ACTA STEREOL 1999; 18/1: 1-14
ORIGINAL SCIENTIFIC PAPER

SEGMENTATION OF CYTOLOGICAL IMAGES USING COLOR AND MATHEMATICAL MORPHOLOGY

Olivier Lezoray*[1,3], Abderrahim Elmoataz[1,2], Hubert Cardot[1], Gilles Gougeon[1], Michel Lecluse[3], Hubert Elie[3], and Marinette Revenu[2]

[1]Laboratoire Universitaire des Sciences Appliquées de Cherbourg (L.U.S.A.C), E.I.C, Site Universitaire, B.P. 78, 50130 Octeville, France
[2]Groupe de Recherche en Informatique, Image et Instrumentation de Caen (G.R.E.Y.C), I.S.M.R.A., Bd du Maréchal Juin, 14050 Caen Cedex, France
[3]Department of Pathology, « Centre Hospitalier Louis Pasteur » of Cherbourg, Rue du Val de Saire, 50100 Cherbourg, France
*responsible for correspondence (e-mail: lezoray@info.unicaen.fr)

ABSTRACT

Screening is a manual activity which involves its subjectivity. A semi-automated computer-based system could contribute to the detection of screening errors by the way of a greater reliability. We intend to design such a system operating on color images from serous cytology. The aim of this paper is to present the strategy of the first part of the system : segmentation. It is based on mathematical morphology tools such as watersheds using color information in several color spaces. An extension of watershed to an optimal region-growing operator has been used. A pool of cells has been evaluated by experts to score the segmentation success rate. All cells have been isolated whatever their spatial configuration may be. The average success rate is 94.5% for the nuclei and 93% for the cytoplasm. Our morphological color segmentation of cytological serous images is accurate and provides a good tool for the extraction of cells.

Keywords: cluster division, color, cytology, image analysis, mathematical morphology, region growing, watersheds.

INTRODUCTION

Human screening is a manual activity which has its limits. The aim of this step is the detection of abnormal or suspect cells to set an accurate diagnosis. This manual screening of cytologic slides is described as « intense, particularly complex and whose result relies on the human being interpretation » (Koss, 1989). The relatively small number of abnormal cells overlooked in the screening process involves a high concentration from the cytotechnologist. Because of the screening subjectivity, some errors can occur and cause false-negative

diagnoses. The risk of false-negative slides is currently the main criticism of conventional cytological diagnosis. None of the palliative methods are completely accurate (rescreening of a certain number of slides, screener's training...) (Revision, 1967). A promising approach is to assist the cytotechnologist in finding the abnormal cells in the smear. Many attempts have been studied to use semi-automated screening systems of slides (Tanaka et al., 1977; Tolles, 1955). Those systems aids detecting abnormal or suspect cells in the slides : they allow the recognition of possibly abnormal cells and select the more significant for a cytotechnologist review. Those semi-automated computer-based systems contribute therefore to the detection of screening errors by the way of a greater reliability. Some of those automated rescreening systems have been developed, they are designed to operate in conjunction with a human screening process by rescreening all standard smears that are determined to be normal by cytotechnologists (Knesel, 1996; Patten et al., 1996 ; Rosenthal et al., 1996). Looked at from that point of view we are developing a system (Lezoray et al., 1998) as an assistance to screening using image analysis cellular classification : it is called A.R.C.T.I.C (Aide à la Recherche en Cytopathologie par le Tri Informatique Cellulaire). The particularity of our system is to use color images in various color spaces and in contrary to the existing systems we choose to proceed on serous effusions. In this paper we present the color-based mathematical morphology segmentation method of our system and score the percentage of correctly segmented cells. The segmentation step is the most difficult and the most critical in an automated system. It is intuitive for the human observer to segment an image thanks to the cells color. A machine, however, relies on digital processing techniques to define the set of pixels which stresses each region of the image. Our work proceeds on the elaboration of a segmentation strategy of color images from serous cytology : this strategy relies on region extraction by mathematical morphology using color information in various color spaces. We will describe the whole process performed during the segmentation action so as to obtain cytoplasmic and nuclear regions of the cells from our serous effusions slides.

MATERIALS AND METHODS

The Imaging System

The slides were examined using an Olympus BX-50 microscope, with a 20x objective type, equipped with a Sony 3CCD XC-003P color video camera (Tokyo, Japan). The magnified microscoped image is captured by the three-chip camera and is digitized to a 512x512 pixel color image at a resolution of 24 bits. The pixel separation corresponds to 0.349µm. The digitized images are stored on a Windows NT™ based computer hard-drive and referenced in an ACCESS™ database by : the number of the slide, the code corresponding to the pathology of the cells on the image, comments typed in by a pathologist about the cells, the organ and sample sources. This database has allowed us to test the segmentation process for various cellular types such as normal or abnormal cells, for special spatial configurations such as clusters, overlapping or just-touching cells and for different types of image backgrounds (homogeneous, haemoragic or muco-proteinic).

