16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

DIGITAL DEWAXING OF RAMAN SPECTRAL IMAGES OF PARAFFIN-EMBEDDED HUMAN SKIN BIOPSIES BASED ON ICA AND NCLS

Cyril Gobinet[1], *David Sebiskveradze*[1,2], *Valeriu Vrabie*[2], *Ali Tfayli*[1], *Olivier Piot*[1] *and Michel Manfait*[1]

[1]Unité MEDyC, CNRS UMR 6237, UFR de Pharmacie, Université de Reims Champagne-Ardenne
51 rue Cognacq-Jay, 51096, Reims Cedex, France
phone: +33(0)326918125, fax: +33(0)326913550
emails: cyril.gobinet,david.sebiskveradze,ali.tfayli,olivier.piot,michel.manfait@univ-reims.fr

[2]CReSTIC, Université de Reims Champagne-Ardenne
Chaussée du Port, 51000, Châlons-en-Champagne, France
phone: +33(0)326918221, fax: +33(0)326913106, email: valeriu.vrabie@univ-reims.fr

ABSTRACT

A digital automatic dewaxing method of Raman spectral images acquired on paraffin-embedded human skin biopsies is proposed in this paper. This method is based on pre-processing steps specific to Raman spectra, on the estimation of the paraffin spectra by the Principal Component Analysis (PCA) and Independent Component Analysis (ICA), and on the estimation of the paraffin concentrations by the Non-negatively Constrained Least Squares (NCLS) method. The proposed method allows the automatic extraction of skin signatures which depend on the acquisition points. By this way, further classification of a skin biopsy into tumoral, peritumoral and normal zones should be possible.

1. INTRODUCTION

Raman spectroscopy is an optical vibrational technique based on the interaction between an exciting laser light and a sample to be analyzed. Incident photons coming from the laser are absorbed by the sample. A small part of their energy is dissipated by vibrations of the sample and the rest is returned as scattered photons. The scattered photons are informative about the vibrational states of the sample. Thus Raman spectroscopy is an investigative tool of the molecular composition and the structure of the sample.

The analysis of a sample is based on the investigation of a Raman spectrum that is a recording of intensities of scattered light. Recorded spectra are usually expressed according to the wavenumber shift (generally with the unity cm^{-1}) that is the inverse of the wavelength shift between the incident laser light and the scattered light. A Raman spectrum can be seen as a signal composed of peaks or bands for which the positions, intensities, widths and shapes are informative about the molecular composition and structure of the analyzed sample.

Raman spectroscopy has been widely applied especially for biomedical studies such as human skin characterization [1], cancer [2] and atherosclerosis [3] diagnosis, etc. *In vitro* Raman analysis are realized on frozen or dewaxed paraffin-embedded biopsies. Paraffin is commonly used to preserve samples from decay, but has an intense Raman signature that prevents the study of the underlying tissue. A preliminary chemical dewaxing step, which is not totally efficient and can alter the sample [4], is usually applied to remove the paraffin. An alternative digital solution based on Independent Component Analysis (ICA) of Raman spectra has been developed recently [5, 6]. Digital pre-processing steps specific to Raman spectra have also been proposed in order to enhance the results of this digital dewaxing method [7, 8].

This method supposes an homogenous tissue and allows to extract three paraffin sources and a single spectrum of the underlying tissue per biopsy, which can be informative about the tumoral state of the biopsy [6]. This method thus achieves a global numerical dewaxing since only one spectrum of the tissue is estimated per biopsy. However, the Raman spectrum of the underlying tissue can vary from one acquisition point to another and it is primordial to estimate the tissue spectrum in each acquisition point. For example, a human skin biopsy can be composed of tumoral, peritumoral and normal portions of skin. Thus it is a challenge to be able to discriminate between these states by having access uniquely at the spectral information of the tissue in each acquisition point.

The aim of this paper is to propose a new procedure to automatically estimate the biopsy spectrum in each measurement point, i.e. to completely dewax a Raman spectral image by achieving a local numerical dewaxing (e.g. by dewaxing each acquired spectrum). The ICA is still applied in order to estimate the 3 paraffin sources from the spectra acquired on a paraffin-embedded biopsy [5]. Then the paraffin contribution is estimated and subtracted in each acquisition point by a Nonnegatively Constrained Least Squares (NCLS) method, which has been chosen among other methods for its simple implementation and the positivity constraint it imposes on estimated concentrations.

