Tear Film Maps based on the Lipid Interference Patterns

Beatriz Remeseiro¹, Antonio Mosquera², Manuel G. Penedo¹ and Carlos García-Resúa³

¹Department of Computer Science, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
²Department of Electronics and Computer Science, University of Santiago de Compostela, Campus Universitario Sur, 15782 Santiago de Compostela, Spain
³Faculty of Optics and Optometry, University of Santiago de Compostela, Campus Universitario Sur, 15782 Santiago de Compostela, Spain

Keywords: Tear Film, Interference Patterns, Color Texture Analysis, Image Segmentation, Machine Learning.

Abstract: Dry eye syndrome is characterized by symptoms of discomfort, ocular surface damage, reduced tear film stability, and tear hyperosmolarity. These features can be identified by several types of diagnostic tests, although there may not be a direct correlation between the severity of symptoms and the degree of damage. One of the most used clinical tests is the analysis of the lipid interference patterns, which can be observed on the tear film, and their classification into the Guillon categories. Our previous researches have demonstrated that the interference patterns can be characterized as color texture patterns. Thus, the manual test done by experts can be performed through an automatic process which saves time for experts and provides unbiased results. Nevertheless, the heterogeneity of the tear film makes the classification of a patient’s image into a single category impossible. For this reason, this paper presents a methodology to create tear film maps based on the lipid interference patterns. In this way, the output image represents the distribution and prevalence of the Guillon categories on the tear film. The adequacy of the proposed methodology was demonstrated since it achieves reliable results in comparison with the annotations done by experts.

1 INTRODUCTION

Dry eye is a multifactorial disease of the tears and ocular surface. It is a common complaint among middle-aged and older adults, and affects a wide range of population (DEWS, 2007): between 10% and 20% of the population, although in Asian populations this percentage may be raised up to 33%. This disease worsens with age and so its prevalence may increase in the future (Brewitt and Sistani, 2001). Moreover, it can be a cause of great discomfort and frustration, and symptoms may be modulated by environmental factors such as air conditioning and working at video display terminals. Sufferers will require treatment with a significant potential cost, thus an early diagnosis and the monitoring of the disease are significantly important factors to ensure patients’ quality of life (Bron, 2001).

Diagnosis of dry eye is a difficult task due to its multifactorial etiology, the need for a comprehensive definition, and the use of tests that are limited and variable in their assessment of the tears and ocular surface. Thus, there is no uniform concept for diagnosis or treatment. However, the evaluation of the lipid interference patterns of the tear film lipid layer facilitates the diagnosis of the cause of dry eye symptoms. In this sense, Guillon invented the Tearscope plus as an instrument for rapid assessment of lipid layer thickness (Guillon, 1998), which was classified from open meshwork (very thin) to color fringe (very thick). This grading scale is composed of five categories, with two discriminant features: color and interference patterns. The classification into these five grades is a difficult clinical task, especially with thinner lipid layers which lack color and/or morphological features. The subjective interpretation of the experts via visual inspection may affect the classification, and so a high degree of inter- and also intra-observer variability can be produced (García-Resúa et al., 2013).

The development of a systematic and objective computerized method for analysis and classification is highly desirable, allowing for homogeneous diagnosis and relieving experts from this tedious and time-consuming task. As a consequence, previous researches (Remeseiro et al., 2011; Remeseiro et al., 2013a) have demonstrated that the interference phenomena can be characterized as a color texture pat-
tern. Thus, the classification task can be automatized with a classification accuracy over 95% using the co-occurrence features method for texture extraction and the Lab color space.

On the other hand, the spatial heterogeneity of the tear film lipid layer makes the classification of a patient’s image into a single category, as previous approaches do, not always possible. In fact, performing local analysis of the images in order to detect multiple categories per patient would be more accurate. Furthermore, this kind of analysis would be useful to discern different local states, and so different tear film distribution maps. Accordingly, a first attempt to segment tear film images into the Guillon categories was proposed in (Remeseiro et al., 2013b). This preliminary approach includes the application of previously proposed techniques for color texture analysis and a decision voting system for the final result. Up to the knowledge of the authors, there are no more attempts in the literature to automatically segment lipid interference patterns on tear film images acquired with the Tearscope plus.

