○ **Heritage Science**

Open Access

CrossMark

Assessment of multispectral and hyperspectral imaging systems for digitisation of a Russian icon

Lindsay W. MacDonald[1*], Tatiana Vitorino[2,3], Marcello Picollo[3], Ruven Pillay[4], Michał Obarzanowski[5], Joanna Sobczyk[5], Sérgio Nascimento[6] and João Linhares[6]

Abstract

In a study of multispectral and hyperspectral reflectance imaging, a Round Robin Test assessed the performance of different systems for the spectral digitisation of artworks. A Russian icon, mass-produced in Moscow in 1899, was digitised by ten institutions around Europe. The image quality was assessed by observers, and the reflectance spectra at selected points were reconstructed to characterise the icon's colourants and to obtain a quantitative estimate of accuracy. The differing spatial resolutions of the systems affected their ability to resolve fine details in the printed pattern. There was a surprisingly wide variation in the quality of imagery, caused by unwanted reflections from both glossy painted and metallic gold areas of the icon's surface. Specular reflection also degraded the accuracy of the reconstructed reflectance spectrum in some places, indicating the importance of control over the illumination geometry. Some devices that gave excellent results for matte colour charts proved to have poor performance for this demanding test object. There is a need for adoption of standards for digitising cultural heritage objects to achieve greater consistency of system performance and image quality.

Keywords: Multispectral, Hyperspectral, Imaging, Digitisation, Reflectance, Spectrum, Specularity, Cultural heritage, Icon, Standards

Introduction

Scientific examination and accurate digital photographic recording of cultural heritage (CH) artworks are key for their contextualisation, documentation and conservation. Non-contact scientific examination may enable the identification of an artist's materials and offers insight into the construction of an object without affecting its integrity. The information obtained expands knowledge about the production technology and visual appearance of the artwork, and the artist's working methods. Precise and complete digital archival images also provide a definitive record of the object at the time of acquisition. If the resolution is high enough, the digital record also constitutes a means for the documentation and accessibility of CH [1–3]. For all of these purposes, accuracy and

high quality are important aspects of the data acquired. Records should be true representations of the original artwork, without anything added or taken away.

Acquisition of digital images and associated data serves as a good basis for repeatable research, across both time and place of different institutions, allowing the exchange and safe, non-destructive analysis of obtained results. Digital techniques enable a better and more precise description of all parameters during the acquisition and subsequent processing, including calibration and correction. Applications of spectral image acquisition systems include: imaging in broad bands by multispectral systems; quantification of reflectance spectra at each point by hyperspectral systems; recording of a data cube enabling the reconstruction of images (visible RGB, infrared, greyscale), and statistical processing for mapping, land usage, crop distributions, etc. Many types of spectral image acquisition systems have been applied for the study and digital documentation of cultural heritage artworks.

*Correspondence: lindsay.macdonald@ucl.ac.uk
[1] Faculty of Engineering Sciences, University College London, London, UK
Full list of author information is available at the end of the article

MacDonald *et al. Herit Sci* (2017) 5:41

Page 2 of 16

Depending on the device and workflow, however, the results obtained may vary considerably in quality.

Colour and Space in Cultural Heritage (http://www.COSCH.info) was a European network of researchers, conservators and museum professionals, supported by the COST programme. This 4-year trans-domain Action (TD1201, 2013–2016) explored high-resolution optical techniques, defining good practice and open standards for digitisation and documentation of cultural objects. It was focused especially on practitioners of heritage science, i.e. technical staff with high levels of expertise in laboratories and university research facilities, who were associated with national museums.

The primary objective of COSCH Working Group 1 (WG1) was the "identification, characterisation and testing of spectral imaging techniques in the visible and near infrared field". In this context, a Round Robin Test (RRT) exercise was conducted to compare the characteristics and performance of different spectral imaging systems. Five test objects were analysed at 21 institutions, each with its own instrumentation and workflow. This article is focused on one of the RRT objects by qualitative comparison of the spectral imaging data obtained from ten different imaging systems. The multiple datasets were compared through visual inspection of both global and local areas, and the effectiveness of each system for dealing with this type of object was subjectively assessed. In addition, a reconstruction of the reflectance spectrum at selected points was undertaken. Although several systems in the complete RRT recorded infrared spectral reflectance of up to 2500 nm, the ten systems reported in this article were limited to the visible and near infrared ranges, with a maximum wavelength of 1000 nm.

Test object: a Russian icon

One of the RRT objects, the subject of this article, was a polychrome Russian icon from the late 19th century. Icons constitute a form of religious art, with stylised sacred images of saints, Christ and the Virgin Mary, in which the figures often appear timeless and static against golden backgrounds. The COSCH test object, hereafter described as the Russian icon, is cuboidal with dimensions 267 mm × 222 mm × 23 mm. It was purchased through eBay in January 2014 from a dealer in Tallinn, Estonia. It depicts the traditional subject "Virgin of Kazan" but is printed by chromolithography, not painted by hand, onto a substrate of tinned steel. It gives the impression of the lid of an old biscuit tin, which has been nailed around the edges onto a wooden support. The surface is slightly convex, and its condition is marred by many tiny spots of rust on the underlying steel showing through the printed pattern. The icon was manufactured in Moscow in 1899 by the firm Zhako and Bonaker,

whose speciality was making containers for shoe polish. The owners had invited master-painters from Mstera, a village in Vladimir Province, who were experts in the ancient techniques and styles of icon painting, to produce new designs. This capitalistic approach to mass-production had an adverse effect on the craft and artistry of icon painting, but the iconographic motifs, combined with surrogate gold (*samovarnogo*) and false jewels, made an indelible impression on the mass public, who purchased the products in their millions [4] (Fig. 1).

The Russian icon was chosen as a test object for the study because of its particular physical and material characteristics. Instead of a matte finish, it has a medium gloss on the painted areas and a shiny metallic finish on the gold. There is a wide range of tones and colours, including skin tones and a pictorial realism to the figures, which make it suitable for image quality assessment. The fine spatial detail in both the half-tone colour separations and the filigree gold decoration are appropriate for testing the optical resolving power of an imaging system. The surface is flat, almost planar, with virtually no 3D relief, and it has a convenient size and format. Because it is a common mass-produced object, rather than a fragile artwork, it can be handled freely by researchers in the laboratory without gloves. It turned out to be a very demanding test object, revealing several critical aspects of system performance.
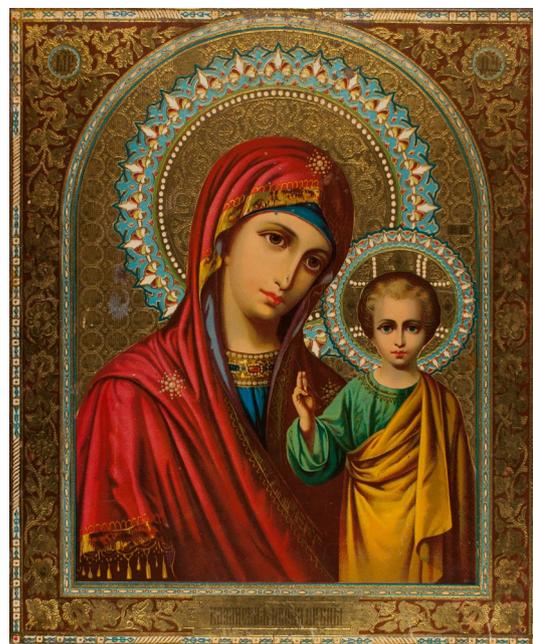
Fig. 1 Photograph of the Russian icon, manufactured in 1899 by the firm Zhako and Bonaker (Moscow), taken by a Nikon D200 camera on a copystand under tungsten illumination (for which see Fig. 13)

MacDonald *et al. Herit Sci* (2017) 5:41

Page 3 of 16

Spectral reflectance image capture

Spectral imaging systems, which scan the entire surface of an artwork without physical contact, enable spectroscopic information to be obtained as an accurate digital record. They have increasingly been included in the range of analytical techniques available for the conservation and study of CH [5–8]. Spectral imaging devices acquire images in many different wavebands, enabling the reflectance spectrum to be sampled at each pixel and hence the colorimetric representation of the image to be calculated for any specified spectrum of visible illumination. Note that in this study we are dealing with multi/hyperspectral imaging based on directional or diffuse reflectance of illumination from a surface as a function of wavelength. As an image acquisition modality, this is distinct from X-ray fluorescence scanning to obtain an XRF spectrum at every single point on an object, which could also be claimed as a type of hyperspectral imaging.

