Electrical Engineering and Computer Science   College of Engineering and Computer Science

Summer 7-5-2012

MathWeb: A Concurrent Image Analysis Tool Suite for Multi-spectral Data Fusion

Tiranee Achalakul
Syracuse University, tachalak@syr.edu

Peter D. Haaland
Mobium Enterprises

Stephen Taylor
Syracuse University, Department of Electrical Engineering and Computer Science

Follow this and additional works at: https://surface.syr.edu/eecs

Recommended Citation

MathWeb^{TM}: A Concurrent Image Analysis Tool Suite for Multi-spectral Data Fusion

Tiranee Achalakul[a], Peter D. Haaland[b], Stephen Taylor[a]

[a]Department of Electrical Engineering and Computer Science, Syracuse University
[b]Mobium Enterprises, Inc. 518 West Linden, Louisville, CO 80027-3124

ABSTRACT

This paper describes a preliminary approach to the fusion of multi-spectral image data for the analysis of cervical cancer. The long-term goal of this research is to define spectral signatures and automatically detect cancer cell structures. The approach combines a multi-spectral microscope with an image analysis tool suite, MathWeb. The tool suite incorporates a concurrent Principal Component Transform (PCT) that is used to fuse the multi-spectral data. This paper describes the general approach and the concurrent PCT algorithm. The algorithm is evaluated from both the perspective of image quality and performance scalability.

Keywords: concurrent computing, image fusion, principal component transform, spectral signature

1. GENERAL APPROACH

This paper describes an approach to multi-spectral analysis of cervical cancer cells. The approach uses a multi-spectral microscope to capture pathology images at selected spectral wavelengths. The resulting multi-spectral image is directly manipulated using an image analysis tool called MathWeb. Image fusion in this tool can be accomplished using a wide variety of techniques that include pixel, feature, and decision level algorithms [3, 4]. At the pixel level, raw pixels can be fused using averaging, ray-casting, wavelet [8], Fourier[9] or Principal Component Transform (PCT) [1]. At the feature level, raw images can be transformed into a representation of objects, such as image segments, signal amplitude, shape, or orientation. At the decision level, images are processed individually and an identity declaration is performed to fuse the results. In this paper, we describe and evaluate a concurrent PCT algorithm. Image registration is not required since the input is already registered [13, 16].

The multi-spectral microscope. Initially, our goal is to detect, isolate, and highlight cell structures by developing spectral contrast; the wavelength dependence of absorption or reflection in different cell structures will then be used to generate color contrast. To achieve this goal, we have developed a multi-spectral microscope for collecting pathological imagery. This capability is shown in Figure 1.
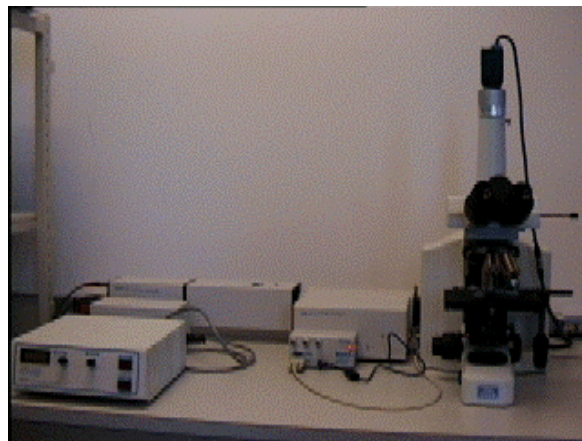

