

A color object recognition scheme: application to cellular sorting

O. Lezoray¹, A. Elmoataz^{1,2}, and H. Cardot¹

¹ Laboratoire Universitaire des Sciences Appliquées de Cherbourg (LUSAC) IUT, Site Universitaire de Bellevue, 120 Rue de l'exode, 50000 Saint-Lo, France

² Groupe de REcherches en Informatique, Image et Instrumentation de Caen (GREYC) Equipe Image, ISMRA, 6 Bd. du Maréchal Juin, 14050 Caen cedex, France

Received: 15 June 2001 / Accepted: 1 July 2002

Published online: 23 April 2003 – © Springer-Verlag 2003

Abstract. This paper presents a color object recognition scheme which proceeds in three sequential steps: segmentation, features extraction and classification. We mainly focus on the first and the third steps here. A color watershed using global and local criteria is first described. A color contrast value is defined to select the best color space for segmenting color objects. Then, an architecture of binary neural networks is described. Its properties relies on the simplification of the recognition problem, leading to a noticeable increase in the classification rate. We conclude with the abilities of such a recognition scheme and present an automated cell sorting system.

Key words: Morphology – Color – Segmentation – Classification – Microscopy

1 Introduction

The problem of object recognition is one of the main objectives in computer vision. In this paper, we focus on the recognition of objects in color images. An object might be described as a set of pixels that a human observer visually considers as an entity or as what a human expert wants a computer to extract in an image. That definition is highly related to human sensitivity. For example, an object can be a cell in a microscopic image, a face, or a person in an outdoors image. To recognize objects in images, a computer vision strategy has to be planned. Such a strategy can be defined in three sequential steps: extraction (i.e. segmentation) of the objects, object descriptor computing and object classification. For each of these steps, some knowledge is necessary: image analysis for the segmentation and the object descriptor computing, data analysis for the classification of the objects, and knowledge on the source of the image (a priori knowledge). The use of such a vision strategy is suggested to build a color object recognition scheme. In this paper, we mainly focus on the segmentation and the classification.

Correspondence to: O. Lezoray
(e-mail: Olivier.Lezoray@info.unicaen.fr)

2 Morphological color object segmentation

2.1 Color segmentation

With color being information directly attached to regions, color segmentation should allow the obtaining of more significant regions. The main methods are histogram analysis (2D or 3D) [1], pixel classification [2], and region growing [3]. For a more precise review of these principal color segmentation methods, see [4]. In this paper, morphological segmentation methods are considered. These latter seem very appropriate to segmenting complex color images [5]. The structure of a mathematic morphology segmentation is twofold:

- *Marker extraction.* The extraction of the markers is the initialization step of the growing: it is oriented toward an imprecise extraction of the objects. A priori knowledge on the objects enables choosing appropriate operators that allow a total or partial extraction of the objects. For color images, the main a priori information relies on color. For that reason, marker extraction has to use color vectors or color vector components.
- *Growing.* Using previously extracted markers, one of the principal morphological operators can be used, such as a watershed. The latter enables the growing of the markers and the obtaining of the final regions.

A priori knowledge is therefore essential to fulfill a mathematic morphology segmentation strategy. The definition of a new aggregation function for a watershed-based color segmentation is now proposed.

2.2 A new aggregation function

The watershed has been extended to color images by Meyer [6]. Color distances between pixels and neighbor regions are considered as the aggregating probability of a pixel. However, a watershed integrating local and global criteria can achieve a segmentation in a very reliable way. The color of objects and the transitions between regions can be directly expressed by a statistical comparison and a color gradient measure. Therefore, a watershed will be very accurate with the following criteria: local information (modulus of the color gradient) and

global information (a statistical comparison between the color of a point p and a neighbor region R). That new aggregation function can be defined in a formal way. Let $\overline{I_{C_1C_2C_3}}(R)$ denote the mean color vector of the region R of the image I in the color space $C_1C_2C_3$, the vector $I_{C_1C_2C_3}(p)$ denote the color at point p , and $\nabla I_{C_1C_2C_3}(p)$ the color gradient at p . The aggregation function is expressed as

$$f(p, R) = (1 - \alpha) \|\overline{I_{C_1C_2C_3}}(R) - I_{C_1C_2C_3}(p)\| + \alpha \|\nabla I_{C_1C_2C_3}(p)\|.$$

The color image and the color gradient image have to be normalized to have equivalent magnitudes. α is a blending coefficient which allows modifying the influence of the local and global criteria during the growing process.

