On the analysis of local and global features for hyperemia grading

L. Sánchez[1], N. Barreira[1], N. Sánchez[1], A. Mosquera[2], H. Pena-Verdeal[3] and E. Yebra-Pimentel[3]
[1]Department of Computer Science, University of A Coruna, Spain luisa.brea@udc.es
[2]Department of Electronics and Computer Science, University of Santiago de Compostela, Spain
[3]Department of Applied Physics, University of Santiago de Compostela, Spain

ABSTRACT

In optometry, hyperemia is the accumulation of blood flow in the conjunctival tissue. Dry eye syndrome or allergic conjunctivitis are two of its main causes. Its main symptom is the presence of a red hue in the eye that optometrists evaluate according to a scale in a subjective manner. In this paper, we propose an automatic approach to the problem of hyperemia grading in the bulbar conjunctiva. We compute several image features on images of the patients' eyes, analyse the relations among them by using feature selection techniques and transform the feature vector of each image to the value in the adequate range by means of machine learning techniques. We analyse different areas of the conjunctiva to evaluate their importance for the diagnosis. Our results show that it is possible to mimic the experts' behaviour through the proposed approach.

Keywords: Hyperemia grading, Optometry, Feature selection, Regression, Medical imaging.

1. INTRODUCTION

The bulbar conjunctiva posses a special relevance in the diagnosis of pathologies such as dry eye syndrome or allergic conjunctivitis. Both pathologies are widely spread around the world population, hence making their early diagnosis a necessity. One of the first symptoms that appear in an unhealthy conjunctiva is the engorgement of the blood vessels, known as hyperemia, that produces a distinctive red colouration in the affected area. In order to measure hyperemia, specialists analyse an image of the patient's eye to find a series of features related with hyperemia, such as hue of the conjunctiva or vessel disposition. Then, they compare the patient's image with a grading scale and decide the grading in consequence. Hyperemia grading scales are collections of images, drawings or photographies, that represent different levels of severity. In this article, we work with the two scales depicted in Fig. 1. The specialists will assign their patient a grade that usually includes a decimal value that indicates the distance between the closest prototype of the scale and the patient's eye. The main drawback of the manual method is the high inter- and intra-expert subjectivity, as well as the lack of repeatability. Therefore, the automation of the process is fundamental to provide objective and repeatable results.

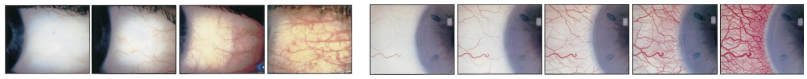

Figure 1: CCLRU (left, values 1-4) and Efron (right, values 0-4) grading scales.

There are several works on the implementation of image features in order to perform an objective grading of bulbar hyperemia [9, 19]. However, the features are compared with the expert evaluation one by one, without having into account neither the effects of their combinations nor the procedure chosen to transform the computed values into the grading scale levels. In [13] we analyse features computed in the whole conjunctiva. However, the specialists sometimes pay special attention to different image features depending on which side of the eye is affected. Note that in Fig. 1 each scale shows a different part of the eye.

In this article, we compute several image features that measure both local and global conjunctiva factors. Then, we analyse the set by means of feature selection techniques in order to obtain a reduced subset without worsening the results of the system. Finally, we apply machine learning techniques in order to transform the feature vector of each image in the value in a given grading scale. These steps are integrated in a fully automatic methodology for hyperemia grading. The system receives a video as input, and produces the grade in the selected scale. It consists of four steps: the selection

This paper has been partially funded by the Ministerio de Economía y Competividad through the research project DPI2015-69948-R

Ninth International Conference on Machine Vision (ICMV 2016), edited by Antanas Verikas, Petia Radeva, Dmitry P. Nikolaev, Wei Zhang, Jianhong Zhou, Proc. of SPIE Vol. 10341, 103411T

of the best frame of the video [15], the segmentation of the conjunctiva [14], the computation of the image features and the transformation of the feature vectors in the final value.

This paper is structured as follows: Section 2 will depict the implemented image features. Section 3 will explain the process to transform these features in a grade in the given scale. Section 4 will show the obtained results and, finally, Section 5 will elaborate our conclusions and future lines of work.