Image Acquisition

Multispectral images are used for further interpretation. The camera is calibrated with a blanked-field image to ensure correct color registration of the intensity levels of each color channel. The acquisition procedure remains always the same : 20x objective, 0.3 aperture and a stabilized light source for constant and linear illumination of the slide field. The voltage is

maintained to a value of 9V corresponding to the color temperature of 5500K. We choose this voltage according to the microscope and the camera constructors specifications and especially to be in respect with the specifications of the CIE standard illuminates. Before each acquisition, we let turned on both camera and microscope for one hour at least : this is necessary to reach the thermal equilibrium of the system. Through this step, we reduce the noise of the set camera and microscope. We have set and verified this fact by comparing the difference of two images of a same blank field taken at different moments after switching on the elements of our system. Then we have studied the remaining noise which corresponds to a gaussian random variable. This noise will be later corrected during the image processing steps thanks to a smoothing. Along with the noise, there remains some non-uniformity on every acquired field. To correct those inhomogeneities (coming from optical aberrations, focusing misalignment of the light source), we perform a division of each image by a background image. This image is obtained from a selected empty image field. This method corrects the background variations due to sample-dependent factors.

COLOR IMAGE ANALYSIS PROCEDURE

Used Formulas

Color of a point is given by a vector, each component giving the amount of Red, Green and Blue of the point. But there are other three-dimensionnal color spaces based on variables obtained from the RGB color space. In our segmentation we will use the HSL (Hue, Saturation, Luminance) and L*u*v* color spaces. They can be computed with the given below formulas (Pauli, 1975 ; Wyszecki, 1982 ; Luong, 1990) :

$$H_0 = \arccos\left[\frac{((R-G)+(R-B))}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right] \text{ and if } G < B : H = 2\pi - H_0$$

$$S = 1 - \frac{3\min(R,G,B)}{L}$$

$$L = R + G + B$$

$$\text{with } \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.76 & 1.7518 & 1.13 \\ 1 & 4.5907 & 0.0601 \\ 0 & 0.565 & 5.5943 \end{pmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$L^* = 116\,(Y)^{\frac{1}{3}} - 16 \quad \text{if } Y > 0.008856$$

$$L^* = 903.3 * Y \text{ elsewhere}$$

$$u^* = 13L^* (u' - 0{,}21)$$

$$v^* = 13L^* (v' - 0{,}47)$$

$$u' = \frac{4X}{X + 15Y + 3Z} \qquad v' = \frac{9X}{X + 15Y + 3Z}$$

The HSL color space has the particularity to be more intuitive and allows an interpretation closer to the human vision. The L*u*v* color space is perceptually uniform. The use of color in image analysis is recent. In cytology we can find some works using color but it is rather used to characterize a texture (Harms et al., 1986) than to segment an image (Garbay, 1986).

Our approach : Color and Watersheds

Color can be used with watersheds (Beucher, 1990 ; Vincent and Soille, 1991) only on one color feature instead of all color components but a watershed using many color features will lead to finer results. We can state that the use of color versus gray-level watershed methods will give lines corresponding more to significant changes in the images (Ohta et al., 1980). This is not related to the fact that color components give new spatial information but that the information obtained is a little better (Ohta, 1980 ; Luong, 1993). This fact is directly

related to the color gradient obtaining which should lead to better watershed lines. But the principle characteristic of color is that it is directly and naturally related to regions (Luong, 1993 ; Shafarenko et al., 1997). So a watershed will be improved by taking into account those two major facts : local information (from the gradient for instance) and global information (from the regions color homogeneity). This extended watershed has been exposed by Belhomme et al. (1997) and proceeds as an optimal seeded region growing algorithm.

The segmentation principle of the extended watershed we use can be decomposed in two main steps. The first step is a marker extraction to avoid over-segmentation. Color can be used to find markers not only by using one but several color features. The second step consists in building the watershed lines of the marker image. To express the use of global and local information during the flooding process of the extended watershed we use the following potential function :

$I_{X_1Y_1Z_1}$ denotes a color image in the $X_1Y_1Z_1$ color space. Given the vector $\overline{I_{X_1Y_1Z_1}(R)}$ giving the mean color of the region R, the vector $I_{X_1Y_1Z_1}(p)$ giving the color of the point p and $\nabla I_{X_1Y_1Z_1}(p)$ the color gradient at the point p we set :

$$f(p,R) = (1-\alpha)\left\|\overline{I_{X_1Y_1Z_1}(R)} - I_{X_1Y_1Z_1}(p)\right\| + \alpha\left\|\nabla I_{X_1Y_1Z_1}(p)\right\| \quad (1)$$

This function combines local criteria (gradient modulus of a color image) and global information (resulting from statistical comparison between a point p and one of its neighboring regions R). α is a weighting coefficient which allows to modify the relationship and the influence between the local and global criteria during the flooding stage. We have to note the use of color in the expression of $f(p,R)$: we use a color gradient which can be performed in any color space and we use color similarity measures before aggregating a point to a region. This similarity measure is computed in any color space by an Euclidean distance between $\overline{I_{X_1Y_1Z_1}(R)}$ and $I_{X_1Y_1Z_1}(p)$. During the flooding stage, each time a point is aggregated to a region R the vector $\overline{I_{X_1Y_1Z_1}(R)}$ is brought up to date.

