Fourier-transform infrared imaging and clustering: toward an automated histology of normal colon

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

Fourier-transform infrared imaging and clustering: toward an automated histology of normal colon

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

[1] Université de Reims Champagne-Ardenne, Equipe Biophotonique et Technologies pour la Santé, UFR de Pharmacie, 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), Strasbourg, France.
[*] Corresponding author (email: cyril.gobinet@univ-reims.fr).

Abstract - Fourier-transform infrared (FTIR) imaging is currently used as a non-destructive and label free method for analyzing biological specimens. Combined with unsupervised clustering method, this biophotonic approach allows to perform a spectral histopathology of human tissues. However, this method requires the subjective choice of the number of clusters. To overcome this problem, we developed a hierarchical double application of 9 cluster validity indices (CVIs) using K-Means clustering. Applying this approach to FTIR images of normal human colon tissue samples, PBM and SI reveal to be the most efficient indices in retreving the main structures of colon histology. These results suggest that the hierarchical double CVI application is thus a promising method for an automated spectral histology.

Index Terms - Optical Imaging, Image Processing, Signal Processing.

I. INTRODUCTION

In biophotonic, Fourier-transform infrared (FTIR) imaging is a non destructive and label-free technique for analyzing samples. This approach spatially records vibrational information which is directly linked to the chemical composition of samples. Recent studies [1] have shown that the combination of this biophotonic approach with unsupervised clustering method, such as K-Means (KM), allows to perform a spectral histology of human tissues. However, this method requires the subjective choice of the number of clusters k. To automatically estimate the optimal k, cluster validity indices (CVIs) have been developed for clustering methods.

In this study, to automate the spectral histology, we developed a hierarchical double CVI application for FTIR imaging. The performance of this approach was then evaluated on 9 different CVIs, using KM partitions estimated on FTIR images acquired on healthy human colon tissue sections.

II. MATERIALS AND METHODS

II.1. FTIR imaging datasets

Five formalin-fixed paraffin-embedded tissue blocks of normal colon zones were obtained from colon cancer surgery of four patients. For each block, two adjacent 6μm thick slices were prepared. The first slice was mounted on a CaF_2 window for FTIR imaging (Perkin Elmer). The second slice was stained with Hematoxylin-Eosin (HE) for conventional histology, and used as a reference for comparison with FTIR imaging.

Each FTIR image was recorded with $6.25\mu m$ spatial resolution, on the mid-IR range of 900 to 1800 cm^{-1} with 4 cm^{-1} spectral resolution. In order to avoid chemical deparaffinization, an in-house Matlab code was applied for neutralizing the spectral interferences from paraffin.

II.2. Hierarchical double CVI application

A CVI is a mathematical function that measures the quality of a partition. By performing a KM clustering for different values of $k, 2 \leq k \leq k_{max}$, this function calculates the ratio between the compactness and separation of clusters for each KM partition. The optimal number of clusters is defined as the number of clusters giving the optimal CVI value.

	Patient sample	1	2	3	4	5
	PBM	**10**	**9**	**8**	**9**	**10**
	SI	**13**	**9**	**9**	**20**	**10**
	Dunn	326	316	180	257	278
	OS	40	392	394	333	374
CVI	SV	228	378	22	58	21
	DB	7	7	**8**	4	4
	COP	5	4	4	7	4
	SWC	4	5	4	4	4
	XB	5	6	8	4	4

