Convex Objects Recognition and Classification Using Spectral and Morphological Descriptors

Steven Le Moan*[1,3]*, Alamin Mansouri*[1]*, Tadeusz Sliwa*[1]*, Madaín Pérez Patricio*[2]*, Yvon Voisin*[1]*, Jon Y. Hardeberg*[3]*;
1 : ***Le2i - Université de Bourgogne, Auxerre, France***
2 : ***Instituto Tecnológico de Tuxtla Gutiérrez, Mexico***
3 : ***ColorLab, Gjøvik University College, Gjøvik, Norway***

Abstract

In this paper, a new approach for the recognition and classification of convex objects in color images is presented. It is based on a collaboration between color quantization, mathematical morphology and reflectance estimation from RGB data. This yields a robust algorithm regarding the conditions of illumination, the color sensor used for acquisition, as well as the shape/overlapping ambiguities among the objects. One singularity of this work is the use of mathematical morphology in two distinct topologies: first in the entire image, for segmentation purposes, then locally, to enhance the classification of each object. A resolution reduction is used to alleviate the effect of local disturbances such as noise or natural impurities on the objects. The method's efficiency and usefulness are illustrated on the particular task of coffee beans sorting.

Introduction

Classification of convex objects in color images is an important problem in many computer vision applications. In the food industry for instance, automatic inspection of products in order to satisfy quality specifications is a common task. A large amount of methods have been developed since the late 90s [1], involving various products such as fruits, vegetables, baked goods, meat, prepared consumer food, etc. In many cases, the shape of the objects, when projected on the image plane provides a useful *a priori* knowledge that is to be properly exploited in the different processing steps. In [2], the author proposes a general model for image analysis. From the problem domain to the results, three main levels of analysis can be highlighted: the low, intermediate and high levels. Low-level analysis include image acquisition and pre-processing, which purpose is to enhance the image in the best possible way according to the application. Intermediate-level analysis involve image segmentation, representation and description. Segmentation is known to be a critical analysis in the overall process since all the following steps are highly dependent on it. It consists in the identification of regions/samples of interest, in order to create a local topology. Representation and description of the segmented image intend to extract meaningful and robust features from the regions, in order to describe them in the best way possible, once again in accordance with the application. Finally, high-level analysis include object recognition and image interpretation. These last steps are generally combined in the classification process. This generic sequence leads to a very large variety of possible methods.

Recently, a great deal of research effort has been directed toward image analysis in both color and spatial domains, for segmentation. In [3] for instance, morphological processings are used for background extraction and color information is used for the neural-network-based classifier. In [4], spatial information is injected in the conventional grey-level histogram in order to create a so-called *compacigram*, in which identification of modes and, hence, segmentation, are eased. These methods have an important drawback which is their dependence on the conditions of acquisition of the image (viewpoint, illumination, sensor). This is the reason why it is important to extract robust features from the image. In [5, 6] for instance, schemes are proposed for object recognition regardless of viewpoint and illumination. Color- and shape-invariant descriptors are used to form robust image features. However, tristimulus-based color (RGB) remains a weak descriptor of the scene and thus, can mislead the segmentation. Indeed, depending on the illumination and the sensor used for acquisition, two objects from two different given classes can have very close RGB representations. However, because reflectance acquisition is usually a very expensive solution, methods for reflectance estimation from RGB pixels have been proposed [7, 8, 9, 10].

The new approach proposed in this document involves a collaboration between color quantization, mathematical morphology and reflectance estimation from RGB data. An originality of it is that two distinct topologies are considered. First, low-, intermediate- and high-level analysis are computed on the overall image, then, remaining classification ambiguities are solved by a local use of mathematical morphology, in the classes' space. A resolution reduction allows to alleviate the effect of local disturbances such as noise or natural impurities on the objects. Estimation of the reflectance is made by means of a learning dataset built upon preliminary spectroscopic acquisitions in the studied scene. An adaptive Principal Component Analysis (PCA) is used to refine the estimation. The classification consists in minimizing the Goodness-of-Fit Coefficient computed with each element from the learning set. The overall procedure can be summarized by the following steps: (1) pre-processing, (2) overall segmentation, (3) resolution reduction, (4) classification and (5) local segmentation. The method is illustrated throughout this article on the particular application of coffee bean sorting.

