

Compressed AFM-IR hyperspectral nanoimaging

B Kästner^{1,*} , M Marschall¹ , A Hornemann¹ , S Metzner¹ , P Patoka² , S Cortes³ , G Wübbeler¹ , A Hoehl¹ , E Rühl^{2,*}  and C Elster¹ 

¹ Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany

² Physikalische Chemie, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany

³ Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade NOVA de Lisboa, Rua Junqueira 100, 1349-008 Lisbon, Portugal

E-mail: bernd.kaestner@ptb.de and ruehl@zedat.fu-berlin.de

Received 28 May 2023, revised 24 August 2023

Accepted for publication 21 September 2023

Published 3 October 2023



CrossMark

Abstract

Infrared (IR) hyperspectral imaging is a powerful approach in the field of materials and life sciences. However, for the extension to modern sub-diffraction nanoimaging it still remains a highly inefficient technique, as it acquires data via inherent sequential schemes. Here, we introduce the mathematical technique of low-rank matrix reconstruction to the sub-diffraction scheme of atomic force microscopy-based infrared spectroscopy (AFM-IR), for efficient hyperspectral IR nanoimaging. To demonstrate its application potential, we chose the trypanosomatid unicellular parasites *Leishmania* species as a realistic target of biological importance. The mid-IR spectral fingerprint window covering the spectral range from 1300 to 1900 cm^{-1} was chosen and a distance between the data points of 220 nm was used for nanoimaging of single parasites. The method of *k*-means cluster analysis was used for extracting the chemically distinct spatial locations. Subsequently, we randomly selected only 10% of an originally gathered data cube of 134 (*x*) \times 50 (*y*) \times 148 (spectral) AFM-IR measurements and completed the full data set by low-rank matrix reconstruction. This approach shows agreement in the cluster regions between full and reconstructed data cubes. Furthermore, we show that the results of the low-rank reconstruction are superior compared to alternative interpolation techniques in terms of error-metrics, cluster quality, and spectral interpretation for various subsampling ratios. We conclude that by using low-rank matrix reconstruction the data acquisition time can be reduced from more than 14 h to 1–2 h. These findings can significantly boost the practical applicability of hyperspectral nanoimaging in both academic and industrial settings involving nano- and bio-materials.

Keywords: AFM-IR, hyperspectral nanoimaging, low-rank matrix reconstruction, *Leishmania* parasites

* Authors to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1. Introduction

Infrared (IR) hyperspectral imaging (HSI), i.e. the recording of a spectrum at each pixel of a 2D specimen, is a powerful approach for non-invasive and non-ionizing materials characterization ranging from analytical chemistry [1, 2], materials sciences [3, 4], life sciences [5, 6] to microelectronics [7], minerals as well as planetary observations [8, 9]. HSI allows for directly correlating sample's morphology and topography with spectroscopic properties and offers the capability of high-throughput spatially resolved analysis when combined with multivariate statistics and machine learning tools, thus, enabling automated detection, pattern recognition, and phenotyping, particularly in the field of biomedical diagnostics [10–12].

Hyperspectral maps are usually obtained from parallel detection, for instance by employing focal-plane-array (FPA) detectors, which can considerably reduce data acquisition times via simultaneous detection of multi-pixel elements [13]. However, the highest spatial resolution of conventional optical techniques is limited by diffraction [14, 15]. The diffraction limit can be overcome by modern scanning-based methods, such as atomic force microscopy-based IR spectroscopy (AFM-IR) [16, 17], scattering-type scanning near-field optical microscopy (s-SNOM) [18, 19], photoinduced force microscopy [20], and tip-enhanced Raman spectroscopy [21]. As these approaches gather spectra in a sequential manner, HSI is highly time consuming and challenging for large sizes of 2D data arrays.

More specifically, AFM-IR is a powerful spectromicroscopy technique that has a spatial resolution below the diffraction limit of optical microscopy and is complementary to s-SNOM, where interferometry is used along with a broadband radiation source, such as synchrotron radiation, to determine local IR spectral properties [22]. This is different from AFM-IR along with HSI used in this work, where a tunable quantum cascade laser (QCL) provides an easily tunable, pulsed photon source that allows for the detection of local IR spectra at selected locations or sample arrays. The specific advantage is found in the ease of operation and the linear correlation of photo-thermal expansion and the local concentration of the species under study.

Typically, HSI measurements using scanning based methods may take several hours to reach a meaningful detection sensitivity (see, e.g. [23]). This, on the other hand, may lead to sample and tip damage as well as drift artifacts [24]. This compromise between sensitivity and data acquisition time inhibits unfolding the potential of HSI for reliably identifying and distinguishing chemical species. To overcome this limitation, we introduce the mathematical technique of low-rank matrix reconstruction [22, 25–27] to AFM-IR spectroscopy and show how to realize a compressed and correspondingly fast HSI-measurement by this method. We apply this technique to single *Leishmania* parasites as a realistic target of biological importance [28] and heterogeneous composition, well beyond simple model structures. This target also relates to our previous work

using diffraction limited FTIR spectromicroscopy [22]. Given the broad range of spectra present in the considered biological specimen, alternative compressed measurement techniques, such as compressed sensing [29, 30], are hardly applicable. This is due to the violation of the underlying sparsity assumption, which is usually not required for the techniques presented here [22].