The main problem with the approach proposed in (Remeseiro et al., 2013b) is that it uses the so-called background category, in addition to the five categories defined by Guillon. This category represents the areas of an image where there is no interference pattern. Since the samples of this unreal category have a high level of variability, they cannot be defined by uniform texture color features and so the accuracy of a classifier may be affected. Furthermore, it uses the class predicted by the machine learning algorithm to segment the images, regardless of the probability of the prediction. Thus, the proposed methodology will provide segmented images by means of a weighted voting system which takes into account the multiclass probability, and a minimum threshold to distinguish the background from the Guillon patterns without including any unreal category.

The outline of the paper is as follows: Section 2 describes a methodology to create tear film maps based on the lipid interference patterns. Section 3 presents the materials and methods employed in this paper. Section 4 shows the experimental results, and Section 5 includes the conclusions and future lines of research.

2 RESEARCH METHODOLOGY

The research methodology to create tear film maps based on the lipid interference patterns consists of five stages (see Figure 1). The input data is a tear film image acquired with the Tearscope plus, and the output data is a labeled image based on the interference patterns defined by Guillon. First, the region of interest (ROI), in which the analysis will take place, is detected. Next, using a specified window size, each window located at the ROI of the input image is analyzed in terms of color and texture features, and its class-membership probabilities are calculated by soft classification. After that, the segmentation is performed with the weighted voting system and a tear film map is obtained, represented as a labeled image. And finally, the tear film map is post-processed in order to eliminate the small regions.

The whole process described in this section can be seen in Algorithm 1. In what follows, each stage will be explained in depth.

2.1 Location of the Region of Interest

Input images acquired with the Tearscope plus include irrelevant areas for tear film segmentation, such as the sclera or the eyelids. Previous works (Remeseiro et al., 2011; Remeseiro et al., 2013a) located the region of interest (ROI) as a rectangle in the bottom part of the iris. A set or ring-shaped templates that define different ROI shapes and the template matching technique were used. Nonetheless, in this research the analysis is taken over the whole tear film and so a new process to locate the ROI is presented.

The tear film can be perceived with the best contrast in the green channel of the input image in RGB, so only this single channel will be considered in this stage. First, the green channel is thresholded using its histogram. Then, the normalized cross-correlation

![Figure 1: Steps of the research methodology, where the colors depict the Guillon categories. Note that the images of the steps “feature extraction” and “soft classification” correspond to the input image zoom, where the squares represent the windows.](image-url)
(Russ, 1999) is applied to the thresholded image, using circles of different sizes as templates. Thus, the circle with the maximum cross-correlation value allows to locate the pupil of the image. Next, a new circle with the same center than the previous one and a radius \(n\) times larger is created. This new circle is used as a first approach to the ROI (see Figure 2).

On the other hand, the area in which the tear film can be perceived is lighter than the iris and the pupil which surrounded it. In this way, a second approach to the ROI can be determined by finding those pixels whose gray level is greater than a threshold:

\[
th = \mu - p \times \sigma\]  

(1)

where \(\mu\) is the mean value of the gray levels of the image, \(\sigma\) is its standard deviation and \(p\) is a weight factor empirically determined.

Since some images can include irrelevant regions, such as eyelashes or shadows cast by them, the morphological operator of erosion (Gonzalez and Woods, 2008) is applied in order to eliminate them from this second approach to the ROI (see Figure 2). Finally, the logical AND operator between the two approaches is calculated. This region is likely to be free of irrelevant features and so, in most cases, could be the final ROI. Despite that, the length of the eyelashes in some cases and specially the irregular shape of this ROI motivate a final adjustment: the biggest circle concentric to the pupil is “divided” in sixteen quadrants and, in each one, the minimum radius is considered in order to simplify the final ROI. See Figure 2 as an example of the whole process. Note that this step of the methodology corresponds to the line 1 of the Algorithm 1.

