SAND99-0148C

Biomedical Applications of the Information-efficient Spectral Imaging Sensor (ISIS)

Stephen M. Gentry[a]
Sandia National Laboratories, Albuquerque, NM 87185-0980

Richard Levenson[b] **M.D.**
Carnegie Mellon University, Pittsburgh, PA 15213

Key Words: spectral imaging, dimensionality reduction, principal component, projection pursuit, basis pursuit, band selection, optimal spectral filter, tunable spectral filter

ABSTRACT

The Information-efficient Spectral Imaging Sensor (ISIS) approach to spectral imaging seeks to bridge the gap between tuned multispectral and fixed hyperspectral imaging sensors. By allowing the definition of completely general spectral filter functions, truly optimal measurements can be made for a given task. These optimal measurements significantly improve signal-to-noise ratio (SNR) and speed, minimize data volume and data rate, while preserving classification accuracy. The following paper investigates the application of the ISIS sensing approach in two sample biomedical applications: prostate and colon cancer screening. It is shown that in these applications, two to three optimal measurements are sufficient to capture the majority of classification information for critical sample constituents. In the prostate cancer example, the optimal measurements allow 8% relative improvement in classification accuracy of critical cell constituents over a red, green, blue (RGB) sensor. In the colon cancer example, use of optimal measurements boost the classification accuracy of critical cell constituents by 28% relative to the RGB sensor. In both cases, optimal measurements match the performance achieved by the entire hyperspectral data set. The paper concludes that an ISIS style spectral imager can acquire these optimal spectral images directly, allowing improved classification accuracy over an RGB sensor. Compared to a hyperspectral sensor, the ISIS approach can achieve similar classification accuracy using a significantly lower number of spectral samples, thus minimizing overall sample classification time and cost.

INTRODUCTION

Complexity of spectral imaging approaches vary from the most simple red, green, blue (RGB) cameras taking three simple fixed measurements, through liquid crystal tunable filters taking a number of arbitrary band-pass images, up to Fourier transform spectral imagers, taking hundreds of spectral samples per scene point. Each approach increases data volume, and each increases measurement time for a fixed signal-to-noise ratio (SNR). Therefore applications requiring higher spectral resolution are burdened by a corresponding increase in data volume, and decrease in speed of operation. Increased data sizes also decrease the rate at which image analysis and classification can occur. Therefore, for clinical biomedical applications, high spectral resolution correlates both to an increase in cost and complexity of the optical analysis system, and to net decrease in speed of sample acquisition and classification. Since sample analysis speed is typically correlated to cost per sample, and speed is correlated to number of spectral measurements, *the spectral method that achieves the maximum classification accuracy in the minimum spectral data samples will ultimately minimize overall analysis cost per clinical sample.*

---

[a] Correspondence: email: smgentr@sandia.gov, (505) 845-9473

[b] Correspondence: email: rml@andrew.cmu.edu, (412) 422-9385

WHAT IS ISIS?

The Information-efficient Spectral Imaging Sensor (ISIS)[1,2,3,4] approach breaks this previously necessary correlation between spectral resolution and data volume, speed, and cost, by allowing truly optimal spectral measurements to be defined for a given application. This is accomplished through optical design concepts that allow arbitrary definition of spectral filter functions that can have positive and negative spectral weights. These optimal measurements carry the maximum discrimination information in the fewest number of data samples. In the context of this paper, such a measurement is coined a spectral basis 'vector'. An example of such a vector is shown in figure 1.

Figures 2 and 3 illustrate one type of optical instrument capable of capturing spectral basis vectors directly. Shown is the ISIS field test instrument, constructed to evaluate the potential of ISIS applications for remote sensing. Such a device would be similar to that used in a biomedical application. Functionally, the instrument operates as follows. The instrument is a push-broom imager that takes a line image and scans it over the illuminated sample. Referring to figure 2, the light enters from the left through collection optics and passes through the input slit. The slit image is then spectrally dispersed and imaged onto a liquid crystal spatial light modulator (SLM) which can be adjusted to transmit or reject the various spectral components. The light is then spectrally de-dispersed and imaged on to a linear array. Since the liquid crystal SLM requires polarized light to operate, two linear polarization components are split off in the fore optics to create two independent channels without loss of light. To achieve the basis vector shown in figure 1, a spatially variable transmission function is applied to each of the SLMs and the resulting signal at the two detectors is differenced to achieve positive and negative weights. This optical design approach is only one of a number of optical design approaches recently included for patent application.[5] Other researchers are investigating similar optimal vector approaches using acousto-optic tunable filters[6] or dielectric coating stacks.[7] The basic thesis of all these approaches is to directly acquire optimal spectral measurements. In doing so, SNR and speed are optimized. For all these systems, development of the optimal spectral measurements would require general spectral process analysis done at moderate to high spectral resolution using traditional HSI or MSI approaches. Following is a biomedical example of developing those optimal measurements.

BIOMEDICAL EXAMPLES OF OPTIMAL BASIS VECTORS

This paper considers two sample ISIS applications: prostate and colon cancer screening. Samples were obtained that had been formalin-fixed and stained with hematoxylin and eosin, standard dyes used in pathology. Even though there are only two dyes in this mixture, they have the potential to interact with sample constituents, and with each other, causing complex spectral features. Hyperspectral images of the two test samples were taken using a Fourier transform spectral imaging system at 53 spectral samples over 400 – 700 nm, and 107 spectral samples over 450 – 700 nm, respectively.

The prostate and colon hyperspectral data sets were then analyzed and expert classified[+]. The RGB images of the samples and the associated hyperspectral reference classifications are shown in figures 4 and 5. These classifications were utilized as the reference classifications for the following evaluations. Based on these classifications, training and test regions were selected in each image as shown in figures 6 and 7. Training and test sample regions were mutually exclusive to reduce the potential for training on data peculiarities. Spectral class statistics were then generated for the training regions using MultiSpec[8] and methods outlined in the software manual.[9] These statistics provide the basis for MultiSpec to compute optimal spectral basis vectors.

