Journal of Informatics and Mathematical Sciences
Vol. 8, No. 4, pp. 281–286, 2016
ISSN 0975-5748 (online); 0974-875X (print)
Published by RGN Publications

http://www.rgnpublications.com

Special Issue
CRIMSTIC 2016
Current Research in Information Technology, Mathematical Sciences, Science and Technology International Conference 2016
April 13-14, 2016, Melaka, Malaysia

Homology Functionality for Grayscale Image Segmentation

Research Article

Rabih Assaf[1,*], Alban Goupil[1], Valeriu Vrabie[1] and Mohammad Kacim[2]

[1] *CReSTIC, University of Reims Champagne-Ardenne, Reims, France*
[2] *Department of Mathematics, Holy Spirit University of KASLIK, Jounieh, Lebanon*
***Corresponding author:** rabih.assaf@etudiant.univ-reims.fr

Abstract. Topological tools provide features about spaces, which are insensitive to continuous deformations. Applied to images, the topological analysis reveals important characteristics: how many connected components are present, which ones have holes and how many, how are they related one to another, how to measure them and find their locations. We show in this paper that the extraction of such features by computing persistent homology is suitable for grayscale image segmentation.

Keywords. Algebraic topology; Persistent homology; Image processing

MSC. 55U99; 62H35; 54H99

Received: March 19, 2016 **Accepted:** August 14, 2016

1. Introduction

In the last years, there have been concerted efforts to use the fundamental theories of topology and developing it to more applied aspects like computational topology and applied algebraic topology. These efforts, introduced by [5] and [10], have made possible the use of powerful

mathematical concepts in many applications such as data analysis, signal processing, machine learning and shape recognition [2]. A new appealing way to perform image segmentation is thus the use of the topological spaces formalism. Indeed, spaces are studied by the mathematical theory of topology [1] and functions between spaces can represent general transformations, which preserve connectivity. The topological spaces formalism thus allows construction of methods and algorithms that are inherently invariant to disruptive transformation such as rotation, translation and other continuous transformations. The objective of this work is to analyze if the association of topological features with classical ones improve the segmentation of gray scale images. This paper introduces the topological spaces formalism and how persistence features can be extracted in Section 2. Section 3 illustrates two applications of this methodology in order to perform image segmentation on an unstained tissue section imaged by a quantitative phase imaging system and on a satellite grayscale image.

2. Topological Background and Methods

The proposed topological features calculation of a gray scale image follows a concept workflow that begins with its "spatialization". Then the linearization of its combinatorial representation permits the use of tools from algebraic topology and, among them, the persistence. Each of these steps are described in extent below.

2.1 Spatialization

The input image is viewed as a continuous function f from the domain $D \subseteq \mathbb{R}^2$ into the real line $\mathbb{R}$, i.e. $f : D \to \mathbb{R}$. This point of view is correct for grayscale images and thus several spaces can be defined regarding f. The sublevel sets are given by all points of the domain whose value does not exceed a level $a : U_a = f^{-1}]-\infty; a]$. The sublevel sets are ordered by their level a under inclusion, hence $U_a \subseteq U_b$ when $a < b$. This permits to define the filtration as the nested sequence of spaces

$$\phi \subset U_a \subset U_b \subset \cdots \subset U_z \subset D\,. \tag{2.1}$$

2.2 Combinatorial Representation

The spaces under study are mathematically well defined but are not suited for algorithmic calculation. Therefore, spaces are decomposed into cells. The set of all cells and the gluing information provided by its boundaries are called the cell complex [7]. This is particularly suited in the case of grayscale image: a pixel is seen as a square, which shares its boundary with its four neighbors and its corner with eight neighbors. The image viewed as a function gives a value for each cell of its domain representation. For example, $f(x)$ is the grayscale value assigned to a pixel x. The value of an edge is the minimum of values of surrounding pixels, while a corner's value is the minimum of values of incoming edges.

According to this procedure, the sublevel sets obtained by selecting the cells whose value is below a constant level a is necessarily a sub-complex. It means that if a p-cell is in the complex,

its boundary is also in the complex. This evaluation of cells implies that the filtration of spaces (2.1) has an equivalent filtration of complex along the level a that preserves the inclusion property.

