Fruit Maturity Estimation based on Color Scales

Marco Mora, Miguel Oyarce, and Claudio Fredes

Laboratory of Technological Research in Pattern Recognition
Universidad Catolica del Maule
Chile
{mora@spock.ucm.cl,moyarce@litrp.cl}
http://www.litrp.cl

Abstract Color is an important parameter used to estimate fruit maturity and optimal harvest time. In this paper a general methodology for estimating fruit maturity based on the development of a color scale is proposed. Maturity estimation is performed by computing of the fruit representative color and its comparison with the colors of the scale. In experimentation the case of grape maturity estimation based on seed color is presented. The method is simple and effective, allowing its application in the real processes.

Keywords: Fruit Maturity, Color Scale, Color Estimation.

1. Introduction

Color is an important parameter in determining fruit maturity. For fruits such as grapes, raspberries, olives, among many others, there are color charts associated with fruit maturity, maturity being the most important indicator in determining the optimal time to harvest.

The color estimation of fruits is performed by various methods. The simplest method is the visual estimation, in which an expert inspects the surface of a fruit and determines its color. This color is compared with those present in a color scale. Is finally decided to wich color inside the scale corresponds the fruit inspected. In general, the visual estimation of color is a very simple and inexpensive method, but at the same time is inaccurate and subjective, as it depends on the experience of a human expert. Besides the above is hardly representative, since in general the expert cannot perform a significant amount of analysis.

Another method of estimating the color of the fruit is based on the use of instruments, such as colorimeters. The use of tools for color estimation allows measurements with high precision. However, suitable instruments for fruits color measurement are expensive, and keeping the disadvantage of not allowing high representation analysis.

In order to address the above problems, Computer Vision Systems (SVC) has been developed based on digital images to inspect and estimate the color of fruits. In [1] an SVC was reported to estimate the antioxidant content and phenols of the carrot based on surface color of the fruit. The color is determined based on the calculation of the center of mass of the bidimensional histogram composed of a and b channels of the CIE Lab color space. In [2] is designed and implemented an inexpensive SVC to measure the color of a heterogeneous food product in shape and color. A method to estimate the representative color through the average per canal in the Cie Lab color space is presented, reporting experiments in potatoes. In [3] an index is proposed to estimate the maturity of sweet cherry, analyzing outdoor images. The system implements a stage of removing the light hit analyzing the G channel and redness index is obtained based on the colors set in a color chart. In [4] an index of maturity of the tomato, which allows classifying the fresh fruit into 6 classes according to the USDA standard, is proposed. Tomato images are transformed to HSI color model and information Hue (channel H) is used to construct the proposed index. As shown in previous studies, the color of the fruits is a robust indicator of maturity, which allows performing accurate estimations of optimal harvest time.

This paper proposes a new and effective method to estimate fruit maturity. The method is based on the comparison of the color of the fruit regarding a color scale constructed with samples representing the color evolution. The method involves the following three stages: (i) A methodological strategy is proposed to determine a robust descriptor of the color of an object, using as ground of truth the information provided by a colorimeter. (ii) An original method is proposed to create scales of fruit colors. (iii) The process of visual fruit color estimation is formulated as an image processing problem. The representative color is calculated and compared with those present in the color scale, using euclidean distance of points in the CIE Lab space. To acquire images the method considers the use of a conventional scanner, allowing a massive acquisition of samples in a controlled lighting environment. The advantages of the method are: Allows automating the process of estimation of maturity at low cost (due to the use of the scanner). The experiments of this work are performed in pictures from grape seeds, being the analysis of the seeds one of the current trends for the estimation of the maturity of the grapes.

The structure of the paper is as follows: In section 2 a robust descriptor of the representative color of an object is determined by comparing various methods with the color given by a colorimeter. Section 3 presents the methodology for creating color scales that represents the maturation process. In section 4 materials and methods are presented which will be used. In section 5 the case of the color estimation of the seed and grape maturity is presented. Finally in section 6 presents the conclusions.

2. Validation of methods for calculating the Representative Color

In this section the methods to evaluate the representative color will be presented. This color will be contrasted with a colorimeter to know which method provides a better approach.

2.1. Methods Presentation

Calculate the representative color is complex due to the objects present many colors. Among the most common techniques for estimating the color of objects simple methods are found such as: average per channel (M1), mode per channel (M2) and median per channel (M3).