The reminder of this paper is organized as follows: section 2 presents a biomedical example in order to illustrate the application of the proposed method. Section 3 deals with the general scheme developed to automatically and digitally dewax paraffin-embedded samples. The 4^{th} section presents the results obtained on the paraffin embedded skin biopsy. The last section concludes the paper.

2. BIOMEDICAL APPLICATION

The study has been realized on a 4.1×4.9 mm^2 and 10 µm thin section of formalin-fixed paraffin-embedded human skin biopsy deposited on a CaF_2 slide. This type of slide has been chosen because its spectral signature is composed of a unique Raman peak around 325 cm^{-1}. This peak being out of the recorded spectral range does not interfere with the Raman signature of the analyzed sample. A Labram microspectrometer equipped with a near infrared excitation wavelength of 785 nm delivered by a titanium-sapphire laser has been

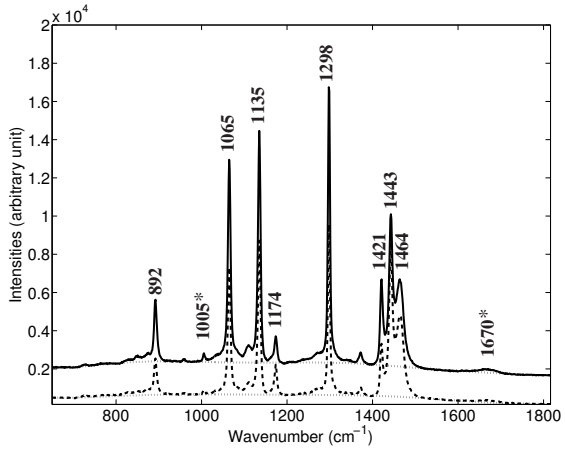


Figure 1: Example of two Raman spectra (solid and dashed lines) recorded on a paraffin-embedded human skin biopsy and their fluorescence background estimations (dotted lines). The Raman features labelled with a star ($*$) are attributed to skin, while the others to paraffin.

used. One spectrum has been acquired every 100 µm in x and y directions leading to the recording of a spectral image composed of $N_x = 41$ spectra according to the x direction and $N_y = 49$ according to the y direction, i.e. a total of $N_x \times N_y = 2009$ Raman spectra. Each spectrum has been recorded in $N_{\tilde{\lambda}} = 990$ different wavenumbers belonging to the spectral range $[650, 1816]$ cm^{-1} with a spectral resolution of 4 cm^{-1}. Thus, a three dimensional dataset has been recorded, which is called a Raman spectral image, and was unfolded to give a matrix dataset $\mathbf{I_t} \in \mathbb{R}^{2009\times 990}$.

Two examples of Raman spectra acquired on the paraffin-embedded human skin biopsy are shown on figure 1 in solid and dashed lines. The paraffin peaks are labelled with their wavenumber written near them. The skin features are marked with their wavenumber followed by a star ($*$). These wavenumbers represent the position of the maxima of the Raman peaks. It can be easily seen in this figure that the skin spectral features are completely hidden by the intense Raman signature of paraffin. The aim is thus to digitally remove the paraffin signal from each recorded spectrum in order to have access to the underlying human skin Raman spectrum.

In the next section we propose a complete dewaxing procedure based on ICA and NCLS. These methods being involved, data must be generated by a linear model. However, due to biological and instrumental distortions, the recorded data are known to be affected by non-linearities. Some specific pre-processing steps, which have been developed to approximate a linear generative model, are also included into the dewaxing method.

3. THE DIGITAL DEWAXING PROCEDURE

The proposed digital dewaxing method is based on the successive application of pre-processing steps and of signal processing tools. Figure 2 presents the flowchart of the digital dewaxing procedure. The application of the pre-processing steps results in a dataset that can be approximately modelled as linear and memoryless [7, 8], leading to the possible use of linear signal processing tools such as Principal Component Analysis (PCA) [9], ICA [9, 10] and NCLS [11].

