

# IMPROVING HISTOLOGY IMAGES SEGMENTATION THROUGH SPATIAL CONSTRAINTS AND SUPERVISION

Nicolas Hervé<sup>a</sup>, Aude Servais<sup>b</sup>, Eric Thervet<sup>b</sup>,  
Jean-Christophe Olivo-Marin<sup>a</sup>, Vannary Meas-Yedid<sup>a</sup>

<sup>a</sup> Institut Pasteur, Quantitative Image Analysis Unit, URA CNRS 2582, 75015 Paris, France

<sup>b</sup> Departments of Nephrology and Renal Transplantation, Necker Hospital, 75015 Paris, France

## ABSTRACT

We introduce two approaches to improve an existing color segmentation technique based on a Split and Merge quantization process for the study of stained histological images. We propose to modify the merge criterion : first, we include a spatial constraints heuristic; then we suggest the use of supervision and a more elaborated visual features representation. We tested these approaches on a renal biopsies dataset to automatically quantify interstitial fibrosis and show that supervision brings very significant improvements.

**Index Terms**— Color segmentation, color quantization, histology, heuristic, learning framework, spatial constraints

## 1. INTRODUCTION

We are interested in automatic color segmentation of stained histological images. Several generic approaches exist in the literature using clustering, edge detection, region growing or hybrid methods [1, 2, 3, 4]. More specifically, in biomedical images, many techniques have been developed for cytology [5, 6], but less for histology [7, 8]. In [5] the color segmentation is based on color component thresholding. In [6], as in [2], the idea is to combine segmentations obtained by standard techniques applied within different color spaces. [6] uses Support Vector Machine for the basis segmentation tool and [2] uses KMeans. These approaches are well adapted to cell image segmentation due to a good contrast between the background and the cell regions and to the fact that the information of the cell shape is well known.

Interstitial fibrosis (IF) is one of the features that characterizes the renal transplant change. Studies using surveillance biopsies have consistently shown that IF correlates with renal graft survival, thus suggesting that early histological detection of IF may be used to predict the risk of subsequent graft function deterioration [9]. The Banff classification [10] was developed to standardize interpretation of renal allograft biopsies. There are several limitations to its use. First, the reproducibility is limited by inter-observer variations. Second, the use of only 3 grades to describe the severity of individual histological abnormalities may lack accuracy.

Thus, quantitative image analysis represents a more reliable and reproducible technique and can be an alternative to the IF Banff quantification. Following the Banff technique, we quantify the fibrosis on Masson trichrome stained biopsy slides (see Figure 1). The IF score is defined as the ratio of the fibrosis surface over the full cortical surface. In a practice of routine, the challenge is to develop efficient algorithms that can quantify the IF with little or no human intervention.

The Masson trichrome stains several types of collagen and not only collagen constitutive of interstitial fibrosis, so in the first step of the analysis, accurate collagen segmentation is required in order to have a precise quantification and to be able to investigate a more elaborated structural analysis than the surface ratio. Then in a second step a processing based on color, shape and texture analysis is performed to remove several elements that are non constitutive of IF in order to keep only fibrosis regions. In this paper, we will only focus on the first step which involves color segmentation.

## 2. OUR APPROACH

We have an image  $\mathbf{I}$  of  $w \times h$  pixels. Each pixel  $p_{x,y}$  has a color which is represented by a multidimensional vector in a color space having generally 3 dimensions. In our case, we will use an hybrid colorspace IH2H3 defined by  $I1 = \frac{R+G+B}{3}$ ,  $H2 = R - G$  and  $H3 = \frac{R+G}{2} - B$  that was shown to be accurate for Masson trichrome histology slides segmentation [11]. Any quantization algorithm tries to find a set  $\mathbf{Q}_n$  of  $n$  representative colors that minimizes the global error between the original image and the quantized image where each pixel color is replaced by the closest one in  $\mathbf{Q}_n$ . The value of  $n$  is either a parameter of the algorithm or is determined automatically during the quantization process.

We use the fact that our histology slides are stained with the Masson trichrome as an *a priori*. The regions of interest have very distinctive colors. The background and lumen are white, the collagen appears in green/blue and the remaining part of the tissue is stained with a red/purple color. Thus, we focus on supervised clustering algorithms and choose  $n = 3$  so as to separate them. A Split and Merge approach [4] has been used and already provides good results [9]. This method reduces the number of colors by splitting the color space uniformly into  $m$  (typically  $m = 2048$ ) color bins and then by merging these clusters into  $n$  final classes by minimizing the mean square error of the partition.