An imaging system is commonly considered to be multispectral if the number of spectral bands is greater than 3 and less than 20, for which the working bandwidth is between 10 and 50 nm. In turn, systems with hundreds of contiguous bands with a bandwidth less than 5 nm are considered hyperspectral [9]. The definitions of multispectral and hyperspectral are very author-dependent. Goetz argues that the term multispectral should be used if no contiguous bands are acquired, irrespective of the range of spectral sensitivity of the equipment, while hyperspectral should apply if contiguous bands are acquired [10]. With this definition in mind, all of the equipment used in the present study may be considered as hyperspectral, some limited to the visible range and others covering the Visible + NIR range. Recently the new term ultraspectral has been proposed for systems where there are thousands of bands with a bandwidth of as little as 0.1 nm [11].

All these image acquisition devices generate datasets with three dimensions (two spatial and one spectral), commonly represented as a data-cube [12]. Their performance, however, is dependent on the many components of the imaging system, including illumination, optics, filters, dispersive gratings, sensor and signal processing, and also on operator choices relating to object positioning, calibration, file format and metadata [13]. Usually, each institution has its own unique set-up and specific workflow, which are dependent on the working principles of the acquiring system, and hence can significantly affect the information acquired for the same object [14]. As an example, the standard procedure in the National Museum in Krakow, Poland, involves a series of tasks requiring manual adjustment and judgement, as follows:

a. As there is no set working distance, first the area to be scanned is defined, bearing in mind that the larger the frame the coarser the spatial sampling.
b. Focusing of the optics to obtain a sharp image.
c. Placement of lights, setting the proper distance and angle (usually 45°).
d. Setting exposure level to obtain maximum signal strength without over-exposure.
e. Regulating movement speed of the camera stage to achieve true 1:1 geometric scale.
f. Adjusting frame rate and sampling time for data acquisition.
g. Recording the image.
h. Recording data for image normalisation: white target scan and dark image scan.
i. Applying a normalisation algorithm to the image data.
j. Image data is saved in a standard format.
k. Entry of metadata relating to image.

This specificity not only makes it difficult to compare data from different systems, as it can influence the accuracy and reliability of the recorded information, but also causes some systems to be inappropriate for particular objects. Spectral imaging as a method of documentary recording of cultural artefacts would be much more useful if standardised procedures could be established for a wide range of systems, users and materials.

Various earlier projects have contributed to the development of spectral imaging technology for the CH sector and have succeeded in designing high-performance hardware and software solutions for accurate image acquisition and processing. These commenced in the early 1990s with the VASARI project at The National Gallery, London, which achieved accurate high-resolution colorimetric images of paintings using a multispectral scanning system with seven filters in the visible region [15, 16]. The CRISATEL project later extended the work into the NIR region and applied basic spectroscopy techniques to the results [17]. Advances in technology made hyperspectral imaging possible in the 2000s [18]. The use of 'pushbroom' techniques was pioneered by IFAC-CNR in Florence [19, 20] and by the National Gallery of Art in Washington [21]. Simple photographic methods with transmission filters applied to either the camera or the illumination remain viable, however, and may be considered as the technological baseline against which more sophisticated systems can be compared [22].

Recent applications of hyperspectral imaging in cultural heritage include two Italian illuminated manuscripts, where the characteristics of the spectral reflectance signal in the visible range were used to identify pigments

MacDonald *et al. Herit Sci (2017) 5:41*

Page 4 of 16

by comparing the obtained spectra with those of a reference library of medieval pigments [23]. In another study, the analytical suitability of two different hyperspectral imaging systems was evaluated for Goya's paintings, the first employing a "push-broom" system and the second a "mirror-scanning" system. The main pigments present were identified by imaging reflection spectroscopy [24].

Qualitative assessment of image quality

The research question within the present study was: to what extent do the operating principles and technologies of a range of spectral image capture systems affect the image quality achieved in the datasets?

Spectral image quality is influenced by a number of factors [25] and understanding their role and how different devices are used in the digital documentation workflow will help to define better procedures for acquisition and processing. Spectral image recording systems are first and foremost imaging devices, and so can be compared against the performance of more familiar digital cameras and scanners. The quality of the images acquired can be assessed by the well-established methods of visual scaling and psychophysics, with observers judging each image either in isolation or side-by-side with a reference image [26]. Various attributes of image quality can be distinguished and scaled, including tonal rendering, colour, sharpness, naturalness, freedom from defects, etc.

Image datasets from ten different systems were compared (Table 1). Four were pushbroom hyperspectral devices, three were multispectral systems based on liquid crystal tunable filter (LCTF), and three were Nikon cameras with external narrow-band transmission filters. The illumination system was different in each case. The institution in Spain carried out measurements with two different systems (D and J). Datasets from France, Italy, Poland and Cyprus were shared as file-cubes. All the others were provided as sets of single TIFF images or as Matlab data files. Because some systems are commercially available while others are custom-built and upgradable and flexible, it is difficult to compare them directly by specification. Nevertheless, the available information about the type of system, working bandwidth and spatial resolution is summarised in Table 1. The spatial resolution is calculated as the number of pixels across the width

Table 1 Participating institutions and characteristics of their systems

Image	Country	System	Range (nm)	# Bands	Working bandwidth (nm)	Illumination source	System	Camera information	Spatial resolution (pixels/mm)
A	Norway	LCTF	400–720	33	10	Fluorescent lamps	Custom built	DVC-16000 M monochrome 4872 × 3248	6.0
B	France	Pushbroom	414–994	160	3.5	Tungsten halogen 150 W fibre optic	Hyspex VNIR1600	SSI CCD 1600 × 1200	8.1
C	Cyprus	Filters	420–1000	30	20	Two 50 W halogen lights (4700 K)	MuSIS HS	CCD and photocathode 1600 × 1200	5.2
D	Spain	Pushbroom	400–1000	61	10	D65—SpectraLight III luminaire	Custom built	Monochrome CCD 1392 × 1040	12.0
E	Poland	Pushbroom	394–1009	776	<1	OSRAM Decostar WFL 12 V 35 W	Specim V10E	Monochrome sCMOS linear 2184 pixels	5.5
F	Italy	Pushbroom	398–961	448	<1	Two fibre-optic Schott 150 W	Custom built	Monochrome CCD 1344 × 1024	2.2
G	Portugal	LCTF	400–720	33	10	Metal Halide—Sylvania 150 W	Custom built	Monochrome CCD 1024 × 1344	4.1
H	UK	Filters	400–700	16	20	Photolux 150 W tungsten halogen	Custom built	Nikon D200 RGB CCD 3900 × 2600	10.4
J	Spain	LCTF	400–720	33	10	VeriVide Daylight Fluorescent	Custom built	Monochrome CCD 1392 × 1040	4.9
K	Switzerland	Filters	420–660	13	20	Two Broncolour flashlights	Custom built	Nikon D3 RGB CMOS 4256 × 2832	10.2

MacDonald *et al. Herit Sci* (2017) 5:41

Page 5 of 16

Fig. 2 RGB images of the Russian icon reconstructed from datasets of Norway (**A**) and France (**B**). The visual assessment always used the same image on the left as reference in a pair comparison technique

of the icon in the digitised image, divided by its physical width of 222 mm.