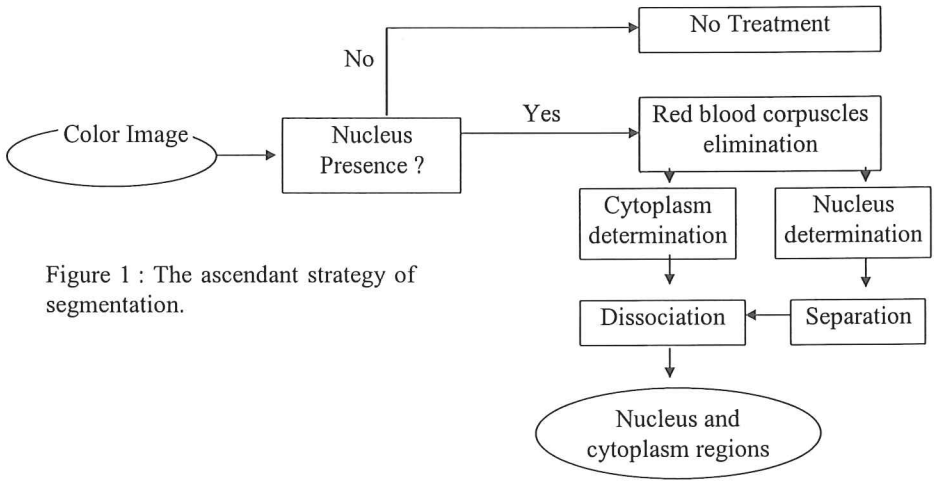

Figure 1 : The ascendant strategy of segmentation.

Finally, contrary to the usual watershed transformation which deals with a fixed potential image, thus constraining to reduce a color image into a single object-representative

component, the extended-watershed allows the processing of global and local information simultaneously by combining them into one function.

AIM OF THE SEGMENTATION OF THE IMAGES

We want to isolate both cytoplasm and nuclei of the cells present on the images : the cytoplasm to get the context information (to characterize isolated or clustered cells) and the nucleus for malignity grading. Both information should lead to a differentiation between the cellular types. To segment the images correctly we have to know precisely the image contexts. Our images are color images which can be divided in three groups according to their backgrounds : homogeneous, haemoragics or muco-proteinics. Cells have a green cytoplasm and a blue nucleus except for the red blood corpuscles which are colored in red. However the spatial configuration and the color of the cells have an extreme variability, it is getting from isolated, just-touching to clusters and overlapping of cells. Our segmentation strategy will proceed in an ascendant way : first we will determine if there are any nuclei on the processed image. In a positive case we will eliminate all the red blood corpuscles because they have no nuclei of interest for the segmentation results and they could be troublesome to determine the cells boundaries. At this step we will separately focus on the cytoplasm and on the nucleus extraction with a color segmentation process. The final thing to do is to separate the touching nuclei and to dissociate the cytoplasm in order to have only one nucleus or nuclei cluster per cytoplasm. The whole strategy is resumed in figure 1.

Nucleated Cells Recognition

The first step in the analysis of an image is to determine if there are nucleated cells of interest on the image. In our case cells of interest are cells with a blue nucleus, so we have to see if there are some blue objects on the images. To perform this, the HSL color space is used and according to the hue intensity, a nucleus presence or non-presence can be decided. A double thresholding of the hue selects the objects having a defined blue hue. The two thresholds have been determined by measuring various nuclei hues. This step doesn't allow to recognize precisely the nucleated cells but it focuses on the blue objects and this focus leads to determine the presence of interesting nucleated cells even if some debris or red blood corpuscles are remaining in the background since they have no real similarities with blue objects.

Red Blood Corpuscles Elimination

From this on we have set up the presence of nucleated cells on the image but we have to recognize the red blood corpuscles in order to eliminate them. We proceed in the same way as in the precedent step : the elements having a hue characterizing red objects are selected. We also perform an opening operation on those objects to refine their boundary definition. The binary mask I_{mask} which is obtained enables to select the image's regions having no red blood corpuscles on it. An example of this mask is given in figure 2 : the image with red blood corpuscles and in figure 3 the corresponding mask. The use of this mask reduces significantly the amount of cells to be segmented but it doesn't penalize the future recognition of the nucleated cells since it doesn't eliminate red blood corpuscles overlapping cells .

