HAL
open science

A unified geometric model for virtual slide images processing

Matthieu Toutain, Xavier Desquesnes, Abderrahim Elmoataz, Olivier Lezoray

To cite this version:

Matthieu Toutain, Xavier Desquesnes, Abderrahim Elmoataz, Olivier Lezoray. A unified geometric model for virtual slide images processing. International Workshop on Adaptation and Learning in Control and Signal Processing (IFAC), Jul 2013, Caen, France. pp.629-634. hal-00932532

HAL Id: hal-00932532

https://hal.science/hal-00932532v1

Submitted on 3 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Unified Geometric Model For Virtual Slide Images Processing ⋆

M. Toutain * **X. Desquesnes** * **A. Elmoataz** * **O. Lézoray** *

* *Université de Caen Basse-Normandie, ENSICAEN, CNRS, GREYC Image Team.*

Abstract:
In this paper, we propose a unified framework to address the problem of cytological computer-aided diagnosis. Such an approach relies on our previously introduced formalisms: general formulation of discrete functional regularization, PDEs based morphology and geometric diffusion on graphs. The approach is illustrated through two applications in cytopathology, with examples of nucleus extraction and 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 partial difference equations (PdEs) based morphology on graphs to address the problem of cytological computer-aided diagnostic through different subproblems: image preprocessing, cell extraction and cell classification. The rest of this paper is organized as follows. In Section 2, we introduce the microscopic imaging problems and propose a strategy to address them. In Section 3, we provide definitions and notations on graphs that will be used in the next sections. Sections 4, 5, and 6 present the different steps of our methodology for processing virtual slide images. Section 7 concludes this paper.

2. MICROSCOPIC IMAGING PROBLEMS

In this section, we briefly introduce pathology and virtual slide images, in order to propose our approach to process such data.

⋆ This work was supported under a doctoral grant of the Conseil Régional de Basse Normandie and of the Coeur et Cancer association in collaboration with the Department of Anatomical and Cytological Pathology from Cotentin Hospital Center.

2.1 Modeling

Pathology is roughly composed of two sections: cytology and histology. For both these sections, the visual inspection of cellular specimens and histological sections through a light microscope plays an important role in clinical medicine and biomedical research. Cytology literally means the study of cells. It studies morphological features of human body fluid cells which are put on a glass slide and stained. The study of the modification of the main cellular components (nucleus and cytoplasm) is the fundament of the cytological study. The morphological features of cells are visually evaluated by cytotechnologists and cytopathologists and these features involve several notions including size, shape, color, texture and topography. The interaction between nucleus and cytoplasm is also of interest: the position of the nucleus in the cytoplasm, the nucleus–cytoplasm area ratio, the position of nucleols in the nucleus, the color and the granularity of the cytoplasm.

Virtual slide images In this paper we focus on virtual slide images of cytological slides stained with Feulgen and Papanicolaou coloration. Virtual slides are digitized images of slides. A Leica SCN 400 digital microscopical scanner is used to completely digitize our slides at 40× magnification scale. The resulting images are compressed with a quality of 75% in JPEG compression format. To facilitate the visualization and the processing, scanned samples acquired by the scanner are directly stored as an irregular pyramid where each level of the pyramid is an underresolved version of the highest resolution image (the pyramid base). The images being very large (about 80000×80000 at the highest resolution), each resolution level is split into image tiles (of size 512×512) in a non-overlapping layout. The usual size of a compressed slide image is about 2 Gygabytes.

Image object oriented modeling As discussed above, image processing methods are of high interest to provide Image Decision Guided Systems (IDGS) to perform prognostic, diagnostic and early detection of cancer. For the case of problems in pathology, image processing can be used for several tasks: quantification of cellular content (DNA, proteins, color), recognition and sorting of cellular

types, extraction of cellular groupings or clusters and of their topographic relations. In terms of image processing problems, the objectives concern segmenting or analyzing objects at different levels: cellular or cellular grouping level. Cytological images have common properties and can be described by an image object-oriented modeling. First of all, the images have to be acquired in color. This is essential to follow the exact visual way the pathologists evaluate microscopic images. For a microscopic image, one can always divide it into two parts: the background and the rest of the image, namely the objects to be extracted. The background is always close to a given color and is usually homogeneous even if some debris or artifacts can occur (some mucus for instance). The rest of the image is composed of the elements of interest for the pathologists. In this work, we use an image object-oriented modeling of the elements of an image: they can be classified and characterized in terms of size, shape, color and texture. Before trying to conceive any image segmentation technique, a pool of representative images has to be built to have a model of the problem which is as close as possible to reality. Once this database has been constructed, a segmentation strategy can be conceived.

