A robust sclera segmentation algorithm

A Robust Sclera Segmentation Algorithm

Petru Radu, James Ferryman and Peter Wild
Computational Vision Group, School of Systems Engineering, University of Reading, U.K.
{p.radu | j.m.ferryman | p.wild}@reading.ac.uk

Abstract

Sclera segmentation is shown to be of significant importance for eye and iris biometrics. However, sclera segmentation has not been extensively researched as a separate topic, but mainly summarized as a component of a broader task. This paper proposes a novel sclera segmentation algorithm for colour images which operates at pixel-level. Exploring various colour spaces, the proposed approach is robust to image noise and different gaze directions. The algorithm's robustness is enhanced by a two-stage classifier. At the first stage, a set of simple classifiers is employed, while at the second stage, a neural network classifier operates on the probabilities' space generated by the classifiers at stage 1. The proposed method was ranked the 1st in Sclera Segmentation Benchmarking Competition 2015, part of BTAS 2015, with a precision of 95.05% corresponding to a recall of 94.56%.

1. Introduction

The sclera region in a human eye is surrounding the iris and, although riddled with blood vessels, appears white. A membrane called conjunctiva, which is a clear mucous membrane, covers the sclera. When an eye image with an off-angle iris is acquired, one may observe the blood vessels from the conjunctiva and sclera. Sclera vessels have therefore a multiple layer structure and change their position when the eye moves [1]. Sclera recognition was proposed as a biometric modality by R. Derakhshani, A. Ross and S. Crihalmeanu in 2006 [2].

Although the accuracies of the visible spectrum iris recognition systems are not comparable to those operating in the near infrared spectrum [3] [19], the visible spectrum iris imaging has the advantage of permitting the integration of additional sources of information, such as eye colour or sclera vasculature [4]. Additionally, reliable sclera segmentation can significantly improve and simplify more complex tasks such as iris segmentation and gaze tracking [5]. Thus, automatic detection of sclera is becoming an important research topic in biometrics. At the moment of writing this paper, the literature is not rich in works depicting sclera segmentation algorithms. To encourage research efforts on sclera segmentation, a competition called Sclera Segmentation Benchmarking Competition (SSBC) 2015 [6] was organized as part of the BTAS 2015 conference.

One of the first papers that describes sclera segmentation employs a modified Self Organizing Map [7] in a gaze tracking approach. The method relies on finding the iris boundary first and fixing two control positions calculated by using iris center and radius. The two control positions are then employed in an active contour model algorithm to fine tune the sclera boundary location. In [1] it is suggested that sclera recognition should be done only on the sclera vein patterns layer, which are stable over time, rather than including the conjunctiva vasculature. The sclera segmentation approach employed in [1] assumes that the images contain frontal-looking eyes and the iris center location is available. Two binary maps are generated based on detecting non-skin area using RBG colour space and white colour using HSV colour space. Furthermore, the convex hull of the two masks is calculated and fused to obtain a final sclera region.

Sclera vasculature as a biometric modality is explored in [4] under different wavelengths. The sclera was segmented by employing a sclera index measure, which relies on multispectral information, i.e. the difference between near infrared and green pixel intensities is larger for the sclera region. In [8] a K-means clustering approach is employed to segment the sclera. A survey of the sclera recognition works until 2013 was made in [9] and with regards to sclera segmentation the survey shows that the few existing approaches are relying on various assumptions, e.g. iris center location is known. In 2014, Abhijit et al proposed a method for sclera segmentation based on Fuzzy logic [10].

Unlike existing sclera segmentation algorithms, the present work relies on machine learning techniques to robustly detect the pixels that belong to the sclera region without employing conventional constraints. The proposed approach employs three feature types: statistical image features, Zernike Moments and Histogram of Gradients (HoG)-like features. The contributions of the present work are twofold: first, a flexible two stage multiple classifier system (MCS) architecture is proposed for pixel-level sclera detection which can be easily configured and adapted to other machine vision tasks; second, a thorough

