Automatic cell nuclei counting: a protocol to acquire images and to compare results between color and multispectral images

Mohamed Bouzid, Ali Khalfallah, André Bouchot, Mohamed Selim Bouhlel, Franck Marzani

▶ **To cite this version:**

Automatic cell nuclei detection: a protocol to acquire multispectral images and to compare results between color and multispectral images

M. Bouzid[a,b], A. Khalfallah[a], A. Bouchot[c], M.S. Bouhlel[a], F.S. Marzani[*b]

[a] Sciences and Technologies of Image and Telecommunications (SETIT) research unit, Sfax University, Higher Institute of Biotechnology of Sfax, route de Soukra km3, 3038 Sfax, Tunisia;
[b]Le2i Lab., UMR CNRS 6306, Burgundy University, B.P. 47870, 21078 Dijon Cedex, France;
[c]CellImaP, Federative Structure for Research 100, Burgundy University, B.P. 47870, 21079 Dijon Cedex, France

ABSTRACT

High-throughput screening in histology and analysis need a necessary automatic cell or nucleus counting. Current methods and systems based on grayscale or color images give results with counting errors. We suggest to use multispectral imaging (with more than three bands) rather than color one for nucleus counting.

A traditional acquisition chains uses a source of white light and a CCD camera in addition to the optical microscope. To pass to a multispectral acquisition, we use a Programmable Light Source (PLS) in the place of the white light source. This PLS is capable of generating different wavelengths in the visible spectrum. So, one color image and four multispectral images have been acquired from histological slices. The four multispectral images contain respectively 3 bands, 5 bands, 7 bands and 10 bands.

To make a proper comparison of data, several considerations have been taken, like camera linearity, intensity difference between the wavebands from the PLS and non uniformity of the light intensity range in the images. So, a set of measures were done for calibrating the system.

An automatic detection method based on segmentation by expectation-maximization and ellipse fitting is used. An extension of this method is proposed in order to be applied to multispectral images. The original and the extended method are then applied to the data previously acquired to have first results regarding the effect of using multispectral images rather than color ones.

Keywords: Automatic cell nuclei detection, multispectral imaging, color imaging, acquisition protocol, image calibration.

1. INTRODUCTION

Histology is the study of organism's tissues and also cells shape composing these tissues. To achieve this study, tissues must be included in resins such as paraffin or tissue Tek, cut in thin section of about 5µm, and colored (Hematoxyline Eosine stain, Masson's Trichrome stain, ...) or immuno-labeled. Histological analysis is performed by examining these sections of tissue under an optical microscope [1]. The study of histology images is implied in the clinical diagnosis of cancers and identification of prognostic, and in the research field in cancer (but not only) model in animals and therapeutic pathways. For cancer diagnosis, histopathologists visually examine the regularities of cell shapes and tissue distributions, then decide whether tissue regions are cancerous, and determine the malignancy level. Such histopathological study has been extensively employed for cancer detection and grading applications [2]. But this manual analysis of histology tissues that is still the primary diagnosis method in the clinical field depends on the expertise and experience of histopathologists. Such manual intervention has the disadvantages of (a) being very time consuming and (b) showing intra- and inter-observation variations among experts. Computer-based image analysis has become an

---

[*] franck.marzani@u-bourgogne.fr; phone +33 380 39 63 33; fax +33 380 39 59 10; le2i.cnrs.fr

increasingly important field because of the high rate of production and the increasing reliance on these images by the biomedical community. Histology images have a large amount of objects of interest (cells, cell compartments such as nuclei…). The first goal of computer-based image analysis is to have a robust and accurate approach for identifying nuclei. Although, in literature [3, 4], a rich set of nuclei segmentation methods has been developed, almost none of them is fully sufficient to handle microscopy images where an enormous number of objects of interest may exist in various forms with large variations.

Many color segmentation algorithms can be found in literature. Color being directly and naturally attached to the regions of an image [5, 6], color segmentation methods might give better results than monochromatic segmentation methods. Color segmentation raises problems related to the representation and the processing of color. One can mention the histogram analysis [7], the pixel classification [8] and the region-growing methods [9].

This paper proposes an efficient method for nuclei segmentation that uses both ellipse fitting and Expectation-Maximization (EM) segmentation. It is first applied on color images.

