



Original software publication

Percolation Features: An approach for evaluating fractal properties in colour images



Guilherme Freire Roberto ^{a,*}, Leandro Alves Neves ^b, Leonardo Henrique da Costa Longo ^b,
Guilherme Botazzo Rozendo ^b, Thaína Aparecida Azevedo Tosta ^c, Paulo Rogério de Faria ^d,
Alessandro Santana Martins ^e, Marcelo Zanchetta do Nascimento ^f

^a Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), Av. Trabalhador São-carlense, 400, 13566-590, São Carlos-SP, Brazil

^b Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), Rua Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto-SP, Brazil

^c Science and Technology Institute, Federal University of São Paulo (UNIFESP), Avenida Cesare Mansueto Giulio Lattes, 1201, 12247-014, São José dos Campos-SP, Brazil

^d Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia (UFU), Av. Amazonas, S/N, 38405-320, Uberlândia-MG, Brazil

^e Federal Institute of Education, Science and Technology of Triângulo Mineiro (IFTM), Rua Belarmino Vilela Junqueira, 38305-200, Ituiutaba-MG, Brazil

^f Faculty of Computer Science (FACOM) - Federal University of Uberlândia (UFU), Avenida João Naves de Ávila 2121, BLB, 38400-902, Uberlândia-MG, Brazil

ARTICLE INFO

Keywords:

Percolation
Fractal geometry
Feature extraction
Pattern recognition

ABSTRACT

Percolation is a physical phenomenon that describes the transport of fluids through porous media. We present a new approach for extracting local and global percolation features from RGB images. The proposed method is based on multiscale and multidimensional approaches for extracting other fractal measures from histology images. Our method is able to obtain three percolation properties: number of clusters, occurrence of percolation, and size of the largest cluster. This approach aimed to provide an alternative to fractal dimension and lacunarity as ways of evaluating fractal properties in an image.

Code metadata

Current code version	v1
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2022-101
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/2522280/tree/v1
Legal Code License	MIT license
Code versioning system used	git
Software code languages, tools, and services used	MATLAB
Compilation requirements, operating environments & dependencies	MATLAB
If available Link to developer documentation/manual	Not provided
Support email for questions	guilherme.freire@unesp.br

1. Introduction

Percolation is a physical phenomenon that consists on the transport of fluids in porous media. Several natural systems exhibit this type of behaviour, such as water flowing through coffee powder or gases through rocks. If the fluid is able to cross the whole system from one

end to the other, a phenomenon known as percolation occurs. These concepts can also be adapted to image analysis.

In natural systems, the distribution of pores happens randomly. Based on this principle, studies have found that each system has a value p , which corresponds to the probability that a space is or is not a pore [1]. In addition to indicating what the ratio between the existence

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

* Corresponding author.

E-mail addresses: guilherme.freire@unesp.br (G.F. Roberto), leandro.neves@unesp.br (L.A. Neves), leonardo.longo@unesp.br (L.H. da Costa Longo), guilherme.botazzo@unesp.br (G.B. Rozendo), tosta.thaina@unifesp.br (T.A.A. Tosta), paulo.faria@ufu.br (P.R. de Faria), alessandro@iftm.edu.br (A.S. Martins), marcelo.nascimento@ufu.br (M.Z. do Nascimento).

<https://doi.org/10.1016/j.simpa.2022.100387>

Received 19 June 2022; Received in revised form 12 July 2022; Accepted 16 July 2022

of pores and non-pores in a system is, p has an important property known as the percolation threshold. This property denotes that, after a certain value of p , there is the guarantee of percolation occurrence. These concepts can also be adapted for images and avoid efforts in searching for clusters that indicate percolation. Each type of system has a different percolation threshold value, wherein the threshold for square matrices, which also fits for images, is 0.59275 [2].

Thus, percolation-based features have become more exploited to quantify images due to the ability to analyse percolating structures and clusters, complementing observations made with other fractal measures such as Fractal Dimension (FD) and Lacunarity (LAC). Percolation has been applied to quantify vascular [3], cardiac [4], and bone [5] images. However, these applications have been limited for quantifying binary or greyscale images. Therefore, we propose an approach based on percolation theory for extracting features from colour images. This approach uses multiscale and multidimensional observations in order to evaluate the following properties: average number of clusters; occurrence of percolation; and coverage ratio of the largest cluster. The method was applied to different histology image datasets and the obtained features were given as input to classifiers such as Random Forest and Multilayer Perceptron, which provided accuracy values ranging from 87% up to 98% [6].

