

# PDES-BASED MORPHOLOGY ON GRAPHS FOR CYTOLOGICAL SLIDES SEGMENTATION AND CLUSTERING

Xavier Desquesnes, Abderrahim Elmoataz, Olivier Lézoray

Université de Caen Basse Normandie, GREYC, ENSICAEN, UMR CNRS 6072, F14032 Caen, France

## ABSTRACT

In this paper, we propose a new graph-based approach to address the problem of cytological computer-aided diagnostic. Such approach uses our previously introduced formalism of PdEs-based morphology and geometric diffusion on graphs. The approach is illustrated through two applications in cytopathology (involving Feulgen and Papanicolaou colorations), with examples of nucleus extraction and classification.

**Index Terms**— PdEs, weighted graphs, morphology, cytology, segmentation, classification

## 1. INTRODUCTION

Traditionally, cytopathologists take a diagnostic decision by studying morphological and texture features of cellular components (cytoplasm and nuclei), seen through a microscope. This is a complex process as a cytological slide can contains millions of cells on which abnormal cells (affected by a cancer) are often very rare or fortunately absent. During the last decade, the advent of fast and efficient high-resolution slide scanners along with the development of computer vision has opened the way to using digital pathology as a diagnosis tool. However, computer-aided diagnosis involves many tasks, from the acquisition of the cytological slide to the final classification of each cell. Indeed, the digital image can be preprocessed for enhancement, and analyzed to extract cellular objects. The extracted objects might also be preprocessed, and analyzed to extract some features in order to perform a classification. In this work, we use our recently proposed framework of PdE-based morphology on graphs to address the problem of cytological computer-aided diagnostic through different sub-problems: image preprocessing, cell extraction or cell classification. The rest of this paper is organized as follows. In Section 2, we provide definitions and notations as well as a recall on PdEs based morphology on graphs and detail our previous works on label propagation on graphs. Section 3 presents two different graph-based methodologies for cytological slide segmentation and classification,

with a particular focus on image preprocessing and nucleus extraction steps. Section 4 concludes this paper.

## 2. PRELIMINARIES ON GRAPHS AND NOTATIONS

As the core structure of our approach, in this Section we provide notations and basics on weighted graphs. We recall our formulations of differences, morphological differences and gradients on weighted graphs [1, 2].

**Notations:** We assume that any discrete domain can be modeled by a a weighted graph. Let  $G = (V, E, w)$  be a weighted graph composed of two finite sets:  $V = \{u_1, \dots, u_n\}$  of  $n$  vertices and  $E \subset V \times V$  a set of weighted edges. The weight function  $w : V \times V \rightarrow [0, 1]$  represents a similarity measure between two vertices of  $G$ . According to  $w$ , the set of edges is defined as :  $E = \{(u, v) \mid w(u, v) \neq 0\}$ . We denote by  $N(u)$  the neighborhood of a vertex  $u$ , i.e., the subset of vertices that share an edge with  $u$ . In this paper, graphs are assumed to be connected, undirected and with no self loops. We denote by  $\mathcal{H}(V)$  and  $\mathcal{H}(E)$  the Hilbert spaces of functions that assigns a real value to each vertex, respectively edge, of  $G$ . Let us consider  $\mathcal{A}$ , a subset of  $V$ . The *outer* and *inner* boundary sets of  $\mathcal{A}$  are respectively denoted  $\partial^+ \mathcal{A}$  and  $\partial^- \mathcal{A}$ . With  $\partial^+ \mathcal{A} = \{u \in \mathcal{A} \mid \exists v \in \mathcal{A}^c, v \sim u\}$  and  $\partial^- \mathcal{A} = \{u \in \mathcal{A} \mid \exists v \in \mathcal{A}^c, v \sim u\}$  where  $\mathcal{A}^c$  is the complement of  $\mathcal{A}$ .

