

Compression of Reflectance Data Using An Evolved Spectral Correlation Profile

Peter Morovič†, Ján Morovič†, Michael H. Brill‡, Eric Walowitz¶, †Hewlett Packard Company, Spain; ‡Datacolor, USA; ¶USA

Abstract

Since spectral data is significantly higher-dimensional than colorimetric data, the choice of operating in a spectral domain brings memory, storage and computational throughput hits with it. While spectral compression techniques exist, e.g., on the basis of Multivariate Analysis (mainly Principal Component Analysis and related methods), they result in representations of spectra that no longer have a direct physical meaning in that their individual values no longer directly express properties at a specific wavelength interval. As a result, such compressed spectral data is not suitable for direct application of physically meaningful computation and analysis. The framework presented here is an evolution and extension of the spectral correlation profile published before. It is a simple model, driven by a few adjustable parameters, that allows for the generation of nearly arbitrary, but physically realistic, spectra that can be computed efficiently, and are useful over a wide range of conditions. A practical application of its principles then includes a spectral compression approach that relies on discarding spectral wavelengths that are most redundant, given correlation to their neighbors. The goodness of representing realistic spectra is evaluated using the MIPE metric as applied to the SOCS and other databases as a reference. The end result is an efficient, yet physically meaningful, compressed spectral representation that benefits computation, transmission and storage of spectral content.

Introduction

Ultimately spectral data - be it reflectance, transmission or power - is at the basis of all color and imaging research and appli-

cations. Even where other types of data are processed, the stimuli they refer to have certain spectral properties that are among their causes and in many cases access to the spectral data itself is needed to address certain challenges or provide certain solutions. Wherever multiple illuminants are at play, multiple observers (either human or machine) or multiple materials are involved, access to spectral data becomes indispensable for the development of robust solutions or for an accurate characterization of an imaging or viewing setup.

Since spectral data is higher dimensional than colorimetric data, any computation that needs to be performed in a spectral domain starts off with a challenge derived from the need of having it applied to vectors with a greater number of members and therefore having to be executed a greater number of times and requiring more operating memory as well as storage.

It is, however, well known that spectral reflectances do not have a dimensionality that matches the number of spectral samples that were used when it was measured or synthesized using some model. Instead of the 31 or even 16 spectral samples typical in imaging applications (corresponding to a 400nm to 700nm range at 10nm and 20nm uniform sampling steps respectively), naturally-occurring spectra tend to have properties that allow for the vast majority of their variance to be accountable for using 3 to 8 basis functions (Krinov, 1947; Vrhel *et al.*, 1994), obtained using Principal Component Analysis (PCA) or Characteristic Vector Analysis (CVA). In other words, linear combinations of 3-8 basic "spectra" can result in approximations that are very close to measurements taken at 16 or even 31 wavelengths.

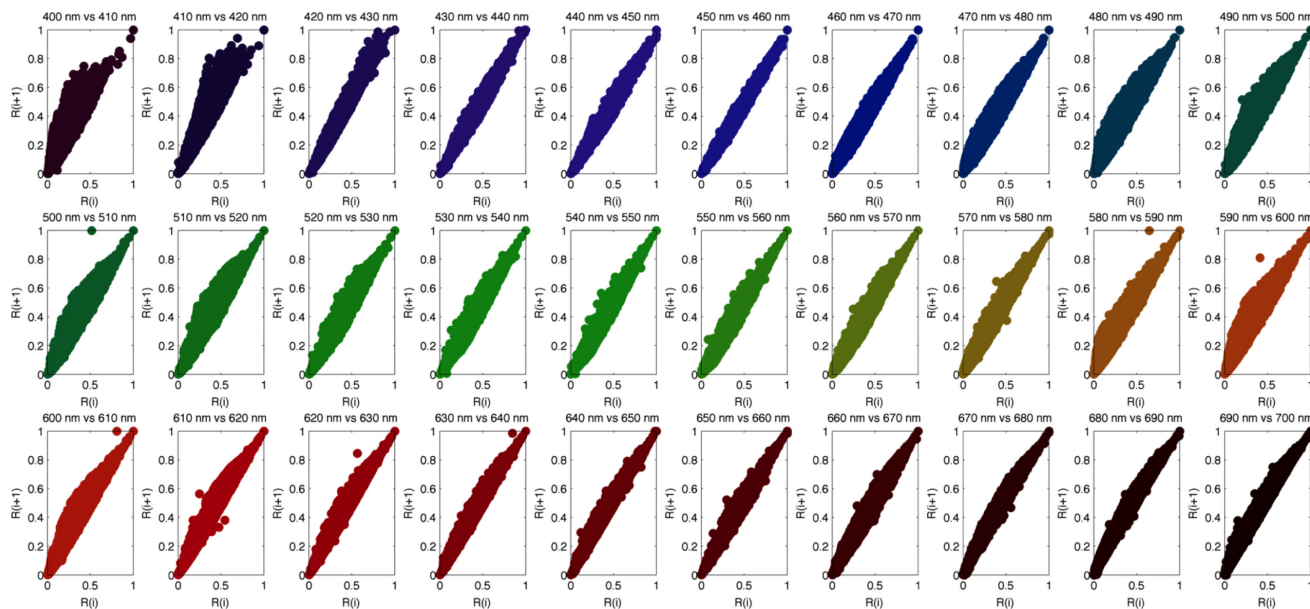


Figure 1: Example of the Relative Spectral Neighborhood Range Profile for the SOCS data set at 31 samples.

While the use of methods like PCA is well established and provides a successful means for compressing spectral data (of which a familiar form is dimensionality reduction), it does so at two costs. First, the representation of spectra as weights in a PCA basis strips them of their range, as was discussed previously (Morović *et al.*, 2013). Second, the basis weights have no direct physical meaning and it is therefore not possible to perform direct, physically-derived analysis or computation on them.

What is meant here by ‘physical meaningfulness’ is that, while a reflectance or spectrum represented as “some values at some sampled wavelengths” has a direct physical meaning, where a given value at a given wavelength expresses the properties of a surface in relation to radiation at that wavelength, PCA weights do not have such direct relationships to physical properties. As a consequence, operations that can be performed on directly measured or synthesized spectral data (e.g., wavelength-independent processing, but also analysis and visualization), cannot be applied to the weights of a PCA-compressed representation.

What would be needed instead would be compressed representation of spectra that, in addition to preserving ranges, would also retain the physical meaning of its values. This would allow for spectral processing to be performed as on a full spectral representation, but only a smaller number of times. The result would be a lower memory and storage footprint, faster processing speed, and the ability to easily generate physically meaningful spectra over a potentially arbitrary range of conditions.

To arrive at such a transformation of spectra, an evolution of the spectral correlation profile (Morović *et al.*, 2013) will be introduced here, which takes advantage of the insight that there is high correlation between neighboring wavelength values in physically-realized spectra (Figure 1). Characterizing the specific correlation profile of a set of data then provides an ordering of the full set of wavelengths according to their similarity to their neighbors and therefore according to how disposable they are.

Spectral Neighborhood Profiles

While the concept of a spectral correlation profile was introduced at last year’s CIC (Morović *et al.*, 2013), the priority there was to explore the framework intuitively, while at the same time providing a taster of the kind of analysis it enables. To make further progress and take greater advantage of the concept’s potential, this section will present a more formal exposition of what will now be referred to as a Spectral Neighborhood Profile (SNP), since it represents the properties of relationships between reflectances from neighboring wavelength intervals.

Beyond formalism, this section will also introduce three new types of SNPs. The first will be the Relative Spectral Neighborhood Range Profile (rSNRP), which expresses the degree of greatest increase and decrease of one interval’s reflectances versus its neighbor (this is what the CIC21 paper labeled “spectral correlation profile”). Second, the absolute values of reflectances within wavelength intervals will be used to yield an Absolute Spectral Neighborhood Range Profile (aSNRP), which will have the further benefit of more closely tracking a data set’s spectral domain. Third, it will not only be ranges, but the specific distributions within them that the profile will express, resulting in the Spectral Neighborhood Distribution Profile (SNDP). Finally, the fourth variant of the SNP will be the Spectral Neighborhood Correlation Profile (SNCP), where it will be Pearson product-moment correlation coefficients that will be computed between the reflectances of neighboring wavelength intervals. The following section will then look at applications of the profiles presented here.

Before proceeding to the details of the alternative spectral neighborhood profiles, it is worth noting that each of them can be thought of a kind of spectral reflectance compression – either at the level of individual reflectances (e.g., the application of the SNDP) or at the level of sets of reflectances (applicable to all four alternatives).

Relative Spectral Neighborhood Range Profile

Given a set of M reflectances $\mathcal{S}$, represented by N equal-interval spectral samples, such that $\mathcal{S}$ can be written as an $M \times N$ matrix, its Relative Spectral Neighborhood Range Profile is defined as a pair of $(N-1)$ – vectors $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ such that at each wavelength λ_i :

$$\mathbf{c}_{\min}(i) = \text{MIN}_{j=1:M} \mathbf{S}(j, i) - \mathbf{S}(j, i+1) \quad (1)$$

$$\mathbf{c}_{\max}(i) = \text{MAX}_{j=1:M} \mathbf{S}(j, i) - \mathbf{S}(j, i+1) \quad (2)$$

Where $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ are the per-neighboring-wavelength minima and maxima of differences between the reflectances at neighboring wavelength samples (λ_i and λ_{i+1} indexed by i and $i+1$ in $\mathcal{S}$) over the entire set $\mathcal{S}$. The index (i) here refers to the range between λ_i and λ_{i+1} or $[i, i+1]$ in matrix indices terms. Here $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ are the lower bound and upper bound of differences. If $\mathbf{c}_{\min}(i)$ is negative it means that there is at least one case where the reflectance is increasing, i.e. that at λ_{i+1} it is bigger than λ_i , while if it is positive it means all reflectances are decreasing between λ_i and λ_{i+1} and likewise for $\mathbf{c}_{\max}$.

Then, a $1 \times N$ reflectance vector $\mathbf{s}$ is said to satisfy the rSNRP profile $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ if for all neighboring wavelengths λ_i and λ_{i+1} the following inequalities hold:

$$\mathbf{c}_{\min}(i) \leq \mathbf{s}(i) - \mathbf{s}(i+1) \leq \mathbf{c}_{\max}(i) \quad (3)$$

Given the rSNRP as defined above, reflectances can be generated synthetically to satisfy Eq (3) by constructing them progressively. Let us assume a synthetic reflectance $\mathbf{q}$ has already been constructed at wavelengths λ_1 to λ_{i-1} , then the value $\mathbf{q}(\lambda_i)$ has to satisfy:

$$\mathbf{q}(i+1) \leq \mathbf{q}(i) - \mathbf{c}_{\min}(i) \quad (4)$$

$$\mathbf{q}(i+1) \geq \mathbf{q}(i) - \mathbf{c}_{\max}(i) \quad (5)$$

The spectral envelope (min/max values per wavelength) cases above result in two branches of the generation process – $\mathbf{q} \rightarrow [\mathbf{q}^1, \mathbf{q}^2] \rightarrow [\mathbf{q}^{11}, \mathbf{q}^{12}, \mathbf{q}^{21}, \mathbf{q}^{22}] \rightarrow \dots$ – one that is at the upper limit of the rSNRP (the max value at i from $\mathbf{c}_{\max}$) and one that is at the lower limit (the min value i from $\mathbf{c}_{\min}$), as:

$$\mathbf{q}^1(i) = \mathbf{q}(i-1) - \mathbf{c}_{\max}(i-1) \quad (6)$$

$$\mathbf{q}^2(i) = \mathbf{q}(i-1) - \mathbf{c}_{\min}(i-1) \quad (7)$$

Starting at an arbitrary or random value $\mathbf{q}(1)$ at λ_1 and repeating the above process, branching at each consecutive λ_i , generates 2^{N-1} (e.g. for $N=16$ where $\lambda_1 = 400 \text{ nm}$ and $\lambda_N = 700 \text{ nm}$ and $(\lambda_{i+1} - \lambda_i) = 20 \text{ nm}$, this is 32768) samples that envelope the possible values while satisfying the rSNRP. Figure 2 illustrates this generative process.

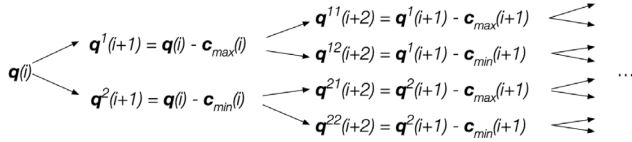


Figure 2. Generating reflectances using an rSNRP.

The two vectors $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ defined in Eq. (1) and (2) then jointly define the rSNRP - a characterization of the relative range of neighboring values - as it was introduced in CIC21, where it was referred to as the “spectral correlation profile”. In the present paper we extend this definition to a broader, more general form that allows for further analysis and synthesis in the spectral domain.

Absolute Spectral Neighborhood Range Profile

The spectral neighborhood profile as outlined so far is relative, as it does not consider the offset of actual values at any one wavelength. To represent the range of data, the profile can be defined in absolute terms such that an additional two (N-1)-vectors $\mathbf{v}_{\min}$ and $\mathbf{v}_{\max}$ (which in the relative case are, implicitly, all at 0 and 1 respectively) are computed in the process of building the correlation profile. These will simply include the minimum and maximum reflectance value at any wavelength λ_i over the whole set $\mathcal{S}$, and can be defined as follows:

$$\mathbf{v}_{\min}(i) = \text{MIN}_{j=1:M} \mathbf{S}(j, i) \quad (8)$$

$$\mathbf{v}_{\max}(i) = \text{MAX}_{j=1:M} \mathbf{S}(j, i) \quad (9)$$

Then, the two pairs of vectors, $[\mathbf{c}_{\min} \ \mathbf{c}_{\max}]$ and $[\mathbf{v}_{\min} \ \mathbf{v}_{\max}]$ jointly extend the spectral neighborhood profile and define an *absolute* variant - aSNRP. This in turn allows for the synthesis of spectral data that mimics both the individual wavelength value ranges as well as the cross-wavelength difference ranges.

Like in the case of the rSNRP the aSNRP can equally be used to synthesize spectral data, starting at $\mathbf{q}^1(1)$ and $\mathbf{q}^2(1)$ with values determined from $\mathbf{v}_{\min}(1)$ and $\mathbf{v}_{\max}(1)$ and generating subsequent steps as before according to equations (6) and (7) and the schematic in Figure 2 with the additional constraint of checking the bounds of each subsequent $\mathbf{q}(i)$ to make sure that at wavelength λ_i the range of $\mathbf{v}_{\min}(i)$ and $\mathbf{v}_{\max}(i)$ is not exceeded.

Spectral Neighborhood Distribution Profile

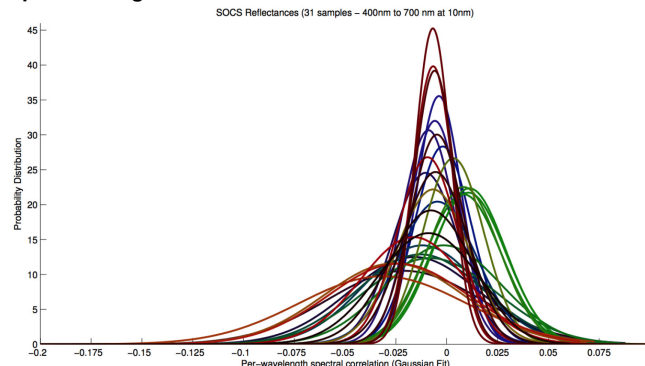


Figure 3. Example of a spectral neighborhood distribution profile for the SOCS data set (31-sample).

The next extension to the concept of the spectral neighborhood profile looks at the distributions of absolute differences across wavelengths (Figure 3). This is useful since the $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ ranges may be large but need not be indicative of the distribution, which is likely to be non-uniform. Thus for the above data set $\mathcal{S}$ we can compute N-1 distributions - for example approximated by a normal distribution - that at each wavelength are defined by the mean and the standard deviation of the pair of wavelengths under scrutiny. Let us denote these as the two (N-1)-vectors $\mathbf{d}_{\mu}$ and $\mathbf{d}_{\sigma}$ such that:

$$\mathbf{d}_{\mu}(i) = \text{MEAN}_{j=1:M} \mathbf{S}(j, i) - \mathbf{S}(j, i+1) \quad (10)$$

$$\mathbf{d}_{\sigma}(i) = \text{STDD}_{j=1:M} \mathbf{S}(j, i) - \mathbf{S}(j, i+1) \quad (11)$$

Extending the definition of the spectral neighborhood profile in this way enables a statistical synthesis of reflectances by drawing samples from the respective distributions at any given wavelength. Not only will such synthesis result in new data that is consistent in terms of the range of values and the behavior at neighboring wavelengths, but also in terms of the probability of values at any wavelength.

Overall, the sequence of the above descriptive characteristics of spectral data can be thought of as sequentially incremental: the simplest and loosest descriptor being the relative SNRP as it disregards absolute values of the data, followed by the absolute SNRP which then takes this into account and finally the distribution based profile which takes both the location, spread and frequency of the neighboring wavelengths into account. The next characterization does not follow this sequence but rather is a simplification of the distribution based normal approximation (SNDP).

Spectral Neighborhood Correlation Profile

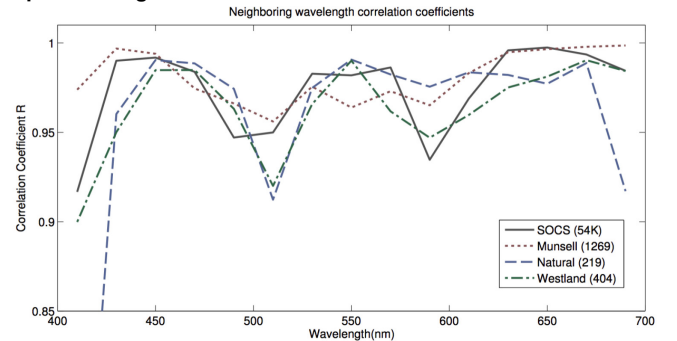


Figure 4. Spectral correlation coefficients for neighboring wavelengths (at wavelength $(\lambda_i + \lambda_{i+1})/2$ as the correlation coefficient for λ reflectance data sets.

While the term *correlation* has been used in a previous instantiation of this approach, that was only done loosely and its strict application is presented next. For a data set whose above defined spectral neighborhood profile results in small per-neighboring-wavelength ranges, its correlation coefficients will be high. The converse however is not true: for large ranges correlation coefficients can be small if the distribution is narrow and the range is wide due to outliers, or they can be big if the distribution represents a spread in the data.

Hence, another view of the same spectral data $\mathcal{S}$ is to compute the correlation coefficients of neighboring wavelengths. The corre-

lation coefficient is commonly denoted as r and is computed as follows:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (12)$$

Here too a $(N-1)$ -vector r is computed and expresses the degree to which neighboring wavelengths are correlated rather than the range or distribution of values these can take as has been done so far. Figure 4 is an example plot computed from a variety of data sets at 16 spectral samples: SOCS (Tajima *et al.*, 2002), Munsell (Parkkinen *et al.*, 1989), Natural (Krinov, 1947), Westland *et al.* (2000).

While in absolute terms all values are high, meaning neighboring wavelengths are correlated to a high degree, there are significant relative differences. For example in the SOCS data set there is a low of 0.9167 between 400 and 420nm and a high of 0.9975 between 640 and 660nm. The ‘Natural’ data set is poorly correlated at the beginning of the spectral range (0.6562), while the ‘Munsell’ data set reaches 0.9986 at the long end of the wavelength range.

While this last characterization of the spectral data set is not directly suitable for synthesis of new data it still expresses an intrinsic property of the data that can be used to determine redundancy present in the data, as outlined next.

Spectral Sampling Analysis

The spectral correlation coefficient analysis introduced above lends itself directly for an analysis of the optimal spectral sampling of reflectances (e.g., Becker *et al.*, 2005).

Given a dataset, its analysis follows these steps:

1. Compute the spectral correlation coefficients for all neighboring wavelengths of the entire data set.
2. Assuming a spectral sampling of N equidistant wavelength samples, for $i=1$ to $N-1$, pick progressively the i wavelengths at which correlation is lowest.
3. For each of the $N-1$ choices of i wavelengths perform the following analysis of precision:
 1. Interpolate the value at the wavelengths i of the entire data set. This is the non-uniform wavelength sampling representation of the data set.

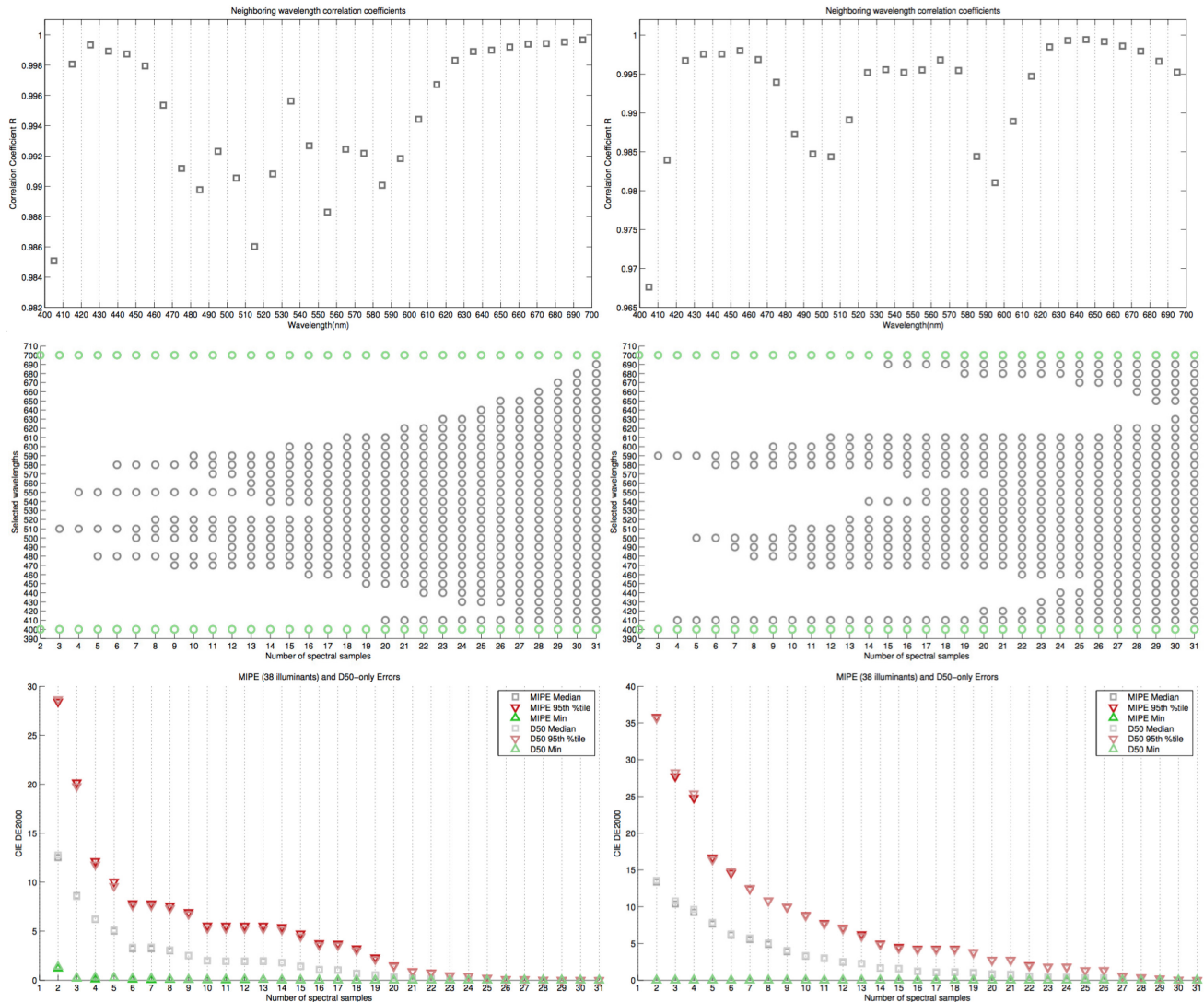


Figure 5: Spectral correlation coefficients (top); progressive wavelength sampling selection (middle); MIPE metric for all spectral sampling representations from 3 to 31 (bottom) for the Munsell data set of 1269 samples (left column) and the SOCS data set of 54K samples (right column).

2. Extrapolate from the above the full, original N -equi-distant wavelength sampling.
4. Compare the original data set at its N -sample representation and the lower and non-uniformly sampled alternatives in terms of their MIPES, whose median ($MIPE_{MED}$) is computed as follows (Morović et al., 2012):

$$MIPE_{MED} = \text{median}_{i=1, s=1}^{i=173, s=n} (\Delta E_{2000}(O_{i,s}, R_{i,s})) \quad (13)$$

where n is the number of reflectances viewed under 173 light sources (Hordley, 2001), $O_{i,s}$ is the s -th uncompressed reflectance, viewed under light source i and $R_{i,s}$ is the corresponding compressed reflectance.

In the above process the two extreme values (e.g. at 400nm and 700nm) are kept by default and only intermediate samples are evaluated. This allows for interpolation even with a single sample (plus the two extremes) and avoids the issue of correlation ambiguity at the edges of the range.

An example of this process and its partial results is shown in Figure 5. Here, three views are presented: first the neighboring wavelength correlation coefficients for each data set (top), then the progressive addition of wavelengths in order of lowest to highest correlation coefficient (middle) with 400nm and 700nm samples added from the beginning and finally the MIPE metric evaluation for each spectral sample number representation in turn (bottom).

It can be readily seen that there is significant room for maneuver and that wavelengths are not created equal. In the 1269 samples from the Munsell book of color (Figure 5 left), of the 31 wavelength samples used to represent it, around 10 can be safely sacrificed without significant impact on accuracy. What the analysis also allows for is to parameterize this, depending on the application and the precision requirements. For example in some cases it may be acceptable to have a 95th %tile error of $5 \Delta E$ in which case 15 wavelength samples are already sufficient.

Using the results from Figure 5, it is then a matter of choice to determine what precision is desired and sufficient and to use the corresponding spectral sampling vector instead of the full uniform-

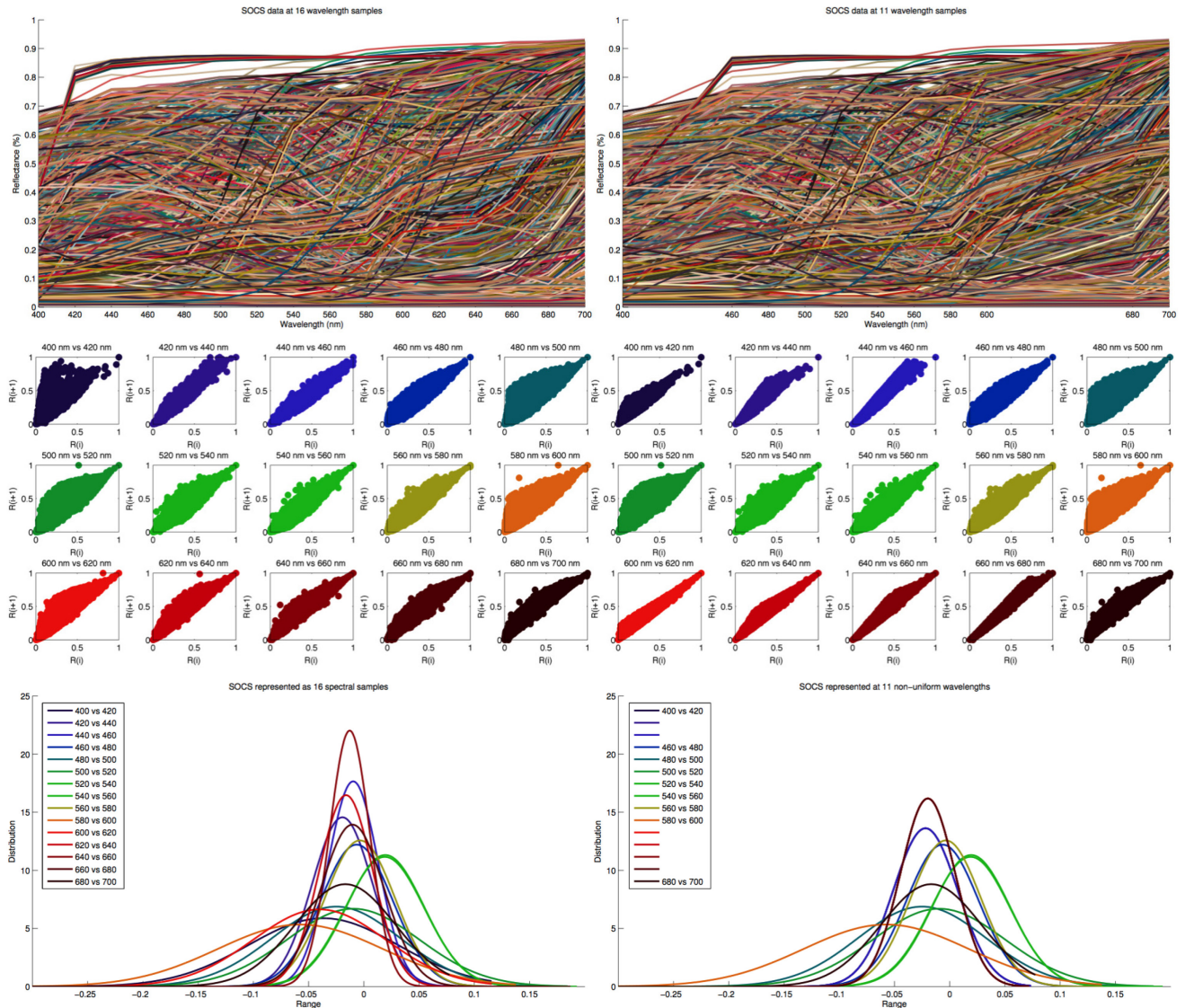


Figure 6: Original (16-sample) reflectance plot and 11-sample reduced reflectance plot (top left and right), $rSNRP$ for the original 16-sample data and the $rSNRP$ for the 11-sample reduced data (middle left and right) and the $SNDP$ view of the same 16-sample and 11-sample data (bottom left and right).

ly sampled data resolution. To verify that using this approach the resulting reflectance data maintains the neighborhood relations such as the relative Spectral Neighborhood Range Profile and the Spectral Neighborhood Distribution Profile, Figure 6 shows the results by performing this analysis starting with a 16-sample representation of the SOCS data set and then comparing it against the best 11-sample representation using non-uniform spectral samples.

The first two plots in Figure 6 show how between the selected samples the variation is minimal is only down to the interpolation between the wavelengths that were not left out in the sampling selection process. This is also mimicked in the rSNRP plot where the wavelength neighborhoods that have been kept (all but 5 in the above case) are identical, while the remaining ones show narrow and synthetic behavior (see 400 vs 420nm, 420 vs 440nm and 440 vs 460nm, 600 vs 620nm etc. in the right-hand-side of Figure 6, compared against the same on the left-hand side), again due to interpolation. Finally in the SNRP plots at the bottom of Figure 6 this is apparent by those neighborhoods that are present being identical, with high and narrow distributions (high correlation) missing in the right-hand-side plot of distributions again as expected,

Conclusions

The specific nature of relationships between reflectances from neighboring wavelength intervals is an inherent characteristic of a reflectance spectrum or of a set of reflectance spectra. By extending the simple notion of neighboring interval difference ranges presented previously, the approaches presented here allow not only for gamut-aware synthesis but also for a synthesis that preserves an original data set's difference distributions (via the Spectral Neighborhood Distribution Profile) and for a dimensionality reduction that takes advantage of highly correlated neighbors (via the Spectral Neighborhood Correlation Profile).

The above dimensionality reduction was shown to perform accurately for a number of data sets, as evaluated using the MIPE metric. Furthermore, the result is a physically-meaningful, compressed representation that is suitable for having further computation performed on it much the same way as with canonical equidistant wavelength sample representations, unlike the weights of multivariate analysis bases.

As such the spectral neighborhood based techniques introduced here are a useful extension to the existing methods available for spectral analysis and processing and provide a good starting point to revisiting various applications in this field in the future. An example here would be the choice of spectral data used for camera characterization, which the authors intend to explore next.

References

- Becker B. L., Lush D. P., Qi J. (2005) Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis, *Remote Sensing of Environment*, vol. 97, no. 2, pp. 238–248.
- Hordley S. D. (2001) *Colour by Correlation*, PhD Thesis, School of Computing Science, University of East Anglia.
- Krino E. L. (1947) Spectral Reflectance Properties of Natural Formations, *Proceedings of the Academy of Sciences of the USSR*.
- Morovič P., Morovič J., García-Reyero J. M. (2013) Spectra From Correlation, 21st IS&T Color and Imaging Conference, Albuquerque, NM, pp. 20–26.
- Morovič P., Morovič J., Arnabat J., García-Reyero J. M. (2012) Revisiting Spectral Printing: A Data Driven Approach, IS&T/SID 20th Color and Imaging Conference, Los Angeles, CA, pp. 335–340.

- Parkkinen J. P. S., Hallikainen J., Jaaskelainen T. (1989). Characteristic spectra of Munsell colors. *JOSA A*, vol. 6, pp. 318–322.
- Tajima J., Haneishi H., Ojima N., Tsukada M. (2002) Representative Data Selection for Standard Object Colour Spectra Database (SOCS), 10th Color Imaging Conference, pp. 155–160.
- Vrhel M. J., Gershon R., Iwan L. S. (1994) Measurement and analysis of object reflectance spectra, *Color Research and Application*, vol. 19, no. 1, pp. 4–9.
- Westland S., Shaw J., Owens H. (2000). Colour statistics of natural and man-made surfaces. *Sensor Review*, vol. 20, no. 1, pp. 50–55.

Author Biographies

Peter Morovič received his Ph.D. from the University of East Anglia, UK in 2002 and holds a B.Sc. in theoretical computer science from Comenius University, Slovakia. He has been a senior color and imaging scientist at Hewlett-Packard in Barcelona since 2007. His research interests include color reproduction, vision, image processing and computational geometry; he has 50+ refereed publications and holds 50+ US patents (pending or granted); he received the IS&T's Raymond Davis Scholarship (2001) and the JSPS Post-doctoral Fellowship (2005).

Ján Morovič received his Ph.D. in color science from the Colour & Imaging Institute, University of Derby (UK) in 1998. After working there as a lecturer in digital color reproduction, he became senior color scientist and later master technologist at Hewlett-Packard in Barcelona, where he has been since 2003. He has also served as the director of CIE Division 8 on Image Technology and Wiley and Sons have published his 'Color Gamut Mapping' book.

Michael H. Brill is Director of Research at Datacolor in Lawrenceville, NJ. A Ph.D. in physics at Syracuse University, he has done theoretical research on color in human and computer vision, geometric/photo-metric invariance, and physics-based vision. He obtained the 1996 ISCC Macbeth Award for work on color constancy and the 2010 ISCC Nickerson Service Award. He is on the Editorial Board of Color Research and Application, and chairs ASTM Subcommittee E12.04 (Color and Appearance Analysis).

Eric Walowitz's interest is in appearance estimation, color management, and digital photography. He is founder (retired) of Color Savvy Systems, a color management company. He currently helps friends and colleagues with interesting color problems and ventures on a friendly basis. He graduated from RIT's Image Science program in 1985, concentrating in Color Science. He has authored more than 50 patents, publications, and presentations. Eric is a member of ICC, ISOTC42, IST, and PMA.