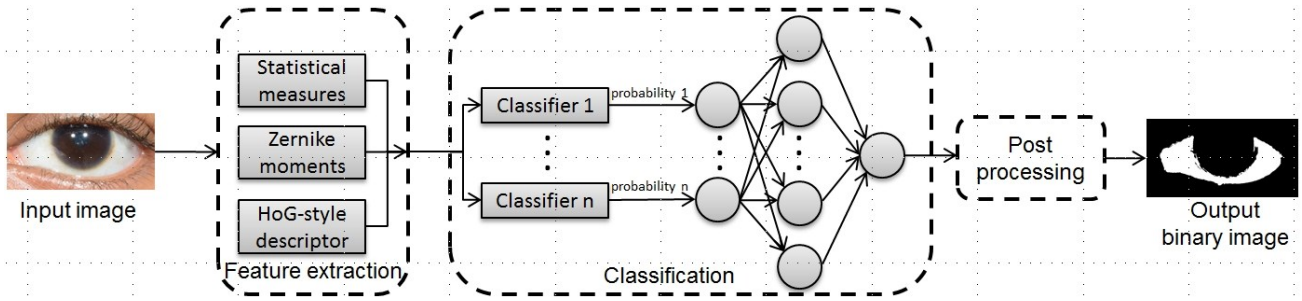


Figure 1: Proposed sclera segmentation system

evaluation of the sclera segmentation is performed, after observing that the existing literature does not focus on performance evaluation of sclera segmentation.

The remainder of this paper is organized as follows: in Section 2, the automated sclera segmentation approach is presented, together with the details of the feature extraction and classifier architecture. In Section 3 the experimental results are presented and the conclusions are drawn in Section 4.

2. Algorithm design

The proposed sclera segmentation algorithm operates on visible spectrum RGB eye images. The present approach was designed to be robust to various factors, e.g. change in illumination, occluded sclera regions or off-angle iris; therefore it does not rely on available prior information such as eyelid detection or iris center coordinates. The block diagram of the proposed sclera segmentation approach is illustrated in Figure 1.

2.1. Feature extraction

In most pattern recognition algorithms, the feature extraction plays an important role towards an enhanced robustness to noisy data. A relevant, diverse, independent and compact set of features is necessary to achieve an increased accuracy of the learning algorithm [11]. Three feature types are employed in the present work to distinguish between sclera and non-sclera pixels of an RGB image. The features have been selected to grasp independent information related to colour, shape and presence of edges.

The *first feature type* employed explores the various relationships between pixels intensities from different colour spaces, as suggested in [12] and [13]. For a 2-dimensional image *I(x, y)*, where *x* and *y* denote respectively the image row and column, an 18-dimensional vector is computed for every pixel as follows:

$$\{d^{\mu\sigma}_{cr-cb(0,2,4)}, nb^{\mu\sigma}_{(0,2,4)}, S^{\mu\sigma}_{(0,2,4)}, d_{RGB}, \mu_{RGB}, \Delta_{Lab(4)}\} \quad (1)$$

where *nb*, *S* denote the scale-normalized blue channel and saturation channel from HSV colour space respectively and d_{cr-cb} and μ_{RGB} denote the difference between chroma red (*cr*) and chroma blue (*cb*) from YCbCr space and the average value between the RGB channels respectively. The subscripts in parentheses represent the radii of the local neighborhood window centered at *(x, y)*. For the value of the subscript equal to 0, only the intensity value of the center of the window is considered. The superscripts μ and σ indicate the mean and standard deviation of the local neighborhood with respect to the radii from the subscripts. The d_{RGB} and Δ_{Lab} are defined by the following formulas:

$$d_{RGB}(x,y) = 2I_R(x,y) - I_G(x,y) - I_B(x,y) \quad (2)$$

$$\Delta_{Lab}(x,y) = \sqrt{L^{\mu}(x,y)^2 + a^{\mu}(x,y)^2 + b^{\mu}(x,y)^2} \quad (3)$$

where I_R, I_G and I_B are the red, green and blue channels respectively and *L*, *a*, and *b* are the channels of LAB colour space.

