Journal of
www. biophotonics-journal.org
BIOPHOTONICS

WILEY-VCH

REPRINT

J. Biophotonics **5**, No. 3, 220–229 (2012) / **DOI** 10.1002/jbio.201100098

Journal of BIOPHOTONICS

FULL ARTICLE

Rapid acquisition of Raman spectral maps through minimal sampling: applications in tissue imaging

Christopher J. Rowlands[1], *Sandeep Varma*[2], *William Perkins*[2], *Iain Leach*[3], *Hywel Williams*[4], *and Ioan Notingher*[*,1]

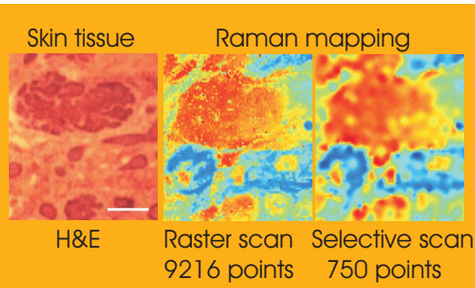
[1] School of Physics and Astronomy, University of Nottingham, University Park, NG7 2RD, Nottingham, U.K.
[2] Dermatology Department, Nottingham University Hospital NHS Trust, QMC Campus, Derby Road, NG7 2UH, Nottingham, U.K.
[3] Histopathology Department, Nottingham University Hospital NHS Trust, QMC Campus, Derby Road, NG7 2UH, Nottingham, U.K.
[4] Centre of Evidence-Based Dermatology, C Floor South Block, Nottingham University Hospital NHS Trust, QMC Campus, Derby Road, NG7 2UH, Nottingham, U.K.

Received 15 October 2011, revised 28 November 2011, accepted 29 November 2011
Published online 20 December 2011

Key words: Raman spectroscopy, image reconstruction, computer assisted diagnosis, microscopy

Supporting information for this article is available free of charge under http://dx.doi.org/10.1002/jbio.201100098

A method is presented for acquiring high-spatial-resolution spectral maps, in particular for Raman micro-spectroscopy (RMS), by selectively sampling the spatial features of interest and interpolating the results. This method achieves up to 30 times reduction in the sampling time compared to raster-scanning, the resulting images have excellent correlation with conventional histopathological staining, and are achieved with sufficient spectral signal-to-noise ratio to identify individual tissue structures. The benefits of this selective sampling method are not limited to tissue imaging however; it is expected that the method may be applied to other techniques which employ point-by-point mapping of large substrates.



A skin tissue section scanned over 10× faster than full raster scanning.

1. Introduction

Raman micro-spectroscopy (RMS) is a well-established technique used to study molecular properties of samples with spatial resolution on the order of micrometers. A key feature of RMS is that chemical components of the sample can be mapped without requiring sample preparation or other contrast-enhancing procedures. However, the conventional approach of raster-scanning the sample through the laser spot (or laser spot across the sample) can make the mapping process very slow, in particular for weakly-scattering samples such as tissue sections [1], for which the signal-to-noise ratio is very low. While photodiode arrays are commonly used in dispersive spectrographs to capture the entire spectrum simultaneously, the required integration times are still often on the order of seconds. With such acquisition time required per pixel means that to scan even a modestly-sized image (e.g. 1×1 mm at 96×96 pix-

* Corresponding author: e-mail: ioan.notingher@nottingham.ac.uk, Phone: +44(0) 1159515172, Fax: +44(0) 1159515180

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

els) can take several hours or even days. One application in which fast spectral mapping of large samples is necessary is the intra-operative imaging of tumour margins. Single-point in-situ measurement based on a handheld Raman probe has already been successfully used for the diagnosis of breast tumours [2]. While numerous studies have also demonstrated the ability of RMS combined with multivariate spectral analysis to provide simultaneous imaging and quantitative diagnosis of tumours in large tissue specimens excised during cancer surgery [3–5], the implementation of intra-operative RMS has not been achieved yet mainly due to the long imaging times.

Techniques such as Stimulated Raman Scattering (SRS) [6] and Coherent Anti-Stokes Raman Scattering (CARS) [7] permit significantly shorter exposure times than conventional Raman scattering. In situations where the intensity at a particular wavenumber can be used as an effective contrast mechanism, then CARS, SRS or Raman imaging using avalanche photodiodes [8] can be effectively employed. However, these techniques require very complex and expensive instruments, can suffer from issues with non-resonant background [9], and have not, to date, demonstrated that they can be used with multivariate statistical techniques, such as linear discriminant analysis, that are necessary for the distinction of subtle changes within the fingerprint region. These issues limit quantitative analysis of the spectrum compared to conventional RMS, which, combined with linear or non-linear multivariate spectral analysis is capable of objective diagnosis of tissue samples with sensitivity and specificity greater than 95% for a large number of tissue types [10]. Similar limitations are found in of wide-field Raman microscopy when baseline variations often obscure the small differences in the Raman bands used for discrimination between different biomolecules [11]. Raman spectral imaging based on line-mapping [12] can considerably decrease the imaging time, up to a factor equal to the number of simultaneously-measured sampling points, provided the laser power used for single-point mapping can be maintained over the whole line. Recent studies have demonstrated that an increase of a factor of 10 can realistically be achieved for imaging tissues, compared to point-by-point scanning [13]. This factor is still too low for intra-operative use however.

