Effects of digital dewaxing methods on K-means-clusterized IR images collected on formalin-fixed paraffin-embedded samples of skin carcinoma

David Sebiskveradze, Cyril Gobinet, Elodie Ly, Michel Manfait,
Pierre Jeannesson, Michel Herbin, Olivier Piot and Valeriu Vrabie

***Abstract*— Mid-IR spectral imaging is an efficient method to analyze biological samples. Several research studies showed its potential to diagnose cancerous tissues. However, some limitations appear when formalin-fixed paraffin-embedded tissues are studied due to the intense IR contribution of paraffin, unless to perform a time-consuming and aggressive chemical dewaxing. We propose in this paper to analyze the efficiency of two digital dewaxing methods developed to remove the paraffin influence on IR images acquired on a cancerous skin sample. The first method is the Extended Multiplicative Signal Correction (EMSC), which is a preprocessing step applied to neutralize the IR contribution of paraffin. The second one, previously developed for Raman spectroscopy of paraffined tissues, is based on the Independent Component Analysis (ICA) and the Nonnegatively Constrained Least Squares (NCLS). ICA+NCLS permits to remove the IR spectral signature from tissue spectra. Both preprocessing methods are compared on the basis of K-means-clusterized IR images in respect to a conventional histopathological staining. In conclusion, these preliminary results show the efficiency of the preprocessing methods; however ICA+NCLS has to be improved to get more relevant outcomes.**

I. INTRODUCTION

Mid-IR or FTIR (Fourier Transform IR) spectroscopy is an optical technique based on the interaction between an incident light beam and matter. If the energy of an incident photon is near from the energy of a vibrational mode of the sample, this photon is absorbed by the sample and the transmitted light will present a decrease of its intensity at the wavelength of the incident photon. An IR spectrum is a recording of this transmitted light for a range of incident light wavelengths and thus gives information about the vibrational modes of the analyzed sample. The vibrational modes of a sample depending on its molecular composition, the analysis of an IR spectrum thus gives information about the molecular composition and structure of the analyzed sample. This technique has been applied in different fields, especially in biomedical research for the identification of different cancerous tissues [1].

D. Sebiskveradze, C. Gobinet, E. Ly, M. Manfait, P. Jeannesson and O. Piot are with Unité MéDIAN CNRS UMR 6237, Faculté de Pharmacie, Université de Reims Champagne-Ardenne, 51 rue Cognacq-Jay, 51096 Reims Cedex, France {david.sebiskveradze, cyril.gobinet, elodie.ly, michel.manfait, pierre.jeannesson, olivier.piot} @univ-reims.fr

M. Herbin and V. Vrabie are with CReSTIC, Université de Reims Champagne-Ardenne, Chaussée du Port, 51000 Châlons-en-Champagne, France {michel.herbin, valeriu.vrabie} @univ-reims.fr

Since few years, IR spectroscopy has been extended with an imaging system that provides hyperspectral images (e.g. datacube). One image is reconstructed for each recorded wavelength. In cancer research, this technical improvement is used as a tool to localize precisely tumoral nests in the surrounding normal tissue and could thus be developed as a diagnostic tool in clinical oncology [2].

Due to subtle IR spectral differences between normal and tumoral tissue, this task can not be easily realized by a visual analysis. This is the reason why digital multivariate analysis methods such as Principal Component Analysis (PCA), which extracts the biological signal of interest, combined with clustering or classification methods such as Hierarchical Cluster Analysis (HCA), K-means or Fuzzy C-means (FCM) [2] have been applied to IR images in order to automatically discriminate the tumor.

However, when formalin-fixed paraffin-embedded biopsies are studied, the paraffin signal has a huge contribution and can lead to the failure of clustering methods. A recent study [3] has demonstrated that a preprocessing of IR spectral images based on the chemometric tool named Extended Multiplicative Signal Correction (EMSC) [4], [5] neutralizes the paraffin intensity variability. This method constrains the spectral bands of paraffin to the same amplitude, meaning that the paraffin spectral signature is not taken into account in the K-means clustering process. Though the IR bands of paraffin are still present in the recorded spectra, this method can be assimilated to a digital dewaxing.