Colour images were initially reconstructed from each dataset using the ENVI 4.7 software. The three wavebands nearest to 605, 530 and 450 nm were chosen for the red, green and blue (RGB) channels respectively to produce a 'false colour' image. Even though the responsivities of these channels are dependent on their bandwidths and peak wavelength, and are not equivalent to the spectral sensitivities of the eye's photoreceptors nor to the primaries of the sRGB standard display [27], they offered a convenient means for qualitative comparison between the disparate imaging systems. Given sufficient time and resources, alternative methods would have been: (a) to make a weighted sum of the spectral bands to approximate the display RGB primaries; or (b) to reconstruct the visible reflectance spectrum for each pixel and then to calculate the sRGB image via the CIE tristimulus values (see Fig. 9 below).

A group of ten observers, five male and five female, with ages ranging from 20 to 50, all with normal colour vision, was chosen. They were asked to evaluate the quality of each of the ten reconstructed RGB images using a five-step psychometric (Likert-type) scale from 1 (completely disagree) to 5 (completely agree), for each of the following parameters: uniformity of illumination, contrast, sharpness, geometric distortion, visibility of fine spatial details, and colour fidelity. The participants were also asked whether they saw any defects in each image, and

to evaluate the overall quality of the images on a five-step scale from 1 (very bad) to 5 (very good). Finally, they were asked to rank the ten images from most preferred to least preferred. As a reference image, the reconstructed RGB image from the dataset from Norway was chosen, as it was considered to lie somewhere between the best and worst images of the whole set. When shown to the observers, each of the other nine test images was displayed alongside the reference image (Fig. 2). It would have been preferable to have judged the images against the actual icon, viewed under controlled illumination, but at the time of the experiment it was elsewhere being digitised.

The diversity of imaging technologies from the different institutions created considerably different visual results, with varying degrees of pleasantness (Figs. 2, 3, 4). The acceptability of the icon images was judged individually by ten observers, viewing them on an uncalibrated LCD display screen of diagonal 27 inches from a distance of approximately 1 m. The display white point was set to D65 with a luminance of 180 cd/m^2, and judgements were made under typical office illumination (combination of overhead fluorescent lights and daylight from west-facing windows). Each of the test images was compared side-by-side with the same reference image (Fig. 2A). It was not always a simple task to assess the differences between the images because in some cases they appeared dramatically different. Uneven illumination in some systems had caused shading effects and over-exposed areas due to specularity. Some images had been cropped to

MacDonald *et al. Herit Sci* (2017) 5:41

Page 6 of 16

Fig. 3 RGB images of the Russian icon reconstructed from the datasets of: Cyprus (**C**), Spain (**D**), Poland (**E**) and Italy (**F**)

different sizes and/or geometrically distorted, and varying resolution affected the definition of fine details.

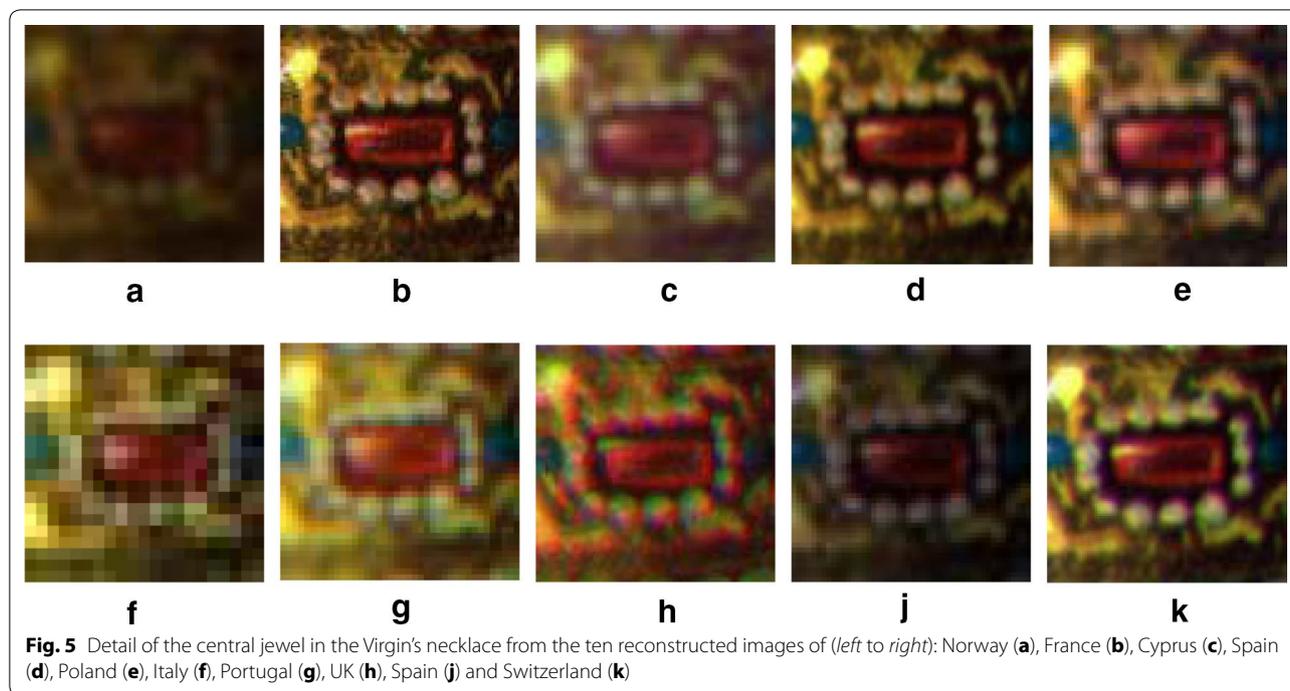The following results emerged from the observer judgements:

- A, C, H and J were considered the worst images in terms of uniformity of illumination, while B, F and G were considered the best;

- B and K were considered the sharpest images, while J did not enable discrimination of fine details;
- D and J were geometrically distorted (both skew and cropping);
- F and G were considered the worst images in terms of colour reproduction;
- B and K were the most preferred images, while J was the least acceptable.

MacDonald *et al. Herit Sci* (2017) 5:41

Page 7 of 16

Fig. 4 RGB images of the Russian icon from the datasets of: Portugal (**G**), UK (**H**), Spain (**J**) and Switzerland (**K**)

The representation of image detail was assessed by enlarging a small square region of the image around the central jewel in the Virgin's necklace, corresponding to an area of 10 mm × 10 mm on the icon surface. Because of the differing spatial resolution of the systems (Table 1), these details ranged in size from 22 × 22 to 120 × 120 pixels, but to enable fair comparison all have been enlarged to the same size (Fig. 5). The variety of rendering of detail, as well as tone and colour, across the different systems is remarkable. It should be noted that the apparent sharpness of image detail depends not only on sampling resolution (pixels/mm) but also on the point

MacDonald *et al. Herit Sci (2017) 5:41*

Page 8 of 16

Fig. 5 Detail of the central jewel in the Virgin's necklace from the ten reconstructed images of (*left* to *right*): Norway (**a**), France (**b**), Cyprus (**c**), Spain (**d**), Poland (**e**), Italy (**f**), Portugal (**g**), UK (**h**), Spain (**j**) and Switzerland (**k**)

spread function (focus) of the optics and registration of the images in the three colour channels.

Spectral reconstruction

A quantitative analysis was performed for the three hyperspectral datasets from France, Italy, and Poland, which were compared with the multispectral data from UK. Six points on the icon were selected (Fig. 6) and the data vector was extracted from the corresponding pixels in each channel of each image using ENVI 4.7 software.