There are many situations where properties of the sample can be exploited to speed up imaging. Hadamard Raman microscopy [14] can speed up the imaging process by sampling the surface in a different, more efficient manner. The technique can improve the signal to noise ratio by sampling from all points in an image simultaneously, in a manner analogous to Fourier transform spectroscopy, although this advantage is only present if the limiting factor is the amount of power that can by withstood by the sample. As a result, total sample times are reduced, as longer exposure times are not necessary to achieve a high signal-to-noise ratio.

In the sampling methods proposed in this paper, the next sample point is selected as the location with the maximum absolute difference between a cubic spline interpolant and a Kriging interpolant. Cubic splines are piecewise-defined polynomial functions which are the smoothest possible functions that interpolate a given dataset; further details on their properties, construction and the definition of 'smoothness' can be found elsewhere [15]. Kriging also generates an interpolating surface, but rather than maximizing smoothness, it minimizes the variance in the prediction error [16]. In regions with few spatial features, the two interpolants should converge on the same result. However, in regions with a higher level of variation, the difference between the two interpolants provides an indication of the optimal point to measure next; the algorithm samples regions of uncertainty with high resolution, at the expense of regions containing relatively few spatial features. In certain cases, this algorithm may miss certain small features, by a combination of chance and the local environment being relatively uniform; in these cases large numbers of points will need to be measured to ensure that the surface is accurately reproduced, and in the ultimate limit, all the points that would have been taken in a raster image will have to be measured. While Reuter et al. [17] used Kriging with a laser fluorometer to map oil spills, calibrating the fluorometer using Raman scattering, to the authors' knowledge there are no examples in the scientific literature of selective sampling with minimal a priori information being used with Raman mapping. It should be noted that this technique is not limited to Raman imaging of tissue samples though, and is expected to generalize readily to other point-by-point mapping techniques.

2. Description of the algorithm

The proposed algorithm is outlined in Figure 1. The basis for the algorithm is that if two different techniques are used to fit an interpolating surface to a dataset, the interpolants will agree in regions where there are few spatial features and deviate more in regions with more features. This approach gives a means by which the algorithm can automatically adjust the sampling density based on the presence of observable spatial features. In this paper the two interpolants are Kriging and a thin-plate spline. It is necessary to perform a data reduction step to obtain a single interpolatable value from the spectrum at each point. For this paper, linear models were deemed sufficient; the spectrum at each point is pro-

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

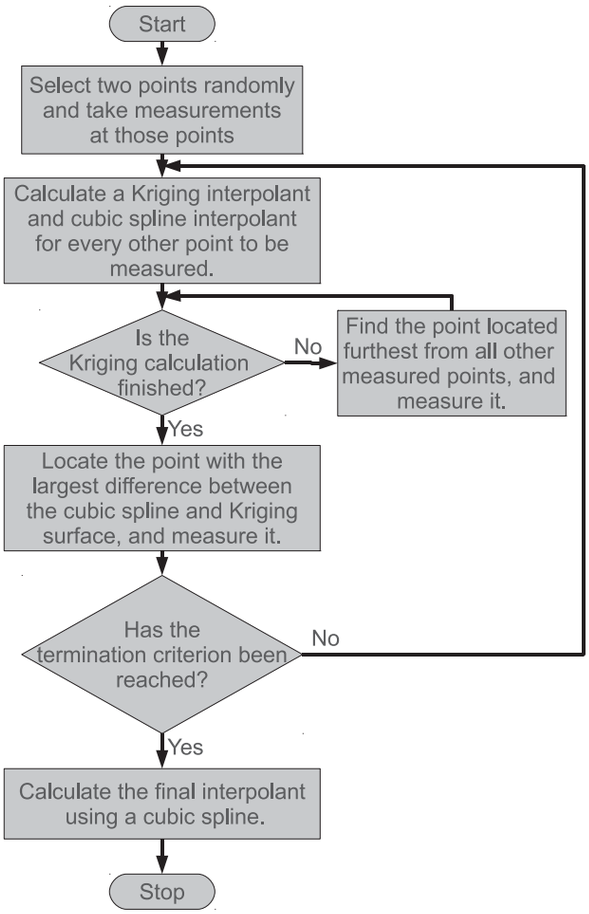


Figure 1 Flowchart illustrating the sampling algorithm.

jected onto a model and the resulting score is interpolated. Depending on the application though, other non-linear models may be equally, if not more applicable. Fitting a thin-plate spline and a Kriging surface can become computationally intensive as the number of measured samples and the number of possible sampling points grows larger. In tests using a 3 GHz Intel Core 2 Duo E8400 processor with 2 GB of DDR2 memory, it took approximately 5 seconds to complete the calculation for 750 measured data points on a 96×96 image. This can lead to considerable wasted time where the imaging system could be taking a measurement but instead must wait for the calculations to complete. The authors propose two solutions to this problem. The first is to exploit data parallelism by evaluating the Kriging surface and thin-plate spline on multiple CPU cores or a graphics processing unit (GPU). Since the evaluation of the Kriging surface is the largest contributor to the total calculation time, and since the interpolant for each remaining point can be calculated independently, the task is expected to scale well. Indeed, almost linear time reduction for this particular part of the calculation can be expected, as the time for data transfer will be negligible. The use of a GPU to accelerate this task is expected to be particularly cost-effective. The other solution to this problem is to sample at the point furthest from all other points that have already been measured (a 'scatter' point) while the interpolant calculations are ongoing, so as to maximise the use of the microscope. This process is sped up by exploiting a kd-tree [18, 19]. Sometimes it is desirable to intentionally sample scatter points to avoid focusing too heavily on local features. The probability of the point being 'scattered' rather than sampled using the interpolants is termed the scatter probability. The authors' experience has shown that a scatter probability of around 0.5 appears suitable for many different samples.

