COMPARISON OF TWO UNSUPERVISED METHODS OF CLASSIFICATION FOR SEGMENTING MULTI-SPECTRAL IMAGES

Danielle Nuzillard[†] *and Cosmin Lazar* [‡]

CReSTIC, University of Reims Champagne-Ardenne,
[†]UFR Sciences, Moulin de la Housse, 51687 Reims cedex 2, France
[‡]IFTS, 7 Bd Jean Delautre, 08000 Charleville-Mézières, France
danielle.nuzillard@univ-reims.fr, vasil-cosmin.lazar@etudiant.univ-reims.fr

ABSTRACT

Some clustering algorithms require assumptions (such as number and shape of classes), which limit their performances or provide wrong results. On the contrary, methods based on the estimation of the probability density function (*pdf*) do not make any assumption neither on the classes shape nor on their number. Two methods based on the *pdf*, are explored and applied to the segmentation of a multi-spectral image of a cereal grain. The first one is inspired from the estimation of the *pdf* Parzen-Rosenblatt and the second one estimates the support of the *pdf* through the support vector theory.

1. INTRODUCTION.

The knowledge of biological structures in plants is a very important task to valorize agronomical ressources. Our work is in keeping with the identification and the repartition of various tissues in cereals. Actually cereal grains are constituted by tissues which are superimposed. The central part contains starch and the external layers serve as protection. Those contain pure natural fluorescent components : cutin, ferulic acid, lignin whose signatures recover partially. Confocal laser microspectrofluorometry enables to visualize autofluorescence of tissues. Many laser excitations associated with a set of various filters in reception allow to get pseudo multi-spectral images.

In a former work, techniques of independent component analysis have been applied to isolate pure chemical compounds and their proportion map from a multispectral image of a barley grain [1], [2]. In [3], the relation between Blind Source Separation methods and classification has been examined through an application in teledetection for which a ground analysis was given. Now, in the framework of the study of the barley grain, we compare two methods of pixel classification which do not make any assumption neither on the shape of classes nor on their number.

Each tissue enclosed in the studied vegetal has a homogeneous chemical composition whose corresponding pixels are gathered in the same class, each being characterized by its concentration map. There are several ways to go from the set of elementary maps to the unique map. A first group of methods is interactive and called interactive correlation partitioning. It consists in selecting the different classes of pixels interactively, within the two- or three-dimensional scatterplot, and back-mapping the selected areas into the real image space. The second group of methods is automatic and sometimes is called automatic correlation partitioning. It consists in automatically grouping similar pixels (that possess similar elementary composition) into one class, the number of classes being *a priori* unknown. This process is known as clustering in artificial intelligence and data analysis communities.

This paper is organised as follows: in the next two sections we will present a briefly overview of the Parzen-watersheds and Support Vector Classification algorithms. Both methods applied to artificial data sets and to a multispectral image of a cereal will be presented in section 4. Finally we will discuss on the results and draw some perspectives of research.

2. THE PARZEN-WATERSHEDS ALGORITHM

Parzen-watershed clustering algorithm is an unsupervised data classification technique which does not make any assumption concerning the shape nor the number of classes [4], [5], [6]. The first step of this approach consists in estimating the *pdf* $p(x_i)$ of the whole data set in the feature space. This can be done according to the Parzen method [7] for which the point distribution (one object corresponds to one point x_i in feature space) is transformed into a quasi-continuous distribution $p(x_i)$ through a convolution by a smoothing kernel:

$$p(x_i) = \lambda \sum_{k=0}^{N} K(\frac{x_i - y_k}{h}) \qquad (1)$$

where K is a smoothing kernel of size h and λ is a normalizing factor. The estimated *pdf* $p(x_i)$ is generally characterized by several modes, or local maxima, separated by valleys or local minima. Each mode is considered to match with a class of objects.

Thus, next step consists in estimating the position of the boundaries between the different classes in the feature space. Although this task is trivial for a one dimensional feature space, it is much complicated for an n dimensional feature space, ($n > 1$). To split the feature space, the watershed function issued from mathematical morphology [8] or the SKIZ (SKeleton by Influence Zones) method [9] can be used.

We retained the SKeleton by Influence Zones procedure. This procedure consists in detecting firstly the connected components using an iterative thresholding of the estimated *pdf*. The connected components are defined as one-piece subsets partitioning the whole data space. For a two dimensional data space, we use the eight-connected neighbours in order to define these subsets. These connected components will be used as seeds in order to determinate the influence zones. Each point in the feature space will belong to the influence zone of that connected component which is the closest to the current point. The number of influence zones will be equal to the number of the connected components.