The spectral reflectance distribution of the object at each point should ideally be independent of the characteristics of the imaging system and illumination and observer. To provide the 'ground truth' for reference, spot measurements were made at the six positions shown in Fig. 6 using a hand-held X-rite i1Pro spectrophotometer. This instrument is designed for quality control of glossy prints in the graphic arts, so has 45/0 geometry, i.e. it illuminates the surface at an angle of 45° and senses the light reflected perpendicular to the surface, in order to avoid specular reflections. The light is averaged over the area of the surface covered by a circular aperture of 4.5 mm diameter. The reflectance factor is reported at 10 nm intervals over the range 380–730 nm, with results shown in Fig. 7. The spectra for all colours are quite broad and smoothly changing. The gold was also measured, and was surprisingly low in reflectance factor because its lustre was excluded by the measurement geometry. It was not possible to measure the white

or turquoise paints because nowhere on the surface was there an area large enough to span the instrument's aperture.

The reflectance spectrum at each location of multispectral dataset H was estimated by averaging a square region of 30 × 30 pixels in each of the 16 spectral channels, processing the image data in Matlab. As the image resolution was approximately 10 pixels/mm, this corresponds to a region of 3 × 3 mm on the icon surface. The raw images from the camera (in NEF format) were converted to 3 × 16-bit RGB TIFF files by the utility DCRAW, to ensure tonal linearity of the data. Corresponding values were obtained for both the icon and a matte grey card used as the reference, and the reflectance factor R at each wavelength was calculated as:

$$R(\lambda_k) = G(\lambda_k)(i(\lambda_k) - b)/(g(\lambda_k) - b)$$

where filter k (range 1–16) corresponds to peak wavelength $\lambda_k = 400, 420, 440 \ldots 700$ nm; G is the reflectance factor of the grey card, measured by the spectrophotometer; i is the mean pixel value of the icon image in the sample area; g is the mean pixel value of the grey card image in the sample area; b is the mean pixel value of the black image in the sample area.

All images were taken in a dark room to minimise the ambient light. The grey card was placed in the plane of the top surface of the icon, covering the same area, and served to correct both non-uniformity of illumination across the surface of the object and vignetting of the lens.

MacDonald *et al. Herit Sci (2017) 5:41*

Page 9 of 16
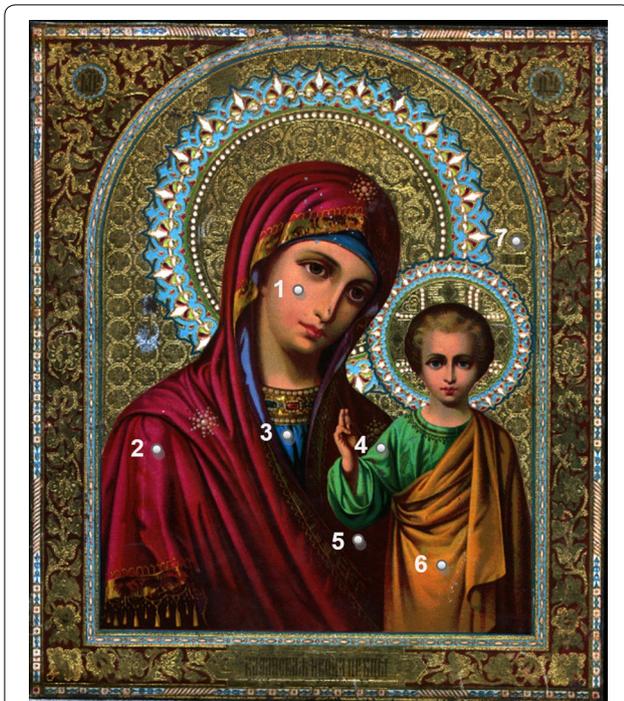
Fig. 6 Reconstructed RGB image of the Russian icon (from dataset of France), showing the different locations from which spectral reflectance was sampled: *1* skin; *2* red robe; *3* blue vest; *4* green shirt; *5* brown shadow; *6* yellow robe; *7* gold
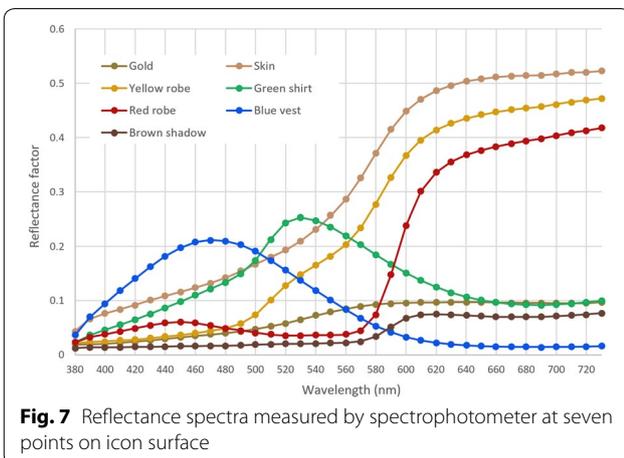
Fig. 7 Reflectance spectra measured by spectrophotometer at seven points on icon surface

The black image was taken with the grey card in place but the copystand lights off, and was an average of ten successive image frames, to reduce the effects of noise. The blue camera channel was used for *k* in the range 1–6 (filter wavelengths 400–500 nm), the green channel for *k* in the range 7–10 (520–580 nm), and the red channel for *k* in the range 11–16 (600–700 nm). The end values at 400 and 700 nm were unreliable because of the very low

signal level through the filters. The resulting reflectance factors are plotted in Fig. 8 as a function of wavelength against the reference measurements from the spectrophotometer for three sample points.

The same calculation was applied to every pixel in the image and the resulting reflectance factors interpolated to 5 nm intervals over the range 400–700 nm. Tristimulus values were calculated by multiplying the reflectance factor at each pixel by the spectral power distribution of the D65 illuminant and the responsivity functions of the CIE Standard Observer. The resulting image was converted to the sRGB display colour space [27], giving the result in Fig. 9.

Comparison with the conventional RGB photograph from the same Nikon D200 camera (Fig. 1) shows that the computed colorimetric image is lower in contrast and colour saturation, but is a good match in hue. Comparison with the 'rough and ready' RGB image synthesised from three bands of the multispectral image set (Fig. 4H) shows a great improvement in overall colour balance, but a loss of sharpness because of uncorrected chromatic aberration.

The three hyperspectral systems from France (B), Poland (E) and Italy (F) all used a Spectralon tile as the white reference. Comparison of their reflectance spectra reconstructed at six sample points of the images (Fig. 10), reveals that they are in general agreement, with similar shapes of the curves and inflections at the same wavelengths. The differences in amplitude are indicative of how each individual imaging geometry reacted differently to the non-Lambertian surface of the icon.

Analysis of the spectra in Fig. 10 enabled a quantitative assessment, by comparing the reconstructed spectra of the four systems with the reference spectrophotometer data. Using the Matlab function interp1, all spectra were interpolated to 1 nm intervals in the range of the visible spectrum from 400 to 700 nm, and differences were calculated between corresponding points in terms of both root-mean-square error (RMSE) and colour difference (ΔE_{ab}^*). The latter is the Euclidean distance between two stimuli expressed in the CIE $L*a*b*$ colour space, and is scaled so that one unit of ΔE_{ab}^* corresponds approximately to one just-noticeable difference (JND) [28]. For colour photography and print reproduction, values of ΔE_{ab}^* less than 5 would generally be considered acceptable [29]. The results in Table 2 show the target colour in CIELAB coordinates (referred to the D65 illuminant) of each of the seven sample points in Fig. 6 and the corresponding differences for the four systems. In most cases the RMSE errors were less than 0.02 and colour errors were less than 5 ΔE_{ab}^* although larger values occurred for the red robe and gold sample points.