2.3 Adjustment methodology

A color morphological segmentation using a new aggregation function has been introduced; nevertheless, none of the important parameters of the method have been tuned. The color aggregation function needs to be set for the appropriate color space to use, for how to compute the color gradient, and for how to compare the colors of a pixel and a region. α is also a decisive parameter in the utilization of our method, and it has to be set according to the images to be segmented. Likewise, the markers have not been defined. To conclude, the utilization of our color morphological segmentation is based on several choices. In order to carry out these choices, which are essential for obtaining good results, a methodology of adjustment is defined. It is held in several successive stages, each one bringing the establishment of a parameter or data input to our method. Each stage uses either quantitative measurements or a priori information of the color images.

2.3.1 Color space choice. The choice of a color space can improve the color segmentation results, but its choice still remains a problem. For a more precise review on the color spaces, their representation and properties, see [7]. To evaluate the contribution of a color space in the extraction of objects in an image, a description of the objects is used. The method consists of extracting these objects by a manual segmentation and determining which color space best corresponds to the transitions between the regions. First, an image database has to be build, denoted by $D = \{I_i | i = 1, \dots, n\}$ for n color images. The latter has to be representative of the segmentation problem at several points of view: panorama of the objects, of their color, of their configurations. The objects to be segmented in the images are manually extracted by an expert of the domain. That allows building a representation of the a priori information. The database of manually segmented images is defined by $D^S = \{I_i^S | I_i \in D\}$ where I_i^S is the manually segmented image of the image I_i . To choose a color space, a global contrast $\zeta(I_i)$ is computed for each image I_i of the database D . The global contrast ζ is defined as follows. For each pixel p being at the boundary between two regions in the corresponding labeled image I_i^S , the local contrast $c(p)$ is processed. For a color image I_i in the $C_1C_2C_3$ color space, a

labeled image I_i^S and a pixel p , $c(p) = \{\max(d(p, m))/m$ is a 8-neighbor of $p\}$. The global contrast is the sum of all the local contrasts at the location of the boundaries of the manually segmented objects: $\zeta(I_i) = \sum c(p)$, p being a boundary point in I_i^S . p is a boundary point if $\exists m / I_i^S(p) \neq I_i^S(m)$ and m is a 8-neighbor of p . The color space contrast (CSC) for a database D corresponds to $CSC(D) = \frac{1}{n} \sum \zeta(I_i)$, $i = 1, \dots, n$ in the $C_1C_2C_3$ color space. To choose which color space is the most appropriate for segmenting images, the color space contrast is computed for several color spaces normalized between 0 and 255 before the CSC calculation of D in order to have equivalent results. The color space which gives the highest CSC value is that best adapted to our segmentation method. Color contrast characterizes the spatio-colorimetric dispersion in an image, i.e., the transitions between the colors. The color space choice method is directly related to our segmentation method, which mainly uses the concept of distance between the colors.

