A. Brömme, C. Busch, A. Dantcheva, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018 

# Visible Wavelength Iris Segmentation: A Multi-Class Approach using Fully Convolutional Neuronal Networks

D. Osorio-Roig[1], C. Rathgeb[2], M. Gomez-Barrero[2], A. Morales-González[1],
E. Garea-Llano[1], C. Busch[2]

**Abstract:** Iris segmentation under visible wavelengths (VWs) is a vital processing step for iris recognition systems operating at-a-distance or in non-cooperative environments. In these scenarios the presence of various artefacts, e.g. occlusions or specular reflections, as well as out-of-focus blur represents a significant challenge. The vast majority of proposed iris segmentation algorithms under VW aim at discriminating the iris and non-iris regions without taking into account the variability that is present in the non-iris region. In this paper, we introduce the idea of segmenting the iris region using a multi-class approach which differentiates additional classes, e.g. pupil or sclera, as opposed to commonly employed bi-class approaches (iris and non-iris). Experimental results conducted on two publicly available databases show that the use of the proposed multi-class approach improves the iris segmentation accuracy. Simultaneously, it also allows for the segmentation of different non-iris regions, e.g. glasses, which could be employed in further application scenarios.

**Keywords:** Biometrics, iris recognition, semantic segmentation, fully convolutional networks.

## 1 Introduction

More than 20 years after the first iris recognition algorithm was developed by Daugman [Da93], iris biometrics remains an active and rapidly growing field of research. Among others, iris recognition based on visible wavelength (VW) images represents one major challenge. Independent tests [Ph08, Gr12] have confirmed remarkable recognition accuracy for iris recognition for good quality near-infrared (NIR) images. If captured properly, VW iris images might exhibit a similar level of detail within the iris texture, while in general the amount of available information is expected to be less [Pr16]. For instance, within VW iris images possible artefacts, such as specular reflections or shadows, are more pronounced, which generally leads to an increased intra-class variation. The segmentation of the iris involves a detection of inner and outer iris boundaries, a detection of eyelids, an exclusion of eyelashes as well as contact lense rings, and a scrubbing of specular reflections [Da04]. Accurate segmentation of the iris region in VW images represents one of the most critical tasks, since errors in the pre-processing stage significantly impact recognition accuracy [PA10].

In the first part of the Noisy Iris Challenge Evaluation (NICE.I) [PA12] optimised versions of the integrodifferential operator proposed by Daugman, e.g. [THS10, Sa10], were among the best-ranked approaches. More recent works demonstrated that a pixel-wise

[1] Advanced Technologies Application Center (CENATAV), La Habana, Cuba
{dosorio,egarea,amorales}@cenatav.co.cu
[2] da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{christian.rathgeb,marta.gomez-barrero,christoph.busch}@h-da.de

classification of iris and non-iris regions based on machine learning concepts further improves segmentation accuracy [Li16]. Among iris segmentation methods reported in the literature, the ones based on Fully Convolutional Networks (FCNs) have achieved the best results [Li16, JU17]. Jalilian and Uhl [JU17] explored three FCNs, which presented problems when dealing with less constrained iris images such as off-angle images recorded from various distances with different types of occlusions (including glasses). The main limitation of these methods is that they only take into account the pixel-wise prediction of iris and non-iris regions. Similarly, Liu et al. [Li16] depicted several image examples, e.g. images from dark skinned subjects or images not containing eyes, for which their proposed method revealed poor results. According to the authors, those results were due to limited training samples with these characteristics in the NICE.I dataset. To address these problems, a semantic segmentation algorithm for segmenting the iris region was firstly introduced in [ORMGGL17]. In this work, the HMRF-PyrSeg algorithm [ORMGGL17], which was trained on 53 manually annotated eye images taken from the training set of the NICE.I database, was used to segment eye regions. Although experimental results did not improve over state-of-the-art segmentation methods, authors pointed out that the multi-class segmentation idea was a promising path for VW iris recognition.

