Evidential segmentation of microscopic color images with pixel classification posterior probabilities

Christophe Charrier, Gilles Lebrun, Olivier Lezoray
University of Caen Basse-Normandie, LUSAC EA 2607, Vision and Image Analysis group
120, rue de l'exode, 50000 Saint-Lo, France
E-mail:{christophe.charrier,gilles.lebrun,olivier.lezoray}@unicaen.fr

***Abstract*— A tool for diagnosis assistance by automatic segmentation of microscopic cellular images is introduced. This method is based on an automatic segmentation technique combining (with the Dempster-Shafer rule) the results obtained by Support Vector Machines (SVM) applied within different color spaces. This combination is performed by integrating uncertainties and redundancies for each color space. Those uncertainties are computed as *a posteriori* probabilities according to the SVM obtained results. An improvement of the final segmentation quality is performed by taking into account the inconsistencies of several pixel classifications.**

***Index Terms*— Segmentation, microscopic cellular images, Classification, Dempster-Shafer combination, SVM**

I. Introduction

Microscopic image analysis is an important task in cytopathology for the detection of abnormal cells in order to establish a diagnosis. Actually, cells are evaluated by a technician during a screening stage. Yet, visual screening stage is a very difficult task and the low number of abnormal cells with regards to the great number of cells implies that the technician concentrates hard. In that case, there can be false-negative cases due to the subjective aspect of screening. To avoid this, one approach consists in helping the technician by developing a semi-automatic screening system to confirm the visual analysis. Such a system is deeply linked to the used segmentation scheme. For microscopic images containing cells, the major problem lies in the spatial and the colorimetric configuration of the nuclei and the cytoplasm. A particular segmentation scheme can perform well for several images but not for others. This is due to the fact that usually only one scheme is applied in order to obtain a segmentation result as close as possible to the ground truth.

From this remark, one intuitive way can be investigated: as one can acheive a good segmentation scheme for several images, one can use other schemes that are working well for others images. In that case, a benchmark of an exhaustive list of segmentation schemes is created. Then, confidence measures can be associated to each one of the schemes of the list. The main drawback of this approach is that one has to be able to modify the confidence measure with respect to the quality of the obtained results for each microscopic image to be segmented. This means that a classification process has been performed with respect to the kind of microscopic images. Finally, the best segmentation scheme is selected with respect to the initial microcopic image group. This solution is time consuming and is an heavy process.

Yet, another way can be investigated: one can select a particular segmentation scheme that performs well, in average. Microscopic images used in this paper have been stained by with a specific preparation, known as the international Papanicolaou coloration: nuclei are blue and cytoplasm are green. Thus, color information included in the image is important and has to be taken into account. From this, the selected segmentation scheme can be performed through different color spaces in order to obtain different segmented maps. From these maps, a fusion process can be achieved to obtain a final segmented image. This method has the advantage to exploit existing redundancies between the segmented maps to increase the quality of the final segmentation map. Thus, the following process is proposed: one particular segmentation scheme based on a pixel classification method is firstly used through different colorimetric transformations. Then, classified maps are fused to obtain a final classified image.

In this paper, the terms "segmentation" and "classification" are indiscriminately used to represent the map obtain after the classification step. Actually, a segmentation can be obtained from a classification map when the connected components of the map are labeled.

Figure 1 shows the proposed segmentation scheme [1]. The first step is to classify pixels through five selected color spaces. Then, *a posteriori* probabilities are computed and can be interpreted as a confidence measure of the classification of a given pixel. The second step is to categorize pixels in 1) *coherent pixels* when all the classifiers select the same class, and 2) *incoherent pixels* when at least one classifier response differs from the others. Only incoherent pixels are processed through a fusion method to select their final class. The final segmentation map results from the union of the two pixel sets.

The paper is organized as follows: in section II, the trial color spaces are discussed. The used segmentation scheme based on a classification approach is provided in Section III. Section IV presents the process applied

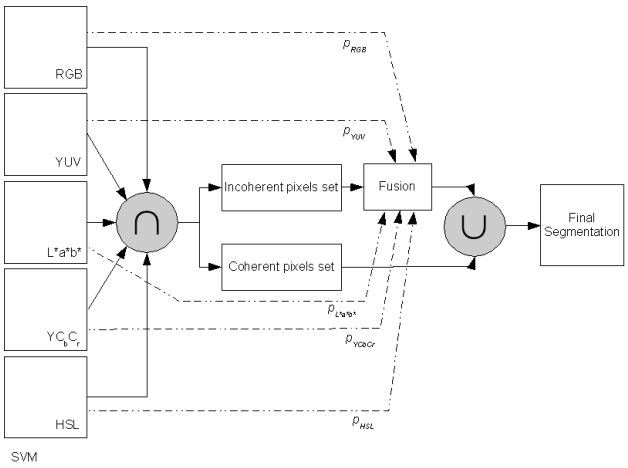


Fig. 1. Synopsis of the proposed segmentation scheme in which p_k represents the *a posteriori* probability computed within the k-*th* color space.

to pixels considered as incoherent ones. The obtained results are presented and discussed in section V. Section VI concludes.