3. Experimental methods

Raman maps were captured using a custom-built microscope setup consisting of a microscope (Eclipse-Ti, Nikon) and automated sample stage (H107 controlled by Proscan II controller, Prior Scientific). The laser (Starbright XM, Torsana) had a wavelength of 785 nm and an output power of up to 500 mW. The microscope objective was either a Nikon $50 \times$ E Plan with a numerical aperture of 0.75 or Leica $50 \times$ N Plan with a numerical aperture of 0.55, depending on the need for a longer working distance. Raman light was dispersed using a spectrograph (77200, Oriel) onto a CCD (DU401-A-BR-DD, Andor Technology). Control software was written in-house using LabVIEW 8.5 and Matlab R2009a. The laser was not interrupted or shuttered during translation of the sample. All calculations were performed in Matlab R2009a, and all parameters were set to their respective defaults unless stated otherwise. Kriging was performed using the DACE toolbox [20] which was modified to permit the definition of a 'nugget' [21]. The Kriging parameters were set using the recommendations of the accompanying user manual; the fitting function was a 0th-order polynomial and the correlation function was Gaussian. The 'theta' parameter for the correlation function was initialized to 100 and bounded between 10 and 104. The cubic spline was calculated using the Thin Plate Spline code of TravisWeins [22], and PCA analysis was performed using the PCA approximation code of Mark Tygert [23], based on work by Rohklin et al. [24]; 10 iterations were used to estimate the principal components.

The algorithm was tested on two types of sample: polystyrene microspheres and human skin tissue sections. The polystyrene microspheres were 1 μm in diameter and cast onto a soda-lime glass microscope slide by placing one drop of the microspheres dis-

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

persed in water onto the slide and allowing the drop to dry. For the mixed paracetamol and polystyrene sample, one paracetamol tablet (Value Health Paracetamol 500 mg Caplets, Galpharm International Ltd.) was ground into powder and a small amount of the powder dispersed onto the drop containing the microspheres before it had an opportunity to dry. The drop was then left to dry as before. The skin tissue sections were obtained from the University Hospitals NHS Trust, Nottingham as part of the routine treatment of basal cell carcinoma by Mohs micrographic surgery. Tissue blocks were excised by the surgeon, mounted on a holder using a tissue mimic (OCT Compound, Tissue-Tek), frozen using cryogenic spray (Frostbite, Surgipath) and sectioned with a microtome (20 μm thickness, CM 1900 UV, Leica). The sections were then mounted on magnesium fluoride disks and kept in the freezer until they were to be used. Adjacent sections were stained using hematoxylin and eosin ('H & E'), and evaluated by a trained histopathologist. Testing of the algorithm on all samples was performed by rastering the laser across the surface of a sample to obtain a full Raman map, and then simulating a sampling procedure. In this procedure, points were selected from the image, the data at each selected point incorporated into the interpolant datasets, the interpolants calculated and the next point selected. This testing methodology necessarily placed a limitation on the size of the spectral maps that could be reasonably captured, as obtaining higher resolution fully-rastered maps required an impractically long time. The integration time for capturing an image of polystyrene microspheres was 0.25 s per point with a laser power at the sample of 310 mW and a spot size of approximately 2 μm diameter. The integration time for tissue sections was 2 s per point with laser power at the sample of 70 mW and a similar spot size.

4. Results and discussion

4.1 Testing the algorithm using polystyrene microspheres

In order to demonstrate the capabilities of the selective sampling method, the algorithm was first tested on randomly-located polystyrene microspheres. The data reduction method was to normalize the dataset by subtracting the mean and dividing by the standard deviation, and then to obtain the second principal component of the normalized dataset. The normalized data was then projected onto this component to yield a score. The second principal component was used rather than the first, since the first corresponded primarily to the glass substrate, whereas the second component served to highlight the regions containing polystyrene. In a practical application, it would be necessary to determine the model prior to measuring the dataset. As this paper is not concerned with the

Table 1 Performance of the selective scanning algorithm when applied to polystyrene microspheres. Data std is the standard deviation of the sample itself, whereas RMSE std is the standard deviation of the root-mean-squared error. The last column represents the number of undersampled points required to equal the performance of theselective sampling algorithm, in % of the number of points used in selective sampling. Note that only a single measurement was taken for the 'B' samples, so mean and standard deviation cannot be calculated.

Identity	Width/μm	Height/μm	X/pixels	Y/pixels	Data std	RMSE mean ± std	Equivalent nr. of points in undersampling (%)
A1	492	492	128	128	0.633	0.41 ± 0.02	183
A2	134	123	128	128	1.08	0.34 ± 0.01	105
A3	70	49	128	128	0.74	0.11 ± 0.01	120
A4	145	137	128	128	1.08	0.31 ± 0.01	71
A5	118	100	128	128	1.24	0.30 ± 0.02	97
A6	127	126	128	128	0.57	0.170 ± 0.007	90
A7	140	94	128	128	2.11	0.28 ± 0.01	137
A8	178	150	128	128	1.14	0.31 ± 0.01	90
A9	170	166	128	128	2.77	0.40 ± 0.01	145
A10	162	122	128	128	4.43	0.64 ± 0.03	145
A11	116	121	128	128	0.782	0.29 ± 0.01	90
B1	973	706	256	256	0.992	0.68	112
B2	263	231	256	256	6.72	0.53	146
B3	390	389	256	256	2.55	0.57	116
B4	806	650	256	256	0.579	0.11	86
B5	155	150	256	256	1.52	0.77	121

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Journal of
BIOPHOTONICS

details of the model, other than that one exists and that it gives good contrast in the resulting image, a principal component of the dataset itself was used.