In this context, the main contributions of this paper are: (1) an exploration of the hypothesis that VW iris segmentation based on a multi-class approach outperforms the common bi-class approach in terms of segmentation accuracy (iris and non-iris), applying knowledge transfer over two FCNs proposed for general semantic image segmentation; (2) a particular improvement of segmentation results for the aforementioned challenging types of VW iris images by using an increased number of semantic segmentation classes; (3) a new annotated database for semantic segmentation of VW iris images, which contains a total number of 10 semantic classes.

The remainder of this paper is organized as follows: in Sect. 2 semantic segmentation for iris segmentation is introduced. Sect. 3 describes the annotated multi-class database for iris segmenation. The experimental evaluation showing our results and a comparison with other state-of-the-art methods is presented in Sect. 4. Finally, conclusions and future work are summarized in Sect. 5.

## 2 Semantic segmentation for iris region localization

*Semantic segmentation* aims at assigning each pixel within an image to a pre-defined object class. In this work, the principle of semantic segmentation is applied for iris segmentation, i.e. pixels of iris images are assigned to diverse pre-defined classes, e.g. iris, pupil or sclera. For this purpose pre-trained FCNs are adapted via *transfer learning*. Transfer learning is the process of learning from a new tasks through the transfer of knowledge from a related task that has already been learned. This process can be classified into inductive learning or reinforcement learning [TS09]. In this work, inductive learning over two FCNs for segmenting several eye regions in multi-class approaches are introduced. Specifically, we select two FCNs for semantic segmentation[3], i.e. `fcn8s-at-once` and `fcn-alexnet` which were fine-tuned from the pre-trained VGG-16 [SZ14] and AlexNet

[3] FCNs for Semantic Segmentation: `https://github.com/shelhamer/fcn.berkeleyvision.org`

Tab. 1: Semantic classes in iris images used for VW iris segmentation

| Nr. | Semantic class | Colour | Nr. | Semantic class | Colour |
|---|---|---|---|---|---|
| 1 | iris | | 6 | eyebrows | |
| 2 | pupil | | 7 | periocular skin | |
| 3 | specular reflections | | 8 | hair | |
| 4 | sclera | | 9 | glasses frames | |
| 5 | eyelids / eyelashes | | 10 | background | |

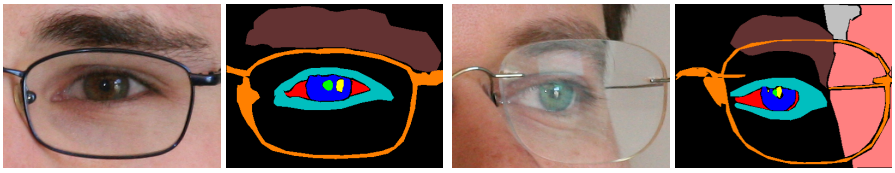

Fig. 1: Examples of iris images and their manual annotations

[KSH12] models, respectively. Both have been trained for general image segmentation of the PASCAL VOC database. The fine-tuning process aims at selecting the most promising hypothesis space which allows to adjust a target knowledge (i.e. eye segmentation) from a source knowledge (i.e. general image segmentation).

The main motivation of using semantic segmentation for iris segmentation is that the introduction of more semantic information is expected to prevent false positive classifications of non-iris pixels, i.e. due to semantic constraints robustness can be improved. Further, it is interesting to explore the learning capabilities of features all at once in different scales for bi-class and multi-class iris segmentation.

## 3 Pixel-wise annotated database

To obtain a ground truth for training classifiers using the semantic classes present in eye images, we manually annotated a training set of 500 eyes images from the NICE.I [Pr10] database. These annotations were built at pixel level[4], i.e. each pixel is annotated with its corresponding class number. Images were annotated with a maximum number of 10 different semantic classes listed in Tab. 1. To achieve comparability of segmentation accuracies, the annotated database[5] builds upon the bi-class ground truth iris segmentation masks available for the NICE.I database[6]. Classes were selected with the aim of obtaining the maximum semantic information which describes the different regions of an eye image. Examples of manual annotations can be seen in Fig. 1.