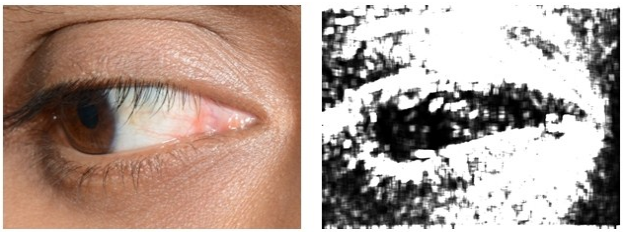
The *second type of features* employed are Zernike moments [14]. Zernike moments are usually employed for rotation invariant shape recognition. The basis set for Zernike moments are Zernike polynomials, which are defined as follows:

$$V_{nm}(x,y) = V_{nm}(\rho\, sin\theta, \rho\, cos\theta) = R_{nm}(\rho)e^{jm\theta} \quad (4)$$

where *m* and *n* are integer numbers and represent the order and repetition of the Zernike moments and $R_{nm}(\rho)$ is called radial polynomial with $R_{n,-m}(\rho) = R_{n,m}(\rho)$ [14]. As observed from (4), the image or region of interest needs to be first expressed as a function *f* of intensity values given polar coordinates ρ and θ. Subsequently, the complex Zernike moments are computed using the following equation:

$$Z_{nm} = \frac{n+1}{\pi}\sum_{\theta=0}^{2\pi}\sum_{\rho=0}^{1} f(\rho,\theta)R_{nm}(\rho)e^{-jm\theta} \quad (5)$$

The absolute values of 9 complex Zernike moments are computed for different values of *m* and *n* for the local



a) b)

Figure 2: HoG-like features. a) original eye image; b) weighted votes for the bin starting at 120 to 140 degrees.

windows with radius 4 centered on every pixel.

HoG-like features are used for the *third type of features*. The motivation behind the choice of this feature type is that the sclera region has significantly less edges than other regions of the eye image. The HoG features are well known for their human detection application [15]. After a Gaussian smoothing operation of the gray scale eye image, an edge detection filter is convoluted with the eye image. Subsequently, the gradients are computed and grouped into 9 bins, according to their orientation, from 0 to 180 degrees, with a step of 20 degrees. For each pixel, a weighted vote corresponding to an edge orientation bin is computed, where the weight is represented by the magnitude of the gradients. Since most of the 9 weighted votes for each pixel are 0, a filtering operation which divides the square of the sum of pixel intensities to the number of pixels within the filtering window is employed to smooth out the values of the weighted votes for each bin. In this way, for the sclera region, the values of the weighted votes have a low value, while for non-sclera regions, the values are spread across a wider range, as shown in Figure 2.

The final feature vector has 36 real valued components and is obtained by concatenating the 3 features types described above.

2.2. Multiple classifier system architecture

For the feature types employed by the proposed sclera segmentation method, a robust classification stage is required, where changes in illumination or skin colour do not have a significant effect on the performance of the algorithm. The performance of a single classifier is likely to be affected more by the noise present in the testing features than the performance of a MCS [11].

At the same time, the principle of diversity of MCS [11] was considered when designing the classifier for the proposed sclera segmentation approach. The MCS employed in the present work can be easily adapted for a broader range of image analysis tasks, where robustness to noisy data is a requirement. The MCS topology proposed in this work is a parallel one, as shown in Figure 1, where a 2-stage operation takes place.

At the first stage, a number of n simple 2 class classifiers are employed. For speed purposes, classifiers that have linear decision boundaries are considered adequate. To ensure that the principle of diversity is respected, a combination of generative classification methods, such as density based classifiers and discriminative classification models is desirable. In the proposed approach, the number of classifiers at the first stage of the MCS is n=3, which operate in parallel on the 36 component feature vector:

1) A density based classifier represented by a Bayes classifier with linear decision boundary [11], which classifies the features to the most probable class;
2) A distance based classifier represented by Fisher Linear Discriminant [16], which projects the data onto a line so that samples from different classes become better separated;
3) A discriminant classifier model represented by regularized logistic regression (LR).

At the second stage, a more complex, nonlinear classifier is recommended to be employed, as it will operate on the probability space generated by the first stage classifiers. The probability space will have a reduced dimension, thus eliminating speed related concerns. In the present work, a feed forward neural network (FFNN) with 1 hidden layer operates on the vector of probabilities generated by the first stage classifiers. The activation function of the FFNN is a classical sigmoid function.

The training of the MCS was performed on randomly selected patches of 100 by 100 pixels from UBIRIS v1 database [17] and the database offered for training for SSBC 2015 [6]. 60 patches are randomly collected from each database for sclera region and 60 patches for non-sclera regions. Examples of sclera and non-sclera patches used for training are shown in Figure 3. Initially, the 36 real valued feature vectors are extracted for all the pixels in the patches. By training the MCS on features obtained from different databases/sensors, an improved modeling of the intra class variability of the data is achieved.

