Vazquez-Fernandez E., Dacal-Nieto A., Martin F., Formella A., Torres-Guijarro S., Gonzalez-Jorge H. (2009) A Computer Vision System for Visual Grape Grading in Wine Cellars. In: Fritz M., Schiele B., Piater J.H. (eds) Computer Vision Systems. ICVS 2009. Lecture Notes in Computer Science, vol 5815. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04667-4_34

A Computer Vision System for Visual Grape Grading in Wine Cellars

Esteban Vazquez-Fernandez[1], Angel Dacal-Nieto[1], Fernando Martin[2], Arno Formella[3], Soledad Torres-Guijarro[1], and Higinio Gonzalez-Jorge[1]

[1] Laboratorio Oficial de Metroloxía de Galicia (LOMG), Parque Tecnolóxico de Galicia, San Cibrao das Viñas, 32901 Ourense, Spain,
evazquez@lomg.net
[2] Communications and Signal Theory Department, University of Vigo, Spain
[3] Computer Science Department, University of Vigo, Spain

Abstract. This communication describes a computer vision system for automatic visual inspection and classification of grapes in cooperative wine cellars. The system is intended to work outdoors, so robust algorithms for preprocessing and segmentation are implemented. Specific methods for illumination compensation have been developed. Gabor filtering has been used for segmentation. Several preliminary classification schemes, using artificial neural networks and Random Forest, have also been tested. The obtained results show the benefits of the system as a useful tool for classification and for objective price fixing.

1 Introduction

At present, grape grading systems at cooperative wine cellars combine several criteria. On the one hand, a class categorization is made by an expert according to a visual inspection of the grape container. This inspection focuses on maturation, presence of rotten or damaged grapes, etc. On the other hand, a sample of grape juice is taken to make a later laboratory study, which is mainly an enzymatic and microbial analysis.

The visual inspection is very important since it determines the separation into classes for the preparation of different quality wines. In addition, the price to pay to the cooperative members also depends on this classification, so it must be as objective as possible. The problem is the inherent subjectivity in classification made by human experts. There are too many factors that can influence the final result, like different technicians, fatigue, lighting conditions, inaccurate perception of proportions, etc. Under these conditions, a computer vision system has been developed to make the classification easier and more accurate.

During the last two decades, the use of computer vision systems in the food and agricultural industries has increased. Some recent reviews are available in [1], [2]. These systems offer many advantages for quality assurance and inspection tasks or grading. However, computer vision has never been used for the

inspection of received grapes in wine cellars (or, at least, the authors are unaware of this use), probably due to difficulties like outdoor conditions, variable illumination, etc.

In this paper, we present an incipient computer vision system for grape grading in cooperative wine cellars. The system is intended to work in the hard conditions of outdoor reception points, so the use of robust image processing techniques becomes a decisive factor. The achieved results suggest that this development can provide an objective and reliable tool for the correct classification and separation of different quality grapes. It also ensures a fair price for the cooperative partners.

The paper is organized as follows. Section 2 provides an overview of the image acquisition system. The preprocessing techniques and segmentation methods applied are presented in Sect. 3 and Sect. 4. Feature selection and classification are explained in Sect. 5. Finally, results and conclusions are drawn in Sect. 6 and Sect. 7.

2 Image Acquisition System

A schematic view of the reception point, where the system is intended to work, is shown in Fig. 1a. The region of interest (ROI) for the inspection is a $2\,\mathrm{m} \times 1.3\,\mathrm{m}$ surface of a container full of grapes (see Fig. 1b). The equipment used for the image capture system is selected according to this real situation, but can be easily adapted to a different reception point.

Fig. 1. Image acquisition system: a) Diagram of the grape reception point; b) Obtained image of the container surface

The camera used is a JAI BB-500 GE [3] with a 2/3" Bayer Color ICX625AQA 5 Mpixel sensor (2456×2048). An easy calculation can be made to choose a camera lens according to the field of view (FOV) and the required working distance [4] given by

$$\beta = \frac{Sw}{FOVw} \tag{1}$$

$$f' = \alpha \frac{\beta}{1-\beta} \tag{2}$$

where Sw is the sensor width in mm, $FOVw$ is the FOV width in mm and α is the distance between the surface to inspect and the lens. For our purposes a Schneider Optics Cinegon 1.4/8mm 2/3" [5] is adequate.