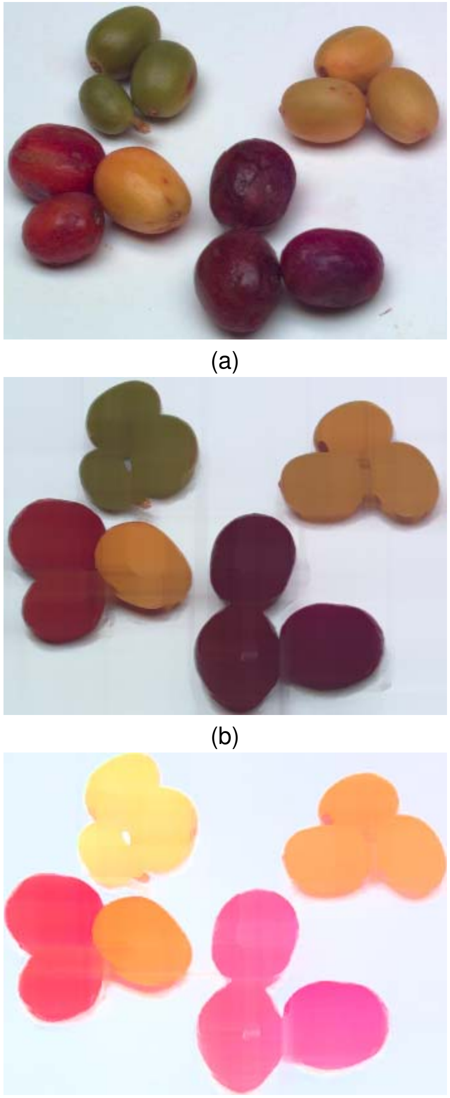
The remainder of this document is organized as follows. Section 2 describes the pre-processings used. Section 3 presents the entire segmentation procedure, in the overall topology. Section 4 explains the extraction of features, while classification and local segmentation are described in Section 5. Eventually, results are presented and discussed in Section 6 before concluding in Section 7.

Pre-processings

Throughout this paper, the image shown in fig 1a will be used as an illustrative application. It represents coffee beans with different levels of maturity. Four classes are clearly identifiable: the green, yellow, red and purple ones. This image has been chosen as it contains all the classes of beans, and includes some heterogenous overlappings. The application has been proposed by the university of Chiapas, Mexico, in the context of the conception of an automated bean sorting machine. The images shown here and in the results section are preliminary acquisitions. However, for the final conditions of acquisition (illumination, camera used) are yet unknown, we propose to keep this description general, which allows for a certain freedom in the possible future applications, whether they are in or outside the scope of the food industry.

First, an anisotropic diffusion [11] allows the elimination of local highlights, but also homogenizes the hue inside each object. Anisotropic diffusion is a specularity removal filtering that homogenizes the texture inside edges so that small shiny areas are removed and the colors are diffused. Contrary to other specularity filters [12, 13], anisotropic diffusion does not require any *a priori* knowledge about the illuminant and has a very low computational burden. The result is shown in Figure 1b.

In addition, we performed an intensity maximisation in the HSV color space (figure 1c). Indeed, the V component of the HSV colorspace contains the data intensity, while the two other components measure its chromacity (hue and saturation). Thus, specularities and shadows are mainly present in V and maximizing it allows to homogenize the intensity in the image and then to minimize the impact of these undesirable effects.