II. On the choice of the trial color spaces

The first step of the proposed scheme concerns the colorimetric transformation of the initial coordinates system, *i.e.*, the RGB space. The question to investigate such transformations is: does exist a color space in which the representation of the color data is the best to optimally perform the segmentation process ? Obviously, many researches have shown that no color space significantly outperforms the others [2]. Nevertheless, each color space has been designed to outperform (under its own hypothesis) the others. For example, the $L^*a^*b^*$ color space is suitable to help a car manufacturer to measure small color differences between two patterns of coachwork of car: are the color of two cars, theoritically identical,visually the same ?

From this remark, the segmentation process is performed through four different colorimetric transformations. The results of this segmentation step are then fused in order to obtain a final segmented image.

One common application of the RGB color model is the display of colors on a cathode ray tube, liquid crystal display or plasma display, such as a television or a computers monitor. Each pixel on the screen can be represented in the computer or interface hardware (for example, a graphics card) as values for red, green and blue. These values are converted into intensities which are then used for display.

By using an appropriate combination of red, green and blue intensities, many colors can be represented. Typical display adapters in 2007 use up to 24 bits per pixel (bpp) that represent 16 777 216 discrete combinations of hue and intensity. It is claimed that the human eye can distinguish as many as 10 million discrete hues (this number varies from person to person depending upon the condition of the eye and the age of the person). However, at the resolution of current screens and at a standard viewing distance people cannot distinguish more than a few hundred hues. Almost all computer monitors around the world use RGB.

The intention of the $L^*a^*b^*$ color space is to produce a color space that is more perceptually linear than other existing color spaces. Perceptually linear means that a change of the same amount in a color value should produce a change of about the same visual importance. When storing colors in limited precision values, this can improve the reproduction of tones. $L^*a^*b^*$ space is relative to the whitepoint of the XYZ data they were converted from. $L^*a^*b^*$ values do not define absolute colors unless the whitepoint is also specified. In practice, the whitepoint is assumed to follow a standard and not explicitly stated (*i.e.*, all ICC $L^*a^*b^*$ values are relative to CIE standard illuminant D50).

Compared to RGB, it is often quicker to make efficient color corrections in $L^*a^*b^*$. The fact that lightness is completely disregarded in the a^* and b^* channels make these much less sensitive to errors. Even though the number of possible numerical values for each pixel is smaller in $L^*a^*b^*$ than for RGB, it is possible to reference a much larger number of colors altogether in $L^*a^*b^*$ - not only colors that cannot be described with RGB, but also sometimes colors that do not appear at all in the real world. In some cases this access to imaginary colors is useful when one goes between several steps in image processing.

The $L^*a^*b^*$coordinates system is the most complete color model used conventionally to describe all the colors visible to the human eye. It was developed for this specific purpose by the International Commission on Illumination (Commission Internationale d'Eclairage, hence its CIE initialism).

The YUV model defines a color space in terms of one luma and two chrominance components. YUV is used in the PAL system of colour encoding in analog video, which is part of television standards in much of the world. YUV models human perception of color more closely than the standard RGB model used in computer graphics hardware, but not as closely as the HSL color space.

The HSL color space (Hue, Saturation, Lightness/Luminance), is quite similar to the HSV space, also known as HSB (Hue, Saturation, Brightness), with "lightness" replacing "brightness". The difference is that the brightness of a pure color is equal to the brightness of white, while the lightness of a pure color is equal to the lightness of a medium gray. The HSL color space is often used by artists because it is often more natural to think about a color in terms of hue and saturation than in terms of additive or subtractive color components.

The YC_bC_r color space is a family of opponent color spaces used in video systems. Y is the luma component and C_b and C_r respectively represent two color difference signals: blue minus Luma (B-Y) and red minus Luma (R-Y). These two components are also called chroma components. It is often confused with the YUVcolor

space, and typically the terms YC_bC_r and YUV are used interchangeably, leading to some confusion. In fact, in this article, the term YUV does not refer to YC_bC_r.

As these color spaces have their own properties, it would be useful to exploit them altogether in a whole segmentation process in order to increase the quality of the results.

III. Segmentation by pixel classification.

From all existing segmentation schemes, an SVM-based technique has been selected due to high classification rates obtained in a previous work [3].

A. SVM Principle

SVMs were developed by Vapnik et al. and are based on the structural risk minimization principle from statistical learning theory [4]. SVMs express predictions in terms of a linear combination of kernel functions centered on a subset of the training data, known as support vectors (SV).

Given the training data $\mathcal{S} = \{(x_i, y_i)\}_{i=\{1,\ldots,m\}}$, $x_i \in \mathcal{R}^n$, $y_i \in \{-1, +1\}$, SVM maps the input vector x into a high-dimensional feature space $\mathbf{H}$ through some non linear mapping functions $\phi : \mathcal{R}^n \rightarrow \mathbf{H}$, and builds an optimal separating hyperplane in that space. The mapping operation $\phi(\cdot)$ is performed by a kernel function $K(\cdot, \cdot)$ which defines an inner product in $\mathbf{H}$. The separating hyperplane given by a SVM is: $w \cdot \phi(x) + b = 0$. The optimal hyperplane is characterized by the maximal distance to the closest training data (see Fig. 2 which presents a 2D illustration of the SVMs principle). The margin is inversely proportional to the norm of w. Thus computing this hyperplane is equivalent to minimize the following optimization problem:

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

where the constraint $\forall_{i=1}^{m} : y_i [w \cdot \phi(x_i) + b] \geq 1 - \xi_i$, $\xi_i \geq 0$ requires that all training examples are correctly classified up to some slack ξ 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 [4] is to maximize the following optimization problem:

$$\mathcal{W}(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \qquad (2)$$

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

The optimal solution α^* specifies the coefficients for the optimal hyperplane $w^* = \sum_{i=1}^{m} \alpha_i^* y_i \phi(x_i)$ and defines the subset SV of all support vector (SV). An example x_i of the training set is a SV if $\alpha_i^* \geq 0$ in the optimal solution. The support vectors subset gives the binary decision function h:

$$h(x) = \text{sign}(f(x)) \text{ with } f(x) = \sum_{i \in SV} \alpha_i^* y_i K(x_i, x) + b^* \qquad (3)$$

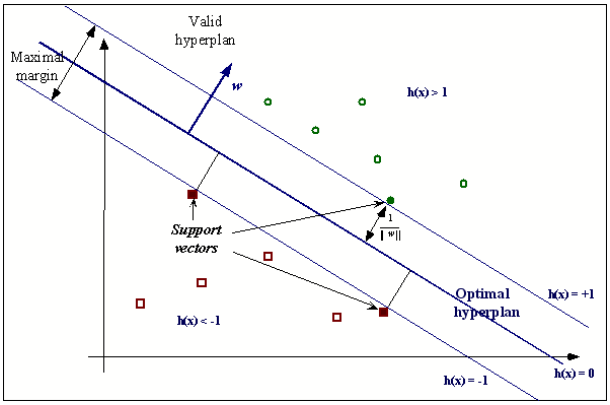
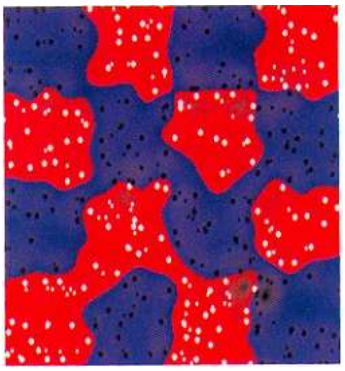


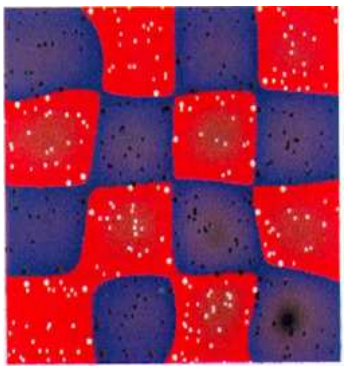
Fig. 2. Synopsis of the SVMs.

where the threshold b^* is computed via the unbounded support vectors [4] (*i.e.*, $0 < \alpha_i^* < C$). An efficient algorithm SMO (Sequential Minimal Optimization) [5] and many refinements [6], [7] were proposed to solve dual problem. SVM being binary classifiers, several binary SVM classifiers are induced for a multi-class problem. A final decision is taken from the outputs of all binary SVM [8].

Figure 3 shows the influence of the parameter σ for a Gaussian kernel function on the shape of the initial decision function. The greater σ, the better the initial decision function.



(a) Small value of σ.



(b) Great value of σ.

Fig. 3. Influence of the parameter σ on the final shape of the initial decision function [9].

B. Computation and use of the a posteriori probabilities

Since SVMs are binary classifiers, the resolution of a multi-class problem is achieved through a combination of binary problems [8]. In our case, three decision functions are created to perform the classification process in three classes {nucleus, background, cytoplasm} corresponding to their one versus all discrimination. Yet, SVMs do not directly provide *a posteriori* classification probabilities. Instead of estimating the class-conditional densities $p(f|y)$, a parametric model is used to fit the posteriori $p(y = 1|f)$ where f represents the uncalibrated output value of SVMs. PLATT [10] has proposed a method to compute the *a posteriori* probabilities from the obtained SVM parameters. The suggested formulae is based on a parametric form of a sigmoïd as:

$$p(y = 1|f) = \frac{1}{1 + \exp(Bf + C)}, \tag{4}$$

where the parameters B and C are fit using maximum likelihood estimation. These parameters are found by minimizing the negative log likelihood of the training data, which is a cross-entropy error function defined as:

$$\min - \sum_i t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \tag{5}$$

where $t_i = (y_i + 1)/2$ represents the target probabilities from a new training set (f_i, t_i), and $p_i = 1/(1 + \exp(Ef_i + F))$. This sigmoid model is equivalent to assume the SVM outputs are proportional to the log odds of a positive example.

A set of segmented images representing a ground truth (*i.e.*, a reference image) is used as training base. Each of the three decision functions is trained on each of the five color spaces. In that case, five segmentation maps are generated where each pixel is associated to an *a posteriori* probability p_k for each class. This probability can be interpreted as a belief measure associated to each class and each color space in conjunction. Each segmentation map (three for each color space) can be considered as an information source that can be imprecise and uncertain. The main idea consists in combining these different sources using the theory of evidence (also known as the Dempster-Shafer theory or the belief functions theory) [11], [12], that yields, on the one hand to combine information from many sources, and on the other hand to process uncertain information.

