal{R}_a$ and $\mathcal{R}_b$ - response maps;
N - window size for key point extraction.

The multi-scale loss function is computed as the sum across all scaling levels:

$$\mathcal{L}_{MSIP}(I_a, I_b, H_{a,b}) = \sum_s \lambda_s \mathcal{L}_{IP}(I_a, I_b, H_{a,b}, N_s) \quad (6)$$

where s - scale level index;
λ_s - scaling factor decreasing with window area, balancing larger losses in larger windows;
N_s - window size at scale s.

Key.Net is trained using the Siamese process, where each key instance passes through the same neural network. This means that two versions of the same input — typically a patch and its geometrically transformed counterpart — are processed in parallel by identical network branches. The model is optimized to produce similar key point responses for corresponding locations in both images.

4. Data Post-processing

After receiving the model results, the data post-processing stage begins. Initial matches of key points are made based on the similarity of their descriptors, which are extracted using HardNet model (Mishchuk et al., 2017). HardNet is a descriptor extraction model based on a convolutional neural network trained to compare small patches around key points. It learns to produce feature vectors (descriptors) such that similar regions in different images are mapped to nearby points in descriptor space, while dissimilar regions are moving apart. The model is trained on patch pairs with known similarity labels, allowing it to learn discriminative representations useful for matching (Figure 7). The matching process is performed using a brute-force approach with cross-checking. Then the matches received are sorted by the distance between them, and only those whose distance is below the specified threshold value are accepted.

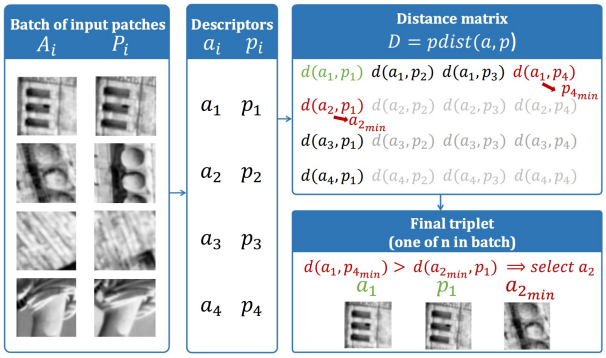

Figure 7. HardNet sampling procedure.

To refine the matches, geometric constraints are applied. If the distance between the key points of two iris images is less than a given threshold, they are considered to belong to the same iris image texture area. We assume that the spatial shift between the matched key points corresponds to an eye rotation angle of no more than ∼20 degrees. Therefore, only key points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ satisfying $|x_1 - x_2| \leq 30$ pixels (taking into account the cyclic shift of the normalized iris image, see Figure 3) and $|y_1 - y_2| \leq 5$ pixels are considered as potential matches. Then we find the most frequent horizontal shift between the matched key points – this shift determines

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-7-2025

the global shift between normalized iris images and the rotation angle between the eyes. The final matches are filtered taking into account the horizontal shift relative to the global one by no more than 5 pixels. These steps effectively remove outliers and apply global geometric constraints, which significantly improves the overall quality of matches.

The resulting key points are sorted by the $score$ value obtained for each key point by Key.Net. This value reflects the confidence level of the model at this key point. At the end, no more than 150 of the best key points are saved.

5. Results

The initial weights of the Key.Net model are obtained from publicly available pretrained parameters (Barroso-Laguna and Mikolajczyk, 2022). We fine-tuned the network on our synthetic dataset pairs generated from normalized iris images. Each patch is of size 40×40 pixels. The fine-tuning is performed using the Adam optimizer for 50 epochs with a learning rate of 0.0001 and a batch size of 32. After training, the model is evaluated on the test subset of the CASIA-IrisV4-Interval database.

Some results are shown in Figure 8 and Figure 9. These examples demonstrate the effectiveness of the proposed method in finding reliable keypoint correspondences for both genuine and impostor image pairs. Importantly, although no eyelid and eyelash masks are applied, the algorithm avoids matching points in these regions. This visual evidence indicates that the model is robust to the presence of eyelids and eyelashes.

