Cellular and MolecularBiology

ISSN 1165-158X
2004Cell.Mol.Biol.

A FAST AND EFFICIENT SEGMENTATION SCHEME FOR CELL MICROSCOPIC IMAGE

G. LEBRUN[1], C. CHARRIER[1], O. LEZORAY[1], C. MEURIE[1] AND H.CARDOT[2]

[1] LUSAC EA 2607, groupe Vision et Analyse d'Image, 120 Rue de l'exode, F-50000 Saint-Lô, France
Fax: +33(0)2 33 77 11 67; Email : {gilles.lebrun, c.charrier, o.lezoray, cyril.meurie}@chbg.unicaen.fr
[2] Laboratoire Informatique (EA 2101), Université François-Rabelais de Tours, 64 Avenue Jean Portalis, Tours, F-37200, France
Email : Hubert.cardot@univ-tours.fr

Abstract – microscopic cellular image segmentation schemes must be efficient for reliable analysis and fast to process huge quantity of images. Recent studies have focused on improving segmentation quality. Several segmentation schemes have good quality but processing time is too expensive to deal with a great number of images per day. For segmentation schemes based on pixel classification, the classifier design is crucial since it is the one which requires most of the processing time necessary to segment an image. The main contribution of this work is focused on how to reduce the complexity of decision functions produced by support vector machines (SVM) while preserving recognition rate. Vector quantization is used in order to reduce the inherent redundancy present in huge pixel databases (*i.e.* images with expert pixel segmentation). Hybrid color space design is also used in order to improve data set size reduction rate and recognition rate. A new decision function quality criterion is defined to select good trade-off between recognition rate and processing time of pixel decision function. The first results of this study show that fast and efficient pixel classification with SVM is possible. Moreover posterior class pixel probability estimation is easy to compute with Platt method. Then a new segmentation scheme using probabilistic pixel classification has been developed. This one has several free parameters and an automatic selection must dealt with, but criteria for evaluate segmentation quality are not well adapted for cell segmentation, especially when comparison with expert pixel segmentation must be achieved. Another important contribution in this paper is the definition of a new quality criterion for evaluation of cell segmentation. The results presented here show that the selection of free parameters of the segmentation scheme by optimisation of the new quality cell segmentation criterion produces efficient cell segmentation.

Key words : cell microscopic image, segmentation, pixel classification, quality criterion, SVM, hybrid color space, vector quantization. Metaheuristic.

INTRODUCTION

In recent years, computer-aided image processing and analysis systems have played a significant part in quantitative pathology. Image analysis in the field of cancer screening is a significant tool for cytopathology (7,15). Two principal reasons can be highlighted. Firstly, the quantitative analysis of shape and structure of nuclei coming from microscopic color images brings to the pathologist valuable information for diagnosis assistance. Secondly, the quantity of information that the pathologists must deal with is increasingly large, in particular when the number of cancer screening increases. That is why, segmentation schemes for microscopic cellular imaging must be efficient for reliable analysis and fast in order to process large quantity of images.

Several studies have shown that segmentation schemes combining color pixel classification and morphological operations are efficient with microscopic cellular images (12,13,14). Cells staining with Papanicolaou international staining (Fig. 1) make it possible to classify the color pixels among three classes (Fig. 2): background, cytoplasm or nucleus, but this classification cannot be perfect. Indeed, a fraction of nucleus pixels have the same color as cytoplasm pixels because of the variability of the nucleus according to the type of cells and of the chromatin distribution. Moreover, for some cytophathologies, the mucus present in the background has the same color as some cells (cytoplasm and nucleus). Morphological operations, like regions growing, which take into account neighborhood relations from the

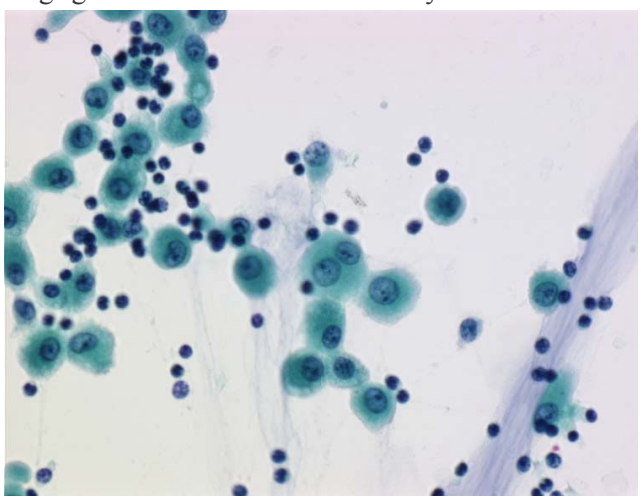

Fig. 1 Microscopic cellular images with international coloration of Papanicolaou.

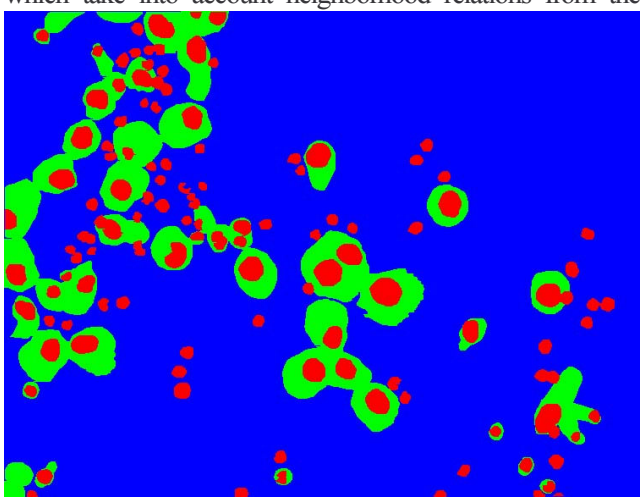

Fig. 2 expert segmentation result on the microscopic image: background (blue), cytoplasm (green), nucleus (red)

spatial repartition of pixels on cells and nucleus, can improve the quality of the segmentation.

The studies mentioned before only focus on improving segmentation quality, this can lead to build a powerful classifier of pixels but the computing time is intractable for a real application. Therefore, a new learning method (11) which uses Vector Quantization (VQ) technique is proposed to build Support Vector Machine (SVM) decision functions of reduced complexity, while preserving efficient generalization. Moreover each decision function uses an adapted hybrid color space (14). The aim is to build a fast and efficient Support Vector Machines classifier of pixels with probabilities estimation. Secondly, a new segmentation scheme is defined, based on the fast and efficient SVM pixels classification. That segmentation scheme has several parameters which must be well tuned in order to optimize the segmentation quality.

Another problem for the design of segmentation schemes is how to evaluate the segmentation quality. Indeed, almost all the segmentation schemes have some parameters. Human observation highlights that the values chosen for those parameters are significant for the quality of the segmentation. However, for an automatic selection of the optimal parameter values, the quality of segmentation must be also automatically evaluated. In literature, there are several quality segmentation criteria: Lui and Borsotti (1), Vinet (23), classification rates and other statistical measures (5). These approaches have drawbacks and are not well designed for the evaluation of cell image segmentation quality.
A new quality criterion is proposed for segmentation quality evaluation. This takes into account the fact that different experts cannot make exactly the same pixel segmentation, nevertheless the differences occur only for pixels near the edge of the shapes. Our new quality criterion weights all pixel classification errors according to its nearer distance from edge shapes of a manual expert segmentation. It is crucial that the segmentation misses the least possible cells and nucleus present in the expert segmentation. In the same way, it also is significant that the number of artifacts produced by the automatic segmentation is as small as possible. Our new criterion takes into account that effect by counting the number of missed or artifact objects in the automatic segmentation by comparing it with the expert segmentation.