The main problem for the capture system is the illumination due to the outdoor placement and the wide period of operation which frequently goes from day to night. In addition, it is not possible to make an indoor reception point, due to the difficulties in manipulating the containers by a forklift truck. Under these conditions, the best results have been obtained with two high power halogen lamps (1000 W).

3 Preprocessing

Preprocessing commonly comprises a series of image operations to enhance the image before the segmentation process starts. In this application, preprocessing plays a fundamental role due to the difficult capture conditions of the images. The preprocessing stage can be divided in three sections: Camera calibration, continuous white balance and illumination uniformity compensation.

3.1 Camera calibration

A standard mapping operation is used to correct lens distortions [6]. For this purpose, the following parameters are needed:

$$\text{Intrinsec Matrix} \equiv \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & x_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$\text{Distortion Coeficients} \equiv (k_1, k_2, p_1, p_2, k_3) \tag{4}$$

which are used to model the focal, center, radial, and tangential distortions of the optical systems. These parameters are obtained offline by using image sequences of a chessboard pattern, stored and applied as proposed in [7].

3.2 Continuous White Balance

Illumination varies in a wide range during the grading session due to the outdoor placement. Although the illumination from the halogen lamps remains constant, the obtained images suffer a wide color variation due to the frequently and unpredictable changes in external light. In this situation, it is necessary to apply a white balance to every captured image.

The use of a fixed target for the white balance produces undesirable errors due to the variable positioning of the container, different light incidence between the grapes and the target, etc. To solve this problem, we decided to use the container itself, which is made of stainless steel, as a pattern. An algorithm to automatically detect the region to use as a gray reference has been developed.

To simplify the operations, only the lower left quarter of the image is used, where both grapes and container appear. To locate the steel gray, a texture filtering is applied. It consists on calculating the range value of a pixel $p(x, y)$ in a neighborhood V (empirically fixed to 13×13 pixels) as

$$p(x, y) = \max(V) - \min(V) \ . \tag{5}$$

Looking at the histogram of the filtered image (Figure 2), two clearly different regions can be seen. The first major peak matches the low textured regions (container walls, etc.), while the extended tail matches the highly textured areas (surface of grapes, edges, etc.). A sliding window method [8] based on derivative variations analysis is applied to locate a binarization threshold.

Fig. 2. Histogram of the range filtered image

After the binarization process, morphological operations of opening and closing and median filtering are applied to reduce noise. This process allows to label the image in a connected regions map. The selection of the region of interest for the white balance is established by features like area and centroid position. Then, the correction factors sr, sg, and sb, which will multiply the R, G and B planes of the entire image ($m \times n$ size), are obtained in the following way:

$$\bar{r} = \frac{1}{m \times n} \sum_{m,n} R \ , \ \bar{g} = \frac{1}{m \times n} \sum_{m,n} G \ , \ \bar{b} = \frac{1}{m \times n} \sum_{m,n} B \tag{6}$$

$$\bar{\imath} = 0.2125\bar{r} + 0.7154\bar{g} + 0.072\bar{b} \tag{7}$$

$$sr = \frac{\bar{\imath}}{\bar{r}} \ , \ sg = \frac{\bar{\imath}}{\bar{g}} \ , \ sb = \frac{\bar{\imath}}{\bar{b}} \ . \tag{8}$$

3.3 Illumination Compensation

Another problem in the obtained images is the non uniform illumination. Grapes fill randomly the container, creating different height zones, so light and shadow regions are produced, which need to be compensated. The proposed method is

intended to modify the V plane of the HSV color space, while the characteristics of color remain stable. A schematic of the process is shown in Fig. 3.

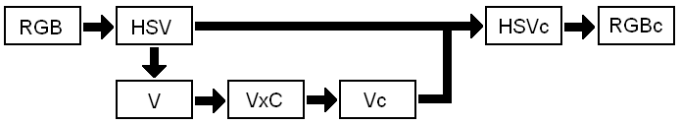

Fig. 3. Schematic view of the illumination compensation process

The most used color space in cameras, screens, etc. is RGB. It also models the way the eye reacts to different wavelength stimulus. However, the high correlation of the three components in RGB linear space makes the three components dependent upon each other and strongly associated with intensity [9]. These relations make it necessary to process the image in a three dimensional color space. To avoid these problems, it is interesting to make a non linear transformation to a HSV color space.

Most of the illumination compensation methods are based on the previous acquisition of the luminance pattern over the scene without the objects of interest [10]. However, we can not take advantage of this method for two reasons: On the one hand, the distribution of illumination depends directly of the objects (grapes) and their placement in the container. On the other hand, the luminance is directly related to the variable incidence of the external illumination all along the day.