MultiSpec contains functionality that enables the definition of two optimal basis sets, principal components[10] (PC), and projection pursuit[11] (PP). The former is a basis set that describes the spectral

---

[+] R. Levenson used pathology experience to select key spectral scene constituents, and representative spectra. The entire scene was then classified using a mean square error classifier.

DISCLAIMER

directions in N space* of decreasing total training set variance. The latter describes the directions that maximize class separation. Both methods are considered potential optimal sets for ISIS consideration. Since these basis functions are simply inner product functions that are applied to the original hyperspectral data, they can be considered to be linear spectral filter functions with positive and negative weights – just the optimal filter functions ISIS needs to optimize information content and minimize data volume. The software computes a number of these basis vectors, N, in either the PC or PP cases that maximally describe the training classes.

MultiSpec also contains the ability to sort through all NxN combinations of these basis vectors to determine the most optimal subset, M, for class discrimination. For this analysis, M was chosen initially to be three, to match the dimensionality of the RGB measurement. Using this method and the training sample regions, three basis vectors sets were evaluated:

- an RGB basis set derived by subdividing the entire spectral region into three equal sections,
- three optimal PC spectral basis vectors,
- three optimal PP spectral basis vectors.

The optimal PC and PP basis vectors for the prostate cancer example are shown in Figures 8 and 9 respectively, and Figures 10 and 11 for the colon cancer example. Note that the PC basis vectors are broad in spectral feature, while the PP basis vectors highlight specific narrower regions in the spectra.

CLASSIFICATION ACCURACY RESULTS

Using the RGB or optimal basis sets defined above, MultiSpec was used to classify the test regions using a standard quadratic gaussian maximum likelihood method.[12] From these results, classification accuracy was determined by the percentage of correctly classified pixels in a test class. The classification accuracy results for the three vector measurements are shown below in figures 12 and 13. The relative improvements of the optimal basis sets over RGB are shown in figures 14 and 15.

In the prostate cancer example, classification accuracy on non-essential components was quite high with all basis sets, above 95%. But on the more critical elements for making malignant determination, epithelial and myoepithelial nuclei, the RGB performance was as low as 80%. Here choice of an optimal basis set improves classification accuracy between 4% and 8%.

In the colon cancer example, classification accuracy on non-essential components was also quite high, all again above 97% for all basis sets. The only anomaly in the results is that the PP basis set appears to have lower classification accuracy than RGB or PC in the Cytoplasm 1 case. Evaluation of the classification data indicates that this reduction in classification accuracy was the result of switching a number of samples between Cytoplasm classes and not reclassifying to more critical classes. This error is discussed in more detail below. Most importantly, discrimination between the malignant and benign cell nuclei is significantly improved by use of optimum basis vector sets. In the case of benign nuclei, the PP optimal vectors achieve 92% accuracy, up from 72% obtained with the RGB measurement. This represents a 28% relative improvement in classification accuracy. For the malignant nuclei, the PP vectors achieve 96% accuracy, up from 79% achieved with the RGB measurement. This represents a 22% relative improvement in classification accuracy.

This increase in accuracy can make a significant difference in the diagnostic outcome of a particular sample. In the colon cancer example here, the difference in final result is dramatic. Figure 16 shows the reference classification as well as that using each of the three basis sets. The important difference is seen between the classification result using the RGB basis set and the PP basis set. In the RGB classification, the cell nuclei in the upper left quadrant of the sample are classified as malignant, while in both the

---

* 'N space' is the N dimensional subspace defined by N independent spectral measurements.

reference and the PP classification, these cells are classified as benign. Obviously, in this example, the choice of an optimal basis set *significantly* effects the diagnostics outcome.*

HOW MANY VECTORS ARE OPTIMUM?

The original choice of three element basis vector sets was based on matching the size of the three element RGB basis set. The question is now posed – how many vectors are actually necessary to maximize classification accuracy? Is it more or less than three?

The same procedure was followed as described previously, using MultiSpec to optimize the vector set of length M, where M varies up from 1 to 9. Figures 17 and 18 show the classification accuracy of the key components of the prostate and colon cancer samples as a function of RGB, PC, and PP basis set length. For the prostate cancer case, figure 16 shows that all three basis sets gain significant discrimination going from a single optimum basis function to two, and little difference between two and three. Therefore in this application, two optimal spectral basis vectors may be enough. In the colon cancer example, figure 18, the RGB and PP basis sets gain significantly from a single measurement up to two. The RGB case maximizes at two, but the PP case gains an additional 5% from two to three. After 3 the PP case is flat. On the other hand, the PC basis set continues to climb up through five basis vectors where the classification accuracy of the best three PP vectors is matched. These results bring two conclusions. First, the dimension of either problem is approximately three, and second that the PC vectors do not contain the same discrimination capability of the more optimum PP vectors. This is not entirely surprising since the PP vectors are designed in MultiSpec to maximally separate the expected classes.

One final problem needs addressing. Previously, it was shown that there was a net decrease in classification accuracy of Cytoplasm 1 in the colon cancer example when using PP vectors relative to RGB. Figure 19 shows the classification accuracy of all the colon cancer sample classes as a function of optimal measurement dimension. The plot indicates that all classes improve in classification accuracy from two measurements up to three, except Cytoplasm 1. Cytoplasm 1 starts high and drops at three basis vectors and then stays completely flat above three dimensions. The authors conclude that this indicates a problem with the Cytoplasm 1 test set possibly having some originally misclassified pixels that should have been correctly classified as a different type of cytoplasm. Thus the drop in classification accuracy is not a reflection on the optimal basis set, but reflects an error in the original reference classification.

