UNSUPERVISED LEARNING FOR RIPENESS ESTIMATION FROM GRAPE SEEDS IMAGES

S. Hernández[1], L. Morales[2] and A. Urrutia[1]

Facultad de Ciencias de la Ingeniería. Universidad Católica del Maule[1]

Facultad de Ciencias Agronómicas, Universidad de Chile[2]

A. San Miguel #3605, Talca. Chile

Email: shernandez@ucm.cl

***Submitted:** May 29, 2017* ***Accepted:** July 25, 2017* ***Published:** Sep 1, 2017*

Abstract- Estimating the current stage of grape ripeness is a crucial step in wine making and becomes especially important during harvesting. Visual inspection of grape seeds is one method to achieve this goal without performing chemical analysis, however this method is prone to failure. In this paper, we propose an unsupervised visual inspection system for grape ripeness estimation using the Dirichlet Mixture Model (DMM). Experimental analysis using real world data demonstrates that our approach can be used to estimate different ripeness stages from unlabeled grape seeds catalogs.

Index terms*:* Mixture model, grape ripening, computer vision.

I. INTRODUCTION

Estimating grape ripeness is an important issue in wine producing. In viticulture, it is well known that near the onset of ripening, the seed coat has a green appearance and browning occurs after berry ripening [1]. In the other hand, determination of the phenolic compounds of is a highly complex process that involves chemical analysis or mass spectrometry techniques. However, a more practical approach to determine grape ripeness is based on visual inspection of the grape [2]. Conversely, color and appearance of seeds can be used as an indicator of grape maturity. Among this line of work, [3] used a simple color scale and demonstrated its use to monitor the evolution of the phenolic maturity of grapes. Visual inspection of grape seeds is easily achieved and does not require any equipment. Nevertheless, this a subjective procedure and depends on the expert perception of color. Therefore, there is a need for automatic methods to perform the testing and provide decision support tools for the wine making process.

Several computer vision techniques can be used for estimating grape ripeness. In [4], the authors developed an image analysis technique for grapes seeds characterization. Discriminant analysis was used to determine whether a sample belongs to any class. A similar approach was taken in [5], where grape seeds are classified into mature or in-mature classes by means of a neural network classifier. Later in [6], the center of mass of the color histogram was proposed as a feature for the fruit maturity classification task. However, it is important to notice that being issued as a classification problem, the system requires labeled images which are difficult to acquire in practice. Supervised learning techniques achieve good results for the fruit maturity estimation problem, however there is still a need for an expert to label each seed into one maturity or ripeness class. This can be cumbersome and the labeling can be also subject to human errors. For this reason, it is interesting to develop an unsupervised method to automatically infer different clusters among the data, while at the same time to associate those clusters to possible ripeness stages.

Other techniques for fruit maturity estimation using color image clustering have been developed in the literature. The main idea is to describe a group of samples using some feature set and then use a similarity measure for finding alike images [7]. A color grading for date maturity estimation

was proposed in [8]. Color grading was achieved by converting the 3D RGB color space into a 1D color index. This approach was later extended to a 2D color index in [8], where a back-projection technique is used to assign a color index to any given sample.

More recently, it has been noted that color histograms are better represented by discrete distributions [9]. Moreover, finite mixtures of distributions are widely used in text and language modeling. In the text domain, the frequency of a word occurrence is proportional to the number of occurrences for that word in other documents sharing the same topic. The Expectation Maximization (E-M) algorithm is a popular estimation method for both, continuous or discrete mixture models [10]. Conversely, if the data points correspond to counts or proportions, the Maximum Likelihood (ML) parameters of the mixture model can be estimated from unlabeled data using the E-M algorithm.

Color histograms are another example of count data that can be modeled as a discrete distribution such as the multinomial distribution. In the case of color image clustering, each (2D or 3D) color histogram sample can be considered as a vector of counts and the resulting statistical model can be regarded as bag-of-words model. However, the ML estimator of a high dimensional multinomial distribution leads to biased estimates [11,12]. This problem has been associated to the independence assumption of the event space (color histogram bins), which means that each color bin is considered as an i.i.d. sample and therefore the estimator is tied up to the binning design. More complications arise when estimating small counts or zero-probabilities from such a high dimensional space such as a color space.

Conversely, we consider a Bayesian approach and the Dirichlet distribution is used as a prior distribution for the multinomial parameters. The Dirichlet is a conjugate distribution for the multinomial distribution, therefore the posterior takes a similar form as the prior. We propose a Dirichlet mixture (DMM) as a generative model for clustering grape seeds. The DMM model allows us to directly model the color histogram of the seeds and to take advantage of the prior specification to estimate ripening class memberships. The rest of the paper is organized as follows. Section 2 reviews the DMM model and develops the estimation framework. Section 3

describes the data used in our examples and Section 4 shows the experimental design. In Section 5 we develop a comparison framework and finally Section 6 presents our conclusions.