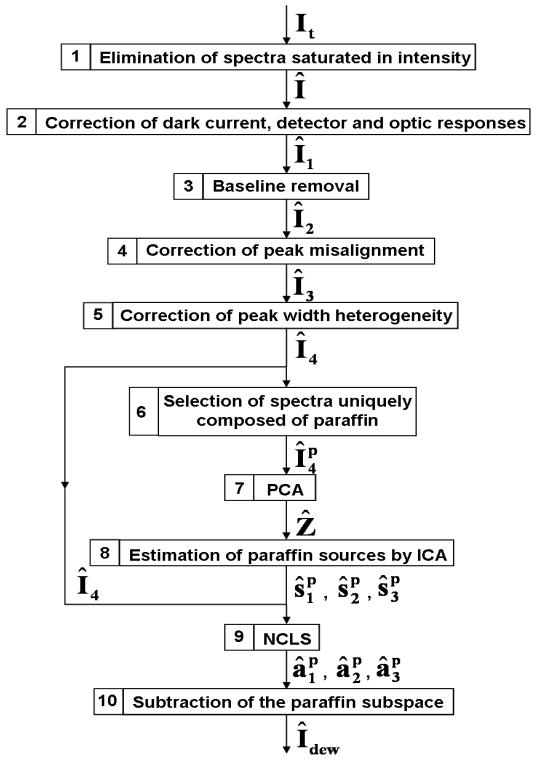


Figure 2: Flowchart of the digital dewaxing procedure.

3.1 Removal of saturated Raman spectra

A Charge-Coupled Device (CCD) detector is characterized by the maximal intensity of scattered light that can be recorded. Thus, if the scattered light is too intense, the recorded spectrum presents saturated intensities, i.e. the spectrum has lost information. A saturated spectrum is automatically detected as having maximal intensity for several successive wavenumbers. The dataset $\hat{\mathbf{I}}$ is obtained by discarding the saturated spectra from $\mathbf{I_t}$, i.e. :

$$\hat{\mathbf{I}} = \mathbf{I_t}(\mathscr{J}), \tag{1}$$

where $\mathscr{J}$ is the subset composed of the Q_1 non-saturated spectra indexes.

The previously described recorded dataset is composed of 95 saturated spectra, i.e. $Q_1 = 1914$. The dataset is thus a matrix $\hat{\mathbf{I}} \in \mathbb{R}^{1914\times 990}$, where each line is one non-saturated recorded spectrum.

3.2 Correction of dark current, detector and optic responses

A noise called dark current corrupts each recorded Raman spectrum and leads to a decrease of the signal to noise ratio. The signal $\hat{\mathbf{I}}_\mathbf{c}$ recorded by the CCD detector without sample, laser or slide is a recording of this noise.

The detector efficacy is dependent on the wavenumber. The recorded Raman spectra are thus distorted by the detector response. An estimation $\hat{\mathbf{I}}_\mathbf{d}$ of this response is obtained by recording the signal detected by the CCD without sample, nor slide, but using a white light.

The optical system of the spectrometer has a response to the laser light which adds to the recorded spectra. An esti-

mation $\hat{\mathbf{I}}_{\mathbf{o}}$ of this signal can be recorded without sample, nor slide during a normal acquisition (i.e. using the laser light).

Once these three recordings are available, their corrections are computed by

$$\hat{\mathbf{I}}_{\mathbf{1}k} = (\hat{\mathbf{I}}_k - \hat{\mathbf{I}}_{\mathbf{o}}) \oslash (\hat{\mathbf{I}}_{\mathbf{d}} - \hat{\mathbf{I}}_{\mathbf{c}}), \tag{2}$$

where $\hat{\mathbf{I}}_k$ is the kth recorded Raman spectrum (with $k \in \{1,\ldots,1914\}$) and $\oslash$ is the elementwise division operator.

3.3 Baseline removal

Due to its biological nature, fluorescence is emitted by the biopsy. This effect is intense compared to the weak Raman spectrum of the skin and has a nonlinear behavior from one recorded spectrum to another.

