

SVD vs PCA: Comparison of Performance in an Imaging Spectrometer

Wilma R. Oblefias¹, Maricor N. Soriano², and Caesar A. Saloma³

National Institute of Physics, University of the Philippines

Diliman, Quezon City 1101

E-mail: ¹woblefias@nip.upd.edu.ph; ²msoriano@nip.upd.edu.ph;

³csaloma@nip.upd.edu.ph

ABSTRACT

The calculation of basis spectra from a spectral library is an important prerequisite of any compact imaging spectrometer. In this paper, we compare the basis spectra computed by singular-value decomposition (SVD) and principal component analysis (PCA) in terms of estimation performance with respect to resolution, presence of noise, intensity variation, and quantization error. Results show that SVD is robust in intensity variation while PCA is not. However, PCA performs better with signals of low signal-to-noise ratio. No significant difference is seen between SVD and PCA in terms of resolution and quantization error.

INTRODUCTION

A common method to measure image spectrum is to use a monochrome charge-coupled device (CCD) camera with narrowband interference filters. For example, to measure the spectrum from 400 to 700 nm at 1 nm resolution requires 301 filters and the same number of images. Each color or grayscale value of the image corresponds to the intensity of the band pass of the filter used.

Singular-value decomposition (SVD) and principal component analysis (PCA) (Hair et al., 1998) are used to compress multivariate data such as spectra. Weighted spectral estimation can be obtained by linear superposition of the basis spectra computed using these techniques.

We have designed, implemented, and characterized an imaging spectrometer microscope for measuring fluorescent and bioluminescent spectra (Oblefias, 2004). An imaging spectrometer microscope, as its

name suggests, is a microscope that not only delivers the magnified image of a slide specimen but gives its spectrum at each image point as well. This paper presents the estimation performance of our device using the basis spectra calculated from SVD and PCA. The two methods are compared by measuring the fidelity of estimation with respect to spectral resolution, noise level, intensity variation, and quantization.

METHODS

SVD is a statistical method that looks for the component space where the data are most efficiently represented. If the data to be analyzed are spectra, each wavelength λ of the spectrum will be considered as one component or axis. Thus, a spectral library with spectrum from 400 to 700 nm at resolution $\Delta\lambda$ equal to 1 nm will have 301 axes. Each spectrum will be considered as data point. Figure 1 shows an example of data set with two components (x_1 and x_2).

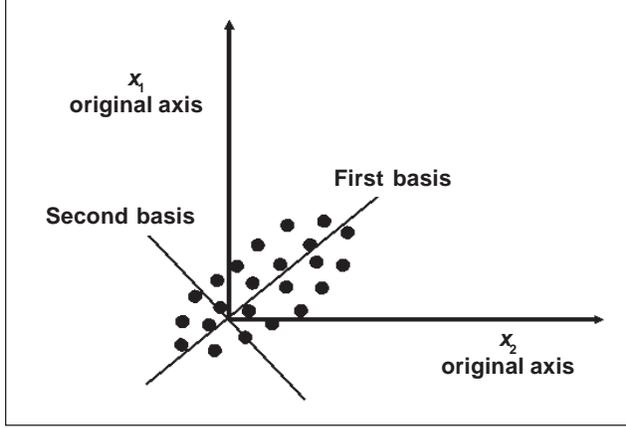


Fig. 1. Diagram of SVD for two-dimensional data.

In SVD the optimum component space is found by rotating the coordinate axes with respect to the origin. The first basis is the axis about which the data are most widely spread, or where the data have the largest variance. The second basis is the second axis, perpendicular to the first, which has the next highest data variance, and so on.

PCA, on the other hand, is a variation of SVD. Its main difference with SVD is that the origin is translated to the mean. Then the first basis is the axis about which the data are most widely spread. Succeeding bases are calculated in the same way as that of the SVD.

