Reducing the Dimensionality of Plant Spectral Databases

Ian E. Bell and Gladimir V. G. Baranoski

***Abstract*—Ground-based measurements of plant reflectance and transmittance are essential for remote sensing projects oriented toward agriculture, forestry, and ecology. This paper examines the application of principal components analysis (PCA) in the storage and reconstruction of such plant spectral data. A novel piecewise PCA approach (PPCA), which takes into account the biological factors that affect the interaction of solar radiation with plants, is also proposed. These techniques are compared through experiments involving the reconstruction of reflectance and transmittance curves for herbaceous and woody specimens. The spectral data used in these experiments were obtained from the Leaf Optical Properties Experiment (LOPEX) database. The reconstructions were performed aiming at a root-mean-square error lower than 1%. The results of these experiments indicate that PCA can effectively reduce the dimensionality of plant spectral databases from the visible to the infrared regions of the light spectrum, and that the PPCA approach can further maximize the accuracy/cost ratio of the storage and reconstruction of plant spectral reflectance and transmittance data.**

***Index Terms*—Leaf, measurements, plant, principal component analysis (PCA), reflectance, spectral databases, transmittance.**

I. Introduction

VEGETATION is arguably the most important remote sensing target, since it is a natural resource on which all human and animal life depends [1]. The applications of plant remote sensing include projects oriented toward agriculture, forestry, and ecology. Viewed in this context, ground-based measurements of plant reflectance and transmittance are essential, since these spectral quantities are directly related to leaf biochemistry and water status. In comprehensive experiments, these measurements are made in conjunction with measurements of leaf biochemical constituents, such as lignin, proteins, and cellulose, as well as pigment and water content. These data allow the application of inversion procedures [2] on spectral data obtained from spaceborne platforms [3]–[5] in order to remotely estimate leaf biochemical content. The major processes involved in the terrestrial ecosystems, such as photosynthesis and foliar decomposition, can be related to these constituents [6].

These measurements, as outlined above, may involve several aspects such as the wavelength of the incident light, the illuminating angle, and the viewing geometry [7]. These measurements can be stored in a spectral database, with multiple entries for each specimen representing each individual species. Sometimes, it is also desirable to store data for a number of specimens of each species in the database, to allow a quantitative spectral analysis of species variability [8]. Furthermore, the number of entries in this database grows as new measurements for different species become available. Eventually, however, its size might become impractical for many applications due to the large number of species found in nature.

Plant spectral data can also be extracted from computer models of plant reflectance and transmittance [9]. In this case, modeling spectral curves for many combination of parameters, such as chlorophyll content, water content, thickness, etc., will consume large amounts of storage space. In theory, one could save storage space by storing these parameters instead of the curves obtained using a computer model. However, the computation of such curves is highly time consuming [7]. For applications that demand high interactivity rates, such data usually need to be precomputed, stored offline, and quickly accessed online.

Clearly, the key aspects in the design of plant spectral databases are compactness and low reconstruction error. Principal components analysis (PCA) [10] techniques can be used to achieve these goals. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a number of uncorrelated variables called principal components [11]. In some application areas, it is also called the (discrete) Karhunen–Loève transform [12].

PCA is widely used in many areas, e.g., signal processing, computer graphics, statistics, colorimetry, and neural computing. In remote sensing, it is mainly applied to the classification and interpretation of satellite images by using spectral data associated with different targets [13]–[18], including regions of vegetation [19], [20]. Furthermore, Price [17], [21]–[23] has proposed a procedure for identifying the independent spectral variability of reflectance spectra of crops and soils, which applies PCA as an intermediate refinement step. This procedure was applied to data obtained from satellite, aircraft, and ground measurements. The crop species used in the ground-based experiments were corn, soy, winter wheat, sunflower, and alfalfa.

Another application of PCA in remote sensing involves spectral band selection for multispectral remote sensor systems. Wiersma and Landgrebe [12] have described an approach that depends on a principal component analysis of a collection of spectra, combined with results from a classification procedure. Recently, Price [24], [25] used his identification procedure to perform spectral band selection for satellite image data.

Manuscript received May 23, 2003; revised October 29, 2003. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant 238337 and in part by the Canada Foundation for Innovation (CFI) under Grant 33418.

The authors are with the School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: gvgbaran@curumin.math.uwaterloo.ca).

Digital Object Identifier 10.1109/TGRS.2003.821697

In this paper, we examine the application of PCA to reduce the dimensionality of plant leaf spectral databases used to store reflectance and transmittance spectra obtained from ground measurements. Vrhel *et al.* [26] analyzed the possibility of representing spectral data of Munsell chips and various natural materials (including plant leaves) with a set of basis functions determined using PCA. Their investigations, however, were aimed at colorimetry applications and were limited to reflectance data in the 390–730-nm range. Grossman *et al.* [27] used PCA in conjunction with stepwise multiple linear regression techniques to examine the statistical relationships among a number of foliar biochemicals. They indicated a reduction of the dimensionality of plant reflectance data in the 800–2498-nm range using PCA.

The investigation presented in this paper extends the observations made by Grossman *et al.* [27], focusing on the compactness and low reconstruction error of plant spectral databases. It covers the visible (400–700 nm), near-infrared (700–1300 nm), and infrared regions (beyond 1300–2500 nm) of the light spectrum, and includes reflectance and transmittance data. Moreover, besides the application of PCA techniques, a novel piecewise PCA approach (PPCA) is proposed. This biologically based algorithm allows us to further maximize the storage savings and to minimize the reconstruction errors associated with plant spectral data. These findings are supported by experiments in which measured and reconstructed spectral curves for different specimens are compared.

