

Spectral Imaging for Classification of Natural and Artificial Turquoise Samples

C. Gurschler, G. Serafino, G. Spöck, A. Del Bianco, M. Kraft and A. Kulcke

Carinthian Tech Research AG
Europastr.4, A-9524 St. Magdalen/Villach

Turquoises ($\text{CuAl}_6[(\text{OH})_2/\text{PO}_4]_4 \cdot 4\text{H}_2\text{O}$) are highly prized as gem stones, but easy to fake. Hence, a simple and reliable method to detect fraudulent jewellery would be of high interest. One possible solution is to use an imaging spectrometer system for the visible and near infrared. This method, in combination with a software toolbox for computer-aided classification, complements the expensive standard methods currently employed to investigate minerals.

A series of experiments was performed to develop a method to distinguish between true turquoises and (potential) fakes. The mineralogical collection of the Vienna Museum of Natural History (NHM Wien) provided referenced samples of (i) natural turquoise, (ii) treated turquoise, e.g. pressed turquoise powder, (iii) artificial, industrial turquoise (Gilson) and (iv) samples of other minerals that are frequently mistaken for turquoises. Spectrally resolved reflectance images of the mineralogical samples were acquired using a hyperspectral-imaging set-up for VIS and near-IR reflectance measurements, and subsequently classified using a spectral imaging toolbox developed in our company. [1]

Hyperspectral imaging is a novel technique developed to simultaneously visualise spectral and spatial properties of the sample surface. The imaging system (Fig. 2) uses a CCD or InGaAs array camera with an ImSpector (Specim Ltd.) optics system attached [2]. The samples are mounted on a motorised linear stage and illuminated by six high-power halogen lamps arranged in a hemispherical geometry around the stage central point. Other than with the older spectral imaging techniques, where a 2D-image is recorded and the spectral information is added by using e.g. filters, here one picture contains the full spectral information for every pixel along one line across the sample. This is achieved by projecting the light reflected from the sample through a slit aperture onto a prism-grating-prism optics system, which projects the spectral information for every pixel along the second axis of the photo detector array. By moving the sample along the axis perpendicular to the slit and taking a separate image for each spatial position, the entire sample surface is recorded. The single images combined to a so-called image cube, as depicted in Fig. 3, contain the entire hyper-spectral information about the sample. This cube is subsequently processed with statistical algorithms for spectral and/or spatial analysis.

The data can be analysed in two different ways. The first possibility is to generate classes in a supervised manner with a training set of data. The second possibility is an unsupervised classification of the data set. In this case only the number of classes has to be fixed and the algorithm generates classes, which have the most spectral differences. A typical diagram for a classification procedure is shown in Fig 4.



Fig. 1. Turquoise samples

SPECTRAL IMAGING LABORATORY SETUP

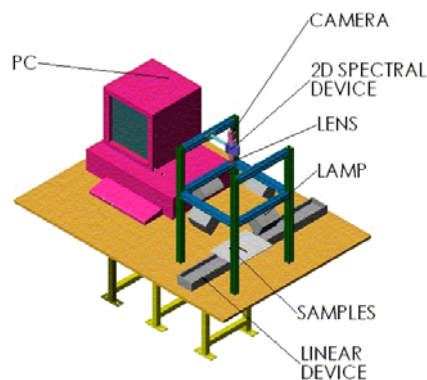


Fig. 2. Principle setup of an imaging spectrometer

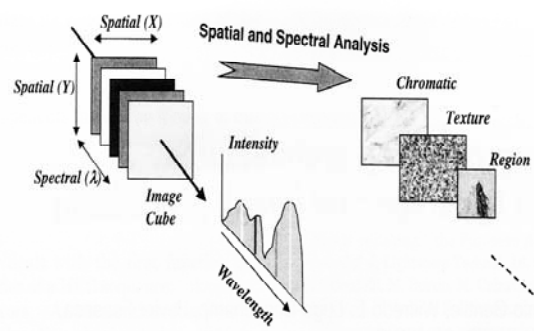


Fig. 3. Principle of an image cube, shown here for an alternative technique of image recording

Data analysis: Turquoise Classification

The IR and visible picture sequences were used for the classification of the turquoise samples. As no spectral information about the samples was available in advance, an unsupervised classification method was chosen. Since it was known from earlier experiments that fuzzy c-means clustering gives the most stable and spectroscopically most meaningful classification results, this technique was used for the data evaluation. The fuzzy c-means algorithm works by assigning class membership values to every single multivariate pixel in the image to be segmented. The membership values describe how strongly each pixel belongs to the different classes. The general fuzzy c-means algorithm assumes that for c initial spectra, the class representatives $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$ are given. Later this assumption will be changed by describing an algorithm that automatically selects these class representatives.