Figure 1: Multi-spectral Microscope

Light from a broadband source is passed through a monochromator to provide a computer controlled, spectrally tunable source. The selected band is then passed into a standard laboratory microscope through a fiber bundle. Images are acquired with a high-resolution black and white CCD camera. A typical multi-spectral image captured from this device consists of 38 gray scale frames, at 640x480 resolution, in the visible spectrum, from 410nm through 780nm at 10nm steps. This multi-spectral image covers the complete detection range of the microscope in steps that are sufficiently fine to capture the primary spectral information. The cell structures we seek to distinguish within the image are nuclei (small dark dots), cytoplasm (oval shape blobs), and cell walls (fine lines). Figure 2 shows three example frames from a typical data set. Unfortunately, the cell features of interests cannot be clearly distinguished in any single image.
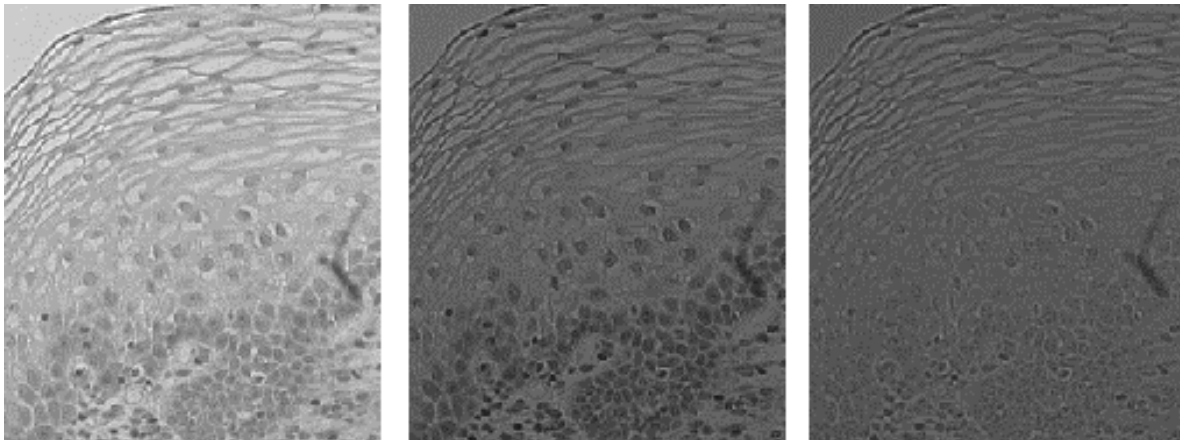

Figure 2: Sample frames at 440nm, 660nm, and 720nm.

MathWeb. MathWeb is an image analysis tool suite based on portable concurrent programming concepts [11, 12]. It operates on symmetric multi-processors (SMP), PC's, and workstations. The tool suite provides a concurrent PCT algorithm for data fusion and interfaces to control laboratory instruments, based on serial and IEEE 488 standards. To represent multi-spectral images, the tool suite generalizes sequential matrix algebra, familiar to users of Mathlab[2], IDL[5], and Mathematica[7], to concurrent tensor algebra. The following hierarchy explains the relevance of this concept to image analysis:

Tensor Dimension	Conventional Interpretation	Alternative Interpretation
0	Scalar	Real or Complex Number
1	Vector	List of Values
2	Matrix	Table of values or Image
3	Tensor	RGB-Image
N indexed by wave length	Tensor	Multi-and Hyper-spectral Image
N indexed by time	Tensor	Video Stream