IV. INCOHERENT PIXEL PROCESSING.

In order to generate the final segmentation map, the intersection of the obtained maps within each one of the trial color spaces is achieved. Only incoherent pixels are processed using the theory of evidence.

A. Elements of theory of evidence.

Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be the set of N possible final classes for an incoherent pixel, called the frame of discernment. In our study, $N = 3$ and Ω corresponds to the three final classes $\{\omega_b, \omega_c, \omega_s\}$ respectively representing the background of the image, the cytoplasm and the nucleus. Instead of narrowing its measures to the set (as performed by the theory of probability constrained by its additivity axiom), the theory of evidence extends on the power set Ω, labeled as 2^Ω, the set of the 2^N subsets of Ω. Then, a mass function m is defined and represents the belief allowed to the different states of the system, at a given moment. This function is also known as the initial mass function $m(\cdot)$ defined from 2^Ω in $[0, 1]$ and coroborating:

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad \text{et} \quad m(\emptyset) = 0 \tag{6}$$

where $m(A)$ quantizes the belief that the search class belongs to the subset $A \subseteq \Omega$ (and to none other subset of A). Subsets A such as $m(A) > 0$ are referred to as *focal elements*. A represents either a singleton ω_j or a disjunction of hypothesis. In the case where the set of hypothesis is exhaustive and exclusive, the mass of the empty set is equal to 0.

Two initial mass functions m_1 and m_2 representing respectively the information providing from two independent sources, can be combined according to Dempster's rule [11]:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K}, \quad \begin{matrix} \forall A \in 2^\Omega \\ A \neq \emptyset. \end{matrix} \tag{7}$$

K is known as the *conflict factor* and represents the discrepancy between the two sources. It corresponds to the mass of the empty set $K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ if the masses are not normalized. VOORBRAAK [13] has justified the Dempster's rule combination even if the normalization step is criticized. The more $k \approx 1$ and the more the combination of the sources is a non sense. When $k = 1$, the fusion process is impossible since the sources are considered in complete opposition. Nevertheless, different solutions have been proposed to process this conflict. SMETS [14] assumes, as DEMPSTER, that the higher the conflict is, the worst the definition of the frame of discernment is. In that case, the value of k simply represents the mass assigned to one or several hypotheses that have not been taken into account.

In this study, the normalization process, *i.e.* devided the mass by the conflict term, is used. In that case, the mass of the empty set is distributed among all the focal elements.

One notes that Dempster's combination, also known as orthogonal sum and written as $m = m_1 \oplus m_2$, is commutative and associative.

After performing the combination, the decision associated to the most "probable" element Ω has to be quantified. Among the existing rules of decision, the most commonly used is the maximum of the pignistic probability. This decision rule, introduced by Smets [15] uses the pignistic transformation that allows to distribute the mass associated to a subset of Ω over each one of its

elements:

$$\text{BetP}(\omega.m) = \sum_{\omega \in A \subseteq \Omega} \frac{m(A)}{|A|}, \forall \omega \in \Omega \quad (8)$$

$|A|$ is the cardinal of A. The decision is executed from the elements of Ω the highest value of which is:

$$\omega^* = \text{Arg} \left\{ \max_{\omega \in \Omega} [\text{BetP}(\omega, m)] \right\}. \quad (9)$$

B. Are the images adapted for the use of the theory of evidence ?

In this subsection, one justifies that the theory of evidence is well adapted to segment microscopic color images. The following remarks can be formulated:

- The images are obtained using the same segmentation technique based on a pixel classification process through different colorimetric transformation. The same regions are observed many times and thus their associated information are redundant,
- Each segmented image yield us to highlight relevant regions that can differ from one image to another one due to its color representation,
- The images are obtained from an acquisition chain composed by an Olympus BX 50 microscope with a Mrzhuser motorized autofocus scanning and a 3CCD JVC KY-F75 camera connected to a computer by IEEE 1394. Before attempting to acquire images, the system needs to be allowed sufficient time to warm up. To determine this thermal equilibrium, a flat field image of a slide has been acquired every 5 minutes during three hours. By computing the difference image between two successive flat field images and taking the mean gray value of the whole difference image, we can plot the time course of color values after the system has been switched on. The data indicates that the thermal stabilization of the system has been located after 90 minutes (only 0.1 mean gray level of difference). The 90 min warming-up period was therefore used in our experiments. The lighting level of the microscope can be modified by the user, but we have fixed it to a constant electric tension which correspond to the D65 illuminant. Once the lighting level has been fixed, the user can adjust the microscope condenser aperture which levels the amount of light passing through the optical lens. A fixed value being more suitable for reproductibility, the aperture has been fixed to 0.25. Therefore, acquiring an image is always performed under constant optical conditions ensuring reproducibility of the further segmentation results. Due to this acquisition step, the obtained images are noisy, *i.e.*, imprecise and uncertain.

Those several constatations let us assume that the frame of the theory of evidence can help us in the refinment of the microscopic color images segmentation process.

C. Mass function design

One of the main drawbacks of the theory of evidence is the design of mass functions: the quality of the fusion process depends on the quality of the mass function. Among all existing modelisations, two models have been compared : 1) the one proposed by DENŒUX [16] has been retained in our study on account of its integration of both the distance to the neighboors and different criteria of neighborhood (*e.g.*, mean luminance, emergence,...) in its definition and 2) the one introduced by APPRIOU [17] based on a likelihood function.