2. The principle of compressed measurements in AFM-IR hyperspectral imaging

A scheme of the setup is shown in figure 1(a). During the experiments pulsed IR laser radiation from a tunable QCL (MIRcat-QTTM, DRS Daylight Solutions Inc. USA) was focused onto the sample in the proximity of the AFM tip of a nanoIR2-s setup (Anasys, Bruker). The corresponding absorption results in local rapid thermal expansion under the condition that the wavelength of the IR laser radiation matches the absorption bands of the samples under study. The induced expansion of the sample generates oscillations of the cantilever which are registered by changes of the deflection laser (DL) reflection spot on the four-quadrant photodiode (FQD) of the AFM. High sensitivity was achieved by using the resonance enhanced mode. In this mode the repetition rate of the laser is continuously matched to the contact resonance frequency of the cantilever which was in contact with the sample. Local absorption spectroscopy is realized by recording changes of the cantilever oscillation amplitude as a function of IR laser emission wavelength. In HSI the full spectral information is gathered in each pixel (distance between the data points: 220 nm) of the topographic map, allowing for a local chemical analysis of the sample related to its topography. Further details on the experimental setup can be found in the supplementary material.

The principle of a compressed measurement using AFM-IR imaging is shown in figures 1(b)–(d). AFM-IR spectral fingerprints were collected via HSI data acquisition in the 1300 to 1900 cm^{-1} spectral window from two single *Leishmania* parasites (*L. braziliensis* parasites, promastigote forms). An illustration of the full HSI data set as a 3D data cube is shown in figure 1(b). It consists of so-called voxels formed by 134×50 points along the (x, y) axes as spatial and 148 points along the wavenumber ($\tilde{\nu}$) axis as the spectral coordinate. The spectral resolution is approximately 4 cm^{-1} , with a data acquisition time of 8 s per spectrum. The total time for recording the full data cube was approximately 14 h.

The projection shown in figure 1(b) views the data cube along the spectral axis, where two spectrally integrated *Leishmania* parasites can be seen. As an example, the yellow circle in figure 1 indicates an arbitrarily chosen position to illustrate the data evaluation.

In a compressed measurement only a small fraction of selected voxels of the data cube are used. This is indicated by the blue dots in figure 1(c). In this example 10% of the full data set were used. For the spatial position indicated by the