The five sections of this paper are organized in the following manner. The first section is the introduction. The second section details the new learning method proposed to build fast pixel classification by using SVMs. The third section describes the segmentation scheme and defines a new quality criterion for cell image segmentation. The fourth section shows some experimental results and the discussion of them. The last section draws a conclusion and refers to future work.

FAST PIXEL CLASSIFICATION

Segmentation processing time is greatly linked to pixel classification processing time when used in a segmentation scheme. As for any classification problem, the choice of an inducer which produces efficient decision functions having good generalization performances is crucial. Working with machine learning algorithms for pixel classification involves to take into account not only the recognition rate of the base inducer but also the processing time needed to perform a single pixel classification. SVM are powerful classifier having high generalization abilities, but the decision function provided by SVM has a complexity which increases with training set size (16,10) (*i.e.* pixel database). Therefore, using SVMs on a huge pixel dataset is not directly tractable for pixel classification.

To this aim, we propose a new learning method which makes it possible to use SVM within the pixel classification framework. This method uses the VQ principle (4) to simplify the training set and thus permits to reduce the complexity of the decision function built by SVM. The Decision Function Quality (DFQ) criterion for the pixel classification depends on two terms: the decision function recognition rate and the decision function complexity. For pixel classification, the decision function complexity depends on the color space used and the number of Support Vectors (see next subsection). Classical color space representation of a pixel is denoted by its *RGB* values, however, depending on the application, another more adapted color space can be chosen (*XYZ*, $L^*a^*b^*$, $L^*u^*v^*$, …). This choice is difficult and subjective; therefore it is more reliable to define a hybrid color space (14) which will be more adapted to optimize the DFQ criterion. For that reason, it is essential that our learning method selects a hybrid color space adapted to each decision function produced by SVM. This hybrid color space is built by selecting a set of color components which can belong to any of the different classical color spaces (14). The mechanism used in our method for the selection of the color components is similar to those which typically used within the feature selection framework (9).

For each binary decision function produced by SVM, our learning method must thus choose the values of the SVM hyper-parameters, the simplification level of the training set and the hybrid color space in order to optimize the DFQ criterion. Exhaustive search for model selection being not tractable, we decided to use *tabu* search meta-heuristic because of its efficiency (8).

Decision functions built by SVM do not provide class probabilities estimation. Pixel class probabilities estimation are useful for the whole segmentation scheme, for example to makers extraction by watershed (12), but also for combination of SVM decision function to produce multi-class classifiers. For those reasons, Platt method (17) has been used to map SVM outputs into posterior probabilities. Two combination schemes have been used to produce multi-class posterior probabilities estimation with SVM decision function.

A method for posterior class probabilities estimation, a multi-class combination scheme, a simplification step by

using VQ, a hybrid color space selection, a new criterion for the DFQ and a *tabu* search meta-heuristic enables us to define a new learning method which produces, in tractable times, an efficient decision function with a reduced complexity and an efficient pixel class posterior probability estimation. The following sub-section goes deeper into the essential information needed to understand and construct that new learning method and ends by a description of that one.

1) SVM OVERVIEW

The SVM were developed by Vapnik and Colaborators (22). They are based on the structural risk minimization principle from statistical learning theory (22). SVM express predictions in terms of a linear combination of kernel functions centered on a subset of the training data, known as support vectors. Given the training data set $S_a : \{(x_i, y_i)\}$, $i \in \{1,\ldots,m\}$, $x_i \in \mathbb{R}^n$, $y_i \in \{-1,+1\}$, SVM maps the input vector x into a high-dimensional feature space H through some mapping functions $\phi : \mathbb{R}^n \to H$ and builds an optimal separating hyper-plane in this space. The mapping ϕ is performed by a kernel function $K(\cdot,\cdot)$ which defines an inner product in H. The separating hyper-plane given by a SVM is: $w \cdot \phi(x) + b = 0$. The optimal hyper-plane is characterized by the maximal distance to the closest training data. The margin (distance to the nearest examples) is inversely proportional to the norm of w. Thus computing this optimal hyper-plane is equivalent to minimize the following optimization problem:

$$\mathrm{V}(w,b,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$$

where the constraint :

$$\forall_{i=1}^{m} : y_i\left(w\cdot\phi(x_i)+b\right) \geq 1-\xi_i,\ \xi_i \geq 0$$

requires that all training examples are correctly classified up to some slack ξ_i and C is a parameter allowing trading-off between training errors and model complexity.

This optimization is a convex quadratic programming problem. Its whole dual problem (21) is to maximize the following constrained optimization problem:

$$\mathrm{V}_{dual}(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$

$$\forall_{i=1}^{m} : 0 \leq \alpha_i \leq C,\ \sum_{i=1}^{m} y_i\alpha_i = 0$$

The optimal solution α^* specifies the coefficients for the optimal hyper-plane $w^* = \sum_{i=1}^{m}\alpha_i^* y_i \phi(x_i) + b^*$ in the feature space H and defines the subset $S_v : (x_i, y_i) \in S_a$, $\alpha_i > 0$. All data examples in S_v are called support vectors. The expression of binary decision function h induces by the dual optimization problem is the follow:

$$h(x) = \mathrm{sign}(f(x))$$

$$f(x) = \sum_{(x_i,y_i)\in S_v}\alpha_i^* y_i K(x_i,x_j) + b^*$$

$f(x)$ is called the output of SVM and its value corresponds to the distance of the point x to hyper-plane separator w^* in the feature space. The threshold b^* is computed via the unbounded (*i.e.* $\alpha_i < C$) support vectors (22). An efficient algorithm SMO (16) and many refinements (2,3) were proposed to solve the dual problem.

An important remark is that SVM algorithm produces a decision function h which only depends on support vector examples in the training set. So complexity of SVMs decision functions is directly linked to the number of support vectors and therefore the training dataset size. The complexity of the decision function is also linked to the dimensionality n of input vector x.

The choice of kernel function K is important for the efficiency of SVM decision function. Radial Basis Function (RBF) is generally greatly efficient with SVM. For incorporation of hybrid color space selection (*see* next subsection 5 and 6 for more details), an extended version of RBF kernel K_β has been used (x^l is the l^{th} attribute of vector x and β_l indicates if the l^{th} attribute is used or not):

$$K_\beta(x_i,x_j) = \exp\left(\frac{-\sum_{l=1}^{n}\beta_l\left(x_i^l, x_j^l\right)^2}{\sigma^2}\right)$$

2) SVM PROBABILITIES ESTIMATION

The output of an SVM is not a probabilistic value, but non-calibrated distance measurement of an example x to the separating hyper-plane w^*. Platt proposed a method (17) to map the SVM into the positive class posterior probability by applying a sigmoid function to the SVM output:

$$p(y=+1 \mid x) = \frac{1}{1+e^{a_1\cdot f(x)+a_2}}$$

The parameters a_1 and a_2 are determined by minimizing the negative log-likelihood under a test set (17).

3) MULTICLASS SVM METHOD

SVM are binary classifiers and multi-class decision functions using that SVM are usually designed by combining several two-class SVM decision functions (19). The *one-versus-all* method using winner-takes-all strategy and the *one-versus-one* (or pairwise) method implemented by max-wins voting are generally used for this purpose. There are many other methods, but those ones could be interesting alternatives when muti-class problems have a greater number of classes, which is not the case here.