1) DENŒUX*'s model:* The mass $m(\{\omega_j\})$ is defined as a decreasing function of the distance d between the pixel to classify and the barycenter of the class:

$$\begin{cases} m(\omega_l) = \alpha \exp(-\gamma_l d^2) \\ m(\Omega) = 1 - m(\omega_l) \end{cases} \quad (10)$$

where $0 < \alpha < 1$ is a constant computed from the obtained *a posteriori* probabilities provided by the SVMs output for the class ω_l within the trial color spaces. In that case, $\alpha = p_k(\omega_l)$. γ_l depends on the class ω_l and is computed by minimization of an error criterion using the EM algorithm.

2) APPRIOU*'s model:* This model is based on likelihood functions $L(\omega_l|X)$ satisfying three axioms [18]: 1) the consistency with the Bayesian approach, 2) the separability of the evaluation of the hypotheses H_n and 3) the consistency with the probabilistic association of sources S_j.

The a priori probability function $f(X|\omega_l)$ is supposed to be known. The conditional likelihood associated to a pattern x could be defined by $L(\omega_l|X) = f(X|\omega_l)$. In this article, each class is assumed to be modeled by a Gaussian distribution. In that way, the conditional density function is

$$f(X|\omega_l) = \frac{1}{(2\pi)^{p/2} |\sum_l|^{-1/2}} . e^{-1/2(x-\mu_l)^{\mathrm{T}} \sum_l^{-1} (x-\mu_l)} \quad (11)$$

where μ_l represents the mean vector and $\sum_l^{-1}$ the inverse covariance matrix associated to the hypothesis ω_l.

According to [18], one model is the most consistent with the Generalized Bayes Theorem introduced by SMETS. Each information sources S_j is associated to N elementary mass function defined by :

$$\begin{cases} m_{lj}(\omega_n) = 0 \\ m_{lj}(\bar{\omega_n}) = \alpha_{nj}(1 - R_j.L(\omega_n|x_j)) \\ m_{lj}(\Omega) = 1 - m(\bar{\omega_n}) \end{cases} \quad (12)$$

where R_j is a normalization factor defined by

$$R_j \in [0, (\sup_{x_j} \max_{l \in [1,\ldots,N]} \{L(\omega_n|x_j))^{-1}\}],$$

and α_j is a reliability factor depending on he hypothesis ω_l and on the source S_j. If the confidence during the training phase is high, $\alpha_j = 1$. Otherwise $\alpha_j = 0.9$ [19].

A mass m is finally obtained from the orthogonal sum of the bbas $m_l j$:

$$m(.) = \oplus_j \oplus_l m_{lj}(.) \quad (13)$$

The p likelihood functions $L(\omega_l|x_p)$ are estimated by means of the EM algorithm while preserving a same frame of discernment for each information source.

3) design of the mass function: For both the models, the five initial mass functions $(m_k)_{k\in[1,\ldots,5]}$ are generated after computing the q candidate regions to the fusion process and before the decision induced by the majority vote. Thus, considering one segmentation map, one pixel associated to one class from $\{\omega_b, \omega_c, \omega_s\}$ can be associated to a subset of classes corresponding at most to Ω. In order to generate such a subset, the affectation constraint has to be loosened. One way to perform that is to generate an interval computed from the maximum value of the *a posteriori* probabilities to generate the subset A such as:

$$A = \{\omega_l \in \Omega / \max(p_k(\omega_l)) - \delta_k \leq p_k(\omega_l) \leq \max(p_k(\omega_l))\} \quad (14)$$

where $k \in \{1,\ldots,5\}$ and δ_k is an ad-hoc constant depending on the perceptual sensitivity of each one of the five trial color spaces. All the classes for which their probabilities are included within this new interval are considered as candidates for classification during the fusion process.

The five mass functions yield to take into account the associated uncertainty to each one of the segmentation maps. Thus, close classes are brought together in the same focal element, and the final decision is taken only after combining the obtained results from other projections.

Concerning the DENŒUX's model, two distance formulas have been investigated: 1) the Euclidean and 2) the Mahalanobis given by $d = \sqrt{(x-\mu_l)^{\mathrm{T}}\sum_l^{-1}(x-\mu_l)}$.

When using the Euclidean distance, the formulae given in 1994 by the Commission Internationale de l'Eclairage (CIE) within the RGB color space has been used. Even if the computation of the distance between colors has to be performed within a perceptual color space different from the RGB coordinates system, the initial formulae is defined for short distance, and has not been validated for long distance. Nevertheless, this formulae has the avantage to be defined by the CIE and, in that way, guarantees the best perceptual color difference.

