# Application of Optimum-Path Forest Classifier for Synthetic Material Porosity Segmentation

Victor H. C. Albuquerque
Technological Research Center
University of Fortaleza
Fortaleza - Brazil
victor.albuquerque@fe.up.pt

João P. Papa
Department of Computing
UNESP – Univ. Estadual Paulista
Bauru – Brazil
papa@fc.unesp.br

Alexandre X. Falcão
Institute of Computing
University of Campinas
Campinas – Brazil
afalcao@ic.unicamp.br

Pedro P. R. Filho
Dep. of Teleinformatic Eng.
Federal University of Ceará
Fortaleza – Brazil
pedrosa@deti.ufc.br

João Manuel R. S. Tavares
Faculty of Engineering
University of Porto
Porto – Portugal
tavares@fe.up.pt

*Abstract*— This paper presents a new application and evaluation of the Optimum-Path Forest (OPF) classifier to accomplish synthetic material porosity segmentation and quantification obtained from optical microscopic images. Sample images of a synthetic material were analyzed and the quality of the results was confirmed by human visual analysis. Additionally, the OPF results were compared against two different Support Vector Machines approaches, confirming the OPF superior fast and reliable qualities for this analysis purpose. Thus, the Optimum-Path Forest classier demonstrated to be a valid and adequate tool for microstructure characterization through porosity segmentation and quantification using microscopic images, manly due its fast, efficient and reliable manner.

***Keywords- Optimum-Path Forest, Synthetic Material Porosity Segmentation, Image Foresting Transform.***

## I. INTRODUCTION

The analysis of porosity on synthetic materials through of its fast and reliable segmentation is an important and crucial task for the evaluation of their mechanical properties, e.g., (i) following the propagation and coalescence of microcracks we can better control the mechanical behavior of brittle regimes, (ii) porous zones are pathways for fluid flow and solute diffusion, (iii) orientation of microfissures controls the anisotropy of mechanical properties of materials that can indicate stress regimes. For example, in hydrothermal systems, porosity segmentation and quantification of altered and unaltered rocks can be used as a tool for prevision of fluid/rock ratio and for the recognition and quantification of pathways to control hydrothermal alterations [1], [2].

Automatic analysis of material porosity from images has been recently addressed. Malcolm et al. [3] used an edge linking approach to refine the image segmentation step, and the porosity analysis results are obtained by the pores 2D characterization. Du and Sun [4] proposed an automatic method for pore characterization of pork ham from images considering three steps: (i) ham extraction, (ii) enhancement of the input image and (iii) pore segmentation. A watershed algorithm was used for segmentation of the input images. Taud et al. [5] used a gray level method that considers the input computed tomography image as a surface. Further, the volumes required for the porosity estimation were obtained by integrating this surface with simple operations applied to the input image histogram.

Recently, a novel graph-based classifier that reduces the pattern recognition problem as an optimum-path forest (OPF) computation in the feature space induced by a graph was presented [6]. The OPF classifier does not interpret the classification task as a hyperplanes optimization problem, but as a polynomial combinatorial optimum-path computation from some key samples (prototypes) to the remaining nodes. Each prototype becomes a root from its optimum-path tree and each node is classified according to its strongly connected prototype. This process defines a discrete optimal partition (influence region) of the feature space. The OPF classifier has some advantages with respect to the traditional classifiers: (i) is free of parameters, (ii) do not assume any shape/separability of the feature space, (iii) run training phase faster and (iv) make decisions based on a global criteria. Results in several applications, such that fingerprint and face recognition, remote sensing image classification, biomedical signal processing and many other works, have been demonstrated that OPF is superior than Artificial Neural Networks using Multilayer Perceptrons and Self Organizing Maps, and similar to SVM, but much faster [6].

In such a way, this paper aims to present and evaluate an innovative computational technique based on OPF classifier for analyzing synthetic material porosity images obtained from optical microscopy, which looks for optimizing the process of segmentation and quantification of such microstructures. The main contribution of this work relies on the fact that we are the first into applying the optimum-path forest classifier in this research field. In order to accomplish comparisons about computational cost and segmentation accuracy through OPF classifier, we also applied the SVM classifier with Radial Basis Function kernel (SVM-RBF) and SVM without kernel mapping (SVM-LINEAR) for synthetic material porosity segmentation. Additionally, visual and analytical comparisons were also addressed. The remainder of this paper is organized as follows. Sections II and III present some review about synthetic materials and porosity and the OPF classifier theory, respectively. Section IV discusses the experimental results and, finally, Section V states conclusions and future works.