Then, last step consists in returning to the real image space. This can be done easily because one knows where any pixel k was mapped in the feature space, when the scatterplot was built. Thus, for any pixel of the image space, it only needs to carry back the label c found in the feature space, to the real space. As for any clustering method, the question of selecting a number of classes is relevant. Within our framework, the number of classes is the number of modes of the estimated *pdf*. The number of modes itself is related to the width of the smoothing kernel h. Estimating the number of classes consists in looking at the number of modes M as a function of the parameter h. The decreasing curve $M = f(h)$ displays some plateaus, which gives stable solutions for the number of classes.

3. THE SUPPORT VECTOR CLUSTERING ALGORITHM, SVC

In this unsupervised classification algorithm, data points are mapped from the data space to a high dimensional feature space using a Gaussian kernel. In this feature space, we look for the hyperplane which separates the data set from the origin with the largest margin. This hyperplane when mapped back to the data space, forms a set of contours which encloses the data points. These contours are interpreted as cluster boundaries. Points enclosed by each separate contour are associated in the same cluster. As the width parameter of the Gaussian kernel decreases, the number of disconnected contours in the data space increases, leading to an increasing number of clusters [10]. SVC can deal with outliers by employing a soft constant margin that allows the hyperplane in feature space not to separate all points. For large values of this parameter, one can deal with overlapping clusters [10]. To separate the data from the origin, one solves the following quadratic equation [11]:

$$min_{w\epsilon F,\xi\epsilon R^l\rho\epsilon R} \quad \left(\frac{1}{2} \parallel w \parallel^2 + \frac{1}{\nu l}\sum_i \xi_i - \rho\right) \tag{2}$$

Subject to:

$$(w.\Phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0 \tag{3}$$

Since nonzero slack variables ξ_i, are penalized in the objective function, we can expect that if w and ρ solve this problem, then the decision function :

$$f(x) = sgn\left(((w.\Phi(x_i)) - \rho)\right) \tag{4}$$

will be positive for the most samples x_i contained in the data set, while $\parallel w \parallel$ will be still small. The actual trade-off between these two goals is controlled by ν . Using multipliers $\alpha_i, \beta_i \geq 0$ we introduce a Lagrangian:

$$\begin{array}{c} L(w,\xi,\rho,\alpha,\beta) = \frac{1}{2} \parallel w \parallel^2 + \frac{1}{\nu l}\sum_i \xi_i - \rho \\ - \sum_i \alpha_i((w.\Phi(x_i)) - \rho + \xi_i)) - \sum_i \beta_i\xi_i \end{array} \tag{5}$$

and set the derivatives with respect to the primal variables w, ξ, ρ, equals to zero, yielding:

$$w = \sum_i \alpha_i \Phi(x_i) \tag{6}$$

$$\alpha_i = \frac{1}{\nu l} - \beta_i \leq \frac{1}{\nu l}, \quad \sum_i \alpha_i = 1 \tag{7}$$

All samples x_i for which $\alpha_i > 0$ are called support vectors. So the decision function f is transformed into a kernel expansion.

$$f(x) = sgn\left(\sum_i \alpha_i k(x_i, x) - \rho\right) \tag{8}$$

Substituting equations 6 and 7 into 5 we obtain the dual problem:

$$min \quad \left(\frac{1}{2}\sum_{ij} \alpha_i\alpha_j k(x_i, x_j)\right) \tag{9}$$

Subjected to:

$$0 \leq \alpha_i \leq \frac{1}{\nu_l}, \quad \sum_i \alpha_i = 1 \tag{10}$$

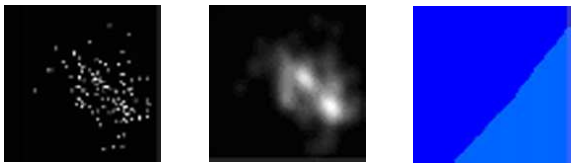
We can recover ρ by exploiting that for any $\alpha_i > 0$ the corresponding pattern x_i satisfies:

$$\rho = (w.\Phi(x_i)) = \Sigma_i \alpha_i k(x_j, x_i) \tag{11}$$

Although this procedure can recover exactly the boundaries of the classes, in our application we used the same SKIZ procedure [9]in order to separate the data space. The binary function corresponds to the connected components in the data space which will be used as seeds in the SKIZ procedure.