Cytoplasm Regions Extraction.

Since the foremost characteristic of the cytoplasm is its green color, the boundary pixels of the cytoplasm are present at the transitions between background and cytoplasm. In other words, boundary pixels normally have higher gradient values than other pixels. In boundary extraction, our approach is to use an extented constrained watershed using a color gradient and cytoplasm markers. This section details markers obtaining, gradient computation and watershed process.

Cytoplasm Markers Obtaining

We want to find one connected element per cytoplasm or cluster of cytoplasm. We proceed by a thresholding on a gray level image obtained from our color RGB image. The difficulty is to find an image which allows a good reproducibility and accuracy of markers obtaining. Those two major constraints are expressed in the image histogram : a good extraction of markers is possible only if the histogram has two peaks and always the same form. Markers obtaining highly depends on the histogram since this one represents the gray level distribution of an image. Finally, our problem is to define an image which has a good gray level distribution (high variation between background and cytoplasm) and which is uniform through all the image we have to segment. To find this image, we have proceeded to the study of all gray level channels of our images in different color spaces (RGB, XYZ, HSL, L*u*v*, L*a*b*). From this study, we set that only the luminance of the HSL color space has a good enough distribution to obtain good markers. What does represent luminance ? Luminance is the sensation to which an area emits more or less light. It is given by an adjective like bright or dark. This corresponds to our images since the background is brighter than the cytoplasm.

After setting the image to be thresholded, we had to find an automatic thresholding method which separates in an accurate way the background from the cytoplasm. To avoid residual noise presence on the luminance image, we perform an exponential smoothing, the resulting smoothed image has sharper and noiseless transitions between background and cytoplasm. Automatic thresholding is a difficult task since the threshold has to give the best results whatever the image is. Several thresholding algorithms were used and compared. The best results were obtained using the variance and the first-valley extraction of the histogram. The final threshold is the maximum of those two latter. This thresholding correctly focus on most isolated cytoplasm or cluster of cytoplasm. One problem remains however : red blood corpuscles have been included in the markers. To eliminate them we perform a morphologic intersection between the markers and the red blood corpuscles mask I_{mask} previously obtained. The resulting image contains only cytoplasm markers without any red blood corpuscles.

Final markers and background are eroded several times to increase their influence zones in order to have a higher regions dissociation.

Cytoplasm Color Gradient Computation

To be able to apply a watershed to segment the cytoplasm, it should be appled on an image where the contours to be calculated correspond to watershed lines and the cytoplasm to catchment basins surrounded by them. The gradient transformation of the original image suits this condition. We use the color gradient in the RGB color space. Gradient is higher at the transition background-cytoplasm since there is a sharp color transition. The gradient was calculated using a 3x3 square structuring element on each of the three channels of the color RGB image. The norm of the gradient is computed as the norm of the three color components.

The retained value is obtained from the maximum of the norm of the gradient according to its direction.

Cytoplasm Extended-Watershed

The final extraction of cytoplasm is obtained by applying the watershed to the color gradient image with the previously obtained markers. The α parameter is set to 1: the potential function uses only the gradient information which is sufficiently accurate for the growing process. Each markers are labeled and the region growing process starts until it reaches a gradient maximum. The resulting image (Fig. 4) is an image of regions giving all the cytoplasm (isolated or clustered). This algorithm correctly segments all cells but is unable to segment distinct cells in clusters since there is no or little difference in intensity between all the cells of the cluster.

Nucleated Regions Extraction

This segmentation step is again based on a color morphologic method using extended constrained watershed. Like cytoplasm, nuclei are recognizable according to their blue color. The determination of the nuclei markers and of the potential function will therefore use this principal knowledge of the objects to isolate.

Nuclei Markers Obtaining

Our aim is to find one connected element per nucleus or cluster of nuclei and this determination is not so easy. There is an extreme variability of nucleated cells on the image because of their texture variations. Those texture and color variations are mainly used by pathologists for classifying the cell. That's why we can have isolated cells with a pale or a dark nucleus (in a blue tone), clusters of cells having very dark nuclei (foremost owing to their overlapping) or chromatin irregularities. This variability makes hard the determination of good nuclear markers since we have to isolate both pale and dark nuclei. On a usual gray level image coming from one color space (whatever it is) this determination is pretty impossible in a reproducible way. The only way to do is to create a synthetic image which emphasizes both pale and dark nucleus. Our approach was to find one image which brings out the brightest blue nuclei and another one which brings out the difference between the nucleus and the cytoplasm and levels the hue difference between all the nuclei. Before obtaining this image, we perform an exponential smoothing on the color RGB image to obtain a noise reduced image. Then we have studied all the gray level channels of our color images in many color spaces and finally we decided to use two images from two different color spaces. The first is v* from the L*u*v* color space and represents a chromatic information of the original color image. The L*u*v* color space has the particularity to be perceptually uniform (Poynton, 1995) and this uniformity allows finer result obtaining, therefore bringing out the brightest nuclei. The second image is the subtracted Blue minus Green image (Ohta et al., 1980) (respectively coming from the RGB color space). This image has the property to emphasize the difference between the nucleus and the cytoplasm but also levels the hue difference between all the nuclei. The final synthetic image is obtained from the mean of those two previous ones :