Recently, a digital dewaxing method based on the combination of Independent Component Analysis (ICA) and Nonnegatively Constrained Least Squares (NCLS), which is here denoted as the ICA+NCLS method, has been developed in order to efficiently remove the paraffin signal from Raman spectra acquired on paraffin-embedded human skin tissues [6]. This methodology has not yet been tested on IR spectra.

In this paper we analyze the effects of the ICA+NCLS preprocessing method on the K-means clustering of IR spectral images acquired on a human skin Basal Cell Carcinoma (BCC) sample and compare the results to those obtained after the application of the EMSC method.

The remaining of this paper is organized as follows. Section II presents the raw dataset and the conventional Hematoxylin-Eosin (H&E) -stained tissue section used for histopathological diagnosis. Section III describes the proposed ICA+NCLS preprocessing method initially developed

for Raman spectroscopy and the reference EMSC preprocessing method. Section IV outlines the K-means clustering algorithm. The clustering results obtained on the raw dataset and after the application of the EMSC and ICA+NCLS preprocessing methods are described and discussed in Section V. We show that both methods give similar results, we explain the reasons of such results and we propose some future extensions of the ICA+NCLS method in order to be adapted to FTIR spectroscopy. Finally, the last Section concludes the paper.

II. RAW DATASET

The study has been performed on a ten micron-thick section of formalin-fixed paraffin-embedded superficial human skin BCC provided by the Pathology Department of Reims University Hospital (Reims, France). An adjacent section has been H&E-stained for histopathological recognition and a visible image of this stained section is shown in Fig. 1. Visual morphology reveals two tumoral areas (outlined) and the normal tissue which is composed of the epidermis (*) and the dermis (+).

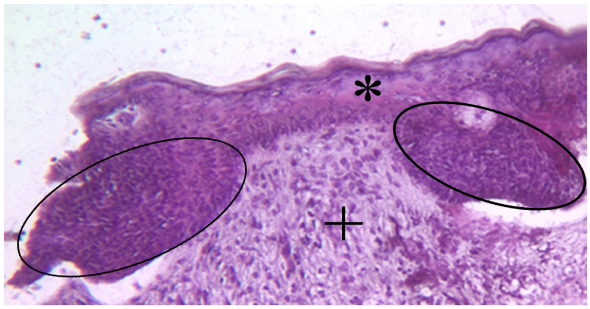


Fig. 1. H&E-stained section of a superficial human skin BCC sample ($*$ epidermis, + dermis, BCC is outlined).

An IR spectral image has been collected with a Spectrum Spotlight 300 FTIR Imaging System coupled to a Spectrum One FTIR spectrometer (Perkin Elmer Life Sciences, France) using the transmission mode. This system provides a spectral resolution of 4 cm^{-1} and a spatial resolution of 6.25 μm. Each spectrum composing the spectral image (e.g. the datacube) has been recorded in the wavelength region 900-1800 cm^{-1} since it is the most informative region. As acquisitions are made in the XY-plane defined by the sample, the datacube is composed of 104 spectra along the x-axis and 61 along the y-axis at 451 different wavenumbers. Then, the datacube is unfolded into a matrix format since the spatial vicinity is not taken into account by preprocessing or clustering methods.

The initial transmittance (T) spectra were converted in absorbance (A) using the following formula: $A = -\log_{10} T$. Characteristic IR spectra of paraffin and paraffin-embedded BCC are shown in Fig. 2. The paraffin IR spectrum is composed of three main bands (labeled 'p') localized at 1378, 1463 and 1471 cm^{-1} with an overlapping of the two last ones. The tissue spectrum is mainly composed of two large bands localized around 1540 and 1650 cm^{-1}.

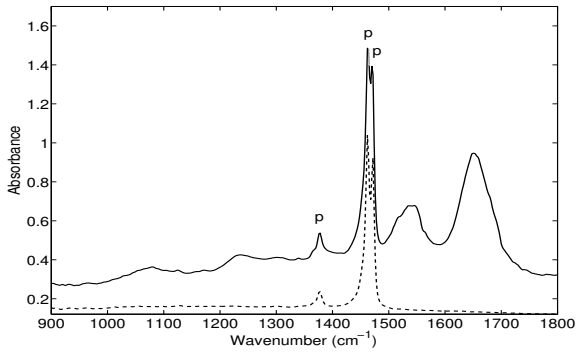


Fig. 2. Example of absorbance IR spectra acquired on the superficial BCC sample. Solid line: from a region where there is paraffin and BCC tissue. Dashed line : from a region where there is only paraffin. The characteristic bands of paraffin are labeled with 'p'.

