Abstract—This paper presents a comparative study of different texture extraction methods for the automatic classification of the tear film lipid layer based on the categories enumerated by Guillon [1]. From a photography of the eye, a region of interest is detected and its low-level features are extracted, generating a feature vector that describes it, to be finally classified in one of the target categories. This paper discusses several texture analysis methods and colour spaces to generate the feature vectors. The proposed methods have been tested on a dataset composed of 105 images, with a classification rate of over 95% in some cases.

Index Terms—Eye lipid layer, Guillon categories, Butterworth filters, the discrete wavelet transform, co-occurrence features, Markov random fields, Gabor filters, machine learning.

I. INTRODUCTION

The tear film consists of a superficial layer, an intermediate aqueous phase and an underlying mucous layer [2]. The tear film provides a smooth optical surface by compensating for the micro irregularities of the corneal epithelium and plays an essential role in the maintenance of ocular integrity by removing foreign bodies from the front surface of the eye.

The lipid layer of the tear film plays a major role in limiting evaporation during the inter-blink period and also affects the tear film stability. The lipid layer thickness can be evaluated by the observation of the interference phenomena [1], which correlates with tear film quality [3], since a thinner lipid layer speeds up water evaporation, decreasing the tear film stability. Thus, a deficit of this layer can cause the evaporative dry eye syndrome, a disease which affects a wide sector of the population, specially among contact lens users, and worsens with age.

In order to classify the tear film lipid layer as a function of its thickness, we consider the five main grades of lipid layer thickness interference patterns proposed by Guillon [1]: open meshwork, closed meshwork, wave, amorphous and colour fringe. The amorphous category has not been included in this study due to the lack of images from this category in the clinical image dataset used for validation. The four categories considered in this study are illustrated in the Fig.1. In general, thicker lipid layers (≥ 90nm) are readily observed since they contain colour and wave patterns. However, thinner lipid layers (≤ 60nm) are difficult to visualise, since the colour fringes and other distinct morphological features are not present and classification is affected by the subjective interpretation of the observer. This has led to the development of techniques to objectively calculate lipid layer thickness using sophisticated optic systems [4], or techniques that use an interference camera to assess lipid layer thickness by the analysis of the interference colours [5].

As previously mentioned, the lipid layer thickness can be evaluated by the observation of the interference phenomena. The purpose of this study is to demonstrate that the interference phenomena can be characterised as a texture pattern. Thus, we present a methodology to classify the eye lipid layer into four of the five categories defined by Guillon, based on the analysis of their texture and colour. We will also compare different texture feature extraction methods in different colour spaces. This automatic classification is very important to the experts, who do this time-consuming task by hand. Thereby, we could not only eliminate the subjectivity of the process but also save time for experts.

This paper is organised as follows: section II describes the methodology for the automatic classification of the lipid layer; section III compares the classification results obtained by different texture and colour analysis methods; finally section IV exposes our conclusions and future work.

II. METHODOLOGY

The methodology we describe in this section has been tested on interference images captured using a Tearscope-plus [6] attached to a Topcon SL-D4 slit lamp [7] and a Topcon DV-3 digital video camera [8]. The slit lamp’s magnification was set at 200X and the images were stored via the Topcon IMAGEnet i-base [9] at a spatial resolution of 1024 × 768 pixels. Since the tear lipid film is not static between blinks, a video has been recorded and analysed by an optometrist in order to select the
In 1D, a Butterworth bandpass filter is defined as:

\[ f(\omega) = \frac{1}{1 + (\frac{\omega - \omega_0}{\omega_c})^{2n}} \quad (1) \]

where \( n \) is the order of the filter, \( \omega \) the angular frequency, \( \omega_0 \) the cutoff frequency and \( \omega_c \) the centre frequency. The order \( n \) of the filter defines the slope of the decay; the higher the order, the faster the decay.

In this paper, we have used a bank of Butterworth bandpass filters composed of 9 second order filters, with bandpass frequencies covering the whole frequency spectrum [15]. The filter bank maps each input image into 9 result images, one per frequency band.