II. DIRICHLET MIXTURE MODEL

Traditionally, when dealing with mixtures of continuous distributions, the Gaussian is assumed as a base distribution. In cases when the data is not continuous but discrete, the Gaussian assumption is no longer valid. In the case of color histograms, the multinomial distribution can be used to specify the probability of a vector with discrete elements $\mathbf{x} = (x_1,...,x_W)$, containing counts on each one of the W bins of the color histogram. Therefore, we can write down the probability of an image given a certain parameter vector $\boldsymbol{\theta} = (\theta_1,...,\theta_W)$ as:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{n!}{\prod_w x_w!} \prod_w \theta_w^{x_w} \tag{1}$$

where n corresponds to the total number of pixels in the image.

If we now consider a set of D images $X = \{\mathbf{x}_1,...,\mathbf{x}_D\}$, we can calculate the ML estimator of the sample set as:

$$\hat{\theta}_w = \frac{\sum_d x_{dw}}{\sum_j \sum_d x_{dj}} \tag{2}$$

As mentioned before, this method can produce unreliable results since $\theta w \approx 0$ for any unobserved or rarely observed color [13,14]. The extent of this problem is increased with a fine binning of the color space since many bins would become empty and count vectors would become high dimensional and sparse. In the other hand, coarse binning of the color histogram would not provide enough information about the true color distribution.

The Dirichlet distribution is the multivariate version of the beta distribution and defines a probability density function on the simplex [15]. As a compound distribution, the Dirichlet-multinomial prevents the problem of estimating parameters from small quantities by setting a w-dimensional hyper-parameter vector $\boldsymbol{\alpha}$ that controls the smoothing level of the model:

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_w \alpha_w)}{\prod_w \Gamma(\alpha_w)} \prod_w \theta_w^{\alpha_w - 1} \tag{3}$$

The marginal likelihood of one sample can be now written as:

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \int p(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\alpha}) d\boldsymbol{\theta} \tag{4}$$

Using Equations 1 and 3 into Equation 4 leaves the marginal distribution of **x** as:

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\alpha}) =& \frac{n!}{\prod_w x_w!} \frac{\Gamma(\sum_w \alpha_w)}{\prod_w \Gamma(\alpha_w)} \int \prod_w \theta_w^{x_w + \alpha_w - 1} d\boldsymbol{\theta} \\ =& \frac{\Gamma(n+1)\Gamma(\sum_w \alpha_w)}{\Gamma(\sum_w x_w + \alpha_w)} \prod_w \frac{\Gamma(x_w + \alpha_w)}{\Gamma(x_w + 1)\Gamma(\alpha_w)} \end{aligned} \tag{5}$$

The integral part of Equation 6 corresponds to an unnormalized Dirichlet density so can now write the MAP estimator as follows:

$$\hat{\theta}_w^{MAP} = \frac{\sum_d x_{dw} + \alpha_w}{\sum_j \sum_d x_{dj} + \alpha_j} \tag{6}$$

It is interesting to notice that as the size of the data size grows $D >> 0$, the MAP estimator of θ becomes peaked at the value of the ML $\hat{\theta}_w^{MAP} \approx \hat{\theta}_w$. This feature allows us to effectively estimate parameters from high-dimensional and sparse data. If we are interested in computing a Bayesian estimate for α, we would have to consider the posterior distribution over the α hyper-parameters $p(\boldsymbol{\alpha}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$, however in this case exact marginalization is intractable. Instead, we can consider a flat prior and directly maximize $E(\alpha) = \log p(\boldsymbol{X}|\boldsymbol{\alpha})$:

$$\begin{aligned} \nabla E &= \frac{\partial E(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ &= D\Psi(\sum_w \alpha_w) - \sum_d \left[\Psi(\sum_w x_{dw} + \alpha_w) + \sum_w (\Psi(x_{dw} + \alpha_w) - \Psi(\alpha_w))\right] \end{aligned} \tag{7}$$

with $\Psi(x) = \frac{d \log \Gamma(x)}{dx}$.