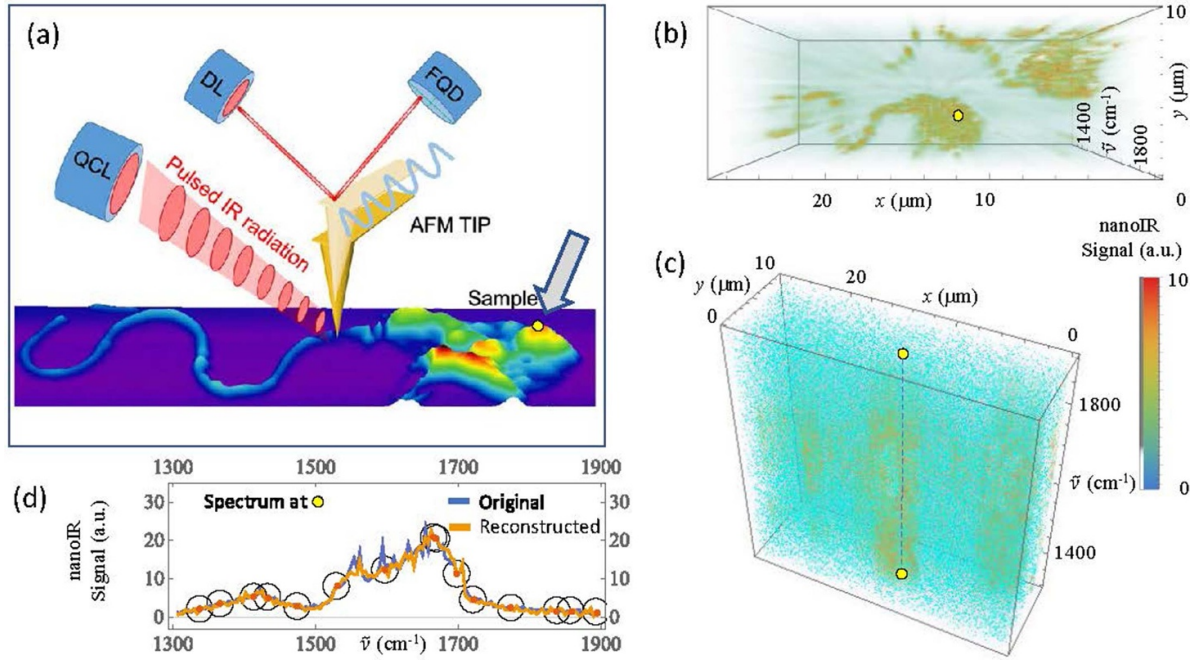


Figure 1. (a) Photothermal expansion experimental setup: QCL—tunable quantum cascade laser, DL—deflection laser, FQD—four-quadrant photodiode. The topography of the sample is shown, with the arrow pointing to the indicated sample spot marked by a yellow circle; (b) 3D representation of the data set viewed along the spectral axes. The color scale shown below encodes the signal strength both as color and transparency, so that the spatial distribution of two *Leishmania* promastigotes can be seen. The yellow circle corresponds to the indicated position indicated in (a); (c) illustration of the compressed measurement, showing the randomly chosen voxels in the data cube, to which low-rank matrix reconstruction was applied. The dashed line connecting the two yellow circles corresponds to the IR spectrum shown in (d); (d) nanoIR spectrum at the position marked by a yellow circle in (a)–(c). The original and the reconstructed data set are indicated by blue and orange curves, respectively. The marked points belong to the randomly chosen set of data used for reconstruction.

yellow dot, only 15 out of 148 wavenumber-values are actually considered during the measurement, cf the dashed line in figure 1(c).

By applying the method of low-rank matrix reconstruction [22, 26, 30], as detailed below, the full spectrum can be retrieved. Figure 1(d) shows, for example, an IR spectrum taken at the point marked by the arrow in figure 1(a) and yellow circles in figures 1(b) and (c). The full and reconstructed spectra are shown in blue and orange, respectively. Both curves are in agreement despite the fact that the reconstructed data is based on only 10% of the complete, original data set. These randomly selected points are marked by gray circles in figure 1(d).

3. Low-rank matrix reconstruction

Low-rank matrices arise in many settings related to mathematical modeling and data compression. Applications range from signal processing [31] to image restoration [32] and machine learning [33]. A relevant example of tasks includes the recovery of a data matrix given only by incomplete observations [22]. The data matrix is then approximated by a matrix product with each factor having lower dimensionality, resulting in a recovery result of lower rank.

The main idea is that a low-rank approximation already captures the main characteristics of the data and less informative dimensions will be removed. In [22], an algorithm for low rank matrix reconstruction has been presented and successfully applied to sub-sampled (FPA)-FTIR data. For a matrix of observations $X \in \mathbb{R}^{n,m}$, the task is to find matrices $U \in \mathbb{R}^{n,r}$ and $V \in \mathbb{R}^{m,r}$ by minimizing $\sum_{i,j} |X_{i,j} - \sum_{k=1}^r U_{i,k} V_{j,k}|^2$. The rank r approximation \hat{X} of X is then given by the matrix product $\hat{X} = UV^T$. The problem depicted above is in general ill-posed. Hence, an additional Tikhonov regularization is utilized [34] together with a smoothness constraint to the spatial domain. The resulting loss functional that needs to be minimized over the matrices U and V is then given by

$$J(U, V) = \sum_{i,j} |X_{i,j} - \sum_{k=1}^r U_{i,k} V_{j,k}|^2 + \lambda \{ \|KU\|_2^2 + \|V\|_2^2 \}, \quad (1)$$

where $\lambda > 0$ is chosen by an L-curve criterion and the matrix K denotes the precision matrix of a Gauß-Markov random field [35] to model smoothness in the spatial domain. To solve the minimization problem, an alternating algorithm is employed, since the optimization functional above reduces to a linear problem, when fixing one of the matrices U or V . The starting values for U and V are taken from a singular value

Table 1. Overview of relevant vibrational modes detected in the mid-infrared regime [39, 40].

Observed modes (cm ⁻¹)	Molecular vibration	Band assignments
1400–1420	$\nu(\text{C-N})$	Primary amide
1550–1562	$\delta(\text{N-H}), \nu(\text{C-N}), \nu(\text{C}=\text{N})$	NHR, secondary amine, protein and nucleic acid (amide II)
1614–1640	$\delta(\text{N-H})$	Primary amide (amide II)
1640–1677	$\nu(\text{C}=\text{O})$ or $\nu(\text{C}=\text{C}), \nu(\text{C}=\text{N})$	Amide I
1705–1725	$\nu(\text{C}=\text{O})$	ketones, -COOH
1735–1753	$\nu(\text{C}=\text{O}), \nu(\text{COOH})$	Saturated esters

decomposition of the zero-filled matrix X . For details on the algorithm and a corresponding Python implementation, see [22, 36].

