

Effective Compression and Reconstruction of Human Skin Hyperspectral Reflectance Databases

Tenn F. Chen¹, Student Member, IEEE, and Gladimir V. G. Baranoski¹, Senior Member, IEEE

Abstract—In this paper, we investigate techniques for reducing the dimensionality of skin hyperspectral reflectance databases and maintaining a high degree of fidelity during data reconstruction. We compare results obtained using principal components analysis (PCA) with results provided by a piecewise PCA approach that explores the different roles performed by the main light attenuation agents acting within the cutaneous tissues in the ultraviolet (UV), visible and near-infrared (NIR) domains. Our investigation encapsulates not only skin spectral responses obtained by varying the contents of these agents, but also responses resulting from the absence of melanin pigmentation associated with the vitiligo condition.

Index Terms—skin, melanin, vitiligo, piecewise PCA.

I. INTRODUCTION

The access to robust skin hyperspectral reflectance databases is essential for the enhancement of a wide range of biomedical applications, from the assessment of biophysical properties of cutaneous tissues (*e.g.*, [1], [2], [3]) and the investigation of photobiological processes affecting skin health and appearance (*e.g.*, [4], [5], [6]), to the prevention and screening of diseases (*e.g.*, [7], [8], [9], [10]). Although such databases are still not readily available, this situation is likely to change rapidly with the development of new technologies for the accurate measurement and predictive modeling of skin responses within and outside the visible domain [11], [12].

Databases composed of measured reflectance data need to consider several dimensions associated with data capture procedures such as spectral resolution, illumination and viewing geometry. In the case of databases composed of modeled reflectances, one may need to account for additional dimensions associated with the combination of distinct specimen characterization parameters. Although, in theory, space could be saved by storing these parameters instead of the modeled reflectance data, the accurate computation of this data is often highly time consuming. Hence, for applications that demand high interactivity rates, it usually needs to be precomputed, stored offline and quickly accessed on the fly.

Regardless of the nature of the stored data, measured or modeled, clearly the main guidelines in the design of hyperspectral reflectance databases are compactness and low reconstruction error [13]. Here, we investigate how PCA [14], which is widely employed in many areas (*e.g.*, signal

processing [15], pattern recognition [16], colorimetry [17], applied meteorology [18] and remote sensing [19], just to name a few), can be used to achieve these goals with respect to hyperspectral skin responses. We also propose the application of a piecewise PCA approach (PPCA) based on characteristic cutaneous light attenuation profiles, which can lead to higher accuracy to computational cost (memory space) ratios for human skin hyperspectral data.

II. METHODS AND DATA

Consider a skin hyperspectral reflectance database represented by a $s \times w$ matrix M , where each row contains the reflectance spectrum of a given specimen, and each column stores the reflectance values with respect to the sample wavelengths. The goal is to reduce the dimensionality of M , *i.e.*, the PCA approach is applied to the entire dataset, rather than to each specimen's reflectance spectrum.

In order to implement PCA, we choose to use the singular value decomposition (SVD) technique [15] due to its numerical stability. Using this technique, one can obtain the $w \times w$ matrix V and the matrix $C = M * V$. While V stores the basis of principal components (as column vectors), C stores the coordinates (coefficients) relative to the basis (as row vectors).

Since the goal is to reduce the dimension of the data, one usually chooses a reduced number of components, $s' \ll s$, and stores smaller versions (the first s' columns) of C and V , more specifically the $s \times s'$ matrix C' and the $w \times s'$ matrix V' , respectively. The reconstructed reflectance spectrum of a selected specimen i , where $1 \leq i \leq s$, is then given by:

$$\rho_r = C'_i * (V')^T, \quad (1)$$

where C'_i corresponds to the i th row of matrix C' . It is worth noting that different implementations of PCA using SVD are possible. In particular, the implementation used in this investigation, whose mathematical details can be found elsewhere [20], aims to minimize the Euclidean norm of the difference between C and C' .

