Optimal Spectral Histology of Human Normal Colon by Genetic Algorithm

Ihsen Farah, Thi Nguyet Que Nguyen, Audrey Groh, Dominique Guenot, Pierre Jeannesson, Cyril Gobinet

Optimal Spectral Histology of Human Normal Colon by Genetic Algorithm

Ihsen FARAH[1,2], Thi Nguyet Que NGUYEN[1,2], Audrey GROH[3], Dominique GUENOT[3], Pierre JEANNESSON[1,2],Cyril GOBINET[1,2*]

[1] Université de Reims Champagne-Ardenne, Equipe MéDIAN-Biophotonique et Technologies pour la Santé, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims, France.
[2] CNRS UMR 7369, Matrice Extracellulaire et Dynamique Cellulaire (MEDyC), Reims, France.
[3] Université de Strasbourg (UdS), EA 3430 Progression tumorale et microenvironnement. Approches translationnelles et Epidémiologie. Fédération de Médecine Translationnelle de Strasbourg (FMTS), Bâtiment U1113, 3 Avenue Molière, 67200 Strasbourg, France.
[*] Corresponding author (email: cyril.gobinet@univ-reims.fr).

Abstract *- **Fourier transform infrared (FTIR) imaging combined with unsupervised clustering method, such as k-means, achieves a real histology of human tissues. This technique has been successfully applied to diagnose different cancers. However, the clustering methods used in spectral histology are local search algorithms, i.e. these methods converge to a local optimum. Metaheuristics are effective methods to overcome this problem and to reach the optimal solution.***
In this work, we propose a genetic algorithm for the optimal clustering of FTIR images of normal human colon tissues. The obtained results show the efficiency of the proposed genetic algorithm to retrieve more precisely than k-means the structures of normal colon.*

Index Terms *- **Image Processing, Signal Processing, Optical imaging.***

I. INTRODUCTION

Fourier transform infrared (FTIR) imaging is a technique used to measure the absorption of light by a sample. This technique acquires the IR spectra of the bioactive molecules, giving an overall molecular information of the sample, in a non-destructive and non-invasive manner without staining or marking. By FTIR spectroscopy, the structural and metabolic changes in cells and tissues can be detected. This technique has been successfully applied to diagnose different cancers, such as esophageal, stomach and colorectal cancers [1].
FTIR imaging is a recent extention of FTIR spectroscopy for the mapping of samples. This technology provides a data cube, composed of two spatial and one spectral dimensions. Each pixel of a FTIR image is a IR spectrum containing information characterizing the molecular composition of the tissue in that point. Applied to human tissue sections and associated with unsupervised clustering methods such as k-means, IR spectral imaging carries out an authentic spectral histology since different tissue structures can be determined according to their biomolecular composition. This approach has recently been applied for different types of tissues, such as colon, skin and cervix [2].
However, unsupervised clustering methods are local search methods that converge to a local optimum and depend on initialization. In this work, we propose a genetic algorithm to estimate a global optimum for the spectral histology of human normal colon.

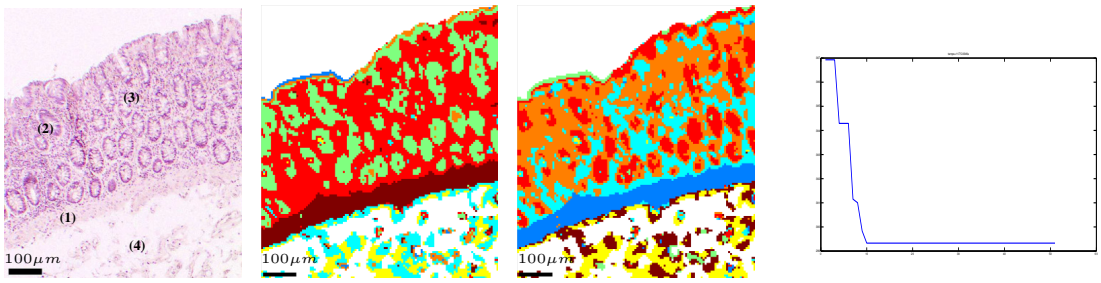
II. MATERIALS AND METHODS

II.1. FTIR imaging and processing

Five formalin-fixed paraffin-embedded tissue blocks of normal zones were prepared from a colon part surgically removed from five patients with a colon cancer. For each tissue block, two consecutive $6\mu m$ thick slices were cut with a microtome. The first slice was mounted onto a calcium fluoride (CaF_2) window. The second slice was mounted on a glass window and stained by Hematoxylin and Eosin (HE) for conventional histological analysis. For data preprocessing, three steps are applied. First, atmospheric correction was performed on each FTIR image by Spectrum Image software to remove water vapor and CO_2 contributions. Second, the spectral range was limited to the 900 - 1800 cm^{-1} fingerprint. Third, the spectra were numerically corrected from paraffin signal and baseline, and normalized using Extended Multiplicative Signal Correction (EMSC) method.

