Abstract—The Phenolic Maturity of the grape is one of the most important parameters to determine the optimal time for harvest. In this paper an innovative methodology to estimate grape maturity is proposed. In particular, the method is based on pattern recognition techniques to analyze seed images and to classify them into immature, mature, and overmature states by means of a supervised learning neural network. The presented methodology provides objective information about grape maturity, which is useful for deciding the moment when the harvest should be performed.

Keywords—Grape Maturity Estimation, Neural Networks, Appearance Descriptors.

I. INTRODUCCIÓN

Chile is one of the principal exporters of wine in the world. The wine market is very dynamic and demanding; therefore it is important to conduct research applied to this field in order to improve the level of competitiveness of local producers. To obtain a wine of good quality, it is necessary to harvest the grape at its optimal point of maturity. Traditionally, the estimation of the maturity is performed by chemical analysis in laboratories at high cost. Alternatively, it is possible to employ the judgment of an expert (enologist), but this alternative has a high degree of subjectivity and a low level of representativity.

Traditionally, the estimation of the Phenolic Maturity of the grape is performed by an analysis of the skin and the pulp of the fruit. A recent line of research analyses the grape seeds to estimate its level of maturity. In [1] and [2] a high correlation between the appearance of the seed and its level of maturity is demonstrated. In [3] one of the first studies that analyze the seed using digital image processing techniques and pattern recognition is presented. The aforementioned study examines the relevance of certain descriptors such as Lightness ($L_*$), chroma ($C_{*ab}$), seed length, roundness and aspect ratio, applied to the problem of seed maturity.

The present paper addresses the problem by means of the adequate application of pattern recognition methodology. Instead of studying a limited set of descriptors, an elevated number of appearance descriptors are computed (form, color, texture). The relevance of descriptors is studied with the Sequential Forward Selection algorithm (SFS), which allows the identification of a reduced and adequate set of descriptors. Furthermore, a method for classifying seeds according to their degree of maturity is proposed, which provides very useful information for the timing of the harvest. For classification, a Multi-Layer Perceptron (MLP) with a training algorithm that avoids overfitting is adopted.

The paper is structured as follows: Section II explains the applied methodology. Section III shows the obtained results, and Section IV presents the conclusions.

II. GRAPE MATURITY ESTIMATION METHOD BASED ON SEED IMAGES

The applied methodology comprises the following stages: segmentation, calculation and selection of descriptors, and training and classification.

A. Segmentation

The images under study in this paper present two defects reported in the literature on Image Processing: shadows and highlights. Therefore, a segmentation method robust to these defects is proposed. For seed segmentation from RGB images, the $c_1c_2c_3$ invariant color model reported in [4] is adopted, in order to avoid the aforementioned problems. The expressions of the model are as follows:

\begin{align}
    c_{1i,j} &= \arctan \frac{R_{i,j}}{\max(G_{i,j}, B_{i,j})} \\
    c_{2i,j} &= \arctan \frac{G_{i,j}}{\max(R_{i,j}, B_{i,j})} \\
    c_{3i,j} &= \arctan \frac{B_{i,j}}{\max(G_{i,j}, R_{i,j})}
\end{align}

Channel $c_3$ is chosen due to the good results obtained in [5]. The automatic segmentation of the channel $c_3$ is performed using the well-known Otsu method.

B. Descriptors and Selection using SFS

After segmenting the seed, an elevated number of appearance descriptors are computed, and the descriptors relevant for classification are selected by the SFS algorithm, following the method proposed in [6]. In particular, 1158 appearance descriptors were computed, including those proposed in [6] and in [3]. In this paper, seeds were classified by an expert according to their degree of maturity. Figure 1 shows the classes of seeds, Figure 1(a) corresponds to the immature class, Figure 1(b) to the mature class, and Figure 1(c) to the over-mature class. 20 samples per class were collected...
and the relevant characteristics were selected by using the SFS algorithm in combination with the Fisher discriminant [7]. The selection process generated 58 descriptors for the immature and mature classes, and 59 descriptors for the over-mature class. Considering that certain descriptors are repeated in the description of each class, a total of 154 descriptors was reached.

Figure 1: Classes of seeds according to their maturity degree.

C. Training and Classification

The architecture of the classifier designed for this problem comprises 3 neural networks: one for each class. The training set consists of 20 seeds for each class and the test set has 10 seeds for each class. To determine the number of neurons on each hidden layer, the Bayesian Regularization algorithm [8] was used. This algorithm generated 2 neurons on the hidden layer for the immature and mature classes, and 3 for the over-mature class.

III. RESULTS

Figure 2 shows the process of seed segmentation. Figure 2(a) presents images of the seed in the RGB model, which shows the presence of shadows and highlights. Figure 2(b) shows the channel c3, Figure 2(c) shows the result of the segmentation of this channel, and finally Figure 2(d) shows the result of the contour detection overlaid on the original color image. The last Figure shows that the segmentation process select only pixels belonging to the seed, excluding pixels of the shadows. Additionally, the segmentation method eliminates pixels that are defective due to highlights.

Table I shows the hit rates for each class, using the proposed neuronal network classifier. As expected, the hit rates for the training set are very high. In the test set, the hit rates are acceptable for the problem being studied.

<table>
<thead>
<tr>
<th></th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immature</td>
<td>100%</td>
<td>97%</td>
</tr>
<tr>
<td>Mature</td>
<td>100%</td>
<td>93%</td>
</tr>
<tr>
<td>Over-mature</td>
<td>100%</td>
<td>97%</td>
</tr>
<tr>
<td>Mean</td>
<td>100%</td>
<td>96%</td>
</tr>
</tbody>
</table>

Table I: Hit Rates of Classification using Neural Networks.

IV. CONCLUSIONS

In this paper, an original method for grape maturity estimation based on pattern recognition techniques is proposed. The designed segmentation stage is robust to the problems present in the images (shadows and highlights). Relevant descriptors are determined for the classification based on characteristics selection techniques. The proposed classifier provides high hit rates. In general, the presented methodology provides relevant information to objectively estimate the optimal time for harvest.

ACKNOWLEDGMENT

This research is financed by the Project FONDEF IDeA CA12110236, “Estimación de la Madurez Fenólica de la Uva basada en Imágenes de la Semilla”. National Commission of Science and Technology (CONICYT), Government of Chile.

REFERENCES