To enhance the robustness of the segmentation algorithm, the training of the MCS is done in 2 phases. In the first phase, the first stage classifiers are trained on a subset of the training patches. The subset of patches for training the first stage classifiers consists of half of the total

a) b)

c) d)

Figure 3: Training patches. a) sclera patches from UBIRISv1; b) sclera patches from SSBC 2015; c) non-sclera patches from UBIRISv1; d) non-sclera patches from SSBC 2015

number of patches in the present work, i.e. 60 patches from sclera regions and 60 from non-sclera regions. In the second phase, the remaining unseen patches are tested on the first stage classifiers to generate training probabilities for the FFNN. The proposed MCS has therefore a data independent second stage classifier [11]. The motivation behind this two phase training procedure is to correct the mistakes done by the first stage classifiers by remapping the deviated probabilities to the correct labels.

The proposed MCS architecture and training strategy led to the following research question: *is the accuracy of the segmentation algorithm insensitive to the size of the training dataset for the first stage classifiers?* In other words, if the training size for the FFNN is large enough to correct the deviated probability values of the first stage classifiers, the precision of the decision boundary tuned in the first stage classifiers may not compromise the accuracy of the segmentation algorithm. This research question will be addressed and answered in the experimental results section, where it is shown how significantly decreasing the training size for the 2 stages of the MCS considerably decreases the training time of the segmentation algorithm while keeping its accuracy almost unchanged.

2.3. Post processing

As the proposed sclera segmentation algorithm does not rely on finding the location of the iris or the shapes of the eyelids, reflections from the skin areas or iris regions might be classified as sclera pixels. To reduce the effect of reflections on the performance of the algorithm, two image processing techniques are employed.

Initially, the aim is to eliminate falsely classified sclera pixels from the iris region. For these pixels to be detected, a simple dynamic contrast adjustment operation is applied on the gray scale eye image. The anatomy of the human eye, which exhibits a significant contrast difference between sclera and iris, allows for course detection of the iris disc by employing contrast adjustment operations. The dynamic contrast adjustment parameters depend on the mean value of the pixel intensities of the gray scale eye image. The parameters of the contrast adjustment were empirically found for different ranges of the average value of the mean intensity of the eye image. After the contrast adjustment operation is complete, a binary image is generated by applying Otsu's thresholding technique [18]. The obtained binary image is subsequently used in a masking operation with the binary image that the FFNN generates to mask out the iris disc from the sclera.

Further, to reduce the effect of skin reflections on the algorithm's performance, the binary image resulted from the masking operation is used to find the connected components regions. From all the connected components, only the ones which have the area above a certain threshold are considered as sclera candidates. The threshold employed depends on the size of the eye image. The small remaining connected components are filled with zeros, where 0 represents the intensity value for black colour. The effect of dynamic contrast adjustment operation and small connected components removal is illustrated in Figure 4.

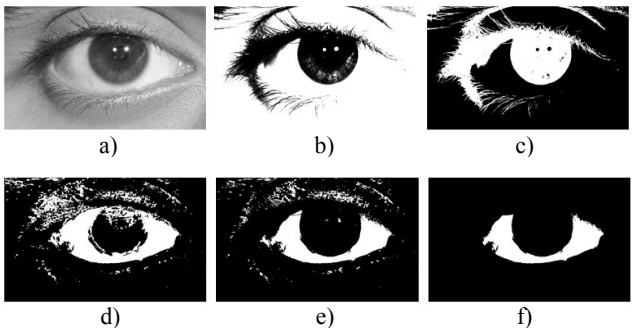


Figure 4: Post processing. a) initial gray scale eye image; b) contrast adjusted image; c) binary mask obtained after contrast adjustment; d) output of the FFNN; e) FFNN output after masking pixels indicated by c); f) final segmented image

3. Experimental results

In this section, the systematic evaluation of the proposed sclera segmentation algorithm is presented. As reproducibility and repeatability of algorithms' evaluation is becoming a highly desirable property of the published biometrics research works [19], the need of a database dedicated to sclera segmentation, where ground truth mask indicating binary class is provided becomes apparent.