$$I_S = \frac{\left(I^B_{RGB} - I^G_{RGB}\right) + I^{v^*}_{L^*u^*v^*}}{2} \quad (2)$$

$I^{X_1}_{X_1Y_1Z_1}$ *denotes the* X_1 *component of the image* I *in the* $X_1Y_1Z_1$ *color space.*

This image has a better gray level distribution between all the nuclei and the rest of the image: nuclei are contained in a shortest gray level interval. Therefore it allows the extraction of all the nucleated regions in a reproducible an reliable way.

The marker extraction from the image I_S is performed through an automatic thresholding. Since the histogram has always the same form and since our nuclei are contained in a little interval corresponding to the main peak, the threshold obtaining is more safe. The threshold is chosen from two values : variance of the histogram and first valley to the left of the histogram peak. After using those two thresholds on our I_S images, we notice that in accordance with the image distribution the best threshold is either the former either the latter. To have a quantitative choose between those two ones, we use this property : a thresholded nucleus image is better than another one if it has less cytoplasm included in the markers. The cytoplasm being characterized by its green color, we use the threshold which minimizes the amount of green included in the markers.

The obtained threshold has the property to fit better the nucleus boundary than the other possible thresholds. The image I_S is then thresholded giving the nuclei markers image and we can assume that both pale and dark nuclei have been included in. In those markers some red blood corpuscles may have been included so we proceed to their elimination by a morphological intersection with the I_{mask} image. Markers and background are eroded several times to increase their influence zones.

Nucleus Color Gradient Computation

The gradient is computed in the RGB color space. A gradient image performed on this image has large edge magnitudes at the cytoplasm-nucleus transition. This color gradient image allows therefore a good determination of the watershed process. The gradient is processed in a similar way than for the cytoplasm, we compute its amplitude and keep the maximum in its direction. Direction masks are the same as that for the cytoplasm gradient image obtaining.

Nucleus Extended-Watershed

The final nuclei are obtained by applying an extended watershed to the gradient image with the nuclei markers image. The parameter α has been set to 0.5. This allows to take into account the gradient amplitude but also the color of the point to be added to the nuclei regions. The use of those two information yields better nuclei regions determination because the gradient has some limitations during the growing process. On the first hand the transition between the cytoplasm and the nuclei is not always sharp enough to have high gradient magnitudes and on the other hand some high gradient magnitudes can occur elsewhere than at the nucleus-cytoplasm transition (because of texture variations inside of the nuclei). Therefore the use of the color difference between the point to be added to the region and the color of the region allows a better nuclei regions determination. Finally all nuclei are segmented whatever their spatial configuration could be (Fig. 5). The resulting objects are either isolated, in clusters or in overlapping clumps.

Nuclei Separation

Certain arrangements of cell nuclei have a high complexity, this include touching, overlapping nuclei or nuclei which are so close together that even a human observer would have difficulties in segmenting them. Hence we will separate only the nuclei which could be visually separated, they are recognizable to their concavities in their nucleic boundaries. Our nucleus separation task uses those concavities to dissociate touching cells. For instance let's

take two touching cells and compute the distance function to the border for each point of the nucleus. If we visualize this function we obtain maxima at the center of each distinct nucleus, so the minima are the points of concavities of the touching nuclei. One simple method to separate them is to use a watershed on the inverse of the image distance to the nuclear boundary with the ultimate erosion set of the image distance as markers. With this method the different nuclei tend to separate on the level of the narrowing. By this method a noticeable limit has emerged at the corresponding location where the cells touch themselves (Fig. 6).

Our method works for all the nuclei types described above. In the other cases nuclei are not possibly separable because of the overlapping of many nuclei. At this step all nuclei have been determined, they are either isolated or in overlapped clusters.