2.3.2 Gradient computation. Cumani has defined an operator based on a contrast function computed on several channels [8]. On the base of the works of Cumani, Drewniok has shown how to extend the gradient concept to multi-spectral images [9]. The latter can be used under two assumptions: the different channels have to be correlated, and the level of noise has to be equivalent in all of them. The first assumption is an important limitation since a change of color space allows, according to the color space, uncorrelating the channels. We therefore suggest using a color gradient which can be performed whatever the color space. The color gradient might be computed as follows in the $C_1C_2C_3$ color space:

$$\|\nabla I_{C_1C_2C_3}(x, y)\|^2 = (\|\nabla I_{C_1C_2C_3}^{C_1}(x)\| + \|\nabla I_{C_1C_2C_3}^{C_2}(x)\| + \|\nabla I_{C_1C_2C_3}^{C_3}(x)\|)^2 + (\|\nabla I_{C_1C_2C_3}^{C_1}(y)\| + \|\nabla I_{C_1C_2C_3}^{C_2}(y)\| + \|\nabla I_{C_1C_2C_3}^{C_3}(y)\|)^2$$

Filters used in the X and Y directions are FX and FY respectively:

$$FX = \frac{1}{12} \begin{pmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ -1 & -4 & -1 \end{pmatrix} \quad FY = \frac{1}{12} \begin{pmatrix} 1 & 0 & -1 \\ 4 & 0 & -4 \\ 1 & 0 & -1 \end{pmatrix}$$

To have an orientation-independent gradient image, it is necessary to make convolutions with all the filters obtained from each other by rotation. The final gradient value is the maximum of all the different values obtained for each direction. Since that gradient is sensitive to the noise, it is preferable to preliminary smooth the color image. The computing of the gradient has to be done in the color space obtained at the color space choice stage to ensure a coherence in the computing of the aggregating function, which takes values of same magnitudes.

2.3.3 Metric choice. The choice of metrics depends on a priori information. The Mahalanobis distance is better adapted for regions having fine variations, but for homogeneous objects, the Euclidean distance is accurate enough. The latter can thus be chosen.

2.3.4 Marker choice. Color being a priori information with regards to the objects to be extracted in an image, the marker extraction must be done on color components which correspond to the required a priori information. If a scene presents color information directly related to a color space component, a component which contains that information will be preferred. If it is not the case, the use of color should not be limited to the use of only one color space: the combined use of various components of several color spaces will bring better results. To summarize, the extraction of the markers can be made on a color component or on a combination of several components coming from various color spaces.

2.3.5 α choice. α is tuned with respect to a priori knowledge of the objects to extract. If the boundaries of the objects are well defined, α should be close to 1; if boundaries are less well defined and if the regions are not very homogeneous, α should be close to 0.5; and if boundaries are not well defined but the objects are homogeneous, α should be close to 0. To choose α , it is necessary to tune it until a satisfactory fine segmentation is obtained. However, an adaptable segmentation which modifies the value of α along the iterations can be more suitable (see [10]).

3 Object descriptor extraction

The choice of appropriate descriptors for the objects to be recognized and the appropriate representation for these descriptors is a crucial problem in computer vision. Various types of descriptors have been suggested in literature, ranging from shape descriptors to region descriptors for structural and non-structural properties. Which of the above descriptors and representations are most appropriate for each case depends on the vision task performed. For that reason, we do not focus here on the extraction of appropriate descriptors for the objects previously segmented. However, the one used in our experiments will be described in the corresponding section.

4 Object classification

4.1 Data classification

Once characterized, a color object is no more seen as an object but as a set of numerical data. It represents a given description of the object. That description is used as an input set for a classification method. The majority of the tasks of classification require a phase of training, generating a model which associates an input (represented by a vector of descriptors) with a class. There is a great number of algorithms for the classification of data. Mainly, one finds decision trees [11], statistical methods [12] (dynamic clouds, K-nearest neighbors, Bayes theory) and neural networks [13]. We are interested in this last type of method of classification.