2. HYPEREMIA FEATURES

We analysed previous works on the subject of automatic feature computation [9] and combined them with insight from the optometrists from the Optometry Group of the University of Santiago de Compostela in order to implement the image features depicted in Fig. 2.

$$B_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((R_{ij}+G_{ij})\overline{VE}_{ij})}{nm} \qquad B_8 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((V_{ij}+S_{ij})\overline{VE}_{ij})}{nm} \qquad V_6 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{H_{ij}VE_{ij}}{\mu_{ij}} \qquad I_5 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}a_{ij}}{nm}$$

$$B_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((|240-H_{ij}|)\overline{VE}_{ij})}{nm} \qquad B_9 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((L_{ij})\overline{VE}_{ij})}{nm} \qquad \mu_{ij} = \frac{\sum_{k=-s/2}^{s/2}\sum_{l=-s/2}^{s/2}\overline{VE}_{ij}Hi+k,j+l}{s^2} \qquad C_v = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}E_{ij}M_{ij}}{n_r}$$

$$B_3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(b_{ij}\overline{VE}_{ij})}{nm} \qquad V_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{ij}VE_{ij}}{R_{ij}+G_{ij}+B_{ij}}\right) \qquad V_7 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ij}VE_{ij})}{nm} \qquad M_{ij} = \begin{cases} 0 & i \bmod step \neq 0 \\ 1 & i \bmod step = 0 \end{cases}$$

$$B_4 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(R_{ij}\overline{VE}_{ij})}{nm} \qquad V_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((R_{ij}-G_{ij})VE_{ij})}{nm} \qquad I_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{ij}}{R_{ij}+G_{ij}+B_{ij}}\right) \qquad A_v = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}VE_{ij}}{nm}$$

$$B_5 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(|128-H_{ij}|)\overline{VE}_{ij})}{nm} \qquad V_3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((R_{ij}-B_{ij})VE_{ij})}{nm} \qquad I_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(R_{ij}-G_{ij})}{nm} \qquad P_v = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}VE_{ij}}{nm}100$$

$$B_6 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((a_{ij})\overline{VE}_{ij})}{nm} \qquad V_4 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}R_{ij}VE_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m}VE_{ij}}100 \qquad I_3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(R_{ij}-B_{ij})}{nm} \qquad W_v = \frac{\sum_{r=1}^{\rho}\sum_{c=1}^{\kappa}W_{rc}}{\rho\kappa}$$

$$B_7 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}((R_{ij}+G_{ij}+B_{ij})\overline{VE}_{ij})}{nm} \qquad V_5 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}H_{ij}VE_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m}VE_{ij}}100 \qquad I_4 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}|128-H_{ij}|}{nm}$$

Figure 2: Implemented hyperemia features.

B_n, V_n and I_n measure the colour levels in the background of the conjunctiva, the vessels and the whole conjunctiva respectively. The first and second types of features apply the Canny algorithm in order to distinguish vessels from background. E is the edge image, VE is the set of vessel edges inside the conjunctiva and VE comprises the remaining pixels of the conjunctiva. The third group obtains the average value in all the pixels of the conjunctiva. Also, there are three vessel quantity-related features: C_v, A_v and P_v . Finally, we measure the average width of the vessels (W_v). In the formulas, *n* and *m* indicate the size of the input image *I*, but considering only the pixels that belong to the conjunctiva; *i* and *j* represent the position (row, column) of the current pixel in the image; *R*, *G* and *B* indicate the channel value in RGB colourspace; *H*, *S* and *V* represent the channel value in HSV colourspace; *L*, *a* and *b* represent the channel value in L*a*b* colourspace. In the feature C_v, n_r is the number of image rows considered, and *M* is a mask. The feature V_6 computes the red hue value taking into account the values of the neighbouring pixels. μ is the value for this neighbourhood, where *s* is the size of the considered window. Finally, the feature W_v measures the average with of the vessels. In order to perform this task, a set of κ circumferences are defined, with radius ρ ranging from $n/2 \square \kappa$ to $n/2$. *W* represents the width values for the cut points. These values are computed using an active contour algorithm [18].

The same set of features was computed in the whole image and in both sides of the image: the pupil side and the lacrimal side. Subscripts $1 - 25$ refer to the features computed in the whole conjunctiva, $26 - 50$ to those computed in the pupil side, and $51 - 75$ to those computed in the lacrimal side.