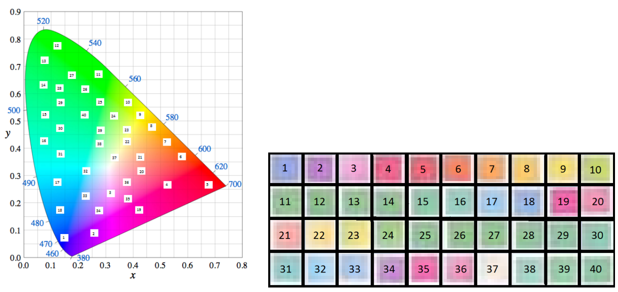
There are also more complex methods that use the information of the three channels to estimate the representative color. One of these methods consists in calculating a vector median of the colors through morphological ordering (M4). The method of vector median consists in ordering the color vectors, and then chooses the central element. It differs from the computation of the scalar median in that the elements that are ordered are vectors of three components. A work that reports very good ordering results of color are reported in [5]. The previous work corresponds to a generalized proposal of ordering the lighter colors to darker. This method is suitable for the case of the seed, since it is known that grape seed undergoes a color change that occurs in light green and ends in dark brown.

2.2. Validation of Methods

M1 to M4 methods presented above will be implemented to estimate the color on samples images which will be compared with the estimated color by the RGB-1002 Lutron colorimeter.

It is noteworthy that the colorimeter is adopted as area of truth and the comparison method is based on the calculation of the Euclidean distance between the color determined by the colorimeter and the estimated color by each method. This is possible due to the use of the CIE Lab color model.

40 samples were selected from the chromaticity diagram, which is adopted to give wide coverage to all colors. In figure 1(a) you can see the labeling of the samples obtained..



(a) (b)

Figure 1. Color samples in Chromaticity Diagram.

The samples obtained from the chromaticity diagram can be seen in figure 1(b). It is observed that the samples do not have a single color, existing a gradient of colors in each of them.

The previous samples were printed on paper and the color was measured using the colorimeter. After the above samples were scanned and the representative color of the samples was estimated based on the four methods described in the previous section.

The distances between the color obtained by the colorimeter and each of the methods being compared are shown in table 1.

Sample	M1	M2	M3	M4	-	Sample	M1	M2	M3	M4
1	31.62	35.39	34.76	35.25	-	21	25.17	32.23	27.64	29.81
2	40.90	45.49	45.82	47.50	-	22	24.15	37.32	24.58	25.07
3	29.14	31.37	30.61	31.35	-	23	32.23	40.11	33.95	35.36
4	46.86	64.69	53.86	55.80	-	24	26.25	25.33	28.18	30.85
5	46.14	61.98	49.88	53.21	-	25	26.31	29.03	28.54	29.14
6	40.96	34.88	45.07	46.28	-	26	27.24	32.67	30.14	31.89
7	34.67	26.10	35.14	33.47	-	27	26.31	33.09	28.73	30.37
8	34.70	21.07	36.26	39.16	-	28	25.01	27.88	28.59	30.78
9	34.23	32.38	34.37	33.26	-	29	24.60	27.76	27.51	30.49
10	32.55	51.22	34.40	35.80	-	30	22.77	25.21	25.60	27.64
11	27.23	31.75	30.01	31.92	-	31	21.00	24.43	23.67	24.59
12	27.10	33.58	31.08	33.02	-	32	19.28	18.06	20.38	21.76
13	25.06	30.09	27.56	28.39	-	33	20.49	21.75	22.36	22.48
14	24.87	26.23	28.23	30.62	-	34	31.61	33.66	34.45	33.30
15	23.43	30.70	27.11	28.26	-	35	43.07	8.00	49.17	50.94
16	20.12	29.42	23.59	24.26	-	36	30.24	27.54	31.25	34.61
17	20.60	25.01	22.02	22.67	-	37	6.51	9.73	6.34	8.02
18	25.45	29.24	28.92	29.26	-	38	17.45	20.01	19.44	21.22
19	46.81	40.84	55.81	56.57	-	39	23.26	24.36	26.26	27.88
20	32.28	31.48	34.88	37.09	-	40	24.42	32.18	28.65	31.78
						Mean	**28.55**	**32.78**	**31.08**	**31.37**

Table 1. Error of the color estimation methods.

By averaging the distances of the 40 samples results that the method based on the average per channel has the lowest value. Therefore, the averaging method per channel is chosen for estimating the representative color of an object.

