ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

Learning features on tear film lipid layer classification

Beatriz Remeseiro, Verónica Bolón-Canedo, Amparo Alonso-Betanzos and Manuel G. Penedo *

Departamento de Computación, Universidade da Coruña
Campus de Elviña s/n, A Coruña 15071, Spain

Abstract. Dry eye is a prevalent disease which leads to irritation of the ocular surface, and is associated with symptoms of discomfort and dryness. The Guillon tear film classification system is one of the most common procedures to diagnose this disease. Previous research has demonstrated that this classification can be automatized by means of image processing and machine learning techniques. However, all approaches for automatic classification have been focused on dark eyes, since they are most common in humans. This paper introduces a methodology making use of feature selection methods, to learn which features are the most relevant for each type of eyes and, thus, improving the automatic classification of the tear film lipid layer independently of the color of the eyes. Experimental results showed the adequacy of the proposed methodology, achieving classification rates over 90%, while producing unbiased results and working in real-time.

1 Introduction

Dry eye syndrome is recognized as a growing health problem, and one of the most frequent reasons for seeking eye care. One of the clinical tests used for dry eye diagnosis is the *lipid layer pattern assessment*, which consists in categorizing tear film images acquired with the Tearscope Plus and using the Guillon classification system [1]. This system is based on a grading scale composed of different categories: open meshwork, closed meshwork, wave, and color fringe. There is no doubt that this classification system is a valuable technique which provides relevant information about the tear film, and so allows clinicians to diagnose dry eye syndrome. Furthermore, the Tearscope Plus and the grading scale defined by Guillon have proven their validity to carry out this task [2].

Lipid layer patterns do not depend on the eye color, although it has been said that optometrists find more difficult to categorize them in light eyes. In fact, Efron [2] suggests the use of two different grading scales based on the Guillon tear film classification system, one for dark eyes and another for light eyes. Therefore, the main objective of this paper is to select the most relevant features for the

*This research has been partially funded by the Secretaría de Estado de Investigación of the Spanish Government and FEDER funds of the European Union through the research projects TIN2012-37954 and PI14/02161; and by the Consellería de Industria of the Xunta de Galicia through the research projects GPC2013/065 and GRC2014/035. We would also like to thank the Optometry Service of the University of Santiago de Compostela (Spain) for providing us with the annotated datasets.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

task of tear film classification and to adapt the existing methodology, making it adequate to work also with light eyes.

Related work: A real-time application for tear film classification was presented in [3], using texture analysis and feature selection. Due to the heterogeneity of the lipid layer, tear film maps were presented in [4] to detect several patterns in one single image by applying the global methodology at a local level, using decision voting systems. However, these automatic tools were designed for dark eyes, the most common ones in human beings. To the best knowledge of the authors, there is no attempt in the literature to automatically classify the tear film patterns observed in light eyes.

Our framework: The framework proposed herein takes advantage of feature selection techniques to consider both dark and light eyes for tear film lipid layer classification, thus it can be used in clinical and research settings to improve the diagnosis and treatment of dry eye. Moreover, by reducing the number of features required for classification, the time to extract the features from the images is reduced accordingly. So, our framework makes three important contributions: (1) it is able to tackle tear film classification regardless of the eye color; (2) it provides a reliable tear film classification for Tearscope images; and (3) it works in real-time.

2 Research methodology

A four-step methodology is proposed to obtain an efficient system for automatic tear film classification regardless of the eye color. In what follows, every step will be explained in depth.

2.1 Region of interest

Experts that analyze these images focus on the bottom part of the iris, in which the tear film can be perceived with higher contrast. Thus, the whole analysis takes place in this area called the *region of interest* (ROI), which can be selected as follows: (1) the green channel of the input image in RGB is selected in this stage; and (2) a rectangular ROI is located inside the lightest area of the iris by analyzing the value of the pixels from the center of the pupil to the bottom part of the image (see Figure 1).

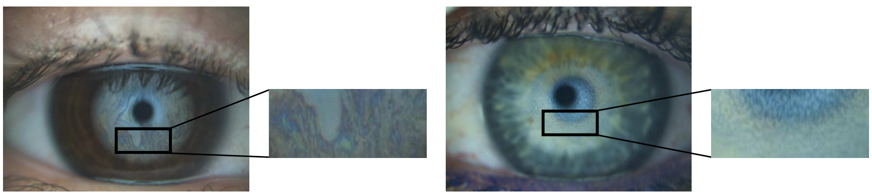

Fig. 1: ROIs of two representative images, from left to right: a dark eye with a color fringe pattern, and a light eye with a closed meshwork pattern.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

2.2 Feature vector

The ROI is analyzed in terms of color and texture, and a feature vector is obtained using the following methods [3]:

- *CIE 1976 L*a*b** [5]. It is a chromatic color space used in this research for color analysis. It is perceptually uniform, an important characteristic since clinicians' perception is being imitated. Thus, the ROI is transformed from RGB to CIELAB, to subsequently analyze the texture of its three channels.
- *Co-occurrence features* [6]. It is an effective method for texture analysis, which generates a set of gray level co-occurrence matrices, and extracts 14 statistical measures from them. The mean and range of these statistics are calculated across matrices to obtain a texture descriptor.

