SSBC 2018: Sclera Segmentation Benchmarking Competition

Abhijit Das[a,d,f], Umapada Pal[b], Miguel A. Ferrer[c], Michael Blumenstein[a,d], Dejan Štepec [e], Peter Rot[e], Žiga Emeršič [e], Peter Peer and Vitomir Štruc [e]

[a]Institute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia
abhijit.das@griffithuni.edu.au, m.blumenstein@griffith.edu.au

[b]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India, umapada@isical.ac.in

[c]IDeTIC, University of Las Palmas de Gran Canaria, Las Palmas, Spain, mferrer@dsc.ulpgc.es

[d]School of Software, University of Technology Sydney, Australia, michael.blumenstein@uts.edu.au

[e]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana
dejan.stepec@xlab.si, Peter.Peer@fri.uni-lj.si, Ziga.Emersic@fri.uni-lj.si, *peter.peer@fri.uni-lj.si*, vitomir.struc@fe.uni-lj.si

[f]Inria Sophia Antipolis – Méditerranée, France

Abstract

This paper summarises the results of the Sclera Segmentation Benchmarking Competition (SSBC 2018). It was organised in the context of the 11th IAPR International Conference on Biometrics (ICB 2018). The aim of this competition was to record the developments on sclera segmentation in the cross-sensor environment (sclera trait captured using multiple acquiring sensors). Additionally, the competition also aimed to gain the attention of researchers on this subject of research.

For the purpose of benchmarking, we have developed two datasets of sclera images captured using different sensors. The first dataset was collected using a DSLR camera and the second one was collected using a mobile phone camera. The first dataset is the Multi-Angle Sclera Dataset (MASD version 1), which was used in the context of the previous versions of sclera segmentation competitions. The images in the second dataset were captured using .a mobile phone rear camera of 8-megapixel. As baseline manual segmentation mask of the sclera images from both the datasets were developed.

Precision and recall-based statistical measures were employed to evaluate the effectiveness of the submitted segmentation technique and to rank them. Six algorithms were submitted towards the segmentation task. This paper analyses the results produced by these algorithms/system and defines a way forward for this subject of research. Both the datasets along with some of the accompanying ground truth/baseline mask will be freely available for research purposes upon request to authors by email.

1. Introduction

In the recent literature, sclera trait-based ocular biometrics in the visible spectrum has been investigated. Iris, sclera and the peri-ocular are the ocular traits that can be captured in the visible spectrum. Peri-ocular is the area around the eye, which consists of a pattern that can be used as a biometric trait [1]. Whereas, the sclera is the white region in the eye, which contains blood vessel patterns that can be employed for personal identification [3]. The very recent literature refers to the development of sclera biometrics [4-7]. The major reason for this development on sclera was due to its potentiality to improve the performance of iris recognition in the context of off-angle iris recognition and iris recognition in the mobile environment [12, 13]. For the majority of mobile phones available, their cameras capture images in the visible spectrum. The performance of iris biometrics in the visible spectrum can get affected. The sclera or peri-ocular biometric in conjunction with the iris biometric can enhance the relevance of the iris biometric in the mobile environment. For sclera, as an emerging trait, it is first necessary to assess its biometric usefulness independently [9]. Moreover, the research conducted on this subject is not very extensive. Additionally, sclera segmentation is found to be a significantly important part to establish sclera biometrics. However, sclera segmentation has not been investigated as a separate topic in most of the works reported in the literature, but mainly summarised as a component of the main task addressed.

Additionally, many works on sclera segmentation, which is addressed in the literature, were evaluated employing independent in-house datasets or on public datasets with fewer challenging sclera images. In this context, it is also worth mentioning that cross-sensor environments can work as an additional challenge if supervised learning techniques are used for segmentation. Incidentally, cross-sensor images were also not present in those datasets used in the literature. Therefore, we proposed to host a competition on sclera segmentation in the cross-sensor environment.

To set a common platform for the evaluation of sclera segmentation the 1st Sclera Segmentation Benchmarking Competition (SSBC 2015) [8], the 2nd Sclera Segmentation and Recognition Benchmarking Competition (SSRBC

0-7695-6403-8/18/31.00 ©2018 IEEE
DOI 10.1109/ICB2018.2018.00053

2016) [2] and the Segmentation and Eye Recognition Benchmarking Competition (SSERBC 2017) [14] were organized in the context of the IEEE 7th International Conference on Biometrics: Theory, Applications and Systems BTAS 2015, the 9th IAPR International Conference on Biometrics (ICB 2016), and the IAPR/IEEE International Joint Conference on Biometrics (IJCB 2017), respectively. The successful organisation and the appreciating impact of these competitions have inspired the organisers to plan further competitions on sclera segmentation in cross-sensor environment namely: SSBC 2018.