II. PCA

The main purpose of PCA is to reduce the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much as possible the variations presented in the dataset [11]. In this section we outline the idea behind this approach, and describe the technique used in our work to implement it. The reader interested in a more detailed description of the PCA approach is referred to various comprehensive texts on this subject [11], [28].

Generally, we have a set of p measurements (e.g., reflectances of p specimens) and n variables (e.g., wavelengths). We arrange the data in a $p \times n$ matrix M, and we wish to represent each set of measurements using $n' \ll n$ components. The PCA identifies n' so-called modes, being defined as the n'-vectors u_j, $j = 1 \ldots n'$ corresponding to the directions in a n-dimensional space where the measurements exhibit the maximum variance. In other words, the first mode u_1 corresponds to the direction of maximum variance, the second mode u_2 corresponds to the direction of maximum variance uncorrelated to u_1, and so on [29].

Singular value decomposition (SVD) is a technique widely used to implement this analysis. Although there are techniques that present a lower time complexity [30], we choose to use SVD in this work because of its numerical stability. Numerous variants of the SVD algorithm exist [10], [11]. We apply here the version proposed by Pratt [31], which is used in the numerical computing system Matlab [32]. This technique consists on forming the singular value decomposition of M

$$M = USV^T \tag{1}$$

where $U[p \times n]$ and $V[n \times n]$ are orthogonal matrices, and $S[p \times n]$ is a real, nonnegative, and diagonal matrix. The columns u_i of $U = [u_1, \ldots, u_n]$ are the left singular vectors, and the columns v_i of $V = [v_1, \ldots, v_n]$ are the right singular vectors. The diagonal of S contains the so-called singular values σ_i, where $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$. When M is symmetric and semidefinite, the singular values σ_i are the eigenvalues of M; otherwise, the singular values are the square roots of the eigenvalues of $M^T M$.

Consider a plant spectral reflectance dataset given by a matrix M_R, where each row contains the reflectance spectrum of a given species, whose values with respect to the sample wavelengths are stored column by column. The goal is to reduce the dimensionality of M_R, i.e., the PCA approach is applied to the entire dataset, rather than to each specimen's spectrum. The computation of the principal components of M_R is given by

$$\begin{aligned} [U_R, S_R, V_R] &= \text{SVD}(M_R) \\ d_R &= \text{diag}(S_R) \\ c_R &= M_R V_R \end{aligned}$$

where d_R represents the singular values, V_R the basis of principal components (as column vectors), and c_R the coordinates relative to the basis (as row vectors). The same procedure can be applied to a plant spectral transmittance dataset represented by a matrix M_T, where the rows represent transmittance spectra for different species, and the columns represent the sample wavelengths.

Once the data matrices have been decomposed, the n_R basis components of V_R can be stored. This basis matrix may be as large as the original dataset M_R. However, since the goal is to reduce the dimension of the data, one usually chooses a reduced number of components, $n'_R \ll n_R$, and stores the smaller basis V'_R (the first n'_R columns of V_R). The selection of n'_R can be made according to the fraction of the singular value sum accounted for by the first n'_R singular values (see Fig. 4). The reconstruction of the spectral data, i.e., the formation of a $p \times n'_R$ matrix $\hat{M}_R$, is given by

$$\hat{M}_R = c'_R * (V'_R)^T .$$

It can be proven that the representation given by PCA is an optimal linear dimension reduction technique in the mean-square sense [11], although of course once the basis is chosen, any other basis spanning the same PCA subspace will have the same data compression or noise reduction capabilities.

III. Piecewise Principal Components Analysis (PPCA)

Experiments [33], [34] show that the reflectance and transmittance of plant leaves are relatively low in the visible (Vis) region of the light spectrum, and that absorptance dominates. Most of the absorption in this region is caused by pigments present in the leaf tissue [2]. As pointed out by Devlin and Baker [35], the chlorophylls are by far the most important and abundant of these pigments. The near-infrared (NIR) region, however, is characterized by a lack of absorption and high values for reflectance and transmittance [33], while in the infrared (IR) region (beyond 1300–2500 nm), the absorption is controlled by the water content [2].

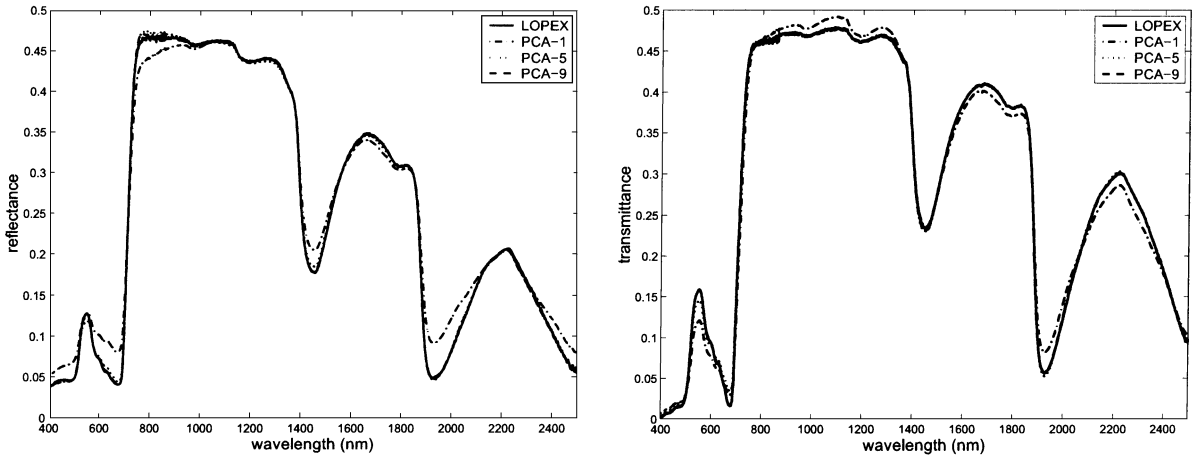

Fig. 1. Measured (LOPEX) and reconstructed (PCA) spectral curves for a fresh soy leaf (*Soja hispida*). Curves were reconstructed using 1, 5, and 9 components.