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be the spectra in the image to be classified. Fuzzy c-means clustering is then locally minimizing the cost function

$$J = \sum_{i=1}^n \sum_{j=1}^c U_{ij}^m \|\mathbf{x}_i - \mathbf{v}_j\|^2$$

where U_{ij} is the class membership value of spectrum \mathbf{x}_i to belong to the j^{th} class. The class memberships take values between 0 and 1. It is furthermore assumed that the class membership values of every spectrum sum up to 1. m is a real number larger than 1 and is a weighting exponent on each fuzzy membership. In this work $m=1.2$ has been used. If all the class representatives \mathbf{v}_j kept fixed, the cost function is minimized by selecting

$$U_{ij} = \frac{\|\mathbf{x}_i - \mathbf{v}_j\|^{-\frac{1}{1-m}}}{\sum_{j=1}^c \|\mathbf{x}_i - \mathbf{v}_j\|^{-\frac{1}{1-m}}} \quad \text{for } i=1, 2, \dots, n \text{ and } j=1, 2, \dots, c.$$

Fixing the class memberships U_{ij} , the cost function becomes minimal if

$$\mathbf{v}_j = \frac{\sum_{i=1}^n U_{ij}^m \mathbf{x}_i}{\sum_{i=1}^n U_{ij}^m} \quad \text{for } j=1, 2, \dots, c$$

got chosen. The idea of fuzzy c-means clustering is to use the two latter equations iteratively until convergence. The Segmentation of the hyperspectral image is performed by assigning each spectrum to the class where it has the highest class membership.

In this study a modification of the fuzzy c-means algorithm, fuzzy c-means clustering with automatic selection of class representatives, was used. A spectrum \mathbf{x}_i that does not belong very strongly to any of the classes $1, 2, \dots, c$, can be chosen as the class representative of a separate, new class.

Cosic et al. [3] propose to select a $(c+1)^{\text{th}}$ class representative \mathbf{v}_{c+1} as $\mathbf{v}_{c+1} = \mathbf{x}_p$, where

$$p = \arg \left(\min_i \left(\max_j U_{ij} \right) \right).$$

The fuzzy c-means algorithm is then restarted with the old c class representatives plus the new one.

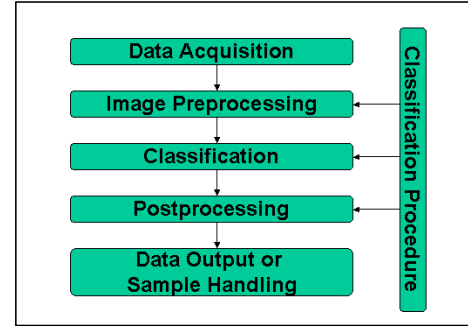


Fig. 4. Principle of image analysis and classification

Turquoise Classification Results

To get more initial information about the spectral properties of the data, a simple principal component transformation of the normalised or of the differentiated data can be applied. Using the principal component analysis (PCA) of the data and plotting the first 3 components of the new coordinate system in combination with an RGB colour scheme, a colour picture containing the spectral characteristics of the samples can be generated. This picture already shows, which samples have spectral similarities and which samples are spectrally highly different. Thus, these plots visualise the spectral characteristics of each sample. To get an impression of the distribution of the spectral properties, score density plots can be used. An example of such a plot for the first four principal components plotted against each other is displayed in Fig. 5.

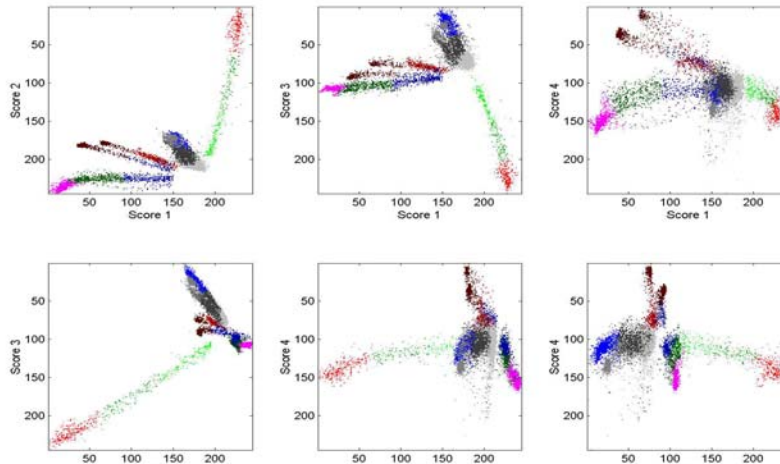


Fig. 5. Score density plot of the first principal components.