When multi-class probabilities must be estimated, those two previous mentioned methods must be adapted. Let ω be the set of all multi-class label ($i, j, c \in \omega$, $n_\omega = |\omega|$). Let (a) $p_{(i,j)}^{\mathrm{SVM}}(x)$ and (b) $p_{(i)}^{\mathrm{SVM}}(x)$ the class posterior probably of x produced by mapping the SVM output with the Platt method after training SVM on a modified version of the initial training set.

For (a) training sets that have only examples of class i and j with respectively class label of +1 and –1. SVM binary decision functions products with those training sets are used in the *one-versus-one* combination scheme . For (b) all examples in training sets are kept but the ones which belong to the class i have the labels +1 and the others have the label –1 in binary corresponding problem. SVM binary decision functions products with those other training sets are used in the one-versus-all schemes.

For one-versus-all combination schemes, all classes probabilities estimation are computed by:

$$p(c=i \mid x)=\frac{p_{(i)}^{\mathrm{SVM}}(x)}{\sum_{j=1}^{n_\omega} p_{(j)}^{\mathrm{SVM}}(x)}$$

For one-versus-one combination, there are several methods for class probabilities estimation (24). All of them must take into account the restricted two class probability estimation values ($p_{(i,j)}^{\mathrm{SVM}}(x)$) for computing the individual class probabilities. One fast and efficient way to do this, is to use the following Price formulation (18):

$$p(c=i \mid x)=\frac{1}{\sum_{j=1, j\neq i}^{n_\omega} \frac{1}{p_{(i,j)}^{\mathrm{SVM}}(x)}-(n_\omega-2)}$$

4) DATASET REDUCTION SIZE METHOD

The main idea for reducing the complexity of SVM decision functions is to reduce the training set size. One possibility for doing this is to produce prototypes which sums up efficiently many examples near of those ones. The Vector Quantization is a classification technique used in the compression field (4) which can perform this. VQ maps a vector x to another vector x' that belongs to m' *prototypes* vectors which is called *codebook*. The *codebook* S' is built from a training set S_a of size m ($m \gg m'$). The algorithm must produce a set S' of prototypes x' which minimizes the distortion d':

$$d'=\frac{1}{m}\sum_{i=1}^{m} \min_{1\leq j\leq m'} d(x_i, x_j')$$

with $d(\cdot,\cdot)$ a distance norm (Euclidian distance). LBG is one of those algorithms (4) which can build this *codebook*. It is an iterative algorithm which produces 2^k prototypes after k iterations.

The dataset simplification method produces a more or less simplified version (S_a^k) of the initial training set in function of the parameter k value and can be formalized by:

$$S_a^k=\bigcup_{c\in C_l} LBG\left(S_{(c)},k\right)\times\{c\}$$

with $S_{(c)}=\{x \mid (x,c)\in S_a\}$ is the feature subset of examples belonging to the class c and $LBG\left(S_{(c)},k\right)$ is the result of k iterations of LBG algorithm with feature.

As the level of simplification k cannot be easily fixed in an arbitrary way, a significant concept in our method is to consider k as variable. In our new learning method k is a parameter which must be tuned by a model selection process.

5) HYBRID COLOR SPACES

The pixels of a color image are digitized in (R,G,B) color space. However, this color space is not always the best suited one for image processing problems and especially for pixel classification. There are many different color spaces and each one presents specific colorimetric, physical and physiological properties (21). For our study, we have retained the most commonly used color spaces[1]: (X,Y,Z), (L^*,a^*,b^*), (L^*,u^*,v^*), (L_1,C,H_1), (Y_2,Ch_1,Ch_2), (I_1,I_2,I_3), (H_2,S,L_2), (Y_3,C_b,C_r). Moreover, in some experiments, it was shown that by combining color components from several color spaces, it is possible to build a hybrid color space more suitable than the initial ones (20). Let E be the space regrouping all n_E distinct color components from e different classical color spaces. By definition a hybrid color space H_E^β is composed of a set of n_β components from E and the vector β indicates which components from E are used (*i.e.* $i\in\{1,\ldots,n_E\}$, $\beta_i=1$ if the i^{th} color component of the space E is used in H_E^β and $\beta_i=0$ in the other case).

In our study, $e=9$, $n_E=25$ and $E=$ (R, G, B, X, Y_1,Z, L^*, a^*, b^*, u^*, v^*, C, H_1, Y_2, Ch_1, Ch_2, I_1, I_2, I_3, H_2, S, L_2, Y_3, C_b, C_r). Then, the objective of our new learning method is to find an adapted hybrid color space H_E^β (*i.e.* the boolean value of β) which improves recognition rate of SVM decision function and allows high level dataset reduction. The β parameter must also be automatically tuned by a model selection process.

6) MODEL REPRESENTATION

Each SVM binary decision function h involved in a multi-class combination scheme is linked to the hybrid color space selection (β), the level of training dataset reduction (k) and hyper-parameter values of SVM (C,σ). All those parameters are summed up in a global model noted θ and a binary decision function produced by SVM for a given model θ is noted h_θ. The representation choice for a model θ is a integers vector ($\beta_1,\ldots,\beta_{n_E},k,C',\sigma'$) with $\beta_i\in\{0,1\}$, $C'\in\{-5,\ldots,15\}$[1], $C=2^{C'}$, $\sigma'\in\{-10,\ldots,10\}$, $\sigma=2^{\sigma'}$. The quantization of C and σ is a classical reduction of model space used in SVM field and it is commonly called "grid search" techniques (2).

7) DECISION FUNCTION QUALITY (DFQ)

The DFQ q of a specific binary decision function h produced by SVM algorithm for a given model θ depends on the recognition rate R_R but also on the complexity C_p of a decision function h_θ when processing time is crucial. The DFQ q can be modeled by: $q_\theta=R_R(h_\theta)-C_p(h_\theta)$. When the decision function is built by SVM with a fixed kernel, the complexity of this decision function depends on

[1] We have added indices for some color components to differentiate them when being denoted by the same letter but not being identically computed.

the number of support vectors and the hybrid color space used. We have chosen to model this complexity as following:

$$C_p(h_\theta) = c_{p_1} \log_2(|S_v(h_\theta)|) + c_{p_2} \log_2(\text{cost}(H_E^\beta))$$

Constants c_{p_1} and c_{p_2} are weighting coefficients which respectively represent the importance of the number of support vectors and the choice of the hybrid color space ($\text{cost}(H_E^\beta)$) in the complexity of h_θ.

The i^{th} color components ($i > 3$) of a pixel are computed by linear or not linear transformation of the first three RGB components (20). The time cost to compute a given color component is more or less expensive as regards the kind of transformation (linear or not, software or hardware). Let κ_i denote the transformation cost to compute the value of i^{th} color components, the value of $\text{cost}(H_E^\beta)$ linked to the hybrid color space H_E^β has been defined as:

$$\text{cost}(H_E^\beta) = \sum_{i=1}^{n_E} \beta_i \kappa_i$$

For hardware transformation, the number of the color components used have a great impact on silicon implementation cost. We have chosen to crudely model it as: $\forall i : \kappa_i = 1$. For software transformation, the processing time (t_i) necessary to compute a specific color component (i) has an impact on the cost of a hybrid color space using it. We have chosen to crudely[2] model it by: $\kappa_i = n_E \cdot t_i / \sum_{j=1}^{n_E} t_j$

[2] Crudely because some color components are helpful to compute other color components.