Cytoplasm Division

We proceed here at the final step of our segmentation strategy : we wish to divide all the cytoplasm so that they contain only one nucleus or one cluster of nuclei, which is not really the case actually because of touching cytoplasm and of the nucleus separation step. The lines of separation are the points where the transition between two touching cytoplasm is outlined . But the most difficult thing is to respect those transitions for the division of the cytoplasm. We can only use a gray-level method based on one color channel but there is a real problem: when a cytoplasm cluster has no visible transitions between the nuclei, the separation cannot be satisfactory unless we use the positions of the nuclei in the cytoplasm. Therefore the separation method must take care of those two main facts. A way to do is to use a watershed with one color channel and nuclei markers. We use the smoothed blue channel as a potential function and the previously obtained nuclei as markers to perform the watershed. This method has two main advantages. The use of the nuclei for markers allows to take care of their spatial configuration and position in one cytoplasm. On the other hand the use of the smoothed blue channel permits to take into account the cytoplasm form and internal local transitions. By this treatment cytoplasm of many nuclei with real concavities or sharp green transitions will be separated as well as cytoplasm without any semantic information related to their form or intensity transitions but with distinct nuclei inside.

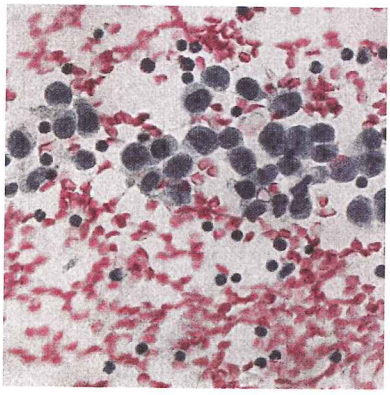

Figure 2 : The original color image

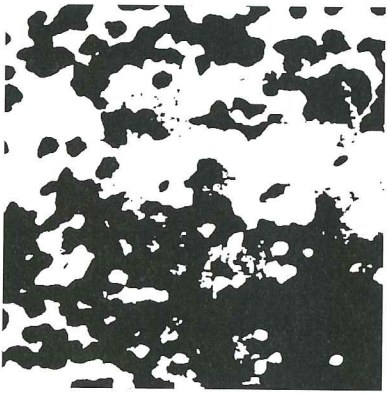

Figure 3 : The Binary red blood corpuscles mask.

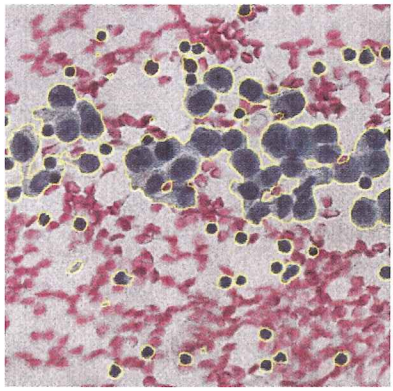

Figure 4 : The Cytoplasm Regions.

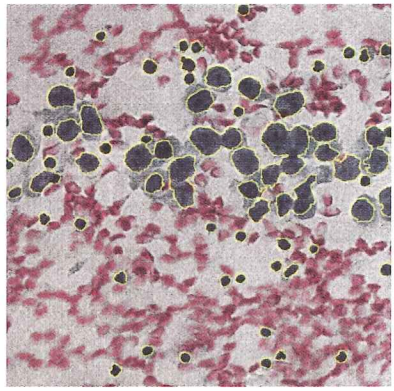

Figure 5 : The Nuclei Regions.

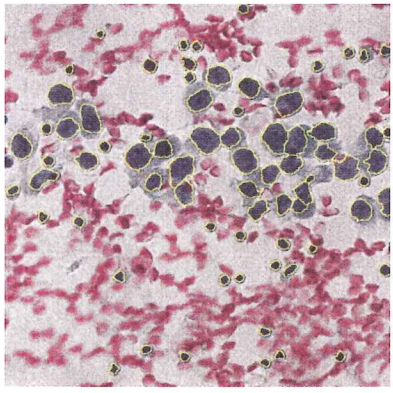

Figure 6 : The Separated Nuclei Regions.

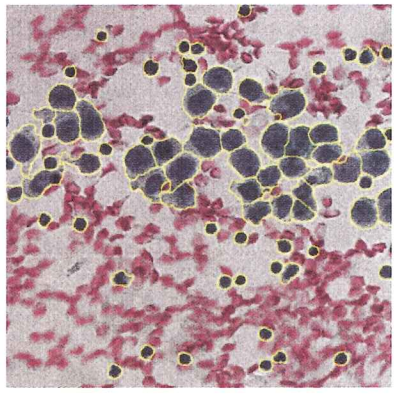

Figure 7 : The Divided Cytoplasm Regions.

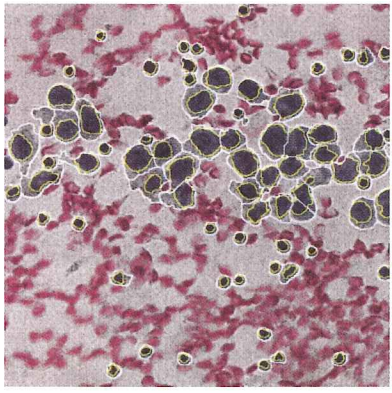

Figure 8.