These well-defined regions suggest that the use of an adaptive PCA approach may result in more compactness and lower reconstruction errors. In other words, instead of an integral application of PCA over the whole light spectrum, we can perform a piecewise application of PCA, in which each piece corresponds to one of the three regions (Vis, NIR, and IR). Since the effectiveness of the PCA method depends on the representativity of the reflectance or transmittance data used in the singular value decomposition, applying PPCA to matrices for the Vis, NIR, and IR regions yields different components than applying PCA to a matrix for all three regions. After eliminating components that contribute little to the overall reflectance or transmittance curves, reconstructing a particular curve with PPCA, therefore, gives pieces for each region that may not join smoothly at region boundaries. Although it would be possible to use a constrained PPCA method to ensure smooth joins, we found the discontinuities to be noticeable only with a few species, and only when using minimal numbers of components.

IV. Data

In our experiments, we considered chlorophyllous leaf specimens whose spectral data are available in the Leaf Optical Properties Experiment (LOPEX) database [6]. This database consists of leaf samples representing woody and herbaceous species that were obtained from trees and crops near the Joint Research Centre in Ispra, Italy. Spectra were originally scanned in 1-nm steps, but the wavelength interval was averaged over 5 nm to reduce noise. The measurements considered fresh and dry individual leaves and optically thick samples (stacked leaves plus needles or powders).

The matrices M_R and M_T described earlier were formed using respectively the reflectance and transmittance spectra of all individual fresh leaves stored in the LOPEX database. For the illustration of the reconstruction experiments, we considered a soy (*Soja hispida*) specimen representing an herbaceous species, and a poplar specimen (Populus canadensis) representing a woody species. In order to increase our scope of observations, we also considered an iris specimen (*Iris germanica L.*) whose reflectance and transmittance curves have magnitudes markedly different from soy and poplar. For completeness, we also included summaries of the results obtained considering spectral data for all species represented in the LOPEX database, and derived from measurements on fresh individual leaves, dry individual leaves, fresh stacked leaves, and dry stacked leaves.

In order to be consistent with the plant remote sensing literature, we use the root-mean-square-error (RMSE) measure in our experiments. As one increases the number of coefficients used in the reconstruction of the spectral curves, the RMSE of these curves decreases. The goal is to obtain the best compromise between the number of coefficients and the RMSE. Therefore, we need to select a threshold RMSE value to be used as reference in the evaluation of our results. Jacquemoud *et al.* [3] state that in terms of plant reflectance and transmittance reconstruction, an RMSE less than or equal to 0.01 indicates a reasonably accurate reconstruction. Hence, we select this value as the upper bound for our RMSE comparisons.

V. Results

Figs. 1–3 illustrate the application of PCA to the reconstruction of spectral reflectance and transmittance curves for the three specimens considered. The original curve, and reconstructions with 1, 5, and 9 principal components are shown (intermediate numbers of components are possible, but omitted here to avoid clutter). A visual inspection of these curves shows that there is good qualitative agreement between the original and the reconstructed curves. Although reconstruction errors vary from one specimen to another, by increasing the number of principal components, one can obtain an asymptotic reduction of the error as illustrated in Fig. 4.

Tables I and II show the number of components required to reconstruct spectral curves within specified RMSE bounds. For PPCA, numbers of components are given for the three characteristic regions (Vis, IR, and NIR). For PCA, numbers of components are given for the whole spectrum of interest, i.e., the union (all) of the these regions. From the figures presented in these tables, it can be clearly observed that the number of principal components varies from one species to another and from one region to another.

The iris, poplar, and soy examples used in this paper were chosen to be representative of our results with the LOPEX database. To demonstrate that the PCA and PPCA methods are vi-

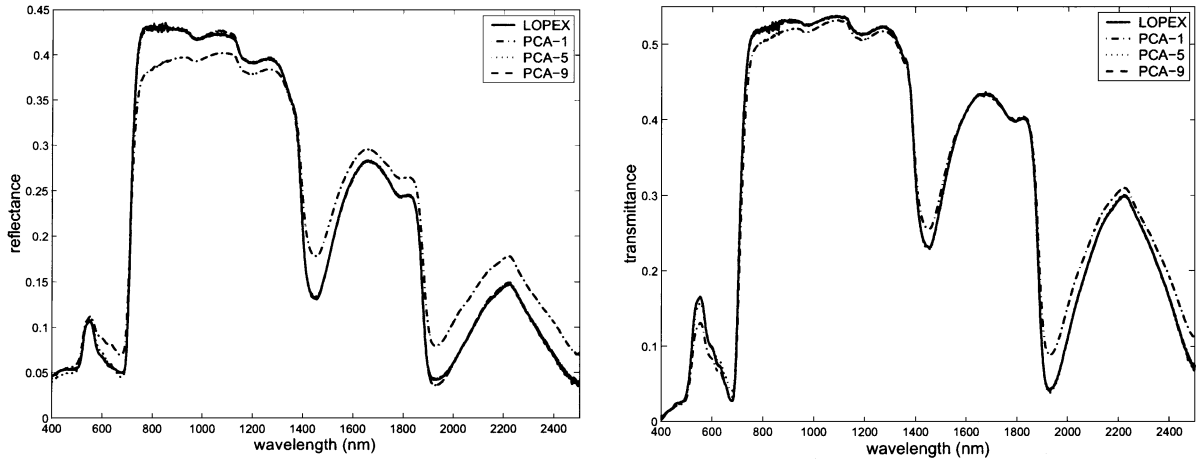

Fig. 2. Measured (LOPEX) and reconstructed (PCA) spectral curves for a fresh poplar leaf (*Populus canadensis*). Curves were reconstructed using 1, 5, and 9 components.