The fluorescence baseline is traditionally modelled by a polynomial function. Its coefficients are commonly estimated by an iterative least squares scheme [12] or the minimization of a truncated quadratic cost function [13]. These methods require the choice of the polynomial order by the user. This choice is dependent of the considered application [12]. Figure 1 presents two fluorescence backgrounds in dotted lines estimated by the method developed in [13] with a polynomial order equal to 7.

Once the fluorescence backgrounds $\hat{\mathbf{F}}_k$ are estimated, they are removed by computing

$$\hat{\mathbf{I}}_{\mathbf{2}k} = \hat{\mathbf{I}}_{\mathbf{1}k} - \hat{\mathbf{F}}_k, \tag{3}$$

with $k \in \{1,\ldots,1914\}$.

3.4 Correction of peak misalignment

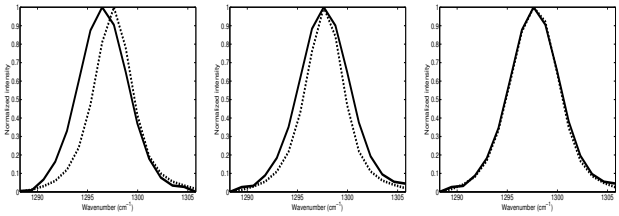
Due to the limited spectral resolution of the spectrometer, the position of each intense Raman peak is affected by a random shift that is different from one recorded Raman spectrum to another. This variability is clearly visible on figure 3(a), where two normalized to unit maximum Raman peaks are shown around 1298 cm^{-1}, after the correction of the dark current, detector and optic responses and the removal of the baseline.

To correct these local nonlinearities, a procedure of peak alignment [7] has been applied to each spectrum $\hat{\mathbf{I}}_{\mathbf{2}k}$ to give the spectrum $\hat{\mathbf{I}}_{\mathbf{3}k}$. This method is based on the computation of the intercorrelation function between a reference peak and a peak to be aligned. The shift between these peaks is estimated as the abscissa that maximizes the intercorrelation function. Then the shift is compensated by the use of the Fourier transform properties. The misaligned peaks of figure 3(a) are shown after the correction of this nonlinearity on figure 3(b). They are now aligned on the same wavenumber.

3.5 Correction of peak width heterogeneity

Due to the irregularity of the sample surface, the laser can be defocused from one acquisition point to another leading to a variability of the width of each intense Raman peak. This local nonlinear effect is clearly visible on figure 3(b) where the aligned and normalized peaks have different widths.

A peak width homogenization procedure has been developed in order to correct this distortion [7, 8]. A reference peak is automatically computed as the mean of the peaks with the largest width. Then, each peak is convolved with a kernel that minimizes the distance with the reference. This



(a) (b) (c)

Figure 3: Raman peaks around 1298 cm^{-1}, normalized to unit maximum: (a) obtained after the correction of dark current, detector and optic responses and the removal of baseline, (b) aligned by the alignment procedure, but having different widths, (c) aligned and having the same width by the peak width homogenization procedure.

technique is applied to each spectrum $\hat{\mathbf{I}}_{\mathbf{3}k}$ to give the processed spectrum $\hat{\mathbf{I}}_{\mathbf{4}k}$. The efficacy of this method is shown on figure 3(c) since the peaks have now approximately the same width.

Note here that the pre-processing steps discussed in 3.4 and 3.5 need the use of a smoothing window to remove the edge effects that might appear.

3.6 Selection of pure paraffin Raman spectra

Paraffin-embedded biopsy sections are always composed of parts of pure paraffin, i.e. where the biopsy is not present. These spectra are selected as those having a near zero energy of the 1670 cm^{-1} Raman band, which represents uniquely a skin feature. A submatrix $\hat{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$ is thus formed from $\hat{\mathbf{I}}_{\mathbf{4}}$ as :

$$\hat{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}} = \hat{\mathbf{I}}_{\mathbf{4}}(\mathscr{K}), \tag{4}$$

where $\mathscr{K}$ represents the subset composed of the Q_2 indexes of the acquired Raman spectra selected as purely composed of paraffin signature. In our study, the dataset $\hat{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$ is composed of $Q_2 = 563$ pure paraffin signatures.