3. HYPEREMIA GRADING

Once we have computed the required image features, we obtain a vector with 75 values for each image. Each feature provides insight to the hyperemia level, but we need to output a single value in the range of the grading scale in order to establish a comparison with the experts' ground truth. This task is not covered in the bibliography, but poses a strong relevance in the process, as it will decide the contribution of each feature in the final result. As several previous works hint at the relationship between automatic features and scale values being complex, we decided to tackle the problem by using machine learning techniques. Specifically, we tested several regression techniques, that can be divided in artificial neural networks (ANNs) (multi-layer perceptron (MLP) [2], radial-basis function network (RBFN) [10], self-organising map (SOM) [8]) and classifiers (random forests (RF) [3], support vector regression (SVR) [17], k-nearest neighbours (KNN) [4], partial least squares (PLS) [1], naive bayes (NB) [7]).

By looking at the feature's formulation, we can see that some of the features compute similar parameters, which can be a hint of the existence of a relationship. Moreover, each feature is computed in three areas of the image, so we expect some redundancy to appear. This situation is far from ideal, as it can add noise to the system or, in the best case scenario, redundant information that do not contributes to improve the system. Therefore, we need to consider only the relevant features as the input of the systems, a goal that can be achieved by means of feature selection. Feature selection techniques analyse each feature of the initial set in order to find certain relationships with other features (such as correlation or information gain). Their aim is to decide which features are more informative, obtaining a small subset to work with, but maintaining nearly the same information than the initial set.

There are several approaches to feature selection, divided in three main groups: filters, wrappers, and embedded methods. Filters compute a certain parameter on each analysed feature subset. Wrappers, however, build a predictive model for the given feature subset, and then analyse the accuracy of the model. Finally, embedded methods do both the training process of the prediction model and the feature selection simultaneously. Filters are usually the fastest of the three. However, as they do not posses knowledge of the predictive model, they can provide less accurate results. Embedded methods have the highest computational cost. We tested five methods [13]: two filters, CFS [6] and Relief [12]; two wrappers, M5 [11] and SMOReg [16]; and an embedded method, SVR-RFE [5].

4. RESULTS

For the computation of the image features, our database consisted of 163 images of the bulbar conjunctiva. Two specialists graded the whole set. Besides, both specialists performed a second grading months apart. As hyperemia grading presents a high intra- and inter-expert subjectivity, we analysed the evaluations of the experts in order to find those where the experts differ less. We considered a variation of 0.5 points in the scale between different experts or times as the maximum allowed, and discarded the images that present a larger variation. In the end, our image set consisted of 114 images that fulfilled the condition. We used the mean value of the four evaluations (two gradings of each expert) as our ground truth. After computing the 75 features for each of the images, we used that vectors as input for the feature selection techniques. Regarding the values for the parameters, s, n_r, *step* and κ were set to 3, 10, 10 and 10, respectively.

As we used 10-fold cross validation during the process, we obtain 10 different sets of features that we need to combine. Although there are several options to perform this combination, we have chosen to include in the subset the features that appear in at least the 70% of the folds. This percentage was empirically selected, as it is inclusive enough while maintaining a small number of features in the result. We tested lower percentages, but the mean square error (MSE) values were similar. Besides the subsets of features selected by each method, we decided to combine the outputs and create a last feature subset as the union of all the former subsets. The resulting subsets are depicted in Fig. 3.

Method	Efron	CCLRU
CFS	V_{13}, V_{715}, V_{128}, I_{539}	V_{13}, V_{715}, I_{539}, I_{154}
Relief	C_{v1}, A_{v2}, V_{13}, P_{v10}, V_{613}, A_{v27}, V_{128}, P_{v35}	C_{v1}, A_{v2}, V_{13}, P_{v10}, V_{613}, V_{128}
M5	V_{13}, V_{715}, I_{539}	V_{13}, V_{715}
SMOReg	V_{13}, V_{715}, V_{537}	V_{13}, V_{715}
SVR-RFE	C_{v1}, I_{49}, W_{v25}, V_{537}, W_{v50}, B_{873}	C_{v1}, I_{49}, W_{v25}, V_{537}, W_{v50}, V_{765}
Union	C_{v1}, A_{v2}, V_{13}, I_{49}, P_{v10}, V_{613}, V_{715}, W_{v25}, A_{v27}, V_{128}, P_{v35}, V_{537}, I_{539}, W_{v50}, B_{873}	C_{v1}, A_{v2}, V_{13}, I_{49}, P_{v10}, V_{613}, V_{715}, W_{v25}, V_{128}, V_{537}, I_{539}, W_{v50}, I_{154}, V_{765}

Figure 3: Selected features for each method.