The main aim of the competition was to establish a standard benchmark for sclera segmentation in the visible spectrum with a common dataset in the cross-sensor environment and also to record the recent developments of sclera segmentation that took place after SSERBC 2016. In addition, this competition was also aimed to attract the interest of researchers on this particular subject. As the conceived competition is related to biometric research, so it was organised in the context of the 11th IAPR International Conference on Biometrics (ICB 2018).

The rest of this paper is organised as follows. In Section 2 the competition schedule, the dataset for the competition and the performance evaluation technique adopted to evaluate and rank the participant's algorithms are described. In Section 3, various algorithms from the participants are described in detail, in Section 4 the results achieved from the submitted algorithms and their detailed analysis is summarised. Finally, the last section i.e. in Section 5, the overall conclusions are drawn and the future work is discussed.

2. The SSBC 2018 competition

The competition schedule is shown in Table 1.

Table 1: Schedule of the competition

Different Phases	Dates
Competition website opens	20th August 2017
Registration starts	21st August 2017
Test dataset available	21st August 2017
Registration closes	29th November 2017
Algorithm/system submission	29th November 2017
Results announcement	30th November 2017

The competition was promoted through a website and further communications were made by email to the researchers. Ten participants registered for the competition from distinguished laboratories of academia and industry, from around the globe. Among them, six submitted their segmentation algorithms. Table 2 reflects the name and the affiliation of participants who submitted their algorithms.

Table 2. Descriptions of the teams details those who submitted their system

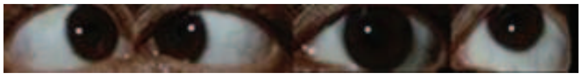
Teams	Name (Institution)
1	Dejan Štepec, Peter Rot, Žiga Emeršič, Peter Peer, Vitomir Štruc, (Faculty of Computer and Information Science, Ljubljana)
2	Somenath Chakraborty (Harirampur Government ITI)
3	Chandranath Adak (Griffith University, Australia)
4	Anonymous

The Multi-Angle Sclera Database (MASD version 1) used in SSBC 2015 is employed in this competition for the segmentation task [8]. It consists of 2624 RGB images taken from 82 identities. The images were collected from both the eyes of each individual so 164 different eyes were available. Here for each individual image, four multi-angles (looking straight, left, right and up) are considered and for each angle 4 images are considered. The individuals are comprised of both male and females with different skin complexions, and a few of them were wearing contact lens and images were taken at different times of the day. The database contains images with blinking eyes, closed eyes and blurred eye images. High-resolution images are provided in the database (300 dpi resolution and 7500 x 5000 dimensions). All the images are in JPEG format. A NIKON D 800 camera and 28300 lenses were used for image capturing.

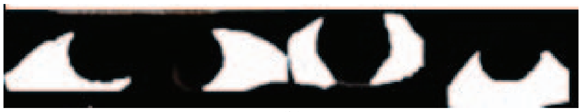
The second database consists of 400 RGB images from both eyes of 25 individuals (in other words 50 different eyes). For each eye, 8 sample images were captured. The database contained blurred images and images with blinking eyes. The individuals were comprised of both males and females (12 males and 13 females) of different ages and different skin colours, 2 of them were wearing contact lenses and the images were taken at different times of the day. Variations in image quality (blur, lighting condition etc.) and different acquisition conditions were included intentionally in the database to investigate the performance of the framework in non-ideal scenarios. High-resolution images (3264 × 2448) of 96 dpi are included in the database. All the images are in JPEG format. The images were captured using a mobile camera with an 8-megapixel rear camera.

A graphical application was developed using Matlab 7 in the Windows 7 Operating System environment to generate manually segmented masks or ground truths of these sclera images in the dataset, in order to obtain a baseline to evaluate the automatic segmentation algorithms submitted by different participants.

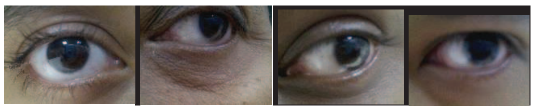
A set of eye images with their respective manually segmented masks from both the datasets are given in Figure 1.