V. EXPERIMENTAL RESULTS

A. Segmentation Quality Measure

When a ground truth image is available, quality measures usually integrate at least a factor to take into account the region size and another one to compute the recovery rate between the regions R_i of the segmented image $\mathcal{I}$ and the regions V_j of the groundtruth image $\mathcal{J}$. Among all the proposed quality metrics, the one developed by MARTIN [20] has been used because this metric is insensitive to the granularity variation levels induced by the manual segmentation produced by different experts. Indeed, even if two human observers have the same perceptual organization of an image, they may choose to segment it at different levels (e.g., a bird can be segmented as only one object or as a set of many sub-objects containing the beak and the rest of the body, and so on.). This measure is based on the computed error $E(s)$ on each pixel as:

$$E(s) = \frac{\text{card}(V_j \backslash R_i)}{\text{card}(V_j)} \quad \text{and} \quad E'(s) = \frac{\text{card}(R_i \backslash V_j)}{\text{card}(R_i)}. \quad (15)$$

The disimilarity measure is provided by the local consistency error as the segmentation quality measure:

$$\text{LCE}(I,V) = \frac{1}{h \times w} \sum_s \min\{E(s), E'(s)\} \quad (16)$$

where h and w respectively denotes the image height and width. The lower the LCE value, the better the segmentation quality is.

In addition, a recognition rate (RR) is used to measure the performance of the proposed scheme. This measure allows to know how many pixels have been classified as the expert does.

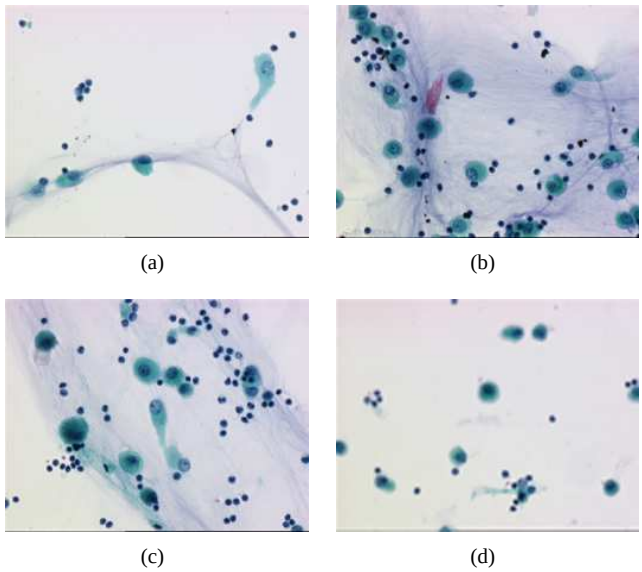
B. Results

The proposed technique has been applied to an image database containing 50 microscopic cell images. Figure 4 presents a panel of four selected images from the database. One observes that the background of the images varies from quite homogeneous (Fig. 4(d)) to highly textured (Fig.4(b)). The number of pixels corresponding to the background, the cytoplasm and the nuclei is not balanced across images: 89% of pixels are located to cytoplasm, 7% concerns the cytoplasm and only 4% of pixels represents nuclei.

Figure 5 shows the difference maps obtained between the classified map and the ground truth one (Fig. 5(b)) from the five trial color spaces (Fig. 5(c) to Fig. 5(g)) and the map of incoherent pixels (Fig. 5(h)) obtained after the intersection of the segmentation maps through colorimetric transformations and before processing the pixels labelled to as incoherent. For Fig. 5(c) to Fig. 5(g), the white pixels correspond to an incorrect classification while the black ones correspond to a correct classification with respect to the ground truth. In Fig. 5(h), incoherent pixels are black ones.

One can observe from Figure 5(h) that major disagreements are very close to cytoplasm and nuclei boundaries. Furthermore, one can observe that concerning the pixels that have been misclassified (white pixels within Fig 5(c) to Fig. 5(g)), the error is the same in each case. In that case, the classification is incorrect but not incoherent since all the five classifiers have selected the same but wrong class.

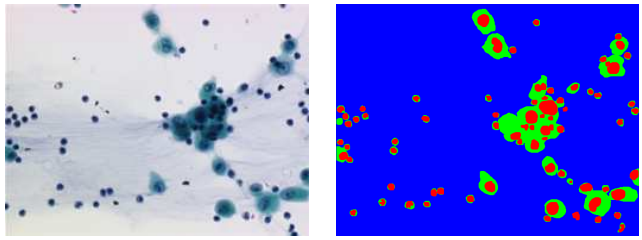
Table I presents 1) the mean (in percentage) of correctly and incorrectly classified pixels and 2) the mean (in percentage) of incoherent pixels for each one of the trial color spaces and all the images of the ground truth database. For example, one can observe that, for $L^*a^*b^*$, only a limited percentage of pixels have been incoherently classified (about 3.38%). The segmentation quality gain can only be obtained from these pixels. Actually, even if 86.52% pixels have been correctly classified, 9.1% remain incorrectly classified. This misclassification cannot be



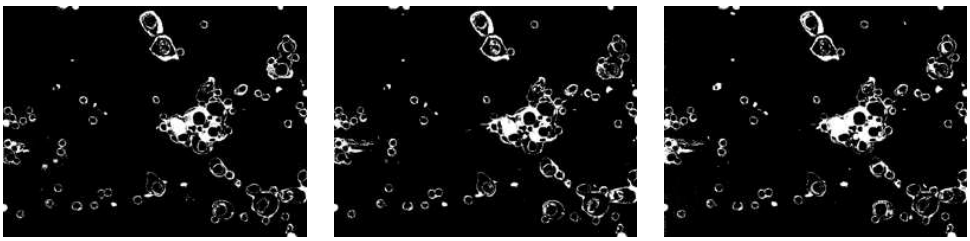
(a) (b)

(c) (d)