2.3 Feature selection

Feature selection techniques can be divided into filters, wrappers and embedded [7]. Both wrappers and embedded methods have the risk of overfitting when having more features than samples [8], as in this research. Therefore, the following filters were chosen for a comparison study after subsequent classification:

- *Correlation-based feature selection* (CFS) [9]. This multivariate filter ranks feature subsets according to a correlation based heuristic evaluation function, whose bias is toward subsets that contain features that are highly correlated with the class and uncorrelated with each other.
- *Consistency-based filter* [10]. This algorithm evaluates the worth of a subset of features by the level of consistency in the class values when the samples are projected onto the subset of attributes.
- *Mutual Information Maximization* (MIM) [11]. This univariate filter simply ranks the features in order to their mutual information score between each feature and the class.
- *minimum Redundancy Maximum Relevance* (mRMR) [12]. This multivariate method selects features that have the highest relevance with the target class and are also minimally redundant, i.e. selects features that are maximally dissimilar to each other. Both optimization criteria (Maximum-Relevance and minimum-Redundancy) are based on mutual information.

2.4 Classification

Finally, the feature vector has to be classified into one of the four Guillon categories. In this stage, five popular machine learning algorithms were selected aiming to provide different approaches of the learning process: C4.5, Naive Bayes, Ib1, Random Forest (all of them available in Weka [13]), and libSVM [14]. A 10-fold cross validation is performed.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

3 Experimental results

The proposed methodology has been tested on two datasets acquired and annotated by experienced optometrists, with subjects aged from 19 to 33 years. These two datasets are: (1) VOPTICAL_I1 dataset, composed of 105 images of dark eyes –29 open meshwork, 29 closed meshwork, 25 wave and 22 color fringe–; and (2) VOPTICAL_L dataset, composed of 108 images of light eyes –30 open meshwork, 28 closed meshwork, 27 wave and 23 color fringe–.

Table 1 shows the test accuracies for all pairwise feature selection methods and classifiers applied to both datasets, and the combination of them (last column). Note that "No FS" is the baseline classification without feature selection. Since MIM and mRMR return a ranking of features, a threshold is necessary, and in this work we have opted for choosing the top 5%, 10%, 15% and 20% of the ranked features. For the sake of brevity, only the best result among the different thresholds is shown in the table. As can be seen, all classifiers perform quite well providing results over 80% accuracy in some combinations, although the best results for each dataset (marked in bold face) are obtained with the libSVM classifier. Notice that, in general, the classification accuracies achieved on the dataset with dark eyes are higher than those obtained on the light eye dataset, confirming the added difficulty of this task. Moreover, it seems that each type of eyes requires a custom treatment, since the best results were obtained when applying the mRMR filter (threshold 15%) for dark eyes, while light eyes' best results were obtained with the CFS method. Additionally, the best accuracy achieved for the combined dataset is lower that the best result obtained with any of the other two datasets, which also seems to indicate the adequacy of using a different treatment for each type of eye color.

Now that we have obtained acceptable classification accuracies for the three datasets considered, we will focus on the time required to extract the corresponding features, since the temporal cost for obtaining the whole set of features (588) is not homogeneous. In fact, it has been demonstrated in [3] that the key is to remove the features related with the so-called 14^{th} statistic, which corresponds with the maximal correlation coefficient [6]. Consequently, we have eliminated the 14^{th} statistics and analyzed the impact on both accuracy and processing time. Table 2 shows the results (in terms of classification accuracy, time to extract the features and number of selected features) for the best configuration for each dataset. Since the best result for the dataset which combines dark and light eyes was obtained without feature selection, leading to a high extracting time, we have also included the second best configuration which in fact uses feature selection (last row). Notice that reducing the time required to extract the features is paramount since (1) the automated tool has to work in real-time; and (2) the methodology has to be applied at a local level, i.e. over thousands of windows, to create the tear film maps (as mentioned in the Introduction). Consequently, a non real-time approach would make the problem of tear film maps unapproachable. Note that the results obtained when removing the 14^{th} statistic showed a noticeable reduction in processing time while the accuracy did

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

not drop to inadmissible values –in one case it even improved–. Furthermore, the classification rates surpass now 92% regardless of the type of eye color.