8) MODEL SELECTION METHOD

The research of the exact θ^* which optimizes the DFQ criterion is not tractable by exhaustive search and by another way classic gradient descent cannot be applied, so a meta-heuristic technique must be used.

Tabu Search (TS) is a meta-heuristic approach for difficult optimization problems. The roots of TS go back to the 1970s; it was first presented in its actual form by Glover (6). TS belongs to iterative neighborhood search methods. The general step, at the it iteration, consists in searching from a current solution θ^{it} a next best solution θ^{it+1} in the neighborhood. This new solution may be less efficient than the previous one, however it avoids local minimum trapping problems. That is why, TS uses short memory to avoid creating cycles. The use of this short memory is helpful to avoid moves which might lead to recently visited solutions (*tabu* solutions). Although the basic idea of TS is straightforward, the choice of solutions coding, objective function, neighborhood, *tabu* solutions definition depends on the application problem.

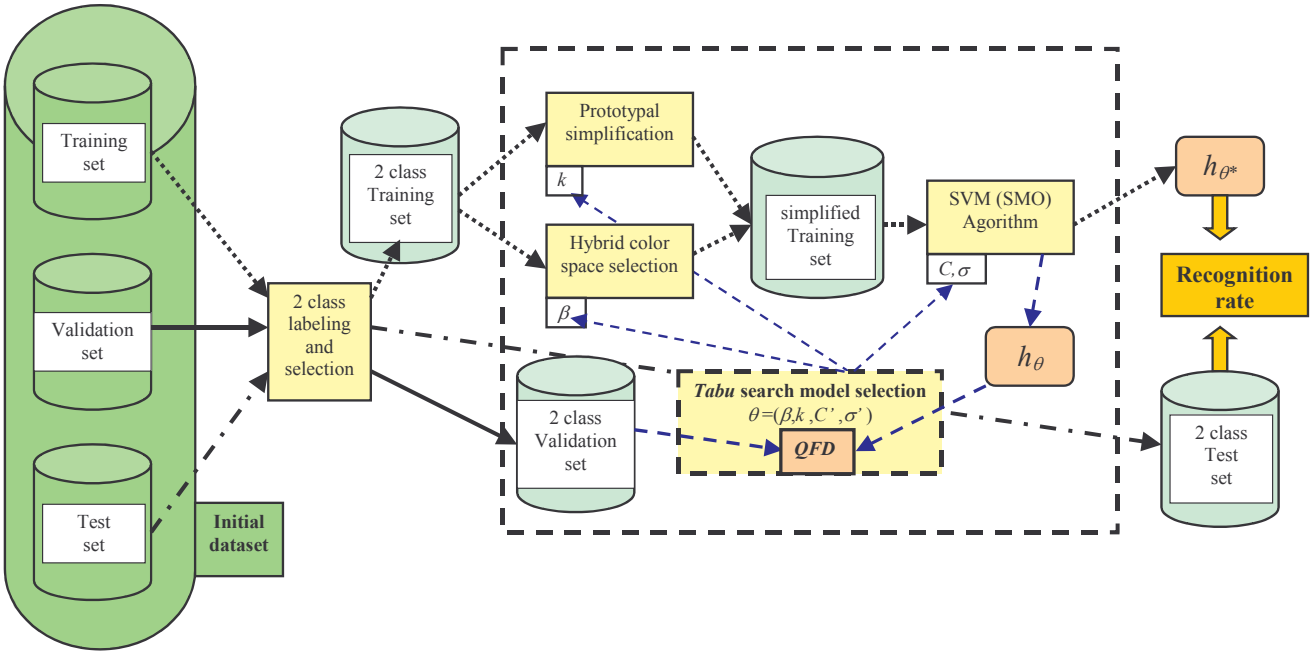

Fig. 2 Synopsis of the new learning training method: selection of binary decision function h_θ which optimize the DFQ criterion.

Our problem is to choose an optimal model (solution) θ which can be represented by a set of integer variables $\theta = (\theta_1, \ldots, \theta_{n'})$ with $\forall i \in \{1, \cdots, n'\}$, $\theta_i \in \{\min(\theta_i), \ldots, \max(\theta_i)\}$. The objective function q_θ to be optimized represents the quality of the binary decision function h_θ. One move in TS corresponds to adding $\Delta \in \{-1, +1\}$ to the value of θ_i, while preserving the constraints on that model. From these constraints, the list of all possible neighborhood solutions is computed. From these possible solutions the one which has the best DFQ and which is not *tabu* is chosen. The set of all Θ_{tabu}^{it} solutions θ which are *tabu* at the it iteration step of TS is defined as following:

$$\Theta_{tabu}^{it} = \{\theta \in \Omega \mid \exists i, t' : t' \in \{1,\ldots,t\},$$
$$\theta_i \neq \theta_i^{it-1} \wedge \theta_i = \theta_i^{it-(t'+1)} \wedge \theta_i^{it-t'} \neq \theta_i^{it-(t'+1)}\}$$

with Ω the set of all solutions and t an adjustable parameter for the short memory used by TS.

9) NEW LEARNING METHOD

Using previously mentioned details of different techniques and theory domains, the synopsis of the new learning method is the following:

1) To build pixel dataset using expert segmentation of several microscopic images.
2) To select a SVM combination multi-class scheme (*c.f.* subsection 3), a software or hardware configuration for hybrid color space preference (κ_i values, *c.f.* subsection 5) and fix values of c_{p1} and c_{p2} constants.
3) For each binary decision function involved in the multi-class scheme the algorithm illustrated in figure 3 is used (see previous subsection for details), then an efficient decision function as regards QFD criterion is produced.
4) For each binary decision function, to determine values of coefficients a_1 and a_2 by minimizing the negative log-likelihood under *2 class validation set* (*c.f.* Fig. 3).
5) At this final step an efficient and low complexity pixel SVM with posterior class probabilities estimation is built (*c.f.* Fig. 4 for illustration).

CELL SEGMENTATION

10) SEGMENTATION SCHEME

The segmentation scheme globally is classical and close to segmentation schemes described in (12,13) for pixel classification. The main difference is that at first step it compute pixel class probabilities directly from color image (with help of hybrid color space) without using previously color smoothing process (13). In the next steps, it use only probabilistic pixel class estimations without returning to initial image color information, in particular for regions growing (whatershed). The segmentation scheme have the following five steps:

1) Color image Pixel classification with posterior probabilities estimation.
2) Smoothing pixel probabilities image with gaussian filter.
3) Extraction of confident marker by using different threshold probability values for each class.
4) Removal of not sufficiently confident markers by morphological operations.
5) Marker growing with watershed using smoothed probabilistic image.

The first step use a fast probabilistic SVM pixel decision function which is built by the new learning algorithm presented in the previous section. The aim of the first step is to capture color information by using adapted hybrid color space for a primary identification of semantic object: background, cytoplasm and nuclei. The next four steps use spatial information for refined shape of identified objects and remove artifact elements. Mucus presence, carbon remains and intrinsic noise in image acquisition system are the principal raison of artefact effect.