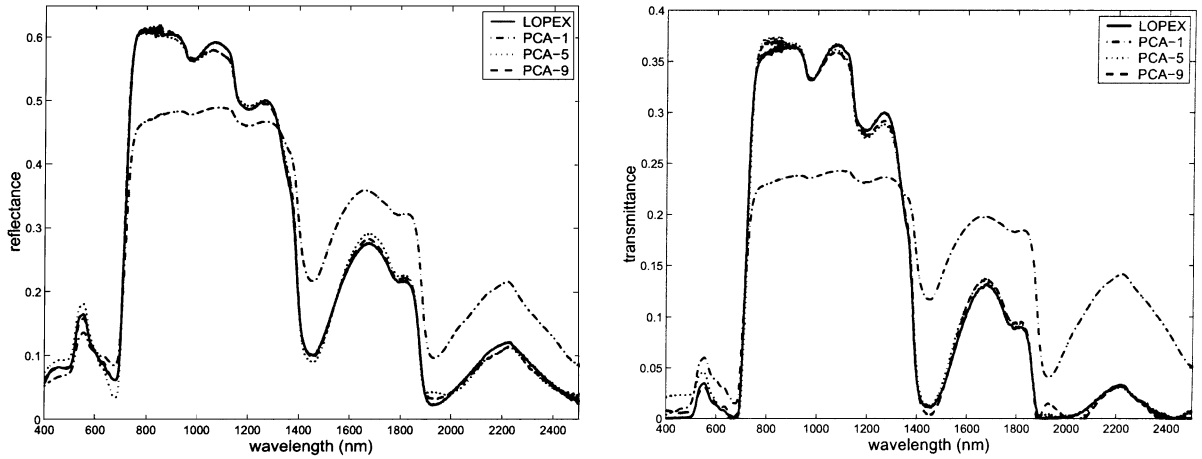

Fig. 3. Measured (LOPEX) and reconstructed (PCA) spectral curves for a fresh iris leaf (*Iris germanica L.*). Curves were reconstructed using 1, 5, and 9 components.

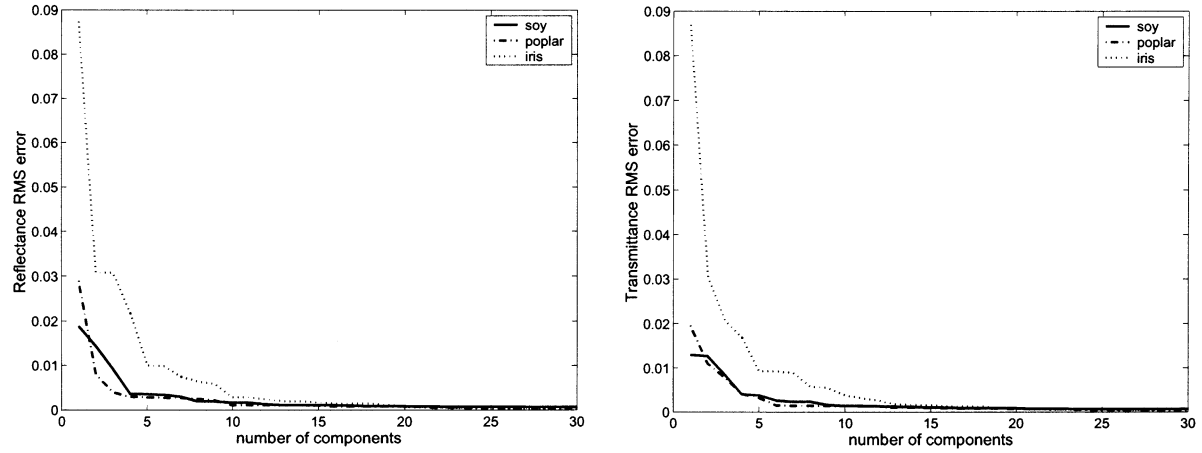

Fig. 4. RMS errors associated with the reconstruction (PCA) of (left) spectral reflectance curves and (right) spectral transmittance curves for soy, poplar, and iris leaves, within the visible to infrared (400–2500 nm) region of the light spectrum, and using different numbers of PCA components.

able over all LOPEX data, we computed RMSE errors for 1-, 5-, 9-, and 20-components PCA and PPCA. Tables III, IV, VII, and VIII show the results for all LOPEX leaf reflectances (fresh, dry, stacked fresh, and stacked dry). Tables V and VI show the results for all LOPEX transmittances (fresh and dry). The PPCA method was applied to the Vis, NIR, and IR spectral regions, while PCA was applied to the union of these regions (all). Larger numbers of components clearly reduce the average and maximum errors in all cases. The PPCA method, by targeting shorter wavelength intervals, gives smaller errors than the PCA method, which is over all wavelengths. For example, in Table III, to achieve a maximum RMSE error of 0.0038 we must use nine components with PCA, but fewer than five components would suffice for PPCA on the Vis or IR regions.

