Supplementary material

Experimental

The VNIR Spectral Camera HS V10E by Specim (Oulu, Finland) of the RIS instrument is sensitive to the spectral range from 400 to 1000 nm. Its sensor is a Si Charge-Coupled Device (CCD) with 1600 × 1200 pixels (lateral × spectral). The spectral dimension was binned down to 212 channels with a resulting spectral sampling of 2.8 nm. Under the chosen experimental conditions, the pixel size was 250 µm, but with a macro objective, a lateral resolution of 40 µm was reached in other measurements.

For normalization, a Spectralon white reference (Specim) placed next to the investigated area was used.

Reference spectra were acquired in the laboratory using the same RIS instrument used on site. Commercial pigments were applied on polished marble with different binders. Red Ochre, Egyptian blue (both from COLORARE, Villenave-d'Ornon, France) and Chalk (Marin Beaux Arts, Arcueil, France) were applied with limewater. Orpiment and yellow ochre (both from Laverdure & fils, Paris, FR) were applied with animal glue as limewater did not provide satisfying coverage on this substrate. A final reference spectrum was acquired from a pressed pellet of Egyptian green (Kremer Pigmente, Aichstetten, Germany) via a fiber optics reflectance spectrometer (FORS) (USB-4000 spectrometer and HL-2000 light source, Ocean Optics, Largo, FL, USA).

The in-house built XRF scanner made use of a 3W Pd transmission anode X-ray tube (Moxtek, Orem, UT, USA). A 0.8-mm pure Pd collimator (Goodfellow, Huntingdon, UK) embedded in a block of silver was used. This yielded a beam of approx. 1 mm. The use of pure Pd prevented the presence of blind signals in the excitation spectrum. The silver block was obtained by a lost-wax casting from a 3D printed wax model (Sculpteo, Villejuif, France). The same company manufactured the 3D printed holder for collimator, X-ray tube and detector.

As detector, a X-123FAST SDD by Amptek (Bedford, MA) was used. Its active area of 25 mm^2 was collimated to 17 mm^2 and an energy resolution of 129 eV was measured at 5.9 keV. The motorized stages of 30 cm travel range were supplied by Thorlabs (Newton, New Jersey, USA).

The sensitivity and correct alignment of the XRF instrument was controlled, every morning and after every re-assembling, by measuring a National Institute of Standards and Technology Standard Reference Material (NIST SRM) 610 "Trace elements in glass" with a nominal concentration level of 500 ppm for a wide range of elements. Its precise characteristics are given elsewhere[1].

From these, limits of detection were calculated. For 300 s of measurement they were found to be between 120 ppm (Ca, $Z = 20$) and 22 ppm (Cu, $Z = 29$) for elements heavier than Ca ($Z = 20$).

Data evaluation

The RIS was normalized to an external Spectralon white reference (Specim) and dark images acquired with closed shutter to take the noise of the camera into account. The normalized spectra were smoothed by a Savitzky–Golay filter.

For the calculation of *endmembers* use was made of the "Spectral Hourglass Wizard" of the ENVI software package (Harris Corporation, Melbourne, Florida, USA). The hyperspectral data cube was represented by 6 principal components yielded by minimum noise fraction transformation (MNF). With these the Pixel Purity Index (PPI) for each pixel was calculated to identify potential *endmembers*. The 10.000 "purest" pixels were displayed in the n-dimensional visualization tool and *endmembers* were selected manually in this representation. Finally, the hyperspectral data was represented making use of the SAM routine of ENVI. The further evaluation and presentation of the spectra was done in Python 2.7.12.

For alignment of the images we made use of the Open Source Computer Vision Library (OpenCV, opencv.org, version 2.4.13.6) in python. Two grey scale representations of XRF and RIS were calculated. The RIS representation is an integration of all channels in the visible wavelength, while for the XRF data Fe and Cu distribution images were subtracted from the Ca distribution image and the result normalized. To have a comparable lateral resolution in both images, the XRF representation was enlarged by a factor of 3, while the RIS representation was smoothed with a Gaussian filter. These blurred and interpolated representations were used only for the detection of features. The fusion of the hyperspectral data is described further below.

Making use of SIFT (Scale-invariant feature transform), common features in both representations were detected. These were inspected, and misaligned features manually removed, a process that will be improved in the future.

Afterwards the transformation matrix for alignment was calculated via homography for the original RIS data to the original sized XRF data set, making use of OpenCV. The RIS data with its higher lateral resolution was aligned to the XRF data as in this way the transformed RIS data featured weighted averages of several pixels and not simply and interpolation, as if a data set is enlarged. Flat objects investigated by a rotating line scanner often have "pin-cushion" distortion as the field of view is dependent on the distance between camera and surface. However, this distortion was negligible for the investigated area.

From the aligned hyperspectral data cubes, a fused set was calculated. Care was taken to balance the statistical weight put on different spectral features. For this, meaningful spectral ranges (465-840 nm for RIS and 0.9 to 18.31 keV for XRF) free of artefacts were selected.

The RIS data was added twice to the fused data. The smoothed reflectance spectra were scaled to have values between 0.0 and 1.0. The first derivative of the smoothed spectrum, also scaled to values between 0.0 and 1.0 was added. This put statistical weight on the edge of the reflectance spectra.

The XRF data was re-binned, reducing the number of channels by a factor of 5, preserving a bin size (37.5 eV) smaller than the energy resolution of the detector (129 eV). These spectra were smoothed by a Savitzky–Golay filter and, instead of simply adding the smoothed XRF data to the fused data, its logarithm was calculated and scaled to have values between 0.0 and 1.0 to give weight to weak lines.

The resulting fused data cube had 32,000 pixels (160 × 200) with 784 channels. 464 (59%) of these channels represented the XRF data and the remainder (320 channels, 41%) representing the RIS data. As the data sets were scaled to be between 0. and 1., this gave both methods a roughly equal statistical weight, especially as a significant part representing of the XRF data just containing "empty" spectral background. The number of channels was reduced using the 25 first principal components of principal component analysis (PCA), that explained more than 90% of the variance of the data. This data was processed in t-SNE with a perplexity of 30, using 5000 iterations with the Barnes-Hut method. This evaluation took, on an early 2015 Macbook Pro, around 2000 s. This calculation was repeated 10 times and the *representation* showing the clearest features was selected for this paper. Clusters were manually selected from the obtained representation of the data. For PCA and t-SNE scikit-learn (scikit-learn.org, version 0.19.1) was used.

Reference

[1] K.P. Jochum, U. Weis, B. Stoll, D. Kuzmin, Q. Yang, I. Raczek, D.E. Jacob, A. Stracke, K. Birbaum, D.A. Frick, D. Günther, J. Enzweiler, Geostand. Geoanal. Res. 35 (2011) 397–429.