III. PREPROCESSING

All acquired IR spectra were preprocessed by two different methods in order to digitally remove or neutralize the paraffin signal from each recorded spectrum. The preprocessing is necessary to take into account only the underlying human tissue IR spectrum in the clustering K-means step. The first method, EMSC, does not remove the paraffin spectrum but neutralizes its contribution in each recorded spectrum. The second method [6] developed for Raman spectra extracts the paraffin contribution. This method is based on the successive application of preprocessing steps and source separation techniques such as ICA and NCLS.

A. EMSC

Commonly, the EMSC method [4] is used to separate and to characterize physical and chemical information in spectra from IR microspectroscopy. This method is useful for applications in FTIR spectroscopy [5] where the scatter (physical) variance in spectra changes with the chemical variance in the sample set. In our study we adapt the EMSC model for correcting mid-IR spectra from the contribution of the paraffin signal [3]. The dataset estimation for the spectral image was chosen as the average spectrum within the image. Light-scattering effects were modeled by a fourth-order polynomial function. This choice gives the best result in terms of K-means clustering for our dataset. Due to the paraffin heterogeneity, the paraffin contribution was modeled by PCA to keep the maximum variance in the paraffin dataset while reducing the amount of data. The 10 principal components with the highest eigenvalues were then introduced in the EMSC model [3]. The result given by EMSC followed by a clustering step will be considered here as the reference since this result has already been inspected by pathologists and confirmed by the biomedical community [3].

B. ICA+NCLS

The ICA+NCLS method can be summarized by the following preprocessing steps:

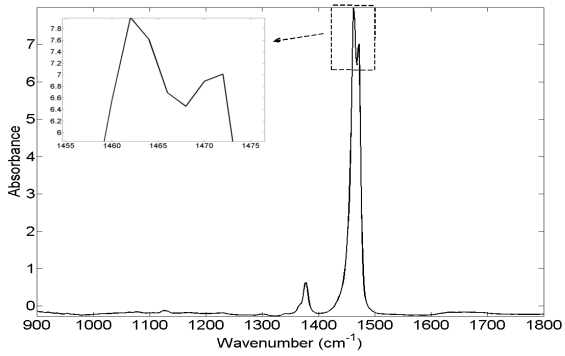
1) Baseline removal: In IR spectra, the background, or baseline, can be due to a scattering of IR beam caused by heterogeneities in the solid, the external light or the source of non-specific absorption. Subtracting the estimation of the background from the raw spectrum leads to a more interpretable signal. This step also allows the application of source separation methods such as ICA since the dataset can be modeled as a linear mixing of spectra of constituents [7]. In our study, the background was estimated and modeled by a polynomial function of order O. The coefficients were estimated by the minimization of a non-quadratic cost function [7], [8]. The choice of the polynomial order depends on the considered application. In our case, the order $O = 7$ provides the best K-means clustering result. Then, the estimated backgrounds were removed from the raw data.

2) Correction of peaks misalignment and width heterogeneity: Due to the limited spectral resolution of the spectrometer, the position of each IR peak of paraffin is affected by a shift that is different from one recorded IR spectrum to another. Note that the shifts are random and usually smaller than the spectral resolution. A method based on the computation of the intercorrelation function between a reference peak and a peak to be aligned has been used to correct these nonlinearities [6], [7].

