Quality Evaluation in Spectral Imaging – Quality Factors and Metrics

Raju Shrestha, Ruven Pillay, Sony George and Jon Y Hardeberg

The Norwegian Colour and Visual Computing Laboratory, Gjøvik University College, Norway
Email: raju.shrestha@hig.no

Spectral imaging has many advantages over conventional three channel colour imaging, and has numerous applications in many domains. Despite many benefits, applications, and different techniques being proposed, little attention has been given to the evaluation of the quality of spectral images and of spectral imaging systems. There has been some research in the area of spectral image quality, mostly targeted at specific application domains. This paper seeks to provide a comprehensive review on existing research in the area of spectral image quality metrics. We classify existing spectral image quality metrics into categories based on how they were developed, their main features, and their intended applications. Spectral quality metrics, in general, aim to measure the quality of spectral images without considering specifically the imaging systems used to acquire the images. Having many different types of spectral imaging systems that could be used to acquire spectral images in an application, it is important to evaluate the performance/quality of these spectral imaging systems too. However, to our knowledge, not much attention has been given in this direction previously. As a first step towards this, we aim to identify different factors that influence the quality of the spectral imaging systems. In almost every stage of a spectral imaging workflow, there may be one or more factors that influence the quality of the final spectral image, and hence the imaging system used for acquiring the image. Identification of these factors, we believe, will be essential in developing a framework, for evaluating the quality of spectral imaging systems.

Received 5 January 2014; revised 26 May 2014; accepted 26 May 2014

Published online: 15 July 2014

Introduction

Spectral imaging has received much attention in recent years for its advantages over conventional three channel based colour imaging (usually RGB), and because of applications in a number of domains such as remote sensing, medical imaging, cultural heritage, biometrics and many more. A spectral imaging system captures image data at specific wavelength intervals (narrow or somewhat wider) across the electromagnetic spectrum. Based on their number of spectral bands, spectral imaging systems can be divided into two major types: multispectral and hyperspectral. There is no fine line separating the two; however, spectral imaging systems with more than 20 bands are generally considered as hyperspectral, and less than 20 as multispectral. We use the term spectral throughout this paper, to refer to both of them in a general sense. Hyperspectral imaging deals with imaging narrow spectral bands over a contiguous spectral range, and produces the spectra of all pixels in the scene. Multispectral imaging systems typically acquire images in a wider and limited number of spectral bands. They do not produce the spectrum of an object directly; but rather use spectral estimation algorithms to obtain spectral reflectances from the sensor responses. Hyperspectral imaging systems produce high measurement accuracy; however, the acquisition time, complexity and cost of these systems are generally high compared to multispectral systems.

 ISSN 2227-1309

A number of different acquisition techniques exist for both multispectral and hyperspectral imaging. Multispectral systems, for example, can be a multi-sensor based [1, 2], filter based using a filter wheel [3, 4] or tunable filter [5, 6], multispectral filter array (MSFA) based [7-9] , or light emitting diode (LED) based [10-13]. A filter-less and demosaicking-less colour sensitive device which uses transverse field detectors or tunable sensitivity sensors has also been proposed [14]. Many of these approaches require multiple shots in order to acquire a multispectral image, whereas a filter based one-shot solution which uses a stereo camera has been recently proposed [15-17]. Hyperspectral imaging systems can also be based on tunable filters and multiple shots, but are more often based on gratings, and used using a pushbroom technique [18].

Having many different types of spectral image acquisition systems, an important question that arises is how do we evaluate the quality of the spectral image data captured? In other words, there should be a way to evaluate the quality of a spectral image, and possibly of the spectral imaging system used to acquire the image. Much research has been carried out on image quality for classic three channel colour images and recently, in particular, based on perceptual quality [19, 20]. And there has been research intended for spectral image quality in a number of specific application domains [21-26]. However, to our knowledge, little work has been done on evaluating the quality of spectral imaging systems. This paper provides a comprehensive review of the research carried out so far in the field of spectral image quality metrics, and also identifies important factors involved in describing the quality of spectral imaging systems and spectral image data acquired with them. We believe that this work can further help in the development of general and/or application specific spectral image quality frameworks.

The next section reviews research that has been carried out on spectral image quality metrics. We then describe and classify the metrics into different categories before identifying and discussing the main quality factors and attributes that could form a basis of a global framework for spectral imaging quality.

Spectral image quality metrics

Many studies have been carried out on the quality evaluation of colour images. However, very little has been done on the evaluation of quality of spectral images. Before discussing the quality evaluation of spectral images, it would be useful to first discuss the notion of *quality* in the context of colour imaging and spectral imaging. There is no single universally accepted definition of colour image quality (CIQ), and this is even more true for spectral image quality. A number of tentative definitions can be found in the literature, and most of the definitions depend on particular applications. This is somewhat inevitable as quality always implies some application, nevertheless we think a more general approach will be beneficial.

One such definition adopted by many researchers is that by ISO [27], which defines image quality as the impression of the overall merit or excellence of an image, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted. One recent and extensive work that has been done on image quality by Pedersen [20] adopted this definition of image quality. Most of the previous work on image quality considered the visual perception of images (perceptual quality), either on a display or printed images. Unlike colour image quality, the definition of spectral image quality should not be limited to perceptual quality only. Since spectral imaging is used in wider application domains, spectral image quality may be defined differently based on its application domain. For instance spectral imaging can be used for producing more accurate colour reproduction. In this case, a spectral image and the imaging system, which produces the most accurate colour reproduction, may be considered as having the best quality. In the case of a spectral imaging system which is used to detect and classify the material composition of an object, the system which accurately detects and classifies can be considered as having the best quality. Thus, the notion of the quality of a spectral image should vary depending on the application. Several studies have been made on spectral image quality based on the application, purpose, and type of spectral imaging systems.

There are fundamentally three different types of colour image quality metrics: no-reference, reduced-reference, and full-reference [20]. In the no-reference metric type, only the reproduction is available, and the calculation of quality is based only on the reproduction without use of the reference (i.e. the original). In the reduced-reference metric type, some information of the reproduction and the original is used in the calculation of quality. Both the reference and the reproduction are available in the full-reference metric type. These could be true in the case of spectral image quality metrics also. Unlike for colour images, subjective quality assessments or vision based models are not sufficient to measure the quality of spectral imaging. Moreover, an end user may not be a human as in the case of colour imaging. Quality can be assessed either in the spatial or spectral domain and it is highly driven by the application. A spectral image quality metric, thus, could be calculated pixel-wise in the whole image globally, or it could be calculated based on spatial pixel values in local segments in the image. Based on whether the purpose of a metric is to evaluate the quality of a spectral or perceptual response, or how good a certain task can be carried out, whether it is a full-reference type, and if the calculation is based on the global or local spatial context in a test image, we propose to classify spectral image quality metrics into five categories: global, full reference spectral quality (GFSQ); global, full reference, perceptual quality (GFPQ); spatial, full reference spectral quality (SFSQ); spatial, full reference spectral quality (SFPQ); and task based quality (TBQ) metrics. Before going into the details of the classification, we first define terms and conventions to be used in describing different spectral image quality metrics, in a consistent way. We use I to denote an image. A spectral image of dimension $m \times n \times l$ can be defined as $I(x, y, \lambda_i)$, where $x = 1, \ldots, m$, and $y = 1, \ldots, n$. λ_i ($i = 1, \ldots, l$) denotes a spectral band or wavelength in a l-band spectral image. $I(x, y)$, thus, corresponds to the spectral reflectance at pixel (x, y) in the image. We denote the original reference image as I_r and the test image acquired by an imaging system as I_t.

We now describe the spectral image quality metrics that fall into the five categories below.

1. **Global, Full-reference, Spectral Quality (GFSQ) metrics:** Spectral image quality metrics, which are based on the calculation of the spectral response for every pixel of the test and reference images in a global context, fall into this category. In these types of metrics, metric values are calculated for every pixel of an image globally and the mean value (and possibly additional statistical information for instance, minimum and maximum values, standard deviation etc.) is computed by averaging over the metric values for all the pixels in the image.

   A number of quality metrics exist to evaluate the quality of spectral images based on spectral responses. One of the most widely used of these is the root mean square (RMS) error, which provides a statistical estimation of the difference between the spectral responses of test and reference images. RMS calculates the cumulative squared error between the original image and the test image. RMS has been widely used due to its easy calculation and analytical tractability. The mean RMS for the test spectral image I_t with respect to the given reference image I_r is given by:

$$RMS = \frac{1}{m \times n} \sum_{x=1}^{m} \sum_{y=1}^{n} RMS(x, y), \tag{1}$$

where $RMS(x, y)$ is the RMS error at pixel (x, y), and is calculated as:

$$RMS(x, y) = \sqrt{\frac{1}{l} \sum_{i=1}^{l} \left[I_t(x, y, \lambda_i) - I_r(x, y, \lambda_i) \right]^2}. \tag{2}$$

Another metric, the peak signal-to-noise ratio (PSNR) is also widely used, and it can be considered as a GSFQ as it is calculated from the RMS:

$$PSNR = 20 \log_{10} \left(\frac{1}{RMS} \right). \tag{3}$$

An alternative to the RMS is the goodness of fit coefficient (GFC), proposed by Romero et al. [28] . Unlike RMS, GFC is insensitive to the shift in magnitude, and its value is normalised to the range 0 to 1, with 1 indicating the perfect estimation and 0 (zero) indicating the worst estimation. The GFC value at a pixel (x, y), $GFC(x, y)$, is calculated as:

$$GFC(x, y) == \frac{\sum_{i=1}^{l} I_t(x, y, \lambda_i) I_r(x, y, \lambda_i)}{\sqrt{\sum_{i=1}^{l} I_t(x, y, \lambda_i)^2} \sqrt{\sum_{i=1}^{l} I_r(x, y, \lambda_i)^2}}. \tag{4}$$

Spectral Angle Map (SAM) [29] is another widely used metric, which is usually used for spectral segmentation, but which provides a measure of the difference in terms of spectral angle (α) between two spectra. SAM is nothing but an inverse cosine of the GFC metric, and hence SAM at a pixel (x, y) is calculated as:

$$\alpha(x, y) = \cos^{-1} \left[GFC(x, y) \right]. \tag{5}$$

Spectral Information Divergence (SID) is another metric used to compare spectral image data [30]. SID views each pixel spectrum as a random variable, and then measures the discrepancy of probabilistic behaviors between two spectra, thereby determining similarity and variability more effectively than SAM. SID at a pixel (x, y) is calculated using the equation:

$$SID(x, y) = \sum_{i=1}^{l} \left(\frac{I_t(x, y, \lambda_i)}{\sum_{j=1}^{l} I_t(x, y, \lambda_j)} - \frac{I_r(x, y, \lambda_i)}{\sum_{j=1}^{l} I_r(x, y, \lambda_j)} \right) \left(\log \frac{I_t(x, y, \lambda_i)}{\sum_{j=1}^{l} I_t(x, y, \lambda_j)} - \log \frac{I_r(x, y, \lambda_i)}{\sum_{j=1}^{l} I_r(x, y, \lambda_j)} \right). \tag{6}$$

Yet another metric that falls into the category GFSQ is the spectral similarity value (SSV), proposed by Sweet et al. [22]. SSV combines magnitude (m) and shape (s) differences between two spectral vectors, giving each equal weighting. SSV at a pixel (x, y) is computed as:

$$SSV(x, y) = \sqrt{m(x, y)^2 + s(x, y)^2}, \tag{7}$$

where $m(x,y)$ is computed as the root mean square value, $RMS(x,y)$:

$$m(x,y) = RMS(x,y) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}\left[I_t(x,y,\lambda_i) - I_r(x,y,\lambda_i)\right]^2} \tag{8}$$

and

$$s(x,y)^2 = 1 - \left(\frac{\frac{1}{l}\sum_{i=1}^{l}\left[I_t(x,y,\lambda_i) - \mu[I_t(x,y)]\right]\left[I_r(x,y,\lambda_i) - \mu[I_r(x,y)]\right]}{\sigma[I_t(x,y)]\sigma[I_r(x,y)]} \right)^2 . \tag{9}$$

Here $\mu[I_t(x,y)]$ and $\mu[I_r(x,y)]$ are means; and $\sigma[I_t(x,y)]$ and $\sigma[I_r(x,y)]$ are the standard deviation of two spectra at a pixel (x,y) in the test and the reference images, computed across the l wavelengths. The SSV approach is appropriate for use with hyperspectral images.

GFSQ metrics are computed per pixel in an image, and therefore, do not take into account variability in specific regions or spatial information from the whole area of the image. Moreover, these metrics are useful only if reference spectral data/image is available, which would not be the case in many situations.

2. **Global, Full-reference, Perceptual Quality (GFPQ) metrics:** These metrics are similar to the GFSQ metrics, except that in this case the metric calculation is based on visual perception (most commonly, the colour), instead of spectral responses. This type of metric is, therefore, useful in applications where we are interested in how a perceptually accurate image can be reproduced from the acquired spectral images. As in GFSQ, the GFPQ metric values are calculated for each pixels of the image, and then the mean value is calculated by averaging over all the values. Once a spectral image is transformed to a 3-band colour image for visual perception, depending on the need of an application, any colour image quality metric can then be used to evaluate its quality. As an illustration we will discuss a colour difference metric.

   Colour difference metrics are used to measure differences in colour between two colour patches. One of the most commonly used colour difference formula is CIE ΔE^*_{ab} [31] which is based on a perceptually uniform colour space, namely the CIELAB colour space. The colour difference is computed as the Euclidean distance between the two colours in the CIELAB space. Extensions of ΔE^*_{ab} have been proposed when it became apparent that it had problems, especially in the blue region with the CIE first proposing ΔE^*_{94} [32] and later ΔE^*_{00} [33]. Since these are increasingly complex compared to ΔE^*_{ab}, ΔE^*_{ab} is still widely used. The mean ΔE^*_{ab} between the test image (I_t) and the original reference image (I_r) is computed by averaging the colour differences in each pixel in the two images. ΔE^*_{ab} at a pixel (x,y) is calculated using the formula:

$$\Delta E^*_{ab}(x,y) = \sqrt{\left[\Delta L^*(x,y)\right]^2 + \left[\Delta a^*(x,y)\right]^2 + \left[\Delta b^*(x,y)\right]^2} , \tag{10}$$

   where

   $\Delta L^*(x,y) = L^*_t(x,y) - L^*_r(x,y)$,

   $\Delta a^*(x,y) = a^*_t(x,y) - a^*_r(x,y)$, and

   $\Delta b^*(x,y) = b^*_t(x,y) - b^*_r(x,y)$

are differences in luminance (L^*) and chrominance (a^* and b^*) channels in the CIELAB space, at pixel (x, y) in the test and the reference images.

Metamerism index based metrics which compare the extent to which two spectra have a different colour between a reference condition and a test condition under different illuminants and observers have also been proposed [34, 35]. Viggiano's metamerism index [35], at pixel (x, y) in an image, $M_v(x, y)$ is computed using the equation:

$$M_v(x,y) = \sum_{i=1}^{l} w(x,y,\lambda_i) \,\|\, \Delta\beta(x,y,\lambda_i) \,\|, \tag{11}$$

where $\Delta\beta(x,y,\lambda_i) = I_t(x,y,\lambda_i) - I_r(x,y,\lambda_i)$, and $w(x,y,\lambda_i)$ are weights computed as follows:

$$w(x,y,\lambda) = \sqrt{\left(\frac{\Delta L^*(x,y)}{\Delta\beta(x,y,\lambda_i)}\right)^2 + \left(\frac{\Delta a^*(x,y)}{\Delta\beta(x,y,\lambda_i)}\right)^2 + \left(\frac{\Delta b^*(x,y)}{\Delta\beta(x,y,\lambda_i)}\right)^2}\,. \tag{12}$$

Perception based quality metrics including GFPQ work only in the visible part of the spectrum, and thus ignore important information in invisible bands such as the infrared and ultraviolet. This limits the use of such metrics in object detection and classification. Moreover, one single image quality metric is inadequate to indicate the quality of an image [36]. None of the perception based spectral image quality metrics takes into account this fact, and therefore cannot be considered as complete spectral image quality metrics.

3. **Spatial, Full-reference, Spectral Quality (SFSQ) metrics:** Some full-reference metrics aim to calculate spectral quality by taking into account the spatial distribution in images. These metrics can be categorised as SFSQ metrics.

One such metric is the $Q2^n$ index, proposed by Garzelli and Nencini [37]. The $Q2^n$ index extended the universal quality index (UQI) proposed for monochrome images [38], as a generalisation to multispectral and hyperspectral images, through a hypercomplex correlation coefficient (CC) between the reference (I_r) and the test images (I_t). The index jointly measures both spectral and spatial distortions. The $Q2^n$ index is derived from the theory of hypercomplex numbers, particularly of 2^n-ons (two-to-the-any-ons) [39], and made up of different factors to take into account for correlation, mean of each spectral band, intra-band local variance, and the spectral angle. Two hypercomplex image maps corresponding to the test and the reference images, $I_{t,h}$ and $I_{r,h}$ are obtained from a 2^n-on hypercomplex number at each pixel from the 2^n spectral bands. If the number of bands is not a power of two, the image bands are appropriately zero-padded, to analyse the overall data with 2^n spectral bands. The null bands do not influence the image quality measurement. A 2^n-on hypercomplex number for an image, I at a pixel (x, y) is represented as:

$$I_h(x,y) = I(x,y,\lambda_1) + \sum_{i=2}^{n} I(x,y,\lambda_i)\, j_{2^i}\,, \tag{13}$$

where $j_2, j_3, \ldots, j_{2^n}$ are hypercomplex unit vectors. Analogously to complex number, the conjugate I_h^* is given by:

$$I_h(x,y) = I(x,y,\lambda_1) - \sum_{i=2}^{n} I(x,y,\lambda_i)\, j_{2^i}\,, \tag{14}$$

and the modulus by

$$| I_h(x,y) | = \sqrt{\sum_{i=1}^{n} I(x,y,\lambda_i)^2} \ . \tag{15}$$

The $Q2^n$ computes the correlation, the mean of each spectral band, and the intra-band local variance, at each pixel (x, y) using a sliding window of size $N \times N$ in the hypercomplex image map. Let $\mathbf{t}$ and $\mathbf{r}$ be the pixel arrays within the sliding window in the test and the reference image maps respectively. The $Q2^n$ at pixel (x, y) is then computed using the equation:

$$Q2^n(x,y) = \frac{\sigma_{\mathbf{tr}}}{\sigma_{\mathbf{t}}\sigma_{\mathbf{r}}} \cdot \frac{2\overline{\mathbf{t}}\overline{\mathbf{r}}}{\overline{\mathbf{t}}^2 + \overline{\mathbf{r}}^2} \cdot \frac{2\sigma_{\mathbf{t}}\sigma_{\mathbf{r}}}{\sigma_{\mathbf{t}}^2 + \sigma_{\mathbf{r}}^2} \ , \tag{16}$$

where $\overline{\mathbf{t}}$ and $\overline{\mathbf{r}}$ are means, and $\sigma_{\mathbf{t}}$ and $\sigma_{\mathbf{r}}$ are the standard deviations of $\mathbf{t}$ and $\mathbf{r}$ respectively. $\sigma_{\mathbf{tr}}$ is the hypercomplex covariance between $\mathbf{t}$ and $\mathbf{r}$. These terms are defined as:

$$\overline{\mathbf{t}} = E[\mathbf{t}] \ ,$$

$$\overline{\mathbf{r}} = E[\mathbf{r}] \ ,$$

$$\sigma_{\mathbf{t}} = E[|\mathbf{t}|^2] - |\overline{\mathbf{t}}|^2 \ ,$$

$$\sigma_{\mathbf{r}} = E[|\mathbf{r}|^2] - |\overline{\mathbf{r}}|^2 \ ,$$

$$\sigma_{\mathbf{tr}} = E[(\mathbf{t} - \overline{\mathbf{t}})(\mathbf{r} - \overline{\mathbf{r}})] = E[\mathbf{tr}^*] - \overline{\mathbf{t}}.\overline{\mathbf{r}}^* \ .$$

Among the three terms in Equation (16), the first term measures the hypercomplex CC, the second term measures the mean of the spectral band, and the third term computes the intra-band local variance. The mean $Q2^n$ is then obtained by averaging the magnitudes of all $Q2^n$'s over the whole image:

$$Q2^n = E\left[|Q2^n(x,y)| \right]. \tag{17}$$

The $Q2^n$ index assumes real values in the interval [0, 1], with 1 being the best value, which can be achieved if and only if the test image is identical to the reference image. This metric is useful to measure the fidelity of a spectral image with respect to a known reference in terms of both spatial and spectral distortions.

4. **Spatial, Full-reference, Perceptual Quality (SFPQ) metrics:** Full-reference metrics which compute the quality of spectral images based on perceptual quality, taking into account local spatial information, can be classified into this category. The SFPQ metrics, therefore, are mainly based on spectral images in the visual range.

   LeMoan and Urban [40] recently proposed an evaluation technique of the perceptual quality of spectral images, which is based on pooling (averaging) the image quality indices proposed by Lissner et al. [41], computed under a set of different illuminants, with the images rendered in the perceptually uniform LAB2000HL colour space [42], whose perceptual uniformity is based on the CIE ΔE^*_{00} colour difference formula [33]. Based on the assumption that given the difference of two images under a certain illuminant, the error added by considering other illuminants can be summarised solely in terms of chroma and hue difference, they proposed an approximation of the spectral image difference (SpID) between the two images as the average of colour image differences (CID) under different viewing conditions (VC). SpID at a pixel (x, y), $SpID(x, y)$ is calculated using the equation:

$$SpID(x,y) = \frac{1}{N_{VC}} \sum_{i=1}^{N_{VC}} CID_{VC_i}(x,y), \tag{18}$$

where N_{VC} is the number of viewing conditions considered. CID under a visual condition, $CID_{vc}(x, y)$ is computed by transforming spectral image to a CIE XYZ image, and then incorporating the chromatic adaptation transform (CAT) employed by CIECAM02. It is calculated using the equation:

$$CID_{VC}(x,y) = 1 - l_L(x,y) l_C(x,y) l_H(x,y) c_L(x,y) s_L(x,y), \tag{19}$$

where l_L, l_C, l_H, c_L, and s_L are image difference features (IDFs): lightness-difference, chroma difference, hue difference, lightness-contrast and lightness-structure respectively. These features are derived from the LAB2000HL images using the structural similarity index (SSIM) [19]. These terms at a pixel (x, y) are computed within a sliding window in the two images. Let $\mathbf{t}$ and $\mathbf{r}$ are the pixel arrays within this window in the test and the reference images. Among the five IDFs, lightness, chroma and hue differences are calculated as follows:

$$l_L(x,y) = \frac{1}{c_1.\overline{\Delta L(\mathbf{t},\mathbf{r})}^2 + 1} \tag{20}$$

$$l_C(x,y) = \frac{1}{c_4.\overline{\Delta C(\mathbf{t},\mathbf{r})}^2 + 1} \tag{21}$$

$$l_H(x,y) = \frac{1}{c_5.\overline{\Delta H(\mathbf{t},\mathbf{r})}^2 + 1} \tag{22}$$

where $\overline{f(\mathbf{t},\mathbf{r})}$ denotes a Gaussian-weighted mean of $f(x, y)$, computed for the pixel (x, y), using all the pixel pairs in the two images, within the sliding window. ΔC is the chroma difference, the chroma being defined as $C = \sqrt{a^2 + b^2}$. Hue difference, ΔH is computed using the equation:

$$\Delta H = \sqrt{(a_t - a_r)^2 + (b_t - b_r)^2 + \Delta C^2}\,. \tag{23}$$

Lightness-contrast and lightness-structure are computed as follows:

$$c_L(x,y) = \frac{2\sigma_{\mathbf{t}}\sigma_{\mathbf{r}} + c_2}{\sigma_{\mathbf{t}}^2 + \sigma_{\mathbf{r}}^2 + c_2} \quad (24)$$

$$s_L(x,y) = \frac{\sigma_{\mathbf{tr}} + c_3}{\sigma_{\mathbf{t}}\sigma_{\mathbf{r}} + c_3} \quad (25)$$

where $\sigma_{\mathbf{t}}$ and $\sigma_{\mathbf{r}}$ are the standard deviations of the lightness components in the sliding windows. $\sigma_{\mathbf{tr}}$ corresponds to the cosine of the angle between $\mathbf{t} - \bar{\mathbf{t}}$ and $\mathbf{r} - \bar{\mathbf{r}}$ in the lightness component [19]. $c_1, \ldots, c_5$ are parameters that are adjusted for the colour space used, and large colour differences.

Like GFPQ, being perceptual quality based metrics, SFPQ metrics also do not take into account information in the invisible bands.

5. **Task based (Functional) Quality (TBQ) metrics:** There are some spectral image quality metrics which are aimed at an evaluation based on their performance for a certain task or function. These metrics can be categorised as TBQ metrics. One of the most common task is to detect targets/objects in a scene. Several spectral image quality metrics have been proposed in order to evaluate the performance for object detection and/or classification. In most cases, spectral data which cover a wider spectrum including the ultraviolet and/or infrared ranges are used as this allows detection even of targets invisible to the human eye.

   Kerekes and Hsu [23] proposed a model-based spectral quality rating scale (SQRS) for target detection in VNIR (Visible Near InfraRed)/SWIR (ShortWave InfraRed) hyperspectral images. The higher the SQRS value, the better the spectral quality of the image. In their latest version of the work, SQRS is computed using the empirically-derived equation:

$$SQRS_{SCR-detection} = 10.6 - 1.6\log_{10}(t) + 3.3\log_{10}(GSD) + 1.6\log_{10}(SCR) \quad (26)$$

   where GSD is the ground sample distance in cm, SCR is the signal-to-clutter ratio defined for a target and background having spectral mean vectors μ_t and μ_b, and the background having a spectral covariance matrix, $\sum_b$. t is a threshold on the normalised match filter output test statistic θ that leads to a specified false alarm rate on the image background. *SCR,* and $\theta(p)$ (θ for a pixel p) are computed using the equations:

$$SCR = \sqrt{(\mu_t - \mu_b)^t \Sigma_b^{-1} (\mu_t - \mu_b)} \quad (27)$$

$$\theta(p) = \frac{(\mu_t - \mu_b)^t \Sigma_b^{-1} (p - \mu_b)}{(\mu_t - \mu_b)^t \Sigma_b^{-1} (\mu_t - \mu_b)} \quad (28)$$

   Shen [43] analysed a large number of images with varying spectral image parameters and proposed a target detection probability measure (P_D) based on a regression between some metrics and image parameters:

$$P_D = 6.25 - 0.811\log_{10}(GSD) + 0.12\log_{10}(SNR) - 0.20\log_{10}(\Delta\lambda) - 2.43\log_{10}(\sigma_{scene}) \quad (29)$$

where GSD is the ground sample distance in cm, SNR is the signal-to-noise ratio, $\Delta\lambda$ is the average spectral resolution of the channels in nm, and σ_{scene} is the average standard deviation in HYDICE scaled radiance units (1 HYDICE = 4/3 μW/cm^2-sr-μm), of the pixels in the scene across all spectral bands.

Martin *et al.* [21] defined spectral quality as the extent to which an image or data set precisely replicates the scene represented by the image or data set. They proposed an approach to subjectively determine the utility through analyst assessments, calculate the quality of an image, and then relate these two metrics to obtain an objective quality metric. The postulated probability of correct material identification, P_{CI} is defined as a function of a number of parameters including the accuracy of signature definition, the sensor performance (spatial, spectral, and radiometric), the analysis of system performance, the sample abundance, and a decision criterion.

Simmons *et al.* [44] tried to combine spectral and spatial information with the aim of a general quality metric based on semantic transformations of the spatial and spectral quality [44]. It calculates spatial and spectral confidences, and a single total confidence value is obtained by combining the two confidences:

$$C_{Total} = 1 - (1 - C_{Spatial}).(1 - C_{Spectral}), \quad (30)$$

where $C_{Spectral}$ is the spectral confidence, which is obtained through an assessment of the separability of target and background spectral distributions or from results of hyperspectral image analysis techniques. $C_{Spatial}$ is the spatial confidence, which is largely driven by the size of the target relative to the image resolution, and is computed using the equation:

$$C_{Spatial} = \frac{(N/N_{50})^E}{1 + (N/N_{50})^E}, \quad (31)$$

where $E = 2.7 + 0.7(N/N_{50})$, N is the number of resolutions cycles per minimum dimension of the target, and N_{50} is the cycle criteria for 50 percent success and has values of 1.0±0.25, 4.1 ± 0.35 and 6.4 ± 1.5 for detection, recognition and identification respectively.

Most of the TBQ metrics are based on empirical modeling from a limited set of data, and therefore may not work well in a general sense. Since they use some of the information available from the original scene, they are of reduced-reference type. These metrics, in general, do not take into account spectral and colour accuracies.

Purpose	Full reference	
	Global	Spatial
Spectral	RMS, PSNR, GFC, SAM, SID, SSV	$Q2^n$
Perceptual/Colour	ΔE^*_{ab}, M_v, other CIQ metrics	SpID

Table 1: Summary table of the first four categories of the spectral image quality metrics.

Among the five categories of spectral image quality metrics just discussed above, we can summarise the first four as shown in Table 1. From the table we see that all the metrics belonging to these four categories are full-reference type. The TBQ metrics ($SQRS$, P_D, P_{CI}, Simons et al.'s C_{Total}) are mostly reduced-reference type, and they are calculated globally or spatially or a mix of both. We have found no no-reference type spectral image quality metric. From the review, we have seen that all the spectral image quality metrics are aimed at certain application requirements. For instance, GFSQ metrics are aimed at evaluating the accuracy of spectral responses of the test images compared to the original images. None of the metrics so far proposed has been universally accepted as a general spectral quality metric. A comparative study on GFSQ and GFPQ metrics showed that none of the metrics are superior to others for all purposes, and that the choice of metric should be made based on appropriateness to the application [45]. There is, therefore, a need to do further research towards a more effective and possibly a more general spectral image quality metric.

Towards a framework for spectral imaging system quality

We have reviewed a range of spectral image quality metrics, which essentially aim to objectively evaluate the quality of spectral images. As there are many different types of spectral imaging systems that can be used to acquire the spectral images, it is increasingly important to evaluate the quality of these imaging systems themselves. In general one or more appropriate spectral image quality metrics are used to evaluate the quality of the imaging systems also. But from the review of the existing spectral image quality research, we see that the spectral image quality metrics/techniques so far proposed do not take into account all of the parameters that can influence the overall quality, and hence these metrics are not sufficient to fully evaluate the quality of the imaging systems. There is, therefore, a need for a spectral image quality framework which takes into account all of these factors including the characteristics of the scene, acquisition system, algorithms, application requirements etc. Information from the different stages of the spectral imaging workflow such as: spectral acquisition, processing, and resulting spectral data could provide information on these attributes. In this section we try to identify some of the most important of these as a basis for developing a framework for spectral imaging system quality.

- An important attribute that measures quality is the *spectral accuracy*. GFSQ metrics could be used to evaluate this attribute.
- The *perceptual quality* of a colour image rendered from a spectral image can be an important broad level attribute. This could be further detailed using effective image quality metrics from GFPQ, including colour accuracy.
- *Spatial* and *spectral resolution* are two important attributes whose information is available at the very beginning of the acquisition process. There could be a tradeoff between the spatial and spectral resolutions. A good example is the MSFA based spectral imaging system, where there is a need to compromise spatial resolution in order to increase the number of spectral bands.
- Many spectral acquisition systems rely on *image fusion* or *registration*. In such systems, pixel-to-pixel registration is important to obtain a high quality result. In some spectral imaging techniques, for example in satellite imaging, image fusion is one integral part of the spectral

imaging process and some research exists on the evaluation of the spectral images based on the quality of image fusion [25, 26]. Many of the spectral imaging systems use a combination of optical components that can produce both spatial and spectral distortion. These require careful characterisation, including the determination of a sensor model, in order to calibrate fully.

- *Noise* is an inevitable part of digital imaging. Different capturing methods involve different noise characteristics. An effective noise measurement and model should be developed in order to take noise into account more realistically.
- The processing stages may involve *geometrical corrections*, *spectral estimation methods*, and *demosaicking algorithms*, and these algorithms play a vital role in the final quality of a spectral image.
- *Target detection and recognition* capabilities could be other attributes to be considered and TBQ metrics are useful here.
- Repeatability and reproducibility are also very important for a good spectral imaging system. Vilaseca *et al.* [46] studied and analysed the repeatability, reproducibility and accuracy of a pushbroom hyperspectral system, and from their study they concluded that hyperspectral systems have good repeatability, adequate reproducibility and good accuracy. They used the spectral metric *RMS* and the colorimetric metrics ΔE^*_{ab} and ΔE^*_{00} in order to evaluate the accuracy.

We believe that taking into account all of these factors in a general quality framework would lead to a more effective evaluation of the quality of spectral imaging systems.

Conclusions

We have carried out a comprehensive review of previous studies on spectral image quality research. From this review, we found that most of the spectral image quality evaluations are intended for a number of specific domains and/or applications. They do not take into account all the key attributes that influence the quality of the resulting spectral data, and hence are not sufficient to be used to fully evaluate the quality of the spectral imaging systems. We have established a need for a generalised spectral image quality framework, and as a basis for its development, we have identified some of the important factors and attributes that might be involved in one or more of the steps in the workflow of the spectral acquisition process and which will, in turn, influence the overall quality of the spectral image data. Development of a general spectral image quality framework and metric, taking into account those attributes, will therefore be an important and useful area of future research.

References

1. FluxData, *Fluxdata FD-1665 series of 3 CCD cameras*, [www.fluxdata.com – last accessed 31st January 2014].
2. Murakami Y, Yamaguchi M and Ohyama N (2012), Hybrid-resolution multispectral imaging using color filter array, *Optics Express*, **20** (7), 7173-7183.
3. Martinez K, Cupitt J, Saunders D and Pillay R (2002), Ten years of art imaging research, *Proceedings of the IEEE*, **90** (1), 28-41.
4. Ribés A, Schmitt F, Pillay R and Lahanier C (2005), Calibration and spectral reconstruction for CRISATEL: An art painting multispectral acquisition system, *Journal of Imaging Science and Technology,* **49** (6), 463-473.

5. Miller PJ and Hoyt CC (1995), Multispectral imaging with a liquid crystal tunable filter, *Proceedings of SPIE: Optics in Agriculture, Forestry, and Biological*, **2345**, 354-365.
6. Hardeberg JY, Schmitt F and Brettel H (2002), Multispectral color image capture using a liquid crystal tunable filter, *Optical Engineering*, **41** (10), 2532-2548.
7. Miao L and Qi H (2006), The design and evaluation of a generic method for generating mosaicked multispectral filter arrays, *IEEE Transactions on Image Processing,* **15** (9), 2780-2791.
8. Shrestha R, Hardeberg JY and Khan R (2011), Spatial arrangement of color filter array for multispectral image acquisition, *Proceedings of SPIE: Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII*, **7875**, 787503-787503-9.
9. Shrestha R and Hardeberg JY (2013), CFA based simultaneous multispectral imaging and illuminant estimation, *Computational Color Imaging Workshop (CCIW)*, **7786**, 158-170.
10. Park JI, Lee MH, Grossberg MDZD and Nayar SK (2007), Multispectral imaging using multiplexed illumination, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1-8.
11. Shrestha R, Hardeberg JY and Boust C (2012), LED based multispectral film scanner for accurate color imaging, *Proceedings of the Eighth International Conference on Signal Image Technology and Internet based Systems (SITIS)*, 811-817.
12. Shrestha R and Hardeberg JY (2013), Multispectral imaging using LED illumination and an RGB camera, *Proceedings of the Twenty-first Color and Imaging Conference (CIC)*, 8-13, Albuquerque, USA.
13. Parmar M, Lansel S and Farrell J (2012), An LED-based lighting system for acquiring multispectral scenes, *Proceedings of SPIE: Digital Photography VIII*, **82990**, 82990P-82990P-8.
14. Langfelder G, Longoni AF and Zaraga F (2011), Implementation of a multispectral color imaging device without color filter array, *Proceedings of SPIE: Digital Photography VII*, **7876**, 787608-787608-9.
15. Shrestha R, Mansouri A and Hardeberg JY (2011), Multispectral imaging using a stereo camera: Concept, design and assessment, *EURASIP Journal on Advances in Signal Processing,* **2011** (1), 57-57.
16. Shrestha R and Hardeberg JY (2012), Simultaneous multispectral imaging and illuminant estimation using a stereo camera, in *Image and Signal Processing*, Lecture Notes in Computer Science (LNCS), **7340**, 45-55, Springer Berlin Heidelberg.
17. Shrestha R, Hardeberg JY and Mansouri A (2011), One-shot multispectral color imaging with a stereo camera, *Proceedings of SPIE: Digital Photography VII*, **7876**, 787609-787609-11.
18. Goetz AFH, Vane G, Solomon JE and Rock BN (1985), Imaging spectrometry for earth remote sensing, *Science,* **228** (4704), 1147-1153.
19. Wang Z, Bovik AC, Sheikh HR and Simoncelli EP (2004), Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing,* **13** (4), 600-612.
20. Pedersen M (2011), Image quality metrics for the evaluation of printing workflows, *PhD Thesis*, University of Oslo and Gjøvik University College.
21. Martin L, Vrabel J and Leachtenauer J (1999), Metrics for assessment of hyperspectral image quality and utility, *Proceedings of International Symposium on Spectral Sensing Research*, Las Vegas, USA.
22. Sweet J, Granahan J and Sharp M (2000), An objective standard for hyperspectral image quality, *Proceedings of AVIRIS Workshop*, Jet Propulsion Laboratory, Pasadena, California, USA.
23. Kerekes JP and Hsu SM (2004), Spectral quality metrics for VNIR and SWIR hyperspectral imagery, *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, **5425**, 549-557.
24. Kalenova D, Toivanen P and Bochko V (2005), Preferential spectral image quality model, in *Image Analysis*, Lecture Notes in Computer Science (LNCS), **3540**, 389-398, Springer Berlin Heidelberg.
25. Babawuro U and Beiji Z (2011), Satellite imagery quality evaluation using image quality metrics for quantitative cadastral analysis, *International Journal of Computer Applications in Engineering Sciences*, **1**, 391-395.

26. Al-Wassai FA, Kalyankar NV and Al-Zaky AA (2012), Spatial and spectral quality evaluation based on edges regions of satellite: Image fusion, *Proceedings of the Second International Conference on Advanced Computing Communication Technologies (ACCT)*, 265-275.
27. ISO (2004), ISO 20462-2 photography - psychophysical experimental methods to estimate image quality - part 2: Triplet comparison method.
28. Romero J, García-Beltrán A and Hernández-Andrés A (1997), Linear bases for representation of natural and artificial illuminants, *Journal of the Optical Society of America A*, **14** (5), 1007-1014.
29. Kruse FA, Lefkoff AB, Boardman JW, Heidebrecht KB, Shapiro AT, Barloon PJ and Goetz AFH (1993), The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data, *Remote Sensing of Environment*, **44** (2-3), 145-163.
30. Chang CI (1999), Spectral information divergence for hyperspectral image analysis, *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, **1**, 509-511.
31. CIE (2004), CIE 15:2004: Colorimetry.
32. CIE (1995), Industrial colour-difference evaluation, publication CIE 116-95.
33. Luo MR, Cui G and Rigg B (2001), The development of the CIE 2000 colour-difference formula: CIEDE 2000, *Color Research and Application*, **26** (5), 340-350.
34. Fairman HS (1987), Metameric correction using parametric decomposition, *Color Research and Application*, **12**, 261-265.
35. Viggiano JAS (2002), Perception-referenced method for comparison of radiance ratio spectra and its application as an index of metamerism, *Proceedings of SPIE: The Ninth Congress of the International Colour Association*, **4421**, 701-704, Rochester, USA.
36. Pedersen M and Hardeberg JY (2012), Full-reference image quality metrics: Classification and evaluation, *Foundations and Trends in Computer Graphics and Vision*, **7** (1), 1-80.
37. Garzelli A and Nencini F (2009), Hypercomplex quality assessment of multi/hyperspectral images, *IEEE Geoscience and Remote Sensing Letters*, **6** (4), 662-665.
38. Zhou W and Bovik AC (2002), A universal image quality index, *IEEE Signal Processing Letters*, **9** (3), 81-84.
39. Smith WD (2004), *Quaternions, octonions, and now, 16-ons and 2n-ons; new kinds of numbers*, [http://www.scorevoting.net/WarrenSmithPages/homepage/nce2.pdf – last accessed 31st January 2014].
40. LeMoan S and Urban P (2013), Evaluating the perceived quality of spectral images, *Proceedings of the International Conference on Image Processing (ICIP)*, 2024-2028, Melbourne, Australia.
41. Lissner I, Preiss J, Urban P, Lichtenauer MS and Zolliker P (2013), Image-difference prediction: From grayscale to color, *IEEE Transactions on Image Processing*, **22** (2), 435-446.
42. Lissner I and Urban P (2012), Toward a unified color space for perception-based image processing, *IEEE Transactions on Image Processing*, **21** (3), 1153-1168.
43. Shen SS (2003), Spectral quality equation relating collection parameters to object/anomaly detection performance, *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, **5093**, 29-36.
44. Simmons RE, Elder TD, Stewart DJ, Cincotta EJ, Kennedy CS and Van Nostrand RC (2005), General spectral utility metric for spectral imagery, *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, **5806**, 457-468.
45. Imai FH, Rosen MR and Berns RS (2002), Comparative study of metrics for spectral match quality, *Proceedings of the First European Conference on Colour in Graphics, Imaging and Vision (CGIV),* 492-496, Poitiers, France.
46. Vilaseca M, Schael B, Delpueyo X, Chorro E, Perales E, Hirvonen T and Pujol J (2013), Repeatability, reproducibility, and accuracy of a novel pushbroom hyperspectral system, *Color Research and Application*, doi: 10.1002/col.21851.