Multispectral imaging is a new technique; it uses precise measurements of optical spectra at each pixel of an image. Compared to the color image, the multispectral image can provide more information for each pixel.

The multispectral dimension can be achieved by placing many filters between the sample and the monochromatic sensor. Each filter provides specific information about the sample in the spectral band considered. Such a device can sample the visible spectrum with a higher resolution than a color camera, and avoid the subjectivity of color imaging. The spectral dimension comes from a filter system positioned in the optical path of light. Therefore, a multispectral acquisition can be done by using filters between the light source and the sample or between the sample and the sensor. The technology we used to perform the filtering is based on the use of a Programmable Light Source (PLS).

The method developed and used for nuclei segmentation based on color images has been extended to be applied to multispectral images. Its results, both on color and multispectral data, are compared.

The remainder of this paper is structured as follows. In section 2, we present on one hand the protocol and the acquisition system used, and on the other hand the method chosen to detect the nuclei both on color and multispectral images. In section 3, the experimental results for real images are presented in detail. Finally, we give conclusions in section 4.

2. MATERIAL AND METHOD

2.1 Data acquisition

Cell imaging platform (CellImaP) of SFR 100, University of Burgundy, currently has a conventional acquisition of cell images (Cell Observer station from Zeiss). This system is essentially composed of an optical inverted microscope equipped with a white light source illuminating the sample (halogen source) and a CCD camera for the acquisition. We use a Programmable Light Source (PLS) that we place in the excitation part of the station. Unlike liquid crystal filters, the PLS is capable of generating rays of the visible spectrum with flexibility (amplitudes and widths of the bands).

To connect the PLS, we designed a mechanical part between the optical fiber end of the PLS and the optical inverted microscope. The dimensions of this piece were taken with good precision for securing the fiber center and then the light beam. Unfortunately, this system degrades image quality.
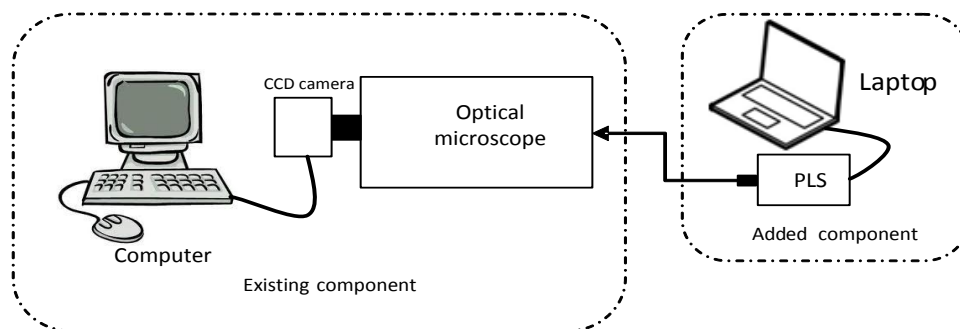

Figure 1. Acquisition chain.

The degradation is mainly due to the weakness of the light energy from the PLS and more, to the illumination of the sample in reflection and not in transmission. We decided to take a 3-bands multispectral image (MS3) to simulate the color image. Then we decided to acquire 5-bands multispectral images (MS5), 7-bands multispectral images (MS7) and 10-bands multispectral images (MS10) covering the visible spectrum.

To interpret the results and provide data acquisitions in conditions similar to the existing professional system, we had to characterize the system.

Linearity of the camera:

To analyze the linearity of the camera, we made acquisitions without any sample with integration times ranging from 300 ms to 10,100 ms (with 200 ms steps). To this end, 50 images were performed. For each image, an area of 100 by 100 pixels in the center was taken and averaged, plotted as a function of integration time.

Figure 2 shows the linearity curve amplitude grayscale function of time integration.

Figure 2. Response curve of the camera.
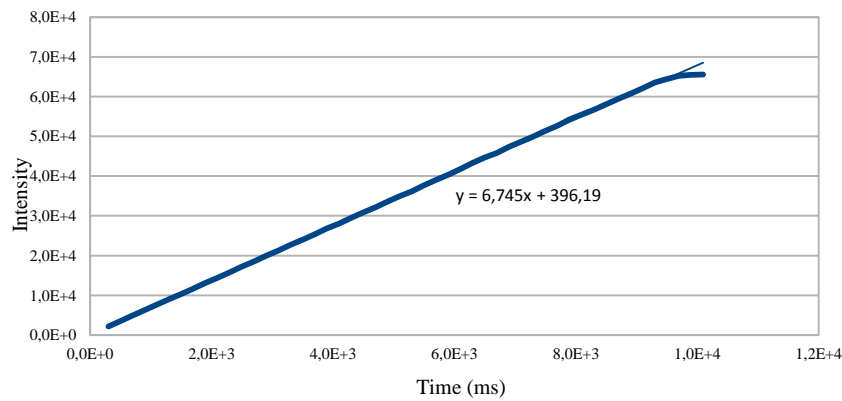
The linearity of the camera is clearly proved in Figure 2.