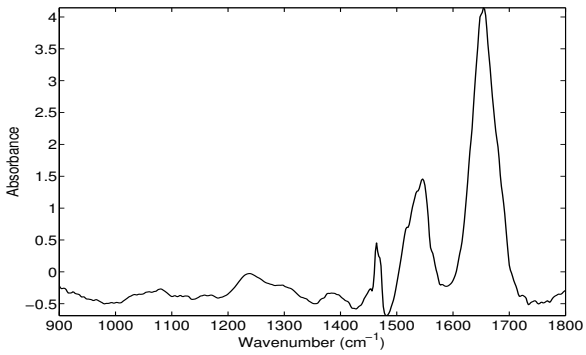
In addition, a variability of the width of each intense IR paraffin peak from one acquisition point to another is observed. A method based on a convolutive transformation of peaks [6], [9] was used to homogenize the peak width. Once the IR spectra peaks were corrected from these misalignments, the ICA+NCLS method was applied.

Note that these corrections are not required by EMSC since this method does not remove the paraffin spectrum but neutralizes its influence on the recorded data. However, these corrections are mandatory for source separation techniques such as ICA or NCLS. These methods are based on a linear mixing of spectra of constituents meaning that each spectrum exists at different concentrations in the recorded dataset. A shift in position or a width deformation of the peaks of paraffin will influence the result of source separation techniques and, consequently, the result of a K-means clustering.

3) ICA: The ICA method was used to estimate the paraffin spectrum from the spectral image [10], [11]. The aim of ICA is to estimate unknown sources (spectra in our case) from observations by supposing that these sources are independent and mixed linearly. A first step, assimilated to the well-known PCA, is used to extract decorrelated spectra from the observations. This is used to reduce the dimensionality of the dataset by keeping only the first decorrelated spectra, which can be viewed as a denoising of the data. In a second step, independent spectra are estimated from the decorrelated ones. Different approaches can be used to carry out this step [11]. We have chosen here the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [10], which consists in jointly diagonalizing cumulant matrices extracted from the fourth-order cumulant tensor of the decorrelated spectra. The ICA method was applied uniquely on pure paraffin spectra extracted from tissue-free regions of the paraffin-embedded BCC section. These spectra were automatically selected by computing the relative energy of the human tissue band around 1650 cm^{-1} in respect to the global energy of recorded spectra (see Fig. 2). A low value of this relative energy at a specific acquisition point means that there is only paraffin on the biopsy at this point. One independent source was finally estimated by ICA. It is important to notice here that three different independent components are necessary to model the paraffin signal in Raman spectroscopy since it is composed of 7 thin and intense peaks. In IR spectroscopy, the paraffin signal is less complex since it is composed of only one small band and two overlapping bands and only one independent source is thus required to model it. The spectrum of paraffin estimated by ICA on a region that does not contain any information related to the human tissue is represented in Fig. 3(a).



(a)



(b)

Fig. 3. The paraffin (a) and human tissue (b) spectra, extracted by ICA and by NCLS respectively.

4) NCLS: Once the signal of paraffin is estimated by ICA, it must be subtracted from each spectrum of paraffined tissue. We have chosen here to use the NCLS method [12], which is a spectral unmixing technique that aims at estimating the concentrations of known spectral signatures in measured linear mixings. A spectrum of the human tissue estimated by NCLS on a zone that contains the tissue embedded in paraffin is represented in Fig. 3(b).

IV. K-MEANS CLUSTERING

K-means clustering is a nonhierarchical and non-supervised clustering algorithm [13]. It is also a "hard"

clustering method because the membership value of each datum to its cluster center is either zero or one. The aim of the K-means algorithm is to minimize a squared-error objective function, based on the Euclidean distance between the datum and its center. In our study, for the first iteration, k spectra (k being the number of clusters selected by the user) were chosen randomly and the algorithm computed the distance between those k spectra (initial cluster centers) and the remaining dataset. Spectra closest to a particular cluster center were grouped into the same cluster. Then the new cluster centers were computed by averaging spectra for each cluster and a new reassignment was processed. The algorithm stopped when no more spectra changed clusters.

K-means clustering has been preferred to other techniques such as hierarchical clustering or fuzzy C-means [2], [3] because it is the computationally fastest, easiest and most popular clustering method. K-means maps were calculated several times after the EMSC and ICA+NCLC preprocessings to make sure that a stable solution was reached independently of the k initial spectra. The number of clusters was set to 11 in agreement with previous studies [2], [3].