(a) (b) (c)

Figure 1. Pre-processings (a) Original image (b) Result of an anisotropic diffusion (c) Result of a V-component maximization

Segmentation

This step aims to separate the convex objects in the overall image, in the best possible manner. Mathematical morphology [14, 15] is used for for shape analysis. In order to stay in a simple framework, only binary morphology is considered. In this purpose, a background extraction is computed on the pre-processed image, by means of a minimum variance algorithm [16], as shown in Figure 2a. In order to separate the objects, while taking into account overlapping and contact between them, a pertinent morphological opening is required, and preferred to a closing, which carries more favor to the grouping of regions. The adequate structuring element's properties are set by the *a priori* information, which, in our case, are:

- *Convexity*, which allows to choose a linear structuring element $s_{\vartheta,l}$ for a better handling of the cavities separating the objects.
- *Minimal and maximal sizes* of a bean, determined according to the acquisition device properties (resolution, focal distance, etc.), allows to set the size of $s_{\vartheta,l}$. Of course, this assumes that the shapes under study are of the same scale.

Since we know that the final objects are convex and in finite number, the image present some directions that favorise the objects separation. Therefore, and in order to take advantage of this, a series of morphological openings is used to filter the shapes, by means of using $s_{\vartheta,l}$ with variable orientation ϑ. Its size is set according to the minimal size of an object. These directional openings have been combined by means of a pixel-by-pixel product (equivalent to an AND operation) as described by equation 1, in order to take into account all directions together.

$$I_l = \Pi_{\vartheta \in [0;179]}(I \ominus (s_{\vartheta,l}) \oplus (s_{\vartheta,l})) \quad (1)$$

where $\ominus$ and $\oplus$ respectively represent morphological erosion and dilation. After that, a removal step consists in eliminating small components that contain less pixels than a pre-fixed threshold. The latter depends obviously on the application, and is set according to the minimum size of a bean. The result is shown in figure 2c.

Using the same method as the one used for binarization, an L-level quantization is also computed (L being the number of classes of beans plus one corresponding to the background). This analysis allows the identification of overlapping beans of different classes (Figure 3).

From the morphologically treated binary image in Figure 2c, only connected components with a size superior to an acceptable threshold are kept for further treatments. The threshold being, in our case, the maximal size

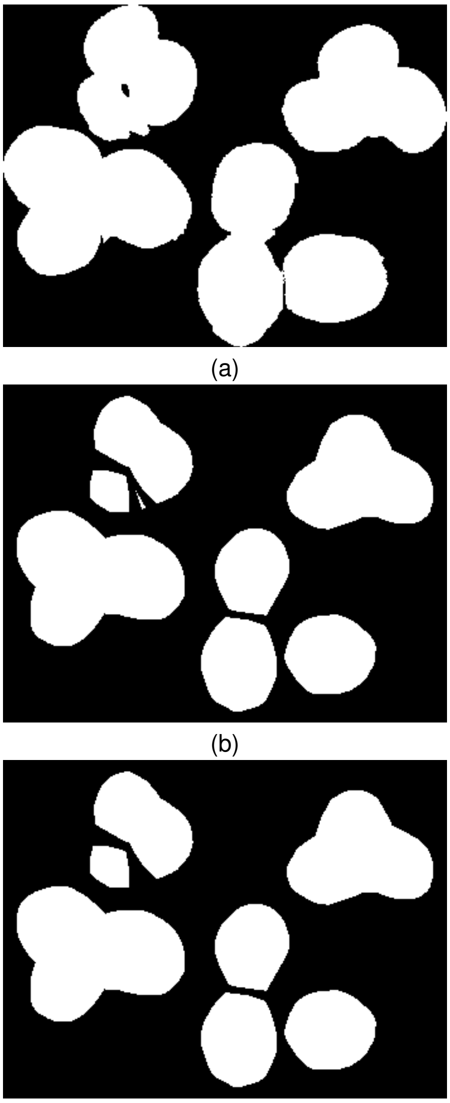


Figure 2. Morphological processings (a) Binarized image (b) Result of the minimal directional opening (c) Elimination of small regions

of a bean. Figure 4a shows an example. Then, a fusion step consists in merging the morphologically treated large region with its corresponding 5-levels quantization (Figure 4b). This merging is achieved by a simple product. Each class is then isolated, by considering all the others as background, in order to compute as many minima of directional openings as there are classes. Non-significative classes (according to the size of the structuring element) will yield an empty image, whereas preponderant ones will allow for a better separation of objects, when the results are merged, as shown in Figure 4c.