2.3 Linearization

Even if it is possible to develop the following theory for general coefficients, we limit our work over $\mathbb{Z}_2$ for the sake of simplicity. The complex previously built gives birth to several vector spaces C_p that are finite formal sums of p-cells. We call the elements of C_p a p-chain [6]. That is, $c = \sum a_i \sigma_i$, where the σ_i are the p-cells and the a_i are the coefficients in $\mathbb{Z}_2$. Its boundary operator is $\partial pc = \sum a_i \partial p \sigma_i$, where $\partial_p \sigma_i$ represents the boundary of the p-cell σ_i and it's the sum of the boundaries of its cells. Hence, by taking the boundary function we map a p-chain to a $(p-1)$-chain. We write this homomorphism as $\partial_p : C_p \to C_{p-1}$. A chain complex is a sequence of chain groups connected by boundary homomorphisms such that $\partial_p \partial_{p-1} = 0$ for all dimensions p:

$$\cdots \xrightarrow{\partial_{p+2}} C_{p+1} \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \xrightarrow{\partial_{p-1}} \cdots \tag{2.2}$$

2.4 Homology

Homology is an algebraic and topological tool to detect connectivity of topological spaces. Boundary less p-chains are meaningful and form a subgroup of C_p that we call the p-th cycle group Z_p:

$$Z_p = \{x \in C_p \mid \partial_p x = 0\} = \ker \partial_p \,. \tag{2.3}$$

Among these cycles, we consider the ones that surround chains. They form a subgroup called the p-boundary group B_p.

$$B_p = \{x \in C_p \mid \exists\ y \in C_{p+1},\ x = \partial_{p+1} y\} = \operatorname{im} \partial_{p+1} \,. \tag{2.4}$$

The p-th homology group H_p is defined as the quotient group Z_p/B_p. It's the group of non-bounding cycles. The homology group H_p keeps the count of essentially different cycles that are interesting by distributing all cycles into equivalent classes. An element of H_p gathers together equivalent cycles, which can be deformed continuously one onto the other. In other words, two cycles are equivalent if their difference is a boundary. In addition, the dimension of H_p is called the p-th Betti number, β_p. The Betti numbers in dimensions $0, 1$, and 2 are the number of connected components, tunnels, and voids of the complex, respectively. Because of the linearity, homology groups H_p can be easily computed by standard matrix manipulations given a combinatorial representation of the chain complex.

2.5 Persistent Homology

Persistent homology comes from the ideas of filtration and the functionality of homology described above. Let $K = \{\sigma_1, \ldots, \sigma_i\}$ a cell complex of dimension d. We assume an ordering on

the cells such that for each $i \leq n$, $K_i = \{\sigma_1, \ldots, \sigma_i\}$ is a cell complex. The chain $\phi = K_0 \subset K_i \subset \cdots \subset K_n = K$ is a filtration of K. In our case, such a filtration is defined according to a function $f : K \to \mathbb{R}$ that orders the cells of K by function value. In addition, by tracking the topological evolution of this filtration using homology, we get a sequence of homology groups that are connected by linear maps induced by inclusions:

$$H(K_0) \to H(K_1) \to \cdots \to H(K_i) \to \cdots \to H(K_n). \tag{2.5}$$

Persistent homology tracks the appearance of classes in this sequence. As we go from K_{i-1} to K_i, we gain new homology classes and we lose some [4]. This is clear following this procedure: if $f_i^j : H(K_i) \to H(K_j)$, we say that an element $\alpha \in H(K_i)$ is born in $H(K_i)$. If it does not belong to the image of the map f_{i-1}^i, we say that α dies at $H(K_j)$ if $f_i^j(\alpha) \in \operatorname{im} f_{i-1}^j$ but $f_i^{j-1}(\alpha) \notin \operatorname{im} f_{i-1}^{j-1}$. If there is a class α born in $H(K_i)$ that dies in $H(K_j)$, we record this as a pair (i, j).