Table 1: k_{opt} estimated by hierarchical double CVI application on the FTIR images. Bold values represent the optimal partitions that retrieved at least the main histological structures of normal human colon.
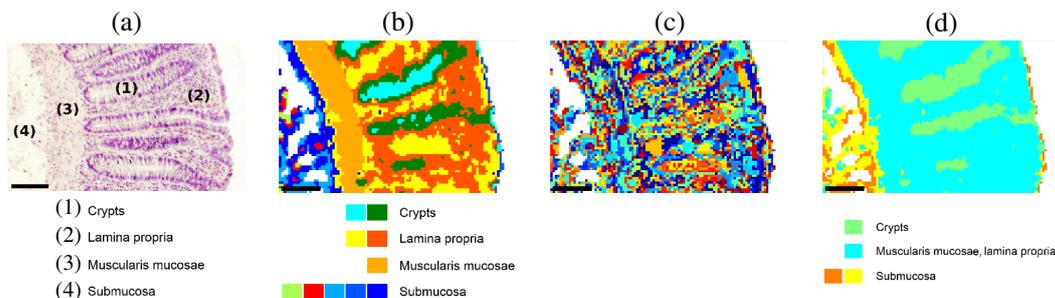
Figure 1: Comparison between conventional histology and pseudo color-coded images reconstructed by double CVI applications for patient sample 1. (a) HE stained image of tissue with 4 histological structures of colon. Optimal partition estimated by (b) PBM ($k_{opt} = 10$), (c) Dunn ($k_{opt} = 316$) and (d) SWC ($k_{opt} = 4$). Scale bars indicate $100\mu m$.

In this study, the nine following CVIs [2,3] were applied: Dunn, Davies-Bouldin (DB), Silhouette-Width-Criterion (SWC), Xie-Beni (XB), Pakhira-Bandyopadhyay-Maulik (PBM), Sym-Index (SI), Context-independent Optimality and Partiality (COP), Separation-Variance (SV) and Overlap-Separation (OS).

To realize an objective and automated spectral histology, a hierarchical double application of CVIs to KM results with $2 \leq k \leq 20$ is proposed. For each spectral image, this method is applied using the following steps: i) The CVI is first applied on the KM results of the spectral image. The optimal number of clusters estimated by the CVI is the number of main clusters k_{main} composing the dataset. ii) Then, on each main cluster, KM is applied for $2 \leq k \leq 20$, and a second application of the CVI on these KM results estimates the optimal number of sub-clusters. iii) The final optimal number of clusters k_{opt} is thus the sum of estimated optimal numbers of sub-clusters. The corresponding optimal double CVI partition is obtained by assembling all the k_{main} estimated optimal sub-partitions together.

III. RESULTS

The k_{opt} estimated by this approach applied on the 5 spectral images are listed in Table 1. Data show that PBM and SI are the most effective indices since their k_{opt} varied between 8 and 20 while exactly matching the main colon tissue structures (Figure 1(a) and (b)) for the 5 samples.

Concerning Dunn, OS and SV, they often exhibited dramatically high k_{opt} ($k_{opt} \geq 40$), thus complicating the assignment of clusters to the corresponding histological structures (Figure 1(c)). By contrast, DB, COP, SWC and XB estimated low k_{opt} inducing optimal partitions which partially correspond to the histological structures. For example, the optimal SWC partition (Figure 1(d)), assigned the cyan cluster to both the muscularis mucosae and the lamina propria.

IV. DISCUSSION-CONCLUSION

In this study, the traditional single application of CVIs has been tested on the FTIR images (data not shown) and mainly leads to an under-estimation of the number of clusters, thus preventing retrieval of the main structures of human normal colon tissue. Theoretically, CVI works correctly for dataset composed of compact and separated clusters, which is not a property fulfilled by FTIR spectral datasets acquired on normal human colon.

For this, the proposed hierarchical double application of CVI succeeds in retrieving the structures of colon tissue, permitting to realize an automatic spectral histology. However, the number of layers of the hierarchical application of CVI is obviously dependent on the considered dataset. An objective criterion needs thus to be defined.

In conclusion, spectral histology is a new concept associating FTIR imaging and clustering. To overcome the subjective choice of the number of clusters, a hierarchical double CVI application for KM partition was proposed. This procedure achieves an automated spectral histology, since the main human normal colon tissue structures are detected for all the analyzed FTIR images.

REFERENCES

[1] J. Trevisan, P.P. Angelov, P.L. Carmichael, et al. *"Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives"*, *The Analyst*, 2012, Vol. 137, pp. 3202-3215.

[2] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, et al. *"An extensive comparative study of cluster validity indices"*, *Pattern Recognition*, 2013, Vol. 46, pp. 243-256.

[3] W. Wang, Y. Zhang, J. Muguerza, et al. *"On fuzzy cluster validity indices"*, *Fuzzy Sets and Systems*, 2007, Vol. 158, pp. 2095-2117.