There is no closed form for the maximum likelihood value of α, however using a lower of bound for the gradient of the log-likelihood ∇E, Minka [16] derived a fixed-point iteration technique based on the following update rule:

$$\alpha_w = \alpha'_w \frac{\sum_d \{\Psi(x_{dw} + \alpha_w) - \Psi(\alpha_w)\}}{\sum_d \{\Psi(\sum_w x_{dw} + \alpha_w) - \Psi(\sum_w \alpha_w)\}} \tag{8}$$

Now we want to model grape seeds images whose color distribution represents different ripening stages. However, instead of considering a single prior distribution, we consider a mixture of K Dirichlet distributions with parameter $\boldsymbol{A} = \{\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_K\}$, where each component of the mixture is a Dirichlet distribution with parameter $\boldsymbol{\alpha}_k$. The distribution for the θ parameter can be written as:

$$p(\boldsymbol{\theta}|\boldsymbol{A}, \boldsymbol{\lambda}) = \sum_k \lambda_k \frac{\Gamma(\sum_w \alpha_{kw})}{\prod_w \Gamma(\alpha_{kw})} \prod_w \theta_w^{\alpha_{kw}-1} \tag{9}$$

where $\lambda = (\lambda_1,...,\lambda_K)$ is vector containing the mixture weights.

The marginal likelihood of the observed data given the DMM prior is now:

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{\lambda}) &= \sum_k \lambda_k p(\boldsymbol{x}|\boldsymbol{\alpha}_k) \\ &= \sum_k \lambda_k \frac{\Gamma(n+1)\Gamma(\sum_w \alpha_{kw})}{\Gamma(\sum_w x_w + \alpha_{kw})} \prod_w \frac{\Gamma(x_w + \alpha_{kw})}{\Gamma(x_w + 1)\Gamma(\alpha_{kw})} \end{aligned} \tag{10}$$

III. MATERIALS

The data set contains 289 images from grape seeds acquired using a conventional scanner (Canon MG-3110), obtained between February and May 2013 (For a complete description of the dataset, the reader is referred to [5]). Figure 1 shows three images corresponding to different maturity stages. Figure 1(a) corresponds to an image where most seeds have a low grade of maturity. Figure 1(b) contains seeds with mixed maturity grades and finally Figure 1(c) shows an image where most seeds are mature. It is possible to appreciate how seeds with lower level of maturity shows green colors shades, but these colors disappear in posterior images, where brown tones have bigger influence.

(a) Feb. (2013-02-14) (b) Mar. (2013-03-20) (c) May (2013-05-06)

Figure 1: Grape seeds images with different maturity levels. There is a predominant presence of light green colors in the images taken during February. These tonalities begin to disappear during March and May where brown colors are gradually introduced.

From Figures 1(a), (b) and (c) we can appreciate the color difference for the different ripeness stages. Therefore, it is possible to represent the color histograms using a Dirichlet mixture model. Using this method, we expect a group of grape seeds collected at any stage to be represented by a single Dirichlet distribution.

IV. COLOR HISTOGRAM

To achieve an effective representation of the color distribution, the first step was to separate the background from the seed. A well-known method for image segmentation is therefore proposed for this task [17]. Figure 2 shows the result after applying the Otsu method.

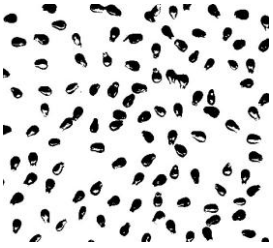
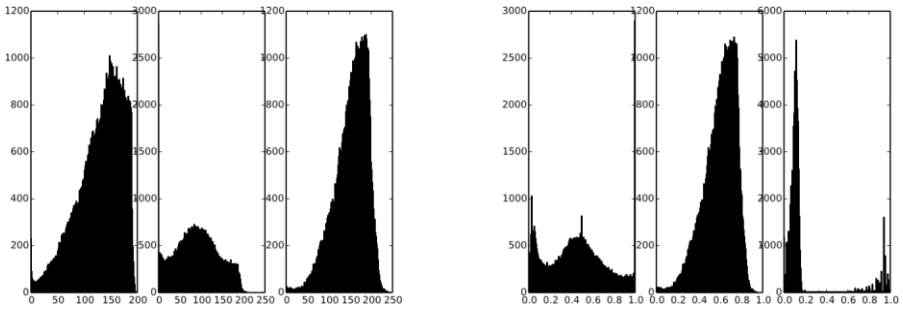


Figure 2: Segmentation using the Otsu method.