II.2. Genetic algorithm for the clustering of FTIR images

For most of combinatorial optimization problems, which are NP-hard, it is very difficult to find the optimal solution especially for large-scale problems. Indeed, the exact methods require a lot of memory and take a lot of run time. It is thus necessary for practitioners to use methods providing a (sub) optimal solution within a reasonable time, such as metaheuristics.
In this work, we propose a genetic algorithm (GA), whose fitness is the objective function of k-means, to achieve the IR spectral histology of human normal colon. The initial phase of our algorithm is to randomly construct P chromosomes. Each chromosome is represented by a vector



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

Figure 1: (a) HE stained image of tissue where the muscularis mucosae, the crypts, the lamina propria, and the submucosa are respectively annotated by numbers (1), (2), (3) and (4). (b) k-means partition with k = 7 (fitness = 307.3424). (c) GA partition with k = 7 (fitness = 299.3287). (d) Fitness variation of GA between the best solution in first population (fitness = 306.9284) and the last population (fitness = 299.3287).

of integers (genes) which represent the class to which belongs each spectrum. Next, three phases are applied iteratively. During the first phase, each chromosome is represented on a wheel by a portion proportional to the value obtained after applying an exponential scaling to its fitness. $P/2$ homogeneous random draws are made on the wheel leading to the selection of $P/2$ chromosomes (the parents). In the second step, $P/2$ multiple crossovers are made. In fact, two parents, selected randomly, swap a part of their genes between two points selected randomly. At the end of the second stage, $P/2$ children are produced. During the third step, each chromosome has a probability p of mutation. When a chromosome is chosen, two genes (from two different classes) are selected and exchanged. At the end of these phases, a new population, constitued by $P/2$ chromosomes selected in the first step and $P/2$ chromosomes obtained after application of crossover and mutation, is obtained. The three steps are repeated M times.

III. RESULTS

After tests, we set the parameters of our algorithm as follows : P = 100, M = 50, p = 0.1.
Our results show on 5 different sections that our GA provides a better clustering than k-means, compared to the reference image. Figure 1 shows an example. The crypts are more precisely defined in Figure 1 (c) (red cluster) than Figure 1 (b) (green cluster). Also, the solution given by the GA is able to reveal the heterogeneity of the lamina propria by representing two classes in Figure 1 (c) (cyan and orange clusters) instead of one in Figure 1 (b) (red cluster). The comparison of fitness between k-means (307.3424) and GA (299.3287) confirms the better quality of the solution provided by GA.

IV. DISCUSSION-CONCLUSION

Applied on FTIR images of human normal colon, the proposed GA estimates more precise results than a classic clustering method such as k-means. The different structures of the human normal colon and the heterogeneity of the lamina propria are detected by GA. However, the computational time is greater for the GA. One way of overcoming this problem is to hybridize GA with a local search method to accelerate the convergence to a global optimum.

In conclusion, the spectral histology is an innovative technique that proposes an alternative for conventional histology. For optimal classification, a genetic algorithm was proposed. The results show that genetic algorithm is effective to optimally detect all structures of human normal colon tissue and their heterogeneity unlike k-means method.

ACKNOWLEDGMENTS

Authors thank Cancéropôle Grand-Est, Ligue contre le Cancer, the URCA technological platform of cellular and tissular imaging PICT-IBiSA, Région Champagne-Ardenne, Région Alsace and Ministère de l'Enseignement Supérieur et de la Recherche for financial support.

REFERENCES

[1] O. J. Old, L. M. Fullwood, R. Scott, G. R. Lloyd, L. M. Almond, N. A. Shepherd, N. Stone, H. Barrd and C. Kendall. *Vibrational spectroscopy for cancer diagnostics*, *Analytical Methods*, 2014, Vol. 6, pp. 3901-3917.

[2] M. J. Walsh, R. K. Reddy and R. Bhargava. *Label-Free biomedical imaging with Mid-IR spectroscopy*, *IEEE Journal of selected topics in quantum electronics*, 2012, Vol. 18, n° 4, pp. 1502-1513.