2.2 Our proposed scheme

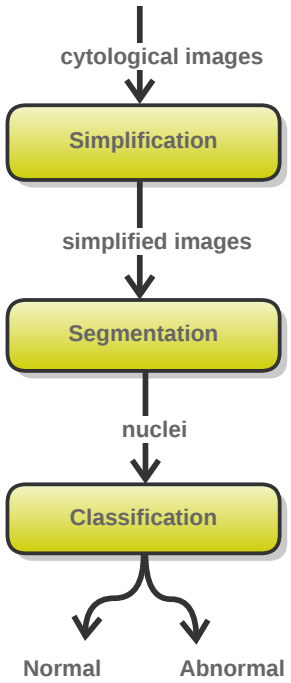


Fig. 1. System overview.

Our approach for processing cytological virtual slide images is illustrated in fig. 1. First step is the simplification of images that directly come from the slide's images at the highest resolution. Simplification can be seen as a pre-processing step, in order to ease the second step : segmentation. Here we separate nuclei from anything else in the slide. Once segmentation is done, we have detected all the nuclei in the slide image. The last step of our strategy (classification) is first to extract features of the segmented nuclei, then to classify these nuclei, according to their feature vector. Simplification, segmentation, and classification are performed with partial differential equations (PDEs) on graphs.

3. 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 (Elmoataz et al. [2008], **?**]).

Notations : We assume that any discrete domain can be modeled by a weighted graph. Let $G = (V, E, w)$ be a weighted graph composed of two finite sets: $V = \{u_1, ..., u_n\}$ of n vertices and $E \subset V \times V$ a set of weighted edges. The weight function $w : V \times V \to [0, 1]$ represents a similarity measure between two vertices of G. According to w, the set of edges is defined as: $E = \{(u, v)|w(u, v) \neq 0\}$. We denote by $N(u)$ the neighborhood of a vertex u, i.e., the subset of vertices that share an edge with u, and the notation $v \sim u$ means that $v \in N(u)$. 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 assign 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 $\partial^+\mathcal{A} = \{u \in \mathcal{A}^c | \exists v \in \mathcal{A}, v \sim u\}$ and $\partial^-\mathcal{A} = \{u \in \mathcal{A} | \exists v \in \mathcal{A}^c, v \sim u\}$ 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 $\mathcal{G}_w : \mathcal{H}(V) \to \mathcal{H}(E)$ is defined on an edge $(u, v) \in E$ by

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

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 $\mathcal{G}_w^{\pm} : H(V) \to H(E)$, by

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

with the following notations: $(x)^+ = \max(x, 0)$ and $(x)^- = -\min(x, 0)$. The weighted gradient of a function $f \in \mathcal{H}(V)$ at vertex u is defined as the vector of all weighted gradients with respect to the set of edges $(u, v) \in E$

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

This operator corresponds to the local variation of the function f at the vertex u and measures the regularity of f in the adjacent neighborhood $N(u)$. Hence the $\mathcal{L}_2$-norm of the weighted gradient operator is

$$||(\nabla_w f)(u)||_2 = \left[\sum_{v \sim u} w_{uv}(f(v) - f(u))^2\right]^{1/2}. \tag{4}$$

In the sequel, the 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

$$(\nabla_w^{\pm} f)(u) = ((\mathcal{G}_w^{\pm} f)(u, v))_{v \in V} \tag{5}$$

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

$$\begin{aligned} ||(\nabla_w^{\pm} f)(u)||_p &= \left[\sum_{v \sim u} w_{uv}^{p/2} |(f(v) - f(u))^{\pm}|^p\right]^{1/p}, \\ ||(\nabla_w^{\pm} f)(u)||_\infty &= \max_{v \sim u}(\sqrt{w_{uv}} |(f(v) - f(u))^{\pm}|). \end{aligned} \tag{6}$$