		Dataset		
Classifier	Filter	Dark eyes	Light eyes	Dark & Light eyes
	No FS	75.24	78.70	74.18
	CFS	78.10	75.93	80.75
C4.5	Cons	72.38	80.56	80.75
	MIM	74.29	72.22	70.89
	mRMR	76.19	67.59	77.47
	No FS	79.05	75.00	73.71
	CFS	81.91	85.19	77.93
Naive Bayes	Cons	70.48	78.70	76.53
	MIM	78.10	67.59	69.01
	mRMR	80.95	75.00	73.24
	No FS	88.57	75.00	83.57
	CFS	87.62	83.33	84.04
Ib1	Cons	77.14	77.78	75.12
	MIM	86.67	71.30	73.71
	mRMR	86.67	73.15	77.93
	No FS	84.76	78.70	82.16
	CFS	86.67	81.48	82.63
Random Forest	Cons	80.95	81.48	85.45
	MIM	83.81	71.30	77.93
	mRMR	81.91	72.22	80.75
	No FS	92.38	90.74	**90.14**
	CFS	93.33	**91.67**	89.67
libSVM	Cons	88.57	90.74	85.45
	MIM	91.43	84.26	84.51
	mRMR	**94.29**	74.07	86.86

Table 1: Test classification accuracy.

Dataset	Filter	with 14^{th}	without 14^{th}
Dark eyes	mRMR 15%	94.29% 8.06 s 88 feats.	92.38% 0.80 s 76 feats.
Light eyes	CFS	91.67% 2.97 s 37 feats.	92.59% 0.62 s 34 feats.
Dark & Light eyes	No FS	90.14% 38.27 s 588 feats.	89.67% 8.62 s 546 feats.
	CFS	89.67% 2.36 s 30 feats.	87.32% 0.52 s 28 feats.

Table 2: Case study with libSVM and best configuration for each dataset.

4 Conclusions

The tear film classification systems proposed in previous studies are focused on dark eyes since these are the most common. In this research, a methodology to solve this classification problem for any type of eye color is presented. This methodology includes the use of color and texture analysis techniques to obtain the learning features, feature selection filters to select the most relevant ones, and different classifiers. All these methods have been tested on three datasets: dark eyes, light eyes, and their combination.

Results obtained with this methodology proved its adequacy for tear film classification when using Tearscope images of any eye color. Additionally, they suggest the use of two different systems, one per eye type color, and demonstrate that it is possible to classify any kind of eye with a classification accuracy over 92% in less than one second, thanks to the effectiveness of feature selection thus allowing real-time use. The clinical relevance of these results should be highlighted since the agreement between subjective observers is over 91%. As future research, we plan to use the proposed methodology at a local level and create tear film maps for light eyes too.

References

[1] J.P. Guillon. Non-invasive Tearscope Plus routine for contact lens fitting. *Cont Lens Anterior Eye*, 21 Suppl 1, 1998.

[2] N. Efron. *Contact Lens Complications (3rd edition).* Elsevier, 2012.

[3] B. Remeseiro, V. Bolon-Canedo, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdiñas, A. Mosquera, M. G. Penedo, and N. Sánchez-Maroño. A Methodology for Improving Tear Film Lipid Layer Classification. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1485–1493, 2014.

[4] B. Remeseiro, A. Mosquera, M.G. Penedo, and C. García-Resúa. Tear film maps based on the lipid interference patterns. In *6th International Conference on Agents and Artificial Intelligence*, pages 732–739, 2014.

[5] K. McLaren. The development of the CIE 1976 (L*a*b) uniform colour-space and colour-difference formula. *Journal of the Society of Dyers and Colourists*, 92(9):338–341, 1976.

[6] R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621, 1973.

[7] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. *Feature Extraction: Foundations and Applications*. Springer Verlag, 2006.

[8] J. Loughrey and P. Cunningham. Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets. *Research and Development in Intelligent Systems XXI*, pages 33–43, 2005.

[9] M. A. Hall. *Correlation-based feature selection for machine learning.* PhD thesis, The University of Waikato, 1999.

[10] M. Dash and H. Liu. Consistency-based search in feature selection. *Artificial intelligence*, 151(1-2):155–176, 2003.

[11] G. Brown, A. Pocock, M. Zhao, and M. Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research*, 13(1):27–66, 2012.

[12] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[14] C.C. Chang and C.J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1–27, 2011.