PLS calibration:

To ensure a good response with the Programmable Light Source, we conducted some series of tests on the PLS at rates of different times in a perfectly dark room in our lab (Le2i-France). The measurements were made using a perfect white patch and a spectrophotometer.
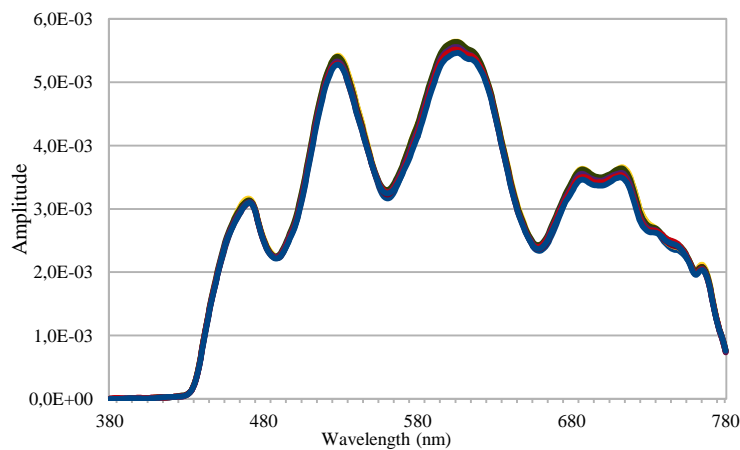

Figure 3. Response curve of the PLS.

Figure 3 shows the curves of repeatability of the PLS. These curves represent measurements taken at different moments simulating the real acquisitions. The x-axis represents the wavelength and the y-axis represents the amplitude of the signal. The measurements showed that the response of PLS is stable but not linear.

Uniform Light Range (ULR)

After acquisition of images by our system, we noted that these images have an intensity distribution default of light and may affect future treatments. So to have a Uniform Light Range (ULR), images acquired are corrected by a correction matrix, one coefficient per pixel. The coefficients of this matrix are computed using acquired images by our system in absence of sample. This correction is a kind of normalization which allows images acquired in absence of sample to become uniform after correction. In figure 4, we can see that the dark part of the image (top left corner) has been lightened.

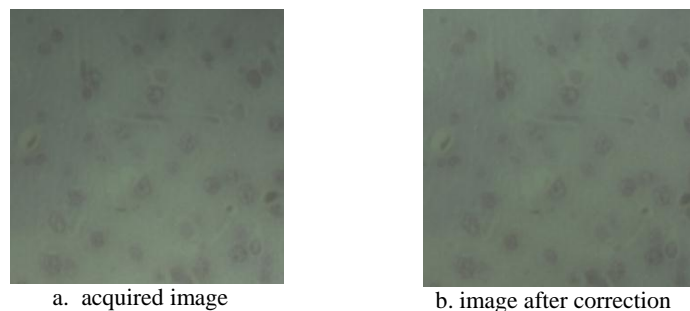a. acquired image   b. image after correction

Figure 4. Impact of Uniform Light Range.

For PLS, measures taken by our system allowed us to prepare the set command, to verify the stability and the nature of the signal emitted during the image acquisition time. They also allowed us to determine the amplitude of signal emitted in each band which still enables us to calculate multiplicative coefficients. These coefficients can serve in case we want to compare the bands of a multispectral image.

For the camera, measures taken have allowed us firstly to check the linearity of the camera, and on the other hand to compute the coefficients of the correction matrix described previously.

2.2 Method

In the literature there are many methods for counting cell nuclei; among these methods we decided to follow the one proposed by Park et al. [5]. The principle of this method relies on the use of 3-bands RGB color images. Image segmentation is one of the key processes in the field of cell nuclei counting. The choice of the segmentation method can affect the counting results. Our method is based on Expectation-Maximization (EM) segmentation and ellipse fitting for edge detection and contour selection.