V. RESULTS AND DISCUSSION

Data processing was realized with programs written in Matlab 7.2 (Mathworks, Natick, MA) running on a AMD Athlon 64 (2.23 GHz) with 1 GB of RAM. The EMSC and ICA+NCLS preprocessing steps followed by the K-means clustering require around 2 and 10 minutes respectively. These different computational times are due to the different methodological principles of these methods. The ICA+NCLS method is unsupervised as only the recorded IR image and the wavenumber position of paraffin bands are required to construct the paraffin model. On the contrary, the EMSC method is supervised because an IR image of a pure paraffin block must be recorded independently from the considered experiment in order to model the paraffin signal with PCA. Furthermore, the ICA+NCLS method is based on successive non-linear corrections of the recorded spectra, which are computationally expensive. On the contrary, the EMSC method is based on a linear model of recorded spectra. A final difference between the two preprocessing steps is that the EMSC-like method neutralizes the paraffin signal by roughly imposing the same paraffin contribution to every recorded spectra. On the opposite, the ICA+NCLS method removes the paraffin signal from each recorded spectrum in order to have access to the underlying skin signal, which is computational more expensive since the contribution of paraffin must be estimated in each recorded spectrum.

We present here the results of K-means clusterings obtained on the raw dataset and after the two preprocessing steps. The cluster map obtained on the raw dataset is depicted in Fig. 4. As we can see, it is not possible to dissociate the different parts of the human skin: epidermis and dermis. The use of the non-preprocessed spectra leads to an unexploitable cluster map. A preprocessing step is thus mandatory.

On the superficial BCC human skin section, satisfactory results of the K-means clustering after the EMSC preprocessing were obtained (Fig. 5). The efficiency of the EMSC model was already shown in previous studies [3]. Indeed, the spectral clusters correspond to the main histological structures visible on the H&E-stained section (see Fig. 1). We can identify two tumor sites represented by yellow and dark blue colors (Fig. 5), but also the epidermis (dark red) and the dermis (green, orange and cyan colors) layers. The red and light blue clusters correspond to the stratum corneum.

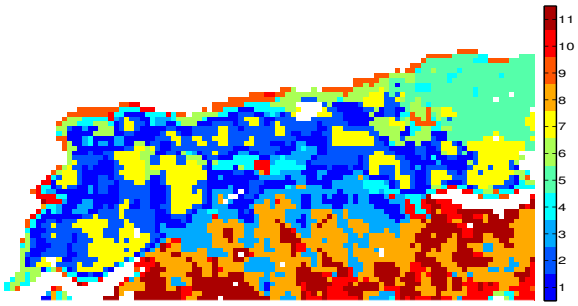


Fig. 4. K-means clustering on the raw dataset of the superficial BCC without any preprocessing.

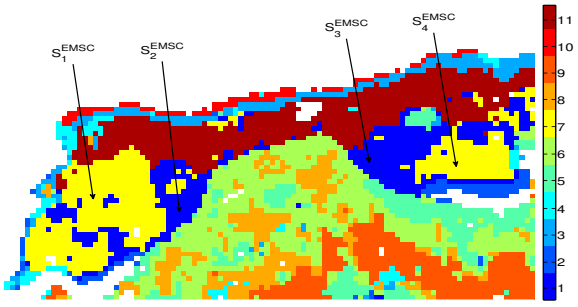


Fig. 5. K-means clustering on the BCC after the EMSC preprocessing. The S^{EMSC} point out the positions of 4 spectra preprocessed by this method that are depicted in Fig. 7(a).

The first step of the ICA+NCLS method is a removal of a background modeled by a polynomial function from the raw dataset. The best estimation of the background has been obtained for a polynomial order of $O = 7$. The corresponding K-means map is shown in Fig. 6.

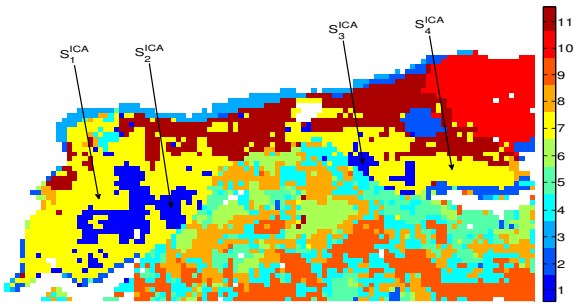


Fig. 6. K-means clustering on the same BCC after the ICA+NCLS preprocessing (baseline removal by a polynomial function of order $O = 7$ and completed by the correction of peaks misalignment and width heterogeneity). The S^{ICA} point out the positions of 4 spectra preprocessed by this method that are depicted in Fig. 7(b).