3. Creation of the Color Scale

A computational method for estimating color scales based on samples and Support Machines for Regression (SVR) is proposed.

For a set of acquired samples is estimated the representative fruit color by the proposed method in [5]. After estimating the color of the samples a scale of continuous color is constructed, for which a regression based on the estimated SVR considering the color samples is performed. In figure 2(a) can be observed the dataset, where each point represents a color in the Cie LAB model.

The SVR continuous outputs are generated based on the input samples similarly to the case of the classification [6].

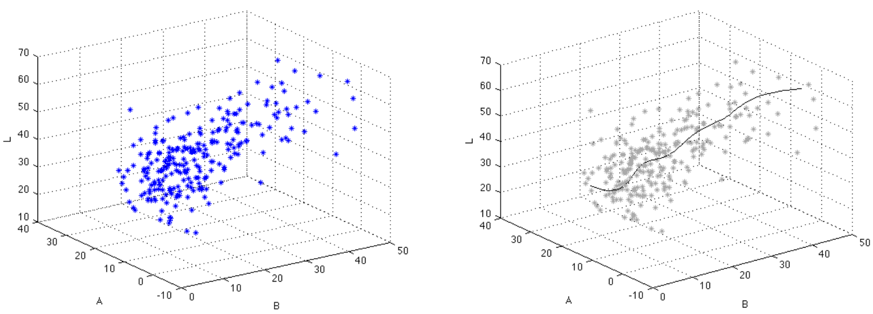
It provides a set of data as $\mathbf{x} = (x_1, x_2, x_3)$. These values correspond to the channels L,A and B respectively. Moreover, the traditional model of the SVR has an output y corresponding to a real value, which limits us if we seek based on a single color to estimate the other two with a single model. Because of this it is proposed to create two estimation models, which based on a channel estimate another 2 channels of the color model. The proposed model is defined as follows:

$$L_E(B) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(B_i, B) + b \tag{1}$$

and

$$A_E(B) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(B_i, B) + b \tag{2}$$

where B corresponds to the values of the channel B and L_E y A_E the estimated values for the channels L y A respectively. Figure 2(b) shows the results of the regression applied to the sample set. It is possible to appreciate how the generated curve continues in soft form the central tendency of the samples.



(a) Data Set (b) SVR Curve

Figure 2. Color Scale Estimation Model

4. Method for estimating maturity

The image processing methodology involves the steps of image acquisition, segmentation, descriptors calculation and classification. These stages are extensively described in classical texts of image processing and pattern recognition [7].

- Controlled environment for image acquisition: The use of a controlled environment allows maintaining constant lighting conditions for measurements over time. Overall the images acquired by optical devices have two problems reported in the literature of image processing, namely, the shadows and light bumps. The above problems lead to difficulties in the correct segmentation of objects.
- Robust segmentation to image processing problems: As mentioned above, the images present two defects: shadows and light bumps. For robust segmentation to the above defects from a digital RGB image, the color invariant to lighting model is adopted. $c_1c_2c_3$ [8]. The expressions of the model are as follows:

$$c_{1i,j} = arctan\frac{R_{i,j}}{max(G_{i,j}, B_{i,j})} \tag{3}$$

$$c_{2i,j} = arctan\frac{G_{i,j}}{max(R_{i,j}, B_{i,j})} \tag{4}$$

$$c_{3i,j} = arctan\frac{B_{i,j}}{max(G_{i,j}, R_{i,j})} \tag{5}$$

  where $R_{i,j}$, $G_{i,j}$ y $B_{i,j}$ represent a pixel in the red, green and blue components in the image I. It has been chosen the channel c_3 due to the good results obtained in the work proposed in [9]. The automatic segmentation of the channel c_3 is performed by the known Otsu method [10].
- Descriptor calculation: After the images of the fruit are segmented, we proceed to the representative color computation, by calculating the average per channel. Using this method is adequately justified in section 2.
- Estimation of the degree of maturity: After the computation of the representative color the Euclidean distance is used to determine which the fruit color in the color scale is.

5. Results

The results of this paper show the application of the method to the case of estimating grape maturity based on images from the seed. The analysis of the seed is a recent line of research present in the literature. In a very recent study, conducted by authors of this paper [11], a method to distinguish two states of seed maturity, namely immature seeds and mature seeds is presented. By using a simple neural classifier, a method that achieves high hit rates, which allows provide objective information that can be used to decide the optimal time to harvest is proposed.

In figure 3 a set of 27 seeds which are used in the experimentation of this work are presented.