Steps 2, 3 and 4 have several parameters which must efficiently be selected to improve the segmentation quality of that scheme. For step 2, the widow size (w_s) of gaussian filter. For step 3, the thresholds (T_i) at which a pixel is belong potentially to a marker. For step 4, the number (n_e) of

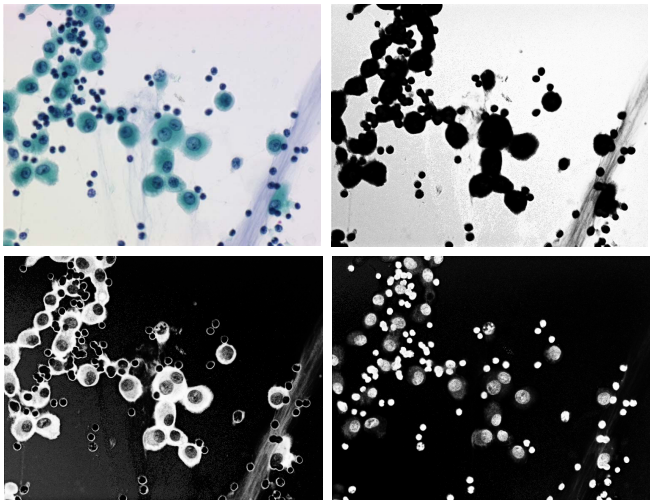

Fig. 4 Pixel probabilities estimation of background (up-right), cytoplasm (down-left) and nucleus (down-right) by decision function produce with our learning method and applied on a cell microscopic image (up-left). Processing time of the whole image take less at 2 seconds.

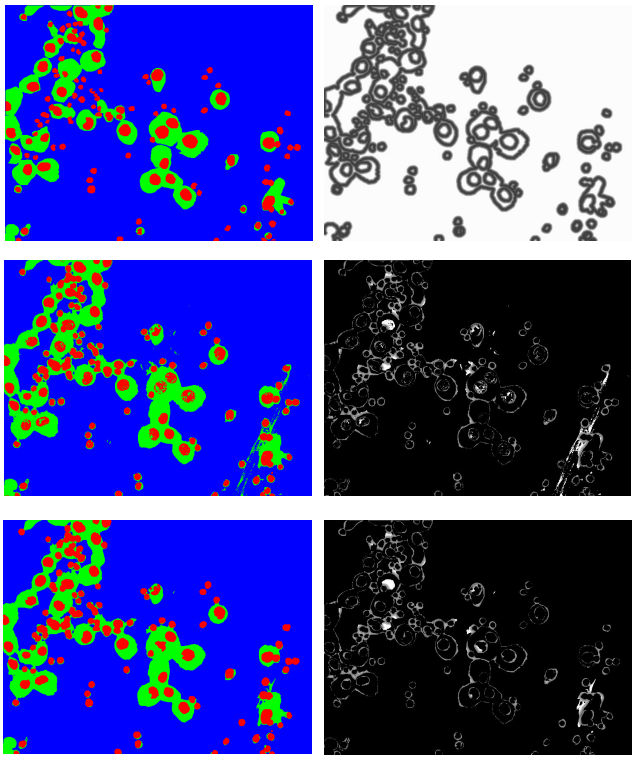

Fig. 5 From expert segmentation (up-left) pixel error weighting map image (up-right) is compute in accordance to q_{shape} formulation. Direct pixel classification image (middle-left) has a $q_{shape} = 0.769$ (middle-right) and image segmentation (down-left) with efficient parameters choice has a $q_{shape} = 0.492$ (down-right). Visual improves of segmentation by comparison of direct pixel classification is also highlight by q_{shape} criterion.

morphological opening operations (those are used in order to remove the too small markers which could have a high probability to being artifacts). All those parameters are chosen by a *tabu* search method which optimizes the q_{seg} cell segmentation quality criterion as defined in the next subsection.

11) CELL SEGMENTATION QUALITY

This new cell segmentation quality criterion (q_{seg}) must take into account the good adequacy (q_{shape}) between the shape of the objects produced by the automatic segmentation (I_a) and those produced by the expert segmentation (I_e). That criterion must also take into account the number of missing objects (n_{missing}) and the number of artifact objects (n_{artifact}). The definition of that criterion is:

$$q_{\text{seg}} = q_{\text{shape}} + \lambda \cdot n_{\text{missing}} + (1-\lambda) n_{\text{artifact}}$$

with $\lambda \in [0,1]$. The constant λ makes it possible to favor a segmentation which limits the number of missing objects as compared to the number of artifact objects and vice versa. In the case of the segmentation of cells, it is essential that no cell is lost, even if that forces to keep some artifacts like bronchial carbon remains (those artifacts are identified by a next module which works on more high level information than pixels), so we have chosen $\lambda = 0.9$.

The good shape adequacy q_{shape} is defined as following:

$$q_{\text{shape}} = \frac{1}{|I_a|} \sum_{\substack{I_a(p) \neq I_e(p) \\ p \in I_a}} \min\left(d_e\left(p, I_e\right), d_{\max}\right)^2$$

where $d_e\left(p, I_e\right)$ corresponds to the distance between the pixel p and the nearest pixel of p belonging to the shape edges in expert segmentation I_e. $d_{\max}$ value aims at restricting the effect of weighting decrease when pixels are close to expert segmentation boundaries. Taking into account the actual size (752x574) of microscopic images, the current value of $d_{\max}$ is 4.

Figure 5 illustrates where the errors of classifications are the least serious in function of expert segmentation.

EXPERIMENTAL RESULTS

12) PIXEL CLASSIFICATION RESULTS

To build the pixel dataset ground truth, 8 microscopic images of bronchial tumors (RGB, 752x574 pixels) with expert segmentation have been used. Pixels corresponding to background, cytoplasm or nucleus objects have respectively class label 1,2,3. Pixel dataset has been split to produce training, validation and test sets by using respectively the 2 first , 2 second and 4 last images. As the number of pixels in each class is not balanced in images (1: 89%, 2: 7%, 3: 4%), only a subset of pixel of classes 1 and 2 was selected by random to build the training and validation sets, so that each class has the same number of examples (around 60000 pixel examples by class). This procedure has also the advantage to speed-up vector quantization process and *tabu* model search.

To reduce biais in model selection the recognition rate (R_R) is evaluated from testing set. For validation or test datasets S recognition rate is evaluated by:

$$R_R(h,S) = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{|S|}{\left|S_{(y_i)}\right| \cdot C_l} l(h(x_i), y_i)$$

with $S_{(y_i)}$ a subset of S which only containe examples of class y_i. In next tables, some abbreviations have been used: Binary Decision Function (BDF), Hybrid Color Space (HCS), Classical Color Space (CCS), Training Time (TT), Classification Time by Image (CTI), hardware (HW), software (SW) .

The first experiment (Table 1) illustrates the automatic tuning of binary decision functions involved in an one-again-all combination scheme (c_{p2}=0.01 and κ_i =1). The first remark is that globally the increase of penality for number of support vectors reduce greatly that one, but recognition rates have very low decrease. So, it is possible to produce efficient decision functions with low complexity. Moreover, the simplification (k) and hybrid color space selection depend on discrimination problems. For example, the discrimination of cytoplasm versus background and nucleus ($h_{(2)}$) is the most difficult. The decrease of penalty has for effect to increase recognition rate near 2.3%, but only by using a lesser-simplified training set. The direct drawback is an expensive increase to the evaluation cost of the decision function.

c_{p1}	BDF	HCS	$\lvert S_v \rvert$	k	R_R
0.003	$h_{(1)}$	RY_1	12	3	96.57 %
0.003	$h_{(2)}$	GLI_2	2623	11	87.02 %
0.003	$h_{(3)}$	B	6	2	89.85 %
0.010	$h_{(1)}$	u^*S	2	1	96.24 %
0.010	$h_{(2)}$	u^*Ch_1	8	3	84.94 %
0.010	$h_{(3)}$	B	2	0	89.56 %

Table 1 Automatic model selection of hybrid color space (HCS) and level of simplification (k) in function of discrimination problem ($h_{(i)}$) and penalty level (c_{p1}). The number of support vectors ($|S_v|$) and the recognition rate (R_R) performance are also recorded.