Having obtained the targeted zone in each image, we extract the pixels from the seeds and remove the pixels from the background. From this data, we can discretize the color distribution of

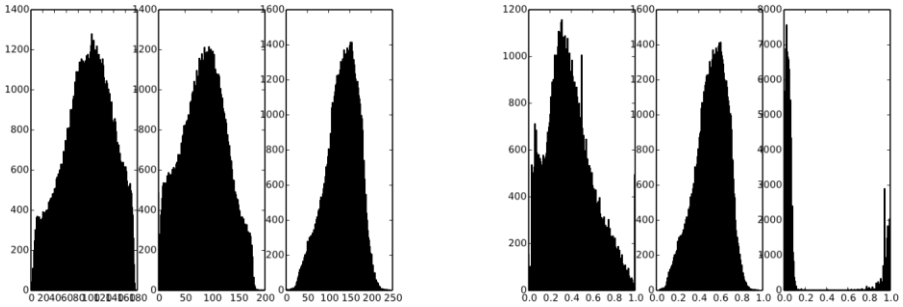
the seeds by partitioning the color space axes at regular intervals and therefore producing rectangular bins. Usually, one or two axes having more information are selected and more bins are chosen these axes. The histograms from Figure 3 corresponds to grape seeds in an early ripeness stage, while histograms from Figure 4 corresponds to a late ripening stage. Figures 3(a) and 4(a) shows the color histograms for channels Red, Green and Blue in the RGB color space. Figures 3(b) and 4(b) shows the color histograms for Hue, Saturation and Value in the HSV color space.

Notwithstanding the correctness of the approach, the resulting histograms from the RGB color space are sparse and there are well known drawbacks when making computations for fitting mixture models to this data [16]. The HSV color space is primarily different from the RGB color space, due to its separation of color information (chromacity) and intensity (luminance). Because of its invariance properties, is widely used in content-based image retrieval applications. Figures 3 and 4 show RGB and HSV histograms of grape seeds images obtained in two different stages.
In our experiments, we use the RGB and HSV color spaces for histogram generation. Different number of bins are selected and compared in terms of the sparseness of the resulting histograms. Figures 5 show the histograms for the RGB and HSV color spaces using a total of 512 bins, where 16 bins where used for the R and G channels and only 2 bins were used for the B channel. Figures 5(a) and 5(b) corresponds to the RGB color space of low maturity grape seeds and 5(c) and 5(d) to a posterior stage. There are many bins without any data and some bins contains most of the information. This problem stems from nature of the RGB color space and the data. Grape seeds images contain light green and brown colors, which is only a small part of the full color spectrum.



(a) RGB (b) HSV

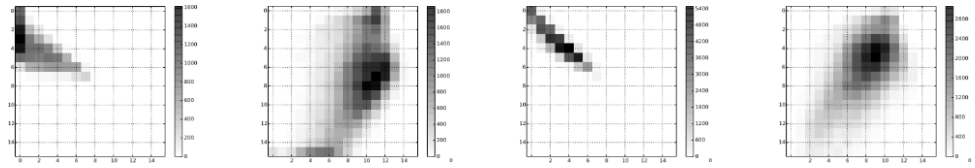
Figure 3: Color histograms of grape seeds samples obtained in February (2013-02-14)



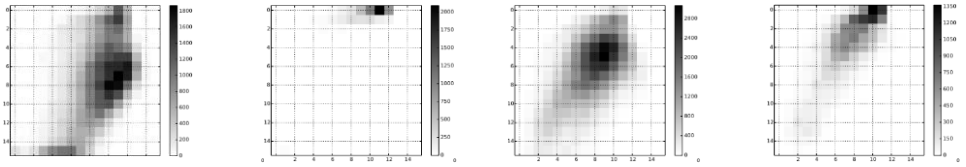
(a) RGB (b) HSV

Figure 4: Color histograms of grape seeds samples obtained in May (2013-05-06).

The HSV color space separates the intensity (luminance) and color (chromaticity) [18,19]. As with the RGB discretization, we used 16 channels for the S and V channels and only 2 bins for the H channel. Using this discretization, from Figures 5(e) it can be seen how most bins are occupied. This is also true for late maturity stages whose color discretization can be seen in figures 5(g) and 5(h).



(a) RGB-I-I (b) RGB-I-I (c) RGB-II-I (d) RGB-II-II



(e) HSV-I-I (f) HSV-I-II (g) HSV-II-I (h) HSV-II-II

Figure 5: 2-D color histograms for grape seeds images obtained at different maturity stages. (a) and (b) RGB color discretization for early ripening stage, (c) and (d) RGB color discretization for late ripening stage, (e) and (f) HSV color discretization for early stage of ripening, (g) and (h) HSV color discretization for late ripening stage.

V. RESULTS

In this section, we apply the Dirichlet mixture model to the data obtained from color discretization. The model was tested using different number of bins and different number of mixture components. As mentioned in Section 4, the discretization of the HSV color space was performed, leaving an effective number of colors (ENC) corresponding to the number of bins with non-zero data. Table 1 summarizes the number of bins per channel and the ENC value obtained for each discretization.