TABLE I
NUMBER OF PRINCIPAL COMPONENTS REQUIRED TO RECONSTRUCT SPECTRAL REFLECTANCE CURVES FOR SOY (S), POPLAR (P), AND IRIS (I) LEAVES CONSIDERING DIFFERENT RMSE BOUNDS. THE PCA METHOD APPLIES TO THE VISIBLE-TO-INFRARED DATA (ALL). THE PPCA METHOD APPLIES TO THE THREE REMAINING CATEGORIES (VIS, NIR, AND IR)

RMSE	All			Vis			NIR			IR		
	s	p	i	s	p	i	s	p	i	s	p	i
0.010	3	2	5	2	2	2	1	2	4	2	2	4
0.009	4	2	7	2	2	2	2	2	4	2	2	4
0.008	4	2	7	2	2	2	2	2	4	3	2	4
0.007	4	3	8	2	2	2	2	2	4	3	2	5
0.006	4	3	9	2	2	2	2	2	5	3	2	5
0.005	4	3	10	2	2	2	2	2	5	3	2	5
0.004	4	4	10	3	2	3	4	2	5	3	3	5
0.003	8	5	10	4	3	3	4	3	5	3	3	5
0.002	8	12	13	4	4	5	4	4	5	4	3	7
0.001	18	18	20	7	4	10	12	6	12	8	10	10

TABLE II
NUMBER OF PRINCIPAL COMPONENTS REQUIRED TO RECONSTRUCT SPECTRAL TRANSMITTANCE CURVES FOR SOY (S), POPLAR (P), AND IRIS (I) LEAVES CONSIDERING DIFFERENT RMSE BOUNDS. THE PCA METHOD APPLIES TO THE VISIBLE-TO-INFRARED DATA (ALL). THE PPCA METHOD APPLIES TO THE THREE REMAINING CATEGORIES (VIS, NIR, AND IR)

RMSE	All			Vis			NIR			IR		
	s	p	i	s	p	i	s	p	i	s	p	i
0.010	3	3	5	2	2	1	3	2	3	2	2	4
0.009	3	3	7	2	2	1	3	3	3	3	2	4
0.008	4	3	8	2	2	1	3	3	3	3	2	4
0.007	4	4	8	2	2	1	3	3	4	3	2	4
0.006	4	4	8	2	2	1	3	3	4	3	2	4
0.005	4	4	10	2	2	2	3	3	5	3	2	5
0.004	4	5	10	3	2	2	3	3	5	3	3	5
0.003	6	6	11	3	2	2	3	3	5	3	4	7
0.002	9	6	13	3	2	3	3	3	6	4	4	7
0.001	16	13	19	6	5	4	9	9	7	7	7	8

TABLE III
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 83 LOPEX FRESH LEAF REFLECTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.02490	.01152	.01043	.01792	.19478	.09115	.04857	.17898
5	.00453	.00099	.00137	.00168	.01260	.00305	.00468	.00371
9	.00189	.00029	.00094	.00089	.00380	.00079	.00122	.00139
20	.00073	.00011	.00065	.00056	.00108	.00017	.00092	.00075

TABLE IV
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 75 LOPEX DRY LEAF REFLECTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.02567	.01062	.02215	.01697	.09808	.05439	.07576	.06741
5	.00642	.00258	.00184	.00230	.01850	.00575	.00397	.00479
9	.00303	.00087	.00104	.00113	.00526	.00181	.00149	.00202
20	.00097	.00017	.00065	.00058	.00155	.00025	.00088	.00079

TABLE V
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 83 LOPEX FRESH LEAF TRANSMITTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.02862	.00919	.01017	.02251	.08791	.09683	.03415	.06917
5	.00402	.00098	.00116	.00154	.01497	.00350	.00238	.00305
9	.00161	.00030	.00083	.00083	.00267	.00070	.00112	.00123
20	.00067	.00008	.00057	.00054	.00087	.00013	.00079	.00072

This aspect is also illustrated in Figs. 5 and 6, which show comparisons involving the reconstruction of spectral reflectance and transmittance curves for the iris specimen applying PCA and PPCA, with the former using five components for the visible-to-infrared spectrum and the latter using five components for each region (Vis, NIR, and IR). The error bars used in these graphs represent the quantitative differences (magnified by a factor of ten) between the original and reconstructed curves at various wavelengths. We used the iris specimen because, from the species considered in our experiments, it presented the lower sconvergence rate (Fig. 4). A visual inspection of these graphs shows that considerably lower reconstruction errors can be obtained using the proposed PPCA approach.