4.2 MONNA:

a multiple ordinate neural network architecture

MONNA (Multiple Ordinate Neural Network Architecture) is made of several small neural networks to solve a problem of

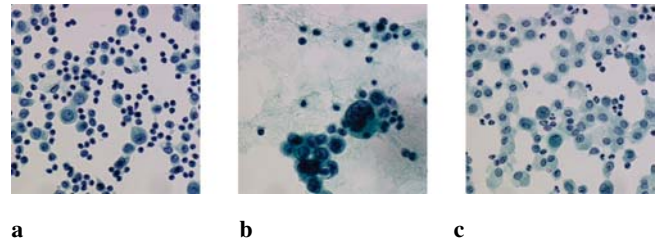


Fig. 1a–c. The images of the different types of cells. **a** Type 1; **b** type 2; **c** type 3

classification. The neural networks used are multi-layer networks with back-propagation of the gradient error (multi-layer perceptrons, MLP). MONNA is based on three steps: (1) construction, (2) training and (3) decision. A more precise description and experimentation of MONNA can be found in [14]. The classification of data by only one large network can present difficulties of generalization when the number of classes of objects to be separated is high. Therefore a simplification of the problem can be a reduction of the number of classes having to be recognized by this network. The idea consists in using only neural networks of simple structure. For a problem of classification having to separate n classes, MONNA is composed of a set of unconnected networks, each one intended to separate the elements into two distinct classes. If the problem has n classes, that leads us to have $(n(n-1))/2$ neural networks. The difficulty in separating n classes is simplified by the specialization of each network, because a network is interested only in separation of two classes. The structure of the neural networks used is the following one: a layer of inputs containing as many neurons as the number of descriptors associated with the object to be classified, an hidden layer containing a variable number of neurons and one output neuron. The value of the output neuron is in the interval $] -1, 1[$. According to the sign of the result associated with this single neuron, an object is classified in one of the two classes that the network separates. During the training step, each network learns to recognize only examples of two classes: the set of data to be learned is restricted. This implies, on the one hand, simplifying the training, and, on the other hand, making easier the discrimination between these two classes since the network learned how to recognize only those. The global training dataset containing patterns of all the different classes is divided in several subsets for each neural network. $S_T(c_i, c_j)$ is the dataset which corresponds to the neural network which differentiates the classes C_i and C_j . This initial training data associated with each neural network is split into two subsets: a learning set ($S_L(c_i, c_j)$) and a validation set ($S_V(c_i, c_j)$), consisting in 80% and 20% of $S_T(c_i, c_j)$ respectively. The learning of a neural network is performed on $S_L(c_i, c_j)$ and the $S_V(c_i, c_j)$ validation set is used to evaluate the classification rate of the network during the training. An automatic method is used to find the number of hidden neurons that gives the best classification rate [14]. Once the training of a network is carried out, the classification rate of this network is available. The latter is obtained on the $S_V(c_i, c_j)$ validation dataset and thus relates only to data that have not been learned. We must now specify how to use the results coming from each network in order to classify an object. With that intention, a hierarchy of neural networks is built. The networks can be ordered according to their quality

of classification (the rate of classification obtained by each network at the end of the training), the first network being that which learned best, i.e., that for which the decision of classification is most reliable. To choose the final class of an object to be classified, a selection by elimination is performed. The networks are used one by one along the hierarchy, eliminating progressively the classes according to the output of the neural networks until only one class is available: this is the final class of the object to be classified. The advantages of MONNA are improved classification, and faster and incremental learning since the whole inducer is not totally rebuilt each time new training data is available.

5 Experiments: an automated cell-sorting system

5.1 Problem description

In serous cytology, samples are smeared over slides, fixed, colored by the Papanicolaou international standard of coloration and examined under the microscope in order to locate cells of interest. That slide-reading stage consists of a visual evaluation of the cytological slide and is called *screening*. The goal of this stage is the detection of abnormal or suspect cells. It is thus of capital interest for the pathologist, who must establish a reliable and valid diagnosis. We suggest applying our color object recognition scheme to build an automatic cell-sorting system for quality control in serous cytology. For a more precise description of the system, see [15]. In this section, each step of our color object recognition scheme is presented.