2.6 Computation of Persistent Homology

We can compute the homology $H_p(K_i)$ for all sublevel sets K_i of (2.1) in order to depict the evolution of the number of topological features of an image. However, we lose the information concerning the evolution of each particular cycle. Indeed, a cycle may emerge at a given level i and die further at the level j. Recording the "life duration" of each characteristic cycle is more informative than recording the evolution of Betti numbers. The life duration of a cycle is given by the difference between the death time and its birth time along the filtration. The persistence, and its algorithm [3], gives this recording of the evolution of cycles along the level. The evolution of lifetimes of 0-cycles and 1-cycles can be represented using a persistent diagram or a barcode with respect to filtration time, which is not shown here, due to the limited space.

3. Application to Image Segmentation

The topological features calculated using persistent homology have a big importance in image segmentation. More explicitly, we manipulate images by an overlapping square sliding window. And after computation of persistent homology in each window we can get the life duration of 0-cycles and of 1-cycles as well as the persistent entropy, that is $-\sum_{i \in I} p_i \log p_i$, where I represents the intervals of life durations, $p_i = \frac{l_i}{L}$, l_i = death time − birth time and $L = \sum_{i \in I} l_i$ [9]. Besides topological features, we calculate the mean and the standard deviation of the life durations of 0-cycles and 1-cycles in each window, their persistent entropies for dimensions 0 and 1, and the mean and standard deviation of pixel values, which will form a set of 8 features calculated in each patch. Then we perform the standard clustering method k-means that classify the observations of the features calculated into classes [8]. Finally, we obtain a classification of the data of the image, which lead to a segmentation based on topological and statistical features.

This methodology was applied on a quantitative phase image of a prostate gland, the grayscale levels reflect the refractive index map of the unstained histopathology slide (Figure 1). The segmentation of the gland shows four classes corresponding to the main types of

tissue areas. The same methodology was applied on a satellite image (Figure 2). In these applications, topological features show their discriminative power in image segmentation. These characteristics give more refined measures than only using statistical ones in sake of detecting texture features of images.

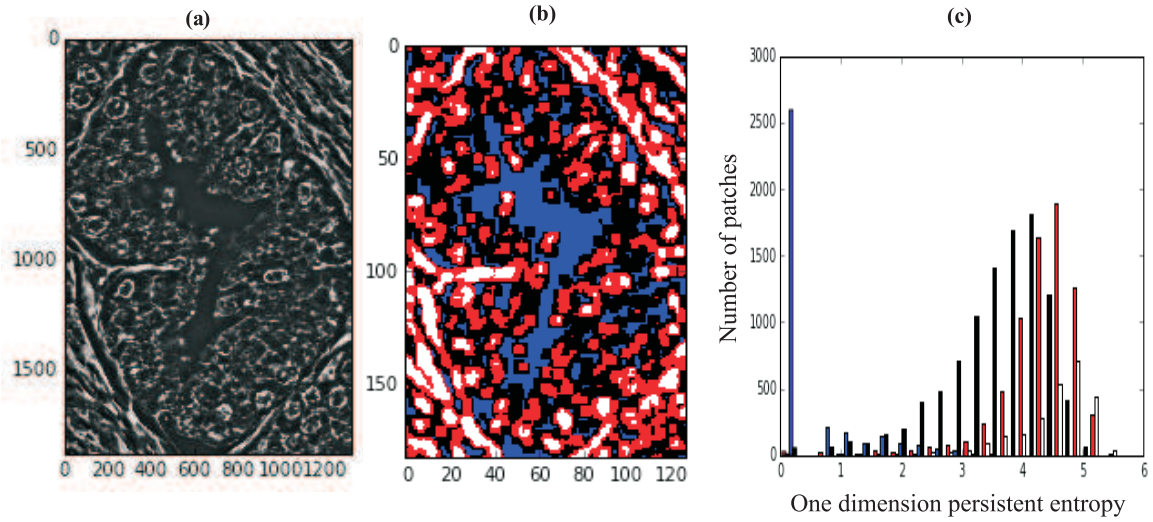

Figure 1. (a) Quantitative phase image of a prostate gland of size 1880 × 1324 pixels, (b) Segmentation of the gland after processing with an overlapping square sliding window. The size of the window was chosen to 50 × 50 pixels and the overlapping to 10 pixels. The eight features were calculated for each window after filtration of non-interesting cycles. (c) Histogram of the one dimension persistent entropy for each class. We see clearly that values of persistent entropy define the distribution of classes, which show the power of topological features calculated in the sake of image segmentation.