**Operators on Weighted Graphs:** Given a weighted graph  $G = (V, E, w)$  and a function  $f \in \mathcal{H}(V)$ , the *weighted gradient operator* or *weighted difference operator*, noted  $\mathcal{G}_w : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$  is defined on an edge  $(u, v) \in E$  by

$$(\mathcal{G}_w f)(u, v) \stackrel{def.}{=} \sqrt{w(u, v)}(f(v) - f(u)). \quad (1)$$

Based on the *weighted gradient operator* definition, two *weighted directional gradient operators* are defined. The weighted directional *external* and *internal* gradient operators are defined as  $\mathcal{G}_w^\pm : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$ , by

$$(\mathcal{G}_w^\pm f)(u, v) \stackrel{def.}{=} \sqrt{w(u, v)}(f(v) - f(u))^\pm \quad (2)$$

with the following notations:  $(x)^+ = \max(x, 0)$  and  $(x)^- = -\min(x, 0)$ . The *weighted gradient* of a function  $f \in \mathcal{H}(V)$

This work was supported under a doctoral grant of the Conseil Régional de Basse-Normandie and of the Cœur et Cancer association.

at vertex  $u$  is defined as the vector of all *weighted gradients* with respect to the set of edges  $(u, v) \in E$

$$(\nabla_w f)(u) \stackrel{def.}{=} \left( (\mathcal{G}_w f)(u, v) \right)_{v \in V}. \quad (3)$$

In the sequel, *weighted gradient* will refer to this gradient defined on vertices. Similarly, the *weighted morphological internal and external gradients* at a vertex  $u$  are expressed as

$$(\nabla_w^\pm f)(u) = \left( (\mathcal{G}_w^\pm f)(u, v) \right)_{v \in V} \quad (4)$$

with the following  $\mathcal{L}_p$  ( $p \in \{1, 2\}$ ) and  $\mathcal{L}_\infty$  norms

$$\begin{aligned} \|(\nabla_w^\pm f)(u)\|_p &= \left[ \sum_{v \sim u} w_{uv}^{p/2} |(Df(u))^\pm|^p \right]^{1/p}, \\ \|(\nabla_w^\pm f)(u)\|_\infty &= \max_{v \sim u} (\sqrt{w_{uv}} |(Df(u))^\pm|), \end{aligned} \quad (5)$$

**PdE based Morphology** Discrete erosion and dilation on weighted graphs are respectively defined by

$$\partial_t f(u) = +\|(\nabla_w^+ f)(u)\|_p \text{ and } \partial_t f(u) = -\|(\nabla_w^- f)(u)\|_p. \quad (6)$$

These equations (6) constitute a PdEs based framework [2] that extends algebraic and continuous morphological operators to graphs. Such a framework has been involved in the transcription of geometric PDES from continuous domains, to graphs of arbitrary topology. Indeed, given, a set of vertices  $\mathcal{A} \subset V$  and using external and internal graph boundaries, equation of dilation over  $\mathcal{A}$  can be intuitively interpreted as a growth process that adds vertices from  $\partial^+ \mathcal{A}$  to  $\mathcal{A}$ . By duality, erosion over  $\mathcal{A}$  can be interpreted as a contraction process that removes vertices from  $\partial^- \mathcal{A}$  to  $\mathcal{A}$ .

Let  $\Gamma$  be a parametrized curve evolving on a domain  $\Omega$ . A very common way to describe the evolution of  $\Gamma$ , introduced by Osher-Sethian [3], is to embed the curve in a function  $\phi(x, t)$  such that the evolving curve  $\Gamma_t$  can be provided by the 0-level set of  $\phi$ . Once formulated as a level set problem, the curve evolution can be done by solving the following equation

$$\frac{\partial \phi}{\partial t} = \mathcal{F} |\nabla \phi|. \quad (7)$$

In a previous paper [4], we have proposed a transcription of (7) to weighted graph, that can be expressed as a morphological process with the following sum of dilation and erosion.

$$\frac{\partial \phi}{\partial t}(u) = (\mathcal{F})^+ \|(\nabla_w^+ \phi)(u)\| + (\mathcal{F})^- \|(\nabla_w^- \phi)(u)\| \quad (8)$$

where  $\mathcal{F} \in \mathcal{H}(V)$  controls the front propagation. Such a formulation enables to recover geometric diffusion models as mean curvature motion, active contours, or a graph based transcription of the eikonal equation.