4. IMAGE SIMPLIFICATION / FILTERING

As described in the previous section, our proposed scheme to address the problem of processing virtual slide images is to separate it into the three following problems: preprocessing, cell extraction, and classification, all under our

graph based framework. In this section, we focus on the preprocesing step. While working on very large data (such as virtual slides), many algorithms become inefficient due to the mass of data to process, and the variability of the data. A simple and efficient way to avoid this problem is to work on a simplified version of the data. In this paper, we consider cytology slide images, with Feulgen and Papanicolaou coloration. Feulgen coloration is quite computer-vision-friendly, because the slide can be seen as a set of two easily separable classes: the background (in grey) and the nucleus (in pink). Papanicolaou coloration is the most used in cytology, and stains cytoplasms in light green and nuclei in dark blue. However, the nuclei boundaries are often unsharp due to refractiveness or mucus artifacts standing around the nucleus. In order to ease the segmentation step, we use filtering as a preprocessing step for this kind of coloration.

4.1 Simplification of the images

In image processing, the graph data-structure is commonly used to represent a digital image. For our filtering process, we use a grid graph, which is a natural structure corresponding to the definition of digital images: vertices represent pixels and edges represent pixel adjacency relationship. From the previous definitions, we introduce here the *weighted p-Laplace operator* $\Delta_w^p f(u)$ at vertex u :

$$\Delta_w^p f(u) = \sum_{v \sim u} \gamma(u,v)(f(v) - f(u)), \tag{7}$$

where

$$\gamma(u,v) = w(u,v)(||(\nabla_w f)(u)||_2^{p-2} + ||(\nabla_w f)(v)||_2^{p-2}). \tag{8}$$

In the case where $p = 1$ and 2, we have the definitions of the standard graph curvature $\Delta_w^1 f(u) = \kappa f$ and graph Laplace $\Delta_w^2 f(u) = \Delta f$ operators. More details on these definitions can be found in Elmoataz et al. [2008].

Discrete regularization framework To regularize a function $f^0 \in \mathcal{H}(V)$ using the p-Laplace operator (7), we consider the following general variational problem on graphs:

$$\min_{f \in \mathcal{H}(V)} \{E_{w,p}(f, f^0, \lambda) = R_{w,p}(f) + \frac{\lambda}{2}||f - f^0||^2\}. \tag{9}$$

The intuition behind regularization is to provide a smoother version of an initial function f^0 (with the regularizer term) while keeping it close to the initial function (with the fitting term). The first term, $R_{w,p}(f)$, is the regularizer and is defined as the discrete Dirichlet form of the function $f \in \mathcal{H}(V) : R_{w,p}(f) = \frac{1}{2}\sum_{u \in V} ||\nabla_w f(u)||_2^p$. The second term is the fitting term. $\lambda \geq 0$ is a fidelity parameter called the Lagrange multiplier which specifies the trade-off between the two competing terms. Both terms of $E_{w,p}(f, f^0, \lambda)$ in (9) are strictly convex function of f (Chan et al. [2001]). By a standard argument in convex analysis, this optimization problem has a unique solution for $p = 1$ and 2 which satisfies, for all $u \in V$:

$$\frac{\partial}{\partial f(u)} E_{w,p}(f, f^0, \lambda) = \Delta_w^p f(u) + \lambda(f(u) - f^0(u)) = 0. \tag{10}$$

Equation (10) can be viewed as the discrete analogue of the Euler-Lagrange equation. Using the p-Laplacian formulation (equation. (7)) in equation (10), the optimization problem solution is also the solution of the following system of equations. For all $u \in V$,

$$\left(\lambda + \sum_{v \sim u} \gamma(u,v)\right) f(u) - \sum_{v \sim u} \gamma(u,v) f(v) = \lambda f^0(u). \tag{11}$$

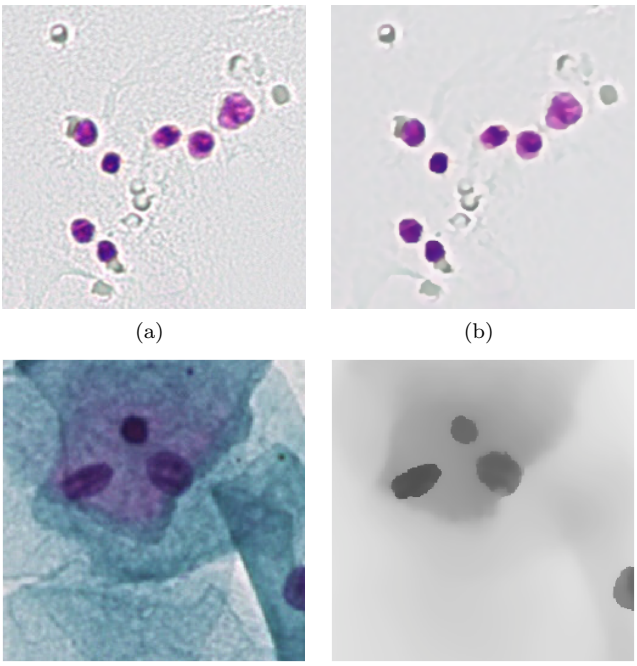
To approximate the solution of the minization (9), we can linearize this system of equations and use the Gauss-Jacobi method to obtain the following iterative algorithm:

$$\begin{cases} f^{(0)}(u) = f^0(u), \\ f^{(t+1)}(u) = \dfrac{\lambda f^0(u) + \sum_{v \sim u} \gamma^{(t)}(u,v) f^{(t)}(v)}{\lambda + \sum_{v \sim u} \gamma^{(t)}(u,v)}. \end{cases} \tag{12}$$

where $\gamma^{(t)}(u,v)$ is the γ function (in equation (8)) at the iteration step t. At each iteration of the algorithm, the value of f at step $(t+1)$, for a vertex u, only depends on two quantities: the original value f^0 and the sum of the weighted existing values $f^{(t)}$ in the neighborhood of u. By using different formulations of w and different values of p, a family of linear and nonlinear filters is obtained. Indeed, when $p = 2$ and $w(u,v) = 1$ one obtains the linear diffusion on graphs. When $p = 1$ and $w(u,v) = 1$ one recovers the TV digital filter (Chan et al. [2001]). The reader can note that this isotropic regularization corresponds to the weighted discrete transcription of the regularization functional in the continuous case. The interested reader can refer to Elmoataz et al. [2008] for more details on the formulation and the connections with other formalisms.

Image filtering Through the values of the p parameter, the discrete regularization (10) describes a family of linear and nonlinear filters. This image filtering/denoising can be viewed as an image simplification that can ease the seed extraction step. Before processing acquired cytological whole slide, they are split into thousands (it varies, depending on the size of the acquired virtual slide) of 1024×1024 tiled images. All images presented in this paper are very small details extracted from some of the processed tiled images and obviously can not represent the hugeness of digital slides. Fig. 2(a) and 2(b) show a filtering with $p = 2$ and $w(u,v) = \mathrm{e}^{\frac{-(d(u,v))^2}{\sigma^2}}$, with $d(u,v)$ the Minkowski distance of order 2 in the RGB color space, on a Feulgen cytological image represented by an 8-adjacency grid-graph, i.e., a graph that connects a pixel to its neighbors in a 3×3 spatial window centered on the pixel. During regularization, each color channel is processed separately. Fig. 2(c) and 2(d) show a filtered version of the v* channel of the CIELUV color space (see Ohta and Robertson [2006] for details on color spaces) of a Papanicolaou stained image (v* caracterising a blue-green opposition, and being the most discriminant to extract nuclei for Papanicolaou staining), with $p = 2$ and $w(u,v) = \mathrm{e}^{\frac{-(f(v)-f(u))^2}{\sigma^2}} \mathrm{e}^{-\left(\frac{(x_v - x_u)^2}{2\sigma_s^2} + \frac{(y_v - y_u)^2}{2\sigma_s^2}\right)}$, with (x_u, y_u) and (x_v, y_v) the corresponding spatial coordinates in the image of, respectively, u and v. In this case, we construct a grid graph with edge connections in a spatial

square window of width related to σ_s. One can note that this last weighting function permits to recover a bilateral filter (Tomasi and Manduchi [1998]).



(a) (b) (c) (d)

Fig. 2. Examples of cytological image pre-processing.(a) initial Feulgen stained image. (b) Preprocessed image using discrete regularization framework. (c) initial Papanicolaou stained image (gynecology). (d) Filtered version of v* channel of CIELUV color space of the initial image.

5. IMAGE SEGMENTATION

In this section we focus on the cell extraction step of our strategy. To extract cells, we perform a segmentation of the simplified images obtained from the pre-processing step. As the algorithm we detail in this section performs a label propagation, we first extract seeds from the simplified images that will be used as initial labels to propagate. We first provide details and recalls on PdEs based morphology to introduce our methods, which is also based on the previous definitions provided on graphs and operators on graphs. We then make a link with the static version of the eikonal equation (for further details, see Desquesnes et al. [2013]), and show the use of front propagation for initial label propagation. Results are shown in the last part of this section.