Bins	H	S	V	ENC	Sparsity proportion
8	2	2	2	8	0%
32	2	4	4	31	3.125%
128	2	8	8	115	10.16%
288	2	12	12	249	13.54 %
512	2	16	16	434	15.23%
1152	2	24	24	949	11527.62%
2048	2	32	32	1631	20.36%
8192	2	64	64	5983	27.0%
20000	2	100	100	13250	33.75%

Table 1: Quantization of the HSV color space and effective number of colors

To evaluate the quality of a discretization and at the same time to assess the performance of a model (selecting the number K of mixture components), we estimate the parameters A and λ and then calculate the *perplexity* of the model, defined as:

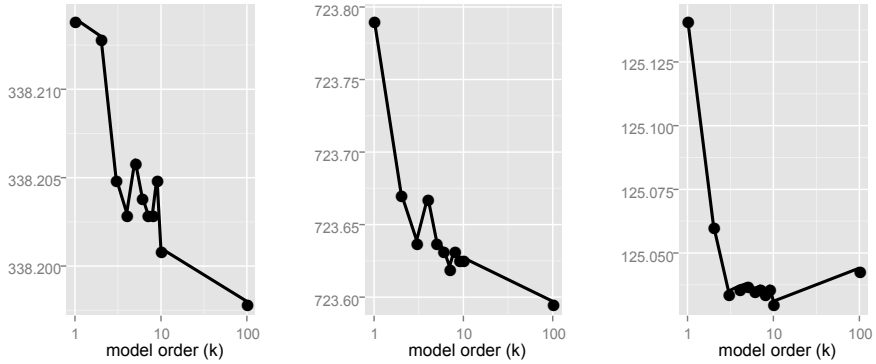
$$pp = \exp\Big(-\frac{\sum_d \sum_w \log(x_{dw})}{D}\Big) \qquad (11)$$

Perplexity is a widely used metric for evaluating language models and represents the branching factor of the model, therefore if a model contains symbols (words or colors) appearing with higher probability than others, the model would have less perplexity [20]. Lower values in this metric indicate better predictive quality, since the model is able to use the observed features to discriminate between different modalities. Table 2 shows the perplexity values achieved for each discretization and model order K considered.

Bins/K	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	100.00
8.00	449.54	449.54	449.54	449.54	449.54	449.54	449.54	449.54	449.54	449.54	449.54
32.00	102.36	102.36	102.36	102.36	102.36	102.36	102.36	102.36	102.36	102.36	102.36
128.00	338.21	338.21	338.21	338.20	338.21	338.20	338.20	338.20	338.21	338.20	338.20
288.00	723.79	723.67	723.64	723.67	723.64	723.63	723.62	723.63	723.63	723.63	723.60
512.00	125.14	125.06	**125.04**	125.04	125.04	125.04	125.04	125.04	125.04	125.03	125.04
1152.00	277.91	277.46	277.15	277.08	277.04	276.98	276.98	276.98	276.95	276.96	
2048.00	486.69	485.36	484.28	483.97	483.82	483.64	483.63	483.63			
8192.00	1932.52	1919.27	1904.82	1899.38	1897.01						
20000.00	4524.73	4471.15	4415.11	4400.17							

Table 2: Perplexity values for different models and quantization levels

Figure 6 shows the perplexity values for distinct number of bins and mixture components. Figure 6(a) shows the perplexity values for different models and a discretization using 128 bins. Using this discretization, the perplexity values tend to decrease when increasing the number of mixture components. However, the improvement is not significant for models with more than 3 components. This tendency is also perceived in figures 6(b) and 6(c) when using finer quantization levels (higher number of bins). From Figure 6(c) the performance of the model with 3 components is good enough for our purposes.



(a) 128 bins (b) 288 bins (c) 512 bins

Figure 6: Perplexity achieved by different number of bins and number of mixture components. Figure (a) corresponds to the perplexity score achieved with 288 bins. Figure (b) shows the perplexity score with 512 bins.

We are also interested in comparing the quality of the different discretization schemes. Table 2 shows no improvement in perplexity for quantization with 8 and 32 bins using a more components, this could indicate that there are no clusters in the data. However, given that the perplexity values tend to increase with a finer quantization. Table 3 shows the average perplexity.

Bins/K	1	2	3	4	5	6	7	8	9	10	100
8	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
32	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
128	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
288	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
512	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
1152	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	
2048	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30			
8192	0.32	0.32	0.32	0.32	0.32						

Table 3: Average perplexity per color for different models and quantization levels.

Figure 7 shows the average perplexity for the different quantization levels with different number *K* of components. Although finer quantization is less sparse, the average perplexity tends to increase. That means that fine grained resolutions bins are less capable to discriminate between different ripening stages. In the other hand, there is no improvement for quantization levels above

512 bins for all models considered. This could be explained by the reduced color spectrum of the images and the sparsity that results from fine-grained resolutions.