The eikonal equation is also a very popular equation in computer graphics and computer vision which is involved in

many applications. Numerous methods have been proposed to solve it on Cartesian grid and some particular non-Cartesian domains (see [5] and references therein). Recently, we have proposed two adaptations of the eikonal equation, first as a time dependent version [2], then as a static version [5] which is expressed as

$$\begin{cases} \|(\nabla_w^- f)(u)\|_p = P(u). & \forall u \in V \\ f(u) = 0 & \forall u \in V_0. \end{cases} \quad (9)$$

Such adaptations are expressed using the PdEs based morphological erosion, and can be linked with the general geometric PdE equation (8). In [5], we have proposed a label propagation algorithm based on the resolution of (9), that enables the propagation of many labels on a graph. The propagation is performed from a set of seeded vertices and until all vertices of the graph are marked with a label. This algorithm enables many applications on graphs, as geodesic distance computation on graphs, image segmentation and data clustering.

In the following Section, we present the interest of the above morphological processes, applied to cytological slides segmentation and clustering.

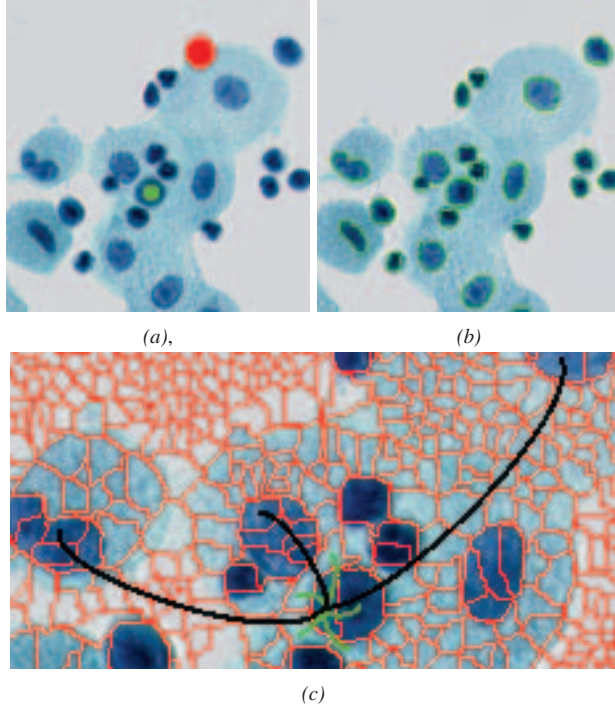
### 3. APPLICATIONS TO CYTOLOGY

In this Section, we focus on the three following problems: preprocessing, cell extraction and classification, under our unified formalism of PdEs-based morphology on graph. The different steps will be illustrated through two methodologies. The first that considers a Papanicolaou coloration along with a RAG representation of images; The second considers a Feulgen coloration along with natural grid-graph representation of images. In both cases, the preprocessing and cellular objects extraction steps are presented.

*Remark :* In this paper, we consider images as grid-graphs where each pixel is represented by a vertex, and edges are added between adjacent pixels (in the sense of 4-adjacency). The weigh function  $w$  is defined as a Gaussian kernel and holds the similarity between the pixels color vectors. Acquired cytological whole slides are split in thousands of  $1024 \times 1024$  tiles images on which the different algorithms are processed. All images presented in this paper are very small details extracted from some of the processed tiled images and can obviously not represent the hugeness of digital slides.

#### 3.1. First method: Papanicolaou coloration

The Papanicolaou coloration is the most used in cytology, and stains cytoplasm in light green and nuclei in dark blue. In this paper, we propose a semi-supervised segmentation approach, based on two different graph representations of the tiles.



**Fig. 1.** Papanicolaou nuclei extraction. (a) Initial image with superimposed user seeds. (b) Extracted nuclei (green boundaries). (c) Detail of the partitioned version of the image (given by the red boundaries). The extended RAG is illustrated for a single vertex with local (green) and non-local (black) edges.