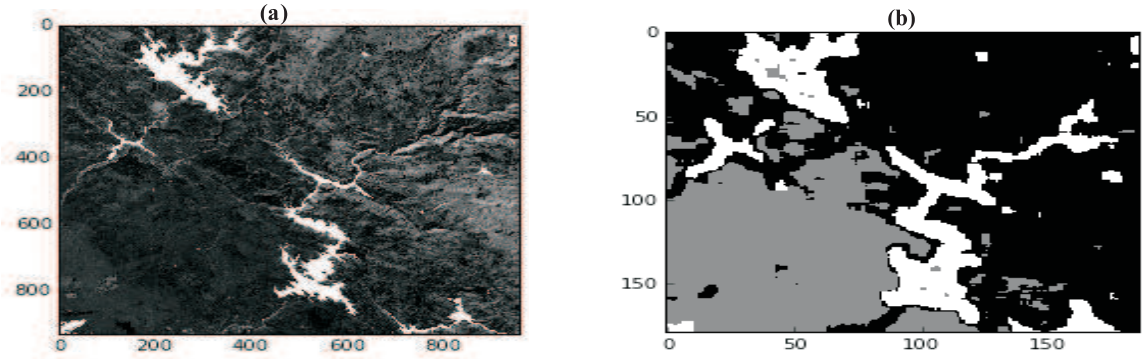

Figure 2. (a) Application of the method described above on a grayscale image of size 929 × 960 pixels taken by a satellite. (b) Segmentation of the image after processing with an overlapping square sliding window. The size of the window was chosen to 30 × 30 pixels and the overlapping to 10 pixels.

4. Conclusion

The topological spaces formalism can be used to represent functions between spaces as transformations preserving connectivity. Algebraic structures allows transforming algebraic topology theorems into computationally feasible algorithms. Indeed, the linearization of

the spatialization permits the use of the persistence, a powerful tool for depicting the evolution of the number of topological features. Translated to grayscale images, this methodology allows extracting features invariant to geometric transformations such as rotation, translation and scaling. Performing image segmentation based on topology and persistent homology characteristics guarantees several nice properties and initial results demonstrate a high potential, as shown the application on two real images, one acquired by a quantitative phase imaging system on an unstained tissue section and one representing a more familiar grayscale satellite image. An improvement in segmentation could be achieved by using other topological invariants and criteria like sheaf theory and multidimensional persistence.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] C. Adams and R. Franzosa, *Introduction to Topology: Pure and Applied*, Prentice Hall (2007).

[2] G. Carlsson et al., Topological data analysis and machine learning theory, *BIRS Workshop*, Alberta, October 15-19 (2012).

[3] C. Chen and M. Kerber, Persistent homology computation with a twist, *27th European Workshop on Computational Geometry*, Switzerland, March 28-30, 2011.

[4] H. Edelsbrunner and J.L. Harer, *Computational Topology*, American Mathematical Society (2010).

[5] H. Edelsbrunner, D. Letscher and A. Zomorodian, Topological persistence and simplification, *Discrete and Computational Geometry* **28** (4) (2002), 511–533.

[6] R. Ghrist, *Elementary Applied Topology*, edition 1.0, Createspace (2014).

[7] A. Hatcher, *Algebraic Topology*, Cambridge University Press (2001).

[8] J. Macqueen, Some methods for classification and analysis of multivariate observations, in *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, 281–297.

[9] E. Merelli et al., Topological characterization of complex systems: using persistent entropy, *Entropy* **17** (2015), 6872–6892.

[10] A. Zomorodian and G. Carlsson, Computing persistent homology, *Discrete and Computational Geometry* **33** (2005), 249–274.