4. Results and discussion

4.1. Spectral assignments

For HSI the following spectral regions were put into our focus (cf table 1): in the 1500–1900 cm⁻¹ spectral window proteins show amide I and amide II bands. Amide bonds are abundant in proteins because of their higher stability and proclivity for forming resonating structures, which influence secondary structure adoption and biological activity [37]. Particularly, band positions, band widths (full-width-at-half-maximum, FWHM), and shifts of amide bands may represent an important indicator for molecular structural (re-)organizations. Furthermore, an analysis of the area ratio between the amid I and II bands can indicate valuable spectral biomarkers and can provide a quantitative tool also for many biomedical applications [38]. Besides the amide window there is the so-called ‘mixed region’ (1300–1500 cm⁻¹), including fatty acid bending vibrations, C–N stretching and N–H deformational modes of proteins, as well as P = O stretching modes of phosphate-carrying species.

4.2. Cluster analysis imaging

Cluster analysis is a statistical approach for multivariate differentiation and classification studies of biological specimens and can be easily applied to IR-based hyperspectral data [41]. Meaningful clusters in the full data set were obtained from the following steps. Prior to clustering the data were pre-processed by smoothing along the spectral axis using the Savitzky–Golay-filter with a window size of 10 pixels and using 2nd order polynomials. Subsequently, the 2nd derivative is numerically computed using a finite difference scheme to reveal curvature information that improves separability of the subsequent clustering. Cluster analysis was performed using the well-known method of k -means clustering applied to all spectra in the data cube [42]. The pre-processing and clustering algorithm was implemented using Mathematica Version 11.0 [43].

The spectra were initially grouped into 15 clusters and the mean spectrum was determined for each cluster. Based on

the spectral position, $\tilde{\nu}_m$, of the maximal value of the mean spectrum, $A(\tilde{\nu}_m)$, the 15 clusters were then combined into 3 clusters. Depending on whether $\tilde{\nu}_m$ lies in the region within 1550–1640 cm⁻¹ or within 1650–1677 cm⁻¹ the corresponding cluster was assigned to cluster number 1 and 2, respectively. In this way, the regions where amide I or amide II are dominant were selected. The background was assigned to cluster number 3, with a maximal value below a threshold of 5 expressed as arbitrary units of the AFM-IR signal (cf figure 2).

4.3. Result of the reconstructions from 10% of the data

It is shown in the following that the information gained from cluster analysis is preserved after the reconstruction of the data from a subset of 10% of the data. Figures 2(a) and (b) compare the clusters obtained from the original and low-rank reconstructed data cubes, respectively. The low-rank reconstructions were performed on a workstation with 4 Intel Xeon CPU E7-8867 nodes and 504 GB of memory. The value of 10% of the data was chosen uniformly random from the complete data set. The mean spectra are shown for cluster 1, 2, and 3 as red, green, and black lines, respectively. The amide I and II spectral region is highlighted in blue and labeled as AI and AII, respectively. In addition, the region of primary amides referring to C–N stretching vibrations located at 1400 cm⁻¹–1420 cm⁻¹ has been highlighted and labeled as P. The cluster mean spectra of the original data cube resemble the reconstructed mean profiles and particularly the peak ratios: the mean spectrum of cluster 1 (red lines), which features an enhancement of the amide II/amide I band ratios compared to profiles of cluster 2 (green lines), are modeled correctly. The band shapes as well as their FWHM, could be successfully reproduced using the low-rank reconstruction. Furthermore, the baselines and spectral offsets can be completely recovered from low-rank modeling with respect to their magnitudes. Also note that the spectral features near 1735–1753 cm⁻¹ referring to the baselines occur at the same position.

Based on the spectral clusters chemically distinct spatial regions have been defined, as shown in the left-hand column of figure 2. The colors of the cluster correspond to the colors of the lines of the mean spectra in the right-hand column. Both cluster maps of the two *Leishmania* parasites contain the three clustered groups, given by the spectral features located mainly in the body (red), spectral features of the body