Figure 19 also reinforces the conclusion that the three optimal basis functions carry virtually all the classification information of the entire data set. With the exception of malignant cell nuclei, all sample components maximize classification accuracy at three measurements or less. For malignant nuclei, the majority of classification can be done with three measurements, and only 2% additional accuracy is added in jumping from three to ten measurements.

CONCLUSIONS

The results presented here are an example of biomedical applications of the ISIS spectral sensing approach where the majority of critical classification information can be achieved with three optimally chosen spectral basis vectors. These optimal basis vectors are particularly useful in discerning the difference between benign and malignant cell nuclei, particularly in the colon cancer application where a 28% improvement was shown over an RGB sensor. Of the two optimal vector sets chosen, projection pursuit vectors are more effective in classifying these samples at lower measurement dimension than principal component vectors. This is expected since projection pursuit vectors are designed in the MultiSpec software to maximally separate the expected classes, not merely describe total training set variance. It is concluded that in practice, some similar optimal spectral basis vector sets would significantly improve the classification accuracy of cancer screening over a competing RGB system. Most importantly, an ISIS style spectral imager can acquire these optimal spectral images directly, allowing improved classification

---

* In actual practice, other spatial morphological features are used to differentiate benign and malignant features and actual cell and sample classification results would be better than represented here.

accuracy over an RGB sensor. Compared to a hyperspectral sensor, the ISIS approach can achieve similar classification accuracy using a significantly lower number of spectral samples, thus minimizing overall sample classification time and cost.

ACKNOWLEDGEMENTS

The authors would like to recognize David Landgrebe and Larry Biehl of Purdue University for use of the MultiSpec software (provided free of charge over the internet - dynamo.ecn.purdue.edu/~biehl/ MultiSpec). The software is both easy to use and well documented. MultiSpec also directly generated many of the figures in this work.

The authors would like to also recognize previous contributors to the ISIS concept development: Andy Boye, Carter Grotbeck, Brian Stallard, Bill Sweatt, (Sandia National Laboratories), Michael Descour, (University of Arizona). The ISIS field test instrument was developed and built with the assistance of the following scientists and engineers: Steve Gentry, Les Krumel, Jim Daniels, David Smith, Dick Blake, Mike Hinckley (Sandia National Laboratories).

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-ACO4-94AL85000.

References:

[1] B. R. Stallard, S. M. Gentry, W. C. Sweatt, S. E. Motomatsu, C. A. Boye, "Principals and Satellite Applications of the Information-Efficient Spectral Imaging Sensor" , SAND97-1493, Sandia National Laboratories Report, Albuquerque, NM, 1997

[2] S. M. Gentry, B. R. Stallard, "Theoretical Analysis of the Sensitivity and Speed Improvement of ISIS over a Comparable Hyperspectral Imager" , SAND98-1887, Sandia National Laboratories, Albuquerque, NM, 1998

[3] W. C. Sweatt, C. A. Boye, S. M. Gentry, M. R. Descour, B. R. Stallard, C. L. Grotbeck, "ISIS; An Information-Efficient Spectral Imaging System", *Imaging Spectrometry IV*, M. R. Descour, S. S. Shen Ed., Proc. SPIE, Vol. 3438, pp. 98-106, San Diego, 1998

[4] B. R. Stallard, , S. M. Gentry, "Construction of filter vectors for the Information-efficient Spectral Imaging Sensor", *Imaging Spectrometry IV*, M. R. Descour, S. S. Shen Ed., Proc. SPIE, Vol. 3438, pp. 172-182, San Diego, 1998

[5] *Programmable Spectral Filter*, Provisional Patent, #60/093,732, 7/17/98

[6] P.I. Shnitser, I.P. Agurok, "Spectrally Adaptive Light Filtering", Proc. SPIE, vol. 3140 , p.117-27, 1997

[7] M. P. Nelson, J. F. Aust, J. A. Dobrowolski, P. G. Verly, M. L. Myrick, "Multivariate Optical Computation for Predictive Spectroscopy", *Analytical Chemestry*, Vol. 70, No. 1, pp. 73-82, Jan. 1998

[8] D. Landgrebe, L. Biehl, MultiSpec, Purdue University, Pittsburgh PA, (dynamo.ecn.purdue.edu/ ~biehl/MultiSpec)

[9] D. Landgrebe and L. Biehl, "An Introduction to MultiSpec" , v. 7.97, (dynamo.ecn.purdue.edu/~biehl/ MultiSpec), 1997

[10] K. Fukunaga , *Introduction to Statistical Pattern Recognition*, Academic Press, p. 417, 1990

[11] L. O. Jimenez, , D. Landgrebe, "High Dimensional Feature Reduction via Projection Pursuit", TR-ECE 96-5, School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285, (dynamo.ecn.purdue.edu/~biehl/MultiSpec), 1995