It has been observed that while the melanins (eumelanin and pheomelanin) and hemoglobins (oxy- and deoxygenated) dominate light attenuation within human skin in the UV-Visible (250-700 nm) range and their influence is increasingly diminished in the NIR-A (700-1400 nm) region, lipids and water start to contribute to light attenuation in this region and have a dominant role in the NIR-B (1400-2500 nm) range [21]. These well defined regions suggest that the use of an adaptive PCA approach may result in lower reconstruction errors and more compactness.

*This work was supported in part by the Natural Sciences and Research Council of Canada (NSERC) under Grant 238337.

¹ Tenn F. Chen and Gladimir V. G. Baranoski are with the Natural Phenomena Simulation Group, School of Computer Science, University of Waterloo, 200 University Avenue, Waterloo, Ontario, Canada N2L 3G1. gvgbaran@cs.uwaterloo.ca

In other words, instead of an integral application of PCA over the entire hyperspectral domain of interest (250-2500 nm), we can perform a piecewise application of PCA, in which each piece corresponds to one of the three regions (UV-Visible, NIR-A and NIR-B). Since the effectiveness of the PCA approach depends on the representativeness of the reflectance data used in the singular value decomposition, applying PPCA to matrices for each of the three regions yields different components than applying PCA to a matrix for the entire domain of interest. After eliminating components that have a small impact on the overall reflectance curves, reconstructing a particular curve with PPCA gives pieces for each region, which might not join smoothly at region boundaries. Although it would be possible to use a constrained PPCA algorithm to ensure smooth joins, we found such discontinuities to be negligible with respect to skin reflectance data.

In our experiments, we considered a skin hyperspectral reflectance database (SHRD) composed of 144 distinct skin directional-hemispherical curves (Figure 1) with a spectral resolution of 5 nm. These curves were obtained using a novel hyperspectral light transport model, henceforth referred to as HyLIoS (*Hyperspectral Light Impingement on Skin*) [12]. In their computation, we considered an angle of incidence of 10° and variations in the parameters associated with the main light attenuation agents: eumelanin (from 24 to 62.5 g/L), pheomelanin (from 1.5 to 5 g/L, dermal blood content (from 0.225 to 3.12 %), water content (from 26.25 to 93.75 %), and lipids content (from 11.32 to 0.25 %). Note that these parameter ranges are within physiologically acceptable limits indicated in the scientific literature [12].

Since relevant biomedical applications (*e.g.*, [1], [3]) require the use of reference reflectance curves obtained at amelanotic skin sites such as those found in vitiligo subjects, we have also included in the database amelanotic curves obtained by removing the melanin pigmentation. Considering that such curves markedly depart from typical skin reflectance curves, they also allowed us to expand our scope of *in silico* experimental observations.

For consistency with related works (*e.g.*, [13], [17]), we use the root mean square error (RMSE) measure in our experiments. The RMSE values were computed using the following expression:

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (\rho_m(\lambda_j) - \rho_r(\lambda_j))^2}, \quad (2)$$

where ρ_m and ρ_r correspond to the reflectance values (at a given wavelength λ_j) extracted from the modeled (HyLIoS) and reconstructed (PCA or PPCA) curves, respectively, and N corresponds to the total number of wavelengths sampled with a 5 nm resolution within a selected spectral region.

As one increases the number of components (and coefficients) used in the reconstruction of spectral curves, their RMSE values with respect to the original modeled curves decreases. The goal is to obtain the best compromise between the number of components and the RMSE. We selected as

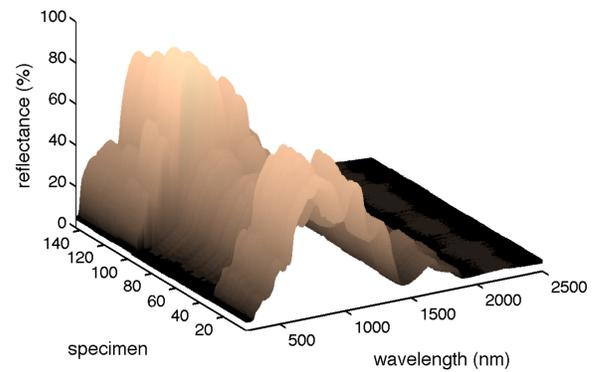


Fig. 1. Plot depicting the skin hyperspectral reflectance database (SHRD) employed in this investigation.