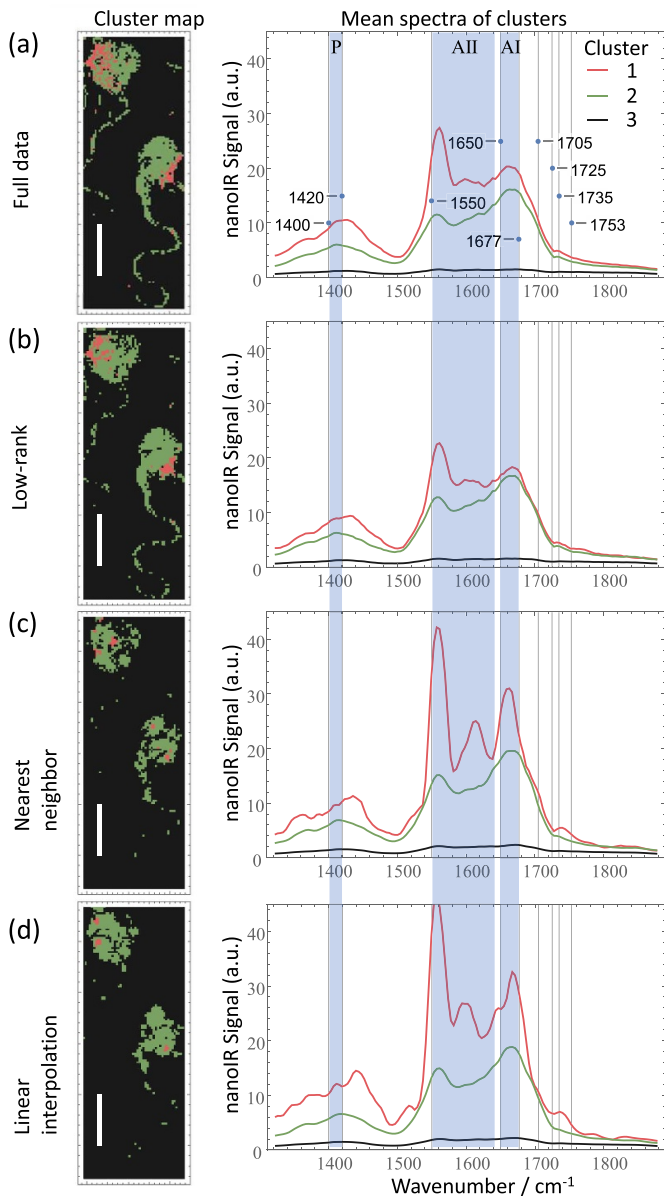


Figure 2. Results of cluster analysis performed using the full data set (a) and the data set reconstructed from 10% of the full data set (b). The left column shows the spatial distribution of the three spectral clusters as determined by the k -means method while on the right the spectra averaged over each of the three clusters are shown; The scale bar corresponds to $5\ \mu\text{m}$. The result for the alternative interpolation approaches of nearest neighbor and linear interpolation are shown (c) and (d), respectively. The relevant spectral regions are shaded and labeled as P: primary amides, AI the amide I, and AII the amide II region. The vertical lines correspond to some of the relevant wavenumber values given in table 1.

and the flagellum regions (green), and the spectral background (black). Notably, the parasites' morphologies are found correctly represented in the reconstructed cluster maps, as can be seen from the flagellum outline and body dimensions. This is a quite important aspect in view of the characterization of biological specimens, particularly in single cell analysis, and

microscopy studies for examining the microorganisms in combination with molecular nanoscopy for a detailed analysis of the contained species. If we compare the cluster maps obtained from the full and the low-rank reconstructed data cubes, the cluster allocations are, to a large extent, recovered at nearly the same pixel coordinates. However, small deviations can be observed. These are mainly due to a smoothing effect resulting from the application of the Tikhonov regularization in our analysis.

We note that while cluster 1 makes up only a small portion ($\approx 1\ \mu\text{m}$ in diameter) of the whole parasite the reversal of the band ratio may be of interest from the biochemical classification point of view [38]. Hence, pixel sizes, corresponding to sub-wavelengths resolution in hyperspectral AFM-IR measurements, are expected to significantly boost the sensitivity in classification as compared to conventional methods with a spatial resolution that is diffraction limited by the wavelength of the incident radiation. Therefore in the context of low-rank matrix reconstruction it is particularly important to note that cluster 1 and the corresponding band-ratio reversal has been correctly reproduced.

4.4. Comparison between different subsamplings and methods

The quality of the results obtained from low-rank matrix reconstruction is compared to alternative interpolation techniques, such as nearest neighbor interpolation and linear interpolation. Both interpolation methods were implemented using Python's SciPy interpolation package for irregular grids and applied to the same subsampled data that were used for the low-rank matrix reconstruction.

The results for these alternative approaches are shown in figures 2(c) and (d). It can be seen that in the cluster map the area of both *Leishmania* regions and in particular of cluster 1 is significantly reduced. The cluster locations and mean spectra deviate visibly from the original data set. As a quantitative measure to evaluate the reconstruction the root-mean-square-error (RMSE) between the available full data set and the reconstruction result has been evaluated for the different reconstruction schemes. Hereby, the RMSE is taken over all pixels and spectral positions. The results are presented in table 2, where we also compare different subsampling ratios. The low-rank reconstruction results in the lowest RMSE at all subsampling ratios. The linear interpolation yields a slightly higher RMSE than the low-rank reconstruction. One has to note that the linear interpolation is performed on the convex hull of the set of data points, defined by the subsampling. Therefore, points outside the convex hull are linearly extrapolated to 0 at the maximal and minimal wavenumbers to compute the RMSE in table 2 for the case of linear interpolation. The nearest neighbor interpolation always yields the highest RMSE.

The uncertainty introduced by the random sampling procedure is negligible. This has been confirmed by repeating the computations ten times with different, randomly drawn

Table 2. RMSE for the low-rank matrix reconstruction quality for different subsampling ratios and compared to alternative interpolation techniques, such as nearest neighbor interpolation and linear interpolation.