Compared to the cluster map obtained on the raw dataset (Fig. 4), the ICA+NCLS method improves the K-means clustering results (Fig. 6). Two tumor sites (yellow and dark blue clusters) and the skin layers such as epidermis (dark red color), dermis (cyan, green and orange color) and stratum corneum (light blue color) can be distinguished.

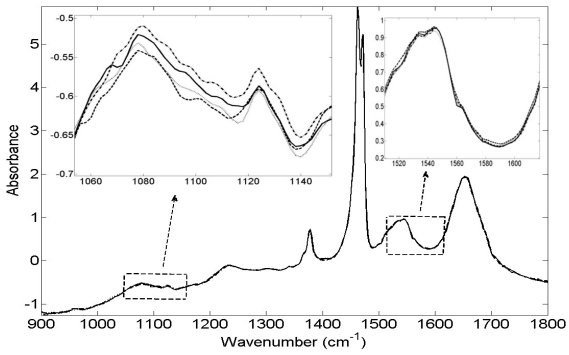
From a general point of view, the two K-means clustering images obtained after the EMSC (Fig. 5) and ICA+NCLS (Fig. 6) preprocessings are equivalent because the same biological structures (i.e. stratum corneum, epidermis, dermis and cancerous tissues) can be extracted. A more detailed analysis reveals however minor differences between these two images. The tumor regions are differently represented by the two clusterings. The left tumor region on the ICA+NCLS K-means clustering image (Fig. 6) is larger than the one preprocessed by EMSC (Fig. 5), and the opposite can be observed for the right tumor region. Furthermore, the left tumor region is composed of two imbricated clusters (yellow and dark blue) and the right tumor region is mainly represented by only one cluster (yellow) on the ICA+NCLS K-means clustering image (see Fig. 6). On the contrary, on the EMSC K-means clustering image (see Fig. 5), the both tumor regions are roughly composed of two juxtaposed clusters (yellow and dark blue). In our opinion, imbricated clusters give a more realistic model of the tumoral growth. However, the EMSC clusters have a quite better spatial homogeneity and better defined boundaries than the ICA+NCLS clusters which give a more pixelized image. The shapes and structures of the tumoral clusters are thus different from one preprocessing method to another.

The same remarks about the spatial homogeneity and boundaries of the EMSC image (Fig. 5) and about the heterogeneity and pixelization of the ICA+NCLS image (Fig. 6) can be made for the epidermis (dark red color) and dermis (cyan, green and orange colors) clusters.

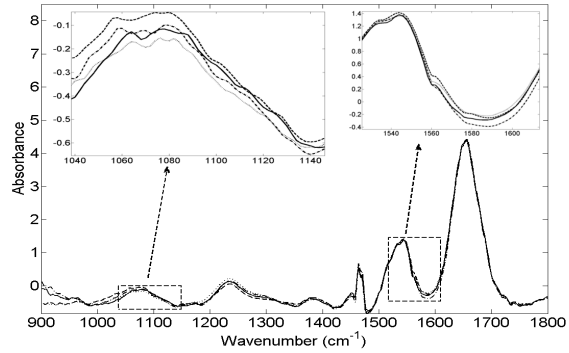
Another difference can be noticed for the stratum corneum. It is represented by two distinct clusters (red and light blue colors) on the EMSC K-means clustering image (Fig. 5), while only one is necessary for the ICA+NCLS image (light blue color, Fig. 6). A decomposition of the stratum corneum into two clusters can be also obtained by increasing the number of clusters. In our opinion, one cluster is sufficient to describe this single part of the human skin, as it is obtained after the ICA+NCLS preprocessing.