TABLE I

MAXIMUM RMSE VALUES COMPUTED FOR THE PCA RECONSTRUCTED CURVES (250-2500 nm) WITH RESPECT TO THE CURVES FOR THE MELANOTIC (96), AMELANOTIC (48) AND ALL SPECIMENS DEPICTED IN THE SHRD OBTAINED USING THE HyLIoS MODEL [12]. THE VALUES IN BOLDFACE FOR THE MELANOTIC AND AMELANOTIC SUBSETS AS WELL AS THE ENTIRE DATABASE CORRESPOND TO SPECIMENS S32, S135 AND S127, RESPECTIVELY.

| PCA Components | Specimens | | |
|----------------|---------------|---------------|---------------|
| | Melanotic | Amelanotic | All |
| 1 | 0.0504 | 0.0337 | 0.0738 |
| 2 | 0.0099 | 0.0172 | 0.0245 |
| 3 | 0.0058 | 0.0099 | 0.0171 |
| 4 | 0.0037 | 0.0030 | 0.0069 |
| 8 | 0.0014 | 0.0012 | 0.0023 |

reference for low-reconstruction error a RMSE value equal to 0.01, which has also been employed for this purpose in works involving other optically-complex natural materials such as plant leaves [13].

In order to further illustrate different levels of reconstruction fidelity, we have also generated skin swatches. In the case of visible reflectance data, we employed a standard XYZ to sRGB conversion procedure [22] considering a light source that approximates a standard D65 illuminant [23]. In the case of UV and NIR reflectance data, we integrated the reflectance values over their respective spectral region and applied a tinted grayscale filter to the resulting values, which are depicted in pseudo colors.

III. RESULTS AND DISCUSSION

Initially, we applied PCA to the SHRD melanotic subset, and computed the RMSE values associated with all reconstructed melanotic curves obtained using different numbers of components. The summary of the resulting maximum RMSE values provided in Table I indicates that only two components were required to reconstruct the melanotic curves with a RMSE below 0.01. This level of compression is significant, especially if compared to similar results obtained for plant leaves, which required at least five components [13]. A comparison between the original and reconstructed curve for specimen S32 (with the maximum RMSE value)

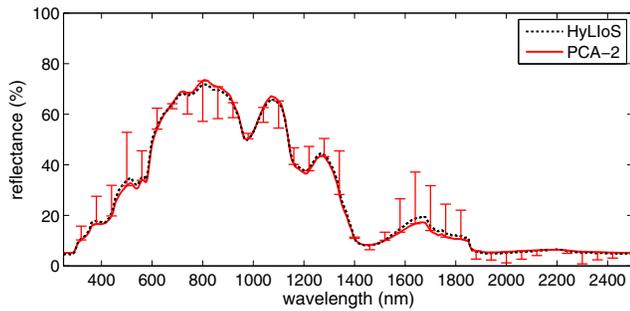


Fig. 2. Modeled (HyLIoS) and reconstructed (PCA using 2 components) curves for specimen S32. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions were 0.0116, 0.0102 and 0.0089, respectively.



Fig. 3. Skin swatches in the UV (left), visible (centre) and NIR (right) ranges obtained for specimen S32 using PCA with 1 component (top), 2 components (middle) and the original modeled curve (bottom) provided by HyLIoS [12].

is presented in Figure 2, and their close agreement is further illustrated by the skin swatches depicted in Figure 3.