## II. SYNTHETIC MATERIALS AND POROSITY

Synthetic materials, i.e., those that have been manufactured or otherwise created by human beings, are extensively used in industrial manufacturing and engineering. In the opposite side, we have the natural materials, which remain in widespread use. In the construction of buildings, wood and stone are used throughout the world. For clothing, wool, cotton, silk, and occasionally fur are still used, sometimes in combination with synthetic fibers. For jewellery, natural stones and pearls remain the ornaments of choice. However, for the overwhelming majority of industrial processes, synthetic materials offer enormous advantages of properties and cost, and in many cases they enable products to be made that simply could not be conceived of using natural materials as, for instance, composite materials, cement, optical fiber, metallic alloy, polymers, semiconductors, among other types of materials used in mechanical and industrial application.

Recently, a notable increasing in the number of applications involving synthetic materials can be stressed. The amount of pores in their structure presents some unique properties that are extremely important for many applications [7]. The presence of pores in materials allow these spaces to be filled with other materials, which in turn can help improve the weaker characteristics of the original material [8]. However, to get benefits from this feature, appropriate porosity of the base material is necessary; high porosity could mean significant alterations in properties of the base material and low porosity is often undesirable in many applications. Therefore, accurate and reliable quantification of the porosity of the material is demanded.

Nowadays, as aforementioned, synthetic materials are often used in several fields, such as instrumentation [9], medicine [10] and engineering [11]. In engineering, for example, some compressors use pistons made from a porous materials. The pores contain oil that appears on the surface of the pistons when the pressure increases and can thus reduce the friction during operation [11]. In instrumentation, gas sensors have been developed by integrating a closed electric circuit made of a porous synthetic material. When there exists gas in the pores of the material, the circuit operates normally; otherwise, the circuit closes and the alarm sensor is activated. In such a way, in this case, the level of porosity is used to control the sensitivity of the sensor [9]. In medical applications, such as in synthetic bones, the pores of the materials allow the live cells to be integrated with the artificial parts, therefore reducing the probability of rejection [10].

## III. OPTIMUM-PATH FOREST CLASSIFIER

Let $Z_1$ and $Z_2$ be training and test sets with $|Z_1|$ and $|Z_2|$ samples of a given dataset. Here, we use samples as pixels of images. Let $\lambda(s)$ be the function that assigns the correct label $i$, $i = 1,2,\ldots,c$, to any sample $s \in Z_1 \cup Z_2$, $S \subseteq Z_1$ be a set of prototypes from all classes, and $v$ be an algorithm that extracts $n$ features (Red, Blue and Green values from each pixel) from any sample $s \in Z_1 \cup Z_2$ and returns a vector $\vec{v}(s)$. The distance $d(s,t) > 0$ between two samples, $s$ and $t$, is the one between their corresponding feature vectors $\vec{v}(s)$ and $\vec{v}(t)$. One can use any distance function suitable for the extracted features, been the most common the Euclidean norm [12].

Our problem consists of projecting a classifier that can predict the correct label $\lambda(s)$ of any sample $s \in Z_2$. Training consists of finding a special set $S^* \subseteq Z_1$ of prototypes and a discrete optimal partition of $Z_1$ in the feature space (i.e., an optimum-path forest rooted in $S^*$). The classification of a sample $s \in Z_2$ is done by evaluating the optimum paths incrementally, as though it were part of the forest, and assigning to it the label of the most strongly connected prototype.

### A. Training

Let $(Z_1, A)$ be a complete graph whose nodes are the training samples and any pair of samples defines an arc in $A = Z_1 x Z_1$. The arcs do not need to be stored and so the graph does not need to be explicitly represented. A path is a sequence of distinct samples $\pi_t = \langle s_1, s_2, \ldots, t \rangle$ with terminus at a sample $t$. A path is said trivial if $\pi_t = \langle t \rangle$. We assign to each path $\pi_t$ a cost $f(\pi_t)$ given by a connectivity function $f$. A path $\pi_t$ is said optimum if $f(\pi_t) \leq f(\pi_\tau)$ for any other path $\pi_\tau$. We also denote by $\pi_s \cdot \langle s,t \rangle$ the concatenation of a path $\pi_s$ and an arc $(s,t)$.