(a)



(b)



(c)

(d)

Figure 1: (a) A set of images with different angles from the first database, (b) their manual mask with only sclera regions of images of the first database from the original image in (a), (c) A set of images with different angles from the second database, (d) their manual mask with only the sclera region of images of the second database respectively according to the original image in (c).

For algorithm or system development purposes of the segmentation task, a subset of the first database and ground truths (1 image for each angle of the first 30 individuals i.e. 120) were provided to the registered participants of the competition. From the second dataset a set of 25 eye images, one image from each individual was provided for algorithm judgement purpose.

The participants were asked to provide a program that can read the images from a directory and which writes the segmented mask in a particular directory with a naming convention. For ease of evaluation and to maintain the real-time property of the submitted algorithm, the participants were asked to submit a system which does not take more than 10 seconds to segment and generate the mask for an image on an Intel Core i7 processor.

The performance of the segmentation task can be evaluated with respect to the manually segmented mask and is a pixel level binary classification, so a precision and recall measure is employed as a performance measure. The recall is considered the measure for ranking the algorithms if the average precision of the submission is found to be the same. The mathematical representation of the precision and recall for our scenario is shown in the following equations.

$$\text{Precision } = \frac{\text{NPAM}}{\text{NPRS}} \quad (1)$$

$$\text{Recall } = \frac{\text{NPAM}}{\text{NRMS}} \quad (2)$$

where,

NPAM = Number of pixels retrieved in the sclera region by the automatically segmented mask

NPRS = Number of pixels retrieved in the automatically segmented mask.

NRMS = Number of pixels in the sclera region in the manually segmented mask

3. Brief description of the submitted algorithms from each team

The six segmentation algorithms submitted by the four participating teams are described in this section.

3.1. First segmentation algorithm: Submitted by participating team 1

This group from the University of Ljubljana (UL) participated in the sclera segmentation challenge used Refinenet [16] which presents a state-of-the-art method for semantic segmentation. The big limitation of simpler methods is repeated subsampling that reduces the image resolution which leads to a significant decrease in the performance. Refinenet explicitly exploits all the information available along the down-sampling process to enable high-resolution prediction using long-range residual connections. At the time of publication, the method presented a state-of-the-art result on a PASCAL 2012 dataset.

They have used the official MATLAB implementation (https://github.com/guosheng/refinenet) and implemented a specific layer for the MASD dataset. They only trained for stage-1 using a fixed learning rate until the performance didn't get stable on the held-out validation set. No post-processing was used for getting the final results.

3.2. Second segmentation algorithm: Submitted by participating team 1

In the second algorithm, the first team used SegNet deep convolutional encoder-decoder [10]. The architecture was implemented in Caffe.

From each image out of those 180 images, 300 perturbations were made. Together they have generated approx. 54000 images. Perturbations were a combination of the following operations: cropping, Gaussian blur, additive Gaussian noise, brightness changes, contrast normalisation and affine transformations (scaling and rotation) similar as in [11]. Original images from the training set and their perturbations were resized to 640 x480 pixels and were used as an input for the SegNet convolutional network. The model was trained with 30000 iterations (with a batch size of 4). In the test phase, the first step is to resize images to 640x480 pixels and store the original size of each input image. The output of SegNet is masked where a pixel value 0 corresponds to the prediction of the background and 1 to the prediction of the sclera. The image size of those masks is as expected 640x480 pixels, so the final step is to resize those masks to the original sizes of input images.

3.3. Third segmentation algorithm: Submitted by participating team 1

For the third method, they have used UNet architecture [15]. CNN architecture was implemented in Tensorflow (https://github.com/jakeret/tf_unet) with encoding and decoding depth 3, 128 features on the first level, 0.5 dropout rate and 0.001 learning rate.

They have trained it with 10 iterations per epoch with 200 epochs, with batch size 4. After prediction phase on validation set post-processing was applied. They have filtered regions smaller than a certain size, and those with contour centre too far away from the middle of the image (assumed that the centre of the eye was near image centre). They added 50 images for all the three algorithms submitted by them (gathered in the wild) with corresponding hand-made annotations to the existing 130 images in the training dataset.