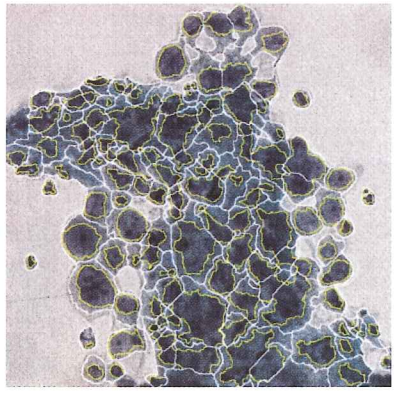

Figure 9.

Final segmentation of touching or isolated cells (Fig. 8) and of clustered cells (Fig. 9). Cytoplasm and nuclei regions are respectively outlined in white and yellow.

This step ends the segmentation strategy. With this reproducible and systematic method we accurately separate all the distinct nucleus or nuclei clusters as to be included in exactly one (Fig. 7).

DATA COLLECTION

In order to test the segmentation of the color images, approximately 2000 cells from serous slides are used. These slides are obtained from the « Laboratoire d'Anatomie et de Cytologie Pathologiques » of the « Centre Hospitalier Louis Pasteur de Cherbourg ». The classification of the slides is known in advance since they have been prepared and screened by a pathologist. The images of the cells have been randomly taken on areas of the slides. Each image is manually focused to obtain the greatest contrast in the image as seen on the monitor (although an automated focusing would have produced a better focus). Color images are acquired and processed by our so called A.R.C.T.I.C segmentation program. The resulting detected cells boundaries are overlaid on the original images and displayed on the monitor for visual inspection.

VISUAL INSPECTION

A method to evaluate the accuracy of a segmentation process is to determine manually the boundaries of all the cells (cytoplasm and nucleus) and to compare the results with the segmentation by counting the incorrectly segmented pixels. This method is very tedious and time-consuming. So we decide to process in a different way : the pathologist visually inspects the boundaries found overlaid on the monitor but he has to choose only the incorrectly segmented or missing cells. For those cells he determines their correct boundaries for the nuclei or/and the cytoplasm. This method is objective and allows to define the segmentation success rate : the percentage of well segmented cells.

All segmented cells are examined by the pathologist even the clusters of cells or overlapping cells. However those cells are considered to be correctly segmented since a human observer could not visually segment them, but they are marked by the pathologist as non segmentable clusters.

RESULTS

In our experiment, segmented cells have been inspected by three experts. All nucleated cells were correctly detected whatever their spatial configuration may be (isolated, just-touching, clusters or overlapping cells). No debris or red blood corpuscles were segmented : all the final objects we obtained with our segmentation strategy are cells. Thanks to the visual inspection process, all incorrectly segmented cells have been listed. The results of the segmentation process for each individual expert are presented in Table 1. The percentage of correct segmentation for the cells range from 89.2% to 98.3% for the nuclei and from 88.7% to 99.1% for the cytoplasm. We note here a noticeable variation between the experts (nearly 10% of difference between all of them) but it is considered to be normal in view of the difficulty of appreciation for minimal errors and for extreme cases.

This fact proves that although a visual inspection of the expert leads to an objective measure of the segmentation error rate, there is still subjectivity intrinsically contained in the human visual inspection: we reach the same limitations than the usual manual screening.

Table 1. Rates of segmentation results for the nuclei and the cytoplasm of serous cells.

Expert N°	Correct Nuclei / %	Correct Cytopl. / %	Nuclei Seg. Errors / %	Cytopl. Seg. Errors / %
1	98.3	99.0	1.7	1.0
2	96.1	91.5	3.9	8.5
3	89.2	88.7	10.8	11.3
Average	94.5	93.0	5.5	6.9

The origins of nuclei segmentation errors are largely due to the lack of contrast between the nucleus and the cytoplasm or similarly to an imperfect focus : an automatic focus will avoid some of those segmentation errors.

The origin of cytoplasm segmentation errors is related to the proximity of the mucus : since it has the same color and texture as a cytoplasm, it is included in the segmented cytoplasm. Other cytoplasm segmentation errors are related to two main configurations : a non-linearity in the definition of the cytoplasm (holes or tearing of a part of it) or a cluster of overlapping cytoplasm with no nuclei superposition. In this last case the cytoplasm division takes advantages of the nuclei distribution but not really respects the cytoplasm transitions since they are hidden by the overlappings. The expert extrapolates because he can focus at different point of view which cannot be done by the computer. Those errors are not really important since they occur only on the overlapping clusters of cytoplasm.