In advance of the classification, the datasets are subjected to a data pre-treatment procedure to get more stable results. A “dark” picture is first subtracted to reduce the dark current noise. In a second step, the data is normalised to standard reflectance using a white reflectance standard. In a third step, the spectra are normalised to unit area under the spectrum. This is a valid procedure, as only the spectral shape of the pixel is relevant, not the absolute intensity. This pre-treated data set is then subjected to the classification procedure. Alternatively, the first derivative spectra can be calculated and used for the classification.

As described above, the classification algorithms allow assigning each pixel to one of c clusters, depending on the similarity or dissimilarity of the spectra. The classification algorithm selects cluster centres and assigns to each spectrum membership values ranging from 0 to 1, describing how strongly the respective spectrum belongs to each of the different clusters. The respective spectrum is then assigned to the class where the membership value reaches a maximum.

Based on the number of samples and the available information about the similarities between the investigated samples, an initial value for the number of classes to be used for the unsupervised classification has to be fixed. In the present case, the procedure was started with 20 classes. During the classification process, the number of classes could be reduced to 11 relevant classes for the NIR measurements and 5 classes for the VIS measurements. After a visual inspection of the typical spectra, the number of significant classes in the NIR could be reduced even further, leaving six significant classes. The results of the classification are shown as 2D-images, where the different classes are assigned specific colours. In Fig. 6, the final classification results for the NIR-measurements are shown.

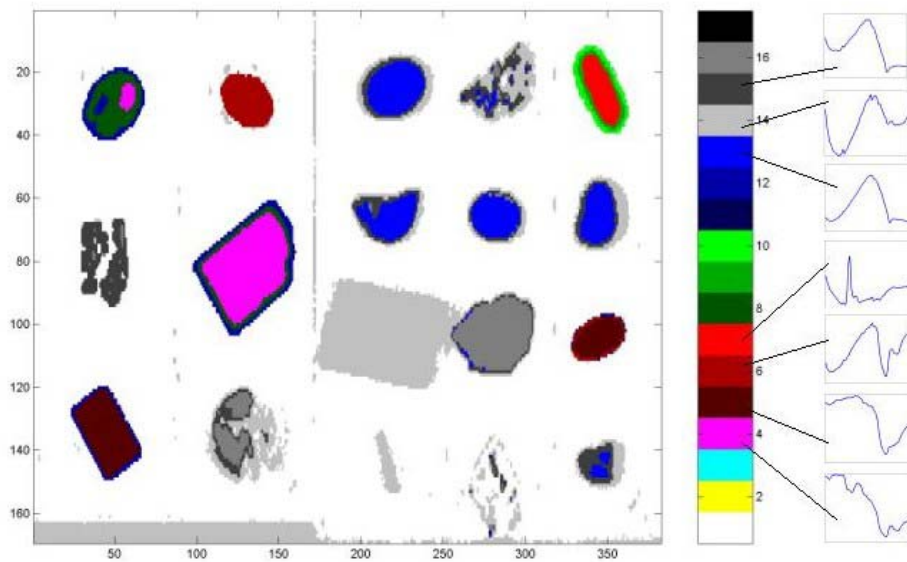


Fig. 6. Results from unsupervised classification of the near-IR spectra from 1000 nm to 1700nm using the fuzzy c-means algorithm. Shown on the right are characteristic spectra of the different classes.

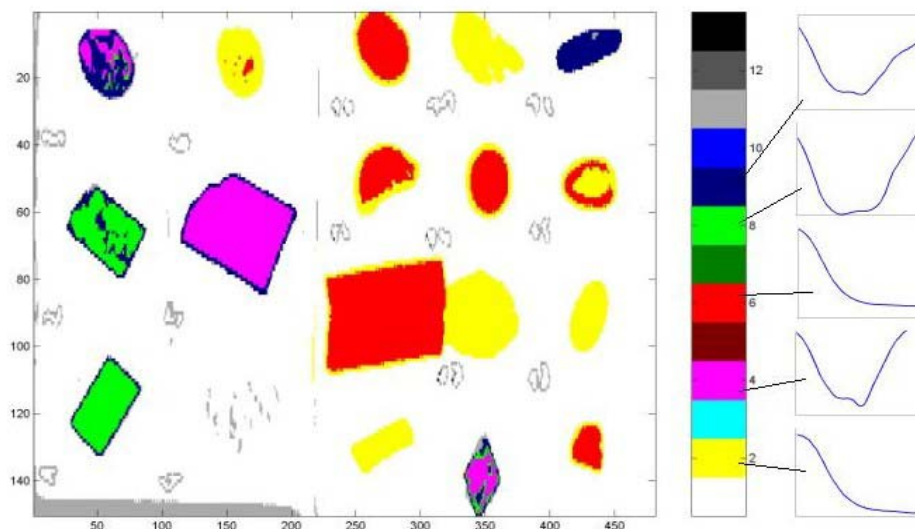


Fig. 7. Results from unsupervised classification of the spectra recorded in the visible region, between 500 nm (green) to 850 nm (near IR). Shown on the right are characteristic spectra of the different classes.