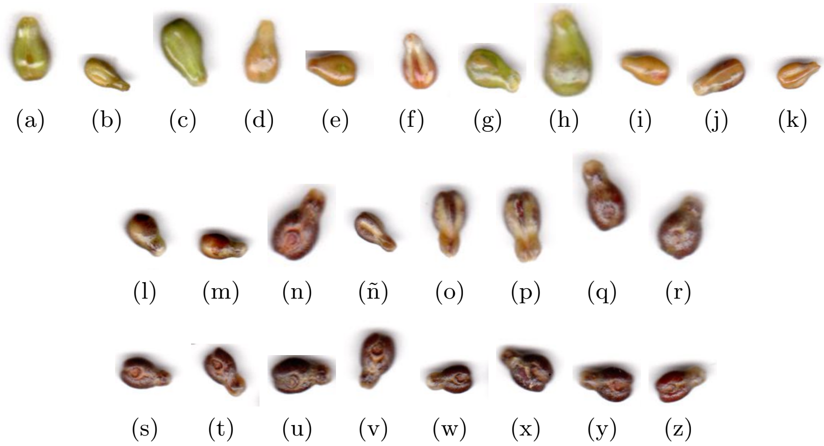
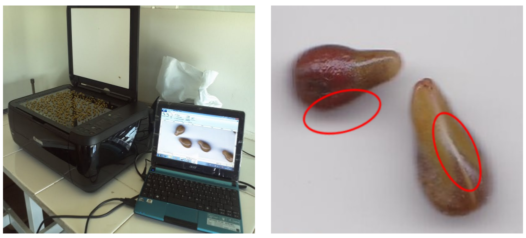


Figure 3. Seed of observation.

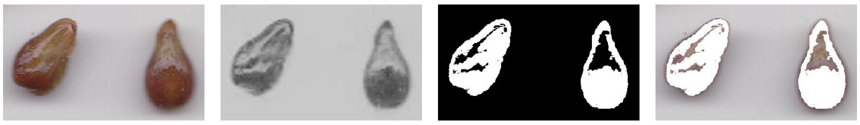
In the case of the seeds has used a scanner to capture images. This device allows the analysis of a large number of samples. Besides the above, since the scanner is a conventional purchase device allows acquisition of images at low cost. The figure 4(a) corresponds to the acquisition system based on the scanner and figure 4(b) shows the aforementioned defects.



(a) Scanner (b) Shadows and highlights

Figure 4. Scanner and problems identified in the acquired images.

The result of the segmentation of the seeds are presented in the figure 5. The figure 5(a) presents seed images in RGB model, in which the presence of shadows and light bumps are observed, figure 5(b) shows the channel c_3, the figure 5(c) shows the result of the segmentation of said channel, and finally figure 5(d) shows the result of the segmentation of the original image. Is observed in the latter figure that the segmentation process allows to select pixels that are only in the seed, excluding pixels of the shadow. Moreover, the segmentation method excludes defective pixels due to light bumps.



(a) Original Image (b) Channel c_3 (c) Otsu method (d) Seed segmentation

Figure 5. Seed Segmentation.

For obtaining information on the maturity of the seed a human expert performs the classification of them into three classes: immature seeds, mature seeds and overmature seeds. In the figure 3 seed images are shown. Expert classification of this set of samples is as follows:

- Immature: From the subfigure (a) to subfigure (k).
- Mature: From the subfigure (l) to the subfigure (r).
- Overmature: From the subfigure (s) to the subfigure (z).

The color scale created from the set of seed samples is as follows:

Figure 6. Color scale for Cabernet Sauvignon strain.

One can appreciate that the pattern meets obscuration in the seed (from light green to dark brown), indicating that the generated scale is consistent with the evolution of seed maturity.

To solve the problem of estimation of seed maturity three reference colors are defined, which represents the immature, mature and overmature classes. To differentiate the stages of maturity well it has been decided on choosing the extremes of the scale colors to represent the immature and overmature class The left end corresponds to the immature class and the right end to the overmature class. It has also chosen the midpoint of the scale for the mature category. These reference colors for the above classes are shown in figure 7.

(a) Inmature (b) Mature (c) Overmature

Figure 7. Color references of Maturity Level in CIE Lab model. Immature = (47.26,7.00,38.00), Mature = (39.48,14.12,20.82), Overmature=(31.70,11.00,4.00).