Using the alternative interpretation, a multi-spectral image of cervical cancer cells taken from the microscope can be represented directly as a third order tensor. Figure 3 explains how such an image can be acquired and manipulated using MathWeb. The *wave* command sets the wavelength of the monochromator to a specified value. The *grab* command grabs an image frame from the CCD camera. Each grabbed image is subsequently inserted into a pre-constructed third order tensor, *Is*. Once all frames have been inserted, the PCT can be performed directly to produce a tensor of components, *Cs*. The final images, *Tc* and *Fc*, are created by mapping the first three principal components to either a true or false color image.

```
Is = create(38)
count = 0
for i = 410:10:780
    wave(i)
    Is[:, :, count] = grab()
    count = count + 1
end

Cs = PCT(Is)

Tc = truecolor(Cs[:, :, 0], Cs[:, :, 1], Cs[:, :, 2])
Fc = falsecolor(Cs[:, :, 0], Cs[:, :, 1], Cs[:, :, 2])
```
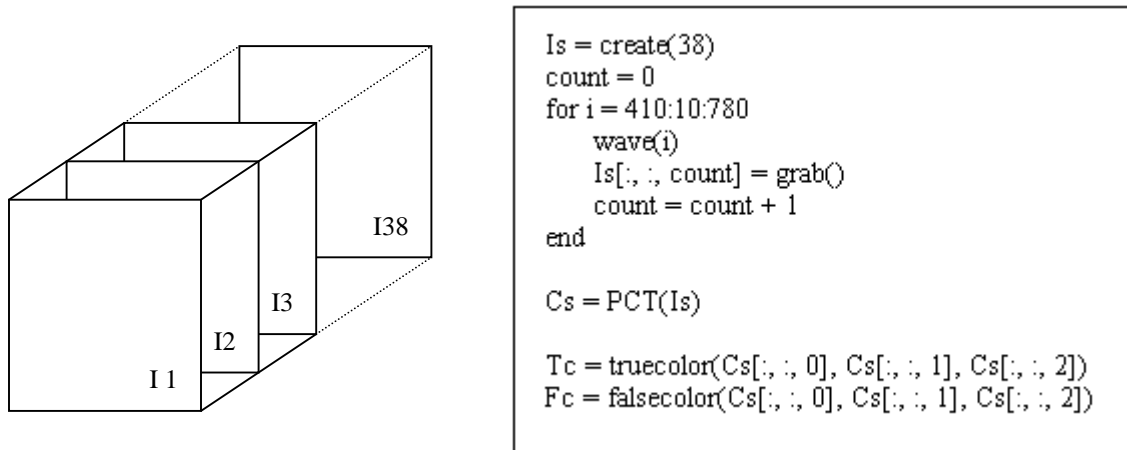
Figure 3: PCT Using MathWeb Script

2. CONCURRENT PRINCIPAL COMPONENT TRANSFORM

The Principal Component Transform [1, 6, 10, 14, 15] is a linear transformation that can be used to compress a multi-spectral image. The coordinate transformation permits identification of information that might not be apparent in any single image, or simple linear combination of images selected empirically. The original multi-spectral image is transformed to a collection of decorrelated images. Each image in the output carries spectral information that collectively can be used to reconstruct the original multi-spectral image. In the cell data set, cell walls and nuclei reflect or absorb differently in different spectra, and thus, the information will be pushed to different components in the resulting output. Since PCT generates a new image set whose bands are rank ordered by the magnitude of their variances (eigenvalues), most of the spectral contrast is pushed forward to the first few components where the first component has the largest variance. As a result, we can ignore many of the latter components when performing image analysis.

$$\vec{Is}_{ij} = [I1_{ij}, I2_{ij}, I3_{ij}, ..., In_{ij}] \qquad \vec{Cs}_{ij} = [C1_{ij}, C2_{ij}, C3_{ij}, ..., Cn_{ij}]$$

$$Cs = PCT(Is)$$

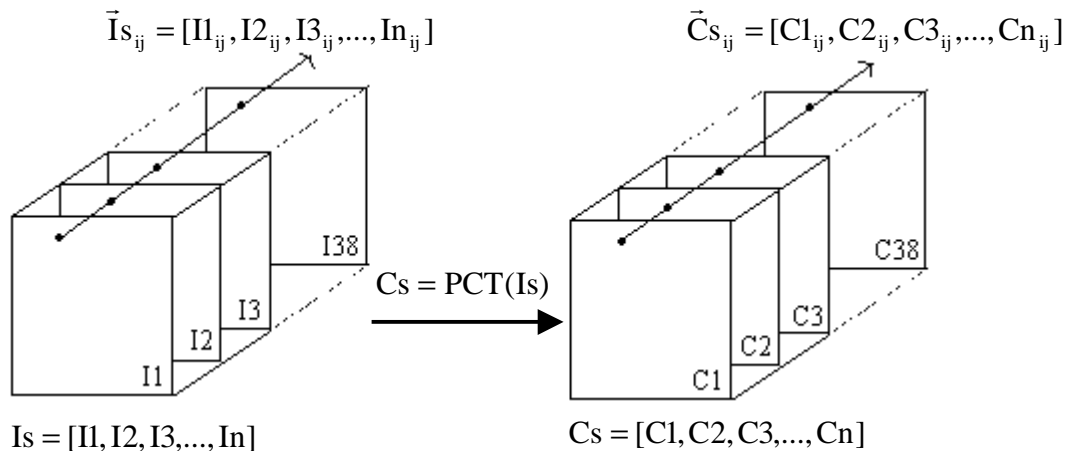$$Is = [I1, I2, I3, ..., In] \qquad Cs = [C1, C2, C3, ..., Cn]$$