In our study we are interested in multispectral images. To apply this nuclei detection algorithm on multispectral image, a pre-treatment phase is necessary to enhance the multispectral image quality.
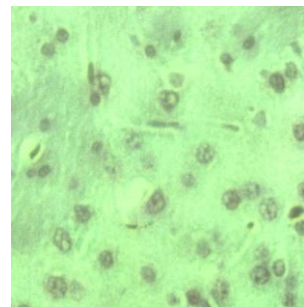
Pre-treatment

The different results for the pre-treatment phase are illustrated in figure 5. In fact, the acquired image are too dark, too noisy (figure 5.a) and seems difficult to use for cell nuclei detection algorithm. The pre-treatment phase is released in 4 steps. The first one is the adaptive histogram equalization: it consists on modifying the image intensity in order to have a uniform histogram. It leads to an improvement in the quality of the processed image by increasing the contrast thus facilitating the distinction between the background and the objects contained in the image (Figure 5.b). The second step is the image intensity values adjustment. This operation consists in modifying the image histogram to let it covering the entire range of intensity. This is realized by a linear change of pixel intensities (Figure 5.c). The third step is a median filtering. It reduces noise without significantly affecting the signal provided by the stained cells (Figure 5.d) [10, 11].
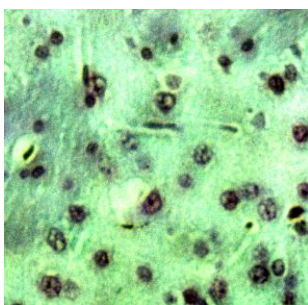
Finally, the fourth step is the Wiener filtering. It attempts to minimize the impact of noise deconvolution at frequencies which have a poor signal-to-noise ratio (Figure 5.e). The wiener method is based on statistics estimated from a local neighborhood of each pixel [12].
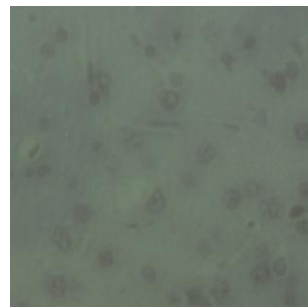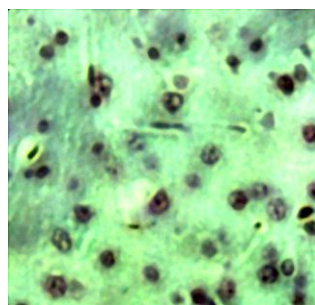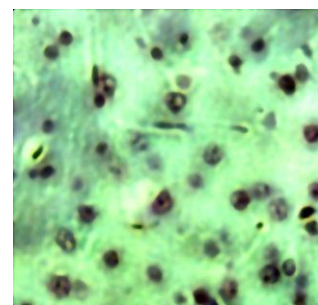

a. image after ULR correction

b. image after histogram equalization

c. image after intensity adjustment

d. image after median filtering
Figure 5. Image pre-treatment.

e. image after Wiener filtering

Automatic cell nuclei detection algorithm

The principle of our approach of automatic nuclei detection is shown in Figure 6. For each band, we apply the segmentation through the Expectation-Maximization (EM) algorithm. From this segmentation we isolate the class relative to the nuclei. We apply on this region a morphological gradient to extract contours. The extracted contours are classified using the ellipse fitting. The selected contours are returned in two categories: contours of single nuclei (RCS) and contours of multiple nuclei (RCM). This process is applied to all bands. Finally, we fuse single nuclei separately from multiple nuclei. This merging is based on the logical OR operator.

After that, we clean the result from isolated regions of small surface areas and combine regions of multiple nuclei with regions of single nuclei to correct the distribution of contours of both single nuclei and multiple nuclei.

- EM segmentation algorithm

Usually, image intensity is considered as a mixture of Gaussians randomly generated. In our case, the application of EM segmentation algorithm in n classes is based on the histogram of the image. This algorithm is an iterative process which is widely used and based on the maximum likelihood. This is achieved by computing a likelihood factor $V(x, i)$ specific to each Gaussian ($i=1..n$) and pixels x which allows, at each iteration, to recalculate the parameters of the Gaussians [10, 12, 13].