12 D. Landgrebe and L. Biehl, "An Introduction to MultiSpec", (dynamo.ecn.purdue.edu/~biehl/ MultiSpec), p. 68, 1997

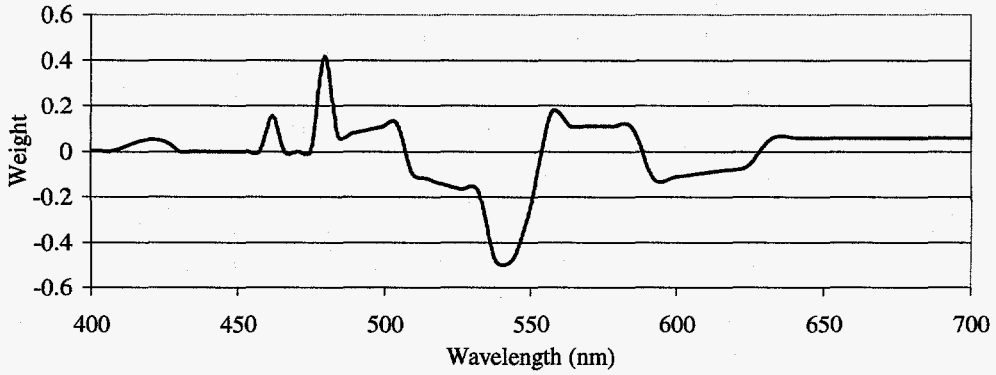


Figure 1: Example Spectral Basis Vector

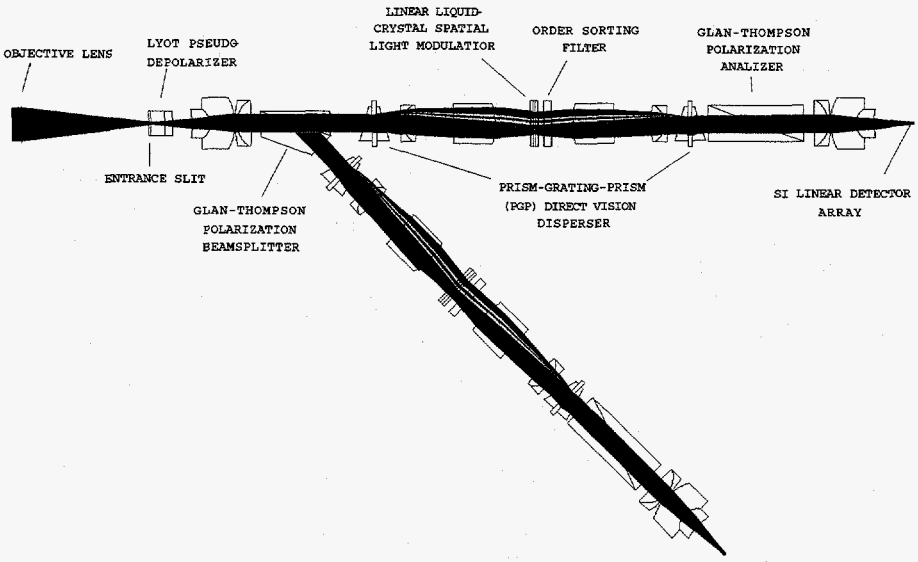


Figure 2: ISIS Field Test Sensor Schematic Layout

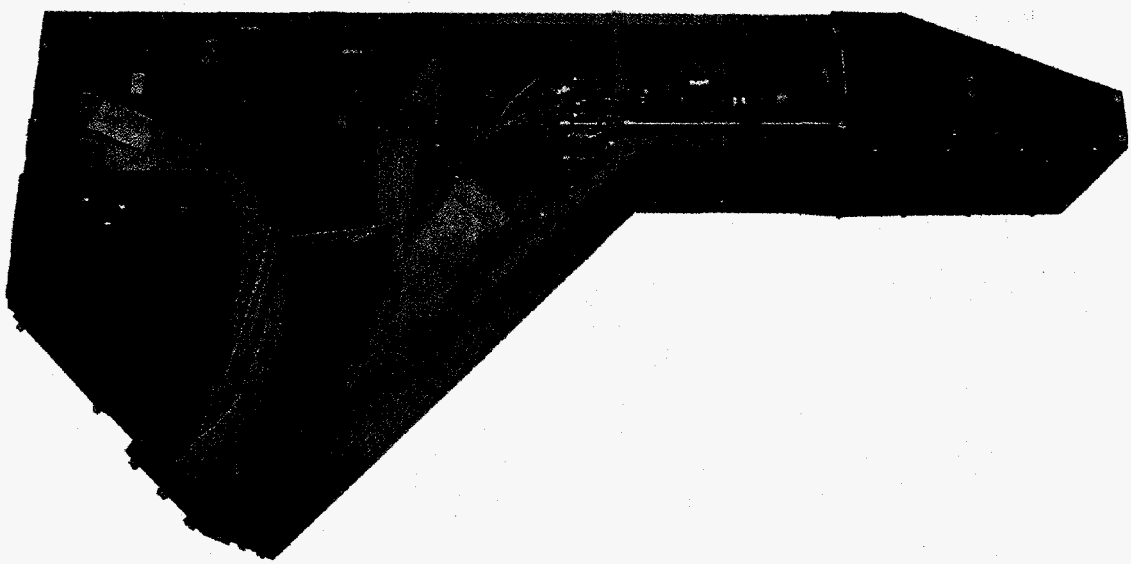
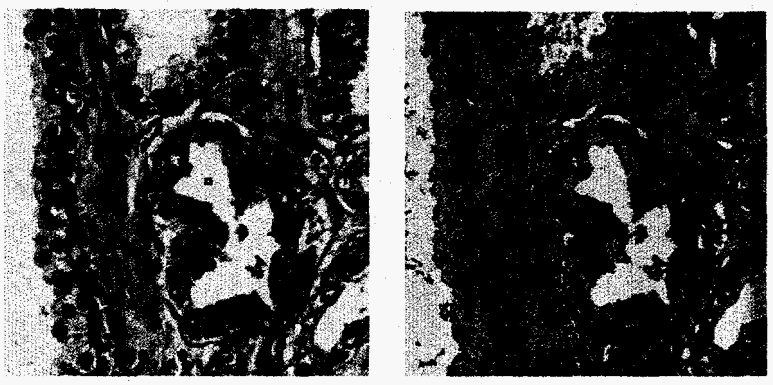
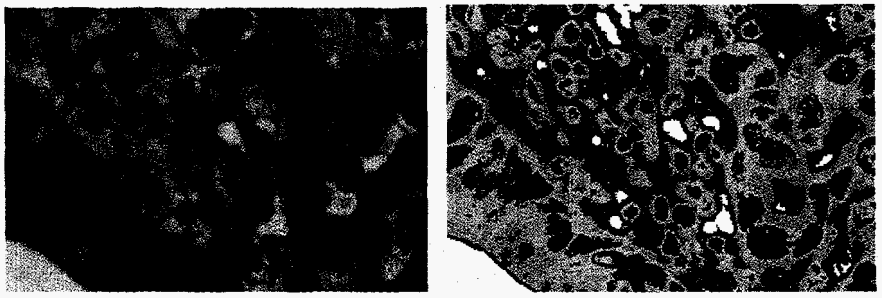