To solve this problem, we have designed a method to perform the illumination compensation by using only the information present in the captured image. First, the illumination plane is modeled. An advantage of the grapes surface is their "homogeneity" and their conformation to the shape of the container (in this sense, they could be approached more like a liquid than like a solid). The luminance distribution is approximated by a Gaussian filtering (equations (9) and (10)) of the V plane (HSV space). A smoothing of the sharp texture of the grapes is achieved, while the influence of luminance in a wider area remains unaltered.

$$h_g(x,y) = e^{-(x,y)^2/2\sigma^2} \tag{9}$$

$$h(x,y) = \frac{h_g(x,y)}{\sum_x \sum_y h_g} \ . \tag{10}$$

To optimize the uniformity of the filtering, its size $n \times n$ should satisfy the following ratio to the standard deviation σ:

$$\frac{n}{2} \approx 3\sigma \Rightarrow \sigma \approx \frac{n}{6} \ . \tag{11}$$

The illumination compensation matrix, $C(x,y)$, is obtained as:

$$\begin{aligned} &C(x,y) = \tfrac{k}{L(x,y)} \\ &k = \alpha \bar{L} \\ &\alpha = 1.2 \end{aligned} \tag{12}$$

where $L(x,y)$ is the filtered image. The compensated V plane is obtained by:

$$V_c(x,y) = V(x,y)C(x,y) \ . \tag{13}$$

A detail of the achieved effect on the color image is shown in Fig. 4.

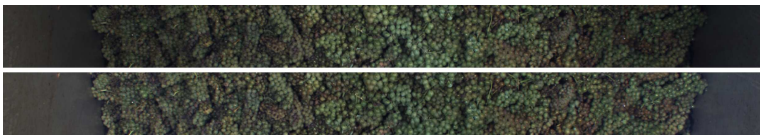

Fig. 4. Comparison detail between images before illumination compensation (above) and after illumination compensation (below)

4 Segmentation

The objective of the segmentation process is to separate the surface of the grapes from the rest of the elements in the image. A priori, a colour based segmentation method seems to be adequate for the purpose but, in practice, it turned out not to be robust enough. Variations in grape varieties, maturation or presence of foreign objects in or around the container makes colour an unreliable criteria, so a texture analysis based on Gabor filtering has been used for the segmentation purpose. A 2D Gabor filter can be thought of as a complex plane wave modulated by a 2D Gaussian envelope and can be expressed in the spatial domain as:

$$\begin{aligned} &G_{\theta,f,\sigma_1,\sigma_2}(x,y) = \exp\left[\tfrac{-1}{2}\left(\tfrac{x'^2}{\sigma_1^2} + \tfrac{y'^2}{\sigma_2^2}\right)\right]\cos\left(2\pi f x'\right) \\ &x' = x\sin\theta + y\cos\theta \\ &y' = x\cos\theta - y\sin\theta \end{aligned} \tag{14}$$

where f is the spatial frequency of the wave at an angle θ with the x axis and σ_1 and σ_2 are the standard deviations of the 2D Gaussian envelope. In most of the texture filtering applications, and also in this case, the Gaussian envelop is symmetric, so we have $\sigma_1 = \sigma_2 = \sigma$.

The segmentation process is based on the hypothesis that the grapes have the highest variation in texture. The texture operations will be applied to the V plane from the HSV image. The parameters for the Gabor filtering are set to the following values:

- Frequency has been set to a value of $f = 1/16$ pixels. The reason is because the value of the average grape diameter is 8 pixels $(1/2f)$.

– For the standard deviation, a value of $\sigma = \lambda/2 = 1/2\,f$ (measured in pixels) has been selected and tested empirically.
– Two orientations are needed and have been fixed to $\theta_1 = \pi/4$ and $\theta_2 = 3\pi/4$.

The grape texture is uniformly distributed in all orientations, while other highly textured elements, like container edges, are oriented in only one direction. Using two opposite orientations for the filtering and combining them reduces the effects of the edges in texture analysis. The best results have been achieved applying first a binarization operation to the magnitude of each filtered image separately and then combining them into one by a logical AND operation. The threshold for the binarization is obtained by a histogram analysis in a similar way to that explained in Section 3.2. The mask in the binary image is filled by mathematical morphology operations of opening and closing. Afterwards, the filled areas are labeled in a connected components map and the biggest one is selected for the mask. Figure 5 summarizes the process.