### 2.2 Feature Extraction

Once the ROI is located, all the windows with a specified size inside it are analyzed and a descriptor per window is obtained as can be seen in the line 5 of the Algorithm 1. This descriptor is a quantitative vector composed of 23 features proposed in (Bolón-Canedo et al., 2012), and obtained as follows:

1. Color analysis. The Lab color space (McLaren, 1976) is used to extract color information, since its use is appropriate in combination with texture analysis according to (Remeseiro et al., 2011). This color space is a 3D model where \(L\) represents the lightness of the color, and \(a\) and \(b\) are the chromatic components. The color analysis step consists in transforming the input image in RGB to the Lab color space in order to subsequently analyze the texture of its three channels.

#### Algorithm 1: Pseudo-code of the weighted voting system for image segmentation

**Data:** input image \(I\), number of classes \(n\), weights \(\omega_1, \omega_2\), threshold \(th\), minimum perimeter \(m\)

**Result:** output image \(O\) (its labels \(\in [0, n]\) indicate the classes, where 0 is the background)

1. \(ROI := locate\_roi(I)\)
2. initialize matrix of votes \(V := 0\)
3. initialize vector of maximum votes \(V_{\text{max}} := 0\)
4. for each window \(w \in ROI\) do
5. \(feats := compute\_features(w)\)
6. \(CP := classify(feats)\)
7. for each pixel \(p \in w\) do
8. \(d := distance(p, center(w))\)
9. for \(k \leftarrow 1\) to \(n\) do
10. \(v = \omega_1 \cdot CP[k] + \frac{\omega_2 \cdot CP[k]}{d}\)
11. \(V[p][k] += v\)
12. \(V_{\text{max}} = \omega_1 + \frac{\omega_2}{d}\)
13. \(V_{\text{max}}[p] += V_{\text{max}}\)
14. end
15. end
16. end
17. \(O := 0\)
18. for each pixel \(p \in ROI\) do
19. \(v := \max(V[p])\)
20. \(i := \text{index}(\max(V[p]))\)
21. if \((v \geq th \cdot V_{\text{max}}[p])\) then
22. \(O[p] := i\)
23. end
24. end
25. \(C := \text{find\_contours}(O)\)
26. for each contour \(c \in C\) do
27. if \(\text{perimeter}(c) < m\) then
28. \(\text{remove}(c)\)
29. end
30. end
31. end

2. Texture analysis. The co-occurrence features technique (Haralick et al., 1973) is used to extract texture information, since it is the most appropriate method at the problem at hand according to (Remeseiro et al., 2011). This method describes texture in terms of statistical measures, and it is based on the computation of the conditional joint probabilities of all pairwise combinations of gray levels. It generates a set of gray level co-occurrence matrices and extracts several statistical measures from their elements. Specifically, a set of 14 statistics proposed by Haralick et al. (Haralick et al., 1973) are computed from each matrix. Next, the mean and the range of these 14 statistics are calculated across matrices in or-
der to obtain a texture descriptor. Note that these statistics represent features such as contrast or homogeneity.

3. Feature selection. The correlation-based feature selection (CFS) method (Hall, 1999) was used for feature selection in order to reduce the number of features and, thus, the computational (memory and time) requirements. It is a simple filter algorithm which ranks feature subsets according to a correlation based heuristic evaluation function. The bias of the evaluation function is toward subsets that contain features that are highly correlated with the class and uncorrelated with each other. An ad hoc feature selection process based on this filter was used for dimensionality reduction, so the descriptor with color texture features was reduced, from 588 to 23 features, with no degradation in performance according to (Bolón-Canedo et al., 2012).

2.3 Soft Classification

Once the descriptor of each window located in the ROI is obtained, a support vector machine (SVM) (Burges, 1998) is used to compute its class-membership probabilities as can be seen in the line 6 of the Algorithm 1. Notice that partial class memberships are used in soft classification to model uncertain labeling and mixtures of classes. In this way, the voting system subsequently presented can vote for each class according to its probability, instead of voting for the class with the highest probability. The selection of the SVM as the machine learning algorithm is based on the results presented in (Remeseiro et al., 2012).