5.2 Object-segmentation strategy

We proceed in five stages. The first stage consists of establishing a presence of cells on the image. In fact, if cells are present, one carries out the elimination of the red blood corpuscles. In the two following stages, the nuclear and cytoplasmic areas are extracted. Once found, at the time of the fourth and fifth stages, touching nuclei and cytoplasm are separated in order to have only one nuclei or cluster of nuclei by cytoplasm. The strategy is a bottom-up one: the first stages find global information in the images, and following stages focus on the objects. The complete segmentation will not be totally described here – for a more precise explanation, see [15]. However, to illustrate the morphological color segmentation and its adjustment methodology, it is described for the segmentation of the nuclei.

5.2.1 Nuclei segmentation. The segmentation of the nuclei is achieved by the means of our color morphological segmentation. To understand it, the adjustment methodology is described.

1. The color space choice is done on representative images of the nuclei of the cells. The nuclei of the cells have an extremely variable color. The images having to be used for the computation of the *CSC* must thus contain enough representative nuclei in order to have a reliable measurement of the contrast. Three images containing three different types of cells were used. Type 1 corresponds to an

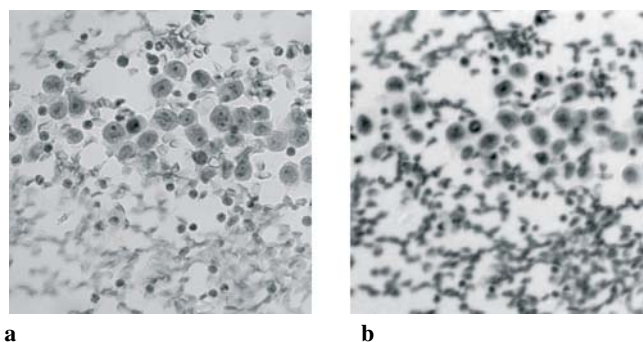


Fig. 2. **a** The blue component of Fig. 3a. **b** The I_S image of Fig. 3b

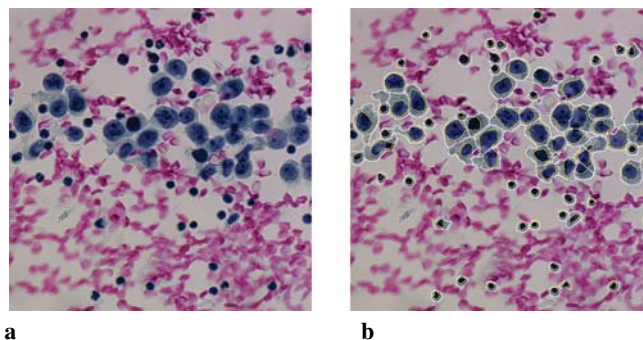


Fig. 3. **a** A cytological color image. **b** A cytological segmented color image. A cytological color image and the corresponding segmentation (nuclei boundaries are in yellow and cytoplasm boundaries in white)

Table 1. Global contrast for each type of nuclei and *CSC* in several color spaces

Nuclei	RGB	XYZ	L*u*v*	L*a*b*	HSL
Type 1	89.63	88.24	62.54	34.74	88.47
Type 2	68.57	67.81	67.31	47.05	68.61
Type 3	98.44	95.25	71.93	36.56	95.76
<i>CSC</i>	85.54	83.76	67.26	39.45	84.28

image containing the cells with isolated or touching dark blue nuclei (Fig. 1a); type 2 corresponds to an image containing the cells with very dark nuclei isolated or in clusters (Fig. 1b); and type 3 corresponds to cells with very clear nuclei isolated or touching (Fig. 1c).

The *CSC* is computed for the database, and the color space RGB arises as giving the highest one (see Table 1). This color space is therefore more suitable for the segmentation of the nuclei.