TABLE VI
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 75 LOPEX DRY LEAF TRANSMITTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.02291	.00622	.01892	.01885	.06484	.05052	.07139	.04814
5	.00464	.00116	.00126	.00194	.01217	.00565	.00241	.00383
9	.00220	.00044	.00077	.00103	.00363	.00117	.00106	.00177
20	.00075	.00007	.00051	.00055	.00116	.00012	.00072	.00076

TABLE VII
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 83 LOPEX STACKED FRESH LEAVES REFLECTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.03822	.01208	.02363	.02390	.29927	.08914	.09402	.25098
5	.00586	.00097	.00221	.00184	.01528	.00328	.00612	.00451
9	.00245	.00031	.00129	.00103	.00549	.00096	.00192	.00150
20	.00085	.00011	.00081	.00058	.00132	.00016	.00116	.00076

TABLE VIII
AVERAGE AND MAXIMUM PCA (ALL) AND PPCA (VIS, NIR, AND IR) RMSE FOR 76 LOPEX STACKED DRY LEAVES REFLECTANCE SPECTRA

Components	Average RMSE				Maximum RMSE			
	All	Vis	NIR	IR	All	Vis	NIR	IR
1	.03733	.01136	.03186	.02250	.13319	.05572	.09907	.10001
5	.00930	.00260	.00239	.00365	.02051	.00610	.00630	.00736
9	.00392	.00088	.00123	.00168	.00777	.00222	.00214	.00293
20	.00123	.00018	.00075	.00067	.00200	.00028	.00112	.00092

VI. DISCUSSION

In this paper, we examined the application of techniques, PCA and PPCA, to reduce the dimensionality of plant spectral databases. Instead of storing the full spectra of plant leaves, one needs to store only a certain number of components and a basis, which requires only a fraction of the memory space. For example, each raw reflectance in the LOPEX database requires 2101 floating-point values; however, Table I shows that for soy, three-component PCA suffices for a RMSE of 0.01, thus three floating-point values (once the SVD basis has been stored). This is a dramatic reduction of data, especially for databases of many reflectances. Using the PPCA method, soy would require only two, one, and two components for each region, respectively, to achieve the same RMSE of 0.01. If all components are stored, this gives five floating-point values, which is a substantial reduction from the original data, but not as small as the three floating-point values for PCA. There are, however, alternative interpretations of this table that show PPCA can be the preferred method.

In applications where very low RMSE tolerances must be met, PPCA can compete with PCA directly in terms of total number of components. Table II shows this trend with iris transmittance for a RMSE of 0.001, where 19 components are required for PCA, and an equal $4 + 7 + 8 = 19$ for PPCA. As another example in Table I, poplar reflectance with an RMSE tolerance of 0.002 requires 12 components for PCA, but only $4+4+3 = 11$ components for PPCA. Further investigation of quantization error at the bit level may reveal other advantages to

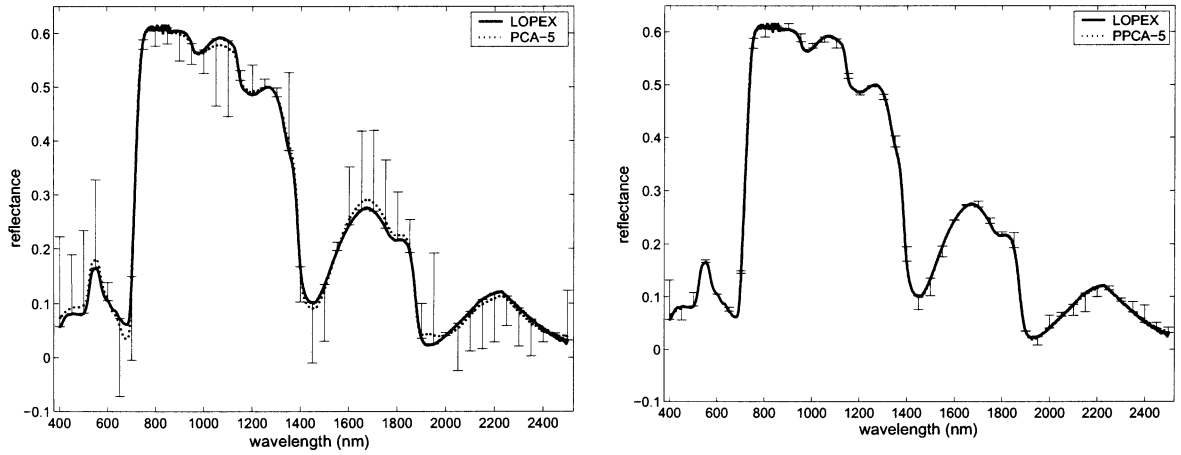

Fig. 5. Measured (LOPEX) and reconstructed [(left) PCA and (right) PPCA] reflectance curves for a fresh iris leaf (*Iris germanica L.*). PCA curves were reconstructed using five components, and PPCA curves using five components for each region (Vis, NIR, and IR). Error bars represent the differences between the curves, magnified by a factor of ten.