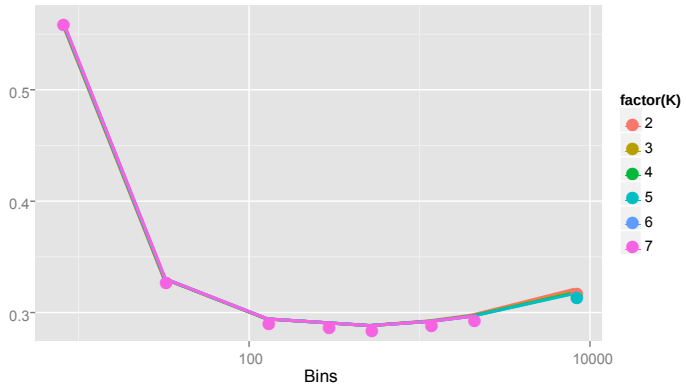


Figure 7: Average perplexity per color for different numbers of bins. Average perplexity depends on the perplexity of the model and the effective number of colors.

Figure 8 shows the class membership values for 24 grape seeds images corresponding to different dates using a 32 bins quantization. Figure 8 (a) corresponds to a model with 2 components and 8 (b) a model with 3 components. The first rows of each image correspond to a late ripening stage, while the last rows correspond to early ripening stages. In the other hand, the columns indicate the cluster membership. Because of the coarse-grained discretization, from figures 8 (a) and 8 (b), there is no apparent transition among the different ripening stages.

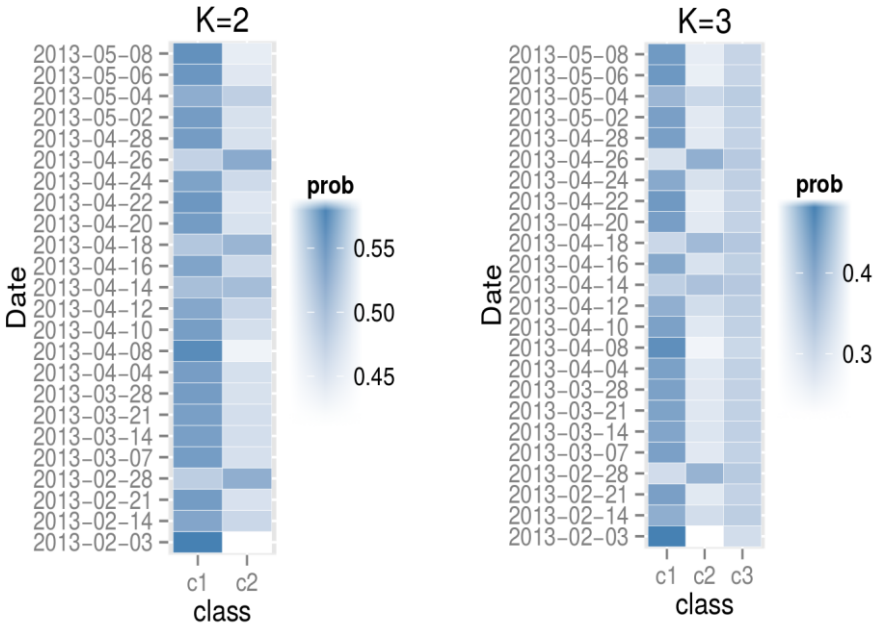


Figure 8: Class membership values using a 32 bins quantization.

When using a 512 bins quantization it becomes possible to observe well defined clusters of data. From Figure 9 (a) we can notice that grape seeds images from early maturity stages being well represented by a single cluster (c1) while images from late maturity stages being represented by cluster c2. When using 3 mixture components (see Figure 9 (b)) it is also possible to assign early maturity seeds to a single cluster (c3). However, clusters c1 and c2 have cannot be identified to different ripening stages.