After repeating this procedure on each large region, another removal is achieved in order to erase small regions, resulting in the image in Figure 5.

This analysis does not favor the separation of beans of a same class, hence some overlapping remains. This is not assumed to be a problem, since the important task is to identify objects according to their class.

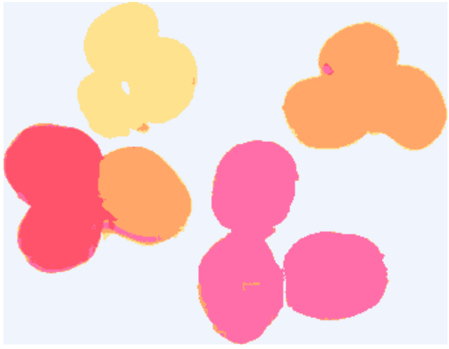


Figure 3. Five-level quantization

Figure 4. Local topology (a) Connected component which size is bigger than the maximal size of a bean (b) Same component with 5-levels quantification (c) Local morphological processing

Extraction of features

Once the objects have been separated, the next step consists in characterizing them, through the extraction of meaningful features. In this paper, RGB information is considered non relevant because metamerism can mislead the color representation. Spectral reflectance then is considered since it is an invariant descriptor of the surface. In order to acquire reflectance, devices such as a multispectral camera or a spectrophotometer are typically used. The problem with the latter is that it achieves only pointwise measurements. Multispectral cameras are appropriate but they are still complex and too costly for many applications, where color cameras are still widely used. Therefore, finding appropriate mathematical methods to estimate the spectral reflectance from color camera output is of great importance as it is crucial for the success of many tasks as classification, segmentation, etc. In [7], an adaptive PCA algorithm was proposed at this aim, which is described now.

First, it is assumed that the reflectance R associated with a pixel can be estimated by a linear combination of a small number k of known basis functions, regrouped in a matrix A:

$$\hat{R} = B.A \tag{2}$$

With B representing the linear combination matrix.

As said before, a set of preliminary acquisitions of the reflectances R_{learn} of the beans is available (6 by class, 4 classes: 24 acquisitions), they are constituted of 31 bands, resulting from a 10nm sampling in the visible range of wavelengths [400-700]nm. These data are matched with randomly chosen RGB data in each class, in order to the pairs reflectance/tristimulus (R_p, T_p) with p representing the association with a given pixel. From this, the following dimensionality reduction mapping is estimated:

$$T_{learn} = H.R_{learn} \tag{3}$$

Where H represents the transfer function corresponding to the conditions of acquisitions and from which data

Figure 5. Result of directional opening minima in restrained topology, after new removal of small regions

must be separated. It contains a product of the illuminant and the sensitivity of each of the three channels of the sensor (R,G,B). The mapping can then be extended to any pair (R_p, T_p), and by using equation 2, it comes that:

$$T_p = H.B_p.A \tag{4}$$

where $H.B_p$ is a 3×3 invertible matrix, so that:

$$A = (H.B_p)^{-1}.T_p \tag{5}$$

and since $R = BA$, the reflectance of each pixel of the image is given by the following equation:

$$\hat{R}_p = B_p.(H.B_p)^{-1}.T_p \tag{6}$$

This represents the model used for reflectance estimation. However, for the reconstruction of a particular spectrum with a particular shape, all the elements of the training set are neither necessary nor appropriate. From this observation, it is proposed to adapt the training-set to each estimated spectrum. To do so, the PCA basis is derived in two steps: firstly, it is derived from the whole training set and a first estimation of reflectance from each tri-stimulus is computed using this basis. Then, the likelihood between the estimated reflectance and each element of the training-set is measured. A new training set is built by keeping only those elements whose similarity falls in the range [95%-100%]. Then, a new PCA analysis is performed from this new set to derive a new basis consisting of new three components with which is performed the final reflectance estimation. The likelihood is calculated using the non centered correlation coefficient, largely used and known in the community as Goodness-of-Fit Coefficient (GFC) expressed by the formula:

$$GFC(R_p, R_{est}) = \frac{\left|\sum_j R_p(\lambda_j) R_{est}(\lambda_j)\right|}{\left(\left|\sum_j [R_p(\lambda_j)]^2\right| \left|\sum_j [R_{est}(\lambda_j)]^2\right|\right)^{1/2}} \tag{7}$$

where $R_p(\lambda_j)$ is the value measured by the spectrophotometer at the wavelength λ_j, and $R_{est}(\lambda_j)$ represents the estimated value at the same wavelength. The algorithm is presented below:

This adaptive approach is preferred instead of others, based for instance on modifications of the Wiener method [9] because of its simplicity, its apparent efficiency, and

Algorithm 1 Adaptive PCA

Derive a basis B_1 from the training set
for $i = 1$ to 24 **do**
  Estimate an intermediate $R_i^{inter} = B_1(HB_1)^{-1}T_i$
  Compute GFC between R_i^{inter} and each R_p
  $G = \max_p(GFC(R_i^{inter}, R_p))$
  $New\_R_p = \{R_p \mid 0.95 * G \leq GFC(R_i^{inter}, R_p) \leq G\}$
  Derive a basis B_2 from the training set New_R_p
  Estimate the final $R_i = B_2(HB_2)^{-1}T_i$
end for

the absence of an artificial data weighting. The latter requires indeed several hypotheses: statistical models, bias coefficients, comparison metrics, etc., whereas the choice of the likelihood threshold in the proposed method is the only arbitrary element and is very general.

Instead of computing reflectance reconstruction for each pixel on the image, a local resolution reduction is performed for two reasons:

- First, local disturbances such as noise or natural impurities can mislead the classification.
- Second, the reflectance estimation method has a consequent computational load, and reducing the number of samples makes it more efficient.

The reduction is performed by means of a regular square grid adapted to each shape (figure 6).

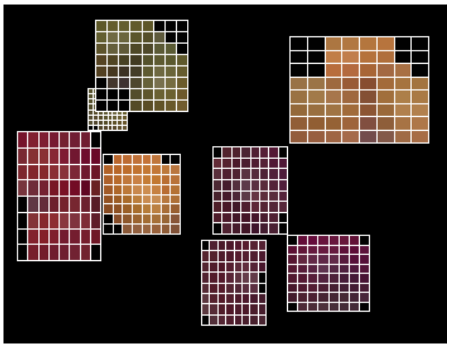


Figure 6. Local resolution reduction

The central limit theorem [17] states that a group of random variables from similar but not necessarily completely identical random distributions can be said to be realizations of the same law, and the accuracy of this approximation grows with the size n of the group. However, the larger n is, the larger the inhomogeneity of the pixels may be. Hence the necessity of finding a good compromise for the size of the grid. For the considered application, it has been determined that an 8×8 grid provides the best results. Values of pixels in each element of the grid are then averaged. Background pixels are neglected during the process.

Classification and local segmentation

The classification is supervised and simply achieved by comparing each feature's reflectance with the known reflectances of the classes of the learning set. The Goodness-of-Fit Coefficient is used to measure the similarity. The features are labeled with the class with which they maximize the similarity. The result is shown in Figure 7, where each gray level represents a class.

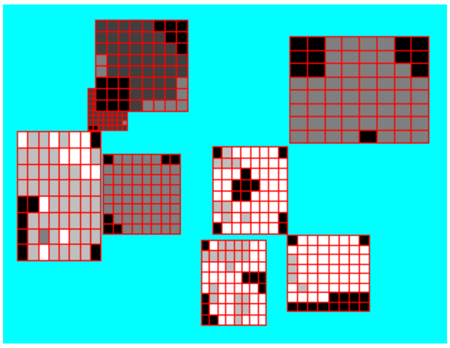


Figure 7. Classification of features

In the case of an ambiguity inside a region, that is, when there are more than one significant class, morphological operators are used once again in order to help making a final decision. A class is assumed to be significant when it represents more than 30% of the region area. X_{ij} is the current classification result of the feature at the coordinates (i, j), which is considered possibly erroneous (mostly because of the central limit theorem approximation), and let S_{ij} be the correct classification result. Let H_0 be the hypothesis that, due to the efficiency of the previous treatments, the number of beans inside a region is low compared to its number of features. Under H_0, it seems reasonable to assume that the accuracy of classification of a feature is better if it matches the classification of its neighbors. In other words, isolated classes in the region have a large probability of being erroneous. The application of a morphological closing on a binary image tends to eliminate the isolated parts inside the image, whereas a morphological opening eliminates isolated parts in the background, thus, by noting U the ensemble of centered strutural elements not reduced to a point, one can assume that:

$$\exists v \in U / Pr\left(((X \bullet v) \circ v)_{ij} = S_{ij}\right) \geq Pr\left(X_{ij} = S_{ij}\right), \quad (8)$$

where $\bullet$ and $\circ$ respectively represent morphological closing and opening. Basically, equation 8 means that the probability of a correct classification is increased after the application of a morphological closing followed by an opening. This treatment is applied on each dominant class of each region using an isotropic structuring element of a diameter $d > 1$. Moreover, in order to retrieve the original resolution of the image, a resolution increase is performed in each region, by interpolation. Figure 8 shows the result on the fifth region (regions are numbered from left to right, starting at number 1), in which there is an ambiguity between two dominant classes (in white and gray).

The final decision is then achieved by retaining the new dominant class. Figure 9 shows the final result on the entire image, where all the beans (or groups of beans) are accurately recognized and classified. The number in red marks the region where a second set of morphological operations was performed.

Results and discussion

As illustrated with the previous and following examples (Figure 10), the proposed method performs well on different images, with different arrangements, and from

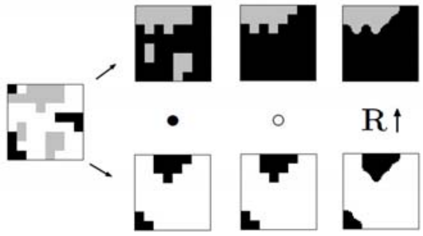


Figure 8. The 3 steps of the local processing on region number 5: closing, opening and resolution increase

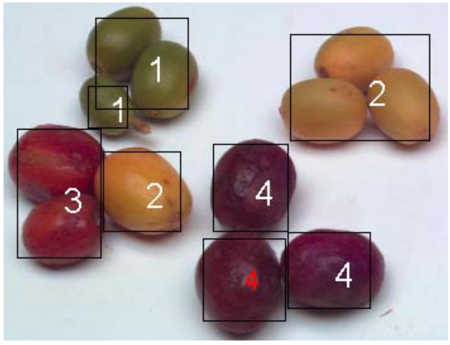


Figure 9. Final classification result

different acquisition devices. Even blurry images yield good classification results, which makes the algorithm adaptive to many situations. Some minor drawbacks are nevertheless apparent, and it is worthwhile to explain and discuss them to allow for further improvements:

- In the first image below, two objects are detected in one, this is because the bean includes a large zone of specularities which is interpreted by the algorithm as a separation between two beans. This undesirable effect should certainly be avoided by a better pre-processing. Since no false result is yielded, only redundancy, this is considered as a minor problem.
- In several cases, the position of the circumscribed rectangles is not apparently accurate. According to the different steps of the algorithm, these rectangles actually happen to contain the regions in the beans that are the most homogenous. A possible improvement would consist in a better fitting of these rectangles with the beans, in accordance with the application. Here again, this drawback is acceptable since the beans are always contained in a reasonable proportion.
- As stated earlier, overlappings are not always well handled, but this concerns only beans of a same class, which makes it a minor problem. However, the resolution of it would be a real improvement of the method. A non-generic solution could consist in finding a gravity center G_r of the region r and computing the points M_{rx} of the contour that are the closest to G_r, by detecting minimas of a contour function. Then, all $[M_{rx}G_r]$ segments would be drawn, allowing a good separation of objects, but only under the condition that the region has certain shape properties.