MacDonald *et al. Herit Sci* (2017) 5:41

Page 10 of 16

Fig. 8 Reconstructed vs measured reflectance spectra for three sample points: (*left*) green shirt; (*centre*) yellow robe; (*right*) red robe. In the legend 'eye one' means the X-rite i1Pro spectrophotometer, used to measure surface reflectance factor
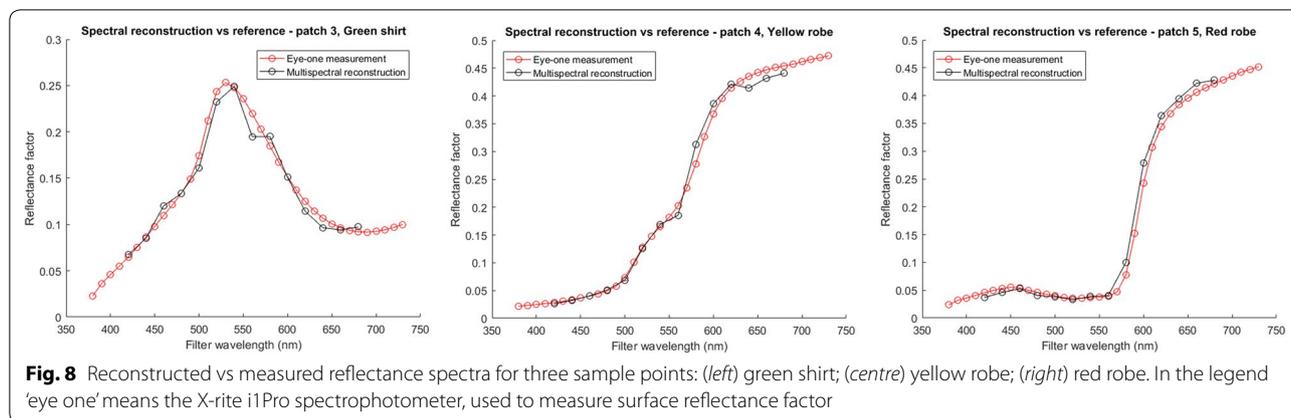
The results agree with analysis of two hyperspectral systems for another of the RRT test objects, the X-rite colour checker chart [30]. There it was found that in CIE $L^*a^*b^*$ coordinates the patches presenting larger colour differences were magenta, blue and red (ΔE^*_{ab} values of 3.7, 3.8 and 5.1 respectively), which lie at the ends of the visible range where the reflectance spectra also exhibited greater variations. On the other hand, the green and yellow patches were more similar between the two systems (ΔE^*_{ab} values of 3.3 and 2.5).

One notable difference between the hyperspectral datasets is the amount of noise present in the signal, especially at the ends of the range. As an example, Fig. 11 shows an enlarged section of the reflectance spectra from the three systems plotted in Fig. 10 for the skin sample between 400 and 450 nm. There is evidently much more 'jitter' in system E than in the other two, especially at the shortest wavelengths, because there is some inherently greater perturbation from one wavelength to the next. The sensitivity of silicon sensors at around 400 nm is very low, resulting in visible image noise. The difference in noise levels is a function of the camera electronics, physical pixel size, whether an equalisation filter has been used and whether noise reduction processing (such as averaging) has been carried out.

Significant differences between the systems were found for the gold area sampled at the location shown in Fig. 6. Although the reflectance spectra all have the same shape, with the characteristic upward slope to a peak at around 610 nm, they vary in amplitude by factors of as much as 3×, both greater and less than the reference measurement of the spectrophotometer (Fig. 12), which leads to the large errors in Table 2. This arises from both the granular nature of the surface and the breadth of the specular peak that gives gold its characteristic lustre [31], causing the image intensity

Fig. 9 Colorimetric image reconstructed from dataset H under D65 illuminant

detected by each system to be critically dependent on its illumination geometry.

An interactive visualisation of high-resolution hyperspectral data from the Russian icon has been developed and can be viewed online [32]. This facility uses the IIPImage software suite [33] to provide interactive online access to a subset of the hyperspectral data. It features an image stack including a colorimetric CIE $L^*a^*b^*$ rendering of dataset B under a D65 illuminant and single frames at various specific wavelengths, which the user can select and smoothly blend between.

MacDonald *et al. Herit Sci* (2017) 5:41

Page 11 of 16
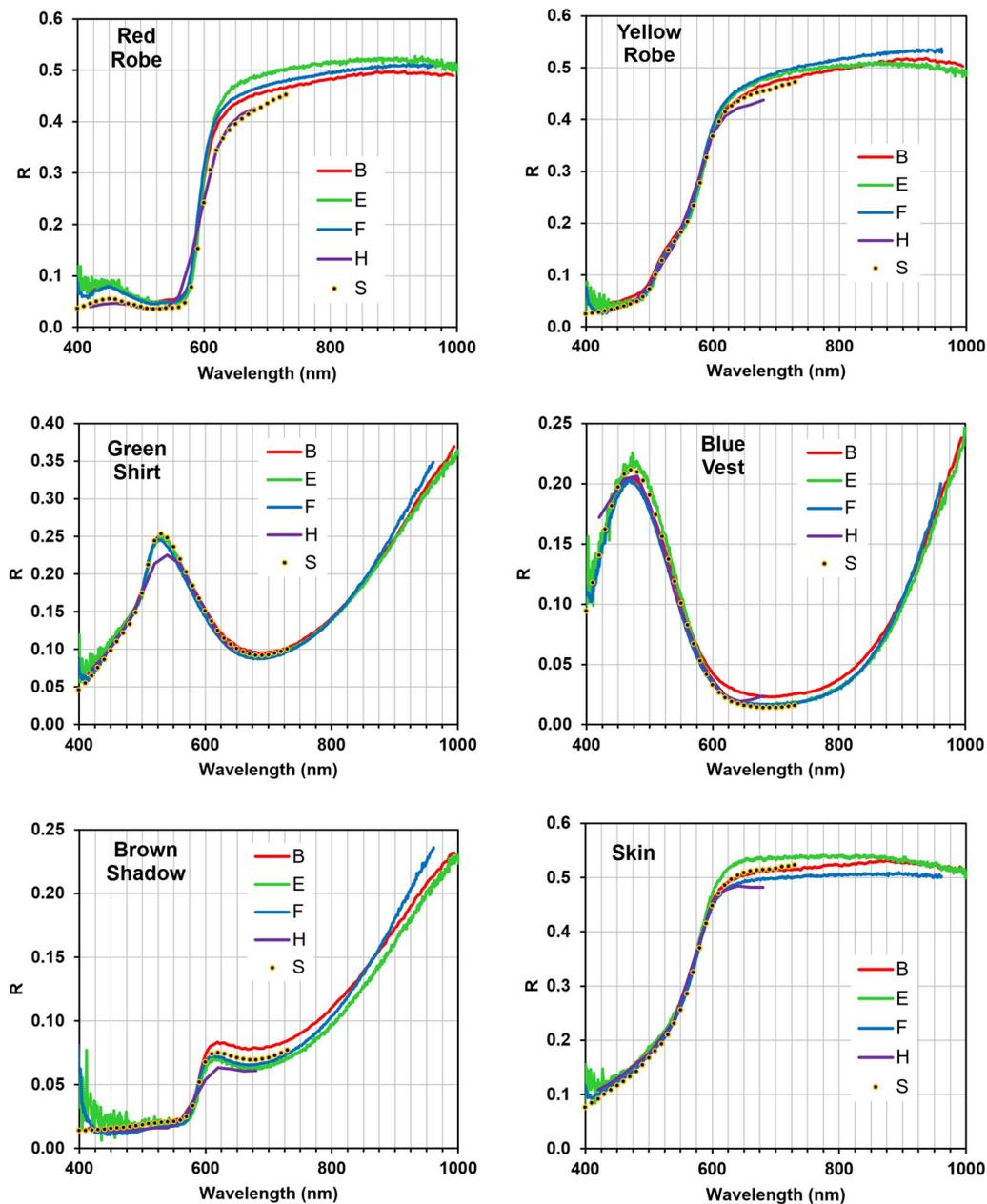
Fig. 10 Reflectance spectra from six different locations of the Russian icon, Reconstructed from the datasets of: France (*B*), Poland (*E*), Italy (*F*) and UK (*H*), with the reference data of a spectrophotometer (*S*)
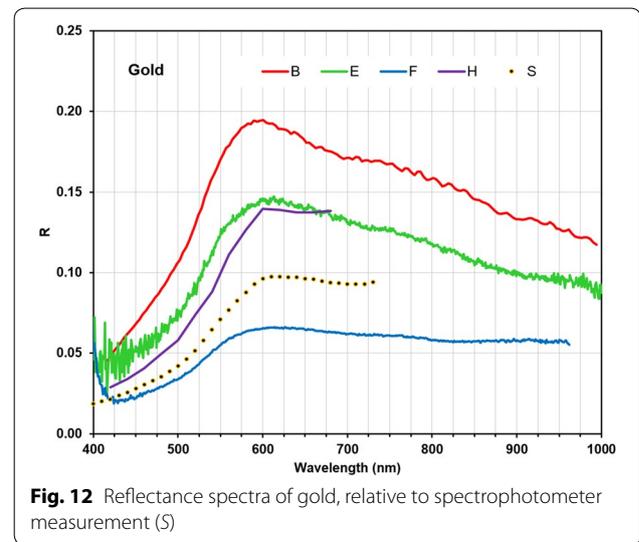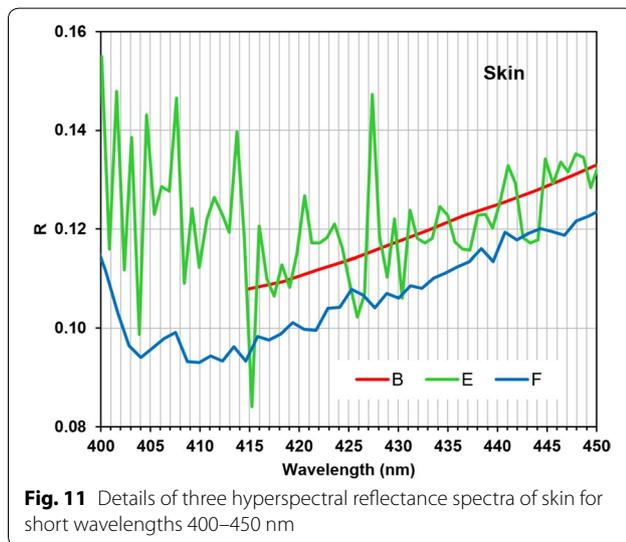
Discussion