The nuclei are seen as dark spots. Based on this difference in brightness, we use the EM algorithm to separate the nuclei from the tissue background. So, we opted for a classification into two classes through the EM algorithm. The obtained results did not allow a good identification of nuclei. To bypass this problem, we actually process a classification into three classes (n=3) where nuclei are located in class 1 with the darkest levels. This change allows efficient localization of nuclei. Finally, the detected edges are computed by morphological operators to extract the contours.

- Ellipse fitting

This step allows computing the characteristic equation of the ellipse relative to a given distribution of points (contours). It allows extracting the parameters of the ellipse passing by the maximum number of points of the given distribution. Among these parameters, we are interested in the lengths of the major and minor axis of the ellipse and its eccentricity.

Applying the ellipse fitting to the contours given by the segmentation and basing on the parameters of the ellipse, we report if the contour may be a single nucleus contour (RCS), a multiple nuclei contour (RCM) or an invalid contour. A RCS contour means that our algorithm has detected one nuclei, a RCM contour means that our algorithm has detected more than one nuclei and 'invalid' contour signifies that the contour given by the segmentation will be eliminated. So, we fixed some thresholds on eccentricity, major and minor axis to sort the contours.
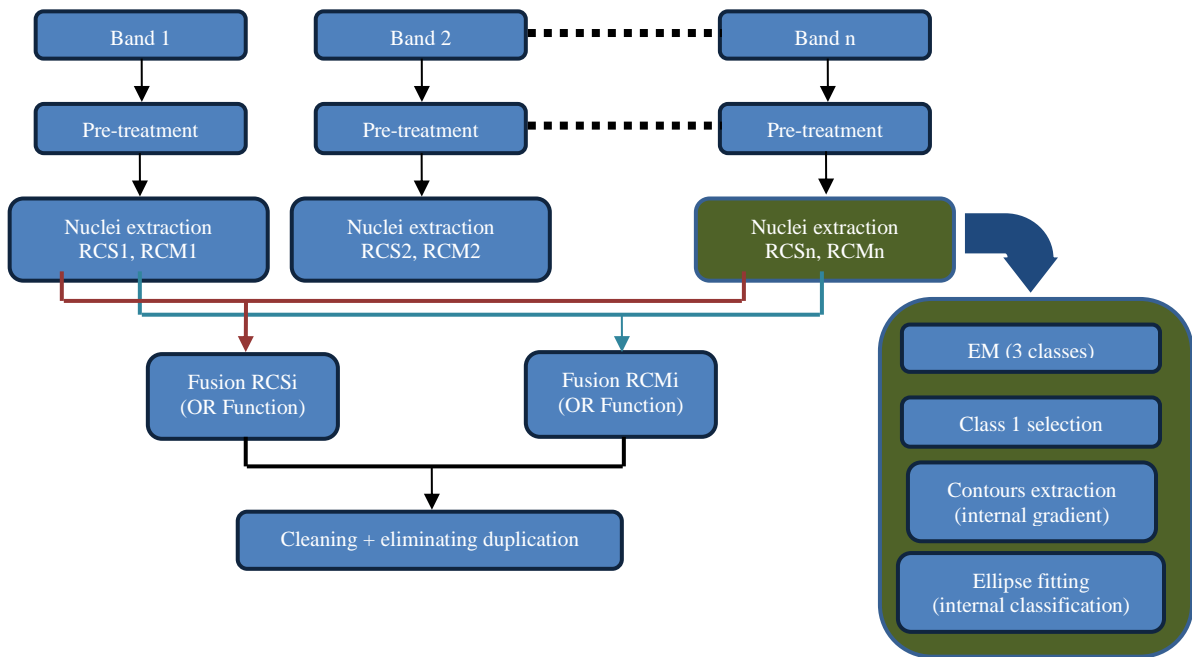
Figure 6. Detection algorithm.

3. RESULTS AND DISCUSSION

As a first assessment of our hardware and our technique for nuclei detection, we chose three areas of a rat brain sample. Capture is made by our multispectral system. So, we got 3-bands, 5-bands, 7-bands and 10-bands multispectral images (MS3, MS5, MS7 and MS10) for each area. To judge our approach, we compared our results with a manual selection done by a biologist.

