FUSION OF SVM-BASED MICROSCOPIC COLOR IMAGES THROUGH COLORIMETRIC TRANSFORMATION

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:{`c.charrier,g.lebrun,o.lezoray`}`@chbg.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.

1. 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 the 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.

Yet another way can be investigated: the combination of obtained results from different schemes. The following process is proposed: different segmentation schemes based on pixel classification methods are firstly used through different colorimetric transformation. The selected color spaces are RGB, $L^*a^*b^*$, YUV, YC_bC_r and HSL. Then classified maps are fused to obtain a final classified image.

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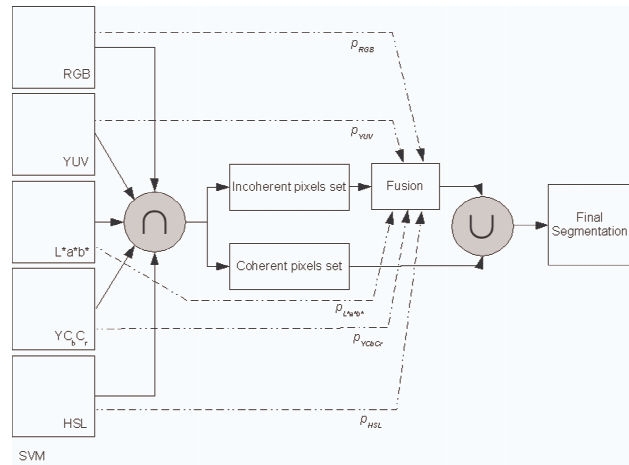
Fig. 1. Synopsis of the proposed segmentation scheme in which p_k represents the *a posteriori* probability computed within the k-*th* color space.

Figure 1 shows the proposed segmentation scheme. The first step is to classify pixels through the 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 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.

2.1. 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 [1].

2.1.1. SVMs Principle

The SVMs developed by VAPNIK ET AL. are based on the structural risk minimization principle from statistical learning theory [2]. 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 $\mathcal{S} = \{(x_i, y_i)\}_{i=\{1,...,m\}}$, $x_i \in \mathcal{R}^n$, $y_i \in \{-1, +1\}$, SVM maps the input vector x into a high-dimensional feature space $\mathbf{H}$ through some non linear mapping functions ϕ : $\mathcal{R}^n \rightarrow \mathbf{H}$, and builds an optimal separating hyperplane in that space. 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. The margin is inversely proportional to the norm of w. An efficient algorithm SMO [3] and many refinements were proposed to solve dual problem.

2.1.2. Computation and use of the a posteriori probabilities

Since the SVMs are binary classifiers, the resolution of a multi-class problem is achieved through a combination of binary problems [4]. 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 provide classification *a posteriori* 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 [5] 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 sigmod as:

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

where the parameters B and C are fit using maximum likelihood estimation.

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) [6, 7], that yields on the one hand to combine information from many sources, and on the other hand to process uncertain information.

2.2. 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.

2.2.1. Elements of theory of evidence.

Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be the set of N final classes possible 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:

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad \text{et} \quad m(\emptyset) = 0 \qquad (2)$$

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 [6]:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K}, \quad \begin{array}{l} \forall A \in 2^\Omega \\ A \neq \emptyset. \end{array} \qquad (3)$$

K is known as the *conflict factor* and represents the discrepancy between the two sources. It corresponds to the mass of the empty set $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$. 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 [8] uses the pignistic transformation that allows to distribute the mass associated to a subset of Ω over each one of its elements:

$$\text{BetP}(\omega.m) = \sum_{\omega \in A \subseteq \Omega} \frac{m(A)}{|A|}, \forall \omega \in \Omega \qquad (4)$$

$|A|$ is the cardinal of A. The decision is executed from the

elements of Ω the highest value of which is:

$$\omega^* = \text{Arg}\left\{\max_{\omega \in \Omega}\left[\text{BetP}(\omega, m)\right]\right\}. \tag{5}$$

2.2.2. 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, the one proposed by DENŒUX [9] 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. Thus 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:

$$\begin{cases} m(\omega_l) = \alpha \exp\left(-\gamma_l d^2\right) \\ m(\Omega) = 1 - m(\omega_l) \end{cases} \tag{6}$$

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.

The five initial mass functions $(m_k)_{k \in [1,...,5]}$ 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:

$$A = \{\omega_l \in \Omega / \max(p_k(\omega_l)) - \delta_k \le p_k(\omega_l) \le \max(p_k(\omega_l))\} \tag{7}$$

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.