Subsampling	Low rank (this study)	Nearest neighbor	Linear interpolation
1%	3.2	3.8	3.4
2%	2.8	3.8	3.3
5%	2.4	3.5	3.1
10%	2.2	3.3	2.9
20%	2.2	3.0	2.7

subsamplings for the case of 1% of the original data. The uncertainty of the RMSE values presented in table 2 is basically smaller than the significant digits indicate. The fact that the uncertainty of the results in table 2 introduced by the choice of the random subsamples is already negligible for 1% of the data suggests that no significant variation is to be expected for the 10% case considered in this work.

The results have demonstrated that the low-rank methodology provides a reliable modeling approach to reconstruct multi-dimensional data of complex biological species, as not only the protein content lying behind the amide I/amide II bands, and their ratios, could be entirely recovered, but also their spectral background and offsets. The latter is likely due to wavenumber-dependent scattering or non-compensated reflection losses [44].

5. Conclusions

We presented for the first time a procedure for compressed AFM-IR spectroscopy using low-rank reconstruction for efficient hyperspectral IR nanoimaging. The suitability and application potential was demonstrated using signatures of the single parasitic specimens *L. braziliensis*. We have shown that the use of only 10% of randomly selected data from the original data cube enabled an adequate reconstruction of the entire data cube. The mid-IR fingerprint spectral complexity could be successfully reproduced, specifically in the amide and ‘mixed’ regions. This was evaluated by comparing cluster maps leading to the same bio-analytical information. Subsequent work should focus on sub-sampled data sets of biological species. This requires, however, that the data acquisition of the instrument also supports a fast approach of randomly selected AFM-tip positions. Translation of data reduction into a similar reduction of data acquisition time is possible by continuously measuring and moving all three axes. Metzner *et al* [45] provides a first description of suitable data acquisition schemes and their impact on data reconstruction quality, which remain to be experimentally realized. We conclude that the data acquisition times required for full hyperspectral measurements in AFM-IR procedures may be significantly reduced by low-rank matrix reconstruction schemes, promoting scanning based methods for hyperspectral imaging for which so far only parallel schemes qualify.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.6832523>.

Acknowledgments

We gratefully acknowledge fruitful discussions with Alexander Govyadinov. We acknowledge financial support by Deutsche Forschungsgemeinschaft (Grants EL 492/1-1, RU 420/13-1). Sofia Cortes was supported by Fundação para a Ciência e Tecnologia (FCT) for funds to GHTM (UID/04413/2020).

Supplementary material

AFM-IR nanoscopy setup

For the studies on single *Leishmania braziliensis* a 450 μm long gold-coated silicon AFM tip with a nominal radius of 25 nm (Bruker, model: PR-EX-nIR2-10, resonance frequency of 13 ± 4 kHz, spring constant: $0.07\text{--}0.4$ Nm^{-1}) was used in a nanoIR2-s spectrometer. The AFM-IR spectra were acquired within a range of $1306\text{--}1894$ cm^{-1} , employing a quantum cascade laser MIRcat-QTTM (DRS Daylight Solutions, USA) equipped with four diode modules. A sweep speed of 100 $\text{cm}^{-1} \text{ s}^{-1}$ was used. The average of the IR laser power applied to the sample was of 1.5 mW. The influence of power variation in the IR source at different wavenumbers was reduced by collecting the laser power spectrum and normalizing the AFM-IR amplitude at the specified wavenumbers. The hyperspectral imaging infrared maps were recorded in the resonance-enhanced mode with a distance between the data points of 220 nm, which was not the resolution limit (estimated to be 40 ± 15 nm) but chosen to achieve a realistic measurement time.











In the resonance-enhanced mode the repetition rate of the QCL laser is tuned to the continuously monitored AFM cantilever oscillation amplitude within a range of ± 30 kHz of oscillation resonance. The oscillation resonance during the scan has a slightly higher frequency than the cantilever’s free resonance in air and is continuously changing. This is due to the fact that the contact resonance of the cantilever in a snap-in position of the probing tip is varying and these changes are caused

by numerous factors, such as the contact surface between the probing tip and the investigated sample, the free resonance frequency of the cantilever, or stiffness of the sample. Thus, the oscillation resonance has to be tuned more frequently for complex samples.

Cultivation and preparation of *Leishmania* parasites

Promastigote forms of *Leishmania (V.) braziliensis* strain (MHOM/PE/94/LC2452c13) were cultured at $24\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ in M199 medium (Sigma Aldrich) supplemented with 10% fetal calf serum (FCS) (Sigma Aldrich), 1% L-glutamine, and 0.5% penicillin/streptomycin (Sigma Aldrich). A neutral pH was ensured by the addition of 0.04 M HEPES-NaOH buffer solution (pH 6.9). Parasites were collected between the 4th and 5th day in culture. The parasites' density was determined with a Casy counter (OLS OMNI Life Science) and concentration adjusted to $5 \times 10^6\text{ ml}^{-1}$ in order to receive thin films for AFM-IR investigations. After two washing steps with 0.5% phosphate-buffered saline (PBS) at 1000 g, for 10 min, parasites were resuspended in 300 μl of 0.5% PBS (or 100 μl for thick drops). A 5 μl drop of the suspension was placed onto a Kevley low-e-slide (Kevley Technologies[®]) and left to air-dry at room temperature. The parasites were microscopically examined with respect to their living or life cycle stages [46]. This should ensure that the parasites shall stay under near-native condition, when prepared and air-dried onto the substrates. The applied salt concentration was carefully evaluated and compromised with regard to keeping the parasites alive during preparation steps. We point out here that no other additives, i.e. FCS or bovine serum albumin, were used, as they contain proteins and may render the IR investigations unreliable with regard to protein detection in parasites.