We will address the connectivity function $f_{\max}$:

$$f_{\max}(\langle s \rangle) = \begin{cases} 0 \; if \; s \in S \\ +\infty \; otherwise, \end{cases} \qquad (1)$$
$$f_{\max}(\pi_s \cdot \langle s,t \rangle) = \max\{f_{\max}(\pi_s), d(s,t)\},$$

such that $f_{\max}(\pi_s \cdot \langle s,t \rangle)$ computes the maximum distance between adjacent samples along the path $\pi_s \cdot \langle s,t \rangle$. The minimization of $f_{\max}$ assigns to every sample $t \in Z_1$ an optimum path $P^*(t)$ from the set $S \in Z_1$ of prototypes, whose minimum cost $C(t)$ is:

$$C(t) = \min_{\forall \pi_t \in (Z_1, A)} \{f_{\max}(\pi_t)\}. \qquad (2)$$

The minimization of $f_{\max}$ is computed by OPF, which is an extension of the general image foresting transform (IFT) algorithm [7] from the image domain to the feature space, here specialized for $f_{\max}$. This process assigns one optimum path from $S$ to each training sample $t$ in a non-decreasing order of minimum cost, such that the graph is partitioned into an optimum-path forest $P$ (a function with no cycles which assigns to each $t \in Z_1 \setminus S$ its predecessor $P(t)$ in $P^*(t)$ or a marker $nil$ when $t \in S$.

The root $R(t) \in S$ of $P^*(t)$ can be obtained from $P(t)$ by following the predecessors backwards along the path, but its label is propagated during the algorithm by setting $L(t) = \lambda(R(t))$.

We say that S is an optimum set of prototypes when OPF Algorithm minimizes the classification errors in $Z_1$. $S^*$ can be found by exploiting the theoretical relation between minimum-spanning tree (MST) and optimum-path tree for $f_{\max}$ [6]. By computing a MST in the complete graph $(Z_1, A)$, we obtain a connected acyclic graph whose nodes are all samples of $Z_1$ and the arcs are undirected and weighted by the distances d between adjacent samples. The MST is optimum in the sense that the sum of its arc weights is minimum as compared to any other spanning tree in the complete graph, and every pair of samples is connected by a single path which is optimum according to $f_{\max}$. That is, the minimum-spanning tree contains one optimum-path tree for any selected root node. The optimum prototypes are the closest elements of the MST with different labels in $Z_1$. By removing the arcs between different classes, their adjacent samples become prototypes in $S^*$ and Algorithm 1 can compute an optimum-path forest in $Z_1$.

### B. Classification

For any sample $t \in Z_2$, one considers all arcs connecting $t$ with samples $s \in Z_1$, as though $t$ were part of the training graph (Fig. 1c). Considering all possible paths from $S^*$ to $t$, one finds the optimum path $P^*(t)$ from $S^*$ and label $t$ with the class $\lambda(R(t))$ of its most strongly connected prototype $R(t) \in S^*$. This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as:

$$C(t) = \min_{\forall s \in Z_1} \{\max\{C(s), d(s,t)\}\}. \quad (3)$$

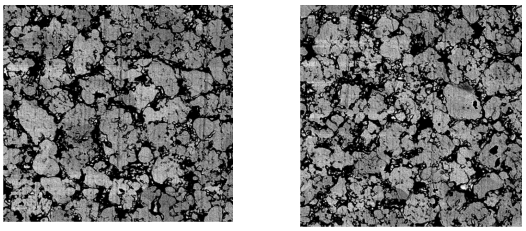
Let the node $s^* \in Z_1$ be the one that satisfies Equation 3 (i.e., the predecessor $P(t)$ in the optimum path $P^*(t)$). Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ as the class of $t$. An error occurs when $L(s^*) \neq \lambda(t)$.

## IV. EXPERIMENTAL RESULTS

In this section, we will describe the dataset used, as well as the evaluation methodology for the synthetic material porosity segmentation.

### A. Dataset

For the application of the computational tool used in this paper, it was necessary to accomplish images acquisition through optical microscopy, accomplishing an ideal adjustment of brightness and contrast. Figure 1 displays the images used in our work, which has been captured with 480x480 pixels.



(a) (b)

*Figure 1.* Synthetic materials porosity images: (a) this image was used in Section IV-C1 for the quantitative evaluation of the robustness of the classifiers and (b) used in Section IV-C2 for visual quality assessment.