In view of the data, we can see how some features appear in most cases, such as V_7, that appears in 3 out of 5 methods for both scales, and V_1, that appears in 4 out of 5. Most of the selected features are computed in the whole conjunctiva. CFS and the two wrappers obtain smaller subsets in comparison to the ranker and embedded method. Half of the features selected by the embedded method and CFS are computed in one side of the conjunctiva. That these features are part of the selection supports the idea that some image characteristics have a higher influence in the grading if they take place in a certain area.

Once we have decided which features are most relevant, the final stage is to use the reduced feature vectors as input to train the regression techniques. We obtained the MSE for each individual feature, and also for the combinations provided by each feature selection technique. The parameters for all the regression techniques are detailed in Fig. 4. 10-fold cross-validation was also used during the configuration of the methods.

Method	Parameters
KNN	number of neighbours = 1, distance = cosine
MLP	configuration = [40 16], activation function = hyperbolic tangent sigmoid training function = Bayesian regularization backpropagation based on Levenberg-Marquardt optimization
NB	determination of prototypes using a KNN model with number of neighbours = 3 distribution type = mvmn (Efron) and kernel with normal kernel (CCLRU)
PLS	number of components = min(number of features, 8)
RBFN	spread = 0.4, error goal = 0.03
RF	number of trees = 60 (Efron) and 40 (CCLRU), minimum leaf size = 10
SOM	competitive layer size = 8 (Efron) and 6 (CCLRU) number of neighbours = 3, topology function = one-dimensional random pattern distance function = Manhattan, configuration of the MLP = [10]
SVR	type = ν-SVR (Efron) and ϵ-SVR (CCLRU), kernel = sigmoid (Efron) and radial basis function (CCLRU) $\gamma = 2^{-12}$ (Efron) and 2^{-10} (CCLRU), C = 2^8 (Efron) and 2^4 (CCLRU)

Figure 4: Parameters of the methods.

In the case of the individual features, the lowest MSE values were 0.0535 and 0.0365, whereas the highest were 0.611 and 0.925 for Efron and CCLRU scales respectively. The average MSE value for the best method was 0.1896 in Efron and 0.1219 in CCLRU. Regarding the combination of features, Fig. 5 depicts the obtained results. We can observe how the best result is improved in both scales if compared with one feature individually. Also, the worst case in the combinations presents a lower MSE than in the one by one test.

	Efron						CCLRU					
Method	All	CFS	Relief	M5	SMOReg	SVR-RFE	All	CFS	Relief	M5	SMOReg	SVR-RFE
KNN	0.152	0.141	0.292	0.182	0.148	0.220	0.083	0.077	0.160	0.105	0.105	0.125
MLP	0.009	0.053	0.224	0.054	0.052	0.075	0.141	0.032	0.141	0.036	0.036	0.141
NB	0.386	0.386	0.360	0.429	0.389	0.405	0.285	0.211	0.235	0.243	0.243	0.259
PLS	0.053	0.064	0.233	0.063	0.070	0.116	0.034	0.038	0.101	0.042	0.042	0.062
RBFN	0.374	0.232	0.473	0.160	0.255	0.225	0.199	0.065	0.327	0.104	0.104	0.143
RF	0.071	0.071	0.127	0.071	0.077	0.129	0.047	0.044	0.073	0.044	0.044	0.063
SOM	0.119	0.092	0.123	0.095	0.100	0.152	0.081	0.058	0.084	0.056	0.056	0.105
SVR	0.223	0.223	0.223	0.223	0.223	0.223	0.142	0.142	0.142	0.145	0.145	0.137

Figure 5: MSE for the combinations of the 75 features for all systems.