**Preprocessing.** While working on very large data (as it is the case of cytological slides), many algorithms become inefficient due to the mass of data to process. A simple and efficient way to avoid this problem is to work on a simplified version of the data, that can be obtained by a partitioning algorithm. In this paper, we propose to use an approach similar to super pixels (which we can call super vertices as we work on graphs). Such an approach uses a set of labeled seeds placed on the image that are dilated using the graph-based propagation algorithm (which preserves local boundaries) until the entire image is labeled. In [5], we used a regular grid of seeds, but to better catch the finest structures in the tiles images we prefer here to use images extrema. Figure 1 presents an image partition computed from the extrema, using our graph-based label propagation approach. From the resulting partition, a Region Adjacency Graph (RAG) can be easily built: distinct regions are considered as vertices and connected regarding their adjacencies. Finally, processing such a reduced version of the image can consequently decrease algorithms computation time as the number of vertices is significantly reduced. In Fig.1 the number of vertices from the first graph (the image grid-graph) to second one (the RAG) has been reduced by more than 95%.

**Nuclei extraction.** The proposed approach uses an ex-

tended version of the previously computed RAG, such that each vertex is also connected to its two most similar vertices in the whole graph. That way, two similar but not adjacent objects can be connected in the graph. The segmentation is performed on the extended RAG by the label propagation algorithm introduced in [5], using a set of user seeds (one for the background and one for the objects of interest). Figure 1 illustrates such a process. Such an approach has two main interests. First, due to extra edges the labels can be propagated to non-adjacent regions and reach distant or not connected objects (as the nuclei). Indeed, considering a labeled vertex, the label propagation is driven along all its edges, including non-local ones which connect distant objects. Second, the algorithm being performed on a very low number of vertices, the method can be used to process very large slides while keeping a reasonable processing time.

### 3.2. Second method : Feulgen coloration

Feulgen coloration is more computer-vision-friendly, because the slide can be seen as a set of two easily separable classes as the background (in grey) and the nucleus (in pink). However, the nuclei boundaries are often unsharp due to refractiveness or mucus artifacts standing around the nucleus. In this paper, we propose an unsupervised segmentation methodology for nuclei extraction, which provides smooth boundaries, following the process detailed in the sequel:

**Preprocessing.** Before the segmentation process, and in order to better extract nuclei, the background homogeneity is enhanced using a weighted morphological closing operation. For a given function  $f \in \mathcal{H}(V)$ , one way to perform a weighted closing operation is to implement it serially as compositions of weighted dilation  $\delta$  and erosion  $\epsilon$ . We recall that dilation and erosion equations (6) can be solved by steepest gradient descent, using the following iterative numerical scheme:  $\forall u \in V$ , with  $f^n(u) \approx f(u, n\Delta t)$

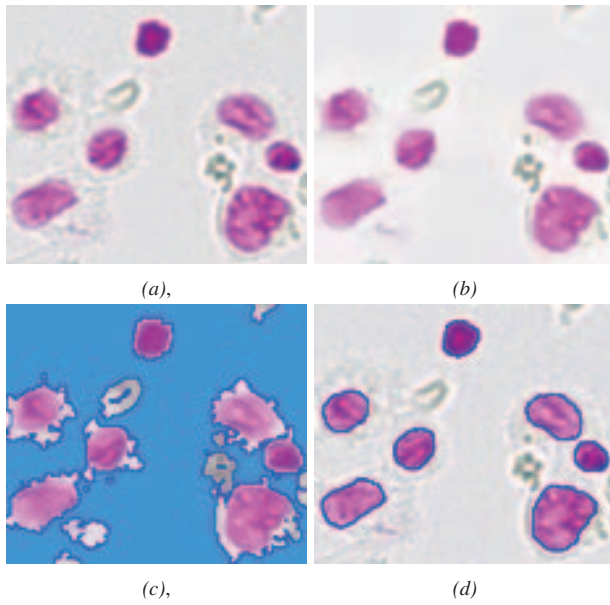
$$f^{n+1}(u) = f^n(u) \pm \Delta t \|(\nabla_w^\pm f^n)(u)\|_p \quad (10)$$

Figure 2 illustrates a closing operation on a cropped part of a cytological slide. This result clearly show that the closing process preserves the important image components, such as boundaries, while filtering and smoothing the homogeneous parts of the images (background and interior of the nuclei).