The most important difference between images depicted in Fig. 5 and Fig. 6 is visible in their right top corner. For the EMSC image (Fig. 5), the stratum corneum (light blue color), the epidermis (dark red color) and the tumoral tissue (dark blue and yellow colors) boundaries are well separated, while only one red cluster is attributed to this gathering region for the ICA+NCLS image (Fig. 6). This problem can be explained as follows: paraffin is more concentrated in this region and the IR bands of paraffin are much more intense than in the other parts of the biological tissue. In spite of the significant decrease of the paraffin signal, the digital dewaxing by the ICA+NCLS method is not perfect and a residual trace of paraffin still remains around 1465 cm^{-1} as can be seen in Fig. 3(b). This trace is certainly a consequence of the paraffin peak shape since around 1465 cm^{-1} there are two adjacent superimposed peaks (see the zoomed window of Fig.3(a)). Thus, in the red region of Fig. 6, the remaining paraffin intensity is higher than in other parts of the tissue and results in the estimation of one cluster specifically affected to this region. This effect is not observed with the EMSC-based preprocessing since the paraffin bands are fixed to the same intensity and shape for the entire dataset.

We will now compare human skin spectra preprocessed by the EMSC and ICA+NCLS methods. Four preprocessed spectra depicted in Fig. 7, which belong to the different tumoral K-means clusters, have been extracted at the positions pointed out by the arrows in Fig. 5 and Fig. 6.



(a)



(b)

Fig. 7. Spectra preprocessed by (a) the EMSC method (S^{EMSC}) and (b) the ICA+NCLS method (S^{ICA}). S_1 is in solid line, S_2 - in dotted line, S_3 - in dashed line and S_4 - in dash-dotted line.

The tumoral spectra from the EMSC model (S^{EMSC}) are very similar and the paraffin intense peaks have the same amplitude (Fig. 7(a)). On the contrary, the ICA+NCLS method results in spectra that present spectral differences between the tumoral clusters (S^{ICA} in Fig. 7(b)), especially in the 1100 cm^{-1} and 1550 cm^{-1} spectral ranges, where the differences between two tumoral spectra are more significant (see the two zoomed regions in the 1100 cm^{-1} and 1550 cm^{-1} spectral ranges in Fig. 7). Note that the same observation can be done between other structures of the skin (results not presented here).

In comparison to the EMSC step, it may seem contradictory that the ICA+NCLS step increases the spectral differences between human tissues while the k-means clustering results are not improved. It can be explained by some limitations of the ICA+NCLS method when applied to IR spectra, as visible on Fig. 7(b). First, edge effects due to the imperfect baseline estimation and correction by a polynomial function can be observed at lower and higher wavenumbers. Second, a residual paraffin contribution remains in the preprocessed spectra. These imperfections introduce errors which reduce the weight of the spectral discriminant informations between tissues during the K-means clustering step.

Future studies will focus on the correction of these effects in order to improve the K-means results and keep the discriminative spectral features highlighted by the ICA+NCLS method. To correct these limitations and to separate properly the two adjacent superimposed paraffin peaks (see the zoomed 1460-1470 cm^{-1} range in Fig. 3(a)), a derivative preprocessing method should be tested in further experiments. In addition, ICA may be replaced by a Bayesian Positive Source Separation (BPSS) method [14]. Contrary to ICA, BPSS is capable to estimate positive spectra that might have a non-vanishing correlation, which means in IR spectroscopy overlapped IR bands.

Another possible way of analysis is to consider the application of the K-means algorithm to the spectra preprocessed by the ICA+NCLS method in the 1100 cm^{-1} and 1550 cm^{-1} spectral ranges. Since this preprocessing method increases the spectral differences between different spectra in these spectral ranges, the clustering results should be improved.

Pathologists traditionally work on the H&E-stained image. However, even for an expert, it is quite difficult to precisely define the boundaries of the tumoral tissue. This may be due to the existence of a transitional tissue between the tumor and the normal tissue. In order to take into account the notion of transition, it may be interesting in a future work to study the application of fuzzy clustering methods.