The purpose of this study was to compare the datasets acquired by different imaging systems, in order to assess the impact of the setup and workflow of each device on the results obtained. In particular, we investigated the capabilities of ten multispectral and hyperspectral devices to digitise a polychrome nineteenth-century

Russian icon with a glossy surface, high specularity and very fine spatial details. We found that the systems and workflows employed by the participating institutions varied widely, including cameras, light sources, imaging geometry and file formats. The cost of the different devices used in the RRT ranged from a few thousand euros, such as for the filter-based photographic systems

MacDonald *et al. Herit Sci (2017) 5:41*

Page 12 of 16

Table 2 Colour coordinates of the sample points and analysis of errors between reconstruction and spectrophotometer reference data

	Red-robe	Yellow-robe	Green-shirt	Blue-vest	Brown-shadow	Skin	Gold
*L**	37.34	54.21	51.53	37.13	21.24	61.67	33.50
*a**	46.02	22.72	−22.68	−20.90	17.72	19.11	3.46
*b**	16.21	51.21	20.92	−23.45	13.86	34.30	20.44
France							
RMSE	0.0376	0.0108	0.0073	0.0108	0.0051	0.0151	0.0742
ΔE_{ab}^*	4.53	4.79	3.13	3.63	2.12	3.12	14.91
Italy							
RMSE	0.0457	0.0160	0.0077	0.0073	0.0068	0.0103	0.0238
ΔE_{ab}^*	5.56	1.90	2.86	2.62	2.79	1.84	6.82
Poland							
RMSE	0.0548	0.0142	0.0094	0.0059	0.0085	0.0213	0.0370
ΔE_{ab}^*	5.88	3.59	3.60	1.51	6.33	2.19	7.71
UK							
RMSE	0.0236	0.0110	0.0109	0.0154	0.0068	0.0185	0.0287
ΔE_{ab}^*	11.05	0.98	4.64	4.13	1.93	2.92	7.41

Fig. 11 Details of three hyperspectral reflectance spectra of skin for short wavelengths 400–450 nm

Fig. 12 Reflectance spectra of gold, relative to spectrophotometer measurement (*S*)

used by Institutions H and K (Table 1), up to 50 thousand euros for some of the specialised hyperspectral imaging systems. All these factors were largely responsible for the wide variations in image quality that were observed. One of the difficulties encountered was the insufficient or, in some cases the complete lack of, written documentation of the procedures employed to produce the image files. Metadata is directly associated with most of the steps in an imaging system workflow and makes possible the identification, management, access, use, and preservation of the digital images. However, it can be time-consuming to produce and is often neglected.

The reflectance factor recorded at each point of the object surface may vary for each instrument or sensor according to its calibration and imaging geometry, i.e. the relative angles of the illumination and optical axis with respect to the tangent plane at the surface. The imaging geometry can have a dramatic effect on glossy surfaces, and for best results it is essential to have some flexibility in the positioning of the light source(s) and object relative to the camera, thereby enabling the operator to exercise judgement in the setup. While this is easily achieved by traditional means in a photographic studio (Fig. 13), it requires careful opto-mechanical design for integrated scanning devices.
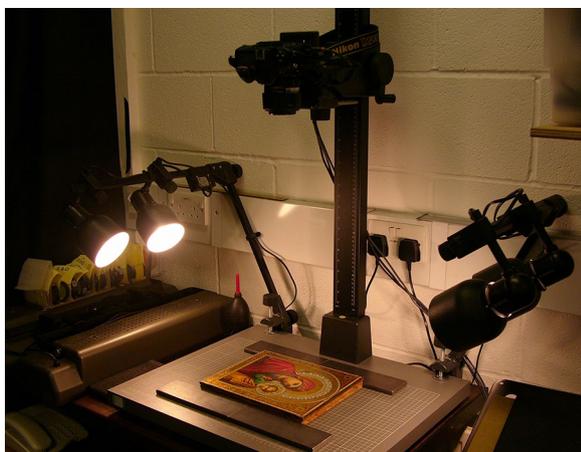
Fig. 13 Multispectral imaging of the Russian icon on a photographic copystand. The Nikon camera is fixed vertically above the object, with four tungsten-halogen lamps arranged symmetrically on either side, at a low angle of elevation to minimise specular reflection

Fig. 14 Hyperspectral imaging of the Russian icon on a system designed by IFAC-CNR in Italy. The assembly of camera and dual lighting moves vertically in front of the stationary object

The hyperspectral imaging instrument at IFAC-CNR is a good example of state-of-the-art spectral imaging with illumination geometry equivalent to that of a spectrophotometer [34]. The system is based on a prism line-spectrograph, connected to a high sensitivity CCD camera. The line-segment is illuminated by two fibre-optic line-lights equipped with focusing lenses fixed to the scan-head, which project their beams symmetrically at angles of 45° with respect to the normal direction of the imaged surface (0°/2 × 45° observation/illumination geometry). The distance of the lens from the object surface is approximately 20 cm. The assembly of lights and spectral head move together during the scanning in order to illuminate only the small rectangular area under acquisition (Fig. 14). The combination of high scan rate and small area of illumination ensures that the overall light exposure levels are compatible with the recommended limits in museums.