In figure 7b and 7c, an area with a single nucleus is marked by a green color and an area with multiple nuclei by a blue color. By comparing these figures, we can deduce that a detection technique analysis can rely on 3 major criteria: detecting defects, labeling defects and valid detection. Detection defects can be divided in two classes:

- False detection: a portion of tissue (background) is considered as nuclei by our algorithm.

- Double detection: a nucleus is marked by more than one contour.

Labeling defects affect mainly the qualification of the detected nuclei. Indeed, the area may have portions for single nucleus and other portions for multiple nuclei. Labeling defects can be divided in two classes too:

• A single contour (green color) including more than one nucleus.

• A multiple contour (blue color) containing a single nucleus.

On the other hand, the algorithm successfully detects nuclei. This detection can be described in two ways:

• General detection is the percentage of nuclei (really existing in the tissue) detected by our automatic detection algorithm (without taking into account the single or multiple nature), compared to the ground truth (manual selection).

• A perfect detection is the percentage of nuclei detected by our algorithm and respecting their nature (single or multiple nuclei), compared to the ground truth.
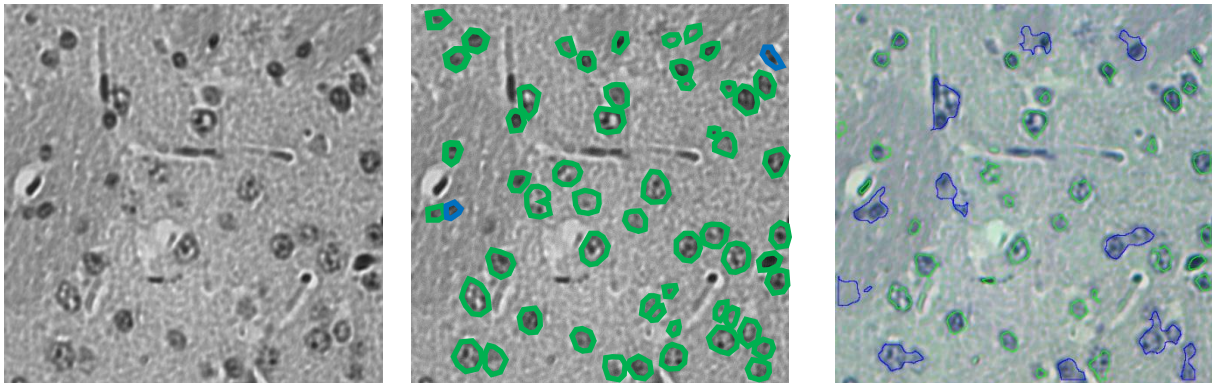
| a. band 2 of MS3 image | b. manual selection of nuclei | c. automatic detection of nuclei |

Figure 7. Comparison between manual selection and automatic detection in a MS3 image.

To evaluate the performance of our acquisition system and detection algorithm and to highlight the effect of the multispectral extension, we use the three criteria previously defined. As mentioned, this study is established on 3 areas of a histological slice of rat brain, each of them containing around fifty single nuclei and some multiple nuclei.

The obtained results are shown in the tables below. Tables 1 and 2 illustrate respectively for MS3 and MS10 images, the wrong detection average (background defects and double detection defects), the wrong labeling average and the valid detection average (perfect detection and general detection) obtained using our automatic nuclei detection algorithm. The average is calculated for the three studied areas (area 1, area 2 and area 3). The results are similar for the other MS5 and MS7 images; for simplicity reasons we chose not to show all tables.

		Wrong detection		Wrong labeling	Detection	
		Background	Double		Perfect	General
MS3	band 1	6,5%	5,7%	27,3%	52,0%	78,4%
	band 2	13,8%	5,9%	34,6%	43,0%	79,6%
	band 3	10,2%	5,5%	29,4%	46,0%	75,4%

Table 1. Results for MS3 image.

		Wrong detection		Wrong labeling	Detection	
		Background	Double		Perfect	General
MS10	band 1	29,0%	0,0%	4,4%	13,2%	17,7%
	band 2	11,7%	3,4%	22,4%	49,1%	71,5%
	band 3	4,6%	3,3%	26,7%	44,9%	71,6%
	band 4	6,1%	1,6%	35,7%	47,1%	82,7%
	band 5	4,5%	1,2%	36,2%	46,2%	82,3%
	band 6	6,2%	2,6%	36,7%	46,0%	82,7%
	band 7	5,7%	2,3%	36,7%	44,9%	81,7%
	band 8	5,2%	0,0%	33,9%	40,4%	74,4%
	band 9	8,6%	0,0%	12,6%	26,2%	38,8%
	band 10	28,5%	1,5%	10,4%	15,6%	26,0%

Table 2. Results for MS10 image.

By analyzing these results, it is clear that some bands have good general and perfect detection; however others show many defects due to background wrong detection (Table 1, band 2). To have a better exploitation of these multispectral images, we decided to fuse the automatic nuclei selection of all bands in same image (MS3 or MS5 or MS7 or MS10). But this merging may increase the wrong detection rate due to background defects. To decrease this effect, we defined a selection criterion of multispectral image bands before starting the fusion process. We take as a reference the MS3 multispectral image which is close to the traditional RGB system. According to the performance of our automatic nuclei selection on MS3 image (Table 1), we set a wrong background detection threshold that selection in bands should not exceed. The band 2 of the MS3 image presents the worst wrong background detection (13.8%) so we set the wrong background detection threshold as 15%.

The lowest general detection is 75.4% given by our selection algorithm in the third band of MS3 image. So, we fixed the general detection threshold as 75%. We consider only the bands where general detection exceeds the fixed threshold to the fusion process.

Following these selection criteria, all bands are taken into account for the MS3 images during fusion process. While, for MS5 images only the bands 2, 3 and 4 bands are selected. For MS7 images only the bands 2, 3, 4 and 5 are selected and finally, for MS10 images, the bands 4, 5, 6 and 7 are used (Table 3). We deduce that relevant information is located in the middle frequencies while low and high frequencies are rejected for their background defects rate or for their low general detection.

	Wrong detection		Wrong Labeling	Detection	
	Background	Double		Perfect	General
MS3 (1,2,3)	12,8%	0,0%	39,1%	49,8%	88,9%
MS5 (2,3,4)	6,7%	1,0%	35,1%	49,2%	84,3%
MS7 (2,3,4,5)	6,7%	0,6%	34,0%	48,7%	82,7%
MS10 (4,5,6,7)	7,5%	0,5%	37,5%	44,8%	82,3%

Table 3. Results for band fusion with maximum threshold for wrong detection of 15% and minimum threshold for general detection of 75%.

According to Table 3, we notice that the merging step decreases double detection rate and increases the general detection percentage. In addition, we notice that the extension in multispectral MS5, MS7 and MS10 reduces wrong detections and the wrong labeling percentage.

From these results, we infer also that the multispectral image does not ensure a best detection. However, this detection is still more than 80% for the general detection and more than 40% for the perfect detection. In other hand, the multispectral image reduces the background wrong detection percentage compared to the MS3.

When the minimum threshold for general detection is decreased to 70%, the number of bands increases in MS5, MS7 and MS10 images. So, bands 1, 2, 3 and 4 are selected in the fusion process of MS5 image. The merging process for MS7 and MS10 images includes respectively bands 2, 3, 4, 5 and 6 and bands 2, 3, 4, 5, 6, 7 and 8 (Table 4).

	Wrong detection		Wrong labeling	Detection	
	Background	Double		Perfect	General
MS3 (1,2,3)	12,8%	0,0%	39,1%	49,8%	88,9%
MS5 (1,2,3,4)	8,2%	1,0%	36,5%	47,8%	84,3%
MS7 (2,3,4,5,6)	7,8%	0,6%	30,9%	53,4%	84,3%
MS10 (2,3,4,5,6,7,8)	10,6%	0,0%	35,8%	48,6%	84,4%

Table 4. Results for bands fusion with maximum threshold for wrong detection of 15% and minimum threshold for general detection of 70%.

Comparing the results illustrated in Tables 3 and 4, we deduce that when we increase the number of bands selected in the fusion process, general detection increases with the background wrong detection. However, the background wrong detection percentage for MS5, MS7 and MS10 image is lower than that obtained for the MS3 image. We note also that the increase of the number of bands improves the perfect detection for MS7 and MS10 images.

We conclude that increasing the number of fused bands boosts the number of detected nuclei but also increases the percentage of background wrong detection.

4. CONCLUSION

There are two main goals in this work, the first one is to calibrate the acquisition system and the second one is to detect automatically cell nuclei inside histological slices of rat brain.