The root-mean-squared error between a full raster scan of the surface and a final interpolation is used as a figure of merit. The results can be seen in Table 1; samples A1–A11 have a resolution of 128 × 128 and samples B1–B5 have a resolution of 256 × 256. The 128 × 128 samples were each tested 8 times and the mean and standard deviation of the root-mean-squared error taken, so as to characterize both the performance and the repeatability. The 256 × 256 samples were too large to undergo so many repeats within a reasonable time frame, so only a single run was performed for each sample. 2150 measured points were taken for the 256 × 256 samples, corresponding to a speed-up of over 30×. Comparisons were also made with undersampling. Resampled images with increasingly high resolutions were taken, until the root-mean-squared error was below that of the newer algorithm. The number of points for undersampling and the newer algorithm were then compared.

Sample sizes were selected semi-arbitrarily in order to achieve step-sizes on the order of 1 μm, but significantly larger (e.g. A1) and smaller (e.g. A3) step sizes were also used, as well as obtaining samples with significantly differing step sizes in the X- and Y-axes (e.g. A7). It would be expected that larger samples give rise to larger errors, as there will be less spatial correlation between pixels, but this effect seems to be relatively small compared to the differences between samples. Readers are recommended to use sample sizes such that the smallest features of interest approximately correspond to a single pixel, but the data suggests that having a sample up to 0.5 to 4 times this size has little or no effect on the algorithm's performance.

An estimate for the noise inherent in the measurement was obtained by performing a raster scan of a surface, followed by another identical scan of the same surface. The second principal component model was taken for the first scan, and applied to both the first and second scans. The root-mean-squared error between the two resulting images was then taken. For a 128 × 128 image, with a data range of 9.11, the root-mean-squared error due to noise was 0.10, or 1.14% of data range.

Table 1 demonstrates that the root-mean-squared error is consistently well below the standard deviation for all samples. In addition, the standard deviation of the RMSE is low, indicating that the selection of the first two random points does not significantly affect the eventual outcome. Comparison with undersampling is generally favourable or comparable; the nature of the experiment (i.e. continuously at-

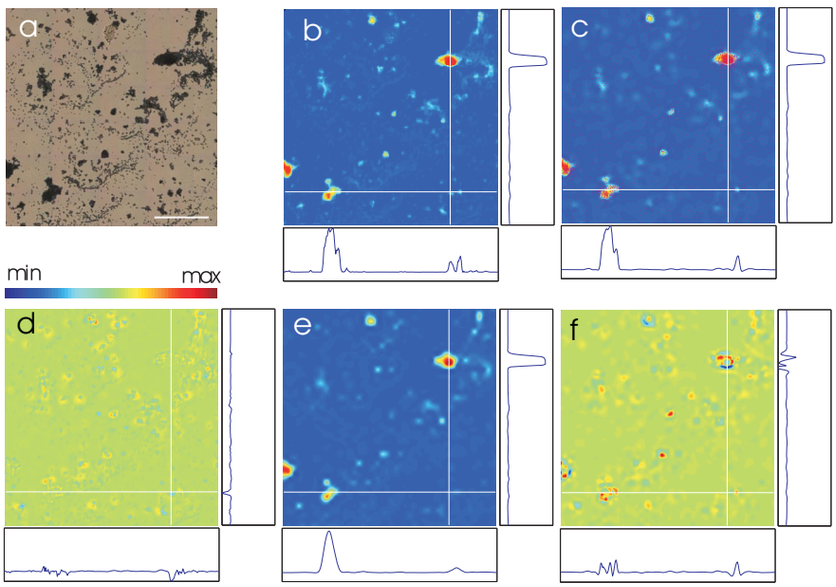


Figure 2 (online color at: www.biophotonics-journal.org) Raman mapping of randomly-distributed polystyrene microspheres (sample B3). (**a**) Brightfield image, scale bar is 100 μm. (**b**) A plot of the second principal component score for a full raster scan of the area depicted in (**a**), with a resolution of 256 × 256. (**c**) A 256 × 256 resolution image interpolated from the sampled locations, which are highlighted using magenta dots. (**d**) A difference image equivalent to (**c**)–(**b**). (**e**) An undersampled version of (**b**), with effective resolution of 48 × 48, resampled to 256 × 256 using a thin-plate spline. (**f**) A difference image equivalent to (**e**)–(**b**). Edge plots are cross-sectional profiles through the white lines illustrated on the main plot, and all axes scales, are equivalent for (**b**), (**c**), (**d**), (**e**) and (**f**). Colourmap scales are equal for (**b**), (**c**) and (**e**), and colourmap scales are also equal for (**d**) and (**f**), but the scale is smaller to better demonstrate differences.