Although only images of coffee beans have been considered here, it is obvious that many other applications

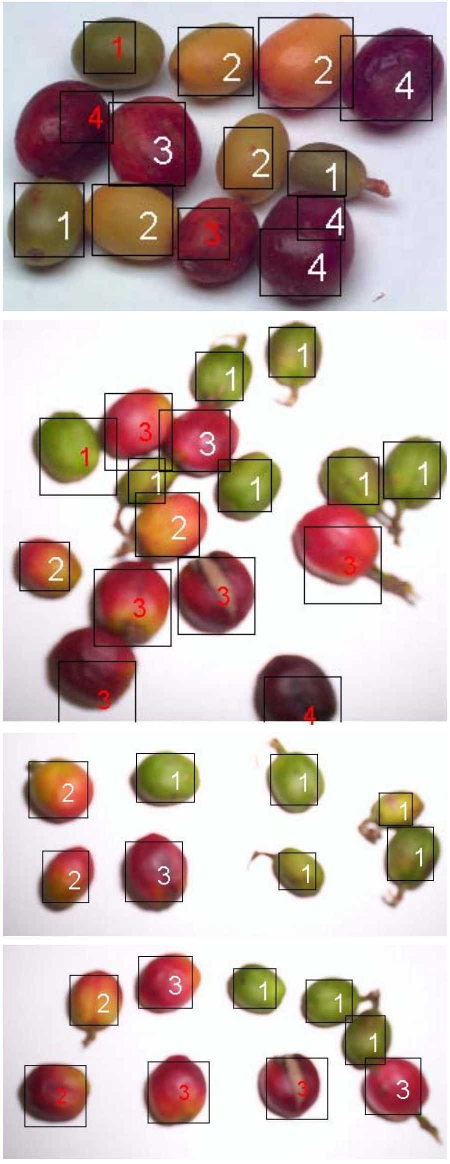


Figure 10. *Other results*

are possible for this method. Food-related applications are probably the most relevant, since the algorithm has been shown to be efficient on objects containing visible imperfections, which is the case of many fruits and vegetables (spots, zones of immaturity or decay, etc..).

Acknowledgements

The regional council of Burgundy supported this work.

Conclusions

In this paper, a new algorithm for recognition and classification of convex objects in color images has been presented. Its originality resides in the collaboration of color and morphological information, as well as in the use of a reflectance estimation method, in order to reduce the influence of the acquisition conditions. The use of *a priori* information about size, shape and convexity of the objects to be classified makes our work very adaptable, which allows it to be applied to other areas such as grapes, nuts, olives, or any other kind of convex-shaped object sorting. Results on several color images of coffee beans have been presented, where the recognized beans are classified with high accuracy. Although only one particular task has been used for illustration, it is obvious that many other applications are possible, the food-related ones being probably the most relevant considering the fluctuations of shape and color that have to be dealt with in this particular area. Further investigations will include a better separation of objects, by means of an angular point detection on the edge of the shapes.

References

[1] T. Brosnan and D.W. Sun. Improving quality inspection of food products by computer vision - a review. *Journal of Food Engineering*, 61(1):3–16, 2004.

[2] D.W. Sun. Inspecting pizza topping percentage and distribution by a computer vision method. *Journal of food engineering*, 44(4):245–249, 2000.

[3] K. Kılıç, İ.H. Boyacı, H. Köksel, and İ. Küsmenoğlu. A classification system for beans using computer vision system and artificial neural networks. *Journal of Food Engineering*, 78(3):897–904, 2007.

[4] C. Botte-Lecocq, O. Losson, and L. Macaire. Color image segmentation by compacigram analysis. In *Proceedings of the 14th International Conference of Image Analysis and Processing-Workshops*, pages 212–215. IEEE Computer Society Washington, DC, USA, 2007.

[5] T. Gevers and A.W.M. Smeulders. Image indexing using composite color and shape invariant features. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 576–581. Citeseer, 1998.

[6] A. Diplaros, T. Gevers, and I. Patras. Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing*, 15(1):1–11, 2006.

[7] A. Mansouri, T. Sliwa, J.Y. Hardeberg, and Y. Voisin. An adaptive-pca algorithm for reflectance estimation from color images. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.