	$h_{(1)}$		$h_{(2)}$		$h_{(3)}$	
CCS	R_R	$\lvert S_v \rvert$	R_R	$\lvert S_v \rvert$	R_R	$\lvert S_v \rvert$
RGB	95.22 %	2	84,35%	13	89,74%	4
XY_1Z	95.29 %	2	84,46%	12	90,10%	5
L*a*b*	94.86 %	2	83,58%	7	88,68%	2
L*u*v*	**96.41 %**	**4**	84,90%	10	89,25%	8
LCH_1	95.86 %	4	84,76%	60	89,25%	4
$Y_2Ch_1Ch_2$	96.00 %	2	85,74%	46	89,82%	4
$I_1I_2I_3$	95.73 %	2	**85,41%**	**6**	89,76%	4
H_2SL_2	95.93 %	3	84,94%	6	89,90%	9
$Y_3C_BC_R$	95.38 %	2	86,17%	149	**90,13%**	**5**
Average	95.63 %	2,6	84,92%	34,3	89,62%	5,0

Table 2 Recognition rate and number of support vector for decision functions produced by our learning algorithm for 9 different classical color space s (c_{p1}=0.01 and c_{p2}=0.01).

The second experiment (table 2) is similar to the first one, but color space choice is fixed. For each color space training, only the level of simplification and SVM hyper-parameters must be tuned. If maximization of QDF criterion of each binary decision function involved in this one-again-all combination scheme is searched, then it is necessary to use 3 different color spaces (L*u*v*, $I_1I_2I_3$, $Y_3C_BC_R$) and 15 supports vectors. In comparison with the solution produced by an automatic hybrid color space selection (Table 1, c_{p1}=0.01), which uses only 4 color components and 12 support vectors for a recognition rate very close (lower decrease around 0.5%), we can conclude that automatic hybrid color space selection work well, even if classical color space (I_1,I_2,I_3) globally work better ($h_{(1)}$ is a little lower but $h_{(2)}$ and $h_{(3)}$ are a little higher, moreover only three color components are used) for this application with this combinational scheme and penalty configuration.

The third experimentation highlights the influence of combination scheme when processing time is critical (table 3 and 4). One-again-all and pairwise combination schemes produce multi-class decision function with similar recognition rate, but time for classifying all pixels of one image (CTI) is generally more important with one-again-all scheme. The main reason is that for pairwise method each binary problem used only a subset of training set (one class is removed), so discrimination problems are globally easier and have fewer examples. Moreover training time is more tractable with pairwise method, especially when hybrid color space selection must be dealt with. Although, in all case training time is not detrimental, especially in consideration of number of parameters that the model search must to adjust. Another important remark is that image processing time is greatly dependant in color space choice for each multi-class combination scheme. Note that when value of c_{p1} penalty increases, the processing time variation between the two combination schemes is reduced, but pairwise method remains better. To conclude for our segmentation scheme, pairwise combination scheme using Price for probability estimation has been retained.

The last experimental results consider the trade-off variation between recognition rate and processing time when penalty c_{p1} is increased with pairwise mutli-class method (Table 5).

CCS	R_R	$\|S_v\|$	TT	CTI	
RGB	86.55 %	479	2639	10.32	
XY_1Z	86.80 %	1364	12017	29.22	
L*a*b*	86.74 %	745	3856	16.80	
L*u*v*	86.35 %	2680	5761	61.98	
LCH_1	85.97 %	1239	6785	27.40	
$Y_2Ch_1Ch_2$	**87.09 %**	**303**	**6404**	**6.58**	
$I_1I_2I_3$	86.85 %	2589	4760	54.11	
H_2SL_2	86.02 %	2520	2899	55.52	
$Y_3C_BC_R$	86.67 %	519	2668	11.08	
Average	86.56 %	1382	5310	30.34	
HCS	**R_R**	**$\|S_v\|$**	**TT**	**IST**	**κ_i**
RGY_1LI_2	87.18 %	2641	75158	63.97	HW
$RGBY_1L$	86.97 %	2532	76256	54.28	SW

Table 3 Recognition rate, number of support vector, training time and classification time by image for multi-class decision function produce by **one-again-all** combination scheme with classical color space and hybrid color space (c_{p1}=0.003 and c_{p2}=0.01).

CCS	R_R	$\|S_v\|$	TT	CTI	
RGB	86.86 %	988	968	21.38	
XY_1Z	87.06 %	880	1258	20.18	
L*a*b*	**86.69 %**	**50**	**1269**	**1.97**	
L*u*v*	80.00 %	13	517	1.07	
LCH_1	86.24 %	128	812	3.81	
$Y_2Ch_1Ch_2$	84.57 %	35	429	1.42	
$I_1I_2I_3$	86.57 %	851	1488	19.04	
H_2SL_2	86.36 %	146	762	3.81	
$Y_3C_BC_R$	86.28 %	36	731	1.51	
Average	85.63 %	347	915	8.24	
HCS	**R_R**	**$\|S_v\|$**	**TT**	**IST**	**κ_i**
RBu^*Y_3	**86.97 %**	**12**	**4622**	**2.13**	**HW**
$RGBu^*$	**86.13 %**	**11**	**4843**	**1.87**	**SW**

Table 4 Recognition rate, number of support vector, training time and classification time by image for multi-class decision function produced by **pairwise** combination scheme with classical color space and hybrid color space (c_{p1}=0.003 and c_{p2}=0.01).

Although QDF with software preference always produces decision functions which are faster than the ones produced by using QDF with hardware preference, the software preference produces also decision function with little decrease in recognition rate. Therefore, we have selected the pixel decision function of table 5 which is obtained by hardware preference for the first step of our segmentation scheme. To finish with pixel classification results, table 5 shows that for some penalty configuration the best hybrid color space could be the initial RGB color space.

HCS	R_R	$\|S_v\|$	TT	IST	κ_i
Ru^*v^*	**86.67 %**	**10**	**1069**	**1.71**	**HW**
RGB	**85.97 %**	**15**	**1128**	**0.81**	**SW**

Table 5 Recognition rate, number of support vector, training time and classification time by image for multi-class decision function produce by **pairwise** combination scheme with only hybrid color space (c_{p1}=0.01 and c_{p2}=0.01).

13) SEGMENTATION RESULTS

Segmentation results are obtained with the segmentation scheme previously presented. Free parameters of this scheme have been selected by a model selection procedure similar to the one used in pixel decision function selection except that the cell segmentation criterion is the one optimized. For model selection, the four first cell microscopic images of pixel database have been used and their average cell segmentation criterion defines the efficacy of the model. Free parameters have been fixed as follows after model selection: w_s=3, T_1=42%, T_2=64% , T_3=56%, n_e=2.

