Multispectral Image Encoding and Compression

Lindsay MacDonald, Stephen Westland and Dongmei Liu
Colour & Imaging Institute
University of Derby, United Kingdom

Abstract

Digital image archives may store images in either device-dependent or device-independent form, but the latter places more stringent demands on the image capture process. Multispectral image archives can reduce the problems of colour inconstancy associated with trichromatic images. An experimental study was conducted to determine the efficiency of a new method for encoding multispectral images, using three orthogonal basis functions derived from cone fundamentals plus additional principal components. The results indicated that the encoding method gave a good approximation to the original, and that images of reasonable quality could be reconstructed from files compressed by spatial sub-sampling of the components, with an overall compression ratio in the range 20-40.

Multispectral Image Capture

The digital reproduction of a colour image typically involves a chain of processes from original scene to destination image. The original scene may be captured directly by means of a digital camera or digitised from a photograph, then stored in a source image file. This file may subsequently be processed by suitable image processing algorithms, often with the interactive control of a human operator for editorial correction or enhancement, and the result stored in a destination image file. This may then be transmitted to a destination device or process to reproduce the image in a particular medium, for example printed with coloured ink on paper.

The multispectral digital image, in terms of its place in the reproduction chain, is an alternative representation of the original scene or object in the digital source image file. Because it carries more detailed spectral information than a conventional three-band digital image, it can provide a more complete representation of the colorants of the original and hence a more accurate record of the reflectance under different sources of illumination. Thus it has potential to overcome many of the problems of metamerism or colour inconstancy associated with conventional image archives.[1]

An ideal archival image would contain all the information present in the original scene or object. If the final medium, rendering intent, application and task are not known at image capture time then the archive image should ideally be able to support any and every possible combination. It should allow the user to 'zoom in' close enough to discern the finest detail present in the original, or to visualise the scene from any angle or viewpoint, or to simulate its appearance under any type of lighting in any viewing environment. In practice, such an objective is only possible for a computer-rendered model, where the image has been synthesised from a specified parameter set.

A real archival image must always differ from the original scene or object, as it is a representation acquired through optical projection and sampling, including:

- geometry of illumination source(s) and camera relative to the scene;
- 3-D original optically projected onto the 2-D image plane within camera;
- sampled in time, either a static (one-off) image or a progressive series;
- sampled in wavelength into specified number of spectral channels;
- sampled spatially across the 2-D image plane.

Limits of Resolution

Because reflectance spectra are functions of wavelength, which can be expressed as cycles per nm, they can be represented by Fourier spectra. The Uniform Sampling Theorem[2] states that "if $x(t)$ is band-limited with no components greater than f_h Hz, then it is completely specified by samples taken at the frequency $f_s > 2\, f_h$ Hz". The minimum sampling frequency f_s is the *Nyquist frequency*. The problem that the spectral reflectance data is not a continuous function of wavelength but a finite set of values sampled at regular intervals (typically 5 or 10 nm) can be overcome via the Discrete Fourier Transform (DFT). The restriction of the limited range of the visible spectrum (approximately 400 to 700 nm) can be corrected by a Hanning window, which prevents the introduction of spurious frequency components into the Fourier spectrum, as shown in Figure 1. If reflectance spectra were sampled by a spectrophotometer at 31 points in the visible range 400-700 nm, i.e. at 10 nm intervals, the maximum detectable frequency would be constrained to 15 cycles per 300 nm or 0.05 cycles/nm.

It is interesting to consider the relationship between spectral frequency and colour saturation. Higher spectral frequencies correspond to steeper slopes and more rapid rates of change of reflectance as a function of wavelength. Hence a colour that has more power at high spectral frequencies will exhibit more chromatic purity, so will be further from the white point and nearer to the gamut boundary of the colour space.[3] The limiting case is for monochromatic colours, where all the power is concentrated at a single wavelength, i.e. at infinite spectral frequency, corresponding to a point on the spectral locus of the CIE chromaticity diagram, as shown in Figure 2. Theoretical 'block dyes', used in the analysis of subtractive colour reproduction systems, also have infinite spectral frequency components, because of their infinite slopes in reflectance space, and lie close to the spectral locus.[4]

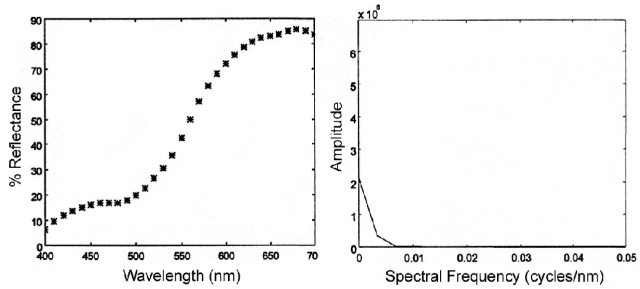


Figure 1 Fourier analysis of a typical reflectance spectrum (left) showing the Fourier amplitude after Hanning correction (right)

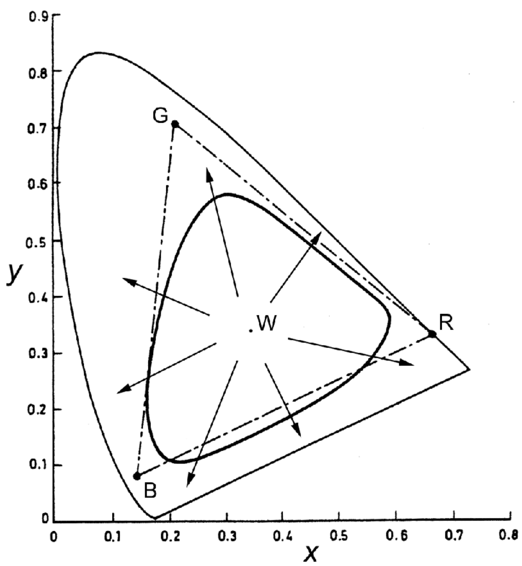


Figure 2 Effect of increasing spectral frequency on chromatic purity. The solid curve shows the gamut of colours corresponding to band-limited spectral frequencies of 0.005 cycles/nm. NTSC primaries are shown for comparison (triangle). (Adapted from Buchsbaum & Gottschalk[3])

For multispectral image capture, a key question is: How many spectral components should be captured? Studies of images using Principal Component Analysis (PCA) have revealed that the spectral reflectances of both natural scenes and man-made scenes can in most cases be approximated accurately with five or six basis functions.[5] It follows that for efficient storage of multispectral images, instead of storing the reflectance of each waveband, it is better to derive and store the coefficients of the basis functions for each pixel of the image, constructed as linear combinations of the spectral channels from the capture device.

Man-made objects, however, have an inherent resolution limit, excluding the micro-structure of the medium. Fine-art painting has a maximum resolvable detail of about 10 lines/mm, and transparency photographs about 50 lines/mm. For man-made objects, therefore, once the medium has been properly characterised, it is possible to establish a spatial sampling limit which can be guaranteed to capture all the spatial information in the original object.

The human visual system has significantly higher spatial sensitivity for luminance components than for chrominance.[6] The contrast sensitivity function (CSF) for luminance is band-pass in nature, peaking at about 6 cycles per degree of visual angle (cpd) and approaching zero at both very low (less than 0.1 cpd) and very high (greater than 50 cpd) spatial frequencies. The peak sensitivity for luminance is about 10 times that for chrominance, and for a given level of sensitivity, the luminance channel can detect spatial frequencies about 10 times higher than the chrominance channel. It follows that the chrominance components of an image can be down-sampled by up to 10 times in linear dimension without perceptible degradation of image quality, provided that the luminance component is maintained. It also follows that image sharpness processing (enhancement of edges or fine detail in an image) can be performed effectively on the luminance component alone, leaving the chrominance components unchanged[7].

Proposal for a Multispectral Image Archive Format

A new technique was recently proposed[8] for encoding multispectral images, with the dual benefits of data compression (reducing the volume of data required for multispectral image archive storage) and greater processing efficiency (being better suited to image processing operations for colour reproduction). The archived image data would consist of:

- A broad-band monochromatic representation of the image at the highest available spatial resolution, designated $lum(\lambda)$, calculated as a weighted sum of the channels to give an overall spectral responsivity close to the standard spectral luminous efficiency function $V(\lambda)$;
- Two orthogonal components closely related to the opponent colour channels of the human visual system, designated $rg(\lambda)$ and $by(\lambda)$.
- A set of additional spectral components representing all of the remaining chromatic information, reduced to a significantly lower spatial resolution, with coefficients derived from principal components analysis.

Such a format would be well suited to the needs of colour image reproduction because of the reduced spatial sensitivity of the human visual system to chromatic information, as described in the previous section. For standard colorimetry, the X and Z tristimulus values could readily be calculated as linear combinations of the chromatic components. The low-resolution chromatic components could be interpolated, if necessary, for the required spatial resolution of the colorimetric image. Spatial sub-sampling of higher-order principal components has previously been tested for multispectral images in which all components were extracted by PCA[9,10].

Opponent coding serves in the human visual system as a mechanism for information compression and redundancy reduction of perceived colour information. The transformation from cone fundamentals can be parameterised in terms of λ when the expected response of the three vision channels for monochromatic stimuli is considered. The transformation can be expressed as:

$$\begin{pmatrix} lum \\ rg \\ by \end{pmatrix} = W' \begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{1}$$

where W' is defined as the transpose of the matrix of eigenvectors, and $L(\lambda)$, $M(\lambda)$ and $S(\lambda)$ are the long-, medium- and short-wavelength cone fundamentals. The Vos and Walraven (1971) cone fundamentals[11] were used in this study, as illustrated in Figure 3a, and can be obtained from the following transformations:

$$\begin{aligned} S(\lambda) &= 0.0073215\bar{z}(\lambda) \\ M(\lambda) &= -0.1551646\bar{x}(\lambda) + 0.4569237\bar{y}(\lambda) + 0.0296946\bar{z}(\lambda) \\ L(\lambda) &= 0.1551646\bar{x}(\lambda) + 0.5430763\bar{y}(\lambda) - 0.0370161\bar{z}(\lambda) \end{aligned} \tag{2}$$

where $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the CIE XYZ color matching functions modified by Judd. Note that $L(\lambda) + M(\lambda) + S(\lambda) = V(\lambda)$.

Applying singular value decomposition (SVD) to the Vos & Walraven cone fundamental functions yields the three principal components illustrated in Figure 3b, as follows:

$$\begin{pmatrix} lum \\ rg \\ by \end{pmatrix} = \begin{pmatrix} 0.887 & 0.461 & 0.0009 \\ -0.46 & 0.88 & 0.01 \\ 0.004 & 0.01 & 0.99 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{3}$$

The eigenvector transformation of $L(\lambda)$, $M(\lambda)$ and $S(\lambda)$ is unique, and has the basic property that the components $lum(\lambda)$, $rg(\lambda)$ and $by(\lambda)$ are mathematically orthogonal, i.e.

$$\int \mathrm{lum}(\lambda)\,\mathrm{rg}(\lambda)\,d\lambda = \int \mathrm{lum}(\lambda)\,\mathrm{by}(\lambda)\,d\lambda = \int \mathrm{rg}(\lambda)\,\mathrm{by}(\lambda)\,d\lambda = 0 \tag{4}$$

The expected monochromatic signal energy of the channels $lum(\lambda)$, $rg(\lambda)$ and $by(\lambda)$ has the ratio $\Gamma_1 : \Gamma_2 : \Gamma_3$ where $\Gamma_1 \geq \Gamma_2 \geq \Gamma_3$ are the eigenvalues, i.e.

$$\int \mathrm{lum}^2(\lambda)\,d\lambda : \int \mathrm{rg}^2(\lambda)\,d\lambda : \int \mathrm{yb}^2(\lambda)\,d\lambda = \Gamma_1 : \Gamma_2 : \Gamma_3 \tag{5}$$

where $lum(\lambda)$, $rg(\lambda)$ and $yb(\lambda)$ are mathematically orthogonal the eigenvalue ratio provides the optimum in signal energy compaction, corresponding to the relative power of the three components:

$$\int \mathrm{A}^2(\lambda)\,d\lambda : \int \mathrm{P}^2(\lambda)\,d\lambda : \int \mathrm{Q}^2(\lambda)\,d\lambda = \Gamma_1 : \Gamma_2 : \Gamma_3$$
$$= 97.2 : 2.78 : 0.015 \tag{6}$$

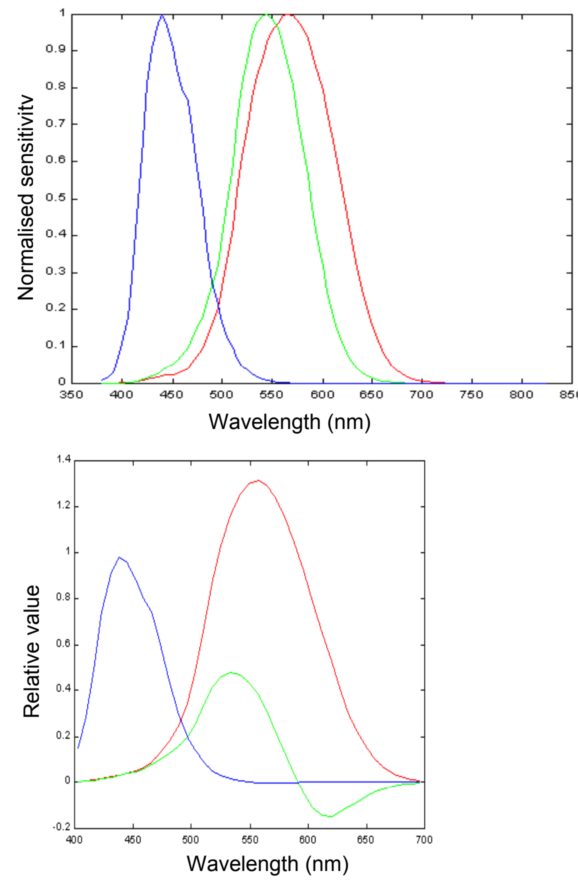


Figure 3 (Top) Vos & Walraven (1971) cone fundamentals; and (Bottom) the three orthogonal basis functions derived from them.

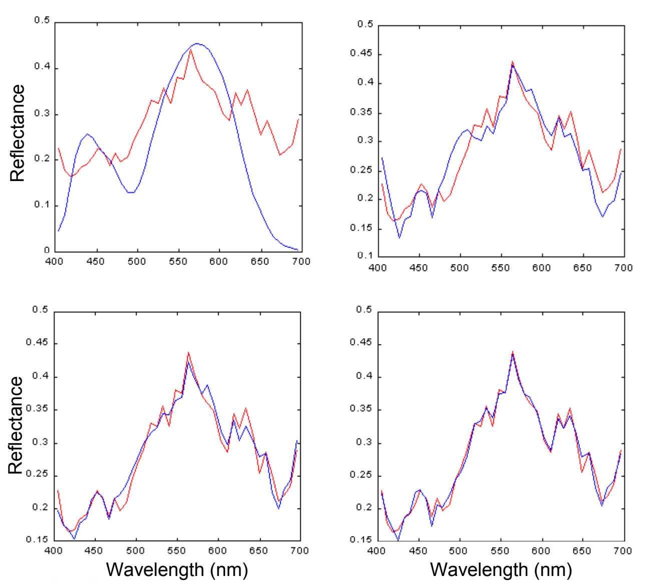
Approximation by Principal Components

A study was undertaken to determine the effectiveness of the proposed image encoding technique. A set of 22 hyperspectral images was selected from the database of natural scenes compiled by Ciao *et al*[12], consisting of 12 images of forest scenes and 10 images of coral reef scenes. Each image was of size 128x128 pixels with 40 channels at wavelength intervals of approximately 7 nm throughout the visible spectrum 403-696 nm.

Each test image was analysed to determine the best fit (by minimisation of mean-square error) to the three orthogonal basis functions defined by Eq. 3. Figure 4 (top left) shows a typical result. The approximation is reasonable in the centre of the visible spectrum but large errors occur at the ends where the cone fundamentals approach zero and the differences are less visible. The mean and maximum colour differences in CIELAB space over all pixels for ten test images were calculated as given in Table 1, indicating a reasonable accuracy of approximation, with a mean ΔE^*_{ab} error value in the range 0.5 to 1.8.

Table 1 Mean and maximum colour differences between original images and approximations with three basis functions, for ten test images.

Test image	Mean ΔE^*_{ab}	Maximum ΔE^*_{ab}
Barrine2	0.88	3.94
Coottha8	1.71	7.96
Hillside	1.59	4.51
Park2	0.85	4.95
Park4	0.89	8.06
Horshe10	1.77	4.96
Horshe29	0.70	3.98
Horshe32	0.69	4.16
Hoshoe12	0.58	4.73
Hoshoe24	0.46	2.49



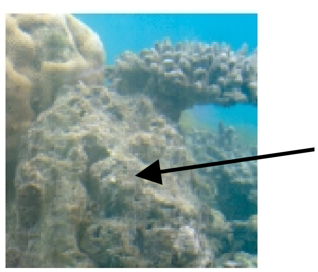


Figure 4 Spectral reflectance and best fit by three basis functions (top left) plus one (top right), two (bottom left) and three (bottom right) principal components, for one pixel at co-ordinates (64,80) of the 'Hoshoe24' test image.

Principal component analysis was then applied to the residues, i.e. the difference between the true spectral reflectance distribution at each pixel and the approximation by the three basis functions. Figure 4 shows clearly how the approximation to the original spectral reflectance distribution successively improved as additional components were included in the summation. The mean ΔE^*_{ab} colour difference reduced to a value less than 0.1 for two or more principal components (i.e. five or more basis functions in the approximation), as shown in Figure 5.

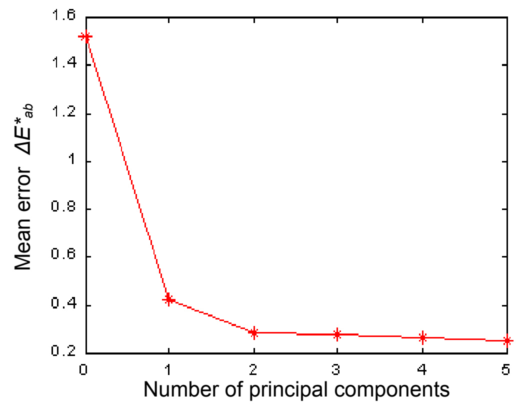


Figure 5 Mean colour difference (ΔE^*_{ab}) between original reflectance spectrum and reconstruction with three basis functions plus 0-5 principal components

Compression by Spatial Sub-Sampling

Use of 3 orthogonal basis functions (derived from the Vos & Walraven cone fundamentals) plus 3 principal components (derived from PCA of the residual differences) gave an approximation of each test image with 6 coefficients per pixel. This represented a data compression factor of 6.7 times relative to the original 40-channel hyperspectral image. If only 2 additional components were used, a compression factor of 8 would be achieved. The proposed image encoding technique was then tested by three different schemes, as set out in Table 2.

Table 2 Encoding schemes

Scheme	1st basis function	2nd, 3rd basis functions	Principal components	Compress-ion ratio
1	1:1	1:4	1:16	3.56
2	1:1	1:9	1:16	4.26
3	1:1	1:16	1:64	5.12

The first basis function (equivalent to luminance $V(\lambda)$) was not changed, but the other two basis functions and the three principal components were sub-sampled. For example in the first scheme, each 2x2 pixel block of the image channels corresponding to the 2nd and 3rd basis functions was averaged. In reconstruction, this mean pixel value was replicated in each of the four corresponding pixel positions. For each of the three principal components a 4x4 pixel block was averaged. The overall compression ratio was 1/(1+2/4+3/16) = 96/27 = 3.56.

Table 3 Mean and maximum colour differences between the uncompressed images and reconstructed images, for ten test images under three encoding schemes.

	Scheme 1		Scheme 2		Scheme 3	
Image	Mean ΔE^*_{ab}	Max ΔE^*_{ab}	Mean ΔE^*_{ab}	Max ΔE^*_{ab}	Mean ΔE^*_{ab}	Max ΔE^*_{ab}
Barrine2	4.04	41.8	5.2	45.2	6.0	62.5
Cootha8	15.2	51.9	15.3	58.4	14.9	68.0
Hillside	12.6	51.6	12.7	67.9	12.9	68.9
Park2	12.9	66.8	12.7	76.9	12.6	74.4
Park4	11.4	79.9	14.0	102.8	15.9	102.3
Horshe10	6.8	76.1	8.5	91.4	10.8	82.3
Horshe29	10.4	64.9	12.9	70.7	15.8	77.1
Horshe32	6.4	69.3	8.7	68.1	10.4	68.6
Hoshoe12	6.9	59.9	9.2	71.4	10.9	72.9
Hoshoe24	5.8	43.7	6.3	60.4	7.6	64.6

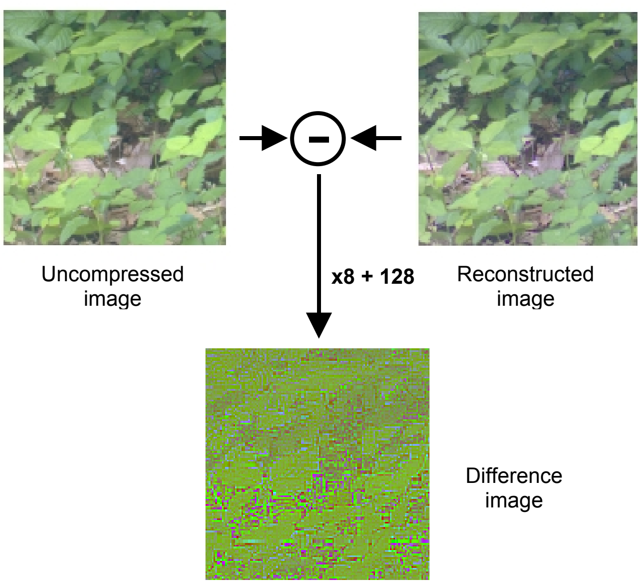


Figure 6 Enhanced difference between an uncompressed and a reconstructed image under Encoding Scheme 1, for the image 'Park2'.

Colour differences between the original and reconstructed images under the three encoding schemes are given in Table 3. The mean errors were relatively large, in the range 4-15 ΔE^*_{ab}, and increased with larger compression ratios. The errors were evenly distributed throughout the image and gave the appearance of high frequency pixel noise. Figure 6 shows an example for the image Park2, in which the values of the pixel differences between the uncompressed image (six-component approximation) and the image encoded and reconstructed under Scheme 1 are multiplied by 8 and added to 128 to make them clearer.

Psychophysical Experiments

To evaluate image quality, two psychophysical experiments were performed, using side-by-side forced-choice comparison of image pairs on a Barco *Calibrator V* CRT display. Image pixel data were converted from spectral reflectance to *XYZ* via the CIE 2° colour matching functions, thence to display *RGB* via the GOG model[13] to ensure colour-accurate rendering of the images. All images were doubled in size to 256x256, using spatial interpolation, to facilitate visual comparison. The experiments were performed in a dark room at normal viewing distance (approximately 60 cm). Eight observers participated, and were instructed to say which image in each pair was of higher quality.

In the first experiment, five versions of each of 10 test images were assessed: the original and the representations using 3 orthogonal basis functions plus 0,1,3 and 5 principal components. This resulted in 100 image pairs per observer. The results are given in Table 4 as mean *z*-scores, averaged over all 8 observers and 10 test images. The *z*-scores are small, resulting from the visual similarity of all the images – differences between the various image versions in this experiment were not easily perceived. Curiously the representation with 3 principal components was judged to be better than that with 5, suggesting that the additional two components carry more noise than information (see also Figure 5).

Table 4 Results of first experiment: mean z-scores and rank.

	Original	3B	3B-1PC	3B-3PC	3B-5PC
MEAN	0.135	-0.072	-0.044	0.050	-0.069
RANK	1	5	3	2	4

Table 5 Results of second experiment: mean z-scores and rank.

	3B-3PC	3B-3PC Scheme 1	3B-3PC Scheme 2	3B-3PC Scheme 3
MEAN	1.365	0.659	-0.214	-1.810
RANK	1	2	3	4

In the second experiment, four versions of each of 10 test images were assessed: the uncompressed representation using 3 orthogonal basis functions plus 3 principal components and the reconstructed images using the three encoding schemes. This resulted in 60 image pairs per observer. The results are given in Table 5 as mean *z*-scores, averaged over all 8 observers and 10 test images. The *z*-scores are larger in this case, indicating greater visual differences between the images. Similar results were evident for all the individual images. The ranking followed the same order as for the magnitude of colour differences of the three encoding schemes, as given in Table 3.

Conclusions

The study demonstrated that it is feasible to encode multispectral images using three orthogonal basis functions derived from human cone fundamentals, plus additional principal components. This is an important conclusion, because it provides a method of multispectral image representation that is 'upward compatible' with conventional trichromatic encoding. The standard colorimetric image, such as *XYZ*, is easily extracted from the first three channels of the encoded image. The experiments showed that very good colorimetric accuracy could be achieved for multispectral images of natural scenes with only two or three additional components derived by PCA of the residues.