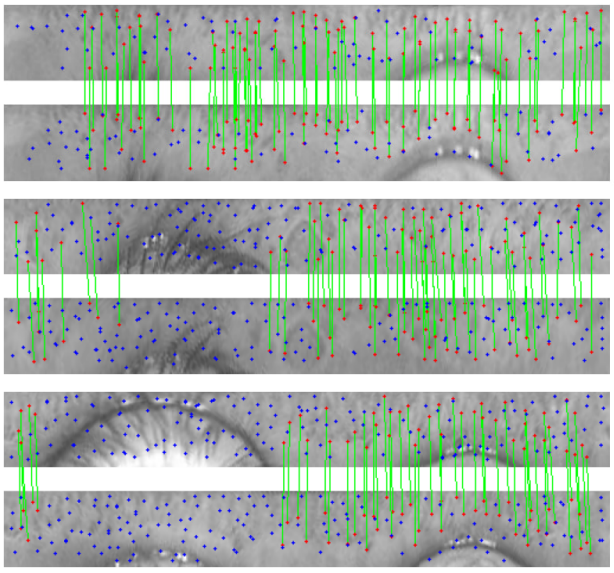

Figure 8. The examples of key points matching for two iris images of one eye.

The following biometric error values are used to evaluate the quality of the algorithm.

- False Acceptance Rate (FAR) — the probability that the system incorrectly matches the input pattern to a non-matching template in the database. It measures the percent of invalid inputs that are incorrectly accepted.
- False Rejection Rate (FRR) — the probability that the system fails to detect a match between the input pattern and a matching template in the database. It measures the percent of valid inputs that are incorrectly rejected.
- Equal Error Rate (EER) is the rate at which FAR and FRR are equal. In general, the lower the EER, the more accurate the biometric system is.

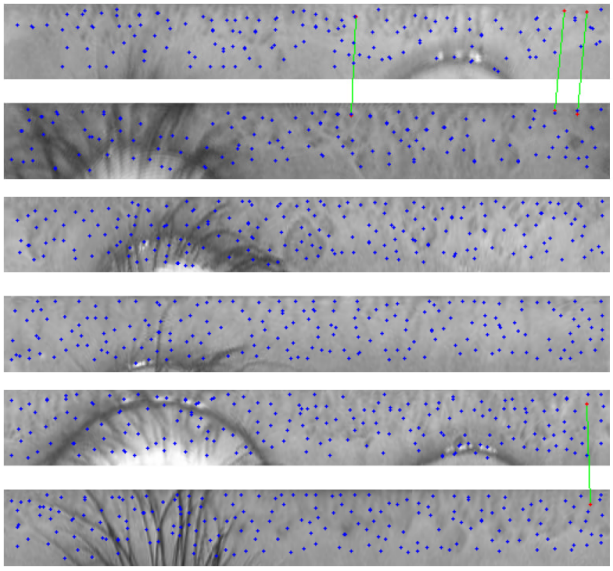

Figure 9. The examples of key points matching for two iris images of different eyes.

The False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER) values are used to evaluate the effectiveness of the proposed method. The similarity between two iris images is defined as the number of matched key points. This value is then compared to a threshold.

- If the similarity score is greater than or equal to the threshold ($\geq$) — the pair is accepted (same identity).
- If the similarity score is less than the threshold ($<$) — the pair is rejected (different identity).

In our experiments, the decision threshold $T = 6$ is selected as the optimal value where FAR and FRR are intersected.

Figure 10 and Figure 11 demonstrate the comparison between the baseline and Hermite-based algorithms. In the baseline algorithm only handcrafted features based on combinations of first and second image derivatives are used, while in Hermite-based algorithm the convolutions with Hermite transform filters are also used to create the handcrafted features. Figure 10 shows the distributions of genuine and impostor scores, and Figure 11 shows the corresponding FAR and FRR curves used to compute the EER.