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

tempting different undersampled images with increasing resolution) favours undersampling slightly since in the cases where there is little to be gained by increasing the number of sampling points, the new algorithm will be fixed in the number of points whereas undersampling will have many 'attempts' to get the lowest possible figure. It was deemed that this was a fairer comparison than the alternative of slightly favouring the newer algorithm. Despite this, in a majority of cases the new algorithm is superior, and clearly so. In only two cases does undersampling require fewer than 90% of the number of points used for the selective sampling algorithm, whereas there are many cases where undersampling required more than 120% of the points needed for the newer algorithm.

An illustration of the algorithm's performance is given in Figure 2(b) and (c), which also demonstrates that the algorithm samples a larger number of points near features of interest (the spot diameter is for illustration only, and does not accurately reflect the laser spot size). A difference image, Figure 2(d), is also provided, along with a map, Figure 2(e), that was obtained by undersampling the image, in order to demonstrate the additional benefit of selective sampling over simply reducing the number of sample points. This benefit is more obvious when comparing Figure 2(d) and (f); undersampling clearly leads to significantly larger erroneous features. Further illustrations with different samples are provided in the supplementary information.

4.2 Selecting the chemical features of interest

As previously noted, this algorithm is designed to operate in situations where the parameter of interest is scalar at all points in the image; for example, the location of a particular spectral component, or the Raman intensity at a particular wavenumber. Alternatively, multivariate measurements can be used to reduce the spectrum to a single number, such as projection onto the first principal component or determining the probability of finding a desired analyte. This criterion is fulfilled in many applications; the user is interested in the spatial distribution of a particular chemical, for example, or a particular combination of molecular spectra. This ability of the algorithm to search for features of interest in an image is demonstrated in Figure 3. The sample consists of a mixture of paracetamol and polystyrene microspheres on a soda-lime glass substrate, and the same dataset is used each time. The full raster image has a resolution of 128×128. In the first case, the linear model is the second principal component; upon investigation this appeared to strongly emphasize spectra with strong contributions from polystyrene, and the root-mean-squared error was 2.55% of data range. The second case used the third principal component as the model, and strongly emphasized paracetamol; root-mean-squared error was 1.00% of data range. A comparison can be made between the white light image and the two resulting hyper-spectral maps, to confirm that the spectral features being

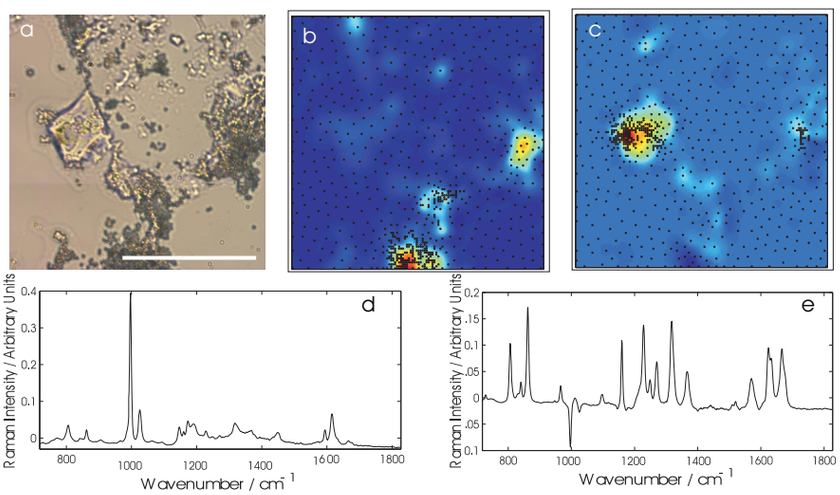


Figure 3 (online color at: www.biophotonics-journal.org) Selection of spectral features of interest for a three-component sample consisting of polystyrene microspheres, paracetamol powder and a soda-lime glass substrate. Red indicates a high principal component score, blue is low. (**a**) Bright-field image, scale bar is 50 μm. (**b**) An interpolated image where the second principal component has been used as the data reduction model. Sampling points are indicated with black dots. (**c**) An interpolated image where the third principal component has been used as the data reduction model. Sampling points are indicated with black circles. (**d**) Plot of the second principal component. (**e**) Plot of the third principal component.

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Journal of
BIOPHOTONICS

detected in the maps really do correspond to real-world features. Other principal components and a SCREE plot for the first 10 principal components is available in the supplementary information.

Of more interest for this particular illustration is the distribution of sampling points. There are 750 sampling points in total for each image, corresponding to approximately 20× fewer sampling points than a full raster scan. The points are observed to cluster around 'complex' regions in the image, where the interpolant varies significantly. As these regions will be subject to increased sampling density, they will consequently be measured with a higher overall accuracy than other less-variable regions of the image, demonstrating that the algorithm can search for certain spectral features in an image and maximize the time spent sampling those features in particular.

4.3 *Tissue sections*

In this section, skin tissue sections were imaged using the selective sampling technique to evaluate the potential for imaging large tissue sections on a timescale suitable for intra-operative use (10–30 minutes). The second principal component was used as a model for data reduction as it discriminated well between tissue structures, and all data were normalised as before. As before, the algorithm was assessed by selecting points from a fully-rastered dataset. The noise level was estimated by taking the second principal component of one 96×96 pixel Raman map and applying it to both that map and another taken at the same points over the same sample. The root-mean-squared error was then taken between the two datasets. In the case of a skin tissue sample, the root-mean-squared error was 7.73 and the data range was 56.1, therefore the root-mean-squared error due to noise is 13.8%, which is considerably higher than in the case of polystyrene samples.