Figure 3: ISIS Field Test Sensor System



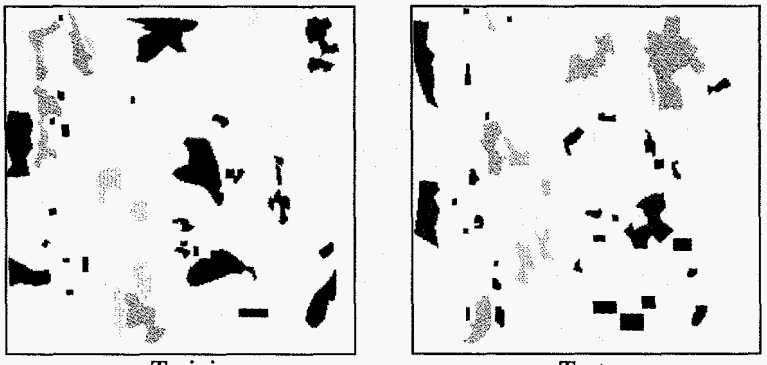
Blue-Myoepithelial Nuclei
Red-Malignant Epithelial Nuclei
Green-Other Epthelial Nuclei
Salmon-Connective Tissue
Yellow-Cytoplasm
White-Background

Figure 4: Prostate Cancer RGB Image and Reference Classification



Green-Benign Nuclei
Red-Malignant Nuclei
Tan-Cytoplasm 1
Lt Blue-Cytoplasm 2
Dk. Blue-Cytoplasm 3
Yellow-Cytoplasm 4
White-Background

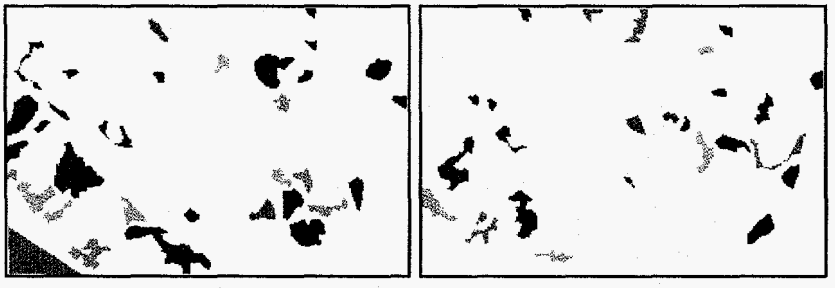
Figure 5: Colon Cancer RGB Image and Reference Classification



Training Test

Blue-Myoepithelial Nuclei
Red-Malignant Epithelial Nuclei
Green-Other Epthelial Nuclei
Salmon-Connective Tissue
Yellow-Cytoplasm
Grey-Background

Figure 6: Prostate Cancer Training and Test Regions



Training Test

Green-Benign Nuclei
Red-Malignant Nuclei
Tan-Cytoplasm 1
Lt Blue-Cytoplasm 2
Dk. Blue-Cytoplasm 3
Yellow-Cytoplasm 4
Grey-Background