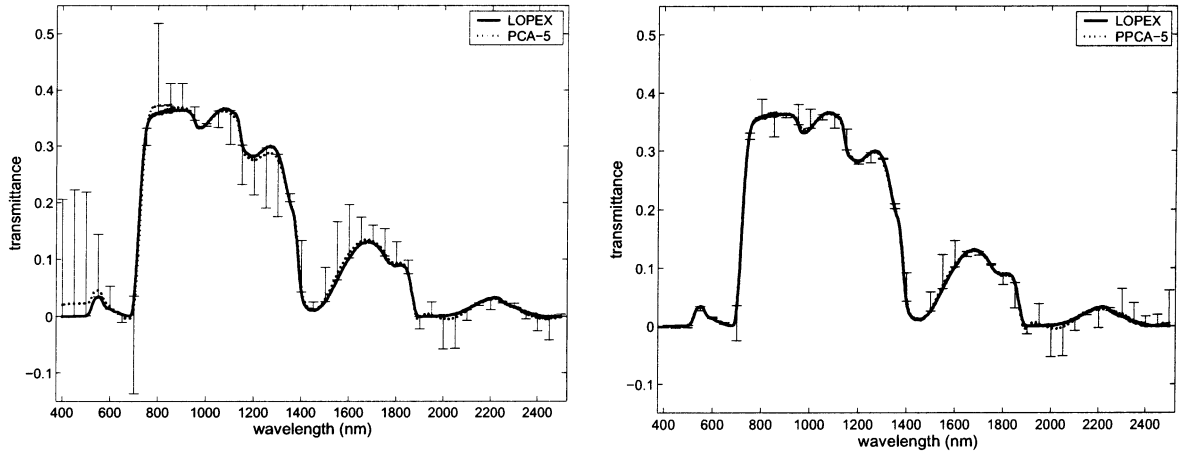

Fig. 6. Measured (LOPEX) and reconstructed [(left) PCA and (right) PPCA] transmittance curves for a fresh iris leaf (*Iris germanica L.*). PCA curves were reconstructed using five components, and PPCA curves using five components for each region (Vis, NIR, and IR). Error bars represent the differences between the curves, magnified by a factor of ten.

the PPCA approach. Even with floating-point values, however, the flexibility of PPCA is an asset.

For some applications, e.g., infrared remote sensing [36], there may be different RMSE requirements for each spectral region. For example, if we consider soy reflectance (Table I), with an RMSE of 0.001 in the infrared, but only 0.01 in the other regions, the PCA method needs 18 components. The PPCA method, however, can meet the infrared RMSE tolerance with eight components, and require only two and one components, respectively, for the visible and near-infrared regions, a total of 11 components. PPCA is, therefore, more adaptable when the spectral regions differ in significance.

These storage savings depend on the number of principal components considered, which in turn depends on the level of accuracy expected for the reconstruction of the original curves. Both techniques examined in this paper allow a significant reduction of the dimensionality of plant spectral databases, while preserving the information represented by the reflectance and transmittance data. The proposed PPCA technique, however, presents a higher accuracy/cost ratio.

Besides contributing to a reduction in the computational overhead associated with the manipulation of such datasets, these techniques may improve the quality of the spectral information stored, since the data not contained in the n selected principal components may be mostly due to noise in the measurements. Another benefit of these methods is that new reflectance or transmittance data may be incorporated into the database without repeating the SVD calculation: it is only necessary to project the new data onto the principal component basis to determine its coordinates, a simple matrix multiplication.

In closing, with increasing amounts of plant spectral data readily available and distributed through the Internet, it is necessary to find efficient storage methods like PCA and PPCA. The PCA method, although not new, may be considered as an approach to compactly storing large databases of plant spectral data, and the novel PPCA method provides more efficient compression, especially when error tolerances vary for different spectral regions. The field researcher or forensic scientist with a hand-held computing device of limited storage would benefit greatly from such compact databases and efficient algorithms. Furthermore, the PCA and PPCA methods support fast addition of new spectra to the database and can be adapted to other remote sensing targets such as soils and natural waters.

ACKNOWLEDGMENT

The authors would like to thank S. Jacquemoud for his helpful suggestions and the anonymous reviewers for their valuable comments. The authors are also grateful to the Joint Research Centre of the European Comission for granting them access to the LOPEX database, which was established during an experiment conducted by the Advanced Techniques Unit of the Space Applications Institute (Italy). Funding is gratefully acknowledged from both the Natural Sciences and Engineering Council of Canada (NSERC Grant 213281) and the Canada Foundation for Innivation (CFI Project 6218)