Our proposal is to use the same color quantization approach but also to apply spatial constraints. Switching from the color domain to the joint color/spatial domain should indeed be more relevant for a segmentation task. Thus, we first start by a simple color quantization to produce  $n > 3$  representative colors. In a second step, these clusters are merged using a new fusion approach, in order to reach our final 3 representative colors. A simple heuristic is used to determine the collagen class. It is represented by the cluster having the minimal value in the H2 color component (close to green).

We performed experiments with the mean shift approach [1], but preliminary results were not conclusive. We also tested the segmentations fusion of Mignotte [2] and texture based approaches without success. Like any algorithm using sliding windows in order

to encompass neighbourhood informations, the smoothing of segmented region boundaries induced by these approaches is too high and, moreover, hardly controllable. The collagen segmentation is the first step of an elaborated process that aims at describing finely the interstitial fibrosis. Thus, we need a very precise segmentation and avoid errors due to coarse approaches.

### 2.1. Improved merging strategy

The result of the Split and Merge color quantization algorithm is a quantized version  $\mathbf{I}_{\mathbf{Q}_n}$  of our image that can be interpreted as a segmentation  $\mathbf{S}$  consisting of an ensemble  $\mathcal{R}$  of  $r_i$  regions or connected components. Each region is characterized by its representative color  $q_i \in \mathbf{Q}_n$  and its set of pixels  $\mathcal{P}_i$ . Each pixel of  $\mathbf{I}_{\mathbf{Q}_n}$  belongs to one, and only one region. We also define  $c_i$  the average color of a region. Finally, the region adjacency graph (RAG) is extracted for this segmentation.

The standard Split and Merge approach quantizes pixel colors at a global level without reflecting any spatial information. We propose a new merging algorithm that jointly considers color and spatial informations. Based on the RAG, we iteratively merge adjacent regions based on a fusion score  $f$ . The distance between average colors is used. It is balanced by a criterion on the size of the regions in order to promote the merging of small regions.

$$f(r_i, r_j) = (1 - \alpha) \|c_i - c_j\| + \alpha \frac{\max(|\mathcal{P}_i|, |\mathcal{P}_j|)}{w \times h}$$

This score is designed to preserve the thin structures having a color gradient. Indeed, the successive merging of small regions having close colors should help distinguishing the collagen structure in the renal biopsies. During the iterative process, when the fusion score reaches a threshold value  $\beta$ , the spatial merging is stopped. We continue the merging without the adjacency constraint until we obtain our final regions. Our algorithm is then :

---

#### Algorithm 1: Spatial merging

---

**Data:** an image  $\mathbf{I}$ , parameters  $\theta = (n, \alpha, \beta)$   
 $\mathbf{Q}_n \leftarrow \text{SplitMerge}(\mathbf{I}, n)$ ;  
 $\mathcal{R} = \{r_i : (\mathcal{P}_i, q_i, c_i)\} \leftarrow \text{Quantize}(\mathbf{I}, \mathbf{Q}_n)$ ;  
 $\text{RAG} \leftarrow \text{AdjacencyGraph}(\mathcal{R})$ ;  
**repeat**  
     $(r_{m_1}, r_{m_2}) \leftarrow \underset{i,j}{\operatorname{argmin}} f(r_i, r_j), (r_i, r_j) \in \text{RAG}$ ;  
     $\text{Fuse}(r_{m_1}, r_{m_2}, \mathcal{R}, \text{RAG})$ ;  
**until**  $f(r_{m_1}, r_{m_2}) > \beta$ ;  
**repeat**  
     $(r_{m_1}, r_{m_2}) \leftarrow \underset{i,j}{\operatorname{argmin}} f(r_i, r_j), r_i, r_j \in \mathcal{R}$ ;  
     $\text{Fuse}(r_{m_1}, r_{m_2}, \mathcal{R}, \text{RAG})$ ;  
**until**  $|\mathcal{R}| = 3$ ;

---

The parameter set  $\theta$  encompasses informations about the staining quality and color variability ( $n$ ), biological structures size and acquisition scale ( $\alpha, \beta$ ). It is, currently, determined empirically.