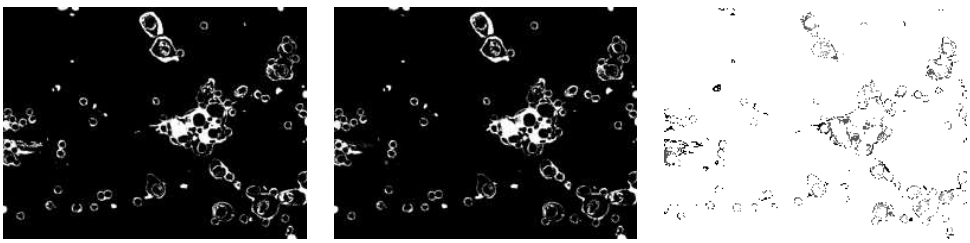
Fig. 4. Panel of four representative images of the used database.



(a) Original Image. (b) The ground truth image.



(c) Segmented map obtained from the $L^*a^*b^*$ color space. (d) Segmented map obtained from the RGB color space. (e) Segmented map obtained from the HSL color space.



(f) Segmented map obtained from the YC_bC_r color space. (g) Segmented map obtained from the YUV color space. (h) Map of incoherent pixels.

Fig. 5. Location of the incoherent pixels after intersecting the five segmentation maps. For Fig. 5(c) to Fig. 5(g), the white pixels correspond to an incorrect classification while the black ones correspond to a correct classification with respect to the ground truth. In Fig. 5(h), incoherent pixels are black ones.

corrected at this stage since all classifiers have selected the same but wrong class. The YC_bC_r color space is the best color space in terms of possible correct classification, since theoritically one can reach a correct classification rate equal to 91.2%. Nevertheless, the spatial distribution of incoherent pixels differs from a color space to another one. In that case, it's quite impossible to choose only one color space. Performing a fusion process applied to the incoherence maps from the five color spaces may be yield us to reach higher classification rates.

	correct	incorrect	incoherent
$L^*a^*b^*$	87.52 $\pm$ 5.97	9.1 $\pm$ 4.21	3.38 $\pm$ 0.96
RGB	78.33 $\pm$ 4.72	10.2 $\pm$ 3.93	11.47 $\pm$ 0.75
HSL	87.11 $\pm$ 5.01	9.5 $\pm$ 4.12	3.39 $\pm$ 0.80
YC_bC_r	87.45 $\pm$ 6.10	8.8 $\pm$ 5.20.	3.75 $\pm$ 0.68
YUV	86.78 $\pm$ 5.87	10.3 $\pm$ 4.25	2.92 $\pm$ 0.84

TABLE I
MEAN OF CORRECTLY, INCORRECTLY AND INCOHERENTLY CLASSIFIED PIXELS WITH RESPECT TO THE GROUND TRUTHS FOR ALL IMAGES AND FOR EACH COLOR SPACES.

	Mean of the RR	Mean of MQ
SVM	87.52 $\pm$ 5.97	0.43 $\pm$ 0.03
Denœux (d_1)	88.99 $\pm$ 5.54	0.40 $\pm$ 0.03
Denœux (d_2)	92.45 $\pm$ 4.78	0.37 $\pm$ 0.04
Appriou	87.68 $\pm$ 5.23	0.42 $\pm$ 0.04

TABLE II
MEAN OF 1) THE RECOGNITION RATE (RR) AND 2) OF THE MARTIN QUALITY (MQ) MEASURE FOR THE SVM-BASED SEGMENTATION SCHEME AND THE PROPOSED ONE BASED ON THE DENŒUX'S MODEL INTEGRATING 1) THE EUCLIDEAN DISTANCE (d_1) AND 2) THE MAHALANOBIS ONE (d_2), AND THE ONE BASED ON THE APPRIOU'S MODEL

Table II shows the mean of the obtained correct classification rates from the database from 1) SVMs, the proposed combination method based on the Denœux's model using 2) the Euclidean distance (d_1) and 3) the Mahalanobis distance (d_2), and the combination scheme based on the APPRIOU's model. SVMs have been trained on a training database where the 20 images are different from those contained within the test database. From the obtained results, one can state that the combination process used in the proposed segmentation scheme outperforms the SVM-based segmentation scheme. Actually, the incoherence is mainly due to a disagreement obtained for the classification of pixels located in cytoplasm or nuclei. That means that for images containing a great number of pixels labeled to as "background", the classification rate is initially high. In that case of images, the expected quality gain will not be significant since only a few number of pixels located around cytoplasm and nuclei boundaries will be labeled to as "incoherent" pixels. In other words, only a few number of pixels will be processed applying the proposed method. The quality gain depends on the number or processed pixels: the higher the number of processed pixels, the higher the quality gain.

Considering the three proposed combination schemes, the one based on the DENŒUX's model implementing the Mahalanobis distance provides the best recognition rate (92.45%). Comparing to the theoritically best score obtained from the "best color space" (91.2% for YC_bC_r), one shows that evidence theory is very promising to segment microscopic color images very carefully. Actually, the obtained recognition rates is greater that the results obtained from only one color space (even if it is the "best").

Since the quality gain essentially concerns the classification of pixels around cytoplasm and nuclei boundaries, the segmentation quality for cytoplasm and thus for nucleus too, increases, and consequently the global segmentation quality.