During the RRT the Russian icon circulated among 21 institutions around Europe with the other COSCH test objects. The working conditions and technical parameters of the acquisition process were not specified. Each institution was asked to image the icon with the setup and methods routinely applied in its own laboratory, including everything from the physical support of the object to the setup of the camera and illumination to the processing of the image. Each group was also asked to provide, together with the image data, a description of the system, setup and procedure used for the acquisition. What was provided in many cases, however, was incomplete metadata, with limited or no information regarding the illumination, calibration, normalisation, and processing details, which made it difficult to interpret the image
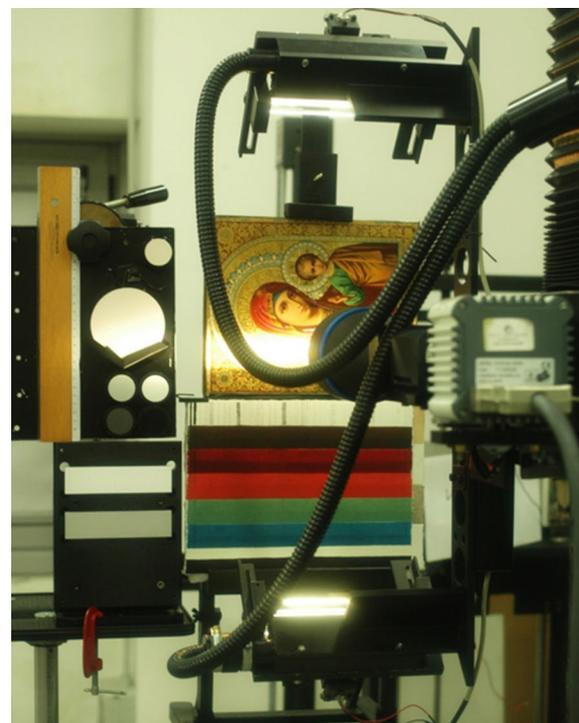
data. In general it needed to be explained more clearly by each provider of a spectral dataset what corrections, if any, had already been applied to the image data.

This study showed that the data obtained from spectral imaging systems is critically dependent on methodological procedures and that it can vary considerably across laboratories implementing different practices. Some harmonisation is therefore desirable to allow proper comparison, adequate use and efficient communication. In every case, image data should be complemented with detailed metadata to enable correct application by users. Both 'raw' and 'corrected' image data should be provided for each object, the latter compensated for degradations related to the image acquisition process, e.g. spatial and spectral calibration, defocus, noise, among others. Specifications regarding the light source are possibly the most important. Its spectral power distribution and spatial variation should be measured and recorded and taken into account in data processing.

The correction of spatial non-uniformity of illumination should be implemented and/or specified using images of a uniform card (normally but not necessarily white) for each spectral band. The noise level should also be estimated by imaging with the system using the same physical setup but optically occluded. Some

MacDonald *et al. Herit Sci* (2017) 5:41

Page 14 of 16

systems have a significant degree of scattered light and its magnitude should be estimated. Chromatic aberrations should be avoided as much as possible as they introduce wavelength-dependent image distortions that are difficult to correct by post-processing, i.e. variable misregistration and defocusing in different spectral bands. Using apochromatic optics will minimize the associated problems. The effects of the non-uniform spectral sensitivity of the sensor and its impact on signal-to-noise ratio should be taken into account. Data from the camera manufacturer or in-house calibration can be used to compensate for this effect. If the system employs different exposure times for different wavelengths this should be specified. We suggest that the image processing procedures should include 'sanity checks' at key points along the processing chain, for example to check on the tonal linearity of the data by using a neutral scale with multiple shades of grey from black to white, and to compare the reconstructed reflectance spectra against spectrophotometer reference measurements, including colorimetric measures.

Conclusions

The ability to digitise and reproduce images of the Russian icon as a truthful representation of the real object proved to be difficult, and significantly different results were obtained from the ten systems and workflows tested. The acceptability of the visual quality of the reproductions was judged by a panel of observers. Obviously, this was a qualitative and subjective assessment, and the difficulty of identifying which image is more realistic limits the value of the results for purposes of examination and documentation. This is particularly true when dealing with coloured artworks or pigmented materials that need to be reproduced with high accuracy.

It must be noted that perceived colour is not simply a property of coloured objects, which could be regarded more as modifiers of the incident light than as themselves the sources of colour. Colour is dependent not only on the reflectance of the object, but also on the properties of the light striking its surface, and the visual sensitivity and adaptation of the observer, all of which are functions of wavelength. Consequently, the retention of descriptive data about the conditions used during acquisition and subsequent data processing should be of great importance. In our case, however, the recording of metadata was generally found to have been neglected, which made the assessment and comparison of some datasets very difficult or even impossible.

The results of the RRT showed a considerable variability in the image data, caused by the diversity of devices (multi- and hyper-spectral systems), their measurement geometries, the methods of data processing, the personnel operating the imaging systems, and the guiding purpose for application of these devices by the different groups (colour reproduction, documentation, identification of materials, etc.). All these factors resulted in varying levels of quality in the final spectral datasets.

The obtained results emphasise the necessity of defining standardised methodologies and best practices for different imaging systems. Specifically, we draw attention to the importance of calibration (spectral, radiometric and spatial) procedures and knowledge of the intrinsic accuracy, precision and limits of the technologies. Furthermore, the RRT made clear the need for regular calibration, validation and testing of every system.

Given the difficulties and ambiguities encountered in this study, it is clear that considerable further refinement of systems and standardisation will be required before multispectral and hyperspectral imaging can be widely adopted within the heritage science community. In the development of standards, the criteria and technical requirements for spectral imaging systems might differ, depending on whether digital documentation or scientific analysis of materials (or both) are considered. It would be advantageous also to consider developments relating to colour management in the graphic arts and media industries, in particular the new iccMAX specification that is currently being adopted as an international standard [35]. This defines a spectral profile together with a spectral profile connection space (PCS), based upon spectral representation of colour rather than colorimetry, as the basis of encoding surface reflectance and hence of colour communication.

In the COSCH RRT, each institution was asked to follow its own practices and to produce a good quality result. Yet the datasets exhibited a wide range of image quality attributes. Much has been made in the discussion above about the need for a standardised procedure, but the fact is that today there exists no such standard. In order to achieve a standard whereby different institutions could achieve consistent results with different optomechanical systems, the imaging geometry, in particular, would have to be closely specified and controlled, analogous to the standards for instrumental spot colour measurement [36].

A new emerging standard, ISO 19263-1, intended for imaging of two-dimensional originals by cultural heritage institutions, specifies a method for analysing imaging systems where it is important to control the degree of accuracy and to ensure that imaging quality is maintained over time [37]. Three aspects of imaging system performance are covered: (a) evaluation—for system benchmarking; (b) optimization—for tailoring the system to a particular job (use case); and (c) monitoring—for

MacDonald *et al. Herit Sci* (2017) 5:41

Page 15 of 16

controlling the quality of the system to remain within specified limits. ISO 19263-1 can be used to establish and maintain image quality in digitisation workflows, and to develop a digitisation strategy including assessment of collections and system selection. Such standards will be essential for generating meaningful digital datasets that can be interpreted and repurposed for many diverse applications.