For Efron scale, the best values are achieved with the MLP with all the features. Taking into account that experts precision is one decimal value, we can assume that an excellent MSE will be $0.1^2 = 0.01$, as it points out that the system is wrong in one decimal at average. Data also shows that the MLP is able to obtain even lower error values. Also, the acceptable threshold between predictions that we are using with the experts' grading is 0.5, which implies that an MSE of 0.25 is the maximum acceptable. There are many approaches that fulfil this rule. The best value for CCLRU scale is obtained with the MLP in the CFS feature set. We can observe that CCLRU obtains worse results than Efron in general, with only a few approaches obtaining a MSE value below 0.25. This is motivated by the fact that Efron scale's prototypes are drawings and are equally spaced. However, the distances between CCLRU's prototypes are not linear and, as a consequence, experts have a harder time grading in CCLRU, a circumstance that is reflected in our system.

5. CONCLUSION

In this work, we analyse the impact of computing local and global features in a hyperemia grading framework. First, we define several image features and compute them in the whole conjunctiva and in its two sides. Then, we apply feature selection techniques in order to discover the most suitable set of features. Finally, we use regression techniques in order to transform the feature vectors in to a value in a grading scale. Our results show that the proposed techniques are adequate to the problem. The lowest MSE is obtained in both scales by the MLP with all the 75 features. However, CFS, M5 and SMOReg in combination with MLP, RF, PLS and SOM approaches produce also an MSE below 0.1. Moreover, half of CFS features are computed in one side of the conjunctiva, reinforcing the idea of the relevance of the local features.

Regarding future lines of work, we intend to improve our database with more images in order to further analyse our system's behaviour. We are also considering to analyse the evaluation of the symptom through time by comparing different checkups for the same patient.

REFERENCES

[1] Abdi, H.: Partial least square regression (PLS regression). Encyclopedia for research methods for the social sciences pp. 792–795 (2003)

[2] Baum, E.B.: On the capabilities of multilayer perceptrons. J. Complexity 4(3), 193–215 (1988)

[3] Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)

[4] Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. IEEE T Syst Man Cyb (4), 325–327 (1976)

[5] Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning 46(1-3), 389–422 (2002)

[6] Hall, M., Smith, L.: Practical feature subset selection for Machine Learning. In: ACSC. pp. 181–191 (1998)

[7] John, G.H., Langley, P.: Estimating continuous distributions in Bayesian classifiers. In: UAI. pp. 338–345. (1995)

[8] Kohonen, T.: The self-organizing map. Neurocomputing 21(1-3), 1–6 (1998)

[9] Papas, E.B.: Key factors in the subjective and objective assessment of conjunctival erythema. Invest. Ophthalmol. Vis. Sci. 41(3), 687–691 (2000)

[10] Park, J., Sandberg, I.W.: Universal approximation using radial-basis-function networks. Neural Computation 3(2), 246–257 (1991)

[11] Quinlan, J., et al.: Learning with continuous classes. In: AI 1992. Vol. 92, pp. 343–348 (1992)

[12] Robnik-Sikonja, M., Kononenko, I.: An adaptation of relief for attribute estimation in regression. In: Machine Learning: Proc. 14th Int. Conf. pp. 296–304 (1997)

[13] Sánchez, L., Barreira, N., Sánchez, N., Mosquera, A., García-Resúa, C., Yebra-Pimentel, E.: On the analysis of feature selection techniques in a conjunctival hyperemia grading framework. In: ESANN. pp. 271–276 (2016)

[14] Sánchez, L., Barreira, N., Pena-Verdeal, H., Yebra-Pimentel, E.: A novel framework for hyperemia grading based on artificial neural networks. In: Advances in Computational Intelligence, pp. 263–275 (2015)

[15] Sánchez-Brea, L., Barreira-Rodríguez, N., Mosquera-González, A., García-Resúa, C., Yebra-Pimentel, E.: Automatic selection of video frames for hyperemia grading. In: EUROCAST, pp. 479–486 (2015)

[16] Shevade, S., Keerthi, S., Bhattacharyya, C., Murthy, K.: Improvements to the SMO algorithm for SVM regression. Neural Networks, IEEE Trans. on 11(5), 1188–1193 (2000)

[17] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing 14(3), 199–222 (2004)

[18] Vázquez, S., Barreira, N., Penedo, M.G., Pena-Seijo, M., Gómez-Ulla, F.: Evaluation of SIRIUS retinal vessel width measurement in REVIEW dataset. In: CBMS. pp. 71–76 (2013)

[19] Wolffsohn, J.S., Purslow, C.: Clinical monitoring of ocular physiology using digital image analysis. Contact Lens and Anterior Eye 26(1), 27–35 (2003)