4. APPLICATION

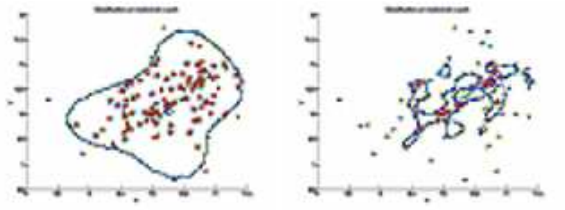
Both procedures described in section 2 and 3, were tested on simulated data and on a real application. Fig.1(a) displays the scatter plot of the simulated data, while Fig.1(b) shows its *pdf* obtained by the Parzen watersheds algorihm and Fig.1(c) the two resulting classes.



(a) *scatter plot* (b) *pdf* (c) classes

Fig. 1. Parzen-watershed on simulatated data.

The SVC method is very robust when tries to separate classes where there is no overlapping, but otherwise it fails as it can be seen on Fig.2 even so both parameters of the method (the kernel size and the outliers' controller) are carefully chosen.



(a) SVM1 (b) SVM2

Fig. 2. SCV results on simulated data.

Now, multispectral image of a cross section of the barley grain acquired in fluorescence, is segmented according to the general scheme of Fig.3 to identify the tissues. The grains were furnished by INRA de Clermont-Ferrand and the images were recorded by INRA Nantes tanks to M.-F. Devaux. The data set contains 19 images of 512x512 pixels. In order to reduce the computing time, we process a sample of 80x10 bidimensional data, obtained by a cross section in the first two principal components. The SKIZ procedure helped us to split the feature space in order to obtain a map of classification which was used to assign the rest of the data. The energy distribution for the four principal components is presented in table 1 and the first two

Table 1. Energy distribution for the PC

Principal components	Energy distribution
PC 1	77.384
PC 2	16.852
PC 3	2.4909
PC 4	1.3034

principal components (they contain more than 90% of the energy) used for the clustering are displayed in Fig.4.

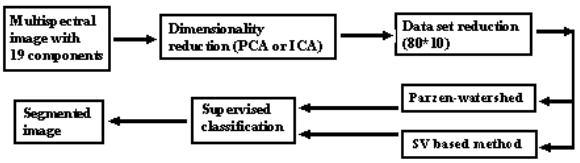
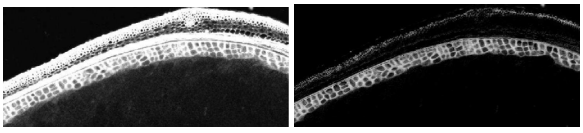


Fig. 3. Scheme of the application.

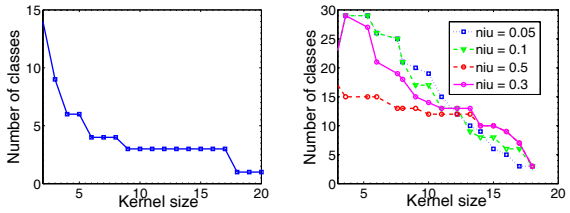


(a) First component (b) Second component

Fig. 4. The two first components.

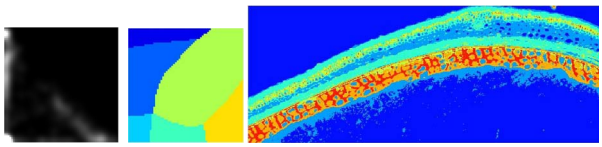
Depending on the kernel size, Parzen-watershed method proposes 6, 4 or 3 classes with significant credibility, as we can see in Fig.5a. The result of the segmentation for 6 classes is presented in Fig.6.
For the SV method, the number of classes depends both on the kernel size and on the outliers' controller. So, the number of classes proposed by this method is 13, 12, 10, 8, and 6, as we can see in Fig.5b. The result of the segmentation for 8 classes is presented in Fig.7. We must say that for 6 classes, SV method cannot reveal the class corresponding to the external tissue (cutin).

The results were compared with the theoretical scheme of the barley grain, and we admitted that a number of six classes of pixels would be enough to identify the tissues. A class does not necessarily correspond to a tissue or to a pure component unless one tissue is constituted only by one chemical component. The tissues may be identifiable by their texture, but the pure components may be achieved by the ICA. As a further research it will be interesting to analyse the classes obtained after the classification process, using the ICA in order to identify the pure compounds presented in each tissue and to compare them to the reference spectra.