- Baseline algorithm: EER=0.103%
- Hermite-based algorithm: EER=0.096%

The marginal improvement achieved by the Hermite-enhanced version highlights its superior discriminative power in iris feature extraction.

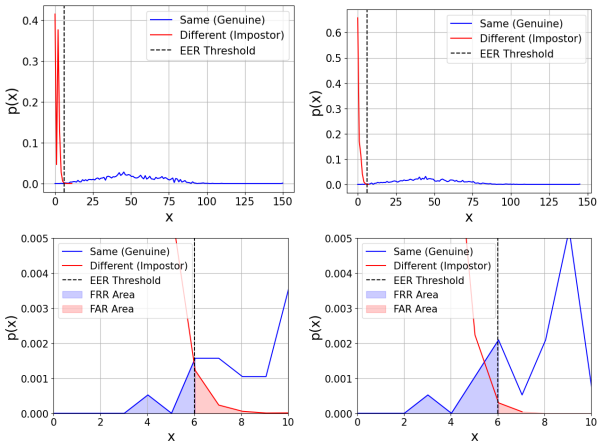

Figure 10. Distribution of genuine (blue) and impostor (red) scores. Left: baseline algorithm. Right: Hermite-based algorithm.

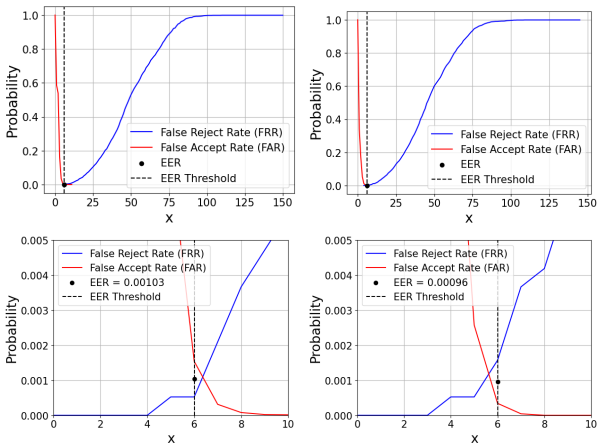

Figure 11. FAR and FRR curves used to calculate EER. Left: baseline algorithm. Right: Hermite-based algorithm.

As shown in Table 1, the proposed method outperforms several iris recognition techniques in terms of EER. All methods in the comparison are evaluated on the CASIA-IrisV3-Interval/CASIA-IrisV4-Interval dataset under the same preprocessing and evaluation protocol to ensure fairness.

6. Conclusions

In this paper the algorithm for effective iris image key points construction and matching is proposed. The algorithm is based on combination of handcrafted and learned features in Key.Net architecture. The proposed algorithm achieves fairly good results and can be promising for the use in biometric identification systems. The method remains robust to the presence of eyelids and eyelashes on iris images, as it avoids matching key points in these regions even without explicit masking.

Method	**EER (%)**
SIFT-based (Alvarez-Betancourt and Garcia-Silvente, 2016)	0.550
SIFT-based (Rathgeb et al., 2019)	0.225
LSC (Sadhya and Raman, 2019)	0.105
DeepIrisNet2-NIR (Gangwar et al., 2019)	0.280
Phase Congruency (Protsenko and Pavelyeva, 2019)	0.226
Dadcnet+VGG (Chen et al., 2022)	0.188
StandardR (Yassir et al., 2024)	0.169
HybridR (Mohammed et al., 2025)	0.161
Proposed	**0.103**
Proposed + Hermite Transform	**0.096**

Table 1. Comparison with existing iris recognition methods on CASIA-IrisV3/V4 datasets.

References

Alonso-Fernandez, F., Tome-Gonzalez, P., Ruiz-Albacete, V., Ortega-Garcia, J., 2009. Iris recognition based on sift features. *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, IEEE, 1–8.