We then applied PCA to the SHRD amelanotic subset, and computed the RMSE values associated with all reconstructed amelanotic curves obtained using different numbers of components. The summary of the resulting maximum RMSE values provided in Table I indicates that three components were required to reconstruct the amelanotic curves with a RMSE below 0.01. A comparison between the original and reconstructed curve for specimen S135 (with the maximum RMSE) is presented in Figure 4.

Finally, we applied the PCA approach to the entire SHRD, and computed the RMSE values associated with all reconstructed melanotic and amelanotic curves obtained using different numbers of PCA components. Due to the noticeable deviation of the amelanotic curves from typical skin reflectance curves, it was expected that a larger number of components would be required to keep the reconstruction error low. Although we needed to use more components (in comparison with the reconstruction of the curves in the melanotic subset), the summary of the resulting maximum RMSE values provided in Table I indicates that only four components were required to reconstruct all curves with a RMSE below 0.01. Hence, even considering extreme responses associated with amelanotic specimens, one can obtain a high compression accompanied by a relatively low-reconstruction error as illustrated by the comparison between the original and reconstructed curve for specimen 127 (with the maximum RMSE value) presented in Figure 5.

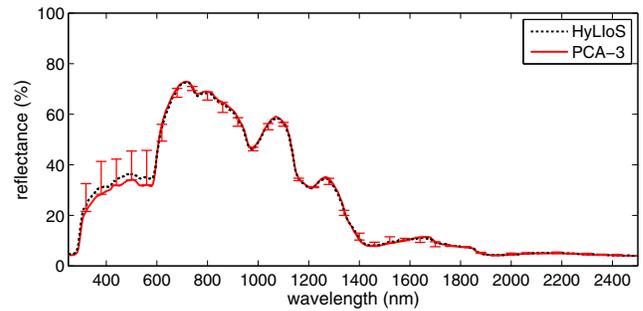


Fig. 4. Modeled (HyLIoS) and reconstructed (PCA using 3 components) curves for specimen S135. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions were 0.0207, 0.0054 and 0.0022, respectively.

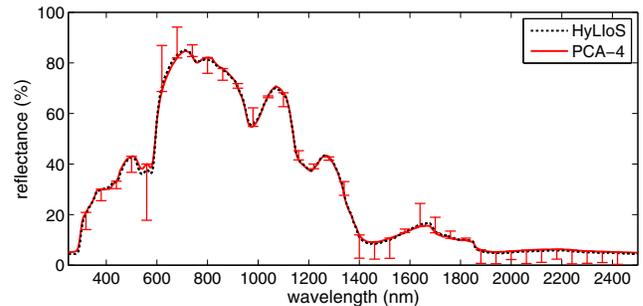


Fig. 5. Modeled (HyLIoS) and reconstructed (PCA using 4 components) curves for specimen S127. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions were 0.0125, 0.0042 and 0.0046, respectively.

When we repeated the experiments employing the PPCA approach (Figures 6 to 8), we were able to obtain even lower reconstruction errors using the same number of components. Besides the overall reduction of reconstruction errors, indicated by the lower RMSE values computed for the PPCA reconstructed curves, a more detailed inspection of the results provided by PCA (Figures 2, 4 and 5) and PPCA (Figures 6, 7 and 8) revealed a closer qualitative agreement provided by the latter approach, notably in the visible domain. This aspect is also highlighted by the comparison of PCA and PPCA derived skin swatches depicted in Figure 9.

We remark that light attenuation is not affected by the melanins and hemoglobins in the NIR-B. As a result, in this region of the light spectrum, one can observe less variations among the spectral profiles of the different specimens (Figure 1). This, in turn, can lead to closest reconstructions (smallest RMSE values) being obtained in the NIR-B.

IV. CONCLUSION

Our findings indicate that substantial storage savings can be obtained using a small number of PCA components in the reconstruction of skin hyperspectral reflectance curves. Moreover, the resulting low reconstruction errors can be further reduced using the same number of components by applying the proposed biophysically-based PPCA approach tailored for human skin hyperspectral reflectance databases.