In order to classify the input images, we must assign each of them a feature vector; in this case, we have used histograms to describe the output images in each frequency level. First, we normalised each frequency band separately and computed the histograms of its output images. Those histograms concentrated most of the information in the lower bins, which made their comparison difficult. In order to increase the relevance of the differences among lower values, we computed uniform histograms with non-equidistant bins; given all the output images from our dataset in a frequency band, we ordered their pixels making sure each bin contained no more than \( \frac{N}{N_{bins}} \) pixels, being \( N \) the number of pixels in the given frequency and \( N_{bins} \) the number of histogram bins. Therefore, the descriptor of an input image in a frequency band is the uniform histogram of the filtering result.

2) The Discrete Wavelet Transform: Mallat [16] was the first to show that wavelets formed a powerful basis for multisresolution theory, defining a mathematical framework which provides a formal, solid and unified approach to multiresolution representations. This wavelet paradigm has found many applications in signal and image processing, such as texture analysis.

The Discrete Wavelet Transform (DWT) generates a set of wavelets by scaling and translating a mother wavelet, which is a function defined both in the spatial and frequency domain, that can be represented in 1D as [17]:

\[ \phi^{a,b}(x) = \frac{1}{\sqrt{a}} \phi\left(\frac{x - b}{a}\right) \quad (2) \]

where \( a \) governs the scale and \( b \) the translation of the function. In 2D, the mother wavelet can be represented as:

\[ \phi^{a,b}(x,y) = \frac{1}{\sqrt{a_x a_y}} \phi\left(\frac{x - b_x}{a_x}, \frac{y - b_y}{a_y}\right) \quad (3) \]

where \( a = (a_x, a_y) \) and \( b = (b_x, b_y) \).

Depending of the values of \( a \) and \( b \), we can define a high pass filter or a low pass filter.

The wavelet decomposition of an image consists of applying these wavelets, generating low pass (L) and high pass (H) result images. This process is performed along the horizontal and vertical directions, obtaining 4 images (LL, LH, HL, HH) which are then subsampled by a factor of 2. In order to construct a multilevel decomposition, the process is repeated iteratively on the LL subimage resulting in the standard pyramidal wavelet decomposition shown in Fig.3.

When using wavelets a fundamental step is the selection of the mother wavelet. There are several alternatives like Haar, Daubechies or Symlet wavelets. In the present work, we have
used the Haar wavelets because they outperform the other wavelet families tested. Thus, we apply a generalised Haar algorithm [16] to decompose the input image into 4 subimages at each scale: LL, LH, HL, HH; using 2 scales we obtain a total of 8 subimages. In this case, no more scales are required because they do not improve the results obtained.

From each result subimage, we compute the mean, energy and absolute average deviation:

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} p(i) \tag{4}
\]

\[
e = \frac{1}{N^2} \sum_{i=1}^{N} [p(i)]^2 \tag{5}
\]

\[
aad = \frac{1}{N} \sum_{i=1}^{N} |p(i) - \mu| \tag{6}
\]

Then, the descriptor of an input image is constructed calculating the \(\mu\) and the \(aad\) of the input and LL images, and the \(e\) of the LH, HL and HH images. Since we use 2 scales, a total of 12 features are obtained.

3) Co-occurrence Features: Co-occurrence features, presented in [18] by Haralick et al., are a popular and effective texture descriptor based on the computation of the conditional joint probabilities of all pairwise combinations of gray levels, given an interpixel distance \(d\) and an orientation \(\theta\). This method generates a set of Grey Level Co-occurrence Matrices (GLCM) and extracts several statistics from their elements \(P_{\theta,d}(i,j)\).

For a distance \(d = 1\) we have a total of 4 orientations (0\(^\circ\), 45\(^\circ\), 90\(^\circ\) and 135\(^\circ\)) and 4 matrices are generated (see Fig.4(a)). For a distance \(d > 1\), the number of orientations increases and, therefore, also the number of matrices. In general, the number of orientations for a distance \(d\) is \(4d\). As an example, Fig.4(b) depicts the orientations considered for the distance \(d = 2\).

From each co-occurrence matrix we compute a set of 14 statistics proposed by Haralick et al. in [18], representing features such as homogeneity or contrast. Then, we compute their mean and range across matrices obtaining a set of 28 features which will be the descriptor of the input image.