5.1 Eikonal equation on weighted graph

PdEs based Morphology Discrete dilation and erosion on weighted graphs are defined by

$$\partial_t f(u) = +||(\nabla_w^+ f)(u)||_p \text{ and } \partial_t f(u) = -||(\nabla_w^- f)(u)||_p, \tag{13}$$

respectively. These equations (13) constitute a PdEs based framework (Ta et al. [2011]) 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. To solve these dilation and erosion process (equations 13), we use the iterative scheme proposed in Ta et al. [2011]. For the dilation : $f^{t+1}(u) = f^t(u) + ||(\nabla_w^+ f^t)(u)||_p$, and for the erosion : $f^{t+1}(u) = f^t(u) - ||(\nabla_w^- f^t)(u)||_p$.

Indeed, given, a set of vertices $\mathcal{A} \subset V$ and using external and internal graph boundaries, the 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 $\partial^-\mathcal{A}$. Let Γ be a parametrized curve evolving on a domain Ω. A very common way to describe the evolution of Γ, popularized by Sethian [1999], is to embed the curve in a function $\phi(x,t)$ such that the evolving curve Γ_t can be provided by the 0-level set of ϕ. Once formulated as a level set problem, the curve evolution amounts to solving the following equation:

$$\frac{\partial \phi}{\partial t} = \mathcal{F}|\nabla \phi|. \tag{14}$$

In Desquesnes et al. [2011], a transcription of (14) has been proposed to weighted graphs, that can be expressed as a morphological process with the following sum of dilation and erosion.

$$\frac{\partial \phi}{\partial t}(u) = \mathcal{F}^+||(\nabla_w^+ \phi)(u)|| + \mathcal{F}^-||(\nabla_w^- \phi)(u)||, \tag{15}$$

where $\mathcal{F} \in \mathcal{H}(V)$ controls the front propagation. Such a formulation enables recovery of geometric diffusion models such 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 grids and some particular non-Cartesian domains (see Desquesnes et al. [2013] and references therein). Recently, two adaptations of the eikonal equation have been proposed, first as a time dependent version (Ta et al. [2011]), then as a static version (Desquesnes et al. [2013]) which is expressed as

$$\begin{cases} ||(\nabla_w^- f)(u)||_p = P(u), & \forall u \in V, \\ f(u) = 0, & \forall u \in V_0, \end{cases} \tag{16}$$

where $V_0 \subset V$ corresponds to the initial set of seeds vertices, and P a potential. Such adaptations are expressed using the PdEs based morphological erosion, and can be linked with the general geometric PdE equation (15). In Desquesnes et al. [2013], a label propagation algorithm has been proposed based on the resolution of (16), 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, such as geodesic distance computation on graphs, image segmentation and data clustering.

5.2 Nuclei extraction by label propagation

Seeds extraction In Desquesnes et al. [2012], a method is proposed to automatically extract seeds from the simplified Feulgen image resulting from the preprocessing step. As we can see in fig. 2, 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 (see fig. 3(a)). For Papanicolaou stained images, an H-minima method (see Soille [2003] for details on H-minima) is used on the simplified version to automatically get seeds from the nuclei. Considering the set S of nuclei seeds, an erosion is performed on S^c to get non-nuclei seeds that do not overlap potential nuclei (see fig. 3(c) for illustration).