When the set of basis spectra $e_i(l)$ are obtained, the estimated spectrum $C_{est}(l; x, y)$ of the reference spectrum $C(l; x, y)$ can be expressed as a weighted linear superposition of the first few significant $e_i(l)$:

$$C_{*1}(\lambda; x, y) = \sum_{n=1}^N a_n(x, y) e_n(\lambda), \quad (1a)$$

$$C_{*1}(\lambda; x, y) = \sum_{n=1}^N a_n(x, y) e_n(\lambda) + \langle C(\lambda) \rangle, \quad (1b)$$

where $\langle C(l) \rangle$ is the mean of the spectral library $\{C_k(l)\}$ and a_n is the coefficient of the n th $e(l)$. The summation is taken from $n = 1$ to $n = N$, where N is the number of $e_i(l)$ that is utilized for estimation. “a” in the equation number refers to SVD, while “b” refers to PCA.

Relating the estimated spectrum to the output of the camera, we obtain

$$Q = Ta, \quad (2a)$$

$$Q = Ta + Q_{\text{mean}}, \quad (2b)$$

where Q and Q_{mean} are M -column vectors containing the channel output of a color camera and average color of the spectral library, respectively. T is an $M \times N$ transformation matrix that maps the expansion coefficients in a to the image space colors Q (Soriano et al., 2002; Saloma et al., 2004).

In spectral imaging, the image output channels $\{Q_m(x, y)\}$, the spectral library $\{C_k(l)\}$, and the basis spectra $e_i(l)$'s are known and the immediate task is to determine the component values of a . After a is known, one proceeds via Eq. (1) to solve for the corresponding spectral estimation $C_{est}(l; x, y)$ which describes the optical spectrum at location (x, y) of the image of the spatially extended fluorescent sample.

For a colored image, the values for the different $Q_m(x, y)$'s for every pixel location (x, y) of the two-dimensional image are taken from the R, G, and B channel outputs ($M = 3$) of the camera. The unknown coefficients $\{a_n(x, y)\}$ are determined via

$$a = T^{-1}(Q) \quad (3a)$$

$$a = T^{-1}(Q - Q_{\text{mean}}), \quad (3b)$$

where T^{-1} is the inverse of T . The inversion matrix T^{-1} is defined only if T is a square matrix because the size of T is equal to $M \times N$, i.e., T^{-1} exists only if $N = M$.

To increase the number M of color channels, the sample is image-captured with a lightly colored transmission filter placed before the camera. With the insertion of a filter, the fluorescent sample is imaged under three more independent channels ($M = 6$), in addition to the original three (for R, G, B) provided by the three CCD camera in the absence of a filter (Imai et al., 2000).

Accuracy of estimation was measured using fidelity f given by

$$f = 1 - \frac{\langle C - C_{*1} \rangle}{\langle C^2 \rangle} \quad (4)$$

where $\langle \cdot \rangle$ is the average value and C is the theoretical spectrum.

Fidelity describes the general similarity between theoretical and estimated spectra. Perfect estimation occurs when $f = 1$.

RESULTS AND DISCUSSIONS

Bases spectra were computed using PCA and SVD from 423 spectra of fluorescence emissions.

A Gaussian emission spectrum

$$C(\lambda; \sigma, \lambda_0) = \exp\left[-\frac{1}{2}\left(\frac{\lambda - \lambda_0}{\sigma}\right)^2\right], \quad (5)$$

where λ_0 is the peak wavelength and σ is the variance related to the width of the spectrum, was used to determine the narrowest spectrum that can be estimated. It was found that 15 basis spectra can be used to estimate a spectrum whose σ is 14 nm and fidelity is approximately equal to unity. This is equivalent to a full width at half maximum (FWHM) of 33 nm. Variance of less than 14 nm, a narrower spectrum, does not give estimation merit of unity even when the number of basis spectra are equal to 15.

Using $\sigma = 14$ nm, two Gaussian curves were superimposed with different λ_0 . The minimum peak separation that can be resolved both by SVD and PCA is 28 nm. Two peaks are said to be resolved if the ratio of the intensity at the midpoint to that at the maxima is 0.811 in accordance with the Rayleigh criterion.