VI. CONCLUSION

IR spectral imaging combined with digital data analysis is a powerful tool to diagnose tumoral tissues. When the analyzed samples are formalin-fixed and paraffin-embedded, the paraffin IR signal can interfere with a correct discrimination of cancerous tissues. The influence of paraffin thus must be neutralized or eliminated from recorded IR images. In this paper, we have compared two digital dewaxing methods. EMSC is a chemometric tool that neutralizes the paraffin variability. A numerical dewaxing method based on ICA and NCLS that has been especially developed for the dewaxing of Raman spectra acquired on paraffin-embedded human skin biopsies has been tested in this paper. After each of these two different preprocessing steps, a K-means clustering has been applied in order to discriminate the tumoral parts of the BCC sample. The results show that EMSC is a well adapted method to analyze IR images of paraffin-embedded tissues even if it does not remove the paraffin signal. The method based on ICA and NCLS is also very efficient even if the clusters are less structured and homogeneous due to an imperfect removing of the paraffin signature from the recorded spectra. Moreover, this preprocessing method increases the spectral differences between clusters compared to the EMSC method. The preliminary results are very promising, however improvements of the ICA+NCLS method appear necessary to get more relevant outcomes.

VII. ACKNOWLEDGMENTS

The authors acknowledge the Pathology Department of Reims University Hospital (Reims, France) for the human skin BCC samples and the Institut National du Cancer (INCa) for financial support. DS is a recipient of a doctoral fellowship from INCa.

REFERENCES

[1] R. K. Dukor, *Vibrational Spectroscopy in the Detection of Cancer*, in Handbook of Vibrational Spectroscopy, J. M. Chalmers and P. R. Griffiths, eds., John Wiley and Sons Ltd., New York; 2001, pp 3335-3361.

[2] P. Lasch, W. Haensch, D. Naumann and M. Diem, Imaging of colorectal adenocarcinoma using FT-IR microspectroscpy and cluster analysis, *Biochimica et Biophysica Acta*, vol. 1688, 2004, pp. 176-186.

[3] E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard and M. Manfait, Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies, *The Analyst*, vol. 133, 2008, pp. 197-205.

[4] H. Martens, J. P. Nielsen and S. B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures, *Analytical Chemistry*, vol. 75, 2003, pp. 394-404.

[5] A. Kohler, C. Kirschner, A. Oust and H. Martens, Extended Multiplicative Signal Correction as a Tool for Separation and Characterization of Physical and Chemical Information in Fourier Transform Infrared Microscopy Images of Cryo-sections of Beef Loin, *Applied Spectroscopy*, vol. 59, 2005, pp. 707-716.

[6] C. Gobinet, D. Sebiskveradze, V. Vrabie, A. Tfayli, O. Piot and M. Manfait, "Digital dewaxing of Raman spectral images of paraffin-embedded human skin biopsies based on ICA and NCLS", *in 16^{th} European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, 2008.

[7] C. Gobinet, Application de techniques de séparation de sources à la spectroscopie Raman et à la spectroscopie de fluorescence. Ph.D. dissertation, Université de Reims Champagne-Ardenne, France, 2006.

[8] V. Mazet, C. Carteret, D. Brie, J. Idier and B. Humbert, Background removal from spectra by designing and minimising a non-quadratic cost function, *Chemometrics and Intelligent Laboratory Systems*, vol. 76, 2005, pp. 121-133.

[9] C. Gobinet, V. Vrabie, A. Tfayli, O. Piot, R. Huez, and M. Manfait, "Pre-processing and source separation methods for Raman spectra analysis of biomedical samples", *in Proceedings of the 29^{th} Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2007)*, Lyon, France, 2007, pp. 6207–6210.

[10] J.-F. Cardoso and A. Souloumiac, Blind beamforming for non-Gaussian signals, *IEE Proceedings-F*, vol. 140, 1993, pp. 362-370.

[11] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, Wiley, New York, USA; 2001.

[12] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji and H.-I Chang, A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, 2006, pp. 409-419.

[13] J. B. MacQueen, "Some methods of classification and analysis of multivariate observations", *in Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, USA, 1967, pp. 281-297.

[14] S. Moussaoui, D. Brie, A. Mohammad-Djafari and C. Carteret, Separation of non-negative mixture of non-negative sources using a bayesian approach and MCMC sampling, *IEEE Transactions on Signal Processing*, vol. 54, 2006, pp. 4133-4145.