Figure 7: Colon Cancer Training and Test Regions

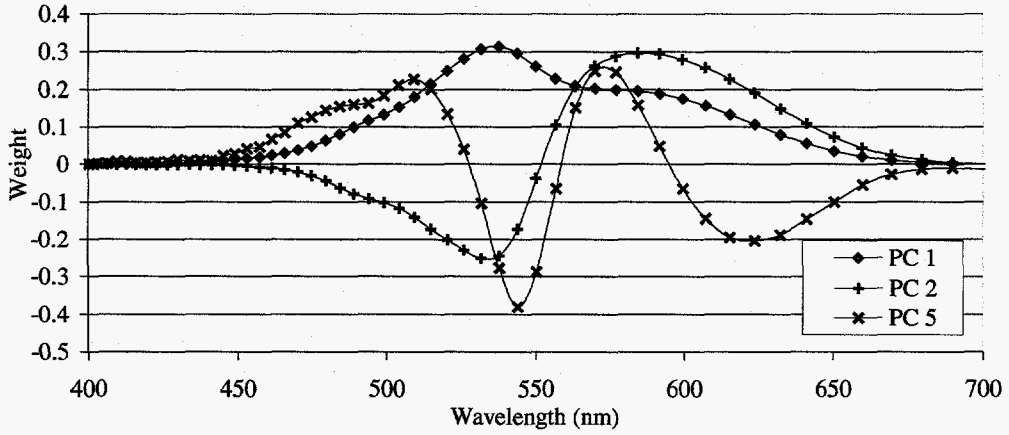


Figure 8: Optimal Three PC Basis Vectors for Prostate Cancer Example

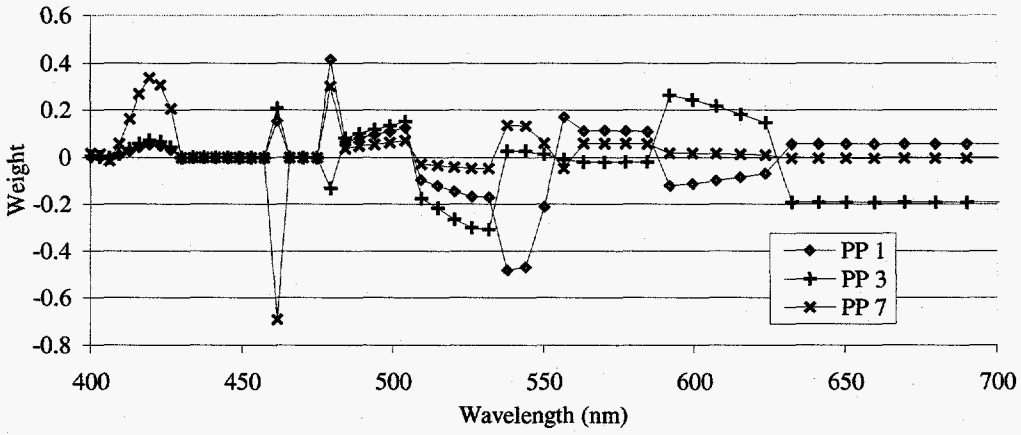


Figure 9: Optimal Three PP Basis Vectors for Prostate Cancer Example

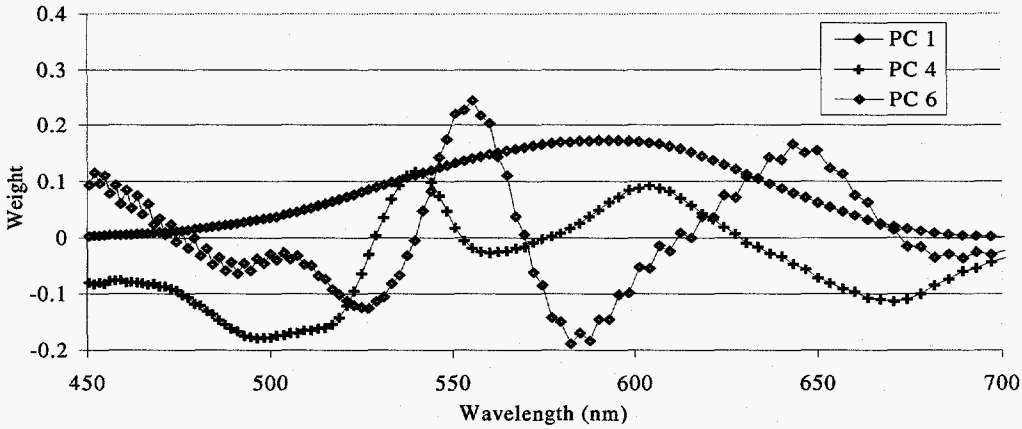


Figure 10: Optimal Three PC Basis Vectors for Colon Cancer Example

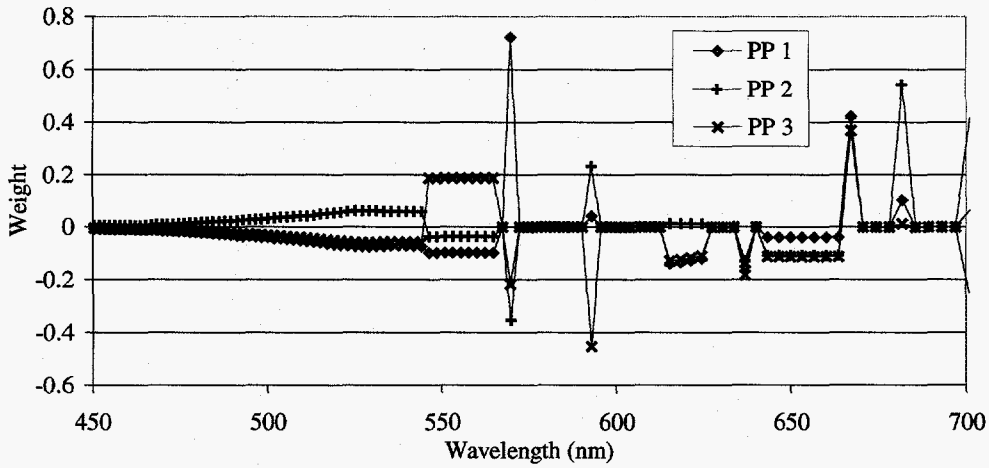


Figure 11: Optimal Three PP Basis Vectors for Colon Cancer Example

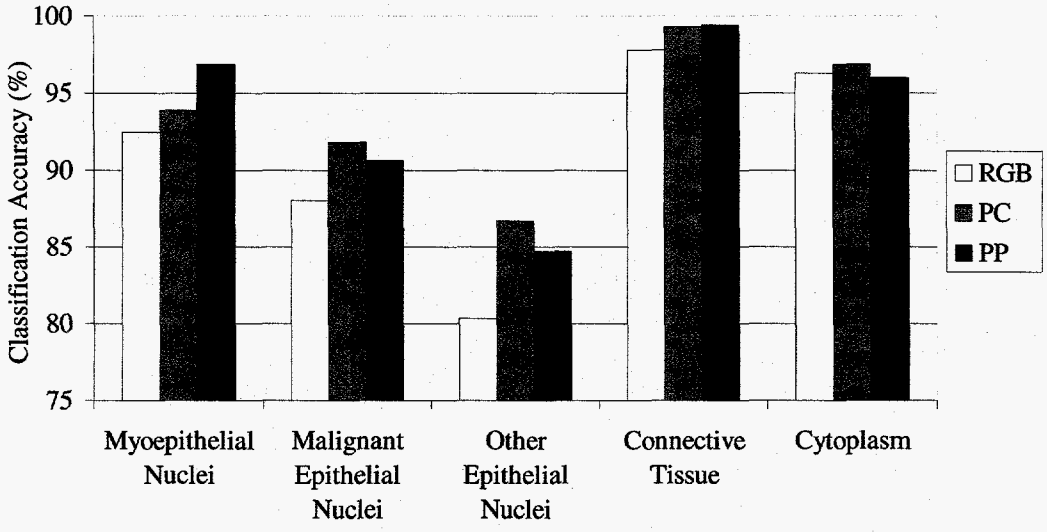


Figure 12: Classification Performance on Prostate Sample with Three Measurement Basis Sets