The results in table 6 correspond to a comparison between optimized segmentation scheme and direct classification (label class of a pixel is the one which has the higher

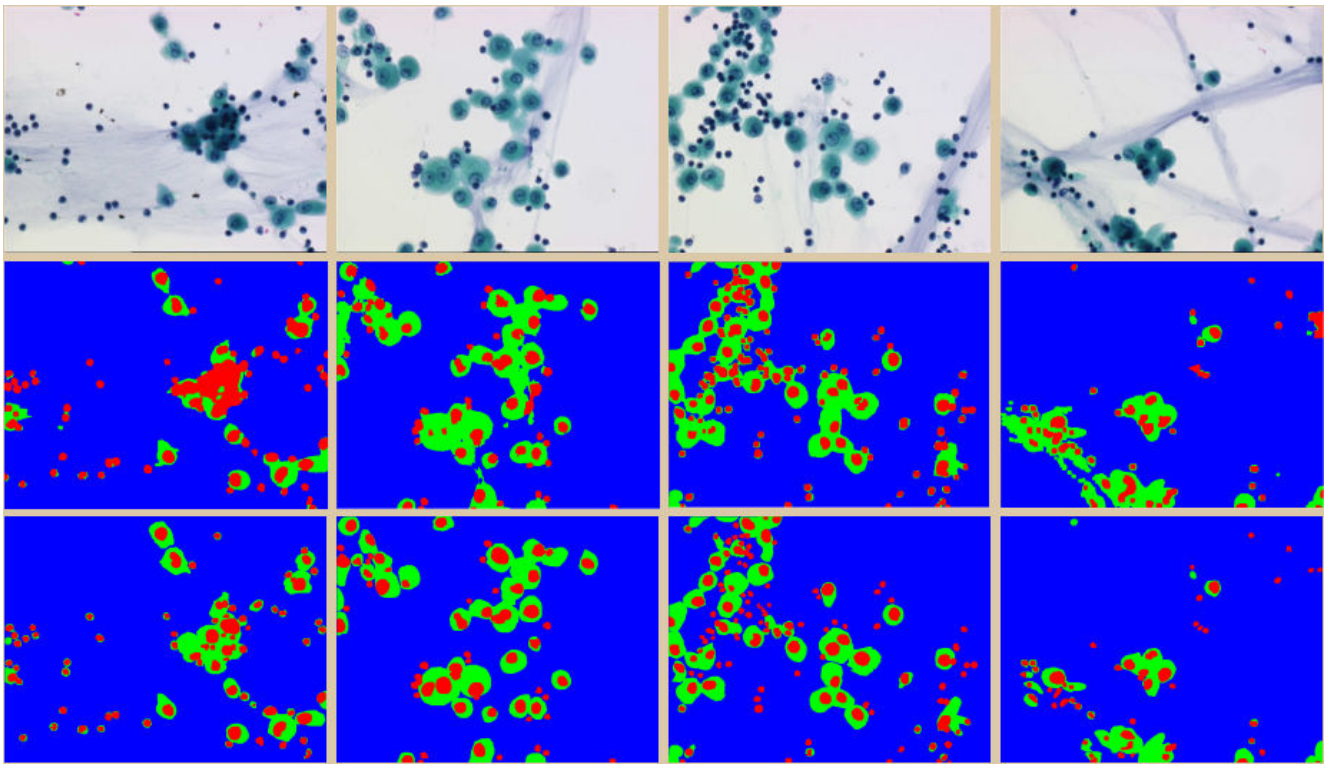

Fig. 6 Cell microscopic images (up), segmentations produce by our method (middle) and expert segmentation (buttom).

probability estimated by pixel decision function). For all images, the segmentation scheme improves cell segmentation quality. Moreover, those results show that cell shapes are improved for all images and at the same time artefact elimination is performed. Those results show the usefulness of spatial information to improve pixel classification as compared to using only color information.

	Direct classification			Segmentation scheme		
Image	q_{seg}	q_{shape}	q'	q_{seg}	q_{shape}	q'
0	17.24	1.04	16.20	**4.60**	**0.80**	**3.80**
1	6.11	1.01	5.10	**4.36**	**0.96**	**3.40**
2	11.00	0.90	10.10	**6.26**	**0.66**	**5.60**
3	24.59	2.19	22.40	**4.89**	**1.49**	**3.40**
4	6.34	0.24	6.10	**2.30**	**0.20**	**2.10**
5	6.42	0.52	5.90	**2.45**	**0.45**	**2.00**
6	7.33	0.73	6.60	**4.58**	**0.68**	**3.90**
7	13.99	0.79	13.20	**3.17**	**0.57**	**2.60**
average	11.63	0.93	10.70	**4.08**	**0.73**	**3.35**

Table 6 Cell segmentation quality (q_{seg}), shape quality (q_{shape}) and miss-artefact trade-off quality ($q' = q_{seg} - q_{shape}$) with the 8 cell microscopic images which have expert segmentation reference for direct classification and segmentation scheme.

Images in figure 6 show segmentation result with our segmentation scheme in comparison with expert segmentation. Globally automatic segmentations have good matching with expert segmentation. Mucus present in all images are correctly identified as background except for some regions which have generally cytoplasm debris produced by medical sampling process (*c.f.* Fig. 7). Nuclei segmentation are mainly well achieved, but some problems remain. When several cells are organized in clusters with overlapping cells, image regions containing the clusters look darker and the cytoplasm color look like nuclei in regions without overlapping. This problem can be avoided by using smear refined technique which limits overlapping. Nuclei

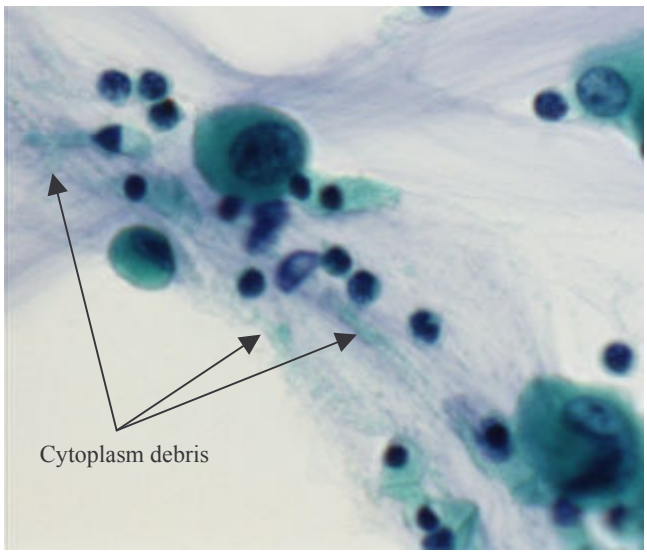

Fig. 7 Cytoplasm debris problem for mucus identification.

segmentation does not work well even when nuclei are very sallow. The main reason is that the chromatin has low density for this type of cell. Most of nuclei pixel have colors similar to

the cytoplasm by transparency effect except nuclide and membrane nucleus regions (*c.f.* Fig 8). Actually this problem can be only corrected by high level module which makes the assumption that a small nucleus which is surrounded by a tall size cytoplasm region could be a sallow nucleus.