3.4. Fourth Segmentation algorithm: Submitted by participating team 2

The algorithm submitted by this team used unsupervised learning which is complementary to the previous three algorithms proposed by team 1, which is assumed to work better in a cross-sensor environment. This participant proposed a scheme to segment the sclera region using Gaussian kernel convolution. The proposed method includes a scale space-based technique to obtain a smooth image. Which was achieved by convolving an original grey image with a Gaussian filter that extracts edges of the image. The implementation includes, repeated convolution operations of gray images with a Gaussian filter at different scales. This generates the difference of Gaussian images from the difference of adjacent blurred images. Further, by wrapping the difference of Gaussian images and using some heuristics detection, sclera region and non-sclera region were located. Further, the threshold was performed using Otsu's binarization (which is unsupervised, assume to work well for the cross-sensor environment) value so that we can detect the sclera and non-sclera portion better.

3.5. Fifth Segmentation algorithm: Submitted by participating team 3

The participants intended to tackle the problem as a colour information-based segmentation problem. Although there were several difficulties, which arose like the eye-brow and iris colour became almost the same or the sclera part where the vessel patterns are quite bundled and became similar to the skin colour. Therefore, instead of approaching it as only as a colour segmentation problem, the participants tried to find the larger white portion that they assume as the sclera part.

The contributor has used an unsupervised technique so that prior knowledge is not required and that will be also helpful to tackle the cross-sensor scenario. In the very beginning, to reduce computation overhead, the method converts the colour image into grey levels. Then all the peaks are found from the grey-level histogram. Fuzzy C-Means (FCM) clustering is used for segmenting the grey image.

Here the number of clusters i.e. C-value is perceived as the number of peaks ± low threshold T of the histogram. Since the aim is to find the sclera, i.e. larger white region, the cluster having the highest grey-value is extracted as the region of interest/sclera region and the remaining portions turned into black. It was further observed that in the segmentation mask, the sclera region does not contain any holes, so all the holes are filled in and very small components of high-intensity vessel patterns are filtered out as noise.

3.6. Sixth Segmentation Algorithm: Submitted by participating team 4

In this algorithm, the participants used the almost same above mention techniques used by team 3. The difference was that instead of taking only the cluster with the highest grey-value, they considered two clusters having the highest and next-to-highest grey value. Now on these two clusters, Otsu's binarization is performed, which is again unsupervised or no prior knowledge is required. Therefore, this is assumed to be helpful for the cross-sensor scenario. Similar to the above approach, here also post-processing/hole-filling and noise (small isolated components) removal are performed to obtain a segmentation mask with a better outcome.

4. Results and Discussion

In this section, we summarised the results achieved after applying the submitted algorithms on the employed dataset for segmentation. A detailed analysis of the result obtained from the submitted algorithms is included in the following subsections.

4.1. Sclera segmentation results and discussion

In this competition, like previous versions of our competitions, we maintained a protocol for submissions of the algorithm/system and then evaluated them by a common framework and ranked them to maintain a fair and unbiased comparison among the algorithms of the participants. Through this publication, the participants and the readers can find the performance of different methods and an analysis of the same.

The results were obtained on the two datasets, comprised of 2624 images from the first dataset out of which 120 images were used by the participants to train their system if the prior knowledge-based model is used and in the second dataset 400 images for testing. In Table 3, the final quantitative results are presented for the six segmentation algorithms submitted in the competition.

Table 3. Final results of the participant's system (in %).

Rank	Algorithms No.	Precision in %	Recall in %
1	1	81.35	75.82
2	2	80.70	48.41
3	5	30.43	18.76
4	4	29.05	15.76
5	3	17.94	52.43
6	6	7.86	9.08

The precision and recall curve for the 1[st] algorithm is in figure 2.

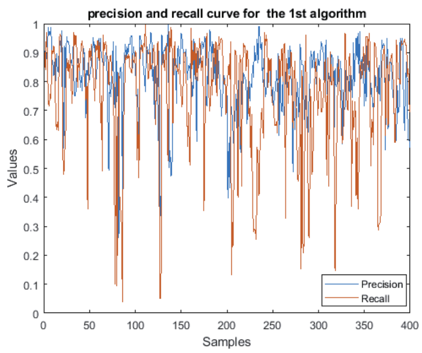


Figure 2: The precision and recall curve for the 1st algorithm.

According to the competition protocol, we ranked the results by the average precision and further ranked by the average recall for any duplicated ranks generated by average precision.