The smaller peak separation of the two Gaussian curves is estimated as a unimodal spectrum. More than 15 basis spectra, equivalent to more than five colored images, are needed to estimate an emission spectrum that has a FWHM of less than 33 nm. The same number of basis spectra can also resolve two peaks that are separated by less than 28 nm.

In this study, however, the above result is already sufficient to estimate the emission of bioluminescent and naturally fluorescing samples since they are not usually narrow. Two emission peaks also rarely occur (Herring, 1993).

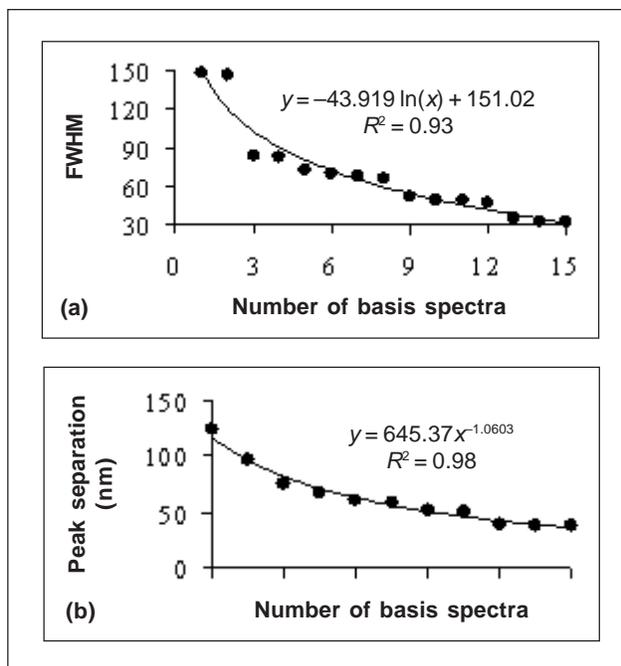


Fig. 2. Resolution with increasing number of basis spectra. (a) minimum FWHM that can be estimated and (b) minimum peak separation that can be resolved.

Figure 2 shows the minimum FWHM and peak separation of a bimodal spectrum that can be resolved as the number of basis spectra increases using SVD and PCA. Bimodal spectrum cannot be resolved with less than five bases no matter how large the FWHM is. Using the best fit curve, a FWHM of 1 nm and a peak separation of 2 nm can be resolved by using 301 basis spectra. However, this is not practical for our purpose because fluorescing bioluminescent samples do not have narrow spectra.

Experimentally, the increase or decrease in the intensity of emission of the sample cannot be fully controlled. The greater the intensity of the excitation light, the greater the intensity of the emission. The sample may also undergo photobleaching after long exposure to the excitation source that decreases the emission intensity. Ideally, the estimated emission spectrum should not change if the intensity of the excitation is varied since emission spectrum is independent of the intensity of the excitation source.

Figure 3 shows the effect in average fidelity computed from 423 spectra in changing the intensity of the sample emission. SVD is robust against changes in intensity.