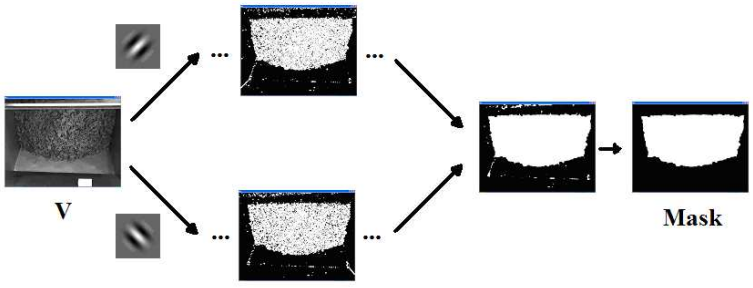

Fig. 5. Mask selection process

5 Feature Selection and Classification

To test the operation of the system, a preliminary classification scheme has been implemented. It is based on low level features. The method is inspired by the one presented in [11]. Jain and Karu suggested a method for learning texture discrimination masks by using neural networks. This approach is very similar to texture filtering approaches, where the filtering, nonlinearity, smoothing and classification are done all in one. However, this previous work is intended to classify only texture characteristics, so it uses only a 5×5 grey level neighborhood (25 features). In our system a more general classification is desired (e. g. also color is important), so the feature vector is formed by the 5×5 neighborhood values of the corrected planes H, S, V, R, G and B of the image (150 features).

The feature vector feeds the classifier, which is intended to learn higher level features from the input ones. The reason for using this scheme is "Black box" simplicity of implementation: the hard work of obtaining representative features

is done by the classifier. Different classification methods have been tested, including Multilayer Perceptron (MLP) and Random Forest classifiers [12] via OpenCV Libraries implementation [6]. For the MLP, several network architectures (number of hidden layers and nodes) have been tested. Divers Random Forest classifiers have been trained with different parameters to obtain several tree populations.

In this first implementation, the classifier is tested with Treixadura, a variety of white grapes grown in Galicia (Spain) and North of Portugal. For this purpose, six classes have been selected: green, ripen, rotten, container fragments, dry leaf and other external elements. The first three ones are used for grape classification, while the others are used to separate from non grading foreign elements.

6 Results and discussion

Representative regions for the different classes have been manually selected by human experts. This process has provided 140333 samples (5×5 pixel neighborhoods), which have been divided in a randomly selected 80 % for training (112266 samples) and 20 % for testing (28067 samples).

In the case of the MLP classifier, several network architectures have been trained to test their performance and to select the most adequate one. The training algorithm is backpropagation. Figure 6 shows the performance of one hidden layer and three hidden layer nets with variable number of nodes. The best results have been achieved by using a three hidden layer network. One can see this approach as a nonlinear generalization of principal components analysis (NLPCA) as proposed by [13], where the first and third hidden layers provide nonlinearity and the second one is the bottleneck. A compromise solution between success rate and computational cost has been achieved by fixing the final architecture to 18, 6 and 18 nodes in the three hidden layers respectively. This results in a 94.1 % and a 93.8 % success rate in the training and test stages. Table 1 shows the confusion matrix for the 18–6–18 three hidden layer MLP.

Table 1. Confusion matrix for the three hidden layer MLP (18, 6 and 18 nodes)

Classified→ / Real↓	Green	Ripen	Rotten	Container	External	Dry Leaf
Green	5450	15	16	113	0	0
Ripen	30	4826	202	16	0	149
Rotten	17	309	4079	118	6	6
Container	10	260	67	4300	1	6
External	0	0	1	2	5799	0
Dry Leaf	16	378	8	0	0	1866

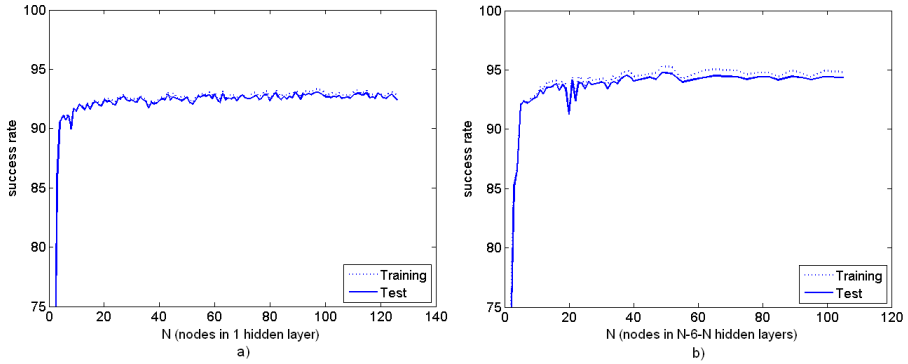

Fig. 6. MLP Network architecture comparison a) MLP one hidden layer net (x–axis represents number of nodes); b) MLP 3 hidden layer net (x–axis represent the first and third hidden layer number of nodes)