4) Markov Random Fields: Markov Random Fields (MRF) are one of the most popular of the model based methods for texture analysis. A MRF [19] is a 2D lattice of points where each point is assigned a value that depends on its neighbouring values. Thus, MRFs generate a texture model by expressing the grey values of each pixel in an image as a function of the grey values in a neighbourhood of the pixel.

Let \(X(i,j)\) be a random variable which denotes the grey value of the pixel \((i,j)\) on an \(N \times M\) image \(I\). For simplicity, \(X\) shall be indexed with one variable: \(X(c)\) where \(c = 1, 2, 3, \ldots N \times M\). Therefore, if \(y\) is a neighbour of \(x\), \(p(X(x))\) depends on \(X(y)\). The MRF is a joint probability density model of the elements of \(I\) such that:

\[
p(X(c)|X(m), m = 1, 2, 3, \ldots N \times M, c \neq m) = p(X(c)|\text{neighbours of } c) \tag{7}
\]

According to this expression, we need to define the concept of neighbourhood as a previous step for creating the MRF model. In this case, we consider the neighbourhood of a pixel as the set of pixels within a distance \(d\). Now, we model the Markov process for textures using a Gaussian Markov Random Field (GMRF) defined as [20]:

\[
X(c) = \sum \beta_{c,m} [X(c + m) + X(c - m)] + e_c \tag{8}
\]

where \(e_c\) is the zero mean Gaussian distributed noise, \(m\) is an offset from the centre cell \(c\) and \(\beta_{c,m}\) are the parameters that weigh a pair of symmetric neighbours to the centre cell. The \(\beta\) coefficients describe the Markovian properties of the texture and the spatial interactions. We can represent equation (8) in matrix notation as:

\[
X(c) = \beta^T Q_c + e_c \tag{9}
\]

Thus, the \(\beta\) coefficients can be estimated using a least squares fitting:

\[
\beta = \left( \sum_{c \in I} Q_c Q_c^T \right)^{-1} \left( \sum_{c \in I} Q_c X(c) \right) \tag{10}
\]

A general GMRF-based approach for texture analysis employs the GMRF model parameters \(\beta\) and the intensity variance \(\sigma\) as the texture descriptor, where the variance can be calculated as:
\[ \sigma = \frac{1}{N \times M} \sum_{c \in I} [X(c) - \beta^T Q_c]^2 \] (11)

Instead of using this feature set, Çesmeli and Wang [21] proposed using the directional variances:

\[ f_i = \frac{1}{N \times M} \sum_{c \in I} [X(c) - \beta_i Q_c]^2 \] (12)

In this paper, we have used the directional variances to generate the input image descriptor because they outperform the model parameters \( \beta \) and the intensity variance \( \sigma \).

5) **Gabor Filters**

Gabor filters are complex exponential signals modulated by gaussians [22] widely used in texture analysis. A two dimensional Gabor filter [23], using cartesian coordinates in the spatial domain and polar coordinates in the frequency domain, is defined as:

\[
g_{x_0,y_0,f_0,\theta_0} = \exp \left\{ i \left[ 2\pi f_0 (x \cos \theta_0 + y \sin \theta_0) + \phi \right] \right\} \text{gauss}(x,y)
\] (13)

where

\[
\text{gauss}(x,y) = a \exp \left\{ -\pi \left[ a^2 (x \cos \theta_0 + y \sin \theta_0)^2 + b^2 (x \sin \theta_0 - y \cos \theta_0)^2 \right] \right\}
\] (14)

\( a \) and \( b \) model the shape of the filter, while \( x_0, y_0, f_0 \) and \( \theta_0 \) represent the location in the spatial and frequency domains, respectively.

In this work, we have created a bank of 16 Gabor filters centred at 4 frequencies and 4 orientations. The filter bank maps each input image to 16 result images, one per frequency-orientation pair.

As with Butterworth Filters, the descriptor of each output image is its uniform histogram with non-equidistant bins.

### B. Colour Analysis

As previously mentioned, we have analysed the texture extraction methods in grayscale, Lab and RGB.

In Lab, we analyse each component separately, generating three descriptors per image corresponding to the L, a and b components. In this case, we concatenate these three descriptors to generate the descriptor that will be classified.

