An Exploration Scheme for Large Images: Application to Breast Cancer Grading

Antoine Veillard  
Dept. of Computer Science  
National University of Singapore  
Singapore  
Email: veillard@comp.nus.edu.sg

Nicolas Loménie  
LIPADE  
University Paris Descartes  
Paris, France  
Email: Nicolas.Lomenie@mi.parisdescartes.fr

Daniel Racoceanu  
IPAL-UMI2955  
CNRS, France  
ASTAR, Singapore  
Email: daniel.racoceanu@ens2m.fr

Abstract—Most research works focus on pattern recognition within a small sample images but strategies for running efficiently these algorithms over large images are rarely if ever specifically considered. In particular, the new generation of satellite and microscopic images are acquired at a very high resolution and a very high daily rate. We propose an efficient, generic strategy to explore large images by combining computational geometry tools with a local signal measure of relevance in a dynamic sampling framework. An application to breast cancer grading from huge histopathological images illustrates the benefit of such a general strategy for new major applications in the field of microscopy.

Keywords—very large image; computational geometry; histopathology;

I. INTRODUCTION

The exploration of large images for object recognition and classification purposes will be a challenging research topic in the coming years. In satellite imaging as well as in microscopic imaging, the size and the scale to which images are acquired raise new issues about efficient strategies to analyze these data in a reasonable time. Most research works focus on algorithms for analyzing small patches while strategies for running efficiently these algorithms over large images are rarely if ever specifically considered. This work proposes a generic algorithmic scheme for such explorations based on computational geometry tools and proceeding by dynamic sampling of the signal. For illustration, a first validation is proposed for microscopic images in the field of computer-aided breast cancer classification and grading.

Analysis of histopathology images is of paramount significance for the diagnosis and prognosis of some of the world’s deadliest diseases such as breast cancer. Today, breast biopsy grading is still manually performed under a light microscope following a standard procedure called Nottingham grading consisting of the assessment of three parameters among which nuclear pleomorphism plays a major role (see [1] for an accessible overview). Grading nuclear pleomorphism involves the classification and the counting of individual cell nuclei in high magnification images. As proved by Dune and Going [2], the outcome of this tedious and time consuming task is highly inconsistent even for well trained specialists. Therefore, a system able to perform the grading from a virtual slide in a more systematic way could bring large benefits in terms of stability and reliability of the prognosis. Numerous algorithmic challenges related to the automated assessment of nuclear pleomorphism are already being actively investigated [3]. Accordingly, the analysis of individual high magnification small patches (referred as frames or hyperfields) is already efficient to some extent [4]–[6].

However, a single biopsy virtual slide is a Very Large Image (VLI) commonly comprising several thousands of high magnification frames, making an exhaustive analysis of all of them not feasible. Our system combines a specific measure of local relevance together with a generic dynamic sampling method based on computational geometry. In the case of our medical application, it is able to provide both an accurate and time efficient solution for the grading of full biopsy slides. Section II details the generic algorithm. Then, in Section III we compare random sampling versus our informed (high-level guided) sampling algorithm.

II. METHODOLOGY

Let $I$ be a VLI split into a large number of rectangular frames $x \in I$. For every frame $x$, a specific measure of local relevance $S(x)$ (referred as “score”) can be computed. The goal of our algorithm (referred as EX-grad) is to efficiently locate the frames in $I$ having the largest relevance score $S(x)$. A dynamic sampling method is first used to identify a subset of the most relevant frames (with high $S(x)$) from which a global map of the score is extrapolated.

A. Local assessment

Ideally, the local relevance score $S(x)$ should be a semantic information specific to the context of the application such as the nuclear pleomorphism in the microscopic image application. Alternatively, when such an information is not available, it can be a low-level feature characterizing the amount of information available such as the compression rate.
**Nuclear Pleomorphism:** In the microscopic images application, the local score is the nuclear pleomorphism \( S_{NP} \) indicating the level of malignancy of a cancer.

First, the high-magnification frames are analyzed following previous work in [6]. Nuclei are segmented using Gaussian color models and classified according to their color distribution as grade 1, 2 or 3. A high grade is a sign of Gaussian color models and classified according to their color application, the local score is the nuclear pleomorphism \( S_{NP} \).

Then \( S_{NP} \) is computed with a score function defined by the formula: \( S_{NP}(x) = \alpha \times p_1 + \beta \times p_2 + \gamma \times p_3 \) where \( p_1 \), \( p_2 \) and \( p_3 \) are the proportions of nuclei of grade 1, 2 and 3 in the frame \( x \) and \( 0 \leq \alpha \leq \beta \leq \gamma \leq 1 \) are fixed parameters. Note that \( S_{NP} \) takes a value from \([0,1]\), high values corresponding to the most malignant levels of pleomorphism.