An illustration of the algorithm's performance on a skin tissue samples is presented in Figure 4, along with an adjacent H & E image. Figure 4 shows the comparison between tumour regions (the darker regions in the H & E image) and the second principal component map for both rastered and interpolated sampling. There is a clear visual correlation between high second principal component scores and tumour regions, consistent with previous literature results [13]. In addition to the correlation between the Raman spectral image and H & E stain, the high signal-to-noise ratio of the measurements is capable of enabling accurate identification of individual skin structures. Typical Raman spectra corresponding to var-

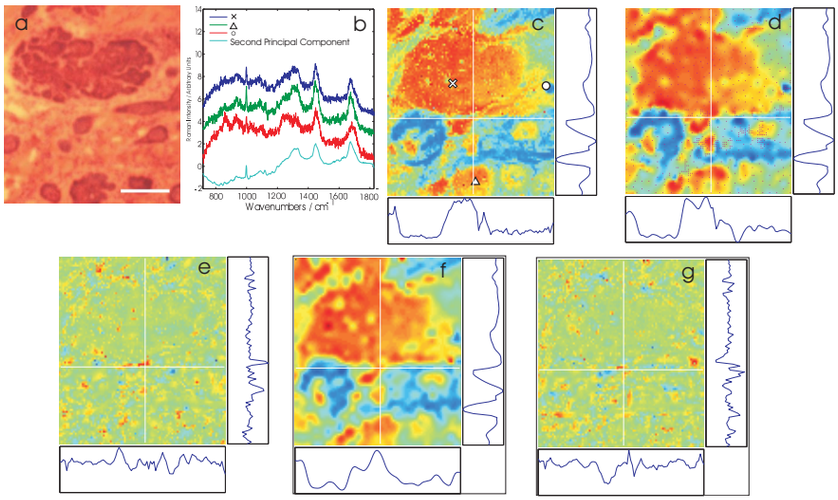


Figure 4 (online color at: www.biophotonics-journal.org) Demonstration of a 96×96 spectral image being reconstructed. Red indicates a high principal component score, blue is low. (**a**) Adjacent H & E image, scale bar is 200 μm. Basal cell carcinoma tissue is stained darker, on account of the larger cell nuclei. (**b**) Spectra measured at locations indicated in (**c**) and the second principal component used to plot images (**c**) and (**d**). (**c**) Full raster scan, with data reduction method being a projection onto the second principal component (total 9216 spectra). (**d**) Reconstructed image with the same data reduction method, using only 750 sampling points. Sampling locations are indicated with magenta dots. (**e**) A difference image, equivalent to (**c**)–(**d**). (**f**) An undersampled version of (**c**), with effective resolution of 48×48, resampled to 256×256 using a thin-plate spline. (**e**) A difference image, equivalent to (**c**)–(**f**). Edge plots are cross-sectional profiles through the white lines illustrated on the main plot, and all axes scales, are equivalent for (**c**), (**d**), (**e**), (**f**) and (**g**). Colourmap scales are equal for (**c**), (**d**) and (**f**), and colourmap scales are also equal for (**e**) and (**g**), but the scale is smaller to better demonstrate differences.

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ious regions identified in the spectral image showed a strong similarity with the Raman spectra of dermis and basal cell carcinoma reported previously [3]. Raman spectra of the large dermis regions are dominated by contributions from collagen I (proline bands around 855 cm^{-1} and 937 cm^{-1}, and a large band around 1267 cm^{-1} corresponding to the amide III region). Spectra from the epidermis contain stronger contributions from nucleic acids at 788 cm^{-1} and 1096 cm^{-1}; because basal cell carcinoma is derived from epidermal cells, the Raman spectra of these cells have also stronger bands associated with nucleic acids. Compared to the polystyrene microsphere samples, the distribution of sampling points in skin sections shows a reduced tendency to cluster at the edges of surface features (see Figure 4(d)). Because polystyrene has a very strong Raman cross-section, the transition between the spectrum of polystyrene and substrate is very sharp, meaning that the sampling points will be clustered strongly in this region. In contrast, the skin sections have comparatively subtle transitions, as well as smaller Raman cross sections and hence noisier spectra, meaning that the sampling is less prone to cluster around small regions.