**Nuclei extraction.** After the preprocessing step, the background is homogeneous and can be automatically and roughly detected as the biggest region of similar pixels (in the sense of color variance). Such pixels become the seeds of the background label which is propagated on the image grid-graph until being stopped by nuclei boundaries. The stopping criterion is held by the potential  $P$  of equation (9), as

$$P(u) = 1 / \left( (\mu_N - f(u))^2 - (\mu_B - f(u))^2 + \mathcal{K}(u) \right) \quad (11)$$

where  $\mu_N$  and  $\mu_B$  are respectively the mean color of nuclei and background,  $\mathcal{K}$  is a curvature term that constraints the

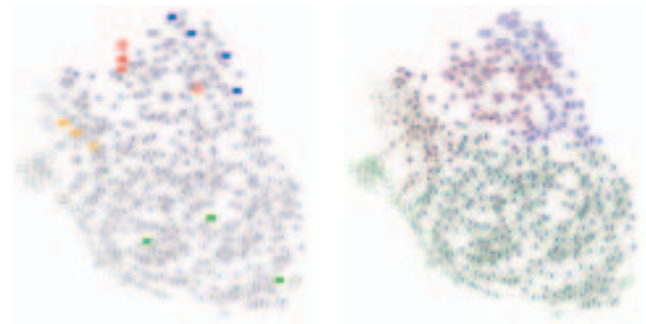


**Fig. 2.** Feulgen nuclei extraction. (a) Initial image. (b) Pre-processed image (using a closing operation). (c) Initial labels for segmentation. (d) Extracted nuclei. See text for details.

label evolution to produce smooth regions. This segmentation scheme has been evaluated on hundred images by a cytopathologist (per nuclei visual assessment) and 98% of the nuclei are extracted and correctly delineated.

### 3.3. Data Clustering

The last step of both methods, which comes after nuclei segmentation and features computation processes, is to determine if some of the extracted nuclei are abnormal. Such problem can be seen as a data classification one, in two (or more) classes: the data to classify is the set of nuclei (represented by their feature vectors) and typical classes are *normal* and *abnormal*. Finally, the classification problem can be seen as a semi-supervised graph clustering problem, where the data is represented as a  $k$ -nearest-neighbor graph (each feature vector is represented by a vertex which is connected to its  $k$  most similar vertices in the whole graph, according to a similarity function). The clustering is performed using our label-propagation algorithm on graphs, with a label per class. Labels seeds are given by a reference data-set (where the class of each nucleus is known) the entries of which are added to the graph. Such a clustering method is presented as a complement of the two nuclei extraction methodologies presented above, and illustrated on Fig. 3 with a 4-labels semi-supervised graph clustering example. Using graph-based label propagation for clustering the set of nucleus has two main advantages. First there is no classifier training, and modifying the reference data-set does not imply to re-train a classifier (as it is the case while using neural networks for exam-



**Fig. 3.** Illustration of nuclei classification using, semi-supervised graph clustering and our label-propagation algorithm. Left: seeds of each class. Right: classification result.

ple). Second graphs intrinsically provide a representation of the organization of the different classes and the position of each nucleus in each class. Such an information can be of crucial importance for cytopathologists on ambiguous cells.

## 4. CONCLUSION

In this paper, we proposed a graph-based approach using our formalism of PdEs-based morphology to address the problem of cytological computer-aided diagnostic. The approach is very general and has been illustrated through two methodologies, using different graph-based image representations (according to the coloration), to perform nuclei extraction. In both cases, the proposed methodologies has given very good and promising results.

## 5. REFERENCES

- [1] A. Elmoataz, O. Lézoray, and S. Boughleux, “Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing.,” *IEEE Trans Image Process*, vol. 17, no. 7, pp. 1047–60, 2008.
- [2] V.-T. Ta, A. Elmoataz and O. Lézoray, “Partial difference equations over graphs: Morphological processing of arbitrary discrete data.,” in *Proc. ECCV*, pp. 668–680, 2008.
- [3] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, 1999.
- [4] X. Desquesnes, A. Elmoataz and O. Lézoray, “PDEs level sets on weighted graphs,” in *Proc. ICIP*, 2011.
- [5] X. Desquesnes, A. Elmoataz, O. Lézoray and V.-T. Ta, “Efficient algorithms for image and high dimensional data processing using eikonal equation on graphs.,” in *Proc. ISVC*, vol. 2, pp. 647–658, 2010.