2. Fractal features for colour images

Fractal features can be obtained from coloured images initially by applying the *gliding-box* algorithm [7]. This algorithm consists in positioning a box of size $L \times L$ in the upper left corner of the image, where L represents the side dimension of the box in pixels. This box moves from left to right to the bottom region of the image, passing through all pixels. After gliding through the entire image, the box is repositioned to the starting point and the value of L is incremented by 2 units. The value of L must always be an odd number, since the existence of a central pixel is required for next steps of the method. In an image of size $H \times W$, the total number T of boxes β_i as a function of L is given by Eq. (1):

$$T(L) = (H - L + 1) \times (W - L + 1) \quad | \quad L \leq \min(H, W). \quad (1)$$

Each time the box β_i is shifted, a colour similarity analysis is performed for each pixel of β_i . This analysis is performed by fixing the central pixel and assigning it to a vector $f_c = \{r_c, g_c, b_c\}$, where r_c, g_c and b_c correspond to the luminous intensity of each of the colour channels, considering RGB images. The other pixels of the box are assigned to another vector $f_i = \{r_i, g_i, b_i\}$ and compared to the central pixel by calculating the Minkowski distance, which allows verifying which pixels belong to the RGB hyperspace formed by the central pixel of the box β_i . The Minkowski distance Δ is given by Eq. (2):

$$\Delta = \max(|f_i(k_i) - f_c(k_c)|), k \in \{r, g, b\}. \quad (2)$$

If the value of Δ corresponding to the distance between f_i and f_c is less than or equal to the scale L , then f_i is labelled 1, otherwise f_i is labelled 0.

After calculating the number of pixels that satisfy Δ for different values of L , different operations can be applied to the labelled boxes to obtain measures such as local FD and LAC.

3. Our approach: Percolation Features

The method is based on Ivanovici and Richard's models for obtaining fractal measures from colour images [7,8]. This approach follows the steps of the *gliding-box* algorithm and calculation of the *Minkowski* distance between a central pixel and the other pixels of a box for colour similarity analysis. Once all pixels in a box are labelled, the procedure to evaluate percolation properties is initiated. We extract two types of features: local and global.

3.1. Local features

In order to evaluate percolation properties, we consider the pixels that satisfied the *Minkowski* distance as pores. Then, the MATLAB function *bwlabel* is applied to each box to label all clusters. For each scale L , three different features are extracted: average number of clusters per box, percolation occurrence rate, and average area of the largest cluster. The average number of clusters $C(L)$ is given by Eq. (3), which consists of the sum of the total number of clusters c in each box β_i divided by the total number of boxes T .

$$C(L) = \frac{\sum_{i=1}^{T(L)} c_i}{T(L)}. \quad (3)$$

In Eq. (5), it is shown how the percolation occurrence rate $Q(L)$ is calculated. This corresponds to the sum of boxes β_i whose number of pixels labelled as pores (Ω_i) satisfy the percolation threshold, as given by Eq. (4).

$$q_i = \begin{cases} 1, & \frac{\Omega_i}{L^2} \geq 0.59275. \\ 0, & \frac{\Omega_i}{L^2} < 0.59275. \end{cases} \quad (4)$$

$$Q(L) = \frac{\sum_{i=1}^{T(L)} q_i}{T(L)}. \quad (5)$$

Eq. (6) corresponds to the calculation of the average coverage area of the largest cluster $M(L)$, given by the sum of the occupancy rates of the largest clusters in each box $\frac{M_i}{L^2}$ divided by the total number of boxes T .

$$M(L) = \frac{\sum_{i=1}^{T(L)} \frac{M_i}{L^2}}{T(L)}. \quad (6)$$

3.2. Global features

Global percolation values can be obtained by generating a curve from the values $\log L \times \log \Phi$, wherein Φ corresponds to one of the three local percolation features ($C(L)$, $Q(L)$ or $M(L)$). Using the method proposed in [8], global values can be obtained by applying the following metrics to such curve: area under curve ($A(\Phi)$), skewness ($S(\Phi)$), area ratio ($A_r(\Phi)$), maximum point ($Max(\Phi)$) and scale of the maximum point ($\sigma(\Phi)$).

The area under the function curve can be obtained by applying the trapezoidal method, as given by Eq. (7), where a and b are the minimum and maximum values of L respectively, Φ is the percolation function and N is the number of observation scales:

$$A(\Phi) = \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(\Phi_n) + f(\Phi_{n+1})). \quad (7)$$

The second global percolation feature obtained is the skewness, which corresponds to a measure of asymmetry. For a percolation curve wherein N local values were calculated, the skewness is obtained according to Eq. (8), wherein Φ_i is the i_{th} value of the sample and $\bar{\Phi}$ is its average value.

$$S(\Phi) = \frac{\frac{1}{N} \sum_{i=a}^b (\Phi_i - \bar{\Phi})^3}{\sqrt{\left[\frac{1}{N-1} \sum_{i=a}^b (\Phi_i - \bar{\Phi})^2 \right]^3}}. \quad (8)$$

Another feature that can be calculated is the ratio between the right and left halves of the curve as shown in Eq. (9).

$$A_R(\Phi) = \frac{A_{(\frac{b}{2}+1, b)}}{A_{(a, \frac{b}{2})}}. \quad (9)$$

The values of $Max(\Phi)$ and $\sigma(\Phi)$ can also be obtained, which gives a total of five global percolation features.

Table 1
Description of software functions.