We can conclude from the Table 3 that appreciable segmentation performance was achieved in the most of the submitted systems. The performance of the system submitted by algorithm 1 by team 1 was the best with respect to both average precisions as well as average recall performance compared to the other algorithms submitted in the competition. The performance of the algorithm was quite good when the cross-sensor scenario was used. The performance gap with the next ranking system was about ~1% for precision and ~20 % for recall, the difference was quite high for the next algorithms.

We have performed a qualitative analysis consisting of three individual observations. With the qualitative analysis also the performance of algorithm 1 was found to be best and outperformed the error achieved in comparison to other algorithms submitted in the competition. Therefore we further analysed the masks generated by the system of algorithm 1.

4.2. Sclera segmentation results analysis

Some examples of the mask where poor segmentation occurred and the corresponding colour eye images are given in Figure 3.

The major misclassification that can be found is: due to the specular reflection (the iris part which contains specular reflection are misclassified as the foreground or sclera part. In Figure 3(a), iris part are misclassified as the foreground or sclera part; the second reason behind the misclassification was due to uneven lighting on the images as shown in Figure 3(b, e and g); in the next category the skin area are misclassified as in Figure 3(c); sclera corner are misclassified due to uneven illumination 3(d); misclassified due to reddish eye 3(f). The performance of the algorithm varied mainly due to illumination change and

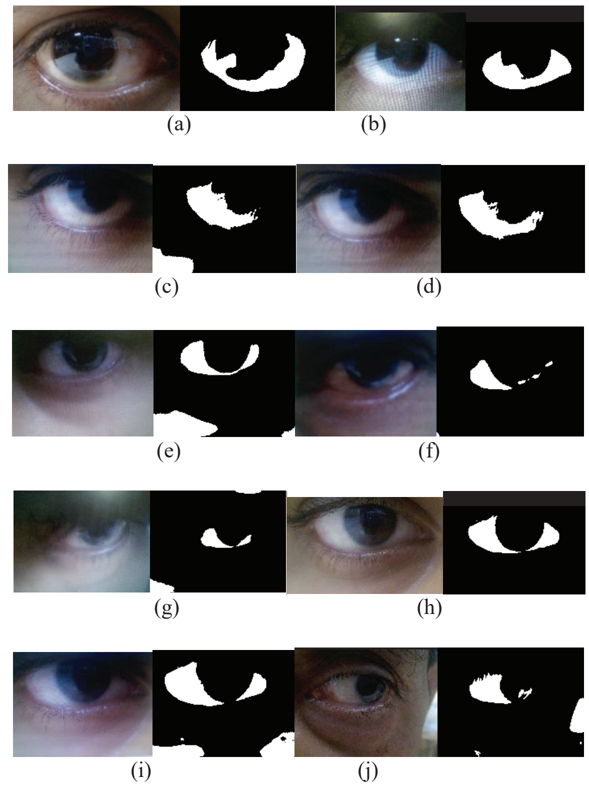


Figure 3: Examples of the masks produced by the algorithm 1 submitted by team 1 and the corresponding eye images from the database.

due to uneven illumination distribution over the image Fig.3 (b and c). The most important highlighting part of the analysis is that in some occasion the performance varied with similar type illumination used to capture both image Fig.3 (e vs. d). Moreover, for a similar type of images in some occasion it worked and did not for the other. Perhaps some pre-processing based on illumination adjustment may highly improve the pitfall or else more rigorous training can solve such misclassifications. Incidentally, it can be concluded that cross-sensor acquiring environment does not have a very high impact on the misclassification found.

We made further analysis, we compared the algorithm 1 with the best algorithms of SSBC 2015 [8], SSRBC 2016 [2] and SSERBC 2017[14]. The performance of the algorithm 1 was found to be better in both the quantitative and qualitative analysis in this aspect also. Moreover, from the detailed analysis, we found that the variations of illumination and other changes highlighted affected those algorithms more in comparison to the algorithm 1. As the context of the competition is to benchmark sclera segmentation in the cross-sensor environment, so it is worth mentioning that the performance of the algorithm 1 while tested with another cross-sensor environment can get affected.

Therefore from the above-mentioned analysis, we can conclude that the remaining challenges i.e. the variation in illumination globally in the image, as well as that locally affecting the performance of sclera segmentation, keeps sclera segmentation as an open research area. To solve some of these challenges, preprocessing of the eye images could help (i.e. avoiding scenarios such as the introduction of sclera vessel patterns in the mask and uneven lighting and illumination scenario, etc.).