DISCUSSION

Computer-based systems for an automatic cellular classification offer a powerful technique to isolate and recognize cells of serous cytology slides. The use of color in extended watersheds has been demonstrated in this paper. Color has multiple advantages : as seen before, a simple thresholding operation is sufficient for eliminating all the red blood corpuscles and the channels of different color spaces offer a different representation of useful information. Our method uses all those possibilities and couples them to an extended powerful mathematical morphology tool: watersheds. This process successfully detects cells whatever their spatial configuration may be (94.5% of nuclei and 93% of cytoplasm correctly segmented).

To conclude, we notice that the variability of the success rate of the segmentation highly depends on the slides cellular populations. If there are a lot of isolated or touching cells, our strategy will be very efficient. But for more complex configurations, the success rate is not really representative since it is even difficult or impossible for a human observer to segment the cells in an absolute way. From this point of view, our system is different from the existing systems. We do not choose to segment only the isolated cells (Figure 8) : all objects are segmented even the most difficult clusters (Figure 9).

ACKNOWLEDGMENTS

This work was funded by «La Ligue Contre le Cancer (Comité de la manche)» in a Ph.D. thesis. It is undertaken under the auspices of the «Pôle Traitement et Analyse d'Images de Basse-Normandie». It is developed more precisely at the «Laboratoire Universitaire des Sciences Appliquées de Cherbourg» and at the «Service d'Anatomie et de Cytologie Pathologiques de l'Hôpital Pasteur de Cherbourg».

REFERENCES

Belhomme P., Elmoataz A., Herlin P. and Bloyet D.: Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images. In: Journal of Microscopy, 186, 1997: 41-50.

Beucher S. : Segmentation d'images et morphologie mathématique, Ph.D. thesis, Ecole Nationale des Mines de Paris 1990.

Garbay C. : Image structure representation and processing : a discussion of some methods in cytology. In :IEEE Trans. On PAMI 8:2, 1986:140.

Harms H., Günzer U. and Aus H.M. : Combined local color and texture analysis of stained cells. In : CVGIP 33, 1986:364-376.

Knesel Jr. E.A. : Roche Image Analysis System. In : Acta Cytologica 40 :1, 1996:60-66.

Koss L.G. : The Papanicolaou test for cervical cancer detection : a triumph and a tragedy. In : J. Am. Med. Assoc 261, 1989:737-743.

Lezoray O., Elmoataz A., Cardot H., Gougeon G., Lecluse M., Elie H. and Revenu M. : Color Images Segmentation : Application to serous effusions cytology for Computer Cellular Sorting. In : Proceedings of RFIA'98, Clermont-Ferrand, 2, 1998 : 287-295.

Luong Q-T. : color in computer vision. In : Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing, 2.3, 1993 : 311-368.

Luong Q-T. : La couleur en vision par ordinateur : 1-une revue. Research report INRIA n°1251, 1990.

Ohta Y., Kanade T. and Sakai T. : Color Information for regions segmentation. In : Computer Graphics and Image Processing, 13, 1980 : 222-241.

Patten Jr. S.F., Lee S.S.J. and Nelson A.C. : NeoPath AutoPap 300 Automatic Pap Screener System. In : Acta Cytologica 40: 1, 1996 : 45-52.

Pauli H. : Proposed extension to the CIE recommendations on uniform color spaces, color difference equations and metric color terms. In : J. Opt. Soc. Am. 66 :8, 1975 :866-867.

Poynton C.A. : Frequently Asked Questions about Color, Internal report, In : <http://www.inforamp.net/~poynton>, 1995.

Revision of Laboratory requisitions for medicine, medicaid and clinical laboratories improvement act of 1967 CLIA'67 programs. Federal Register, 5 :9600, 1967.

Rosenthal D.L., Acosta D. and Peters R.K. : Computed-Assisted Rescreening of Clinically Important False Negative Cervical Smears Using the PAPNET Testing System. In : Acta Cytologica 40 :1, 1996 :120-126.

Shafarenko L., Petrou M. and Kittler J. : Automatic Watershed Segmentation of Textured Color Images. In : IEEE transactions on Image Processing, 6 : 11, 1997 :1530-1543.

Tanaka N., Ikeda H., Ueno T., Mukawa A., S Watanabe., Okamoto K., Hosoi S. and Tsunekawa S. : Fundamental study of cyto-screening for uterine cancer : new system

for automated apparatus (CYBEST) utilizing pattern recognition method. In : Acta Cytologica 1977 :21:85.

Tolles W.E.: The Cytoanalyser : an example of physics in medical research. In :Trans. N.Y. Acad. Sc. 17 :1955 :250.

Vincent L. and Soille P. : Watersheds in Digital Spaces : An efficient Algorithm Based on Immersions Simulations. In : IEEE transactions on PAMI, 13 :16, 1991 :583-598.

Wyszecki G., Styles W.S. : Color Science : Concepts and Methods, Quantitative Data and Formulae. John Wiley & Sons. Second Edition . New-York, 1982.

Presented at the 7th European Congress for Stereology, Amsterdam, April 20th to 23rd, 1998.