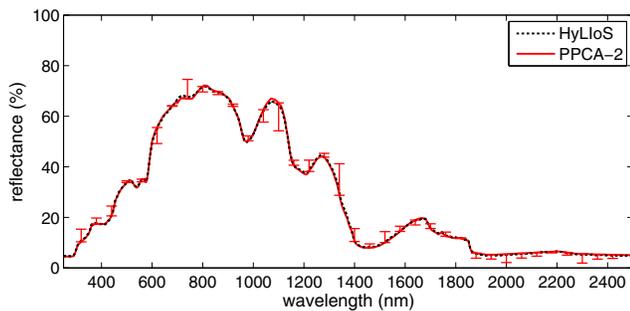


Fig. 6. Modeled (HyLIoS) and reconstructed (PPCA using 2 components) curves for specimen S32. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions are 0.0024, 0.0058 and 0.0024, respectively.

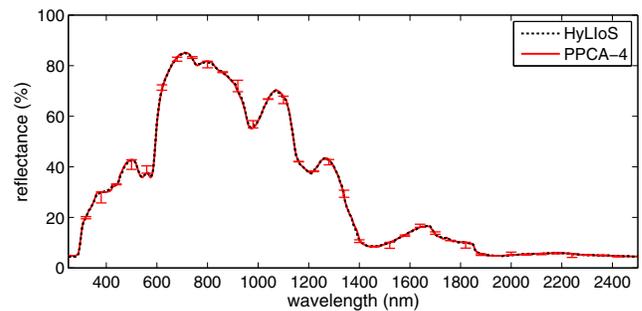


Fig. 8. Modeled (HyLIoS) and reconstructed (PPCA using 4 components) curves for specimen S127. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions were 0.0038, 0.0031 and 0.0010, respectively.

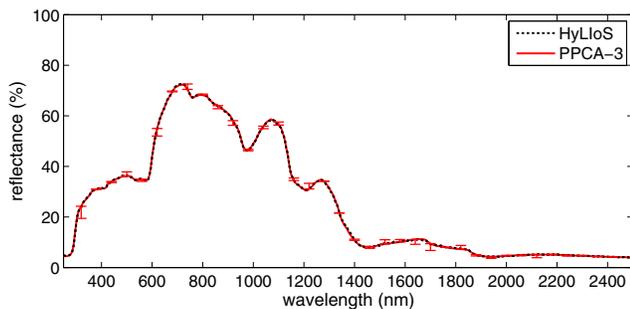


Fig. 7. Modeled (HyLIoS) and reconstructed (PPCA using 3 components) curves for specimen S135. Error bars represent differences between the curves, magnified by a factor of 10. Resulting RMSE values for the UV-Visible, NIR-A and NIR-B regions are 0.0018, 0.0026 and 0.0013, respectively.



Fig. 9. Skin swatches in the UV (left), visible (centre) and NIR (right) ranges obtained for specimen S127 using PCA with 4 components (top), PPCA with 4 components (middle) and the original modeled curve (bottom) provided by HyLIoS [12].