In our study, two distance formulas have been investigated: 1) the Euclidean and 2) the Mahalanobis given by $d = \sqrt{(x - \mu_l)^{\text{T}}\sum_l^{-1}(x - \mu_l)}$ where μ_l represents the mean vector and $\sum_l^{-1}$ the inverse covariance matrix associated to the hypothesis ω_l.

3. EXPERIMENTAL RESULTS

3.1. 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 [10] 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:

$$E(s) = \frac{\text{card}(V_j \backslash R_i)}{\text{card}(V_j)} \quad \text{and} \quad E'(s) = \frac{\text{card}(R_i \backslash V_j)}{\text{card}(R_i)}. \tag{8}$$

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

$$\text{LCE}(I, V) = \frac{1}{h \times w}\sum_s \min\{E(s), E'(s)\} \tag{9}$$

where h and w respectively denotes the image height and width.

3.2. Results

The proposed technique has been applied to an image database containing 50 microscopic cell images. Figure 2(b) shows the incoherent pixels obtained after the intersection of the segmentation maps through colorimetric transformations. One can observe that major disagreements are very close to cytoplasm and nuclei.

(a) Original Image. (b) Map of incoherent pixels.
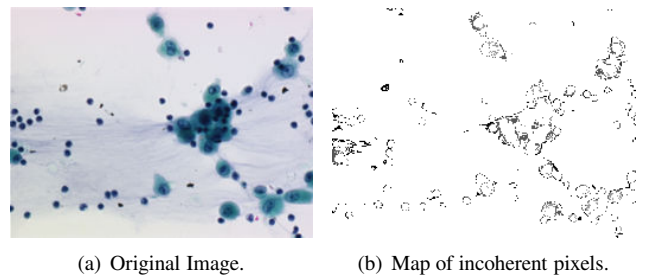
Fig. 2. Location of the incoherent pixels after intersecting the five segmentation maps.

Table 1 presents 1) the mean (in percentage) of correctly and incorrectly classified pixels and 2) the mean (in percentage) of incoherent pixels for all color spaces and for the best

one, *i.e*, the $L^*a^*b^*$ coordinates system. One can observe that, for $L^*a^*b^*$, only a limited percentage of pixels have been incoherently classified (about 3.77%). 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 corrected at this stage since all classifiers have selected the same but wrong class.

All color spaces		
correct	incorrect	incoherent
87.13 ± 6.12	9.4 ± 5.32	3.47 ± 1.13
$L^*a^*b^*$ color space		
correct	incorrect	incoherent
87.52 ± 5.97	9.1 ± 4.21	3.38 ± 0.96

Table 1. Mean of correctly, incorrectly and incoherently classified pixels with respect to the ground truths for all images and for all color spaces (first part) and the best color space (2nd part).

Used method	Mean of the RR	Mean of MQ
SVM	87.52 ± 5.97	0.43 ± 0.03
Combined SVMs (d_1)	87.67 ± 5.62	0.39 ± 0.03
Combined SVMs (d_2)	88.12 ± 5.43	0.38 ± 0.04

Table 2. 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 integrating 1) the Euclidean distance (d_1) and 2) the Mahalanobis one (d_2).

Table 2 shows the mean of the obtained correct classification rates from the database from 1) SVMs, the proposed combination method using 2) the Euclidean distance (d_1) and 3) the Mahalanobis distance (d_2). 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 (whatever the distance formulae used) 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. In that case, since the quality gain essentially concerns the classification of those pixels, the segmentation quality for cytoplasm and thus for nucleus too, increases. 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.

4. CONCLUSION

A tool for diagnosis assistance by automatic segmentation of microscopic cellular images is proposed. The main idea of the method is to process the pixels for which at least one disagreement of classification is observed. The final class for those pixels is determined under the theory of evidence constraint. This allows us to introduce uncertainty on the initial SVM-based segmentation processes. Depending on the used distance measure, the final results show that the proposed method outperforms the SVM-based segmentation technique.

5. REFERENCES

[1] 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.

[2] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[3] J. Platt, *Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods-Support Vector Learning*, MIT Press, 1999.

[4] 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.

[5] 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.

[6] A. Dempster, "Upper and Lower Probablilities Induced by Multivalued Mapping," *Ann. Math. Statist.*, vol. 38, pp. 325–339, 1967.

[7] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.

[8] 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.

[9] 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.

[10] 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.