Function	Input	Output	Description
"percolation.m"	RGB image, <i>maxL</i> parameter	Local and global percolation features	Extracts percolation features from the input image
"getLocalFeatures.m"	Input image in double format and array of <i>L</i> values	Three arrays containing the local features	Calculates the features by applying the process described in Section 3
"getGlobalFeatures.m"	Three arrays for the <i>C</i> , <i>Q</i> and <i>M</i> functions	Struct containing the 15 global features	Applies the procedures described in Section 3.2

4. Software description

The code is written in MATLAB R2019a and it is available on the Code Ocean platform. The software consists of three functions and one script. The functions are described in Table 1. The script *main.m* extracts local and global percolation features from an RGB image given as input according to the parameter *maxL*. Although not mandatory, the value for the parameter *maxL* should be between 41 and 65 for generating a good sample of local values without increasing the running cost by more than necessary. This script can be used by pathologists or computer vision researchers for evaluating percolation properties on medical images, or using the obtained features on a classification system for diagnosis support.

5. Impact

Computer vision methods should be able to represent an object-level analysis of the features present in histological images, based on the criteria used by pathologists such as shape or size of cell nuclei. To quantify such information present in the tissues, computational algorithms must be able to extract features from histological images that explore their morphology, texture, topology and colouration. Fractal measures such as FD and LAC have often been used in this context for the past decade.

The concept of percolation can be explored to define fractal features of images. The information obtained with percolation has the ability to describe properties related to the presence, the dimension and the quantity of clusters in the images. These properties are related to features such as shape, size and quantity of objects, which are relevant information for experts in histopathology. When percolation features were given as input to the classification of non-Hodgkin lymphomas, breast tumours, colorectal tumours and liver tissue datasets, accuracy values of 94.2%, 86.2%, 97.6% and 99.6% were obtained, respectively [9,10]. These results represent an improvement of up to 17% higher when compared to other fractal measures [11]. We believe this new approach can be used as a complementary measure to FD and LAC in order to improve the performance of CAD systems.

6. Limitations & future work

In this work, we have evaluated the proposed technique by applying it to RGB images. Although the use of our method in greyscale is possible with some minor adaptations, other colour models would require major changes to the Minkowski distance calculation. Moreover, a study can be made regarding the ideal value for the input parameter *maxL*, since a greater value would generate more features but significantly increase computing cost. The application of this technique could also be evaluated on non-medical images.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was financed in part by the : Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; National Council for Scientific and Technological Development (CNPq), Brazil (Grants #311404/2021-9 and #313643/2021-0); and State of Minas Gerais Research Foundation - FAPEMIG, Brazil (Grants #APQ-00578-18 and #APQ-01129-21).

References

- [1] Krzysztof Malarz, Simple cubic random-site percolation thresholds for neighborhoods containing fourth-nearest neighbors, *Phys. Rev. E* 91 (4) (2015) 043301.
- [2] N.R.A. Bird, Edith Perrier, Multiscale percolation properties of a fractal pore network, *Geoderma* 160 (1) (2010) 105–110.
- [3] Loretta Ichim, Radu Dobrescu, Characterization of tumor angiogenesis using fractal measures, in: 2013 19th International Conference on Control Systems and Computer Science, IEEE, 2013, pp. 345–349.
- [4] VO Carvalho, LA Neves, MF De Godoy, RD Moreira, AR Moriel, LO Junior, Aprendizado de máquina simbólico e técnicas fractais para caracterizar rejeição em biópsia miocárdica, in: V Latin American Congress on Biomedical Engineering, CLAIB 2011 May 16-21, 2011, Habana, Cuba, Springer, 2013, pp. 272–275.
- [5] S. Sangeetha, C.M. Sujatha, D. Manamalli, Characterization of trabecular architecture in femur bone radiographs using succolarity, in: 2013 39th Annual Northeast Bioengineering Conference, IEEE, 2013, pp. 225–226.
- [6] David Candelerio, Guilherme Freire Roberto, Marcelo Zanchetta Do Nascimento, Guilherme Botazzo Rozendo, Leandro Alves Neves, Selection of CNN, haralick and fractal features based on evolutionary algorithms for classification of histological images, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2020, pp. 2709–2716.
- [7] M. Ivanovici, N. Richard, H. Decean, Fractal dimension and lacunarity of psoriatic lesions-a colour approach, *Medicine* 6 (4) (2009) 7.
- [8] Alexandru Căliman, Mihai Ivanovici, Psoriasis image analysis using color lacunarity, in: Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on, IEEE, 2012, pp. 1401–1406.
- [9] Guilherme F Roberto, Marcelo Z Nascimento, Alessandro S Martins, Thaína AA Tosta, Paulo R Faria, Leandro A Neves, Classification of breast and colorectal tumors based on percolation of color normalized images, *Comput. Graph.* 84 (2019) 134–143.
- [10] Guilherme Freire Roberto, Alessandra Lumini, Leandro Alves Neves, Marcelo Zanchetta do Nascimento, Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images, *Expert Syst. Appl.* 166 (2021) 114103.
- [11] Guilherme F Roberto, Leandro A Neves, Marcelo Z Nascimento, Thaína AA Tosta, Leonardo C Longo, Alessandro S Martins, Paulo R Faria, Features based on the percolation theory for quantification of non-hodgkin lymphomas, *Comput. Biol. Med.* 91 (2017) 135–147.