3.7 Principal Component Analysis (PCA)

The PCA [9] is used here in order to realize a noise subspace subtraction combined with a dimension reduction of the corrected dataset $\hat{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$. First, $\hat{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$ is centered and normalized to unit variance to give the matrix $\bar{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$. Then, $\bar{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}}$ is processed by a Singular Value Decomposition (SVD) [6] giving:

$$\bar{\mathbf{I}}_{\mathbf{4}}^{\mathbf{P}} = \sum_{i=1}^{Q_2} \delta_i \mathbf{B}_i \mathbf{R}_i^T = \hat{\mathbf{Z}} + \hat{\mathbf{N}}. \tag{5}$$

where the δ_i are the singular values, the $\mathbf{B}_i$ and $\mathbf{R}_i$ are the left and right singular vectors, $\hat{\mathbf{Z}}$ and $\hat{\mathbf{N}}$ are the signal and noise subspaces.

From previous studies [5, 6], the paraffin signal subspace has been shown as composed of $M = 3$ principal components. Thus the signal and noise subspaces can be rewritten as $\hat{\mathbf{Z}} = \sum_{i=1}^{M} \delta_i \mathbf{B}_i \mathbf{R}_i^T$ and $\hat{\mathbf{N}} = \sum_{i=M+1}^{Q_2} \delta_i \mathbf{B}_i \mathbf{R}_i^T$.

The next step of the digital dewaxing procedure proposed in this article being based on ICA, the denoised dataset $\hat{\mathbf{Z}}$ must be whitened. This is realized by the SVD since the first 3 principal components $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ are decorrelated so whitened.

3.8 Independent Component Analysis (ICA)

The aim of Blind Source Separation (BSS) is to estimate unknown sources from observations of their mixtures modelled in the linear case as:

$$\hat{\mathbf{Z}} = \mathbf{A}\hat{\mathbf{S}}^{\mathbf{P}}, \tag{6}$$

where $\mathbf{A}$ is the mixing matrix and $\hat{\mathbf{S}}^{\mathbf{P}}$ is the source matrix. Independent Component Analysis (ICA) is a possible method to resolve the BSS problem by assuming that the sources are statistically mutually independent [9, 10].

The entire recorded dataset has been shown as composed of 4 independent sources in [6]. In this previous study, ICA was applied on the all spectra acquired on the biopsy. The resulting dataset $\hat{\mathbf{Z}}$ was composed of $M = 4$ principal components in order to estimate 3 sources to model the paraffin spectrum and one for the human skin spectrum. However, as explained in the introduction, this procedure estimates only a global skin spectrum for the all biopsy, while the molecular composition of the skin can locally vary, especially when a tumoral tissue is analyzed. That's why, in this paper, we applied ICA on the subspace $\hat{\mathbf{Z}}$ of equation 5 considering uniquely $M = 3$ sources in order to only estimate the paraffin sources noted $\hat{\mathbf{s}}_{\mathbf{1}}^{\mathbf{P}}$, $\hat{\mathbf{s}}_{\mathbf{2}}^{\mathbf{P}}$ and $\hat{\mathbf{s}}_{\mathbf{3}}^{\mathbf{P}}$.

In this study, the well known Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [14] has been used since it has been shown as fast and efficient for the digital dewaxing of human skin biopsy [6]. However, other ICA methods such as Maximal Diagonality (MD) [10], Fast Independent Component Analysis (FastICA) [9], Maximum Likelihood Positive Source Separation (MLPSS) [15], etc. may be used since they give similar results [7, 8].

3.9 Nonnegatively Constrained Least Squares (NCLS)

The Nonnegatively Constrained Least Squares (NCLS) method [11] is a spectral unmixing method that aims to estimate the concentrations (or abundance fractions) of known spectral signatures into measured linear mixings of these signatures. The originality of this method is to add, to a classical least squares procedure, a positivity constraint on concentrations.