Finally, for the analysis in RGB we have used the opponent colour theory proposed by Hering [12] in the 1800’s. This theory states that the human visual system interprets information about colour by processing three opponent channels: red vs. green, green vs. red and blue vs. yellow. More precisely:

- \( R_G = R - p \times G \)
- \( G_R = G - p \times R \)
- \( B_Y = B - p \times (R + G) \) (15)

where \( p \) is a low pass filter. We have analysed these channels separately and generated their own descriptor. Again, the descriptor using opponent colours is the concatenation of these \( R_G, G_R \) and \( B_Y \) descriptors.

### C. Classification

The final step of our methodology is the classification of the region of interest into one of the four categories proposed by Guillon. In a previous paper [15], we analysed several machine learning algorithms getting to the conclusion that the Support Vector Machine (SVM) [24] produces the best results and, for that reason, it has been used in the present study.

Next section, we will show all the results in terms of percentage accuracy. Based on the SVM, this measure represents the rate per cent of the images that are correctly classified according to their category.

### III. EXPERIMENTAL RESULTS

We have tested the feature extraction methods previously presented on a dataset composed of 105 images acquired from healthy subjects aged 19 to 33 years. All of them gave their informed consent prior to their inclusion in the study. The dataset includes 29 open meshwork, 29 closed meshwork, 25 wave and 22 colour fringe images. In order to analyse the generalisation of our results to larger dataset, a 10-fold cross-validation [25] has been performed.

#### A. Butterworth Filters

The first experiment in this case was performed on grayscale, Lab and opponent colours and analyses each frequency band separately as well as their combination. In order to combine the adjacent frequency bands, we concatenate their individual descriptors. This experiment will help us to decide which colour space and frequency bands are more appropriate for the task at hand. Table I shows its results in terms of percentage accuracy for all the frequency band concatenations. From top to bottom, each cell shows the results obtained in grayscale, Lab and opponent colours. We have highlighted the best result in each colour space.

Analysing these results we can see that the intermediate frequencies are more discriminative than the lowest and highest ones; achieving results over a 70% of correct classifications in grayscale. The best combinations provide classification rates higher than 80%. Regarding opponent colours, we can see how colour information improves the accuracy of the method compared to grayscale. In this case, the accuracy is almost 90% for the best combinations of frequency bands. Finally, the results show that Lab outperforms opponent colours and produces the best results, that reach 93.33% classification rate. Table I also shows how the results are quite stable; there is a wide range of frequency band combinations where the results are over a 90% accuracy.

#### B. The Discrete Wavelet Transform

Our first experiment aims at determining the relevance of each feature individually by analysing their results separately. Table II shows the results of this experiment, where we can see that the energy performs worse than the rest of features. We have highlighted the best features in the Lab case, all of them achieving over a 70% accuracy.
Our second experiment compares two alternative descriptors: the first one composed of the 12 features in Table II and the second one composed of the six best performing features in Table II. Table III depicts both results in the three colour spaces, all of them with an accuracy of almost 90%. The best distance combinations provide classification rates over a 92%. Opponent colours do not outperform grayscale, being the results quite similar. The best results are obtained using Lab. Almost all the distance combinations obtain over a 90% and some of them around a 95% accuracy.

D. Markov Random Fields

The experiment performed in this case compares different neighbourhoods in grayscale, Lab and opponent colours. Table V shows the results of this experiment; where we can see that there is a range of distances between 3 and 6 that achieve over an 80% accuracy. The lowest and highest distances perform worse, while the use of colour information does not always outperform grayscale. We have highlighted the best results for each colour space, all of them with an accuracy of almost 85%.

E. Gabor Filters

The experiment in this case consists of using a different number of bins to create the uniform histogram which defines the descriptor both in grayscale, Lab and opponent colours. Table VI shows the results for four different number of bins; where we can see that all of them achieve around a 90% of correct classifications. The results are quite stable regardless of the number of bins. We have highlighted the best results for each colour space which we can emphasise the 0.924% of accuracy using the Lab colour space.

### TABLE I

<table>
<thead>
<tr>
<th>Features</th>
<th>Grayscale</th>
<th>Lab</th>
<th>Opponent Colours</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ and σ of image</td>
<td>0.70</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>µ and σ of HH2</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of HH1</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of LL</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
</tbody>
</table>

### TABLE II

The Discrete Wavelet Transform: SVM categorisation accuracy (%) using different features.