The eye image database offered to the participants of SSBC 2015 [6] contains ground truth masks for the sclera regions. The database contains images acquired with a Nikon D 800 camera from 82 individual, therefore 164 different eyes. The SSBC participants were given a subset of the database, containing eye images from 30 individuals with a size of around 3 mega pixels. This subset is used for the experiments in the present work. The acquisition protocol for the SSBC database specifies that the images contain eyes for four gaze directions: straight, up, left and right. Further, the images are acquired under different illumination conditions and contain noise such as reflections or occluded sclera regions, making the SSBC database a challenging one for sclera segmentation. *Note that at the time of submission of this paper, the SSBC 2015 evaluation protocol and database were not yet available*.

The evaluation protocol adopted for the proposed sclera segmentation algorithm consists of generating the precision-recall (P-R) curves for different parameter settings. The precision is defined as the ratio between true positive (TP) and the sum of TP and false positive (FP), while the recall is defined as the ration between TP and the sum of TP and false negative (FN). The P-R curves are chosen to report the results because they offer a holistic

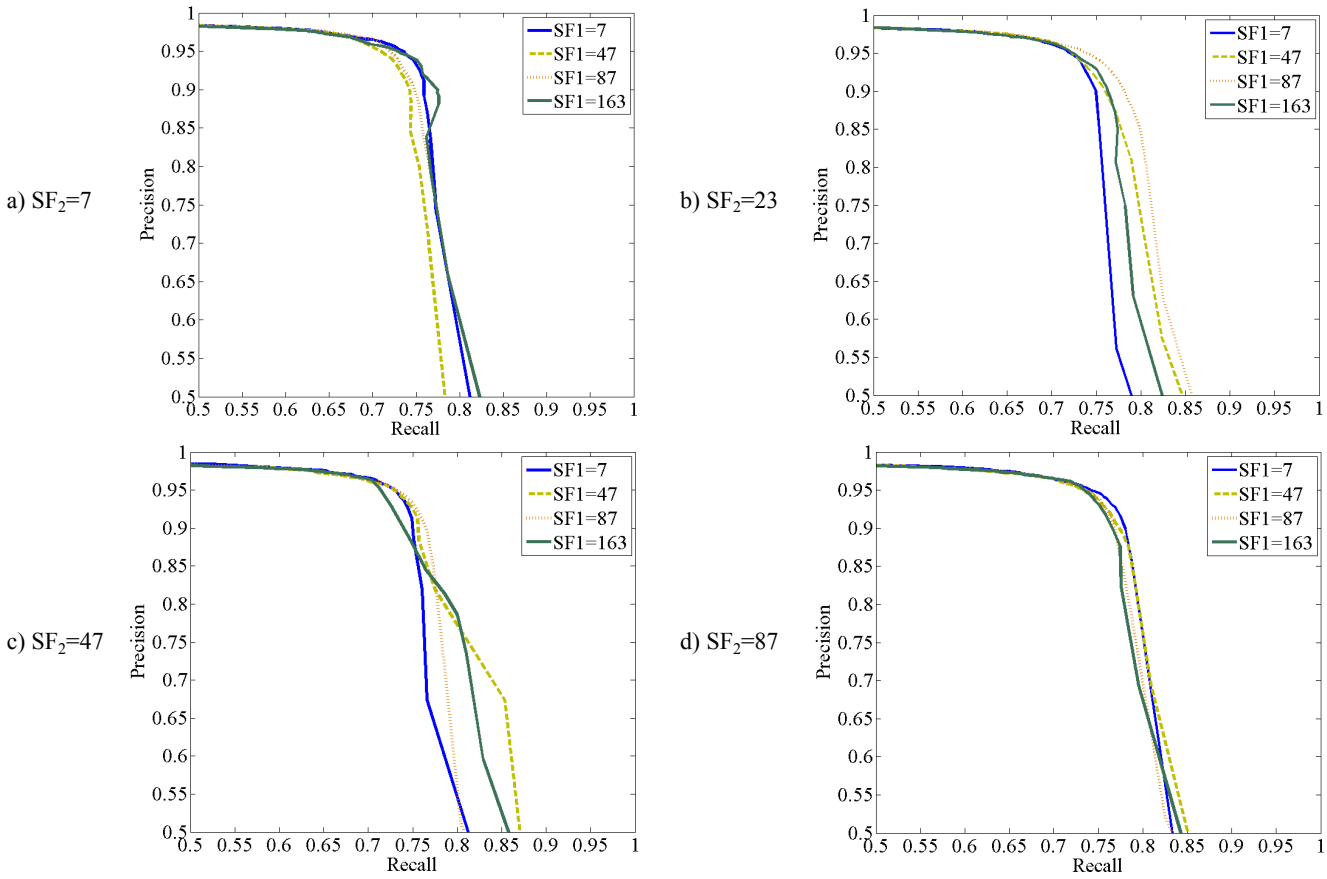


Figure 5: P-R curves for different training sizes of the two stages of the MCS

picture on the segmentation algorithm performance in terms of error rates and operating points for various application requirements.