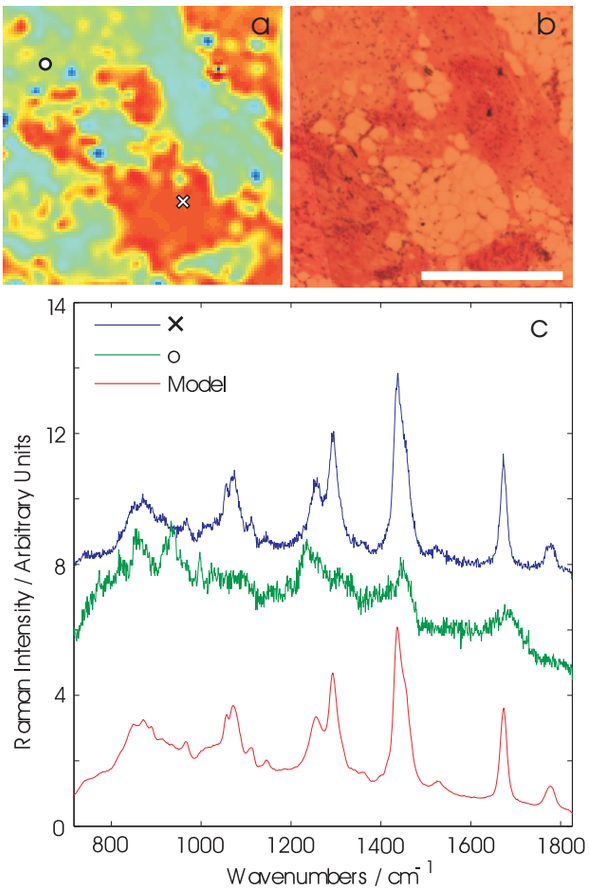


Figure 5 (online color at: www.biophotonics-journal.org) A skin tissue sample imaged using the selective sampling algorithm. (**a**) Interpolated map of fat regions. Red pixels have high fat content. (**b**) H & E stained image of an adjacent skin section; white regions indicate fat cells. Scale bar is 500 μm long. (**c**) Spectrum of the fat model used for data reduction, with two sample spectra for comparison. Sampling locations are indicated with symbols in (**a**).

4.4 Practical implementation

A typical example of a tissue section imaged using the selective sampling algorithm is presented in Figure 5 in which the data reduction model targeted regions of fat. 750 points have been sampled from a grid with a resolution of 100×100 pixels, and the sampled area is 1×1 mm; total sampling time was less than 30 minutes. The spectral feature used for the selective sampling was the score of each Raman spectrum calculated by projecting the spectrum on the Raman spectrum of typical fat regions found in skin tissues, which was measured on a different sample. Comparison between the Raman map and the H & E stained image shows that the targeted fat regions are detected with high contrast. Selected Raman spectra corresponding to fat and dermis regions are provided, along with the model used for data reduction, showing that high-signal-to noise spectra were recorded at each point of the image.

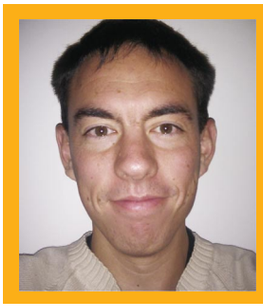
5. Conclusion

This paper outlines the development of a new sampling technique for the fast spectral imaging of large samples. This method has significant advantages compared to conventional raster-scanning: reduced time (in excess of 30-fold in some cases), preservation of high signal-to-noise spectra, and it also enables selection of the spectral features of interest. These advantages are more pronounced in samples with an appreciable degree of spatial correlation, especially cases where there are comparatively large uniform regions separated by well-defined boundaries. This situation might be observed, for instance, in the case of a nodular tumour, permitting the use of this technique in certain types of tumour margin evaluation. The primary assumption underlying this method is that there is a certain degree of spatial correlation in the image; points nearby are expected to have similar values. If this assumption does not hold then the method will be reduced to sampling all of the points in the image in order to achieve an acceptable estimation. Note however that this worst-case scenario still involves sampling the same num-

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ber of points as a conventional raster scan, so even in samples with minimal spatial correlation, the algorithm is still worth employing, provided the time taken to move the stage is negligible. If the integration time per pixel is low then a galvanometer or piezo-driven stage is recommended; galvanometers readily boast step responses of less than a millisecond (GVS001, Thor Labs) and piezo stages can be obtained with travel distances in excess of a few millimeters and a maximum-range settling time of a few tens of milliseconds (P-629.1CD, Physik Instrumente). Good performance has been demonstrated on several polystyrene microsphere substrates of different resolutions, along with mixed-component samples where individual chemical components could be identified and selectively sampled. The potential of this technique for sampling tissue sections is also demonstrated, but it should also be noted that the technique depends strongly on the discrimination ability of the multivariate model. That said, while other fast Raman imaging techniques can be used to provide contrast between different tissue structures, this selective scanning technique can do so while retaining chemical accuracy in the fingerprint region of the spectrum, a feature that supports accurate medical diagnosis and quantitative analysis. While the focus of this paper was on Raman spectroscopy and tissue imaging, it should be noted that these techniques can be applied to other point-by-point imaging techniques, especially those where a large amount of data must be captured at each pixel location.

Acknowledgements This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Invention for Innovation (i4i) Programme (II-AR-0209-10012). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Supporting Information Available: This material is available free of charge via the Internet.



Christopher J. Rowlands Dr. Christopher Rowlands graduated from Imperial College London with first class honours in Chemistry and studied for a Ph.D. at Cambridge University. He has published several papers on Surface Enhanced Raman Spectroscopy, the rapid prototyping of optical waveguides in arsenic trisulfide glass, as well as several new algorithms for smoothing, baseline-correction and separating out overlapping Raman spectra. After completing his thesis he joined the University of Nottingham, developing methods for the rapid acquisition of Raman spectral maps, with a particular interest in the use of Raman microscopy for the evaluation of tumour margins.

Sandeep Varma Sandeep is an experienced consultant dermatologist and dermatological surgeon at Queen's Medical Centre, Nottingham. He was trained in Mohs micrographic surgery (MMS) in Portland, Oregon, USA and undertakes excisional surgery and MMS for skin cancer including BCC. Sandeep is currently section editor for the British Journal of Dermatology and ex-chairman of the Karen Clifford skin cancer charity.

William Perkins William is a consultant dermatologic surgeon at Nottingham University Hospitals NHS Trust, QMC campus. He is also clinical director for dermatology. William leads the multidisciplinary skin cancer treatment service and has a research interest BCC. He is co-author of the Cochrane review on interventions for BCC, a chapter on BCC for the Evidence-Based Dermatology Textbook.