### 2.2. Supervised merging

The previous approach is an elaborated heuristic that uses only a small amount of *a priori* information on the studied domain through  $\theta$ . We also wished to investigate more supervised segmentation approaches, based on learning frameworks. The purpose is still to bypass the last steps of the Split and Merge quantization algorithm and

replace them by a supervised approach. We propose a probabilistic formulation of the merging process. We associate the probability of depicting collagen to each one of the  $n$  color clusters obtained by the quantization step :  $P_{col}(q_i)$ . These probabilities are estimated on ground truth images. For a new image, quantized with the same representative colors  $\mathbf{Q}_n$ , we have the final segmentation function defined for each pixel :

$$P_{col}(p_{x,y}) = \sum_{i=1}^n P_{col}(q_i) P(q_i | p_{x,y})$$

Currently, as we are using a non-fuzzy quantization approach, we have  $P(q_i | p_{x,y}) \in \{0, 1\}$ . The model can easily be extended to fuzzy quantization process and then be equivalent to supervised Gaussian Mixture Models. With this approach, all the color clusters that are likely to depict collagen are merged ( $P_{col}(q_i) \geq 0.5$ ) and the final segmentation in the spatial domain is obtained.

### 2.3. Spatial constraints

Following the idea of [12], we integrate weak geometrical constraints directly in the low-level signal representation by considering co-occurrences of pixels in neighbour pairs. This representation encodes both structural and contextual information. For each pixel, we consider the 4 pairs induced by the 4-connectivity. Thus, we extend the previous model to pairs of pixels.

$$P_{col}((p_{x_1, y_1}, p_{x_2, y_2})) = \sum_{i=1}^n \sum_{j=1}^n P_{col}((q_i, q_j)) P((q_i, q_j) | (p_{x_1, y_1}, p_{x_2, y_2}))$$

The probability for each pixel of being collagen is then averaged over all the pairs to which it belongs :

$$P_{col}(p_{x,y}) = \frac{1}{4} \sum_{|dx|+|dy|=1} P_{col}((p_{x,y}, p_{x+dx, y+dy}))$$

To strengthen the spatial information encoding of our representation, we also used trigrams. Inspired by the center-symmetric local binary patterns of [13], trigrams are triplets of aligned pixels. Following the same formalism defined for the pairs, we have 4 trigrams per pixel.

## 3. EVALUATION

### 3.1. Criterion

Several performance evaluation metrics have been proposed in the literature to assess segmentation results [14]. One of the main issues is often the lack of objective ground truth segmentations to which one can compare the output of the proposed techniques. Even expert pathologists sometimes disagree [9]. Thus, an evaluation metric should at least be able to manage several ground truth segmentations. Another common problem is to decide how to manage several segmentation granularities. In our case, we need to design an algorithm that automatically quantizes the fibrosis surface. Thus, the output of our approach is a single region and we should not worry about over and under segmentation during the evaluation. We therefore choose the Hamming distance to compare two regions. We normalize it by the size of the ground truth segmentation to allow comparisons between images. This measure naturally extends to several ground truths. For a given segmentation represented by its set of pixels  $\mathcal{P}_{seg}$  and a ground truth  $\mathcal{P}_{GT}$ , we have the Normalized Hamming Distance (NHD):

$$\text{NHD}(\mathcal{P}_{seg}, \mathcal{P}_{GT}) = \frac{|\mathcal{P}_{seg} \cup \mathcal{P}_{GT}| - |\mathcal{P}_{seg} \cap \mathcal{P}_{GT}|}{|\mathcal{P}_{GT}|}$$

### 3.2. Dataset

For our evaluation, we use 8 homogeneous biopsy sections. The quality of the staining process has a low overall variability between images of the dataset. We have manually segmented the collagen. Figure 1 shows a distinctive crop and the associated manual ground truth segmentation. All histological slides are fully imaged by a

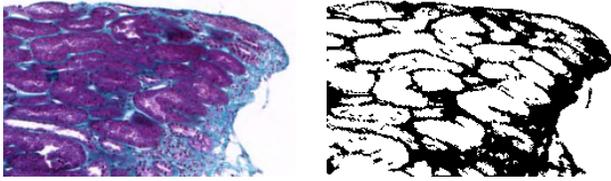


Fig. 1. An example image with manual collagen segmentation

Zeiss Mirax Scan with a 20X objective and a numerical aperture of 0.8. This is a virtual microscope providing several scales at the same time. We choose to export images using the 1:8 scale. The biopsy sections are then automatically recognized, cropped and rotated by standard thresholding and mathematical morphology techniques. This process is completely automatic.