Alvarez-Betancourt, Y., Garcia-Silvente, M., 2016. A keypoints-based feature extraction method for iris recognition under variable image quality conditions. *Knowledge-Based Systems*, 92, 169–182.

Barroso-Laguna, A., Mikolajczyk, K., 2022. Key. net: Keypoint detection by handcrafted and learned cnn filters revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 698–711.

Boyd, A., Czajka, A., Bowyer, K., 2019. Deep learning-based feature extraction in iris recognition: Use existing models, fine-tune or train from scratch? *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 1–9.

Chen, Y., Gan, H., Zeng, Z., Chen, H., 2022. DADCNet: Dual attention densely connected network for more accurate real iris region segmentation. *International Journal of Intelligent Systems*, 37(1), 829–858.

Daugman, J., 2009. How iris recognition works. *The essential guide to image processing*, Elsevier, 715–739.

Gangwar, A., Joshi, A., Joshi, P., Raghavendra, R., 2019. Deepirisnet2: Learning deep-iriscodes from scratch for segmentation-robust visible wavelength and near infrared iris recognition. *arXiv preprint arXiv:1902.05390*.

Hafner, A., Peer, P., Emeršič, Ž., Vitek, M., 2021. Deep iris feature extraction. *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE, 258–262.

Kuhifayegh, F., Rajabi, R., 2025. Iris Recognition via Deep Learning Using Capsule Networks with Enhanced Routing Algorithm. *Intelligence*, 1(2), 91–99.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60, 91–110.

Luo, Z., Li, J., Zhu, Y., 2021. A deep feature fusion network based on multiple attention mechanisms for joint iris-periocular biometric recognition. *IEEE Signal Processing Letters*, 28, 1060–1064.

Ma, L., Tan, T., Wang, Y., Zhang, D., 2004. Efficient iris recognition by characterizing key local variations. *IEEE Transactions on Image processing*, 13(6), 739–750.

Mikolajczyk, K., Schmid, C., 2004. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60, 63–86.

Minaee, S., Abdolrashidiy, A., Wang, Y., 2016. An experimental study of deep convolutional features for iris recognition. *2016 IEEE signal processing in medicine and biology symposium (SPMB)*, IEEE, 1–6.

Mishchuk, A., Mishkin, D., Radenovic, F., Matas, J., 2017. Working hard to know your neighbor’s margins: Local descriptor learning loss. *Advances in neural information processing systems*, 30.

Mohammed, A. H. Y., Dziyauddin, R. A., Kamaruddin, N., Rahim, F. A., 2025. A hybrid ranking algorithm for secure and efficient iris template protection. *Computers & Security*, 150, 104216.

Pavelyeva, E., 2013. The search for matches between the iris key points using Hermite projection phase-only correlation method. *Systems and Means of Informatics*, 23(2), 74–88.

Protsenko, M., Pavelyeva, E., 2019. Iris image key points descriptors based on phase congruency. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 167–171.

Protsenko, M., Pavelyeva, E., 2023. Fractional Wavelet Transform Phase for IRIS Image Key Points Matching. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 201–207.

Quinn, G., Matey, J., Grother, P., Watters, E., 2021. Statistics of visual features in the human iris. *NIST Technical Note*, 932452.

Rathgeb, C., Wagner, J., Busch, C., 2019. SIFT-based iris recognition revisited: prerequisites, advantages and improvements. *Pattern Analysis and Applications*, 22, 889–906.

Sadhya, D., Raman, B., 2019. Generation of cancelable iris templates via randomized bit sampling. *IEEE Transactions on Information Forensics and Security*, 14(11), 2972–2986.

Vijaya Kumar, B., Xie, C., Thornton, J., 2003. Iris verification using correlation filters. *International Conference on Audio-and Video-Based Biometric Person Authentication*, Springer, 697–705.

Yassir, M. A. H., Dziyauddin, R. A., Kamaruddin, N., Ahmad, N., 2024. A standard ranking algorithm for robust iris template protection. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(2), 1214–1225.