2.4 Weighted Voting System

The weighted voting system is presented in this paper for tear film segmentation, although it can be adapted to any image segmentation problem where the classes can be represented by a set of features and classified by a soft classifier. In the problem at hand, the method considers the class-membership probabilities of each window of the ROI, and every pixel in this window receives a vote associated to each class $c$:

$$v_c = \omega_1 \cdot p_c + \omega_2 \cdot \frac{d}{d}$$

where $p_c$ is the probability to belong to the class $c$, $d$ is the distance from the pixel to the center of the window, and $\omega_1$ and $\omega_2$ weight the probability and the distance, respectively.

On the other hand, the maximum vote that every pixel in this window can receive, assuming maximum probability, is also calculated:

$$v_{\text{max}} = \omega_1 + \frac{\omega_2}{d}$$

As windows are overlapped and each pixel belongs to several windows, the votes received from each category are added up. Thus, each pixel will have a set of final votes corresponding to each class and the maximum final votes. First, only the final votes of the classes are considered in order to select the most voted class. Then, the pixel is assigned to this class only if its final number of votes is higher than the maximum number of votes weighted by a threshold $th$. Note that this threshold is used to distinguish the real classes from the background. This main step of the research methodology corresponds to the lines from 2 to 19 of the Algorithm 1.

2.5 Elimination of Small Regions

Once the previous step is performed, a labeled image which represents a tear film map is obtained. This labels corresponds to one of the Guillon categories or represents the background of the image. Small regions may appear in the distribution map, which can correspond to false positives or noisy areas. Thus, a post-processing step is performed in order to eliminate them. This steps consists in extracting the contour of each region and calculating its perimeter. In
this way, the regions whose perimeter is less than a minimum perimeter $m$ previously established are eliminated. Note that this final step appears in the lines from 20 to 23 of the Algorithm 1.

3 MATERIALS AND METHODS

The aim of this work is to create tear film maps based on the lipid interference patterns, using the proposed methodology. The materials and methods used in this research are described in this section.

3.1 Image Data-set

Guillon defined five main grades of lipid layer interference patterns, which in increasing thickness are: open meshwork, closed meshwork, wave, amorphous and color fringe. These patterns can be observed with the Tearscope plus (Tearscope plus, 1997), a handheld instrument to view the tear film non-invasively. It was designed by Guillon (Guillon, 1998) to evaluate lipid layer thickness by interference fringe biomicroscopy, and uses a cold light source to minimize any drying of the tear film during the examination.

In this research, the image acquisition was carried out with the Tearscope plus attached to a Topcon SL-D4 slit lamp (Topcon SL-D4, nd) to gain more magnification. Thus, the interference patterns were observed through this slit-lamp microscope with magnification set at 200X. Since the lipid tear film is not static between blinks, a video was recorded and analyzed by an optometrist in order to select the best images for processing. The interference phenomena was recorded with a Topcon DV-3 digital video camera (Topcon DV-3, nd) and stored via the Topcon IMAGEnet i-base (Topcon IMAGEnet, nd) at a spatial resolution of $1024 \times 768$ pixels per frame in the RGB color space. Finally, images were selected to go through the processing step only when the tear film lipid layer was completely expanded after eye blink.

A bank of images acquired from healthy patients with ages ranges from 19 to 33 years was available. All images in this bank have been annotated from three different optometrists from the Faculty of Optics and Optometry, University of Santiago de Compostela. These annotations consist of delimited regions in the images associated with the Guillon categories. It contains 44 images in which the experts have annotated an average of 87 regions each one. These regions includes samples of the five categories considered.

3.2 Experimental Procedure

The experimental procedure is detailed as follows:

1. Train a SVM with radial basis kernel and automatic parameter estimation with representative samples of the five categories considered. Note that, for this task, the samples correspond to areas in which the three optometrists marked the same category.