ORCID iDs

B Kästner  <https://orcid.org/0000-0002-6575-6621>
 M Marschall  <https://orcid.org/0000-0003-0648-1936>
 A Hornemann  <https://orcid.org/0000-0002-0945-1508>
 S Metzner  <https://orcid.org/0000-0001-8879-1895>
 P Patoka  <https://orcid.org/0000-0002-3866-6903>
 S Cortes  <https://orcid.org/0000-0001-5850-6950>
 G Wübbeler  <https://orcid.org/0000-0002-6871-8903>
 A Hoehl  <https://orcid.org/0000-0003-0813-7514>
 E Rühl  <https://orcid.org/0000-0002-0451-8734>
 C Elster  <https://orcid.org/0000-0003-0113-3713>

References

- [1] Salzer R and Siesler H W 2009 *Infrared and Raman Spectroscopic Imaging* (Wiley)
- [2] Morsch S Lyon S Edmondson S Gibbon S 2020 Reflectance in AFM-IR: implications for interpretation and remote analysis of the buried interface *Anal. Chem.* **92** 8117–24
- [3] Ruggeri F S Longo G Faggiano S Lipiec E Pastore A Dietler G 2015 Infrared nanospectroscopy characterization of oligomeric and fibrillar aggregates during amyloid formation *Nat. Commun.* **6** 7831
- [4] Chae J Dong Q Huang J Centrone A 2015 Chloride incorporation process in $\text{CH}_3\text{NH}_3\text{PbI}_{3-x}\text{Cl}_x$ perovskites via nanoscale bandgap maps *Nano Lett.* **15** 8114–21
- [5] Pilling M Gardner P 2016 Fundamental developments in infrared spectroscopic imaging for biomedical applications *Chem. Soc. Rev.* **45** 1935–57
- [6] Paluszkiwicz C Piergies N Chaniecki P Rekas M Miszczyk J Kwiatek W 2017 Differentiation of protein secondary structure in clear and opaque human lenses: AFM – IR studies *J. Pharm. Biomed. Anal.* **139** 125–32
- [7] Lau W S 1999 *Infrared Characterization for Microelectronics* (World Scientific Publishing)
- [8] Brown A 2006 Spectral curve fitting for automatic hyperspectral data analysis *IEEE Trans. Geosci. Remote Sens.* **44** 1601–8
- [9] Brown A J Viviano C E Goudge T A 2020 Olivine-carbonate mineralogy of the jezero crater region *J. Geophys. Res.: Planets* **125** e2019JE006011
- [10] Lu G Fei B 2014 Medical hyperspectral imaging: a review *J. Biomed. Opt.* **19** 10901
- [11] Lasch P Stämmler M Zhang M Baranska M Bosch A Majzner K 2018 FT-IR hyperspectral imaging and artificial neural network analysis for identification of pathogenic bacteria *Anal. Chem.* **90** 8896–904
- [12] Hu B Du J Zhang Z Wang Q 2019 Tumor tissue classification based on micro-hyperspectral technology and deep learning *Biomed. Opt. Express* **10** 6370–89
- [13] Levin I W Bhargava R 2005 Fourier transform infrared vibrational spectroscopic imaging: integrating microscopy and molecular recognition *Annu. Rev. Phys. Chem.* **56** 429–74
- [14] Gao L Smith R T 2015 Optical hyperspectral imaging in microscopy and spectroscopy—a review of data acquisition *J. Biophoton.* **8** 441–56
- [15] Chan K L A Fale P L V Atharawi A Wehbe K Cinque G 2018 Subcellular mapping of living cells via synchrotron microFTIR and ZnS hemispheres *Anal. Bioanal. Chem.* **410** 6477–87
- [16] Dazzi A Prazeres R Glotin F Ortega J M 2005 Local infrared microspectroscopy with subwavelength spatial resolution with an atomic force microscope tip used as a photothermal sensor *Opt. Lett.* **30** 2388
- [17] Dazzi A Prater C B 2017 AFM-IR: technology and applications in nanoscale infrared spectroscopy and chemical imaging *Chem. Rev.* **117** 5146–73
- [18] Knoll B Keilmann F 1999 Near-field probing of vibrational absorption for chemical microscopy *Nature* **399** 134
- [19] Keilmann F and Hillenbrand R 2009 *Nano-Optics and Near-Field Optical Microscopy*, ed A Zayats and D Richard (Artech House) ch 11, pp 235–65
- [20] Rajapaksa I Uenal K Wickramasinghe H K 2010 Image force microscopy of molecular resonance: A microscope principle *Appl. Phys. Lett.* **97** 073121
- [21] Stöckle R M Suh Y D Deckert V Zenobi R 2000 Nanoscale chemical analysis by tip-enhanced Raman spectroscopy *Chem. Phys. Lett.* **318** 131–6
- [22] Marschall M Hornemann A Wübbeler G Hoehl A Rühl E Kästner B Elster C 2020 Compressed FTIR spectroscopy using low-rank matrix reconstruction *Opt. Express* **28** 38762
- [23] Amenabar I Poly S Goikoetxea M Nuansing W Lasch P Hillenbrand R 2017 Hyperspectral infrared nanoimaging of organic samples based on Fourier transform infrared nanospectroscopy *Nat. Commun.* **8** 14402