For the Random Forest classifier, different values of the maximum depth of the trees have been tested. This provides different forest structures and number of trees, since they are pruned at different depths. Figure 7 shows the evolution of the test through the max depth variation, for different number of variables (Nvar), which are randomly selected at the nodes (and that are used to find the best split). One can see, as the depth level of the trees is increased, an overfitting like effect appears (obtaining much higher rate in the training than in the test).

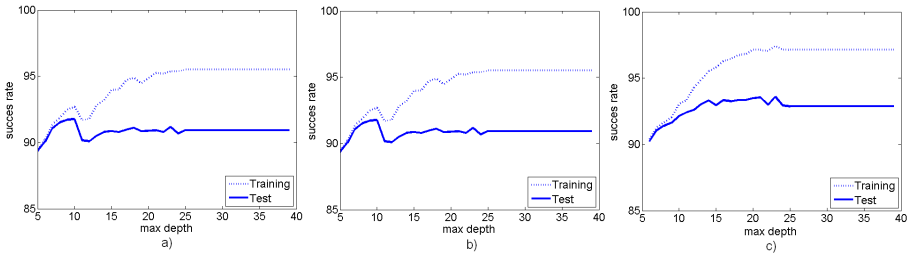

Fig. 7. Random Forest comparison: a) $N\text{var} = 12$; b) $N\text{var} = 6$; c) $N\text{var} = 2$

The obtained results show the feasibility of the system, achieving a success rate over 90 % in all tests. These results lead us to formulate future lines to continue this development:

- Developing specific feature extraction methods to improve the classification.
- Testing the system on different varieties of grapes.
- Extending the classification to other classes (different diseases, abnormal growth, etc.) for obtaining better information for the wine cellar classification, etc.

7 Conclusions

A promising computer vision system for grape grading in cooperative wine cellars has been developed. The system allows grape sorting to produce different quality wines. It also seems to be a useful and objective method to fix an equalitarian price to the cooperative partners.

Working in outdoor environments is a challenge for computer vision systems. Robust preprocessing techniques need to be applied. A method for illumination compensation, which only uses the available information in the captured image, has been implemented.

Texture Gabor filtering has been applied to segmentation. For feature extraction and classification, an all-in-one approach has been presented. It has been tested by using MLP and Random Forest classifiers, which are intended to learn higher level features from low level ones, obtaining a success rate of 94 %.

Acknowledgements This work was partially supported by "Xunta de Galicia" (code project FEADER2008-12).

References

1. Brosnan, T., Sun, D.W.: Improving Quality Inspection of Food Products by Computer Vision - A Review. Journal of Food Engineering 61, 3–16 (2004)
2. Du, D.W., Sun, D.W.: Learning Techniques Used in Computer Vision for Food Quality evaluation: A Review. Journal of Food Engineering 72, 39–55 (2006)
3. JAI - Industrial CCD/CMOS cameras, http://www.jai.com
4. Hornberg, A.: Handbook of Machine Vision. Wiley VCH, Weinheim (2006)
5. Schneider Optics, http://www.schneideroptics.com
6. Bradski, G., Kaehler, A.: Learning OpenCV. O'Reilly Media, Sebastopol (2008)
7. Zhang, Z.: A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 1330–1334 (2000)
8. Martin, F.: Analysis Tools for Gray Level Histograms. Proc. of SPPRA-2003, 11–16 (June 2003)
9. Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J.: Color image segmentation: advances and prospects. Pattern Recognition 34, 2259–2281 (2001)
10. Russ, J.C.: The Image Processing Handbook, Fifth Edition. CRC Press, Boca Raton (2007)
11. Jain, A.K., Karu, K.: Learning Texture Discrimination Masks. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 195–205 (1994)
12. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001)
13. Kerschen, G., Golinval, J.C.: Non-linear Generalization of Principal Component Analysis: From a Global to a Local Approach. Journal of Sound and Vibration 254(5), 867–876 (2002)