### B. Classifiers

In this work, we evaluate the synthetic materials porosity segmentation by three supervised classifiers: Support Vector Machines using Radial Basis Function (SVM-RBF) for kernel mapping, Support Vector Machines without kernel mapping (SVM-LINEAR), and Optimum-Path Forest (OPF). For SVM-RBF, we used the latest version of the LibSVM package [13] with Radial Basis Function (RBF) kernel, parameter optimization and the one-versus-one strategy for the multi-class problem. With respect to SVM-LINEAR, we used the LibLINEAR package [14] with parameter optimized using cross validation. For OPF we used the LibOPF [15], which is a library for the design of optimum-path forest-based classifiers.

### C. Evaluation and Discussion

*1) Robustness of the classifiers:* In this section we used 1% (2303 samples) of the whole image for training classifiers and 99% (228097 samples) to test them (the images were labeled by a technician). Notice that here the SVM-RBF, SVMLINEAR, and OPF algorithms were executed 10 times with randomly generated training and test sets, to compute the mean accuracy and its standard deviation, and the mean training and test execution times in seconds. The accuracy was computed by taking into account that the classes may have different sizes using a methodology proposed by Papa et al. [6]. Table I displays the results for the porosity segmentation accuracy.

We can see that OPF and SVM-RBF achieved similar results and outperformed SVM-LINEAR. However, with respect to the whole execution time, i.e., training and test phases, the OPF was 4.01 times SVM-RBF. Regarding SVM-LINEAR, the former was 1.69 times faster than OPF. If we take into account only the training step execution time, OPF was 125.22 times and 19.98 times faster than SVM-RBF and SVM-LINEAR, respectively

Although the reader argues that training is executed only once, there are several situations in which we need to execute training at a new instance of the problem, such as medical image segmentation and even so synthetic material image segmentation. In this case, the user needs to select some samples for training and further to correct the classification results by marking new samples to improve the classifier's accuracy. This

process is repeated until some convergence criteria to be established.

TABLE I. MEAN ACCURACY AND MEAN TRAINING AND CLASSIFICATION TIMES IN SECONDS FOR OPF, SVM-RBF AND SVM-LINEAR.

| Classifier | Accuracy | Training time | Classification time |
|---|---|---|---|
| OPF | 100% ± 0.0 | 0.55 | 18.15 |
| SVM-RBF | 100% ± 0.0 | 69.02 | 6.00 |
| SVM-LINEAR | 99.53% ± 0.23 | 11.01 | 0.28 |

*2) Synthetic material porosity segmentation analysis:* In this section we show the experimental results for automatic synthetic material porosity segmentation. We used 1% of the image used in the previous experiment (Figure 1a) to train the classifiers. Further, we evaluated them in another image (Figure 1b), in which the results of their segmentation are displayed in Figure 2.

Through experimental analysis, the Optimum-Path Forest classifier and SVM-RBF achieved similar results and outperformed SVM-LINEAR. Additionally, it is important to be aware that specialists in this area of materials science frequently use commercial systems to perform the porosity segmentation. However, these kinds of applications are sometimes inappropriate for this type of analysis, because the porosity evaluation is based on the color histogram built from the image under analysis, thus being less efficient than the solution proposed here.

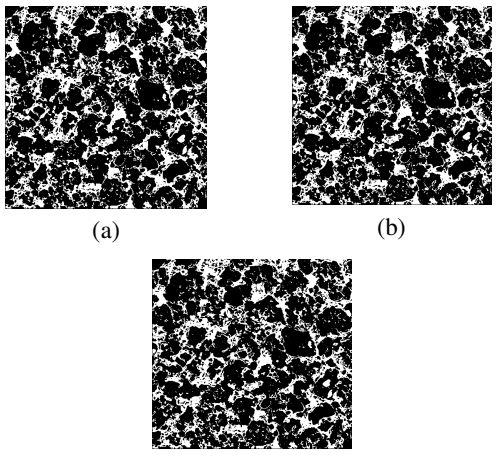



*Figure 2.* Segmentation results: (a) OPF, (b) SVM-RBF and (c) SVM-LINEAR.