3.1. Algorithm's robustness

As mentioned in Section 2.2, the MCS is trained on 100 by 100 pixels patches, but as the sample training size increases, with a number of 10000 features per patch training will be computationally expensive. For this reason, a sampling factor SF_1 is defined for reducing the training size at the stage 1 classifiers and a sampling factor SF_2 for amending the training size for the FFNN. For example, if the sampling factor is equal to 7, the training size for 120 patches (60 for sclera regions and 60 for non-sclera regions) will be 154436, while for a sampling factor equal to 87, the training size for 120 patches will be only 6696.

The research question from the final paragraph in Section 2.2 is addressed now: is the algorithm's performance affected by increasing SF_1, while keeping SF_2 constant? In Figure 5, P-R curves are plotted for different values of SF_1, while keeping SF_2 unchanged. As it may be observed from Figure 5, the performance of the algorithm is not significantly affected by varying SF_1 when SF_2 is unchanged. By further analyzing Figure 5 it may be observed further that the stability of the proposed sclera segmentation algorithm is more pronounced for some values of SF_2 (i.e. 7 and 87), while for other values (i.e. 23 and 47) the P-R curves are not so close to each other. However, in all the P-R curves from Figure 5 it may be observed how the precision of the system remains above 95% for recall values of around 75%. Examples of the output of the algorithm run on poorly illuminated images for 3 gaze directions are shown in Figure 6. The equal error rates (EER) of the system are given in Table 1. As indicated by the values in Table 1, the system tends to have a slightly better performance when SF_1 or SF_2 have large values. The proposed approach is the winning algorithm of SSBC 2015.

Table 1. Performance evaluation on SSBC 2015 database

	EER [%]					
	SF_2=1	SF_2=7	SF_2=23	SF_2=47	SF_2=87	SF_2=325
SF_1=7	19.85	23.40	19.85	20.98	18.74	**17.21**
SF_1=47	24.03	22.42	19.93	24.14	**18.83**	23.63
SF_1=87	22.20	22.25	**19.51**	22.65	22.31	24.66
SF_1=163	20.13	22.25	22.38	20.68	**20.05**	24.57
SF_1=325	21.57	20.09	18.44	20.14	**18.16**	19.89

3.2. Discussion

As illustrated in Figure 5, the behavior of the system is

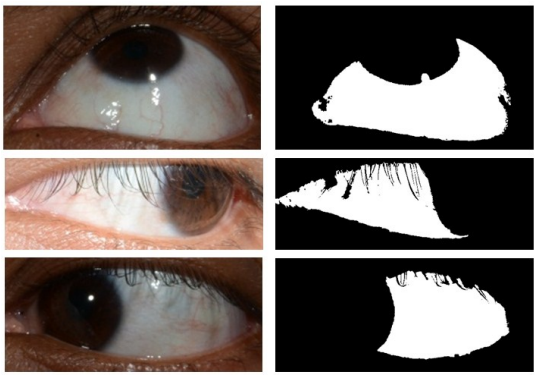


Figure 6: Output of the segmentation algorithm for different gaze directions of images with class id 11 from SSBC 2015 database

not highly sensitive to the choice of the values of SF_1 and SF_2, indicating that the proposed MCS architecture is a highly robust one for the task of sclera segmentation. Therefore, the proposed approach is suitable to be employed in other image analysis tasks, where smooth regions of relatively constant colour have to be found. From the biometric applications' perspective, the complexity of the classification stage of the present sclera segmentation algorithm is compensated by the reduced feature size (only 36 components).

Moreover, as the performance of the system does not rely on the size of the training data, the training time for the proposed MCS can be significantly decreased. For example, for the FFNN, the training time can be lowered from roughly 7500 seconds for SF_2=7 to about 660 seconds for SF_2=87 on an Intel i7 processor. It is also noteworthy that if the operating speed of the first stage classifiers depends on the size of the training size (e.g. k-nearest neighbor classifier), the employed MCS technique is able to decrease the execution speed without compromising the performance of the algorithm.