REFERENCES

[1] H. N. Horler and J. Barber, "Principles of remote sensing of plants," in *Plants and the Daylight Spectrum*, H. Smith, Ed. London, U.K.: Academic, 1981, pp. 43–64.
[2] S. Jacquemoud and F. Baret, "PROSPECT: A model of leaf optical properties spectra," *Remote Sens. Environ.*, vol. 34, no. 2, pp. 75–92, 1990.
[3] S. Jacquemoud, S. L. Ustin, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood, "Estimating leaf biochemistry using prospect leaf optical properties model," *Remote Sens. Environ.*, vol. 56, pp. 194–202, 1996.
[4] R. B. Myneni and J. Ross, *Photon-Vegetation Interactions: Applications in Optical Remote Sensing and Plant Ecology*. Berlin, Germany: Springer-Verlag, 1991.
[5] N. S. Goel, "Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data," *Remote Sens. Rev.*, vol. 4, pp. 1–212, 1988.
[6] B. Hosgood, S. Jacquemoud, G. Andreoli, J. Verdebout, G. Pedrini, and G. Schmuck, "Leaf Optical Properties Experiment 93," Joint Res. Center, Eur. Comm., Inst. Remote Sensing Applications, Ispra, Italy, Tech. Rep. EUR 16 095 EN, 1995.
[7] G. V. G. Baranoski and J. G. Rokne, "Efficiently simulating scattering of light by leaves," *Vis. Comput.*, vol. 17, no. 8, pp. 491–505, 2001.
[8] J. C. Price, "How unique are spectral signatures," *Remote Sens. Environ.*, vol. 49, pp. 181–186, 1994.
[9] S. Jacquemoud and S. L. Ustin, "Leaf optical properties: A state of the art," in *Proc. 8th Int. Symp. Physical Measurements and Signatures in Remote Sensing*. Aussois, France, 2001, pp. 223–332.
[10] G. Golub and C. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: John Hopkins Univ. Press, 1989.
[11] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.
[12] D. J. Wiersma and D. A. Landgrebe, "Analytical design of multispectral sensors," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-18, pp. 180–189, Mar. 1980.
[13] R. P. Gupta, *Remote Sensing Geology*. New York: Springer-Verlag, 1991.
[14] H. Huang and P. Antonelli, "Application of principal component analysis to high-resolution infrared measurement compression and retrieval," *J. Appl. Metereol.*, vol. 40, pp. 365–388, Mar. 2001.
[15] T. M. Lillesand, *Remote Sensing and Image Interpretation*, 3rd ed. New York: Wiley, 1994.
[16] P. M. Mather, *Computer Processing of Remotely-Sensed Images, An Introduction*, 2nd ed. New York: Wiley, 1999.
[17] J. C. Price, "On the information content of soil reflectance spectra," *Remote Sens. Environ.*, vol. 33, pp. 113–121, 1990.
[18] J. S. Tyo, A. Konsolakis, D. I. Diersen, and R. C. Olsen, "Principal-components-based display strategy for spectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 3, pp. 708–718, Mar. 2003.
[19] J. C. Price, "Comparison of the information content of data from the LANDSAT-4 thematic mapper and the multispectral scanner," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, pp. 272–281, May 1984.
[20] E. Simental, D. Ragsdale, E. Bosch, R. Dodge, Jr., and R. Pazak, "Hyperspectral dimension reduction and elevation data for supervised image classification," in *Proc. 14th Annu. Amer. Soc. Photogrammetry Remote Sensing Conf.*, Anchorage, AK, 2003, pp. 3–9.
[21] J. C. Price, "Information content of iris spectra," *J. Geophys. Res.*, vol. 80, no. 15, pp. 1930–1936, 1975.
[22] ——, "On the value of high spectral resolution measurements in the visible and near-infrared," in *Proc. 5th Int. Symp. Physical Measurements and Signatures in Remote Sensing*, Courchevel, France, 1991, ESA SP-319, pp. 131–136.
[23] ——, "Variability of high-resolution crop reflectance spectra," *Int. J. Remote Sens.*, vol. 13, no. 14, pp. 2593–2610, 1992.
[24] ——, "Band selection procedure for multispectral scanners," *Appl. Opt.*, vol. 33, no. 15, pp. 3281–3288, 1994.
[25] ——, "Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 1277–1285, Sept. 1997.
[26] M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Appl.*, vol. 19, no. 1, pp. 4–9, 1994.
[27] Y. L. Grossman, E. W. Sanderson, and S. L. Ustin, "Relationships between canopy chemistry and reflectance for plant species from Jasper Ridge biological preserve, California," in *Proc. IGARSS*, vol. 4, Pasadena, CA, Aug. 1994, p. 2357.
[28] A. S. Milman, *Mathematical Principles of Remote Sensing: Making Inferences From Noisy Data*. Chelsea, MI: Sleeping Bear, 1999.
[29] J. Y. Hardeberg, "Aquisition and reproduction of color images: Colorimetric and multispectral approaches,", Disseration.com, Parkland, FL, 2001.
[30] M. Partridge and R. Calvo, "Fast dimensionality reduction and simple PCA," *Intell. Data Anal.*, vol. 2, no. 3, pp. 1–14, 1998.
[31] W.K Pratt, *Digital Image Processing*. New York: Wiley, 1978.
[32] D. Hanselman and B. Littlefield, *Mastering MATLAB 6 A Comprehensive Tutorial and Reference*. Upper Saddle River, NJ: Prentice-Hall, 2001.
[33] C. J. Tucker and M. W. Garrat, "Leaf optical system modeled as a stochastic process," *Appl. Opt.*, vol. 16, no. 3, pp. 635–642, 1977.
[34] J. T. Woolley, "Reflectance and transmittance of light by leaves," *Plant Physiol.*, vol. 47, pp. 656–662, 1971.
[35] R. M. Devlin and A. V. Barker, *Photosynthesis*. New York: Van Nostrand Reinhold, 1971.
[36] M. Chelle, B. Andrieu, and K. Bouatouch, "Nested radiosity for plant canopies," *Vis. Comput.*, vol. 14, no. 3, pp. 109–125, 1998.

Ian E. Bell received the B.Sc.(Eng.) degree in computer engineering and the M.Sc. degree in computer science from Queen's University, Kingston, ON, Canada, in 1985 and 1987, respectively, and the Ph.D. degree in computer science from the University of Waterloo, Waterloo, ON, in 1996.

He worked in the computer graphics industry, specializing in avionics display simulation, from 1987 to 1996. He is currently a Computer Graphics and Medical Imaging Consultant. He was a Consultant in color and computer graphics for several years before accepting a position at the University of Waterloo from 1999 to 2003. He has developed graduate courses on color and on medical imaging and has given tutorials with Gladimir Baranoski at Eurographics and SIGGRAPH conferences on spectral color reproduction. His doctoral research used spline-based models for conventional and reflective color reproduction, work facilitated by fellowships with IBM's Center for Advanced Studies, Toronto, ON.

Gladimir V. G. Baranoski received the B.Sc. degree in mechanical engineering and the M.Sc. degree in computer science from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, the M.Sc. degree from Indiana University, Bloomington, and the Ph.D. degree from the University of Calgary, Calgary, AB, Canada, in 1998, both in computer science.

He is currently a Faculty Member with the University of Waterloo, Waterloo, ON. Recently, he has coorganized course presentations for conferences (Eurographics 2001 and 2002 and SIGGRAPH 2002 and 2003) and has developed graduate-level courses on topics related to the simulation of natural phenomena. His research interests include biologically and physically based rendering, as well as the simulation of natural processes involving radiative transfer of energy. The main topic of his doctorate was the development of reflectance and scattering models for plants. As a Postdoctoral Fellow at the University of Utah, he continued to work on the modeling and rendering of plasma phenomena.