Classification results compared with chemical information.

Pure $\text{CuAl}_6[(\text{OH})_2/\text{PO}_4]_4 \cdot 4\text{H}_2\text{O}$ crystals are pure blue, due to the copper ions. In real turquoises, Al is substituted up to 5 % with iron, but also chromium and other metal ions can replace Al in the lattice. These ions, in particular iron, are responsible for the typical blue-green colour of the mineral. Related to the ion replacement in the lattice are absorption bands in the UV and blue (323 nm, 427 nm) and red to infrared region (around 800 nm) and an increased reflectivity in the blue-green (500 nm). In the NIR lower than 1400 nm, strong water and hydroxyl ion absorption bands can be observed. These strong spectral features give clear indications for turquoise samples. In the near-IR the low reflection at 1480 nm and high reflection at about 1320 nm are reliable indicators. In the visible, due to the typical colour of turquoise, all samples exhibit a high reflection in the blue-green (500 nm) region and little reflection in the red spectral region, thus rendering this spectral area not very relevant. The significant information is largely restricted to the wavelength region between 700 nm and 850 nm.

The classification showed that most true turquoise samples fall into classes 13 to 16 in the IR and class 2 or 6 in the visible region. Some samples exhibit spectra showing these characteristics, but differ in a other spectral

regions. Although this is an indication that something is wrong, it does not necessarily mean that the investigated stone is a fake. Possible reasons are e.g. a surface varnish to protect the stone, impurities in the mineralogical matrix or higher amounts of rarely encountered ions replacing the Al or Cu ions in the lattice, but also pressed turquoise powder. The latter case is the most commonly encountered faking method, as turquoise powder is significantly cheaper than large stones and the fact that chemically the materials are identical makes a differentiation complicated.

Stones consisting of other chemical substances can be identified very easily and quickly. The most commonly encountered fake materials, e.g. Gibbsite ($\gamma\text{-Al(OH)}_3$), Calcite (CaCO_3) and Prosopite ($\text{CaAl}_2(\text{F,OH})_8$), are easily separated from real turquoise as well as from each other. Sample 16 is Prosopite, exhibiting a single strong peak in the IR, samples 3 and 5 are Gibbsite and 2 and 17 are Calcite.

Conclusion

The combination of spectral imaging in the infrared and visible with suitable classification algorithms provides a reliable, cheap and fast method to distinguish real turquoises from other minerals. Samples belonging to classes 13 or 15 in the infrared and to class 2 or 3 in the visible can safely be classified as real natural turquoises. However, as turquoise is a natural stone, sometimes varieties occur which fall outside these specifications. This is even more complicated by treated turquoises, e.g. made from pressed powder or covered with some kind of varnish, giving very similar spectra and thus making a reliable identification not always possible. However, it should be possible to add an image processing routine as a next step to evaluate the inherent information about the structure or texture of the samples to address this problem. Finally, it proved to be feasible to identify turquoise-alikes, e.g. blue calcite, prosopite or gibbsite. In any case, the possibility to quickly and inexpensively differentiate between true turquoises, could-be turquoises and obviously different materials is a valuable tool, as it allows to reduce the number of samples that have to be analysed in the conventional way, e.g. by X-ray diffraction.

All samples could be classified and grouped in agreement to the identification supplied by the NHM Wien except for one case, where the classification results suggested an accidental exchange of two samples. This could be verified by alternative analysis methods and thus shows the reliability of the method. The trained classification can now be applied to unknown samples, in order to quickly sort unknown mineral samples and thus reduce the need for expensive and time-consuming standard analysis methods.

References:

- [1] G. Spöck and A. del Bianco, *Hyperspectral Imaging Toolbox*, CTR Report ChemSW1-01, Villach, 2001.
- [2] T. Hyvarinen, E. Herrala and A. Dall'Ava, SPIE symposium on Electronic Imaging, **3302**, San Jose, 1998.
- [3] D. Cosic and S. Loncaric, *New Methods for Cluster Selection in Unsupervised Fuzzy Clustering*, *Automatika*, **37**, 133-137, 1996.