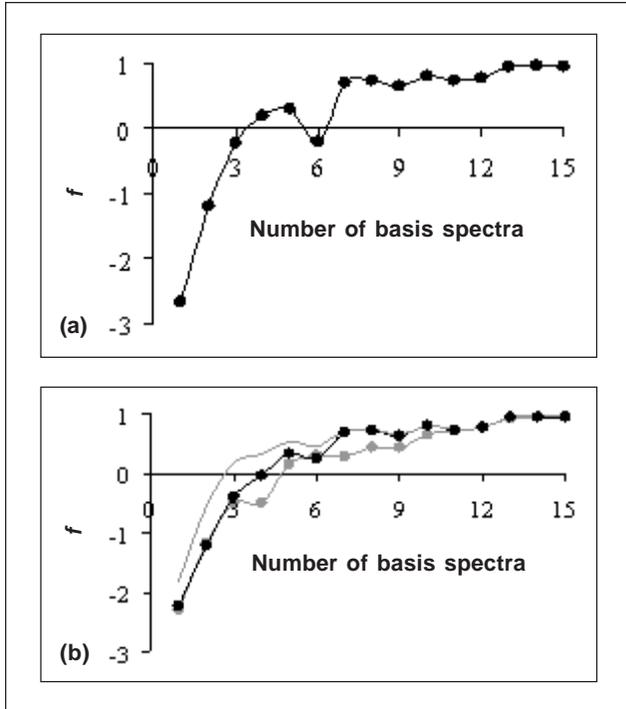


Fig. 3. Fidelity with intensity. Light and dark lines are the result with intensity reduced to one fourth and intensity increased by four, respectively. (a) SVD and (b) PCA. Continuous light line in b is for no variation in intensity. Standard deviation is not shown.

Estimation merit is not affected by either an increase or a decrease of intensity.

PCA, however, is intensity dependent. Increase in intensity has little effect as the number of basis spectra is increased. At the seventh basis spectra, estimation is the same as if there is no change in intensity. To remove the dependency of PCA with intensity, at least 12 basis spectra must be used. This is equivalent to using at least four images.

Noise becomes more apparent with signals of low intensity. Increase in noise results in low SNR. Figure 4 shows the effect of additive white Gaussian noise at estimation merit. The SNR was varied and calculated as

$$\text{SNR} = 20 \log \left(\frac{C^o}{C^D} \right) \quad (6)$$

where C^o is the energy of the theoretical signal and C^D is the difference between the energy of the theoretical signal and the energy of the actual signal. The unit of the SNR is decibel (dB).

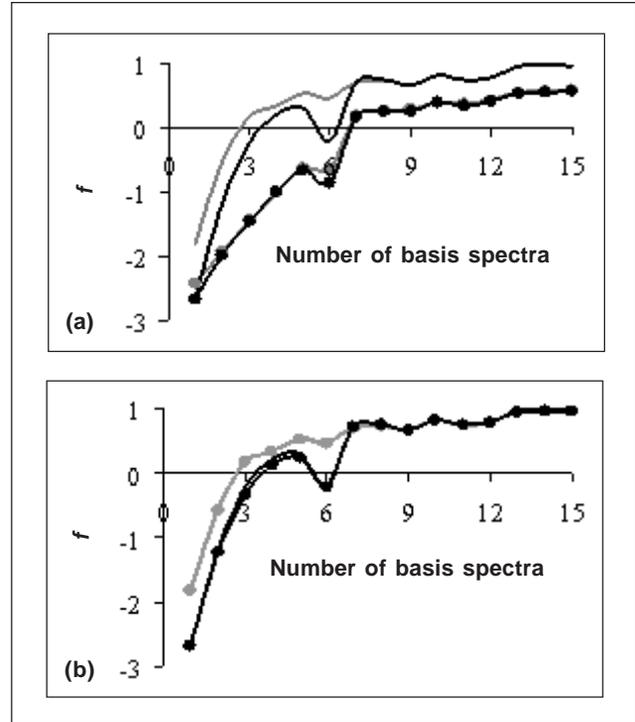


Fig. 4. Fidelity with the addition of white Gaussian noise. (a) SNR = 20 dB, and (b) SNR = 40 dB. Dark and light lines are the result using SVD and PCA, respectively. Continuous dark line is for SVD without noise and continuous light line is for PCA. Standard deviation is not shown.

At low SNR, estimation merit is greatly affected. Increasing the number of basis spectra when SNR is equal to 20 dB does not improve the estimation. As SNR increases, the difference between the estimation merit with noise and that without noise decreases. At high SNR, even if noise is present, its effect is negligible when the number of basis spectra is increased. When SNR is gradually increased, noise is negligible in SVD when SNR is 42 dB, while in PCA, noise is negligible when SNR is 40 dB. Thus, at low SNR PCA performs better than SVD.