## 4 Experimental results

### 4.1 Experimental protocol

Five different class combinations were analysed to find the best subset of labelled classes for representing an eye image and, hence, increasing the inter-class variability as well as decreasing the intra-class variability of the iris region and the remaining non-iris regions.

[4] Object Labeling Tool: `http://dhoiem.cs.illinois.edu/software/`
[5] Multi-class iris segmentation ground truth: `https://dasec.h-da.de/research/biometrics/MCIS/`
[6] NICE.I iris segmentation ground truth: `http://nice1.di.ubi.pt/`

Tab. 2: Summary of different multi-class combinations used for semantic iris segmentation

| Combination | Semantic classes | Combination | Semantic classes |
|---|---|---|---|
| 2-classes (bi-class) | iris, non-iris | 4-classes-b | iris, pupil, sclera, non-iris |
| 3-classes | iris, pupil, non-iris | 5-classes | iris, pupil, sclera, specular reflections, non-iris |
| 4-classes-a | iris, pupil, specular reflections, non-iris | 10-classes | all defined semantic classes |

Tab. 2 lists all analysed combinations. The 2-classes combination represents the common bi-class approach. The 3-classes combination aims at analysing the discriminative capability of the pupil w.r.t. the remaining non-iris classes. Further, two variants of 4-classes combinations were built in order to investigate the impact of the specular reflections and sclera classes, respectively, while the 5-classes combination combines both of the latter classes with iris and pupil. Finally, all annotated classes are combined in the 10-classes combination.

Proposed methods are evaluated on two well-known databases, i.e. NICE.I and MobBIO. The NICE.I database is a subset of the UBIRIS v2 database [Pr10], whose images were acquired under non-ideal and non-cooperative conditions including images at-a-distance and on-the-move. It consists of 500 eye images for training and 445 eye images for testing. NICE.I images present problems of occlusions, specular reflections, off-angle and out-of-focus blur. The MobBIO multi-modal database [Se14] comprises 800 eye images, which were used for testing. MobBIO images were captured in two different lighting conditions, with variable eye orientations and occlusion levels. This database provides manual annotations for every image of both the limbic and pupillary contours.

As evaluation metric we consider the $E^1$ evaluation measure proposed by the NICE.I protocol. This metric estimates the proportion of correspondent disagreeing pixels:

$$E^1 = \frac{1}{N * m * n} \sum_{i,j \in (m,n)} G(i,j) \oplus M(i,j), \tag{1}$$

where $N$, $m$, $n$ are the number, length and width of test images, respectively. $G$ and $M$ are the ground truth and the generated iris mask respectively, and $i$, $j$ are coordinates in pixels of $G$ and $M$. The symbol $\oplus$ represents the XOR operation to assess the mismatching pixels between $G$ and $M$. To apply this metric for the proposed multi-class iris segmentation approaches all iris pixels are set to 1 and all pixels of non-iris classes are set to 0 after semantic segmentation. The resulting $E^1$ error values are in the range $[0,1]$.

As aforementioned, two FCNs were selected for evaluating the proposed approach. The implementations of `fcn-alexnet` and `fcn8s-atonce` architectures are based on Caffe [Ji14]. Different hyper-parameters were taken from [LSD15] and intermediate upsampling layers were initialized by a bilinear interpolation. The output number of each model was fixed according to the employed combination of classes.

### 4.2 Results and discussion

Tab. 3 lists $E^1$ segmentation errors for the proposed multi-class segmentation approaches on the NICE.I and the MobBIO database. Compared to the commonly employed 2-classes approach, most of the defined multi-class combinations achieved improved segmentation