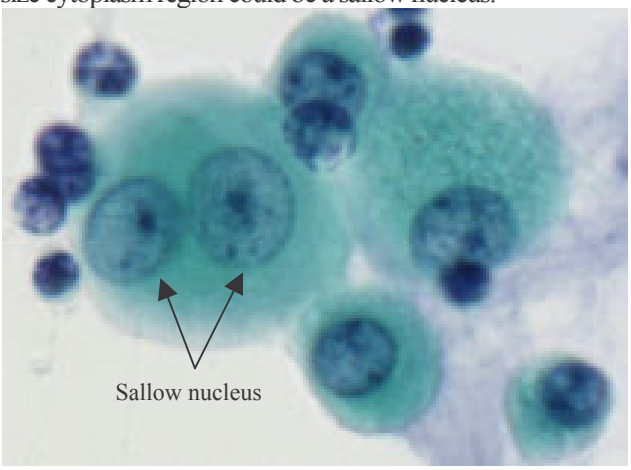

Fig. 8 Sallow nuclei where color of cytoplasm is observed by transparency.

For all images segmentation time is under 3 seconds[3]. Taking into account that great parts of C++ code are not specifically optimized, these results show that fast and efficient segmentation of cell microscopic image can be obtained.

CONCLUSION AND FUTURE WORK

A new learning method is proposed to build fast and efficient decision functions using SVM and estimating pixel class probabilities. A new segmentation scheme using pixel class probabilities has been developed. To tune efficiently free parameters of that segmentation scheme, a new segmentation quality criterion well adapted to characterize cell microscopic image segmentation quality is defined. Experiment results show that segmentation and pixel classification are fast and efficient, in particular when a good combination scheme of binary decision function is used and an adapted color space selected. Resulting segmentation will be helpful for analysis, in particular for cancerous diagnostic-aiding. Automatic hybrid color space selection could be also helpful for testing the segmentation influence of different staining process.

For future work, we have to take into account that even if background identification is achieved, in particular mucus artifacts influence is greatly reduced, the nuclei segmentation of sallow nucleus is problematic. One way we will have to investigate is to split segmentation scheme in two main steps. At first, cell segmentation with no distinction of nucleus or cytoplasm class membership relation. That given a mask where nucleus and cytoplasm object can be found. Secondly, to perform a classification of cells in function of nucleus type (cuts cells which are in contact before, if necessary) and to select an adapted nucleus segmentation scheme (for example: free parameters could have different values as a function of nucleus type). Alternatively, we will classify the small region obtained by an over-segmentation method (partition fine for example) in preference to pixel. The first advantage is that the number of objects to classify is reduced, thus if over-segmentation method is fast enough, it will be possible to speed-up our method. The second advantage is that more features can describe those regions like average and standard deviation of color, size, texture, etc. that could help to classify those regions as background, cytoplasm or nucleus.

To conclude, we have to test theses approaches with several cytopathological specimens and define a global framework to build automatically a segmentation scheme as a function of type of pathology and coloration parameters.

[3] All analyses were performed with a P4 at 2.4 Ghz and with 1 GB of memory.

ACKNOWLEDGEMENTS

The authors thank the reviewers for their valuable comments and suggestions to improve the quality of this paper.

This work was supported by University of Caen Low Normandy under grants of Coeur-Cancer association and the Low Normandy council funds. It was developed in the "Service d'Anatomie et de Cytologie Pathologiques de l'Hôpital Pasteur de Cherbourg". The authors thank their technical staff for time spent to produce manual segmentations of cell microscopic images.

REFERENCES

1. Borsotti, M., Campadelli, P., Schettini, R.: Quanti-tative Evaluation of Color Image Segmentation Results. Pattern *Recognition Letters* 1998, **19**: 741-747.
2. Chang, C.-C. and Lin, C.-J., LIBSVM: a library for support vector machines, *Software Available at* http://www.csie.ntu.edu.tw/~cjlin/libsvm 2001.
3. Collobert, R. and Bengio, S., SVMTorch: Support Vector Machines for Large-Scale Regression Problems, *Journal of Machine Learning Research* 2001, **(1)**: 143-160.
4. Gersho, A. and Gray, R.M., *Vector Quantization and Signal Compression* 1991, **Kluwer Academic edition**.
5. Glory, E., Meas-Yedid, V., Pinset, C., Olivo-Marin, J.-C. and Stamon, G., A Quantitative Criterion to Evaluate Color Segmentations Application to Cytological Images, *ACIVS* 2005, 227-234.
6. Glover, F. and Lagun, M., *Tabu search* 1999, Kluwer Academic Publishers, Boston MA
7. Knesek Jr., E.A., Roche Image Analysis System. *Acta Cytologica* 1996, 40(1): 60-66.
8. Korycinski, D., Crawford , M.M. and Barnes, J.~W., Adaptive feature selection for hyperspectral data analysis, *SPIE* 2004, **5238**: 213-225
9. Kudo, M. and Sklansky, J., Comparison of algorithms that select features for pattern classifiers, *Pattern Recognition* 2000, **33(1)**: 25-41.
10. Lebrun, G., Charrier , C. and Cardot, H., SVM training time reduction using Vector Quantization, *ICPR* 2004, **(1)**: 160-163.
11. Lebrun, G., Charrier, C., Lezoray, O., Meurie, C. and Cardot, H., Fast Pixel Classification by SVM Using Vector Quantization, Tabu Search and Hybrid Color Space. *CAIP* 2005, **LNCS 3691**: 685-692.
12. Lezoray, O. and Cardot, H., Cooperation of color pixel classification schemes and color watershed: a study for microscopical images. *IEEE transactions on Image Processing 2002*, **11(7)**: 783-789.
13. Meurie, C., Lezoray, O., Charrier , C. and Elmoataz, A., Combination of multiple pixel classifiers for microscopic image segmentation. *International Journal of Robotics and Automation, Special Issue on*

Colour Image Processing and Analysis for Machine Vision 2005, **20(2)**: 63-69.

14. Mohr, J. and Obermayer, K, A topographic support vector machine: Classification using local label configurations, *NISP* 2005, **17**: 929-936.
15. Patten Jr., S.F., Lee, S.S.J. and Nelson, A.C., NeoPath AutoPap 300 Automatic Pap Screener System, *Acta Cytologica* 1996, 40(1): 45-52.
16. Platt, J., Fast Training of Support Vector Machines using Sequential Minimal Optimization, *Advances in Kernel Methods-Support Vector Learning* 1999, **MIT Press**: 185-208.
17. Platt, J., Probabilistic outputs for support vector machines and comparison to regulized likehood methods. *In Smola, A.J., Bartlett, P., Schölkopt and Schuurmans, D., eds. Advances in Large Margin Classifiers*, **MIT Press**, 61-74.
18. Price, D., Knerr, S., Personnaz, L. and Dreyfus, G., Pairwise Neural Network Classifiers with Probabilistic Outputs, *NIPS* 1994, (7): 1109-1116
19. Rifkin, R. and Klautau A., In Defense of One-Vs-All Classification, *JMLR* 2004, **5**: 101–141.
20. Vandenbroucke, N., Macaire, L., Postaire, J.-G., Color image segmentation by pixel classification in an adapted hybrid color space: application to soccer image analysis. *Comput. Vis. Image Underst.* 2003, **90(2)**: 190-216.
21. Vandenbroucke, N., Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'image de football, *Thèse de doctorat* 2000, Université de Lille.
22. Vapnik, V.N., *Statistical Learning Theory* 1998, **Wiley edition**, New York.
23. Vinet, L., Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques, *Thèse de Doctorat de l'université de Paris IX Dauphine*, 1991.
24. Wu, T.-F., Lin, C.-J. and Weng, R. C., Probability Estimates for Multi-class Classification by Pairwise Coupling, *JMLR* 2004, **5**: 975-1005.