In this study, the paraffin sources matrix $\hat{\mathbf{S}}^{\mathbf{P}} = [\hat{\mathbf{s}}_{\mathbf{1}}^{\mathbf{P}}, \hat{\mathbf{s}}_{\mathbf{2}}^{\mathbf{P}}, \hat{\mathbf{s}}_{\mathbf{3}}^{\mathbf{P}}]^T$ estimated by ICA is used in the NCLS algorithm as a priori spectra of known chemical species that compose the analyzed paraffin-embedded biopsy. The measured linear mixing is the pre-processed dataset $\hat{\mathbf{I}}_{\mathbf{4}}$. The positive concentrations $\hat{\mathbf{A}}^{\mathbf{P}} = [\hat{\mathbf{a}}_{\mathbf{1}}^{\mathbf{P}}, \hat{\mathbf{a}}_{\mathbf{2}}^{\mathbf{P}}, \hat{\mathbf{a}}_{\mathbf{3}}^{\mathbf{P}}]$ of these sources are estimated in such a way that $\hat{\mathbf{I}}_{\mathbf{4}}$ is best fitted by NCLS, i.e. $\hat{\mathbf{A}}^{\mathbf{P}}$ minimizes the error:

$$e = trace\Big((\hat{\mathbf{A}}^{\mathbf{P}}\hat{\mathbf{S}}^{\mathbf{P}} - \hat{\mathbf{I}}_{\mathbf{4}})(\hat{\mathbf{A}}^{\mathbf{P}}\hat{\mathbf{S}}^{\mathbf{P}} - \hat{\mathbf{I}}_{\mathbf{4}})^T\Big), \tag{7}$$

subject to $\hat{\mathbf{A}}^{\mathbf{P}} \geq 0$.

3.10 Subtraction of the paraffin subspace

Then the paraffin subspace $\hat{\mathbf{I}}_{\mathbf{p}}$ can be estimated by

$$\hat{\mathbf{I}}_{\mathbf{p}} = \hat{\mathbf{a}}_{\mathbf{1}}^{\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{1}}^{\mathbf{P}} + \hat{\mathbf{a}}_{\mathbf{2}}^{\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{2}}^{\mathbf{P}} + \hat{\mathbf{a}}_{\mathbf{3}}^{\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{3}}^{\mathbf{P}}. \tag{8}$$

Finally, the dewaxed subspace $\hat{\mathbf{I}}_{\mathbf{dew}}$ is obtained by the subtraction of the paraffin contribution $\hat{\mathbf{I}}_{\mathbf{p}}$ from the pre-processed dataset $\hat{\mathbf{I}}_{\mathbf{4}}$:

$$\hat{\mathbf{I}}_{\mathbf{dew}} = \hat{\mathbf{I}}_{\mathbf{4}} - \hat{\mathbf{I}}_{\mathbf{p}}. \tag{9}$$