2. The Euclidean metric is used.
3. The gradient is computed in the RGB color space from the original color image smoothed by an exponential filter of value 0.4.
4. The nuclei of the cells are recognizable by their blue color and that a priori information is used. However, the color of the nuclei has very high variability so their color can range from very clear to very dark blue (chromatin texture variations). That makes the systematic extraction of markers difficult because both clear and dark nuclei should be extracted at the same time. On a color component of

a color space, that extraction is difficult in a satisfactory way. Thus an image is created from the fusion of several components of various color spaces. With $I_{C_1 C_2 C_3}^{C_1}$ denoting the component C_1 of the image I in the color space $C_1 C_2 C_3$, the following image is used:

$$I_S = \frac{(I_{RGB}^B - I_{RGB}^G) + I_{L^*u^*v^*}^Y}{2}$$

The image $I_{L^*u^*v^*}^Y$ accentuates the clear nuclei and the image $(I_{RGB}^B - I_{RGB}^G)$ levels the difference between the clear nuclei and the dark nuclei. The image I_S restricts the nuclei within a small gray-level interval (Fig. 2) and the determination of the markers can then be done in an automatic way by a thresholding selected from the values of the interclass variance and the valley (compared to the largest peak) of the histogram of I_S .

5. The parameter α is fixed at 0.5. That makes it possible to take the amplitude of the gradient into account and the color homogeneity of the nuclei. The use of both information allows the obtaining of better nuclear regions because the gradient has certain limitations at the borders of the nuclei.

5.2.2 Segmentation results. An illustration of the complete segmentation strategy result is given by Fig. 3b, which gives the corresponding nuclei and cytoplasm boundaries of the objects (the cells) extracted in Fig. 3a. That segmentation has been evaluated by three experts and the average rate of correct segmentation is 93.0% for the cytoplasm and 94.5% for the nuclei, which is considered accurate for that application [15].

5.3 Object descriptor extraction

Once segmented, the cells have to be described by discriminating enough parameters to allow their good classification. For that experimentation, the various types of objects that can be met in serous cytology were indexed. One distinguishes the isolated cells (made up of 18 classes), the clusters (composed of superpositions of isolated cells) and the debris. That represents a number of classes of cells to be recognized rather important. The criteria which make it possible for the experts to differentiate those are size, color, form and texture. There are 46 parameters which are measured for each cell extracted from the segmentation currently in use [15]: size and form parameters relating to nuclei and cytoplasm, and texture parameters measuring the nuclei because only that makes it possible to evaluate cell malignancy.

5.4 Object identification

To classify an object extracted by the segmentation, it is necessary to initially determine if it is debris, a cluster of cells or an isolated cell. That is carried out by a C4.5 decision tree with an error rate of 1.9%. According to that first classification, the isolated cells must be distributed in the 18 different classes of objects. They are classified by using MONNA with the SFFS feature selection method. The training of our architecture is carried out on a database of 3245 cells (obtained

from the intersection of three experts' independent labeling). The total rate of recognition of MONNA after learning with feature selection by the SFFS wrapper method [16] is 84% for the cells of the test base (the recognition rate for MONNA without feature selection is 66%, which is higher than the one obtained using a single large neural network of 60%). Work on extending the base remains to be carried out in order to balance and to increase the number of cells of each class. According to the optics of our system, namely, the detection of abnormal or suspect cells, our system is very satisfactory because 94.5% of the abnormal cells and 99.3% of the normal cells are recognized, which is considerably higher than the rate of success of an expert. The whole strategy takes 0.5 s for a 512×512 image on a 700 MHz Pentium III. Since our system operates in post-screening, within sight of the results obtained, an abnormal cell omitted by the cytotechnician will be detected by our system.