In addition, using the distance d_2, the mean recognition rate as well as the mean segmentation quality increase according to the use of the distance d_1. This is mainly due to the fact that the distance d_2 takes into account the dispersion of the two compared spatio-colorimetric clouds.

VI. CONCLUSION

A tool for diagnosis assistance by automatic segmentation of microscopic cellular images is proposed. The used segmentation scheme is based on a pixel classification technique developped by VAPNIK known to be the SVMs. This choice is justified by two criteria: 1) a high classification rate and 2) a fast pixel classification process. The main idea of the method is to process the pixels for which at least one disagreement of classification is observed. This allows us to introduce uncertainty on the initial SVM-based segmentation processes. In order to construct different segmentation maps, the SVMs are applied through five different colorimetric transformations. This allows us to take into account the characteristics of each color space. From those five segmentation maps, an *a posteriori* probability is computed which is considered as a belief function associated to each class. Then, an intersection map is generated to detect pixels for which at least one disagreement of classification is observed. Only incoherent pixels are processed under the theory of evidence constraint to determine their final class. Depending on the used distance measure, the final results show that the proposed method outperforms the SVM-based segmentation technique. Results shows that when using a distance that takes into account the neigborhood, the correct classification rate increases.

REFERENCES

[1] C. Charrier, G. Lebrun, and O. Lezoray, "Fusion of SVM-based microscopic color images through colorimetric transformation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006, vol. II, pp. 1112–1115.
[2] V. Meas-Yedid, E. Glory, E. morelon, Ch. Pinset, G. Stamon, and J-C. Olivo-Marin, "Automatic color space selection for biological image segmentation," in *IAPR 17th International Conference on Pattern Recognition*, Aug. 2004, vol. III, pp. 514–517.

[3] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "Fast pixel classification by SVM using vector quantization, tabu search and hybrid color space," in *the 11th International Conference on CAIP*, Rocquencourt, France, 2005, pp. 685–692.
[4] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
[5] J. Platt, *Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods-Support Vector Learning*, MIT Press, 1999.
[6] R. Collobert and S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research*, vol. 1, pp. 143–160, 2001.
[7] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," Sofware Available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm, 2001.
[8] C-W. Hsu and C-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 415–425, 2002.
[9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.
[10] J. Platt, "Probabilities for SV machines," in *Advances in Large-Margin Classifiers*, D. Schuurmans P. J. Bartlett, B. Schlkopf and A. J. Smola, Eds., Neural Information Processing Systems. MIT Press, 2000.
[11] A. Dempster, "Upper and Lower Probablilities Induced by Multivalued Mapping," *Ann. Math. Statist.*, vol. 38, pp. 325–339, 1967.
[12] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
[13] F. Voorbraak, "On the justification of dempster's rule of combinations," *Artificial Intelligence*, pp. 171–179, 1991.
[14] P. Smets and R. Kruse, "The transferable belief model fr belief representation," in *Uncertainty management in Information Systems: from Needs to Solutions*, P. Smets A. Motro, Ed. Kluwer, Boston, 1997.
[15] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," *Uncertainty in Artificial Intelligence*, vol. 5, pp. 29–39, 1990, Elsevier Science Publishers.
[16] T. Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.
[17] A. Appriou, "Multisensor signal processing in the framework of the theory of evidence," *Lecture series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, pp. 5–31, 1999.
[18] A.-S. Capelle, O. Colot, and C. Fernandez-Maloigne, "Evidential segmentation scheme of multi-echo mr images for the detection of brain tumors using neighborhood information," *Information Fusion*, vol. 5, no. 3, pp. 203–216, 2004.
[19] A. Appriou, "Probabilits et incertitude en fusion de donnes multisenseurs," *Revue scientifique et technique de la dfense*, vol. 11, pp. 27–40, 1991.
[20] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images ans its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, Vancouver, BC, July 2001.

Christophe Charrier received his PhD from the university of Saint-Etienne in 1998 in the field of Computer Engineering and Image Processing. From 1998 to 2001, he was a postdoctoral researcher and an associated researcher at the LRTS, Laval University, Québec city, with Pr. Gilles Y. Delisle. He worked both on Radar Cross Section characterization and Color Image Compression. From 2000 to 2001, he worked with the RDDC Valcartier on target recognition task. In 2001, he joined the University of Caen Basse Normandie as an assistant professor. Then from 2003 to 2006, he was the director of the Services and Network Communication as an associate professor. His research concern color image compression, human vision, color image quality and data fusion.

Gilles Lebrun received his PhD from the university of Caen Basse-Normandie in 2006 the field of computer science. He is actually temporary teacher and researcher in computer science in the Services and Network Communication of the university of Caen Basse-Normandie. His research concerns machine learning paradigm and their application to image processing problems. He is the author and co-author of 17 publications.

Olivier Lézoray received the M.S and Doctoral degrees, both in Computer Science, from University of Caen Basse-Normandie, Caen, France, respectively in 1996 and 2000. From 1999 to 2000, he was assistant professor with the Computer Science Department of the University of Caen. Since 2000, he has been an associate professor with the Cherbourg Institute of Technology (University of Caen) in the Communications, Networks and Services Department. His research concern image processing in a discrete settings on graphs and machine learning methods (neural networks and support vector machines) for image data categorization and classification.