5. Conclusions and Future Scope

The 5th Sclera Segmentation Benchmarking Competition, SSBC 2018 was organised with the primary goals to record performance of sclera segmentation in the cross-sensor environment, and the recent advancements in sclera segmentation techniques in the visible spectrum. Moreover, it also aims to provide a common platform to evaluate sclera segmentation algorithms using a unique cross-sensor eye dataset.

Subsequently, the showcasing of the competition in one of the most recognised gatherings in the biometric community i.e. ICB 2018 and promoting them via different electronic means of communications, have also increased the interest of researchers using this particular subject of research in biometric. Furthermore, the conceived competition has satisfactorily fulfilled all of the above aims, and the gain in popularity and interest of the participants were noteworthy.

The algorithms submitted by the participants demonstrate appreciable results on our proposed dataset. We hope the critical analysis was undertaken and results of the different algorithms will also provide a way forward for further research. One very important aspect of the research is the availability of datasets publicly. To the best of our knowledge, no datasets are publicly available which is enriched with a wider variety of multi-angle eye images developed in cross-sensor scenarios, the availability of this proposed dataset will fill that gap.

The successful organisation and the appreciating impact of this competition have inspired the organisers to plan further scientific events on the sclera and eye biometrics paradigm in the near future.

Acknowledgement

The authors sincerely appreciate the volunteers who submitted their samples for datasets development, and those who have helped in the data collection process.

References

[1] U. Park, A. Ross, and A. K. Jain, Periocular Biometrics in the Visible Spectrum: A Feasibility Study, IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-6, 2009.

[2] A. Das, U. Pal, M. A. Ferrer and M. Blumenstein, SSRBC 2016: Sclera Segmentation and Recognition Benchmarking Competition, ICB 2016, 1-6, 2015.

[3] R. Derakhshani, A. Ross, and S. Crihalmeanu. A new biometric modality based on conjunctival vasculature. Proceedings of Artificial Neural Networks in Engineering: 1-8, 2006.

[4] R. Derakhshani and A. Ross. A texture-based neural network classifier for biometric identification using ocular surface vasculature. Proceedings of International Joint Conference on Neural Networks: 2982–2987, 2007.

[5] A. Das, U. Pal, M. Blumenstein and M. A. Ferrer, Sclera Recognition - A Survey, Advancement in Computer Vision and Pattern Recognition: 917 -921, 2013.

[6] A. Das, U. Pal, M. A. Ferrer and M. Blumenstein, Fuzzy Logic Based Sclera Recognition, In FUZZ-IEEE: 561-568, 2014.

[7] Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp. Multi angled sclera recognition. IEEE Workshop on Computational Intelligence in Biometrics and Identity Management: 103 – 108, 2011.

[8] A. Das, U. Pal, M. A. Ferrer and M. Blumenstein, SSBC 2015: Sclera Segmentation Benchmarking Competition, BTAS 2015, 1-06, 2015.

[9] S. Crihalmeanu and A. Ross, Multispectral sclera patterns for ocular biometric recognition, Pattern Recognition Letters, 33: 1860–1869, 2012.

[10] V. Badrinarayana, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, 2015.

[11] Z. Emersic, D. Stepec, V. Struc and P. Peer, Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild, International Conference on Automatic Face and Gesture Recognition -- Workshop on Biometrics in the Wild, 2017.

[12] A. Das, P Mondol, U. Pal, M. F. Ballester and M. Blumenstein, Fast and efficient multimodal eye biometrics using projective dictionary pair learning, Evolutionary Computation (CEC), 2016, 1402-1408, 2016.

[13] S. P. Tankasala, P. Doynov, R. R. Derakhshani, A. Ross and S. Crihalmeanu, Biometric Recognition of Conjunctival Vasculature using GLCM Features, International Conference on Image Information Processing : 1-6, 2011..

[14] A. Das, U. Pal, M. A. Ferrer, M. Blumenstein, D. Stepec, P. Rot, Z. Emersic, P. Peer, V. Struc, S. V. K. Arun and B. S. Harish SSERBC 2015: Sclera Segmentation Benchmarking Competition, IJCB 2017, 1-6, 2017.

[15] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597, 2015

[16] G. Lin, A. Milan, C. Shen, I. Reid, RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation, arXiv:1611.06612, 2016.