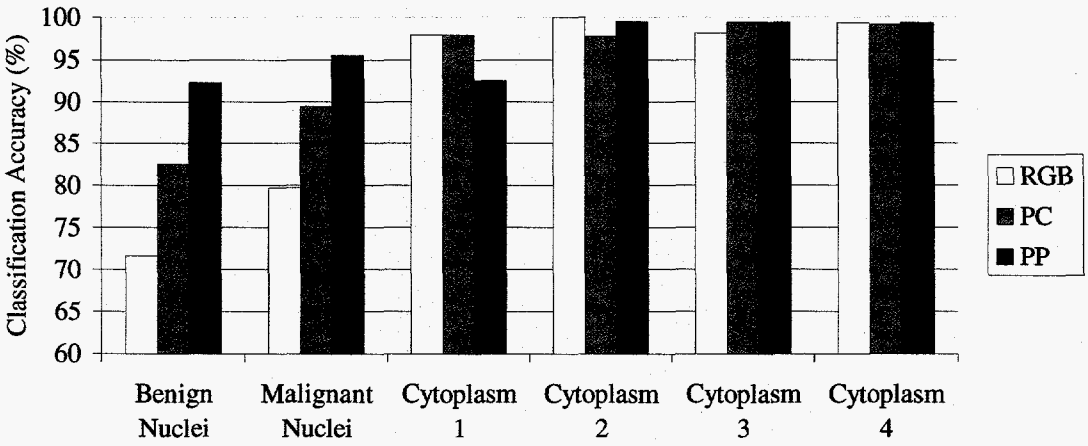


Figure 13: Optimal Basis Set Performance Improvement on Colon Sample

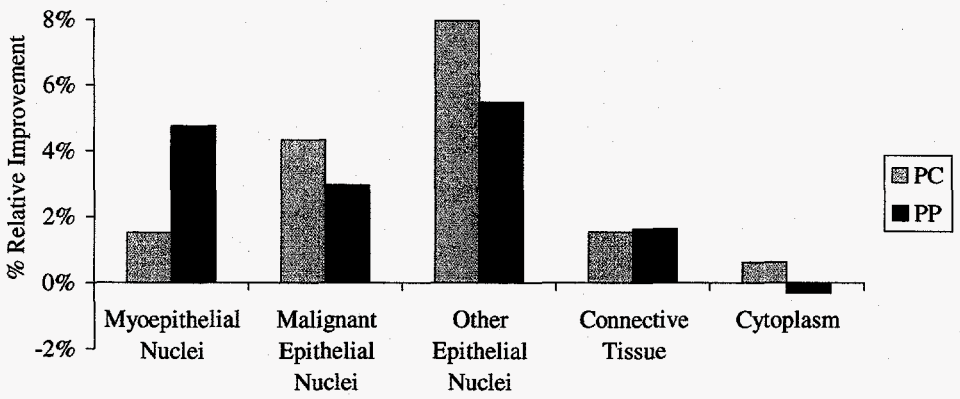


Figure 14: Relative Performance Improvement of a Three Element Optimal Basis Sets over RGB for the Prostate Cancer Example

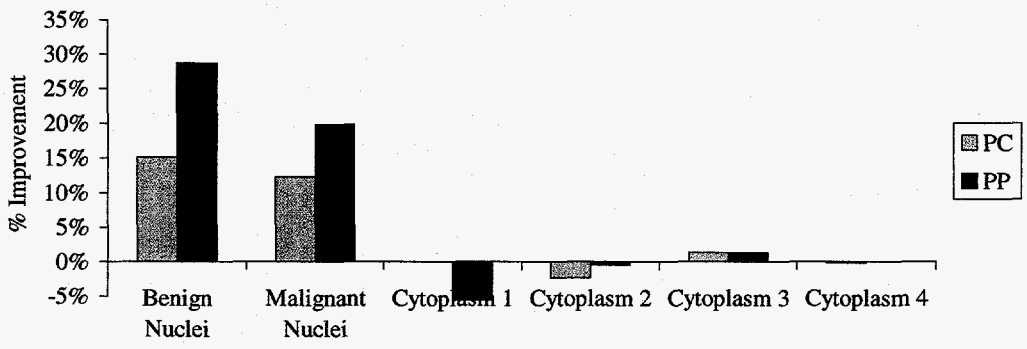


Figure 15: Relative Performance Improvement of a Three Element Optimal Basis Sets over RGB for the Colon Cancer Example

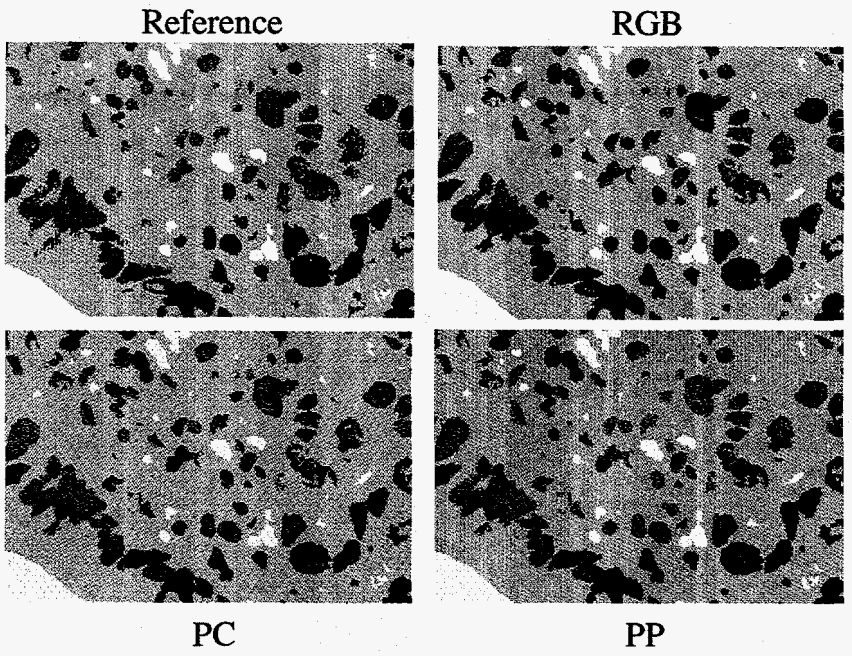


Figure 16: Classification Results for Different Basis Sets
(Yellow – Cytoplasm, Green – Normal Nuclei, Red – Malignant Nuclei, White - Background)