Defined the reference colors for each class we proceed to calculate the maturity of the seed through the Euclidean distance. Counting results of maturity are presented in table 2, which shows the distances calculated between the representative color of each seed regarding the reference color for each class.

Seed	Inmature	Mature	Overmature	-	Seed	Inmature	Mature	Overmature
(a)	**15.78**	32.91	44.47	-	(ñ)	14.54	**6.89**	17.76
(b)	**8.62**	25.37	37.82	-	(o)	11.89	**8.37**	20.53
(c)	**15.78**	32.27	42.73	-	(p)	11.47	**10.01**	21.50
(d)	**17.75**	30.47	42.50	-	(q)	20.05	**3.38**	12.19
(e)	**11.96**	22.43	35.84	-	(r)	21.41	**5.55**	9.93
(f)	**20.36**	26.33	37.21	-	(s)	26.49	9.41	**6.46**
(g)	**14.54**	31.67	43.18	-	(t)	26.85	10.24	**10.15**
(h)	**16.71**	33.60	44.88	-	(u)	41.90	27.16	**17.62**
(i)	**15.71**	27.60	40.93	-	(v)	28.98	12.59	**9.50**
(j)	**11.77**	14.27	27.67	-	(w)	33.44	18.04	**11.76**
(k)	**14.22**	23.60	36.15	-	(x)	31.40	15.56	**6.15**
(l)	23.68	**10.24**	13.39	-	(y)	32.19	16.20	**9.25**
(m)	23.76	**12.15**	19.27	-	(z)	31.66	16.00	**14.85**
(n)	22.66	**6.56**	12.76	-				

Table 2. Distances calculated between the representative color of each seed regarding the reference color for each class.

In the table 2 is highlighted with bold characters of the smallest distance of seed color regarding the reference colors. Given the information we can infer that the method allows to identify from 3 different classes without problem. It is noted that the classification of the seeds by the method previously proposed is entirely consistent with the rating given by the expert.

6. Conclusion

In this paper has been presented a method for the estimation of fruits. First, a search was conducted in the literature in order to adopt common features of similar systems, coming as a result of the acquisition device selection, and the Cie Lab color model. Secondly, the method of estimation of the color to use was determined resulting on the average per channel nearest to the colorimeter measurements. In third place was developing a method for creating fruit color scale based on SVR. Finally the computation of the state of maturity based on Euclidean distance allows solving the problem simply, with wide perspectives for implementation in real systems.

Acknowledgment

This work is funded entirely by FONDEF IDeA CA12i10236 Project: Grape Phenolic Maturity Estimation based on Seed Images. Marco Mora, Claudio Fredes, Sergio Hernandez. National Commission for Scientific and Technological Research (CONICYT), Government of Chile.

Referencias

1. Pace, B., Cefola, M., Renna, F., Renna, M., Serio, F., Attolico, G.: Multiple regression models and computer vision systems to predict antioxidant activity and total phenols in pigmented carrots. Journal of Food Engineering **117** (2013) 74 – 81
2. Pedreschi, F., Leon, J., Mery, D., Moyano, P.: Development of a computer vision system to measure the color of potato chips. Food Research International **39** (2006) 1092 – 1098
3. Wang, Q., Wang, H., Xie, L., Zhang, Q.: Outdoor color rating of sweet cherries using computer vision. Computers and Electronics in Agriculture **87** (2012) 113–120
4. Choi, K., Lee, G., Han, Y., Bunn, J.: Tomato maturity evaluation using color image analysis. Transactions of the ASAE **38** (1995) 171–176
5. Hanbury, A., Serra, J.: Mathematical morphology in the cielab space. Image Analysis and Stereology **21** (2011)
6. Schölkopf, B., Smola, A.J.: Learning with kernels. The MIT Press (2002)
7. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York, NY, USA (1995)
8. Gevers, T., Smeulders, A.W.: Color-based Object Recognition. Pattern Recognition **32** (1999) 453 – 464
9. Avila, F., Mora, M., Fredes, C., Gonzalez, P.: Shadow Detection in Complex Images Using Neural Networks: Application to Wine Grape Seed Segmentation. In: Adaptive and Natural Computing Algorithms. Volume 7824 of Lecture Notes in Computer Science. (2013) 495–503
10. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11** (1975) 23–27
11. Avila, F., Mora, M., Fredes, C.: A method to estimate grape phenolic maturity based on seed images. Computers and Electronics in Agriculture **101** (2014) 76 – 83