4. Conclusions

Sclera segmentation is a relatively new research topic in biometrics. Unlike the traditional approaches, where prior information about iris location or eyelid locations is necessary, this work has developed a novel sclera segmentation approach for pixel-level detection. Employing 3 types of features, the proposed algorithm is robust to noise factors affecting the eye image quality. The 2-stage MCS architecture employed enhances the algorithm's robustness to the size of the training data. The proposed approach was ranked 1st in SSBC 2015 with a precision rate of 95.05% and a recall rate of 94.56%.

Acknowledgements

This work was supported by the EU FASTPASS project under grant agreement 312583.

References

[1] Z. Zhi, E. Y. Du, N. L. Thomas, and E. J. Delp. A New Human Identification Method: Sclera Recognition. IEEE Trans. Syst., Man and Cyb., Part A, 42: 571-583, 2012.

[2] R. Derakhshani, A. Ross, and S. Crihalmeanu. A New Biometric Modality Based on Conjunctival Vasculature. Artificial Neural Networks in Enineering, 2006.

[3] P. Radu, K. Sirlantzis, G. Howells, F. Deravi, and S. Hoque. A Review of Information Fusion Techniques Employed in Iris Recognition Systems. Int'l J. of Advanced Intelligence Paradigms, 4 (3/4): 211-240, 2012.

[4] S. Crihalmeanu and A. Ross. Multispectral scleral patterns for ocular biometric recognition. Pattern Recognition Letters, 33:1860-1869, 2012.

[5] M. Marcon, E. Frigerio, and S. Tubaro. Sclera segmentation for gaze estimation and iris localization in unconstrained images. Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applic, 3: 25-29, 2012.

[6] A. Das, U. Pal, M. Blumenstein, and M. A. F. Ballester. Sclera Segmentation Benchmarking Competition 2015. http://www.ict.griffith.edu.au/conferences/btas2015

[7] M. H. Khosravi and R. Safabakhsh. Human eye sclera detection and tracking using a modified time-adaptive self-organizing map. Pattern Rec., 41(8): 2571-2593, 2008.

[8] S. Crihalmeanu, A. Ross, and R. Derakhshani. Enhancement and Registration Schemes for Matching Conjunctival Vasculature. Adv. in Biometrics. 5558: 1240-1249, 2009.

[9] A. Das, U. Pal, M. Blumenstein, and M. A. Ferrer Ballester. Sclera Recognition - A Survey. 2nd IAPR Asian Conf. on Pattern Rec. *(ACPR)*: 917-921, 2013.

[10] A. Das, U. Pal, M. A. Ferrer Ballester and M. Blumenstein. A new efficient and adaptive sclera recognition system. Computational Intellig. in Biometrics and Identity Management, IEEE Symposium on, 2014.

[11] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms: John Wiley & Sons, 2004.

[12] H. Proenca. Iris Recognition: A method to segment visible wavelength iris images acquired on-the-move and at-a-distance. Adv. in Visual Comp., 5358: 731-742, 2008.

[13] T. Chun-Wei and A. Kumar. Automated segmentation of iris images using visible wavelength face images. Proc. Comp. Vis. Patt. Rec. Wkshp. (CVPRW): 9-14, 2011.

[14] A. Khotanzad and H. Yaw Hua. Invariant image recognition by Zernike moments. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 12: 489-497, 1990.

[15] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, IEEE Comp Soc Conf on, 2005, 1: 886-893, 2005.

[16] R. A. Fisher. The use of multiple measurements in taxonomic problems. Annals of Human Genetics. 7: 179-188, 1936.

[17] H. Proença and L. Alexandre. UBIRIS: A Noisy Iris Image Database. Image Analysis and Processing. In Image Analysis and Processing. 3617: 970-977, 2005.

[18] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst, Man and Cyb. 9: 62-66, 1979.

[19] P. Vandewalle, J. Kovacevic, and M. Vetterli. Reproducible research in signal processing. IEEE Signal Processing Magazine. 26: 37-47, 2009.

[20] D. Yadav, N. Kohli, J. Doyle, R. Singh, M. Vatsa, K.W. Bowyer, Unraveling the Effect of Textured Contact Lenses on Iris Recognition, IEEE Trans. IFS. 9(5):851-862, 2014.