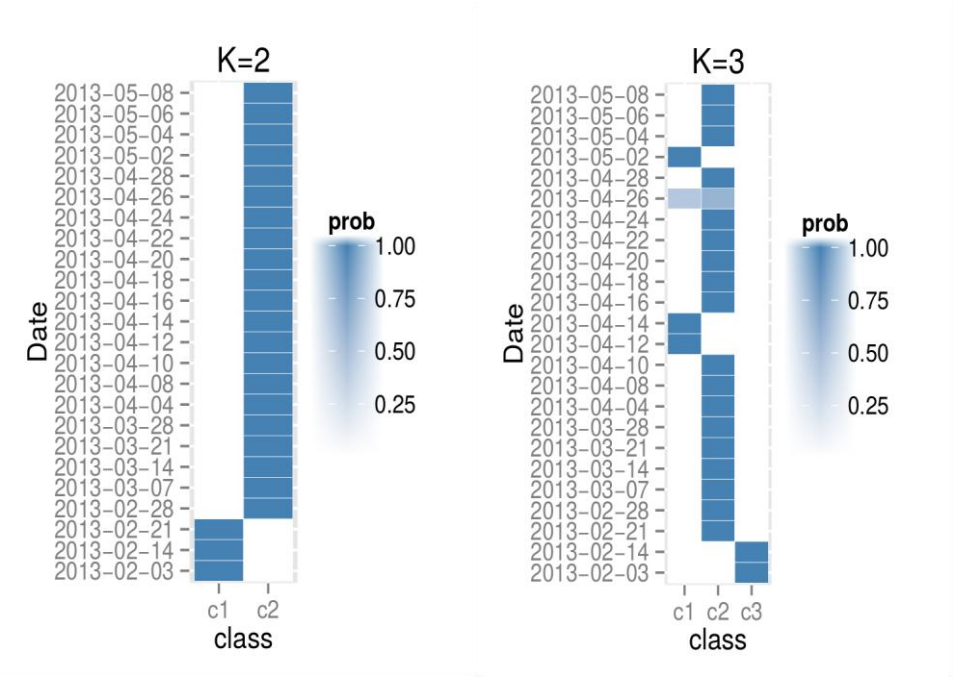


Figure 9: Class membership values using a 512 bins quantization.

Now using a fine-grained quantization (2048 bins) we obtain similar results to the 512 bins. It is possible to observe similar results to the ones observed with 512 bins, where clear separation exists for early ripening stages. However, from Figure 10 (a) and 10(b) we can also observe that it is not possible to distinguish among different late ripening stages using color only information.

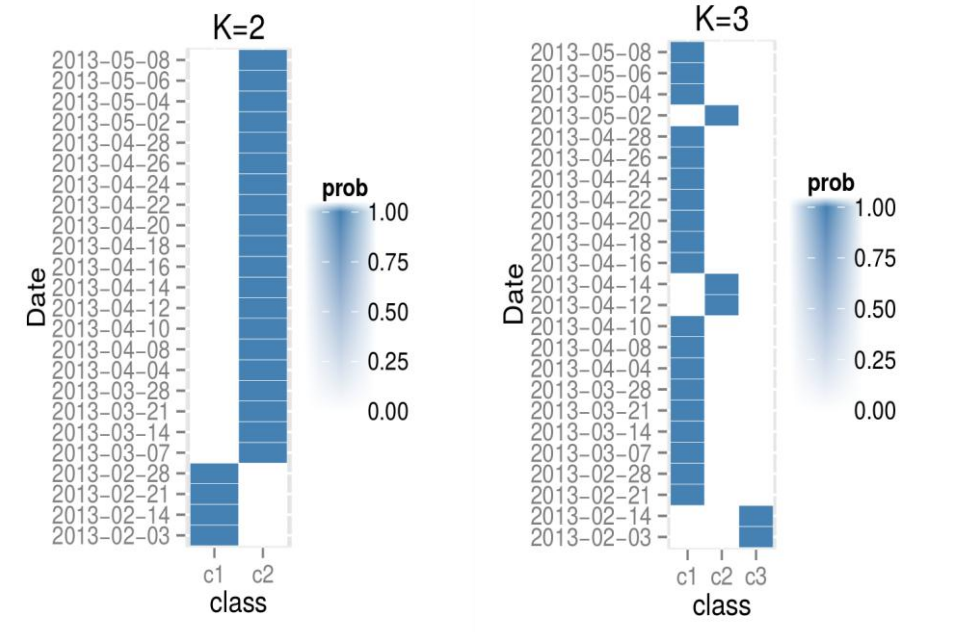


Figure 10: Class membership values using a 2048 bins quantization.

VI. CONCLUSIONS

In this paper, we presented a method for unsupervised learning of ripening stages from grape seeds images. Using a Dirichlet mixture model, the different clusters are estimated from data and each one of the grape images are probabilistically assigned to each cluster using a cluster membership indicator.

Firstly, a discretization with different number of bins was performed on the HSV color space. Then, the data from grape seeds images captured at different time frames (corresponding to different ripening stages) was used to train a Dirichlet mixture model using the EM algorithm. Subsequently, for any grape seeds image we can estimate the membership to any of the K clusters representing a different ripening stage. From our results, it is possible to observe a clear separation from early ripening stages to the following stages. There is a clear distinction between the grape seeds color in this stage and eventually it becomes more complex to estimate later ripening. However, using the perplexity values we could determine the optimal quantization and number of components for the mixture model.