2. Locate the ROI of each image in the data-set.

3. Apply the feature extraction methods to the bank of images, in order to obtain the quantitative vectors of the windows located in their ROIs.

4. Calculate the class-membership probabilities of the quantitative vectors using the SVM previously trained.

5. Apply the weighted voting system using different configurations of parameters ($\omega_1$, $\omega_2$ and $\omega_3$).

6. Eliminate the regions whose contour has a perimeter less than a threshold $m$.

7. Evaluate the effectiveness of the proposed methodology in terms of the similarity between the system and the three experts considered.

4 EXPERIMENTAL RESULTS

In this section, the results obtained with the image data-set will be compared with the annotations made by three experienced practitioners. First of all, it should be highlighted that different parameter configurations were considered in the weighted voting system as can be seen in Section 3.2. However, the rest of the parameters were empirically established in order to carry out the experimentation presented in this section. In this sense, the threshold $m$ used to eliminate the small regions was set to 110 pixels based on the minimum perimeter size of the regions marked by the optometrists. Note also that the window size for segmentation is $32 \times 32$ pixels, since it is the minimum window size which allows a precise segmentation and maintains the texture well-defined according to (Remeseiro et al., 2013b).

First, the research methodology was applied over the data-set of 44 images and so the segmentation process was performed on their ROIs. In this manner, the output of the process is a labeled image, so-called tear film map, in which each label is represented by a color. Notice that each color is associated to a Guillon category in this way: red means open meshwork, yellow means closed meshwork, green means wave, cyan means amorphous, and blue means color fringe.
Additionally, the background of the output image is represented by black.

Figure 3 illustrates a tear film map obtained with the proposed methodology over an input image from the data-set. As it can be seen, there are regions of the image in which the experts agree with the Guillon category, whereas there are other regions in which the agreement is non-existent. The same situation happens if the output map is compared with the experts' annotations.

The validation process consists in dealing with the system as it was another expert, and so a comparison between the “four hypothetical” experts was carried out. In this manner, the regions marked by a reference expert are compared with the regions annotated by the other three experts. In this comparison, the number of pixels of the reference expert that coincide with other experts are counted and graphically represented. Thus, in a comparison of a reference expert versus the rest of the experts, there will be a number of pixels which agree with 0, 1, 2 or 3 experts. Note that the agreement with 3 experts means a total agreement between the four experts considered, and an agreement with 0 experts represents the pixels marked only by the reference expert. The idea is to perform this comparison between the experts in such a way that each expert is the reference expert once, and so a total of four comparisons are obtained for a single parameter configuration.

The parameter values analyzed in this section are in the range [0, 1], with a 0.1 step, for the three parameters considered. It should be highlighted that at least one of the weights, \(\omega_1\) and \(\omega_2\), has to have a non-zero value. Also, if the threshold \(th = 1\), the output image contains no information and all the pixels belong to the background. The reason is that a so high threshold implies that most of the class-membership probabilities have to have the maximum value. Among all the combinations of the three parameters, three of them have been selected to show the validation of the proposed methodology: (i) \([\omega_1 = 1, \omega_2 = 0]\), which only considers the first part of the Equation 2, (ii) \([\omega_1 = 0, \omega_2 = 1]\), which only considers the second part of the Equation 2, and (iii) \([\omega_1 = 1, \omega_2 = 1]\), which equally considers both parts of the Equation 2. In these three cases, the threshold value is \(th = 0.9\). Figures 4, 5 and 6 illustrates through bar plots these three configurations, respectively. If the three configurations are compared, no significant differences can be appreciated which means that considering or not one of the two terms of the Equation 2 is not too relevant. The main reason is that in both terms, the class-membership probabilities of the SVM are taken into account and so their values are important enough to be the key of the weighted voting system.