Iain Leach Iain is histopathology consultant with an interest in dermatopathology at Queen's Medical Centre, Nottingham. Iain has extensive experience in evaluation of skin biopsies and is lead dermatopathologist for skin cancer in Nottingham.

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Hywel Williams Hywel Williams directs the Centre of Evidence-Based Dermatology at the University of Nottingham, which contains the Cochrane Skin Group and the U.K. Dermatology Clinical Trials Network. He trained in dermatology at King's College Hospital and St. John's dermatology Centre in London before taking up his current position in Nottingham. Hywel is national lead for the Comprehensive Clinical Research Network dermatology speciality group and he also works as dermatology expert advisor for NHS Evidence, now run by NICE. He has a particular interest in evidence-based medicine and clinical trials. His main disease interests include childhood eczema, non-melanoma skin cancer and acne. Hywel is a NIHR Senior Investigator and is Chair of the NIHR Health Technology Assessment Commissioning Board and deputy director of the HTA Programme.

Ioan Notingher Ioan is Associate Professor in the School of Physics and Astronomy following a Ph.D. award from London South Bank University and research positions at Imperial College London and the University of Edinburgh. At Nottingham, Dr Notingher's research focuses on biomedical applications of Raman micro-spectroscopy (RMS), ranging from characterisation of bio-nanomaterials, non-invasive imaging of live cells and intra-operative imaging of tumours for cancer surgery.

References

[1] P. J. Caspers, G. W. Lucassen, and G. J. Puppels, Biophys. J. **85**, 572 (2003).
[2] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, Proc. National Acad. Sci. **102**, 12371 (2005).
[3] M. Larraona-Puy, A. Ghita, A. Zoladek, W. Perkins, S. Varma, I. H. Leach, A. A. Koloydenko, H. Williams, and I. Notingher, J. Biomed. Opt. **14**, 054031 (2009).
[4] A. Nijssen, T. C. Bakker Schut, F. Heule, P. J. Caspers, D. P. Hayes, M. H. A. Neumann, and G. J. Puppels, J. Invest. Dermatol. **119**, 64 (2002).
[5] J. Choi, J. Choo, H. Chung, D. G. Gweon, J. Park, H. J. Kim, S. Park, and C. H. Oh, Biopolymers **77**, 264 (2005).
[6] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang, and X. S. Xie, Science **322**, 1857 (2008).
[7] A. Zumbusch, G. R. Holtom, and X. S. Xie, Phys. Rev. Lett. **82**, 4142 (1999).
[8] Olaf Hollricher, Combine & Conquer. oe Magazine (November), 2003.
[9] A. Volkmer, J. Phys. D: Appl. Phys. **38**, R59 (2005).
[10] J. Hutchings, C. Kendall, N. Shepherd, H. Barr, and N. Stone, J. Biomed. Opt. **15**, 066015 (2010).
[11] S. Schlucker, M. D. Schaeberle, S. W. Huffman, and I. W. Levin, Anal. Chem. **75**, 4312 (2003).
[12] C. J. De Grauw, C. Otto, and J. Greve, Appl. Spectrosc. **51**, 1607 (1997).
[13] J. Hutchings, C. Kendall, B. Smith, N. Shepherd, H. Barr, and N. Stone, J. Biophotonics **2**, 91 (2009).
[14] P. J. Treado and Michael D. Morris, Appl. Spectrosc. **43**, 190 (1989).
[15] C. de Boor, A practical guide to splines, volume 27 of Applied Mathematical Sciences. Springer, New York, London, Revised edition, 2001.
[16] Michael L. Stein. Interpolation of Spatial Data. Springer Series in Statistics. Springer-Verlag New York Inc, 175 Fifth Avenue, New York, NY 10010, USA, First edition, 1999.
[17] R. Reuter, R. Willkomm, O. Zielinski, and W. Milchers, Hydrographic Laser Fluorosensing: Status and Perspectives. In J. H. Stel, H. W. A. Behrens, J. C. Borst, L. J. Droppert, and J. P. van der Meulen, editors, Operational Oceanography The Challenge for European Co-operation, volume 62, pages 251, Sara Burgerhartstraat 25, PO BOX 211, 1000 AE, Amsterdam, Netherlands, 1997. Elsevier Science BV.
[18] Jon Louis Bentley, Comm. ACM **18**, 509 (1975).
[19] kdtree code written by Steven Michael, downloaded from http://www.mathworks.com/matlabcentral/_leexchange/7030-kd-tree-nearest-neighbor-and-range-search on 8th March 201.
[20] DACE, by Hans Bruun Nielsen, Soren Nymand Lophaven and Jacob Sondergaard, was downloaded from http://www2.imm.dtu.dk/ hbn/dace/ on 8th March 2011.
[21] N. A. C. Cressie, Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics. (John Wiley & Sons, New York, Revised edition, 1993).
[22] Downloaded from http://www.mathworks.com/matlabcentral/_leexchange/20823-thin-plate-spline-network-with-radiohead-example on 8th March 2011.
[23] Downloaded from http://www.mathworks.com/matlabcentral/_leexchange/21524-principal-component-analysis on 8th March 2011.
[24] V. Rokhlin, A. Szlam, and M. Tygert, SIAM J. Matrix Anal. Appl. **31**, 1100 (2009).

© 2011 by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim