

# Spectral Compression: Weighted Principal Component Analysis versus Weighted Least Squares

Farnaz Agahian<sup>a\*</sup>, Brian Funt<sup>a</sup>, Seyed Hossein Amirshahi<sup>b</sup>

<sup>a</sup>Simon Fraser University, 8888 University Dr. V5A 1S6, B.C., Canada; <sup>b</sup>Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran

## ABSTRACT

Two weighted compression schemes, Weighted Least Squares (wLS) and Weighted Principal Component Analysis (wPCA), are compared by considering their performance in minimizing both spectral and colorimetric errors of reconstructed reflectance spectra. A comparison is also made among seven different weighting functions incorporated into ordinary PCA/LS to give selectively more importance to the wavelengths that correspond to higher sensitivity in the human visual system. Weighted compression is performed on reflectance spectra of 3219 colored samples (including Munsell and NCS data) and spectral and colorimetric errors are calculated in terms of CIEDE2000 and root mean square errors. The results obtained indicate that wLS outperforms wPCA in weighted compression with more than three basis vectors. Weighting functions based on the diagonal of Cohen's  $\mathbf{R}$  matrix lead to the best reproduction of color information under both A and D65 illuminants particularly when using a low number of basis vectors.

**Keywords:** Spectral compression, weighted principal component analysis, weighted least squares

## 1. INTRODUCTION

Principal Component Analysis (PCA) is one of the most widely used techniques in compression of large spectral images and guarantees the best possible representation of the high-dimensional spectra in a low-dimensional eigenvector subspace. This compression method gives an equal treatment to all wavelengths throughout the spectrum and tries to minimize the squared reconstruction errors between the actual and reconstructed spectra:

$$\mathbf{e} = \sum_{\lambda=380}^{780} (\mathbf{R}_{\lambda} - \hat{\mathbf{R}}_{\lambda})^2 \longrightarrow \text{Min} \quad (1)$$

where  $\mathbf{R}_{\lambda}$  is the actual and  $\hat{\mathbf{R}}_{\lambda}$  is the reconstructed reflectance spectrum.

On the other hand, in many applications such as digital image archives, electronic commerce, art conservation science, and telemedicine, it is necessary to compress the spectral data to preserve as much color information as possible. Indeed, to be able to accurately reproduce color under either one or multiple lighting and viewing conditions, the reconstructed spectra must be recovered in some parts of the spectrum more precisely than the others. In the other words, the compression technique should be modified so as to consider individual wavelengths differently depending on their relative importance to the human visual system.

During the last decade, this observation has been developed by extending standard PCA into a weighted version by introducing weights into Eq (1) and minimizing the weighted squared reconstruction error instead<sup>1,2</sup>:

$$\mathbf{e} = \sum_{\lambda=380}^{780} w_{\lambda} (\mathbf{R}_{\lambda} - \hat{\mathbf{R}}_{\lambda})^2 \longrightarrow \text{Min} \quad (2)$$

\*fagahian@sfu.ca

where  $\mathbf{w}_\lambda$  is determined based on the importance of one wavelength relative to the others in visual color perception. More precisely, if a part of the spectrum is not very important to the visual system, its influence can be reduced by decreasing the corresponding weight.

The idea of using wPCA on spectral color data was first recommended by Maloney<sup>3</sup> who employed the weighted least squares technique to fit a linear model to 462 surface reflectances. He used  $v(\lambda)$ , a measure of the sensitivity of the photopic visual system at each wavelength, as the weighting function. Yu et al.<sup>2</sup> proposed multispectral image compression via a weighted Karhunen Loeve Transform (wKLT) for spectral accuracy preservation and resulting high fidelity color reproduction. Later, in 2008, Lammanen et al.<sup>1</sup> presented a PCA-based method for the weighted compression of spectral color data and examined the feasibility of two different weight functions.

In this paper, the idea of weighted spectral compression is examined by making a comparison between the two mentioned weighting strategies, i.e., Weighted Least Squares (wLS) versus Weighted Principal Component Analysis (wPCA). The common goal of both methods is to minimize the weighted errors: in the former a series of weighted linear equations is solved to calculate the principal component coefficients; in the later this goal is achieved by applying an appropriate weighting function on spectral data before forming the correlation matrix and extracting the principal eigenvectors. In the other words, in the case of wLS, the spectral sub-space is formed by *unweighted* ordinary eigenvectors calculated from the *unweighted* spectral dataset and weighting is performed in the process of fitting a linear model to the spectral data.

## 2. EXPERIMENTAL PROCEDURE

A dataset including the reflectance data of 1269 Munsell matte<sup>4</sup> and 1950 NCS standard color samples<sup>5</sup> was used to evaluate the performance of wPCA versus wLS. The wavelength sampling of the spectrum for all samples was in 5nm steps from 380-780nm.

In order to assess the influence of different weighting functions on the spectral and colorimetric errors of the reconstructed spectra, we examined 7 different weighting functions.

WF1 and WF2 – Weighting functions proposed by Laamanen et al.<sup>1</sup>,

WF3, WF4, and WF5 – Weighting functions based on the principal diagonal of Cohen's  $\mathbf{R}$  matrix<sup>6</sup>,

WF6 – Weighting function based on the Gaussian function,

WF7- Weighting function based on  $v(\lambda)$  as recommended by Maloney.<sup>3</sup>

Two weighting functions WF1 and WF2 are illustrated in Fig (1-a). WF1 is a linear combination of the CIE 1931 color-matching functions and defined as:

$$\mathbf{WF1}(\lambda) = \left( \frac{\bar{x} + \bar{y} + \bar{z} + 1}{3.1607} \right) \quad (3)$$

Scaling by the constant 3.1607 is done to make the maximum of the function equal to 1. Due to small values of the CMFs at some wavelengths, a constant is added to the functions to avoid computational instability. For simplicity, we drop  $\lambda$  and call this function WF1 throughout the remainder of the paper.

WF2 is another weighting function proposed by Laamanen et al.<sup>1</sup> To calculate this function, two identical constant spectral curves were considered and then the value of one wavelength was raised by adding a parameter  $w$  until the color difference between two spectra overcame a predefined threshold. This procedure was repeated for all wavelengths and the resulting  $w$  at each wavelength formed the corresponding element in the weighting vector WF2.

Fig (1-b) shows three weighting functions WF3, WF4, and WF5 based on the square root of the principal diagonal of matrix  $\mathbf{R}$  for three different illuminants. Matrix  $\mathbf{R}$  is an orthogonal projector developed by Cohen and Kappauf for decomposing the color stimulus into its fundamental and metamer black.<sup>6</sup> This matrix is calculated from matrix  $\mathbf{A}$  defined as the product of a set of color matching functions (in this study, CIE 2°-1931) and a given illuminant:

$$\mathbf{R} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \quad (4)$$

where prime indicates matrix transpose.

If matrix  $\mathbf{A}$  is defined such that its columns are formed using just three color matching functions, then the diagonal of  $\mathbf{R}$  will represent the squared magnitude of spectral lights in color space.<sup>6,7</sup> Worthey indicates that the square root of the diagonal shows the comparative strength of the colors of the spectrum for the CIE 2° (1931) observer.<sup>7</sup> Imai et al. also used the diagonal of matrix  $\mathbf{R}$  in order to calculate weighted mean square error of spectral matches.<sup>8</sup>

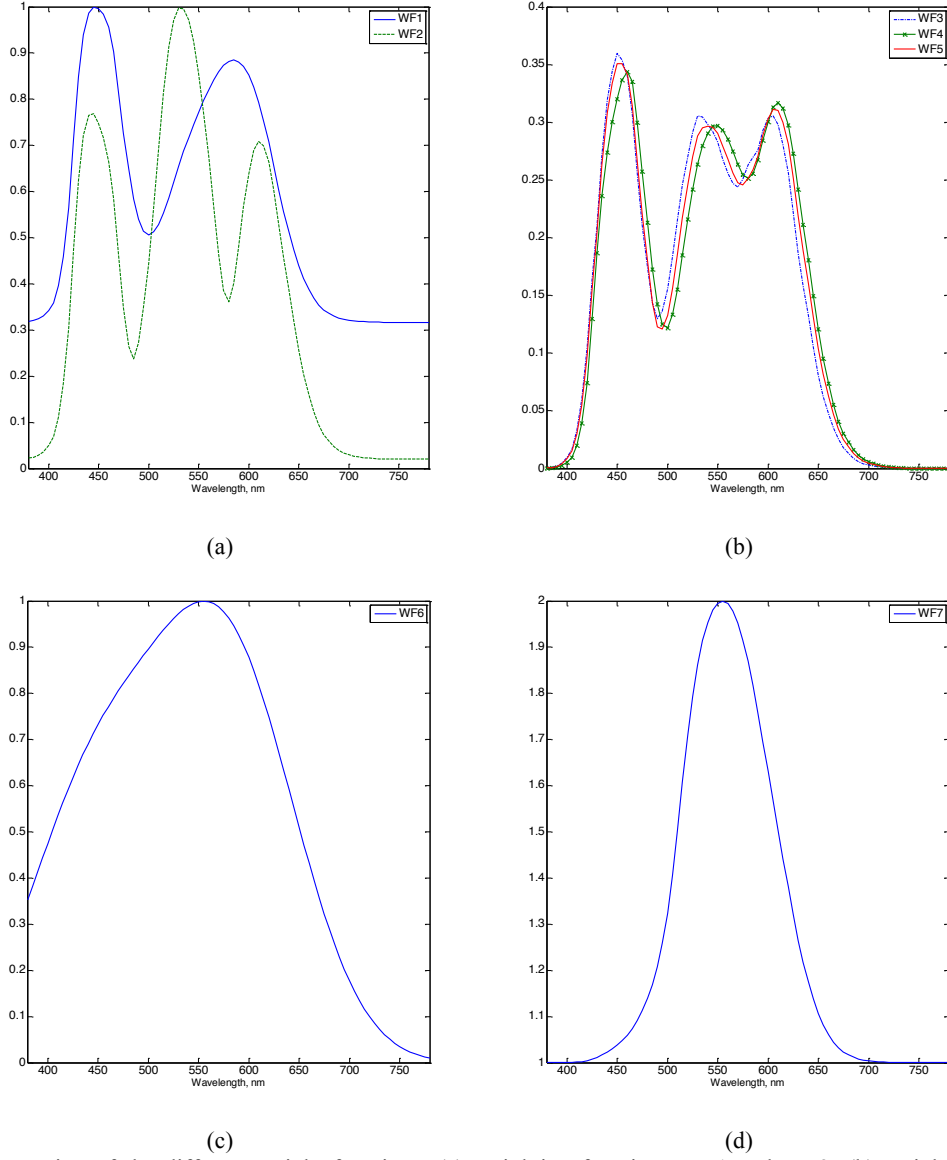


Figure 1. Illustration of the different weight functions: (a) weighting functions WF1 and WF2; (b) weighting functions calculated from matrix  $\mathbf{R}$  for the 2 degree observer and three different illuminants (see text); (c) WF6 calculated from three Gaussian functions; and (d) weighting function based on  $v(\lambda)$ .

As matrix  $\mathbf{R}$  is a function of the illuminant, we can make different matrices and subsequently find different sets of weighting functions for each combination of illuminant and observer. In this study, we use two different weighting functions WF3 and WF4 corresponding to illuminants D65 and A, respectively. WF5 is a synthetic illuminant defined as

the arithmetic mean of D65 and A. In this way, the properties of both illuminants are incorporated in a single synthetic one.<sup>9</sup>

The weighting function WF6 calculated from three Gaussian functions is shown in Fig (1-c). The full width at half height of the Gaussian functions is 125nm and they are centered at 440, 540 and 603 nm. WF6 is the sum of these 3 Gaussians normalized so the maximum is 1.

The last weighting function examined in this study is based on the luminosity function  $v(\lambda)$ . This function shown in Fig (1-d) was also used by Maloney in a weighted least squares manner to fit a linear model to a set of spectral reflectances.

### 3. RESULTS AND DISCUSSION

The error of the spectral reconstruction of the reflectances of the 3219 colored samples (including Munsell and NCS data) compressed using wPCA and wLS based on the seven different weighting functions discussed above are given in Table 1. For comparison, the non-weighted (NW) PCA and LS are included in the first row. The color differences are evaluated for the CIE 1931 standard observer and the CIE standard illuminants D65 and A. The error in the reconstruction is reported in Table 1 in terms of the mean of the Root Mean Square (RMS). The first row of the table represents the results of the non-weighted ordinary PCA/LS compression method. To simplify the comparison, the results of recovery using just *three* basis vectors are shown in this table.

To examine further how different weighting functions influence the performance of wPCA and wLS, the compression is compared for 3 to 10 basis vectors. In addition, the colorimetric errors for the reconstructed spectra versus the input spectra are calculated for two illuminants, A and D65. For each method, Fig. 2 shows for each illuminant the percentage of the reconstructed spectra having  $\Delta E < 0.5$  as a function of the number of basis vectors used. The comparison is limited to the three weighting functions WF2, WF3 and WF5 since they produced the best results according to Table 1.

#### 3.1 Non-weighted versus Weighted Compression – Colorimetric Accuracy

As Table 1 shows, the weighted methods outperform the non-weighted method in terms of colorimetric error under both D65 and A. The amount of improvement depends on the choice of weighting function. In comparison, the spectral dissimilarity (calculated in terms of RMS) between the input and reconstructed spectra is smallest for the non-weighted method. This is to be expected since applying the weights means that the true RMS distance between the two spectra is no longer being minimized. The weightings help reduce colorimetric error at the expense of increasing the spectral error.

Fig. 2 shows that when using a small number of basis vectors non-weighted compression has the lowest percentage of reconstructed spectra having  $\Delta E < 0.5$ . However, as the number of basis vectors increases the percentage of spectra with  $\Delta E < 0.5$  also increases sharply until at 10 basis vectors the non-weighted method works as well as the weighted methods.

#### 3.2 wLS versus wPCA

A comparison between weighted compression via wPCA versus wLS using *three* basis vectors indicates that in all cases wPCA leads to smaller colorimetric errors under D65. However, for illuminant A reconstructing the spectra using wLS with weightings WF2, WF3, WF4, and WF5 leads to a smaller color difference than using wPCA. It appears that wLS more successfully recovers the long wavelength portion of the reflectance spectrum, which is more important under illuminant A. In the other words, applying the weights on the equations corresponding to long wavelengths (wLS) is likely more effective than weighting the data beforehand (wPCA).

Fig. 2 reveals that weighted compression via wLS with more than three basis vectors leads to higher colorimetric accuracy than wPCA under both A and D65. Regardless of the weighting function, the color difference for 100% of the reconstructed spectra under A and D65 is less than 0.5 when the compression is carried out using 10 basis vectors.

Table 1. Spectral and colorimetric accuracy of the spectral reconstruction of 3219 surface reflectances including Munsell and NCS samples using three basis vectors. The reconstructions were performed using the wPCA and wLS techniques. Color differences are based on the color difference formula CIEDE2000 and CIE 2-1931. NW refers to the non-weighted method or, equivalently, weights of unity.

		$\Delta E$ under $D_{65}$		$\Delta E$ under A		RMS
		Mean	Max	Mean	Max	
NW						
	PCA					
	LS	3.45	22.27	2.85	23.73	0.0255
WF1						
	wPCA	2.33	17.37	1.53	18.65	0.0281
	wLS	3.03	20.62	2.31	21.92	0.0262
WF2						
	wPCA	0.7	4.1	0.87	6.40	0.0338
	wLS	1.13	7.72	0.66	7.33	0.0333
WF3						
	wPCA	0.68	5.40	0.66	6.17	0.0350
	wLS	1.00	8.35	0.32	2.84	0.0361
WF4						
	wPCA	1.29	8.19	0.37	5.14	0.0337
	wLS	1.42	10.30	0.38	4.14	0.0346
WF5						
	wPCA	0.94	6.06	0.45	5.73	0.0342
	wLS	1.20	9.11	0.30	3.15	0.0352
WF6						
	wPCA	1.35	11.14	0.82	6.12	0.0325
	wLS	1.93	12.73	1.21	13.30	0.0302
WF7						
	wPCA	3.32	19.73	2.43	20.93	0.0264
	wLS	3.36	20.67	2.60	22.50	0.0257

### 3.3 Comparison between Different Weighting Functions

According to the results summarized in Table 1, of the seven candidate weighting functions, generally WF2, WF3, WF4, and WF5 lead to the smallest colorimetric errors. For the weighting functions based on the principal diagonal of matrix  $\mathbf{R}$  (WF3, WF4 and WF5) the color difference reduction under illuminant A is more significant. This reduction will become even greater when we use this illuminant in making the matrix  $\mathbf{R}$  (WF4 and WF5) and therefore increase its influence on the corresponding weighting function. From the standpoint of colorimetric accuracy under D65, WF3 leads to the least mean color difference, whereas, WF2 leads to the least maximum color difference.

As mentioned earlier, in order to produce more balanced color differences under multiple lighting conditions, both illuminants A and D65 contributed to weighting function WF5. The effect of this contribution is evident in both colorimetric as well as spectral errors. As the results show, WF5 leads to errors between those of WF3 and WF4. This method can be extended to more than two lights. In addition, in some circumstances where different lights have different importance, the contribution of various illuminants can be simply controlled by replacing the ordinary mean by a weighted one.<sup>9</sup>

### 3.4 Effect of the Number of Basis Vectors

#### *wPCA Compression*

A comparison between three different weighted methods WF2, WF3 and WF5 which produced the best results indicates that WF2 and WF3 have the same performance under D65 when only three basis vectors are used. Nonetheless, by increasing the number of components the colorimetric errors drop more quickly when data is weighted by WF3 (2-a). For example in the case of using 5 components the color difference for 83.97% of samples reconstructed by WF2 is less than 0.5 while this percentage will be 96.36 if the recovery is done by WF3. As the graph shows, this trend continues up

to 5 components and then levels off. An almost similar trend is observed under illuminant A, with the exception that WF3 performs better than WF2 even when three components are used, and WF2 works better when the number of components reaches 10.

The graphs reveal that the performance of spectral compression (in terms of  $\Delta E_{D65}$ ) using WF5 is lower than two other methods when only three components are used. However, this method offers much better results when the number of basis components continues to be increased. Figure 2(b) also shows that the colorimetric errors of reconstructed spectra weighted by WF5 are much less than the other methods when the recovery process is done using a small number of basis vectors (up to 7). This is due to the contribution of illuminant A in making matrix  $\mathbf{R}$  upon which WF5 is based. As the graphs show, by increasing the number of components all methods eventually perform almost the same.

#### *wLS Compression*

As can be observed in Fig. 2(c), reconstruction via wLS using WF2 or WF5 shows the same amount of colorimetric error under D65 illuminant for the case of 3 or 4 basis vectors. However, increasing the number of basis vectors improves the performance of WF5 more significantly. Regardless of the number of basis vectors, weighting function WF3 outperformed the others. In Fig. (2d), the superiority of the two weighting functions WF3 and WF5 over WF2 is evident.

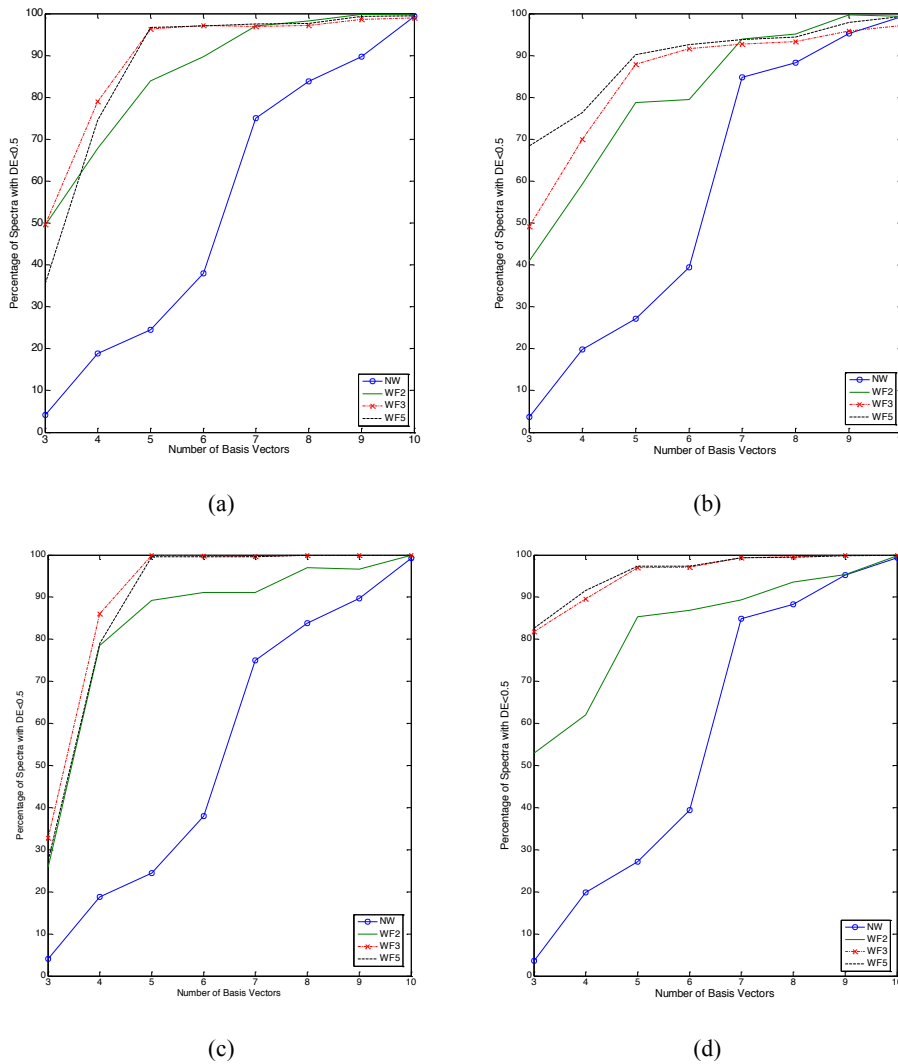


Figure 2. Percentage of spectra reconstructed with  $\Delta E < 0.5$  of the input spectra when observed under D65 (a,c) and A (b,d) versus the number of basis vectors. Reconstruction was performed using wPCA (a,b) and wLS (c,d) with three different weightings (WF2, WF3 and WF5).

### 3.5 Non-weighted versus Weighted Compression – Spectral Accuracy

To evaluate the spectral accuracy of the various recovery methods and illustrate the influence of the weighting factors on spectral errors, we compare the residue of differences between the actual and reconstructed spectra estimated by wPCA and wLS. The weighting function is WF5. The spectral difference as a function of wavelength is shown in Fig. 3.

As can be seen from Fig. 3, the difference between the actual and reconstructed reflectance spectra at the middle of the spectrum—particularly at those wavelengths corresponding to the weighting function’s peak wavelengths—are significantly less than the corresponding errors in the non-weighted reconstruction. However, this accuracy is at the expense of increased error at the end of the spectrum where the human visual system is less sensitive. Therefore, incorporating the weighting function into the compression method selectively alters the relative importance of different wavelengths leading to poorer overall spectral reconstruction, while yielding lower colorimetric error.

## 4. CONCLUSION

Weighted spectral compression is compared for wPCA and wLS with seven different weighting functions. The results show that except when the compression involves a large number ( $>10$ ) of basis vectors, non-weighted (NW) compression always leads to higher colorimetric errors, although its spectral accuracy is better. For spectral compression involving more than three basis vectors, wLS is more successful than wPCA in reproducing the color information under both A and D65. A comparison of the different weighting functions indicates that, although all weighting functions perform similarly when the number of basis vectors goes beyond 8, incorporating weights based on the diagonal of matrix  $\mathbf{R}$  reduces the colorimetric errors more than the other weighting functions whenever 5 or fewer basis vectors are used.

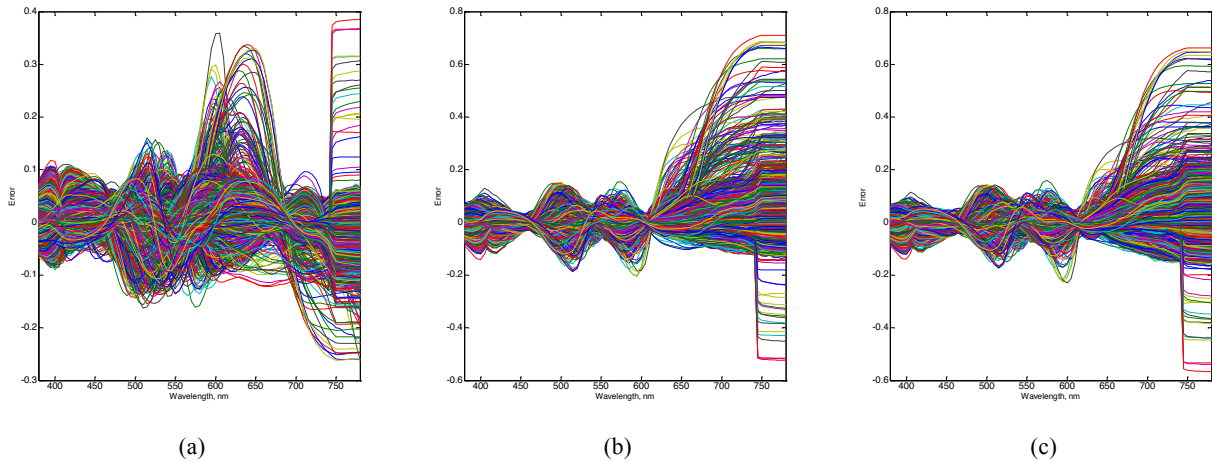


Figure 3. Recovery errors ( $\mathbf{R}_\lambda - \hat{\mathbf{R}}_\lambda$ ) versus wavelength for: (a) NW method; (b) wPCA with WF5; and (c) wLS with WF5.

## 5. REFERENCES

- [1] Laamanen, H., Jetsu, T., Jaaskelainen, T., and Parkkinen, J., “Weighted compression of spectral color information,” *J. Opt. Soc. Am. A* 25 (6), 1383-88 (2008).
- [2] Yu, S., Murakami, Y., Obi, T., Yamaguchi, M., Ohya, N., “Multispectral image compression for high fidelity colorimetric and spectral reproduction,” *JIST* 50 (1), 64-72 (2006).

- [3] Maloney, L. T., "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A* 3(10), 1673-1683 (1986).
- [4] Eastern Finland Spectral Image Database, University of Eastern Finland, Spectral Color Research Group, <https://www.uef.fi/spectral/spectral-database>.
- [5] Hard, A. and Sivic, L., "NCS—Natural color system: A Swedish standard for color notation," *Col. Res. & Appl. J.* 6 (3), 129-138 (1981).
- [6] Cohen, J. B. and Kappauf, W. E., "Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *Am J Psych.* 95 (4), 537-564 (1982).
- [7] Worthey, J. A., "Color rendering: asking the question," *Col. Res. & Appl. J.* 28(6), 403-412 (2003).
- [8] Imai, F. H., Rosen, M. R., Berns, R. S., "Comparative study of metrics for spectral match quality," *Proc. CGIV* 492-496 (2002).
- [9] Agahian, F. and Amirshahi, S. H., "Design of virtual illuminants to control the colors under multiple illuminants," *Col. Res. & Appl. J.* 34(3), 205-209 (2009).