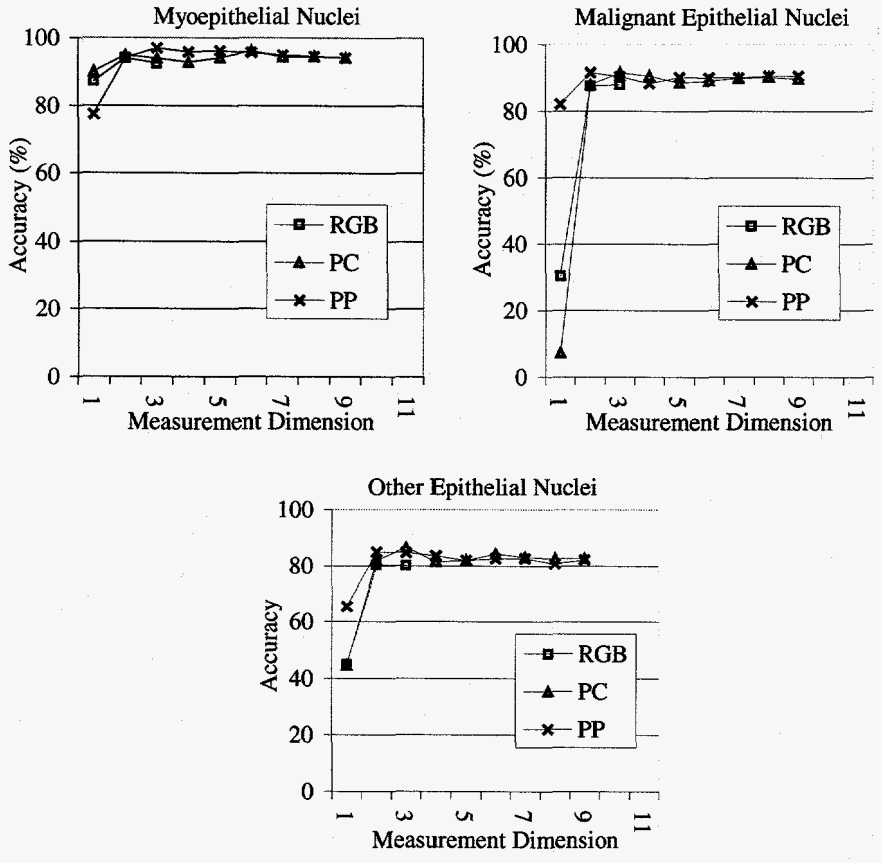


Figure 17: Classification Accuracy of Prostate Cancer Nuclei vs. Measurement Dimension

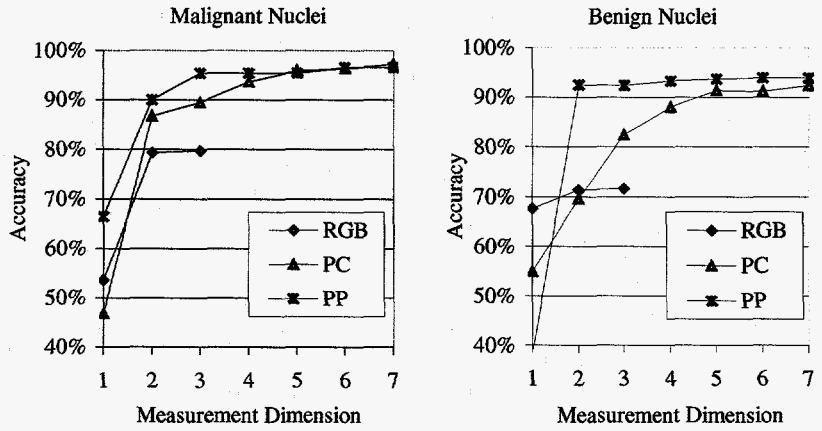


Figure 18: Classification Accuracy of Colon Cancer Nuclei vs. Measurement Dimension

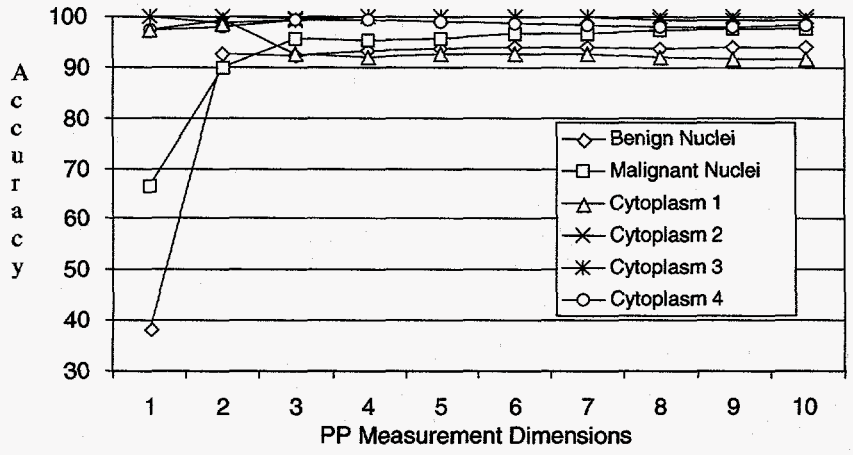


Figure 19: Colon Cancer Classification Accuracy vs. PP Measurement Dimensions