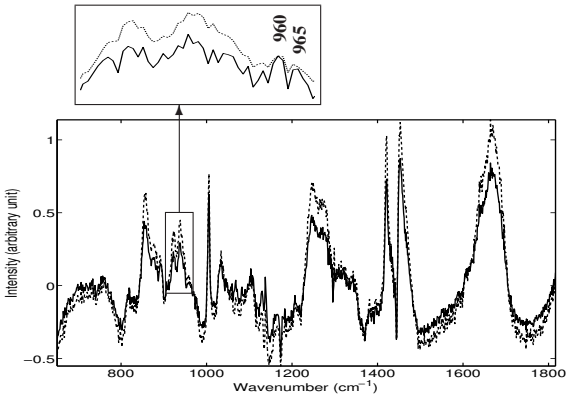
4. RESULTS

We will present first results obtained by the dewaxing method based uniquely on ICA. ICA has been applied on all the recorded spectra of the biopsy after the pre-processing steps described in sections 3.1 to 3.5. In this case a model $M = 4$ sources was chosen, 3 for the paraffin and one for the underlying biopsy, as suggested in [6]. The sources of paraffin ($\hat{\mathbf{s}}_{\mathbf{1}}^{\prime\mathbf{P}}$, $\hat{\mathbf{s}}_{\mathbf{2}}^{\prime\mathbf{P}}$ and $\hat{\mathbf{s}}_{\mathbf{3}}^{\prime\mathbf{P}}$) and the one of the skin ($\hat{\mathbf{s}}^{\prime\mathbf{s}}$) have been estimated at the same time as their respective concentrations $\hat{\mathbf{a}}_{\mathbf{1}}^{\prime\mathbf{P}}$, $\hat{\mathbf{a}}_{\mathbf{2}}^{\prime\mathbf{P}}$, $\hat{\mathbf{a}}_{\mathbf{3}}^{\prime\mathbf{P}}$, and $\hat{\mathbf{a}}^{\prime\mathbf{s}}$. Thus the paraffin subspace $\hat{\mathbf{I}}_{\mathbf{p}}^{\mathbf{ICA}}$ can be estimated by $\hat{\mathbf{I}}_{\mathbf{p}}^{\mathbf{ICA}} = \hat{\mathbf{a}}_{\mathbf{1}}^{\prime\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{1}}^{\prime\mathbf{P}} + \hat{\mathbf{a}}_{\mathbf{2}}^{\prime\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{2}}^{\prime\mathbf{P}} + \hat{\mathbf{a}}_{\mathbf{3}}^{\prime\mathbf{P}}\hat{\mathbf{s}}_{\mathbf{3}}^{\prime\mathbf{P}}$ and the dewaxed data $\hat{\mathbf{I}}_{\mathbf{dew}}^{\mathbf{ICA}}$ by $\hat{\mathbf{I}}_{\mathbf{dew}}^{\mathbf{ICA}} = \hat{\mathbf{I}}_{\mathbf{4}} - \hat{\mathbf{I}}_{\mathbf{p}}^{\mathbf{ICA}}$.

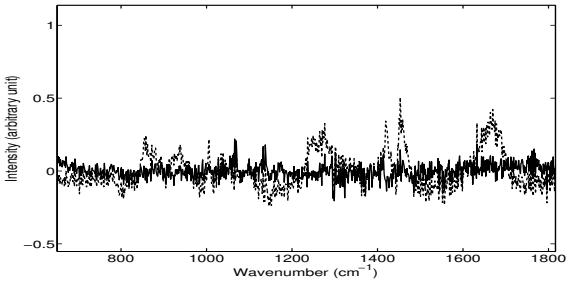
Two dewaxed spectra belonging to $\hat{\mathbf{I}}_{\mathbf{dew}}^{\mathbf{ICA}}$ are plotted in dashed lines on figure 4. The spectrum of figure 4(a) is the dewaxed version of a spectrum which presented spectral skin features (i.e. in a region where the skin biopsy exists), while the one of figure 4(b) is dewaxed from a spectrum which does not have any skin information, i.e. a spectrum recorded on the part of the section where only the paraffin is present. Figure 4(a) clearly shows that the dewaxing by ICA is efficient for spectra which have a contribution of skin, the information linked with the paraffin being completely removed. Figure 4(b) shows that this method is inefficient for spectra which are only composed of paraffin signals because it estimates a non-existent skin spectrum. This observation can be explained as follows. When ICA is applied on the entire dataset $\hat{\mathbf{I}}_{\mathbf{4}}$, 4 independent sources are estimated: 3 sources of paraffin and one source of skin. Modelling all acquired spectra by 4 sources will generate ghost skin features in the parts of the dewaxed data where only a model of 3 sources must be used, as shown on figure 4(b).

The spectra of figure 4 plotted in solid lines are dewaxed by the procedure described by the flowchart of figure 2. It is evident that this method is efficient whatever the considered spectrum. The two methods are quite equivalent when a raw spectrum with a strong contribution of the skin is processed, as shown on figure 4(a). However, the proposed method achieves a more detailed dewaxing as can be seen on the zoom of the spectral range 910-970 cm^{-1} presented on figure 4(a) where the two peaks at 960 and 965 cm^{-1}, which could be assigned to polysaccharides content, are better estimated by the ICA-NCLS-based method. This kind of difference can be very useful in order to discriminate between tumoral, peritumoral and normal portions of skin. When the raw spectrum is only composed of paraffin signals, the dewaxing results in a noise signal as can be seen on figure 4(b).