[8] V. Heikkinen, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, T. Jaaskelainen, and S.D. Lee. Regularized learning framework in the estimation of reflectance spectra from camera responses. *Journal of the Optical Society of America A*, 24(9):2673–2683, 2007.

[9] P. Stigell, K. Miyata, and M. Hauta-Kasari. Wiener estimation method in estimating of spectral reflectance from rgb images. *Pattern Recognition and Image Analysis*, 17(2):233–242, 2007.

[10] N. Shimano. Recovery of spectral reflectances of objects being imaged without prior knowledge. *IEEE Transactions on Image Processing*, 15(7):1848, 2006.

[11] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern Analysis and machine intelligence*, 12(7):629–639, 1990.

[12] K.J. Yoon, Y. Choi, and I.S. Kweon. Fast separation of reflection components using a specularity-invariant image representation. In *International Conference on Image Processing (ICIP)*, pages 973–976. Citeseer, 1999.

[13] S.P. Mallick, T. Zickler, P.N. Belhumeur, and D.J. Kriegman. Specularity removal in images and videos: A pde approach. *Lecture Notes in Computer Science*, 3951:550, 2006.

[14] J. Serra. Advances in mathematical morphology. *Signal*

processing, 16(4), 1989.
[15] J. Serra and P. Soille. Mathematical morphology and its applications to signal processing. *Computational imaging and vision series, Kulwer, Dordrecht, 1994, 383p.*, 1993.
[16] X. Wu. Efficient statistical computations for optimal color quantization. *GRAPHICS GEMS II*, pages 126–133, 1991.
[17] J. Rice. *Mathematical Statistics and Data Analysis (Second ed.).* Duxbury Press, 1995.

Author Biography

Steven Le Moan *is a PhD student at the Le2i, Université de Bourgogne, France. He received his Master's degree in Signal Processing and Embedded Systems from the University of Rennes 1 (2009). He is currently working in cooperation with the ColorLab in Gjøvik, Norway in the fields of multivariate analysis of multispectral images for visualization. His technical interests extends also to the study of 3D meshes with multispectral attributes.*

Alamin Mansouri *received his PhD degree in computer vision (particularily in multispectral imaging) from the University of Burgundy, France, in december 2005. Since september 2006 he is associate professor at the same university. His research is focused on color/multispectral image processing and analysis and their applications.*

Tadeusz Sliwa *received his Ph.D. degree in 2003 from the University of Burgundy, France. Since 2004, he has been an associate professor at the University of Burgundy, France. He is also a member of the Image Processing group of the Laboratory Le2i. His research interests are in the applications of image processing.*

Madaín Pérez-Patricio *received the Diploma in electronic engineering from the Instituto Tecnológico from Tuxtla Gutiérrez in 1994 and the PhD degree in automatic control and computer engineering from the Université Lille 1, France in 2005. He is currently a professor of computer science in the Department of Computer Science of the Instituto Tecnológico from Tuxtla Gutiérrez. The topics he has worked on include pasive and active stereovision and multispectral imagery. He is a member of the IEEE Society.*

Yvon Voisin *is a full professor of signal and image processing at the University of Burgundy and is a member of the Image Processing Group at the Le2i. His research interests are 3D reconstruction and motion analysis. He's also working on the application of artificial vision, especially in biology. Voisin has a PhD in electronic and signal processing from the University of Franche Comté, France.*

Jon Y. Hardeberg *is Professor of Color Imaging at Gjøvik University College. He received his Ph.D from Ecole Nationale Supérieure des Télécommunications in Paris, France in 1999 with a dissertation on colour image acquisition and reproduction, using both colorimetric and multispectral approaches. He has more than 10 years experience with industrial and academic colour imaging research and development, and has co-authored over 100 research papers within the field. His research interests include various topics of colour imaging science and technology, such as device characterisation, gamut visualisation and mapping, image quality, and multispectral image acquisition and reproduction. He is a member of IS&T, SPIE, and the Norwegian representative to CIE Division 8. He has been with Gjøvik University College since 2001 and is currently head of the Norwegian Color Research Laboratory.*