## V. CONCLUSIONS

This paper presents an application of the Optimum-Path Forest classifier to accomplish the porosity segmentation and quantification of synthetic materials obtained from microscopic images. Notice that we are the first into applying the OPF in this research field. Regarding the analysis purposes, the results obtained by OPF were compared against those from SVM-RBF and SVM-LINEAR classifiers. OPF and SVM-RBF achieved similar results, outperforming SVM-LINEAR. However, the OPF execution time was 4.01 times faster than SVM-RBF. Based on this comparison, one can clearly conclude that the Optimum-Path Forest classifier was easier, faster and reliable for the proposed application here. Thus, the Optimum-Path Forest classifier is a solution suitable for researchers and specialists, which work with digital image processing.

## ACKNOWLEDGMENT

We would like to thanks CNPq Proc. 481556/2009-5 and FAPESP Proc. 2009/16206-1.

## REFERENCES

[1] E.M. Bongiolo, D.E. Bongiolo, P. Sardini, A.S. Mexias, M. S.-Kauppi, M.E.B. Gomes, and M.L.L. Formoso, "Quantification of porosity evolution from unaltered to propylitic-altered granites: the 14c-pmma method applied on the hydrothermal system of lavras do sul, brazil," *Brazilian Academy of Sciences*, vol. 79, no. 3, pp. 503–517, 2007.

[2] M. Schild, S. Siegesmund, A. Vollbrecht, and M. Mazurek, "Characterization of granite matrix porosity and pore-space geometry by in situ and laboratory methods," Geophysical Journal International, vol. 146, no. 1, pp. 111–125, 2001.

[3] A.A. Malcolm, H.Y. Leongb, A.C. Spowagea, and A.P. Shacklockc, "Image segmentation and analysis for porosity measurement," *Journalof Materials Processing Technology*, vol. 192–193, pp. 391–396, 2007.

[4] C.J. Du and D.W. Sun, "Automatic measurement of pores and porosity in pork ham and their correlations with processing time, water content and texture," *Meat Science*, vol. 72, no. 2, pp. 294–302, 2006.

[5] H. Taud, R.M.-Angeles, J.F. Parrot, and L.H.-Escobedo, "Porosity estimation method by x-ray computed tomography," *Journal of Petroleum Science and Engineering*, vol. 47, no. 3–4, pp. 209–217, 2005.

[6] J. P. Papa, A. X. Falcão, and Celso T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *Intl. J. of Imaging Systems and Tech.*, vol. 19, no. 2, pp. 120–131, 2009.

[7] T.S. Lee, K.K. Park, K.H. Kim, Y.A. Chu, J.P. Jeon, and M. Hwang, "Protective effect of bioactive ceramics on liver injury: regulation of pro-inflammatory cytokins expression," *Journal of Materials Science*, vol. 20, no. 1, pp. 295–299, 2008.

[8] Y.F. Tang, J. Qi, Z.B. Gu, Z.P. Huang, A.D. Li, and Y.F. Chen, "Fabrication of uniform porous alumina materials by radio frequency (rf) magnetron sputtering," *Applied Surface Science*, vol. 254, no. 18, pp. 2229–2232, 2008.

[9] E. Galeazzo, H.E.M. Peres, G. Santos, N. Peixoto, and F.J. Ramirez- Fernandez, "Gas sensitive porous silicon devices: responses to organic vapors," *Sensors and Actuators B: Chemical*, vol. 93, no. 1, pp. 384–390, 2003.

[10] L.M. Mathieu, T.L. Muellerb, P.-E. Bourbana, D.P. Piolettic, R.M., and J.-A.E. Manson, "Architecture and properties of anisotropic polymer composite scaffolds for bone tissue engineering," *Biomaterials*, vol. 17, no. 6, pp. 905–916, 2006.

[11] R. Langer and D.A. Tirrell, "Designing materials for biology and medicine," *Nature*, vol. 428, pp. 487–492, 2004.

[12] A.X. Falcão, J. Stolfi, and R.A. Lotufo, "The image foresting transform: Theory, algorithms, and applications," *IEEE Transactions on PAMI*, vol. 26, no. 1, pp. 19–29, Jan 2004.

[13] C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001, Software available at url http://www.csie.ntu.edu.tw/˜cjlin/libsvm.

[14] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[15] João Paulo Papa, Celso Tetsuo Nagase Suzuki, and Alexandre Xavier Falcão, *LibOPF: A library for the design of optimum-path* forest classifiers, 2008, Software version 1.0 available at http://www.ic.unicamp.br/_afalcao/LibOPF.