**Compression Ratio:** For more generic applications such as outdoor robotic vision, the local jpg compression rate \( S_{CR} \) can be used. It is defined by the formula: \( S_{CR}(x) = \text{size}(\text{jpg})/\text{size}(\text{bmp}) \) where \( \text{size}(\text{bmp}) \) stands for the byte size of the image stored in the.bmp format.

Maps obtained with the two different score functions on the same biopsy slide are shown on Fig. 1. The high level of similarity between the two maps indicates that the low-level \( S_{CR} \) can be used as an alternative to \( S_{NP} \) when such high-level information is not available.

**B. Dynamic sampling**

The frame sampling procedure is a dynamic and incremental scheme based on computational geometry tools. At each iteration, given \( E \) the frames already sampled in the VLI \( I \), we construct the Voronoi diagram of the centroids of the frames in \( E \) denoted as \( \text{Vor}_E \). \( \text{Vor}_E \) is a collection of Voronoi cells \( \{\nu_x | x \in E\} \), defined by \( \nu_x = \{p \in I | \forall y \in I - \{x\}, \text{dist}(p_x, x) \leq \text{dist}(p, y)\} \). The set of Voronoi vertices, later referred as \( V_E \) are the vertices of the planar graph representation of \( \text{Vor}_E \). Voronoi vertices share the property to be locally the farthest position from their nearest neighbor in \( E \), therefore in the case of our algorithm from already sampled frames.

This geometric construction is aimed at approximating the score \( S \) within a whole Voronoi cell by the score of the frame at its center which results in a nearest neighbor approximation. Accordingly, the most undetermined areas are at the intersection of multiple cells, i.e. frames containing a vertex from \( V_E \). We select our next sample \( x \) out of \( V_E \) following two criteria:

1) **At least one of its neighboring cells has a high score.** Practically, we check that the score \( \text{MaxScore}(x) \) of its highest scoring neighbor in \( E \) is higher than \( p \times \text{max}_E \) where \( \text{max}_E \) is the currently observed maximal score among \( E \) and \( p \in [0,1] \) is a preset parameter defining the selectivity of the algorithm. This condition controls the convergence of the algorithm towards areas with high scores.

2) **The distance between the new sample and its neighbors is not too short.** Practically, we want \( \text{dist}(x, E) \geq d \) where \( d \in [0,\infty] \) is a parameter determining the fineness of sampling. This condition prevents oversampling.

The pseudo-code for one iteration of the sampling algorithm is given in Figure 2.

**Input:** current samples \( E, \text{Vor}_E, p, d, \text{max}_E \)  
**Output:** updated values of \( E, \text{Vor}_E, \text{max}_E \)

1: compute \( V_E \)  
2: sort \( V_E \) according to decreasing distance to \( E \)  
3: for every \( x \in V_E \) do  
4:     if \( \text{dist}(x, E) \geq d \) then  
5:         if \( \text{MaxScore}(x) \geq p \times \text{max}_E \) then  
6:             \( E = E \cup \{x\} \)  
7:         update \( \text{Vor}_E \)  
8:         \( \text{max}_E = \text{max}(S(x), \text{max}_E) \)  
9:     end if  
10: end if  
11: else  
12:     break loop  
13: end if  
14: end for

Figure 2. One iteration of the dynamic sampling algorithm

Incremental construction of the Voronoi diagram ensures that the cost of selecting all the necessary samples remains negligible compared to the cost of frame analysis above all in the case of the informed score \( S_{NP} \). The sampling phase is initialized with three arbitrarily selected frames. Choosing centroids of connected components based on low resolution gray scale analysis has proved to work fast and well. The iterative sampling algorithm is run until depletion of candidate samples. In practice, the parameters \( d \) and \( p \) are adapted during the whole process by successively taking lower values of \( d \) and higher values of \( p \) every time samples are depleted. The rationale behind this is to adapt...
the coarseness of sampling to the score of the regions: regions with homogeneously low scores are assumed to be less interesting and therefore to require less exploration than regions with higher or more heterogeneously distributed scores. Fig. 3 illustrates the evolution of sampling over a biopsy slide.

(a) After 50 samples: the whole VLI is being explored. No area seems favored.

(b) After 150 samples: the algorithm converges towards a high grade area.

(c) After 400 samples: the sampling is very dense around this area and remains sparse in others.

(d) The highest grading area superimposed over a low magnification image of the VLI.