Tab. 3: $E^1$ segmentation errors of proposed multi-class approaches on NICE.I and MobBIO (best results for each model on each database are marked bold)

| FCN model | Combination | $E^1(\%)$ NICE.I | $E^1(\%)$ MobBIO | FCN model | Combination | $E^1(\%)$ NICE.I | $E^1(\%)$ MobBIO |
|---|---|---|---|---|---|---|---|
| `fcn-alexnet` | 2-classes | 1.69 | 3.61 | `fcn8s-atonce` | 2-classes | 1.24 | 2.59 |
| | 3-classes | 1.52 | **3.19** | | 3-classes | 1.15 | 2.51 |
| | 4-clasess-a | **1.51** | 3.58 | | 4-classes-a | **1.18** | 2.70 |
| | 4-classes-b | 1.54 | 3.85 | | 4-classes-b | 1.16 | 2.52 |
| | 5-classes | 1.53 | 3.58 | | 5-classes | 1.16 | **2.46** |
| | 10-classes | 1.58 | 3.61 | | 10-classes | **1.13** | 2.50 |

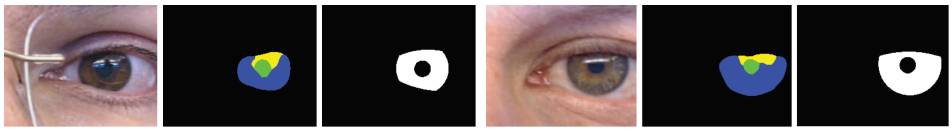

Fig. 2: Two examples of original images from the MobBIO database, segmentation masks obtained by the 4-classes-a combination (`fcn8s-atonce` model) and the available ground truth of iris and non-iris regions

errors by increasing the number of semantic classes for `fcn8s-atonce` and `fcn-alexnet` on both databases. Segmentation errors of multi-class combinations reported by `fcn8s-atonce` on NICE.I always decrease w.r.t. the bi-class combination, when the number of classes in the combination increases, while the `fcn-alexnet` segmentation errors were very similar and did not show an improvement. We believe that higher segmentation errors reported by `fcn-alexnet` result from the smaller architecture of the network and because it performs a very wide interpolation (32-pixels) in the up-sampling process, thereby causing that only higher level features such as lines and boundaries are learned. Due to these reasons, the 10-classes combination on `fcn-alexnet` showed the worst $E^1$ segmentation result w.r.t. the other multi-class combinations.

The inclusion of the pupil class in the 3-classes combination allowed keeping the spatial relation of the learned features for every class, thus achieving a better separability between the iris, pupil and the rest of the non-iris classes. Hence, we can conclude that, among all the defined classes, the introduction of the pupil class can be expected to reveal the highest impact on iris segmentation. On the other hand, `fcn8s-atonce` reported better segmentation errors for almost all our multi-class approaches than the one shown by the 2-classes combination on MobBIO database. Specifically, the 4-classes-a combination showed worse results than the bi-class one. It is important to note, that the available iris and non-iris ground truth masks of the MobBIO database contain errors since they do not take into account the specular reflections. Based on that fact, we believe that the higher segmentation error reported by 4-classes-a combination results from this limitation. Fig. 2 shows two examples, in which the MobBIO ground truth masks provided in [Ho14] do not take into account specular reflection from an eye image while our 4-classes-a combination (iris, pupil, specular reflection and non-iris) does.

Finally, experimental results obtained by `fcn8s-atonce` showed that sclera regions have a greater impact than specular reflection when four classes are used. In this case, we think that this result is given by the fact that specular reflections can have random appearance in

Tab. 4: Proposed method (`fcn8s-atonce` 10-classes) vs. state-of-the-art on NICE.I

| **Segmentation method** | $E^1(\%)$ | **Segmentation method** | $E^1(\%)$ |
|---|---|---|---|
| Liu et al. [Li16] | 0.90 | Tan and Kumar [TK13] | 1.90 |
| Zhao and Kumar [ZK15] | 1.21 | Osorio et al. [ORMGGL17] | 2.38 |
| Tan et al. [THS10] | 1.31 | **Proposed method** | **1.13** |