- [24] Canale C Torre B Ricci D Braga P C 2011 *Atomic Force Microscopy in Biomedical Research: Methods and Protocols* (Humana Press) pp 31–43
- [25] Candes E J and Recht B 2008 Exact low-rank matrix completion via convex optimization *2008 46th Annual Allerton Conf. on Communication, Control and Computing* pp 806–12
- [26] Davenport M A Romberg J 2016 An overview of low-rank matrix recovery from incomplete observations *IEEE J. Sel. Top. Signal Process.* **10** 608–22
- [27] Wübbeler G Elster C 2021 Efficient experimental sampling through low-rank matrix recovery *Metrologia* **58** 014002
- [28] Leishmaniasis 2022 (available at: www.who.int/en/news-room/fact-sheets/detail/leishmaniasis)
- [29] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306
- [30] Kästner B Schmähling F Hornemann A Ulrich G Hoehl A Kruskopf M Pierz K Raschke M B Wübbeler G Elster C 2018 Compressed sensing FTIR nano-spectroscopy and nano-imaging *Opt. Express* **26** 18115–24
- [31] Weng Z and Wang X 2012 Low-rank matrix completion for array signal processing *2012 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 2697–700
- [32] Peng Y Suo J Dai Q Xu W 2014 Reweighted low-rank matrix recovery and its application in image restoration *IEEE Trans. Cybern.* **44** 2418–30
- [33] Yao Q Kwok J T Zhong W 2015 Fast low-rank matrix learning with nonconvex regularization *2015 IEEE Int. Conf. on Data Mining* pp 539–48
- [34] Engl H W Hanke M Neubauer A 1996 *Regularization of Inverse Problems* vol 375 (Kluwer Academic Publishers)
- [35] Rue H and Held L 2005 *Gaussian Markov Random Fields: Theory and Applications* (Chapman and Hall/CRC)
- [36] Marschall M, Wübbeler G, and Elster C 2021 Regression-working group 8.42 (available at: www.ptb.de/cms/nc/en/ptb/fachabteilungen/abt8/fb-84/ag-842/regression-842.html) (Accessed 14 December 2021)
- [37] Mahesh S Tang K-C Raj M 2018 Amide bond activation of biological molecules *Molecules* **23** 2615
- [38] Ricciardi V Portaccio M Manti L Lepore M 2020 An FTIR microspectroscopy ratiometric approach for monitoring x-ray irradiation effects on SH-SY5Y human neuroblastoma cells *Appl. Sci.* **10** 2974
- [39] Parker F S 1975 Biochemical applications of infrared and raman spectroscopy *Appl. Spectrosc.* **29** 129–47
- [40] Helm D Labischinski H Schallehn G Naumann D 1991 Classification and identification of bacteria by Fourier-transform infrared spectroscopy *Microbiology* **137** 69–79
- [41] Hornemann A Sinning D Cortes S Campino L Kuhls K Ulm G Frohme M Beckhoff B 2017 A pilot study on fingerprinting Leishmania species from the Old World using Fourier transform infrared spectroscopy *Anal. Bioanal. Chem.* **409** 6907
- [42] Hartigan J A 1975 *Clustering Algorithms* (Wiley)
- [43] Wolfram Research, Inc. 2021 Mathematica, Version 11 (available at: www.wolfram.com/mathematica)
- [44] Centrone A 2015 Infrared imaging and spectroscopy beyond the diffraction limit *Ann. Rev. Anal. Chem.* **8** 101–26
- [45] Metzner S Kästner B Marschall M Wübbeler G Wundrack S Bakin A Hoehl A Rühl E Elster C 2022 Assessment of subsampling schemes for compressive Nano-FTIR imaging *IEEE Trans. Instrum. Meas.* **71** 4506208
- [46] Wheeler R J Gluenz E Gull K 2013 The limits on trypanosomatid morphological diversity *PLoS One* **8** e79581