Figure 3. Dynamic sampling method applied to a histopathological VLI of size 59,000 x 44,000 pixels. The $S_{NP}$ score has been used. The incrementally constructed Voronoi diagrams are shown in black. Each cell contains a single sample at its center. The maps resulting from the interpolation are shown in colors. Hot colors represent higher grades.

Finally, a global map of the scores over the whole VLI $I$ is extrapolated from the scores of the sampled frames. The map is expected to describe accurately the regions with a high local relevance score.

III. EXPERIMENTS AND DISCUSSION

We evaluate our methodology in the field of breast cancer grading from histopathological images. The grading consists in the assessment of three parameters among which the nuclear pleomorphism indicating the malignity of cancerous cells plays a central role. Therefore, we use the measure of pleomorphism $S_{NP}$ presented in Section II-A as our local relevance score. A medical validation of the computer vision algorithms used to produce $S_{NP}$ is available in [6]. Parameters for $S_{NP}$ have been set to $\alpha = 0$, $\beta = 0.5$ and $\gamma = 1$ following medical advice.

Two different extrapolation paradigms have been used to produce the global map of peomorphism from the samples: (a) a nearest neighbors framework and (b) an analogy with spring mechanics where every frame is linked to its four neighbors by virtual springs of length zero and equal stiffness.

The test set provided by our partner hospital consists of a total of 20,696 HE stained biopsy frames from four different breast cancer patients. The resolution of individual images captured at 40× magnification is 640 by 521 pixels and the field-of-view is one fourth of a standard light microscope which is intentionally narrow in order to increase the meaningfulness of sampling. A typical VLI from the dataset has a 59,000 x 44,000 pixels resolution (8 GB data).

Performances are measured for the retrieval of the set $Rel_f$ of frames having a score of at least $0.8 \times \max$ where $\max$ is the global maximum score in the slide. $Ret_f$ refers to the set of frames retrieved by EX-grad for having an extrapolated score of at least $0.8 \times \max$. In this context:

$$prec = \frac{|Ret_f \cap Rel_f|}{|Ret_f|}, \quad rec = \frac{|Ret_f \cap Rel_f|}{|Rel_f|}$$  \hspace{1cm} (1)
Results are compared to random uniform sampling of the same amount of frames followed by similar extrapolation methods. Figures for random sampling are average values over 100 trials. Comprehensive benchmark results can be found in Table I for precision, recall and F-measure.

As shown on Fig. 4, the nearest neighbor method tends to have a better recall rate whereas the spring based method has much higher precision. Both extrapolation methods eventually converge towards the same results. F-measures are roughly similar at any sampling rate. Nevertheless, given that the recall rate remains at acceptable levels, it is advisable to opt for the more sophisticated spring based approximation since perfect precision is more critical for an accurate diagnosis than better recall.

All results show the excellent overall performances of our algorithm which translate into functioning points distributed well above the first bisector in the Relative Operating Characteristics (ROC) diagram in Fig. 5. Our method has always achieved absolute precision, with as little as 2% of the frames analyzed in half of the cases. Recall spans from 32% to 80% with an average value above 50% which allows the retrieval of enough high score frames to grade the slide. The effectiveness of the dynamic sampling algorithm has been proved by the dramatically lower performances at similar sampling levels with random sampling (followed by any extrapolation method) which appears unequivocally on a ROC diagram.

**IV. Conclusion**

We propose a solution to the problem of pre-attentive vision for very large images and in particular for full biopsy slides that is both time efficient and effective. In the medical application, our method has proved its ability to accurately find and measure the highest levels of nuclear pleomorphism in a biopsy slide within an acceptable time frame as well as to provide a useful, reliable visualization map for the end-user. From a more global standpoint, our method makes it possible to speed up the analysis, enhance the visualization and assist the exploration of very large images by an autonomous robot or with a human in the loop.

---

**Table I**

<table>
<thead>
<tr>
<th>case no.</th>
<th>no. of frames</th>
<th>no. of samples</th>
<th>EX-grad</th>
<th>Random sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>case 1</td>
<td>3648</td>
<td>159 (4%)</td>
<td>1.000</td>
<td>0.650</td>
</tr>
<tr>
<td>case 2</td>
<td>5880</td>
<td>102 (2%)</td>
<td>1.000</td>
<td>0.800</td>
</tr>
<tr>
<td>case 3</td>
<td>2544</td>
<td>527 (21%)</td>
<td>1.000</td>
<td>0.286</td>
</tr>
<tr>
<td>case 4</td>
<td>8624</td>
<td>164 (2%)</td>
<td>1.000</td>
<td>0.318</td>
</tr>
</tbody>
</table>

**REFERENCES**