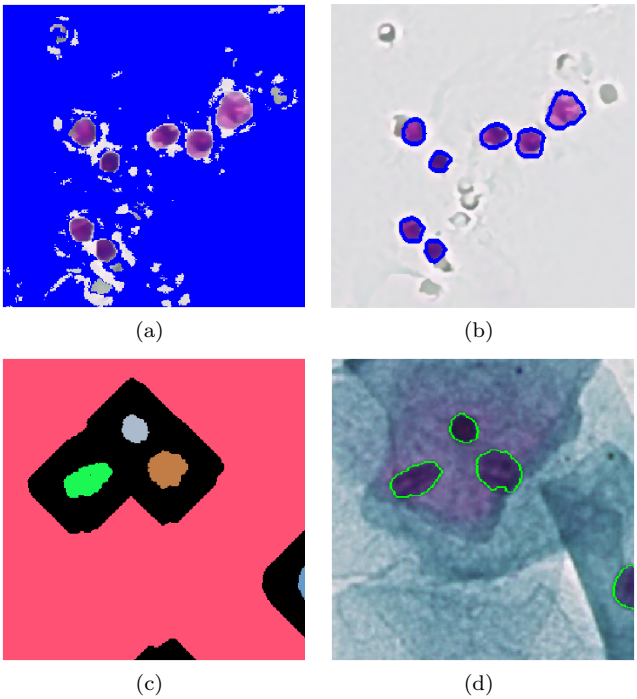


Fig. 3. Feulgen nuclei extraction : (a) Initial labels (in blue) computed on Feulgen stained simplified image. (b) extracted nuclei. (c) Eroded background seeds (in red) and initial nuclei seeds (other colors but black) computed on Papanicolaou stained simplified image. (d) extracted nuclei.

Label propagation Once background seeds have been extracted, they are propagated on the image grid-graph until being stopped by nuclei boundaries. Using equation (6) and equation (16), we get for $p \neq +\infty$:

$$\left(\sum_{v\sim u} w_{uv}^{p/2} \max(0, (f(u) - f(v)))^p\right) = P(u),\ p \in \{1, 2\}, \tag{17}$$

and for $p = +\infty$:

$$\max_{v\sim u}(\sqrt{w_{uv}} \max(0, f(u) - f(v))) = P(u),\ p = \infty. \tag{18}$$

For the sake of clarity, we only present here the solution for the $\mathcal{L}_\infty$ norm formulation (18):

$$\overline{x} = \min_{j=1}^{n}(a_j + h_j C), \tag{19}$$

where $x = f(u)$, $n = \text{card}(N(u))$, $a_i = \{f(v_i)|v_i \in N(u)$ with $i = 1, ..., n\}$, $h_j = 1/\sqrt{w_{uv}}$, and $C = P(u)$.

Further details on the front propagation algorithm can be found in Desquesnes et al. [2013]. The stopping criterion is held by the potential P of equation (16), as

$$P(u) = 1/((\mu_N - f(u))^2 - (\mu_B - f(u))^2 + \mathcal{K}(u)), \tag{20}$$

where μ_N and μ_B are respectively the mean color of nuclei and background, $\mathcal{K}$ is the graph-curvature term that constrains the label evolution to produce smooth regions. The global solution (over the entire graph) is computed in a simple way: let NB be a narrow band of vertices to update (where the distance has to be computed), A a list of already updated vertices, and FA a set of vertices which are not in one of these sets. The algorithm is initialized as follows: the initial seeds set (V_0) is added to A, vertices on the outside boundary of A are added to NB, and the remaining vertices compose FA. The computed distance at a vertex u is indeed the distance to the nearest seed from u (with respect to the graph topology, the weighting function w, and the potential P). So the front arriving at a vertex u is necessarily coming from the nearest seed of u, and it becomes easy to associate the right label with it, looking at the labels of the neighbors of u (at least one being marked by the front already passed). The labeling process can be summarized by the following formula: each time $f(u)$ is updated, the label $L(u)$ is given by

$$L(u) = L(v)|v \in N(u), f(v) < f(u) \text{ and } \frac{f(v)}{w_{uv}} = \min_{z\sim u}(\frac{f(z)}{w_{uz}}). \tag{21}$$

The results of this segmentation scheme are illustrated in fig. 3(b) for Feulgen images and fig. 3(d) for Papanicolaou images. It has been evaluated on a hundred images by a cytopathologist (per nuclei visual assessment) and 98% of the nuclei are extracted and correctly delineated.

6. OBJECT CLASSIFICATION / CLUSTERING

In this section, we show that data classification problems can be addressed using the same framework as in the previous sections. Examples on Feulgen stained cytological images will be used to illustrate our method.

As said previously, the last step of our scheme is to classify nuclei extracted from the segmentation step, to determine if some are abnormal. This is a data classification problem, into two (or more) classes: the data to classify is the set of nuclei, and typical classes are *normal* and *abnormal*.

6.1 Characteristics extraction

To begin with, we have to represent nuclei as data to classify. From the segmentation process, we first compute characteristics on each extracted nucleus (area, compacity, ...), which form a feature vector describing each nucleus.