REFERENCES

- [1] N. Kollias and A. Baqer, "On the assessment of melanin in human skin *in vivo*," *J. Photoch. Photobio. B.*, vol. 43, no. 1, pp. 49–54, 1986.
- [2] J. F. Federici, N. Guzelsu, H. C. Lim, G. Jannuzzi, T. Findley, H. R. Chaudhry, and A. B. Ritter, "Noninvasive light-reflection technique for measuring soft-tissue stretch," *Appl. Opt.*, vol. 38, no. 31, pp. 6653–6660, Nov 1999.
- [3] J. Viator, J. Komadina, L. Svaasand, G. Aguillar, B. Choi, and J. Nelson, "A comparative study of photoacoustic and reflectance methods for determination of epidermal melanin content," *J. Invest. Dermatol.*, vol. 122, pp. 1432–1439, 2004.
- [4] C. Rosen, S. Jacques, M. Stuart, and R. Gange, "Immediate pigment darkening: visual and reflectance spectrophotometric analysis of action spectrum," *J. Photoch. Photobio. B.*, vol. 51, no. 5, pp. 583–588, 1990.
- [5] K. P. Nielsen, L. Zhao, J. J. Starnes, K. Starnes, and J. Moan, "The importance of the depth distribution of melanin in skin for DNA protection and other photobiological processes," *J. Photoch. Photobio. B.*, vol. 82, no. 3, pp. 194–198, 2006.
- [6] M. Yang, V. Tuchin, and A. Yaroslavsky, "Principles of light-skin interactions," in *Light-Based Therapies for Skin of Color*, E. Baron, Ed. London: Springer-Verlag, 2009, pp. 1–44.
- [7] V. Tuchin, *Tissue optics: light scattering methods and instruments for medical diagnosis*, ser. SPIE PM. Bellingham, WA, USA: SPIE/International Society for Optical Engineering, 2007.
- [8] G. Zonios, A. Dimou, I. Bassukas, D. Galaris, A. Tsolakidis, and E. Kaxiras, "Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection," *J. Biomed. Optics*, vol. 13, no. 1, p. 014017, 2008.
- [9] G. Baranoski, T. F. Chen, B. Kimmel, E. Miranda, and D. Yim, "On the noninvasive optical monitoring and differentiation of methemoglobinemia and sulfhemoglobinemia," *J. Biomed. Optics*, vol. 17, no. 9, pp. 097005–1–14, 2012.
- [10] P. Cavalcanti, J. Scharcanski, and G. Baranoski, "A two-stage approach for discriminating melanocytic skin lesions using standard cameras," *Expert. Syst. Appl.*, vol. 40, no. 10, pp. 4054–4064, 2013.
- [11] C. Cooksey, B. Tsai, and D. Allen, "A collection and statistical analysis of skin reflectance for inherent variability over the 250 nm to 2500 nm spectral range," in *SPIE Vol. 9082, Active and Passive Signatures IV*, G. Gilbreath and C. Hawley, Eds., 2014, pp. 908206–1–11.
- [12] T. Chen, G. Baranoski, B. Kimmel, and E. Miranda, "Hyperspectral modeling of skin appearance," *ACM Transactions on Graphics*, vol. 34, no. 3, pp. 31:1–14, April 2015.
- [13] I. Bell and G. Baranoski, "Reducing the dimensionality of plant spectral databases," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 57–577, March 2004.
- [14] G. Golub and C. V. Loan, *Matrix Computations*, 2nd ed. Baltimore, MD, USA: John Hopkins University Press, 1989.
- [15] W. Pratt, *Digital Image Processing*. New York, NY, USA: Wiley & Sons, 1978.
- [16] M. Partridge and R. Calvo, "Fast dimensionality reduction and simple pca," *Intelligent Data Analysis*, vol. 2, no. 3, pp. 1–14, 1998.
- [17] J. Hardeberg, *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches*. USA: Dissertation.com, 2001.
- [18] H. Huang and P. Antonelli, "Application of principal component analysis to high-resolution infrared measurement compression and retrieval," *J. Appl. Meteorol.*, vol. 40, pp. 365–388, March 2001.
- [19] J. Tyo, A. Konsolakis, D. Diersen, and R. Olsen, "Principal-components-based display strategy for spectral imagery," *IEEE T. Geosci. Remote*, vol. 41, no. 3, pp. 708–718, March 2003.
- [20] I. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002, pages 44–46.
- [21] R. Anderson and J. Parrish, "Optical properties of human skin," in *The Science of Photomedicine*, J. Regan and J. Parrish, Eds. N.Y., USA: Plenum Press, 1982, pp. 147–194.
- [22] G. V. Baranoski and A. Krishnaswamy, *Light & Skin Interactions: Simulations for Computer Graphics Applications*. Burlington, MA, USA: Morgan Kaufmann/Elsevier, 2010.
- [23] M. Stone, *A Field Guide to Digital Color*. Natick, MA, USA: AK Peters, 2003.