The study also showed the feasibility of compression of the encoded image by spatial sub-sampling of all components except the first. The reconstructed images were of reasonable visual quality, although artefacts arising from the process were visible. Further study is required with more sophisticated sampling methods (e.g. use of *sinc* functions for decimation and bicubic interpolation for reconstruction) in order to assess the capabilities of the method. Compression ratios of 6-8 were achieved for the encoding and 3-5 for the spatial sub-sampling, yielding overall image file compression ratios of 20-40.

Acknowledgements

The investigation was strengthened through discussions with our colleague Dr Mitch Thomson at CII. We are also indebted to Dr Markku Hauta-Kasari at the University of Joensuu, Finland, for providing the hyperspectral test images and for helpful advice.

References

1. Hubel, P.M., Holm, J., and Finlayson, G., 'Illumination estimation and colour correction', *Colour Imaging: Vision and Technology,* Ed. MacDonald L.W. and Luo M.R., John Wiley, 1999, pp. 73-95.
2. Carlson, G.E., *Signal and linear system analysis*, John Wiley, 1998.
3. Buchsbaum, G., and Gottschalk, A. 'Chromaticity co-ordinates of frequency-limited functions', *JOSA A,* **1**(8), 885-887, 1984.
4. Hunt, R.W.G., *The Reproduction of Colour*, 5th Edition, Fountain Press, Kingston-upon-Thames, UK, 1995, p. 126 and p. 178.
5. Westland, S., Shaw, J., and Owens, H.C., 'Colour statistics of natural and man-made surfaces', *Sensor Review*, **20**(1), 50-55, 2000.
6. Fairchild, M.D., *Color Appearance Models*, Addison-Wesley, Reading MA, 1998, pp. 30-32
7. MacDonald, L.W., 'Framework for an image sharpness management system', *Proc. 7th IS&T/SID Color Imaging Conf.*, 75-79, 1999.
8. MacDonald, L.W., Westland, S., and Shaw, J., 'Colour Image Reproduction: Spectral vs Spatial', *Proc. Intl. Symp. on Multispectral Imaging*, Chiba University, Japan, 81-91, 1999.
9. Parkinnen, J., Hauta-Kasari, M., Kaarna, A., Lehtonen, J., Koponen, P., and Jaaskelainen, T., 'Multispectral Image Compression', *Proc. 2nd Intl. Symp. on Multispectral Imaging*, Chiba University, Japan, 49-58, 2000.
10. König, F., and Praefcke, W., 'Multispectral image encoding', *Proc. IEEE Intl. Conf. on Image Processing*, **3**, Kobe, Japan, 45-49, 1999.
11. Vos, J.J., and Walraven, P.L., 'On the derivation of the foveal receptor primaries', *Vision Res.,* **11**, 799-818, 1971.
12. Chiao, C.C., Cronin, T.W., and Osorio, D., 'Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants', *JOSA A*, **17**, 218-224, 2000.
13. Berns, R.S, Methods for characterizing CRT displays, *Displays*, **16**, 173-182, 1995.

Biographies

Lindsay MacDonald is Professor of Multimedia Imaging at the Colour & Imaging Institute, University of Derby, UK. He received his B.Sc. and B.Eng.(Hons) degrees from the University of Sydney, Australia, in 1970 and 1972. He worked for 18 years 1977-1995 for Crosfield Electronics (UK) developing computer-based imaging systems for the graphic arts industry, for which he holds 15 patents. He is the co-author of 7 books on colour imaging and has been closely associated with the Color Imaging Conference since its inception, including roles as Technical Program Co-Chair in 1996 and General Co-Chair in 1997.

Stephen Westland is Reader in Colour Imaging at the Colour & Imaging Institute. He obtained B.Sc. and Ph.D. degrees from Leeds University, UK, in 1983 and 1988. He developed on-line colour monitoring systems for Courtaulds Research 1996-1990 and was a lecturer in Neurocomputing at Keele University 1990-1999. His research interests include colour chemistry and measurement, colour and spatial vision, and artificial neural networks.

Dongmei Liu was a student 2000-2001 at the Colour & Imaging Institute, and undertook this study as her major project in the M.Sc. Colour Imaging programme. She now works as a software design engineer for Inca Digital Printers Ltd. in Cambridge, UK.