6 Conclusion

In this paper, a color object recognition scheme was presented. That scheme breaks up into three stages useful for the development of a computer vision system: the extraction of the objects, the description of the objects, and the identification of the objects. General methods were put forward for the first and last stages. The extraction of color objects is based on a new aggregation function for watersheds, and uses local and global criteria. In order to be able to adapt the critical points of the method, an adjustment methodology was suggested. The identification of the objects is done by a neural network architecture based on the divide-to-conquer principle. The basis of this neural network architecture consists of simplifying the recognition task by dividing the problem to be solved. These stages defining our color object recognition scheme can be used to solve various computer vision problems and especially biomedical ones as shown with our system of cell sorting. The interaction with the user is reduced to the extraction of a priori knowledge for building the reference image database, the extraction of the markers and object labeling.

References

1. Park S, Yun I, Lee S (1998) Color image segmentation based on 3-D clustering: morphological approach. *Pattern Recogn* 31:1061–1076
2. Campadelli P, Medici D, Schettini R (1997) Color image segmentation using hopfield networks. *Image Vision Comput* 15:161
3. Trémeau A, Borel N (1998) A region growing and merging algorithm to color segmentation. *Pattern Recogn* 30:1191–1203
4. Cappellini V, Plataniotis K, Venetsanopoulos A (1995) Applications of color image processing. In: *Proceedings of the international conference on digital signal processing*, Limassol, Cyprus, 26–28 June 1995
5. Shafarenko L, Petrou M, Kittler J (1997) Automatic watershed segmentation of textured color images. *IEEE Trans Image Process* 6:1530–1543
6. Meyer F (1992) Color image segmentation. In: *Proceedings of the international conference of image processing and its applications*, Maastricht, The Netherlands, 6–9 April 1992

7. Wyszecki G, Styles W (1982) Color science: concepts and methods, quantitative data and formulae, 2nd edn. Wiley, New York
8. Cumani A, Grattoni P, Giuducci A (1991) An edge-based description of color images. *Comput Vision Graph Image Process* 53:313–323
9. Drewniok C (1994) Multi-spectral edge detection: some experiments on data from Landsat-tm. *Int J Remote Sensing* 15:3743–3765
10. Lezoray O, Cardot H (2002) Histogram and watershed based segmentation of color images. In: *Proceedings of the first European conference on color in graphics, imaging and vision, Poitiers, France, 2–5 April 2002*
11. Quinlan J (1993) *C4.5: programs for machine learning*. Morgan Kaufman, San Mateo
12. Jain A, Dubes R (1988) *Algorithms for clustering data*. Prentice Hall, Englewood Cliffs, N.J.
13. Hecht-Nielsen R (1990) *Neurocomputing*. Addison-Wesley, Reading, Mass.
14. Lezoray O, Cardot H, Fournier D, Revenu M (2000) Monna: a multiple ordinate neural network architecture. In: *Second ICSC symposium on engineering of intelligent systems, University of Paisley, Scotland, 27–30 June 2000*
15. Lezoray O, Elmoataz A, Cardot H, Revenu M (1999) Arctic: an automatic cellular sorting system using image analysis. In: *Proceedings of vision interface '99, Trois-Rivières, Canada*
16. Pudil P, Ferri F, Novovicová J, Kittler J (1994) Floating search methods in feature selection. *Pattern Recogn Lett* 15:1119–1125



O. Lezoray received the M.Sc. degree in computer science in 1995 and the Ph.D. degree in computer science from the University of Caen in 2000. From September 2000 to August 2001 he worked as a one year assistant professor in the Computer Science Department of the University of Caen. Since September 2001, he joined the Communication Networks and Services Department of Technology of the University of Caen as an assistant professor. His research focus on image segmentation techniques for color images and data classification methods based on the cooperation of machine learning methods. He is a participant of the GDR-ISIS French National Research Group.



A. Elmoataz has a Ph.D. in image processing. He is an assistant professor in the Computer Science Department of the University of Caen. His research focus on PDE based methods for image simplification and segmentation.



H. Cardot is an engineer of the National School of Engineers of Caen and has a Ph.D. in signal processing. He is an assistant professor at the Electrical and Computer Science Department of technology of the University of Caen. He is the head of the image team of the LUSAC. His research focus on neural networks and their applications in the fields of image processing.