Author details
[1] Faculty of Engineering Sciences, University College London, London, UK. [2] Faculty of Science and Technology, Universidade NOVA de Lisboa, Lisbon, Portugal. [3] IFAC-CNR, Via Madonna del Piano 10, Sesto Fiorentino, Italy. [4] C2RMF, Musée du Louvre, Paris, France. [5] National Museum in Kraków, Al. 3 Maja 1, 30-062 Kraków, Poland. [6] Department of Physics, University of Minho, Campus de Gualtar, Braga, Portugal.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Saunders D. High-quality Imaging at the National Gallery: origins, implementation and applications. Comput Humanit. 1998;31:153–67. doi:10.1023/A:1000696330444.
2. Bianchi C. Making online monuments more accessible through interface design. In: MacDonald LW, editor. Digital heritage—applying digital imaging to cultural heritage. Oxford: Butterworth-Heinemann; 2006. p. 445–66.
3. Boochs F, Bentkowska-Kafel A, Degrigny C, Hauta-Kasari M, Rizvic S, Trémeau A. Towards optimal spectral and spatial documentation of cultural heritage: COSCH—an interdisciplinary action in the COST framework. Proc XXIV Int CIPA Symp. 2013. doi:10.5194/isprsarchives-XL-5-W2-109-2013.
4. Gershkovich E. Mstera's icon painters. In: Agitlak Lacquer Propaganda. Moscow: Gamma-Press; 2009. http://www.agitlak.com/trading_mstiora.html.
5. Fischer C, Kakoulli I. Multispectral and hyperspectral imaging technologies in conservation: current research and potential application. Rev Conserv. 2006;7:3–16.
6. Cucci C, Casini A, Picollo M, Stefani L. Extending hyper-spectral imaging from Vis to NIR spectral regions: a novel scanner for the in-depth analysis of polychrome surfaces. Proc SPIE Conf Opt Arts Archit Archaeol (O3A) IV. 2013;8790:1–9. doi:10.1117/12.2020286.
7. Dyer J, Verri G, Cupitt J. Multispectral imaging in reflectance and photo-induced luminescence modes: a user manual. The British Museum, Charisma Project; 2013. p. 81–5. http://www.britishmuseum.org/pdf/charisma-multispectral-imaging-manual-2013.pdf.
8. Liang H. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. Appl Phys A. 2012;106:309–23. doi:10.1007/s00339-011-6689-1.
9. Cucci C, Casini A, Picollo M, Poggesi M, Stefani L. Open issues in hyperspectral imaging for diagnostics on paintings: when high spectral and spatial resolution turns into data redundancy. Proc SPIE Conf Opt Arts Archit Archaeol (O3A) III. 2011;8084:1–10. doi:10.1117/12.889460.
10. Goetz AF. Three decades of hyperspectral remote sensing of the Earth: a personal view. Remote Sens Environ. 2009;30(113):S5–16.
11. Kudenov MW, Roy SG, Pantalone B, Maione B. Ultraspectral imaging and the snapshot advantage. InSPIE Defense+ Security. Kent: International Society for Optics and Photonics; 2015. p. 94671X.
12. Hagen N, Kester RT, Gao L, Tkaczyk TS. Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems. Opt Eng. 2012;51(11):111702-1. doi:10.1117/1.OE.51.11.111702.
13. Berns RS, Frey FS. Direct digital capture of cultural heritage—benchmarking American Museum Practices and defining future needs, final report 2005. Rochester: Rochester Institute of Technology; 2005.
14. Jung A, Götze C, Gläser C. Overview of experimental setups in spectroscopic laboratory measurements—the SpecTour Project. Photogrammetrie-Fernerkundung-Geoinformation. 2012;2012(4):433–42.
15. Saunders D, Cupitt J. Image Processing at the National Gallery: the VASARI Project. Natl Gallery Tech Bull. 1993;14:72–85.
16. Martinez K, Cupitt J, Saunders D, Pillay R. Ten years of art imaging research. Proc IEEE. 2002;90(1):28–41. doi:10.1109/5.982403.
17. Ribés A, Schmitt F, Pillay R, Lahanier C. Calibration and spectral reconstruction for CRISATEL: an art painting multispectral acquisition system. J Imaging Sci Technol. 2005;49(6):563–73.
18. Kubik M. Hyperspectral Imaging: a new technique for the non-invasive study of art-works. In: Creagh D, Bradley D, editors. Physical techniques in the study of art, archaeology and cultural heritage. Oxford: Elsevier; 2007. p. 199–259.
19. Casini A, Bacci M, Cucci C, Lotti F, Porcinai S, Picollo M, Radicati B, Poggesi M, Stefani L. Fiber optic reflectance spectroscopy and hyperspectral image spectroscopy: two integrated techniques for the study of the Madonna dei Fusi. Proc SPIE Conf Opt Methods Arts Archaeol. 2005;5857:1–8. doi:10.1117/12.611500.

MacDonald *et al. Herit Sci (2017) 5:41*

Page 16 of 16

20. Cucci C, Delaney JK, Picollo M. Reflectance hyperspectral imaging for investigation of Works of Art: old master paintings and illuminated manuscripts. Acc Chem Res. 2016;49:2070–9. doi:10.1021/acs.accounts.6b00048.

21. Delaney JK, Zeibel JG, Thoury M, Littleton R, Palmer M, Morales KM, Rie ER, Hoenigswald A. Visible and infrared imaging spectroscopy of Picasso's Harlequin musician: mapping and identification of artist materials in situ. Appl Spectrosc. 2010;64(6):584–94. doi:10.1366/000370210791414443.

22. MacDonald L, Giacometti A, Campagnolo A, Robson S, Weyrich T, Terras M, Gibson A. Multispectral imaging of degraded parchment. In: Tominaga S, Schettini R, Trémeau A, editors. Computational color imaging. Berlin: Springer; 2013. p. 143–57. doi:10.1007/978-3-642-36700-7_12.

23. Mounier A, Daniel F. Hyperspectral imaging for the study of two thirteenth-century Italian miniatures from the Marcadé collection, Treasury of the Saint-Andre Cathedral in Bordeaux, France. Stud Conserv. 2015;60(Suppl 1):S200–9.

24. Daniel F, Mounier A, Pérez-Arantegui J, Pardos C, Prieto-Taboada N, de Vallejuelo SFO, Castro K. Hyperspectral imaging applied to the analysis of Goya paintings in the Museum of Zaragoza (Spain). Microchem J. 2016;126:113–20. doi:10.1016/j.microc.2015.11.044.

25. Shrestha R, Pillay R, George S, Hardeberg JY. Quality evaluation in spectral imaging—quality factors and metrics. J Int Colour Assoc. 2014;12:22–35.

26. MacDonald LW, Jacobson R. Assessing image quality. In: MacDonald LW, editor. Digital heritage—applying digital imaging to cultural heritage. Oxford: Elsevier; 2006. p. 351–73.

27. Anderson M, Motta R, Chandrasekar S, Stokes M. Proposal for a standard default color space for the internet—sRGB. Proceedings Color and imaging conference. Springfield: Society for Imaging Science and Technology; 1996. p. 238–45.

28. Hunt RWG, Pointer MR. Measuring colour. 4th ed. Chichester: Wiley; 2011. p. 59–60.

29. Song T, Luo R. Testing color-difference formulae on complex images using a CRT monitor. Proceeding of Color and Imaging Conference (CIC). Springfield: Society for Imaging Science and Technology; 2000. p. 44–8.

30. Vitorino T, Casini A, Cucci C, Gebejesje A, Hiltunen J, Hauta-Kasari M, Picollo M, Stefani L. Accuracy in colour reproduction: using a Color-Checker chart to assess the usefulness and comparability of data acquired with two hyperspectral systems. In: Trémeau A, Schettini R, Tominaga S (eds) Proceeding of 5th international workshop on computational color imaging (CCIW). Heidelberg: Springer; 2015. LNCS 9016, p. 225–35. doi:10.1007/978-3-319-15979-9_21.

31. MacDonald LW. The Colour of Gold. Proceeding Conference of the International Colour Association (AIC). Tokyo: Color Science Association of Japan; 2015. p. 320–5.

32. http://merovingio.c2rmf.cnrs.fr/iipimage/iipmooviewer/COSCH_icon.html. Accessed 28 Aug 2017.

33. Pitzalis D, Pillay R, Lahanier C. A new concept in high resolution internet image browsing. Proceeding 10th International Conference on Electronic Publishing, Bansko, Bulgaria; 2006. http://iipimage.sf.net.

34. Vitorino T, Casini A, Cucci C, Melo MJ, Picollo M, Stefani L. Non-invasive identification of traditional red lake pigments in fourteenth to sixteenth centuries paintings through the use of hyperspectral imaging technique. Appl Phys A. 2015;121(3):891–901. doi:10.1007/s00339-015-9360-4.

35. iccMAX. Image technology colour management—extensions to architecture, profile format, and data structure. Specification ICC.2:2016-7. International Color Consortium; 2016. http://www.color.org/iccmax/index.xalter.

36. CIE 015:2004 Colorimetry, 3rd edn. Vienna: CIE.

37. ISO/TR 19263-1:2017 Photography—archiving systems—part 1: best practices for digital image capture of cultural heritage material. Geneva: ISO. https://www.iso.org/standard/64220.html.