First, we calibrated all the components of the acquisition system. The programmable light source has been calibrated by doing many measurements and tests (repeatability, stability, linearity). Next, we checked the camera linearity by measuring its response. Finally we acquired 3-bands, 5-bands, 7-bands and 10-bands multispectral images. Because these acquired multispectral images are quite noisy, we proposed some pre-treatments to enhance the image quality. When this step has been finished, we applied the detection algorithm. It is mainly based on expectation-maximization segmentation and ellipse fitting. The results showed that the detection quality of nuclei augments by increasing the number of bands. Also for the same multispectral image, the general detection percentage increases with background wrong detection percentage.

For the future works, we think working towards two main ways to improve the results. First, in terms of software, we must improve the automatic detection method (segmentation method and ellipse fitting) by better adjusting some parameters of changing some parts of the methodology. Secondly, because current multispectral images contain some noise, we have to change the acquisition mode by using transmission process for the light instead of acquiring images by reflection.

REFERENCES

[1] Saraydin, S. U., Tuncer, E., Tepe, B., Karadayi, S., Őzer, H., Şen, M., Karadayi, K., Inan, D., Elagöz, Ş., Polat, Z., Duman, M. and Turan, M., "Antitumoral effects of Melissa officinalis on breast cancer in vitro and in vivo", Asian Pacific J Cancer Prev 13, 2775-2770 (2012).

[2] Tough, I. C. K., Cater, D. C., Fraser, J. and Bruce, J., "Histological grading in breast cancer", Brit J Cancer 23, 294-301 (1969).

[3] Chen, C., Ozolek, J., Wang, W. and Rohde, G., "An improved method for segmenting cell nuclei from microscopy images", XXVII Congress of the international society for advancement of cytometry, 220 (2012).

[4] Chan, Y., Tsail, M., Huang, D., Zheng, Z. and Hung, K., "Leukocyte nucleus segmentation and nucleus lobe counting", BMC BioInformatics 11, 558-576 (2010).

[5] Park, M., Jin, J.S., Peng, Y., Summons, P., Yu, D., Cui, Y., Luo, S., Wang, F., Santos, L. and Xu, M. "Automatic cell segmentation in microscopic color images using ellipse fitting and watershed", International Conference on Complex Medical Engineering (ICCME)/IEEE, vol. 0, pp. 69–74 (2010).

[6] Kothari, S., Chaudry, Q. and Wang, M.D., "Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques", International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)/IEEE, vol. 2, 795–798 (2009).

[7] Zheng, X., Zhang, Y. and Shi, J. "A new method for automatic counting of marrow cells", 4th International Conference on Biomedical Engineering and Informatics (BMEI)/IEEE, vol. 1, 42-46 (2011).

[8] Arambula Cosio, F., Marquez Flores, J.A., Padilla Castaneda, M.A., Solano, S. and Tato, P., "Automatic counting of immunocytochemically stained cells", Proceedings of the 25th Annual International Conference of the IEEE. Engineering in Medicine and Biology Society, vol. 1, 790 - 793 (2003).

[9] Lezoray, O. and Cardot, H., "Cooperation of color pixel classification schemes and color watershed: a study for microscopic image", IEEE Transactions on Image Processing, vol. 11, 783 - 789 (2002).

[10] Forero, M. G. and Hidalgo, A., "Image Processing Methods for Automatic Cell Counting In Vivo or In Situ Using 3D Confocal Microscopy", [Advanced Biomedical Engineering], Dr Gaetano Giuliano (Ed) InTech, 183-204 (2011),
http://www.intechopen.com/books/advanced-biomedicalengineering/
image-processing-methods-for-automatic-cell-counting-in-vivo-or-in-situ-using-3d-confocalmicroscopy

[11] Forero, M. G., Pennack, J. A., Learte, A. R. and Hidalgo, "A DeadEasy Caspase: Automatic counting of apoptotic cells in Drosophila", PLoS ONE, Public Library of Science 4(5), 5 May 2009,
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005441#abstract0

[12] Sarder, P., Nehorai, A., "Deconvolution methods for 3-D fluorescence microscopy images", Signal Processing Magazine, IEEE 23, 32 - 45 (2006).

[13] Yonghong Z., Yongqin Z. "EM algorithm with improvement for color image segmentation in multiple color spaces", International Conference on Electronics, Communications and Control, IEEExplore,853 - 857 (2011).