### 3.3. Experiments

All the experiments were done in QUIA, our quantitative image analysis software. We first present the results of unsupervised segmentation by our new color clusters fusion heuristic. Standard Split and Merge has a *NHD* score of 0.285. When using spatial merging, we reach 0.233, thus bringing a 18.2% improvement in the precision of the segmentation (see figure 3). We determined empirically the following parameters :  $n = 6$ ,  $\alpha = 0.08$  and  $\beta = 0.09$ . They provide the best results on our dataset. Figure 2 shows examples of segmentations. The ground truth is blue, the proposed segmentation is red,



Fig. 2. (L) Split and Merge, (R) Spatial merge

the black pixels define the correctly segmented regions. We observe that some of the thin structures have been extracted, even if some are still missing. Our approach focuses more on small regions. As a region is growing, it becomes more and more difficult to follow color gradients because of the average color inertia. Indeed, the average color tends to move away from the boundary pixels color.

For all supervised approaches, the model is learned on a single image and used to segment the rest of the dataset. The results are the average for all possible learning configurations. We prefer this approach to traditional leave-one-out as it better reflects real usage conditions with only few data available for learning. Indeed, precise ground truth is hard to acquire as manually segmenting histological slides is very time consuming. To get a baseline result, we tested a standard segmentation based on SVM learning. We consider our segmentation task as a 2-class classification problem (collagen/not collagen). We use directly the real color vectors of the

pixel, without quantization. We use a soft margin support vector machine with a triangular kernel and a fixed relaxation parameter  $C = 1$ . It has the great advantage of being non-parametric and has proven its good performances in similar contexts [12]. We notice

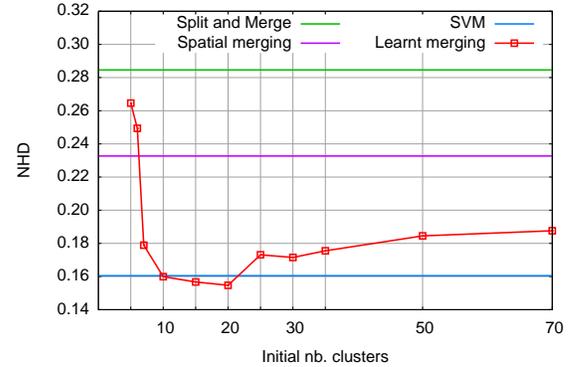


Fig. 3. Merging learnt

immediately that supervised approaches brings very significant improvements. For  $n = 20$ , the gain of our proposed approach over standard Split and merge is 45.6% and 33.5% over spatial merging heuristics. Our probabilistic model obtains comparable results with a standard SVM, but it is significantly faster and provides a more elegant model where parameters are self-explanatory. This approach opens the path to an efficient semi-supervised method where only a minimal human interaction would be necessary.

Similar to spatial relationships introduced in section 2.1, a natural way of considering simultaneously a multivariate gaussian modelling of color clusters and pixel neighbourhood is the use of Markov Random Fields [3]. To obtain baseline results, we use a Potts model with first order cliques. We test different values of the  $\beta$  parameter used to control the homogeneity of the regions and balance the respective influence of color clusters and neighbourhood in the energy function. Thus, we are in a fully supervised context with all model parameters known. We used a Gibbs sampler to minimize the MRF energy function and obtain the segmentation. We present in Figure 4 the results when using the ground truth for the 3 classes in our images. Adding spatial constraints only brings

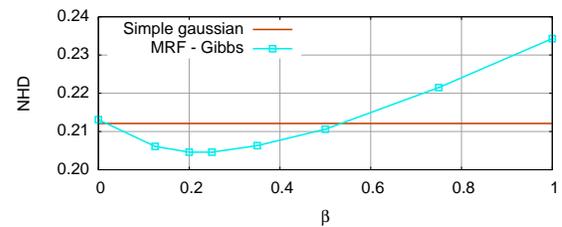


Fig. 4. Simple gaussian and MRF learned on 3 classes

a small improvement of 3.5% ( $\beta = 0.2$ ) at the expense of a huge increase in computation time due to the simulated annealing process induced by the MRF.

The second approach includes spatial constraints directly in the visual representation. Using pairs of pixels brings only a small improvement compared to the single pixel version (3.5% for  $n = 20$ ). As shown in Figure 5, using trigrams does not bring any improvement. We can also clearly observe the overfitting effect when the initial number of color clusters is too high. In this case, the amount

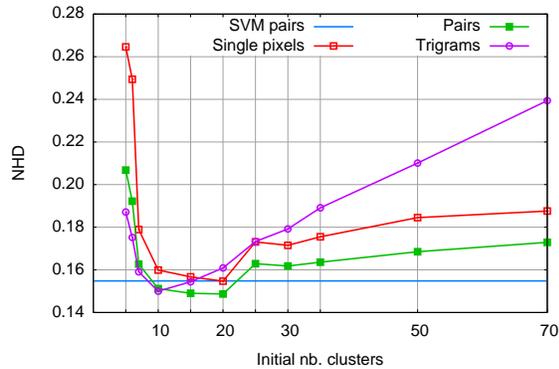


Fig. 5. Using spatial constraints in a supervised context

of learning data is not sufficient to cover our feature space. Segmentation results are shown in Figure 6. We can notice that thin structures are much more preserved. When correctly modeled, the color considered alone provides very good segmentations. The spatial information can help partially. Whatever the learning framework, the improvement is 3.5%, which indicates that the supervised approach leads to better results, both quantitatively and qualitatively.