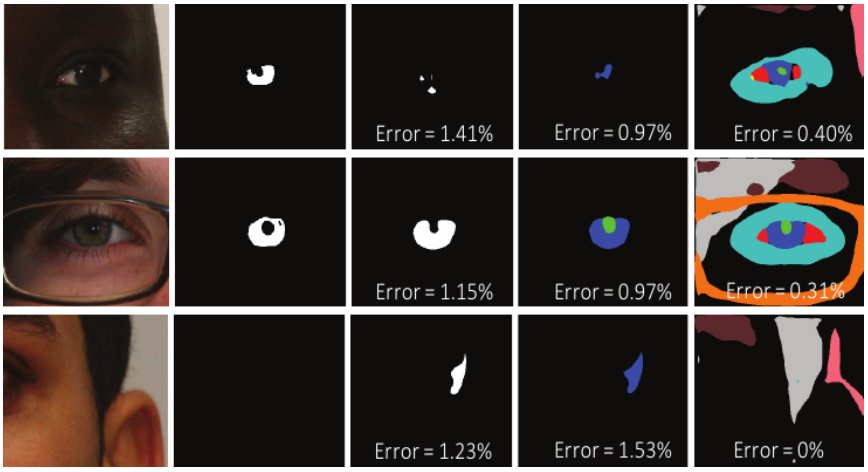


Fig. 3: Examples of $E^1$ segmentation errors; from left to right: original image, ground truth, 2-classes, 3-classes and 10-classes segmentation

an eye image, while the sclera region is present in most eye images, and it has a consistent and stable appearance in all cases. In order to evaluate our results against others reported in the literature, we performed a comparison between our proposal and state-of-the-art segmentation methods on the NICE.I database, see Tab. 4.

As can be observed from Tab. 4, our multi-class approach outperforms most state-of-the-art segmentation methods. Specifically, the system of [Li16] is the only iris segmentation algorithm which achieved better results on the NICE.I dataset compared to our multi-class proposal. To the best of our knowledge, the segmentation algorithm proposed is [Li16] is the best bi-class segmentation algorithm whose $E^1$ error was below 1.0%.

### 4.3 Refinement with 10-classes combination

While several multi-class combinations showed good results, some of them, e.g. 3-classes, 4-classes and 5-classes combinations, could not obtain an appropriate iris segmentation on eye images having the following conditions: eye images at-a-distance, with dark skin, and eye images that do not contain iris regions. In Fig. 3, we show three different examples (one per row) of such cases, that were successfully overcome with our 10-classes combination.

Note that, for the depicted images the proposed 10-classes combination outperforms [Li16] by 5.74, 0.89 and 3.54 percent points, respectively. In particular for the bottom image which does not comprise an iris the 10-classes combination obtains a correct iris segmentation.

## 5 Conclusion

In this paper, we explore the idea of segmenting the iris region under VW using a multi-class approach to improve the common bi-class one. A database of semantic segmentation labels of eye images, which contains 10 semantic classes was presented. Experimental results showed that for iris segmentation employing the information of the different semantic classes present in an eye image (by means of a FCN) is better than iris and non-iris segmentation. In addition, critical cases of misslocalization, such as eye images that do not contain an iris or dark skin images, can be correctly classified when the number of classes in an eye image is increased to 10 semantic classes. In addition, experimental results showed that specular reflections do not contain relevant semantic information with respect to other classes that compose an eye image. Finally, the proposed approach could be employed to detect glasses in VW iris images which have been shown to degrade biometric performance.

## Acknowledgements

This work was partially supported by the German Federal Ministry of Education and Research (BMBF), the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP).

## References

[Da93] Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(11):1148–1161, 1993.

[Da04] Daugman, J.: How iris recognition works. IEEE Trans. on Circuits and Systems for Video Technology, 14(1):21–30, 2004.

[Gr12] Grother, P.; Quinn, G. W.; Matey, J. R.; Ngan, M.; Salamon, W.; Fiumara, G.; Watson, C.: IREX III – performance of iris identification algorithms. NIST interagency report 7836, National Institute of Standards and Technology (NIST), 2012.