Figure 4: PCT on Multi-spectral Images

Unfortunately, PCT is a highly compute intensive algorithm. Unlike Fourier, Walsh, or Hadamard transforms, the PCT transformation matrix is not separable, and thus, no fast algorithm exists [14]. MathWeb approaches the performance problem using multi-threading. The thread creation overheads are minimized by creating a number of threads directly proportional to the number of processors available in a particular machine, and not to the problem size. Figure 4 shows the

general interpretation of the PCT algorithm in which the multi-spectral image, *Is*, is transformed into the tensor, *Cs*, using the transformation:

$$Cs = A(Is - m)$$

where A is a transformation matrix and m is a mean vector. The concurrent algorithm is divided into two parts that calculate the transformation matrix A, and subsequently transform the data as follows:

1. **Mean vector**: Each component of the mean vector, m, is the average of the pixel values of an image in each spectral band. The n-band multi-spectral image produces a mean vector of n elements where each element can be computed independently as follows:

    for all i = 1 to n concurrently

    $$m[i] = \frac{1}{K} \sum_{k=1}^{K} I_k[i]$$

    where K = number of pixels in an image.

2. **Covariance sum**: The covariance matrix for n-band is defined as:

    $$C = \begin{bmatrix} S_1^2 & S_{12} & ... & S_{1n} \\ S_{21} & S_2^2 & ... & S_{2n} \\ . & . & & . \\ S_{n1} & S_{n2} & ... & S_n^2 \end{bmatrix}$$

    where $(S_i)^2$ is a variance of the ith spectral band, and S_{ij} is a covariance between band i and band j. To calculate the covariance matrix, pixels at a specific position in all spectra are related while neighboring pixels in the same image are not. Therefore, the pixels in a multi-spectral image are taken as a sequence from the top left to bottom right. The sequence is divided into P parts using integer division. Each part is allocated to a thread as follows:

    for all p = 1 to P concurrently
        $sum_p = 0$
        for all pixels (i, j) in p

    $$C_{ij} = Is_{ij} Is_{ij}^T - mm^T$$
    $$sum_p = sum_p + C_{ij}$$

    where P = number of parts and sum_p is the matrix sum of the covariance in each part, p.

3. **Covariance matrix**: The covariance matrix is the average of all the matrices calculated in step 2, and is calculated sequentially since its complexity is related only to the number of threads rather than the image size.

    $$C = \frac{1}{K} \sum_{p=1}^{P} sum_p$$

4. **Transformation matrix**: The eigenvectors of the covariance matrix are calculated and sorted according to their corresponding eigenvalues which provide a measure of their variances. As a result, the high spectral contents are forced into the front components. Since the degree of data dependency of the calculation is high, but its complexity is related to the number of spectral bands rather than the image size, this step is done sequentially.

5. **Transformation of the data**: each pixel vector, Is_{ij}, can be transformed independently. Therefore, once again, pixel vectors in the multi-spectral image are taken as a sequence and divided into p parts.

for all p = 1 to P concurrently
   for all pixels (i, j) in p

$$Cs_{ij} = A(Is_{ij} - m)$$

where P = number of parts.