The average perplexity values confirm that there is no performance improvement when using more than $K = 3$ components. This is confirmed when plotting the class membership values (see figures 8,9 and 10) where there are no clear temporal patterns when using more than 3 components.

Finally, it is worth mentioning that this methodology allows us to infer between different ripening stages from grape seeds, without the need of an expert domain who label the samples. However, the performance of the method can be improved by having chemical analysis associated to the grape seeds samples.

Acknowledgment

This work was funded by **CONICYT PAI/ACADEMIA 79112009**. National Commission for Scientific and Technological Research (CONICYT), Chilean Government.

REFERENCES

[1] Adams, D.O. Phenolics and Ripening in Grape Berries. American Journal of Enology and Viticulture **2006**, 223 57, 249–256.

[2] Rodriguez-Pulido, F.J.; Ferrer-Gallego, R.; Gonzalez-Miret, M.L.; Rivas-Gonzalo, J.C.; Escribano-Bailon, M.T.; Heredia, F.J. Preliminary study to determine the phenolic maturity stage of grape seeds by computer vision. Analytica Chimica Acta **2012**, 732, 78 – 82. A selection of papers presented at In Vino Analytica Scientia.

[3] Fredes, C.; Bennewitz, E.V.; Holzapfel, E.; Saavedra, F. Relation between Seed Appearance and Phenolic Maturity: A Case Study Using Grapes cv. Carmenere. Chilean Journal of Agricultural Research **2010**, 70, 381–389.

[4] Rodriguez-Pulido, F.J.; Gomez-Robledo, L.; Melgosa, M.; Gordillo, B.; Gonzalez-Miret, M.L.; Heredia, F.J. Ripeness estimation of grape berries and seeds by image analysis. Computers and Electronics in Agriculture **2012**, 82, 128 – 133.

[5] Avila, F.; Mora, M.; Fredes, C. A method to estimate Grape Phenolic Maturity based on seed images. Computers and Electronics in Agriculture **2014**, 101, 76 – 83.

[6] Avila, F.; Mora, M.; Oyarce, M.; Zuniga, A.; Fredes, C. A method to construct fruit maturity color scales based on support machines for regression: Application to olives and grape seeds. Journal of Food Engineering **2015**, 162, 9 – 17.

[7] Swain, M.J.; Ballard, D.H. Color Indexing. Int. J. Comput. Vision **1991**, 7, 11–32.

[8] Lee, D.J.; Archibald, J.K.; Chang, Y.C.; Greco, C.R. Robust color space conversion and color distribution analysis techniques for date maturity evaluation. Journal of Food Engineering **2008**, 88, 364–372.

[9] Zhang, D.; Lee, D.J.; Tippetts, B.J.; Lillywhite, K.D. Date maturity and quality evaluation using color distribution analysis and back projection. Journal of Food Engineering **2014**, 131, 161–169.

[10] Sefidpour, A.; Bouguila, N. Spatial color image segmentation based on finite non-Gaussian mixture models. Expert Systems with Applications **2012**, 39, 8993 – 9001.

[11] Rigouste, L.; Cappé, O.; Yvon, F. Inference and evaluation of the multinomial mixture model for text clustering. Information Processing & Management **2007**, 43, 1260 – 1280.

[12] Figueiredo, M.A.T.; Jain, A.K. Unsupervised Learning of Finite Mixture Models. IEEE Trans. Pattern Anal. Mach. Intell. **2002**, 24, 381–396.
[13] Bouguila, N.; Ziou, D. Using unsupervised learning of a finite Dirichlet mixture model to improve pattern recognition applications. Pattern Recognition Letters **2005**, 26, 1916 – 1925.
[14] Bouguila, N.; Ghimire, M.N. Discrete visual features modeling via leave-one-out likelihood estimation and applications. Journal of Visual Communication and Image Representation **2010**, 21, 613 – 626.
[15] Ng, K.W.; Tian, G.L.; Tang, M.L. Dirichlet and related distributions: Theory, methods and applications; Vol. 888, John Wiley & Sons, 2011.
[16] Minka, T.P. Estimating a Dirichlet distribution. Technical report, 2003.
[17] Otsu, N. A Threshold Selection Method from Gray-Level Histograms. Systems, Man and Cybernetics, IEEE Transactions on **1979**, 9, 62–66.
[18] Belongie, S.; Carson, C.; Greenspan, H.; Malik, J. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. Computer Vision, 1998. Sixth International Conference on, 1998, pp. 675–682.
[19] Xiao, H.W.; Zhang, J. Image Segmentation algorithm based on color features: Case study with giant panda, International Journal on Smart Sensing and Intelligent Systems **2016,** 9, 799-817.
[20] Murphy, K.P. Machine Learning: A Probabilistic Perspective; The MIT Press, 2012.