Quantization error is the result of the finite number of bits in the digitizer that converts the voltage output of the camera into grayvalue levels. Figure 5 shows the fidelity using 8 bits for each camera channel. Improvement in spectral estimation can be seen from one to seven basis spectra. Spectral estimation degrades after the seventh basis spectra. This means that for such number of bits, using up to three images only is advisable.

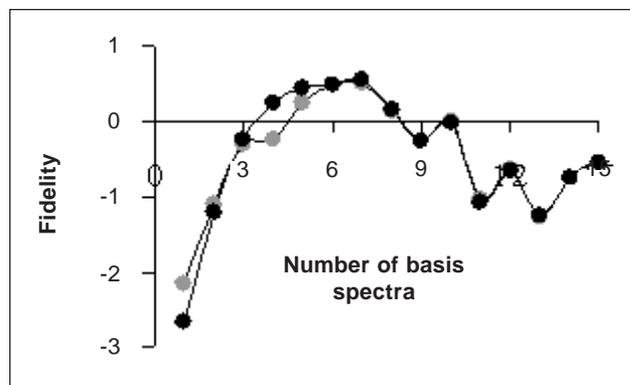


Fig. 5. Fidelity with 8 bits digitizer. The dark line is for SVD without noise and the light line is for PCA. Standard deviation is not shown.

It was observed, however, that using more than three images can be used to get better estimation for samples having a FWHM of greater than 56 nm. Fidelity of narrower spectra follows the curve of Fig. 5.

As the number of bits increases, the graph of fidelity with the number of basis spectra approaches the value of Fig. 3. Significant improvement is observed up to 16 bits. Using more than 16 bits does not give further improvement. Comparison of performance of SVD and PCA when quantization is present shows no significant difference.

SUMMARY

Singular-value decomposition (SVD) and principal component analysis (PCA) were used to calculate the basis spectra.

Using five colored images (15 basis spectra), both SVD and PCA can accurately estimate a unimodal spectrum whose minimum FWHM is 33 nm and a bimodal spectrum with peak separation of 28 nm.

When intensity variation is considered, SVD has a greater advantage than PCA. Thus, in an actual experiment provided that the sample has enough SNR, SVD is recommended since the intensity of emission cannot be controlled due to photobleaching and intensity variation of the excitation source.

The threshold SNR to have good estimation merit is 42 dB for SVD, while it is 40 dB for PCA. A SNR that is lower than the threshold fails to estimate the spectrum correctly even when the number of basis spectra is increased. Thus, PCA should be used for low SNR as long as photobleaching and intensity variation of the sample can be ignored.

The digitizer should have at least 8 bits for samples with FWHM greater than 56 nm. Increasing the number of bits shows improvement in spectral estimation. However, using more than 16 bits does not give further information.

ACKNOWLEDGMENT

This work received financial support from the Office of the Vice Chancellor for Research and Development (OVCRD) of the University of the Philippines Diliman.

REFERENCES

- Hair Jr., J., R. Anderson, R. Tatham, & W. Black, 1998. *Multivariate data analysis*, 5th ed. Prentice-Hall, Englewood Cliffs, New Jersey.
- Herring, P., 1993. The spectral characteristics of luminous marine organisms. *Proc. R. Soc. London*. B220: 183–217.
- Imai, F., R. Berns, & D. Tzeng, 2000. A comparative analysis of spectral reflectance estimated in various spaces using a trichromatic camera system. *J. Imaging Sci. Technol.* **44**: 280–287.
- Oblefias, W., 2004. Spectral imaging of fluorescent and bioluminescent samples. M.S. thesis, University of the Philippines, Diliman.
- Saloma, C., W. Oblefias, & M. Soriano, 2004. Spectral microscopy of live luminescent samples. In *Nanophotonics: Integrating photochemistry, optics, and nano/bio materials studies*. Elsevier: Chap. 30.
- Soriano, M., W. Oblefias, & C. Saloma, 2002. Fluorescence spectrum recovery from image color and non-negativity constraint. *Opt. Express*. 10: 1458–1464.