(a) Parzen-watershed (b) SVC

Fig. 5. Most frequent number of classes.



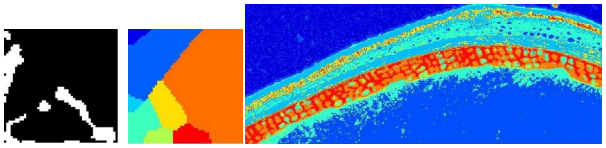
(a) *pdf* (b) classes (c) areas

Fig. 6. Parzen-watershed results for 6 classes.

5. DISCUSSION

Two clustering methods which make assumptions neither on the clusters shape, nor on their number are presented. The Parzen-watershed algorithm (which computes the entire *pdf* for a given data set) is compared with a method based on the Support Vectors theory which computes only the support of the *pdf*.

Using artificial data we simulated the limit case of a two overlapping clusters, a problem which often occurs in data clustering. In that case, the results show that the Parzen-watershed algorithm performs better. One explication of these results can be formulated as follows: estimating the whole *pdf* for a data set reveals more details about the data in study than estimating only the support of the *pdf*. These details (such as peaks in the estimated *pdf*) can lead to class separation. The method based on the SV theory, cannot find such details without loosing reliability.



(a) PDF*pdf* (b) classes (c) areas

Fig. 7. SCV results for 8 classes.

One may expect that the results of both clustering procedures should be the same, as the two methods are somehow related; but is not the case because they perform different when deal with overlapping clusters. Both procedures find 6 classes with significant credibility but in this case, only the Parzen-Watershed method reveals the class corresponding to the external tissue. We explain this situation by the fact that this class is overlapping with the class corresponding to the tissue close to the external layer of the barley grain. In this case, as we saw on the artificial data, SV cannot separate them. The external tissue can be recovered by this method too, but the cost is the increase of the number of classes.

Support Vectors algorithms were successfully used as techniques for supervised learning with very good results. There have been a few attempts to transfer the idea of using kernels to compute inner products in feature space to the domain of unsupervised learning. Kernel learning theory offer further directions: to use or to develop different kernel functions which are suitable for the data set in study.

6. REFERENCES

[1] A. Elhafid, D. Nuzillard, M.-F. Devaux, N. Petrochilos, F. Belloir, "Extraction des signatures de composs purs constituant la couche externe du grain dorge partir dimages de fluorescence", CDRom 449, Belgique, *GRETSI'05*, 2005.

[2] C. Gobinet, A. Elhafid, V. Vrabie, R. Huez and D. Nuzillard, "About importance of positivity constraint for source separation in fluorescence spectroscopy", CDRom 1467, Antalaya Turkey, *EUSIPCO 2005*.

[3] A. Bijaoui, D. Nuzillard, T. Deb Barma,"BSS, Classification and Pixel Demixing", *5rd Int. ICA'04*, pp. 96-103, Granada, Spain, 22-24 september 2004.

[4] J. Cutrona, N. Bonnet, M. Herbin, F. Hofer, "Advances in the segmentation of the multi-component microanalytical images", *Ultramicroscopy*, vol.103, 141–152, 2005.

[5] N. Bonnet, "Artificial intelligence and pattern recognization techniques in microscopic image processing and analysis", *Adv. Imag. Electron. Phys.* vol.14, 1–77, 2000.

[6] M. Herbin, N. Bonnet, P. Vautrot, " Number of clusterings and influence zones", *Pattern Recognition Letters* vol.22, 1557–1568, 2001.

[7] E. Parzen, " On the estimation of a probabitity density function and mode", *Annals Math. Stats.* vol.33, 1065–1076, 1962.

[8] S. Beucher, F. Meyer, "Mathematical Morphology in Image Processing", *Dekker*, New York, p433, 1992.

[9] J. Serra, "Image Analysis and Mathematical Morphology", *Academic Press*, New York, 1982.

[10] A. Ben-Hur, D. Horn, H. Siegelmann, V. Vapnik, "Support vector Clustering", *Journal of Machine Learning Research* 2, 125–137, 2001.

[11] B. Schölkopf, J-C. Plat, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, "Estimating the support of a high dimensional distribution", *Neural Computing* vol.13, 1443–1471, 2001.