Fig. 6. Merging learnt : (L) single pixels, (R) pairs

#### 4. CONCLUSION

We have shown that introducing spatial constraints in an heuristical merge criterion brings significant improvements to a Split and Merge color quantization approach for segmentation. Even if it has been specifically designed, we believe this criterion is generic enough to be applied to other stained biological images. However, the use of supervision clearly brings the most important qualitative shift. In this context, the improvements due to spatial constraints, although still existing, are minimized. The choice to use supervision or not depends on the final application. For our future work, we need to define a segmentation quality assessment criterion that focuses more on thin structures. We also need to investigate the respective influence of our heuristical merge criterion parameters  $\theta$  and study a potential adaptive selection approach. Moreover, we plan to study the robustness of our models over staining variability. A slide normalization approach may be a good choice [15]. Finally, the use of fuzzy clustering is clearly an approach that must be considered for our probabilistic models.

#### 5. ACKNOWLEDGEMENT

We thank Dr L.H. Noël from the Necker hospital in Paris for the biological samples. We are grateful to F. Orioux and N. Chenouard for fruitful discussions. This project is supported by ATN.

#### 6. REFERENCES

- [1] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [2] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Trans. Image Processing*, vol. 17, no. 5, pp. 780–787, 2008.
- [3] Z. Kato, T. Pong, and J. Lee, "Color image segmentation and parameter estimation in a markovian framework," *Pattern Recognition Letters*, vol. 22, no. 3-4, pp. 309–321, 2001.
- [4] J.P. Braquelaire and L. Brun, "Comparison and optimization of methods of color image quantization," *IEEE Trans. Image Processing*, vol. 6, no. 7, pp. 1048–1052, 1997.
- [5] E. Glory, V. Meas-Yedid, G. Stamon, Ch. Pinset, and J.-Ch. Olivo-Marin, "Automated image-based screening of cell cultures for cell therapy," in *Proc. IEEE Int. Symposium on Biomedical Imaging*, 2006, pp. 259–262.
- [6] C. Charrier, G. Lebrun, and O. Lezoray, "Fusion of svm-based microscopic color images through colorimetric transformation," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, 2006, pp. 1112–1115.
- [7] E. Romero, J. M. Raymackers, B. Macq, and O. Cuisenaire, "Automatic fibrosis quantification by using a k-nn classifier," in *Proc. Int. Conference of the IEEE Engineering in Medicine and Biology Society*, 2001, vol. 3, pp. 2609–2612.
- [8] K. Prasad and P. K. Nayak, "Color image analysis to grade shades of a color and its application to quantify stained tissues," in *Proc. Int. Conference on Biomedical Engineering*, 2008, pp. 154–157.
- [9] A. Servais, V. Meas-Yedid, O. Toupance, Y. Lebranchu, A. Thierry, B. Moulin, I. Etienne, C. Presne, de Ligny B Hurault, P. Le Pogamp, Y. Le Meur, D. Glotz, C. Hayem, J. C. Olivo Marin, and E. Thervet, "Interstitial fibrosis quantification in renal transplant recipients randomized to continue cyclosporine or convert to sirolimus.," *American Journal of Transplantation*, vol. 9, no. 11, pp. 2552–2560, 2009.
- [10] K. Solez et al., "Banff 07 classification of renal allograft pathology: updates and future directions.," *American Journal of Transplantation*, vol. 8, no. 4, pp. 753–760, 2008.
- [11] V. Meas-Yedid, E. Glory, E. Morelon, Ch. Pinset, G. Stamon, and J.-Ch. Olivo-Marin, "Automatic color space selection for biological image segmentation," in *Proc. Int. Conference on Pattern Recognition*, 2004, vol. 3, pp. 514–517.
- [12] N. Hervé and N. Boujemaa, "Visual word pairs for automatic image annotation," in *Proc. IEEE Int. Conference on Multimedia and Expo*, 2009, pp. 430–433.
- [13] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, 2009.
- [14] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.
- [15] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symposium on Biomedical Imaging*, 2009, pp. 1107–1110.