6.2 Classification

The classification problem can finally 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 the previously seen label-propagation algorithm on graphs, with a label per class. Label 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 illustrated in fig. 4 with a 4-labels semi-supervised graph clustering example. Using graph based label propagation for clustering the set of nuclei has two main advantages. First there is no classifier training, and modifying the reference data set does not imply to retrain a classifier (as it is the case while using neural networks for example). Second, graphs intrinsically provide a representation of the organization of the different classes and the position of each nucleus in each class. Such information can be of crucial importance for cytopathologists in the case of ambiguous cells.

7. CONCLUSION

In this paper, we proposed a graph based approach using a general formulation of discrete functional regularization, PDEs based morphology and geometric diffusion on weighted graphs of arbitrary topology to address the problem of cytological computer-aided diagnosis. The approach is very general and has been illustrated both on Papanicolaou and Feulgen stained slides. In both cases, the proposed unified methodology has given very good and promising results.

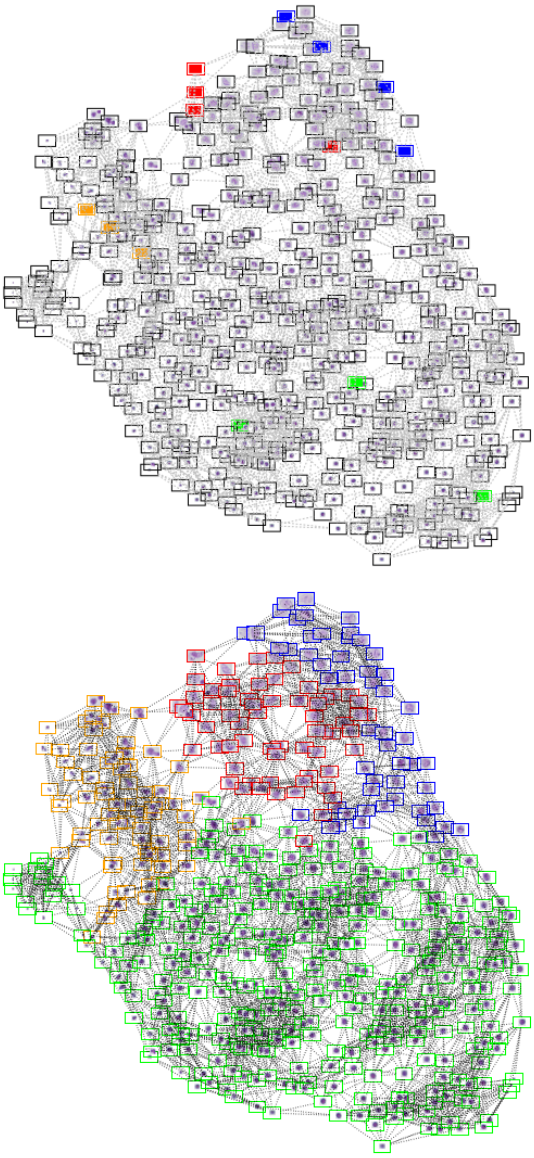


Fig. 4. Nuclei classification using semi-supervised graph clustering via the label propagation algorithm. Top: seeds of each class. Bottom: classification result.

REFERENCES

T. Chan, S Osher, and J Shen. The digital tv filter and nonlinear denoising. *IEEE Transactions on Image Processing*, 10:231–241, 2001.

X. Desquesnes, A. Elmoataz, and O. Lezoray. Pdes level sets on weighted graphs. In Benoît Macq and Peter Schelkens, editors, *ICIP*, pages 3377–3380. IEEE, 2011. ISBN 978-1-4577-1304-0.

X. Desquesnes, A. Elmoataz, and O. Lézoray. Pdes-based morphology on graphs for cytological slides segmentation and clustering. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 1619–1622. IEEE, 2012.

X. Desquesnes, A. Elmoataz, and O. Lézoray. Eikonal equation adaptation on weighted graphs: fast geometric diffusion process for local and non-local image and data processing. *Journal of Mathematical Imaging and Vision*, 2013.

A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE transactions on Image Processing*, 17(7):1047–1060, 2008.

N. Ohta and A. Robertson. *Colorimetry: fundamentals and applications*. Wiley, 2006.

J.A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999.

Pierre Soille. *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc., 2003.

V.-T. Ta, A. Elmoataz, and O. Lézoray. Nonlocal pdes-based morphology on weighted graphs for image and data processing. *IEEE transactions on Image Processing*, 20(6):1504–1516, June 2011.

Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.