### TABLE III

The Discrete Wavelet Transform: SVM categorisation accuracy (%) using combinations of features.

### TABLE IV

Co-occurrence Features: SVM categorisation accuracy (%) in grayscale, Lab and opponent colours; cell $ij$ depicts the results obtained combining the distances ranging from $i$ to $j$.

<table>
<thead>
<tr>
<th>Features</th>
<th>Grayscale</th>
<th>Lab</th>
<th>Opponent Colours</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ and σ of image</td>
<td>0.70</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>µ and σ of HH2</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of HH1</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of LL</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features</th>
<th>Grayscale</th>
<th>Lab</th>
<th>Opponent Colours</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ and σ of image</td>
<td>0.70</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>µ and σ of HH2</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of HH1</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>µ and σ of LL</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
</tr>
</tbody>
</table>
TABLE V

MARKOV RANDOM FIELDS: SVM CATEGORISATION ACCURACY (%) IN GRAYSCALE, LAB AND OPPONENT COLOURS; USING THE DIRECTIONAL VARIANCES.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Grayscale</th>
<th>Lab</th>
<th>Opponent Colours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61.90</td>
<td>86.67</td>
<td>84.76</td>
</tr>
<tr>
<td>2</td>
<td>78.10</td>
<td>87.15</td>
<td>84.76</td>
</tr>
<tr>
<td>3</td>
<td>83.81</td>
<td>90.96</td>
<td>84.76</td>
</tr>
<tr>
<td>4</td>
<td>93.81</td>
<td>92.95</td>
<td>84.76</td>
</tr>
<tr>
<td>5</td>
<td>90.96</td>
<td>93.81</td>
<td>84.76</td>
</tr>
<tr>
<td>6</td>
<td>85.70</td>
<td>95.24</td>
<td>83.81</td>
</tr>
<tr>
<td>7</td>
<td>77.14</td>
<td>94.28</td>
<td>83.81</td>
</tr>
</tbody>
</table>

TABLE VI

GABOR FILTERS: SVM CATEGORISATION ACCURACY (%) IN GRAYSCALE, LAB AND OPPONENT COLOURS; USING DIFFERENT NUMBER OF BINS.

<table>
<thead>
<tr>
<th>Number of bins</th>
<th>Grayscale</th>
<th>Lab</th>
<th>Opponent Colours</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60.57</td>
<td>92.38</td>
<td>86.93</td>
</tr>
<tr>
<td>10</td>
<td>97.62</td>
<td>98.29</td>
<td>85.57</td>
</tr>
<tr>
<td>20</td>
<td>96.67</td>
<td>98.29</td>
<td>85.57</td>
</tr>
</tbody>
</table>

IV. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a study of different techniques to classify the tear lipid film, based on the detection of a region of interest and the analysis of its low-level features through different texture analysis methods and different colour spaces. These analysis show how the problem is feasible with results over 80% accuracy in all the methods tested.

In general, the use of colour information improves the results compared to grayscale because some lipid layers contain, not only morphological features, but also colour features. All the texture analysis methods perform quite well providing results over the 90% in some cases, but Co-occurrence Features generate the best results. Although Markov Random Fields use information of the pixel’s neighbourhood, as the Co-occurrence Features do, this method does not work so well because the statistics proposed by Haralick et al. provide much more information. In short, the combination of Co-occurrence Features and the Lab colour space produces the best classification results with maximum accuracy over 95%.

As a part of our future work, we would like to test the classification rates obtained using the combination of these methods. Also, in many cases, the heterogeneity of the eye lipid layer makes its classification into a single Guillon’s category impossible. Therefore, our future work also involves performing local analysis and classifications in order to detect several categories in each patient.

ACKNOWLEDGEMENTS

We would like to thank the Escuela de Óptica y Optometría of the Universidad de Santiago de Compostela for providing us with the annotated image dataset. This paper has been partially funded by the Ministerio de Ciencia e Innovación of the Gobierno de España and FEDER funds of the European Union through the research project P110/00578; and by the Consellería de Industria of the Xunta de Galicia through the research project 10/CSA918054PR.

REFERENCES