The proposed method is thus able to dewax Raman spectra and allows the study of the skin signal in each acquisition point. This method should be preferred to study paraffin-embedded biopsies composed of different tissues such as normal, peritumoral and tumoral tissues, which is considered as a difficult task. The use of a classification technique on the results obtained by the proposed method should lead to the discrimination and localization of these different tissues.



(a)



(b)

Figure 4: Two examples of spectra dewaxed by ICA (dashed lines) and by the procedure described by the flowchart of figure 2 (solid lines) in a region where (a) the skin biopsy is embedded into paraffin, (b) only the paraffin is present.

5. CONCLUSION

We have presented in this paper a complete digital dewaxing procedure of Raman spectral images acquired on a paraffin-embedded human skin biopsy. This procedure is composed of a pre-processing part to correct the raw dataset from known non-linearities. Then the corrected spectra uniquely composed of paraffin are detected and processed by PCA and ICA to estimate paraffin sources that are linearly mixed into the dataset. The sources of paraffin are injected into the NCLS method as a priori spectra of known chemical species. NCLS estimates the positive concentrations of these sources into the all corrected dataset. These concentrations, combined with the paraffin sources, estimate a paraffin subspace that can be subtracted from the data. The resulting dewaxed dataset allows to access the spectra of the underlying skin in each acquisition point, which should permit the identification of normal, peritumoral and tumoral zones of a biopsy by classification methods.

REFERENCES

[1] P. Caspers, "In vivo skin characterization by confocal Raman microspectroscopy," Ph.D. dissertation, Erasmus University of Rotterdam, The Netherlands, 2003.

[2] M. Gniadecka, H. C. Wulf, N. N. Mortensen, O. F. Nielsen and D. H. Christensen, "Diagnosis of basal cell carcinoma by Raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 28, pp. 125–129, 1997.

[3] H. P. Buschman, J. T. Motz, G. Deinum, T. J. Rmer, M. Fitzmaurice, J. R. Kramer, A. van der Laarse, A. V. Bruschke and M. S. Feld, "Diagnosis of human coronary atherosclerosis by morphology-based Raman spectroscopy," *Cardiovascular Pathology*, vol. 10, pp. 59–68, 2001.

[4] E. Ó. Faoláin, M. B. Hunter, J. M. Byrne, P. Kelehan, H. A. Lambkin, H. J. Byrne and F. M. Lyng, "Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents," *Journal of Histochemistry & Cytochemistry*, vol. 53, pp. 121–129, 2005.

[5] V. Vrabie, R. Huez, C. Gobinet, O. Piot, A. Tfayli, and M. Manfait, "On the modelling of paraffin through Raman spectroscopy," in *Proceedings of 6th IFAC Symposium on Modelling and Control in Biomedical Systems (MCBMS'06)*, Reims, France, September 20-22. 2006.

[6] V. Vrabie, C. Gobinet, O. Piot, A. Tfayli, P. Bernard, R. Huez and M. Manfait, "Independent Component Analysis of Raman spectra: Application on paraffin-embedded skin biopsies," *Biomedical Signal Processing and Control*, vol. 2, pp. 40–50, 2007.

[7] 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 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2007)*, Lyon, France, August 23-26. 2007, pp. 6207–6210.

[8] C. Gobinet, V. Vrabie, O. Piot and M. Manfait, "Prétraitements et méthodes de séparation de sources pour l'analyse des spectres Raman issus d'échantillons biologiques," *ITBM-RBM, Innovation et Technologie en Biologie et Médecine*, doi: 10.1016/j.rbmret.2007.12.004, 2008.

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

[10] P. Comon, "Independent Component Analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[11] 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, pp. 409–419, 2006.

[12] J. Zhao, H. Lui, D. I. McLean and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy," *Applied Spectroscopy*, vol. 61, pp. 1225–1232, 2007.

[13] 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, pp. 121–133, 2005.

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

[15] S. Moussaoui, D. Brie, and C. Carteret, "Non-negative source separation using the maximum likelihood approach," in *Proceedings of the 13th International Workshop on Statistical Signal Processing*, Bordeaux, France, July 17-20. 2005, pp. 1114–1119.