[Ho14] Hofbauer, Heinz; Alonso-Fernandez, Fernando; Wild, Peter; Bigun, Josef; Uhl, Andreas: A Ground Truth for Iris Segmentation. In: Proc. 22th Int'l Conf. on Pattern Recognition (ICPR'14). p. 6pp., 2014.

[Ji14] Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.s; Guadarrama, S.; Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proc. of the 22nd ACM Int'l Conf. on Multimedia (ACMMM'14). ACM, pp. 675–678, 2014.

[JU17] Jalilian, E.; Uhl, A.: Iris segmentation using fully convolutional encoder–decoder networks. In: Deep Learning for Biometrics, pp. 133–155. Springer, 2017.

[KSH12] Krizhevsky, A.; Sutskever, I.; Hinton, Geoffrey E.: ImageNet Classification with Deep Convolutional Neural Networks. In: Proc. of the 25th Int'l Conf. on Neural Information Processing Systems (NIPS'12). Curran Associates Inc., pp. 1097–1105, 2012.

[Li16] Liu, N.; Li, H.; Zhang, M.; Liu, J.; Sun, Z.; Tan, T.: Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In: Proc. Int'l Conf. on Biometrics (ICB'16). IEEE, pp. 1–8, 2016.

[LSD15] Long, J.; Shelhamer, E.; Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'15). pp. 3431–3440, 2015.

[ORMGGL17] Osorio-Roig, D.; Morales-González, A.; Garea-Llano, E.: Semantic Segmentation of Color Eye Images for Improving Iris Segmentation. In: Proc. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP'17). Springer, pp. 466–474, 2017.

[PA10] Proença, H.; Alexandre, L. A.: Iris Recognition: Analysis of the Error Rates Regarding the Accuracy of the Segmentation Stage. Image and Vision Computing, 28(1):202–206, 2010.

[PA12] Proença, H.; Alexandre, L.A.: Toward Covert Iris Biometric Recognition: Experimental Results From the NICE Contests. IEEE Trans. on Information Forensics and Security, 7(2):798–808, 2012.

[Ph08] Phillips, P.J.; Bowyer, K.W.; Flynn, P.J.; Liu, X.; Scruggs, W.T.: The Iris Challenge Evaluation 2005. In: Proc. Int'l Conf. on Biometrics: Theory, Applications, and Systems (BTAS'08). IEEE, pp. 1–8, 2008.

[Pr10] Proença, H.; Filipe, S.; Santos, R.; Oliveira, J.; Alexandre, L. A.: The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans. Pattern Analysis and Machine Intelligence, 32(8):1529–1535, 2010.

[Pr16] Proença, H.: Unconstrained Iris Recognition in Visible Wavelengths. In (Bowyer, W. Kevin; Burge, J. Mark, eds): Handbook of Iris Recognition (2nd Edition), pp. 321–358. Springer London, 2016.

[Sa10] Sankowski, W.; Grabowski, K.; Napieralska, M.; Zubert, M.; Napieralski, A.: Reliable algorithm for iris segmentation in eye image. Image and Vision Computing, 28(2):231–237, 2010.

[Se14] Sequeira, A.; Monteiro, J.; Rebelo, A.; Oliveira, H.: MobBIO: a multimodal database captured with a portable handheld device. In: Int'l Conf. on Computer Vision Theory and Applications (VISAPP'14). volume 3. IEEE, pp. 133–139, 2014.

[SZ14] Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[THS10] Tan, T.; He, Z.; Sun, Z.: Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image and vision computing, 28(2):223–230, 2010.

[TK13] Tan, C.-W.; Kumar, A.: Towards online iris and periocular recognition under relaxed imaging constraints. IEEE Trans. Image Processing, 22(10):3751–3765, 2013.

[TS09] Torrey, L.; Shavlik, J.: Transfer Learning. In: Handbook of Research on Machine Learning Applications. 2009.

[ZK15] Zhao, Z.; Kumar, A.: An Accurate Iris Segmentation Framework Under Relaxed Imaging Constraints Using Total Variation Model. In: Proc. IEEE Int'l Conf. on Computer Vision (ICCV'15). pp. 3828–3836, 2015.