On the other hand, if the bar graphs of each expert are analyzed, interesting differences can be observed. For example, if we focus on the plots “system vs all”, it can be seen that there are a little number of pixels which match with 0, 1 and 2 experts. This fact means that the system has a conservative behavior. In contrast, the graphs “expert3 vs all” show how the expert3 has a completely different behavior since it tends to mark bigger regions. And finally, in the middle of both extremes, the other two graphs “expert1 vs all” and “expert2 vs all”, demonstrate that both expert1 and expert2 have intermediate positions with a really similar plot tendency.

After this analysis focus on the weights \(\omega_1\) and \(\omega_2\), the aim now is to explore the impact of the threshold \(th\). Consequently, only one of the above three parameter configurations is considered but with a lower threshold \(th = 0.8\). As can be seen in Figure 7, the number of pixels in which the agreement is total has increased. Thus, the bigger regions obtained by lowering the threshold still match with regions marked by the three optometrists. However, the number of disagreements has also increased as expected, and now the tendency is more similar to the tendencies of the expert1 and expert2. Finally, two noticeable aspects of the graphs should be highlighted: (i) the system produces some tear film maps in which there are regions associated to the wave pattern that do not match with any expert, (ii) the expert3 marked some regions of the color fringe pattern that do not match with any expert. Not only does this anomalous behavior appear in the images obtained by the system, but also in some of the ones marked by the experts. Thus, it could be said that the system behaves like the experts in a greater or lesser degree depending on the parameter configurations.

5 CONCLUSIONS AND FUTURE RESEARCH

An automatic methodology for tear film lipid layer segmentation is proposed in this paper. From an input image acquired with the Tearscope plus, a distribution map of the tear film is obtained as a result. This methodology includes the application of color texture analysis techniques, soft classification and a weighted voting system focus on two variables (probabilities and distances). It improves a previous approach by eliminating the “unreal” background class, characterized by a high degree of variability, and considering class-membership probabilities. Results obtained with this methodology demonstrate that the
Figure 3: A representative example of tear film map: from (a) to (c) images annotated by the expert1, expert2, and expert3, respectively; (d) tear film map obtained with the proposed methodology. Note that the relation between colors and categories is: red - open meshwork, yellow - closed meshwork, green - wave, cyan - amorphous, and blue - color fringe.

Figure 4: Comparisons between experts when $|\omega_1 = 1, \omega_2 = 0, th = 0.9|$: (a) system vs all, (b) expert1 vs all, (c) expert2 vs all, (d) expert3 vs all.

Figure 5: Comparisons between experts when $|\omega_1 = 0, \omega_2 = 1, th = 0.9|$: (a) system vs all, (b) expert1 vs all, (c) expert2 vs all, (d) expert3 vs all.

Figure 6: Comparisons between experts when $|\omega_1 = 1, \omega_2 = 1, th = 0.9|$: (a) system vs all, (b) expert1 vs all, (c) expert2 vs all, (d) expert3 vs all.

Figure 7: Comparisons between experts when $|\omega_1 = 1, \omega_2 = 1, th = 0.8|$: (a) system vs all, (b) expert1 vs all, (c) expert2 vs all, (d) expert3 vs all.
tear film maps provided by the developed system are similar to the annotations done by three experienced optometrists. In clinical terms, the manual process done by experts can be automated with the benefits of being unaffected by subjective factors and providing a detailed distribution of the interference patterns over the whole tear film lipid layer.

In this research, the proposed methodology processes the whole set of windows inside the region of interest. Although the feature extraction time over one single window is almost negligible (under 1 second), processing the whole set takes too long. Therefore, as future research, we plan to develop an optimization of the proposed methodology focus on the reduction of the processing time. In addition, our future research will also involve proposing new algorithms for tear film segmentations based, for example, on other classical algorithms such as seeded region growing.

ACKNOWLEDGEMENTS

This research has been partially funded by the Secretaría de Estado de Investigación de the Spanish Government and FEDER funds of the European Union through the research projects PI10/00578 and TIN2011-25476. Beatriz Remeseiro acknowledges the support of Xunta de Galicia under Plan I2C Grant Program.

REFERENCES


Topcon IMAGEnet (n.d.). Topcon IMAGEnet i-base. Topcon Medical Systems, Oakland, NJ, USA.