3. EVALUATION

Figure 5 shows the variances of the resulting components from the PCT. Notice that, more than 88% of the information is pushed into the first component. There is almost no significant variance left after the third component. Figure 6 shows the first three resulting components. The first component has a high pixel values (white) in areas that represent cytoplasm, while the second has a high values in cell walls, and the third in nuclei.
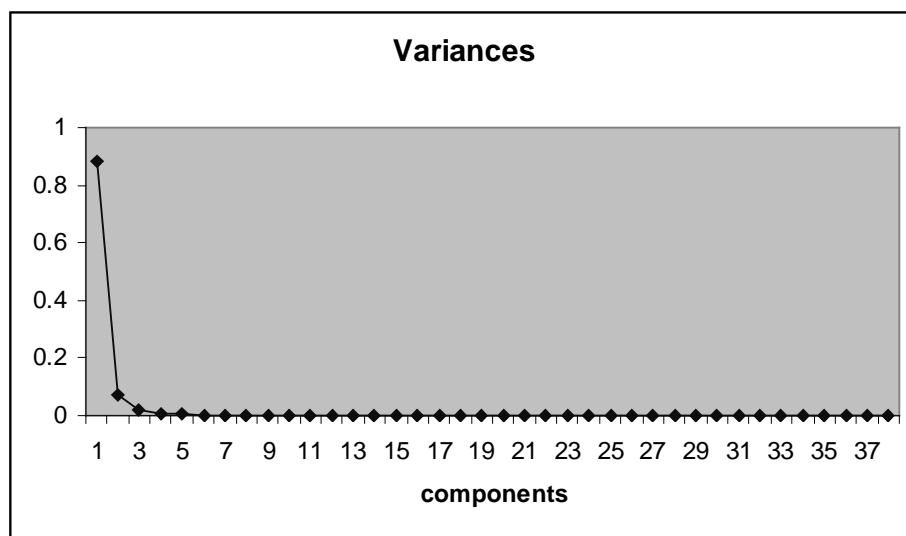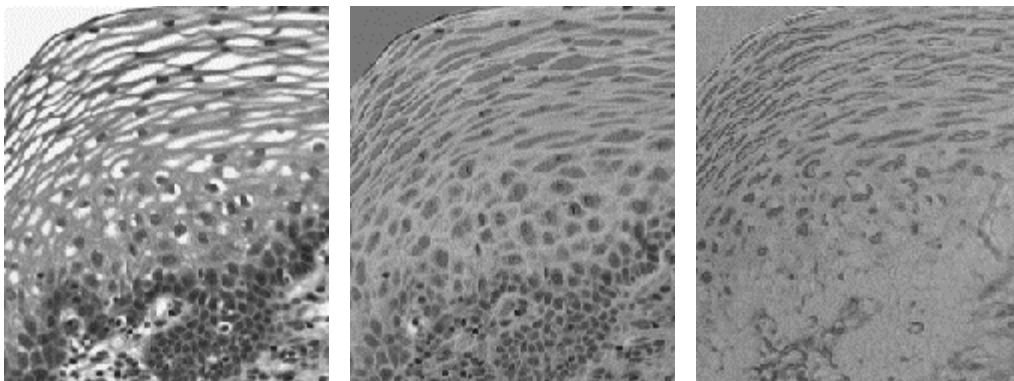

Figure 5: Variances Plot

Figure 6: The First, Second, and Third Principal Components

To display the extracted information in these bands in a perceptible manner, two different methods have been explored: false and true color mapping. The first method falsely colors an image by directly mapping the first component to red, the second

to green, and the third to blue. Thus, we expect the three cell features to be painted in different colors: the nuclei is in high blue, the cell walls in high green, and the cytoplasm in high red. This result is shown on the left of Figure 7.

The true color method maps the first three components into luminance and color opponent channels that produces an image that matches the information processing channels of the human visual system [18]. The first component is mapped to achromatic, the second to red-green, and the third to blue-yellow opponency. The mapping equation is based on a simple matrix multiplication shown below.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.6624 & 0.9271 & 0.1326 \\ 0.6619 & -0.4864 & -0.3017 \\ 0.6634 & 0.0289 & 1.0968 \end{bmatrix} X \begin{bmatrix} PC1 \\ PC2 \\ PC3 \end{bmatrix}$$

The true color mapping produces a high-resolution color image similar to that observed by looking through the microscope eyepiece. This is shown on the right of Figure 7.
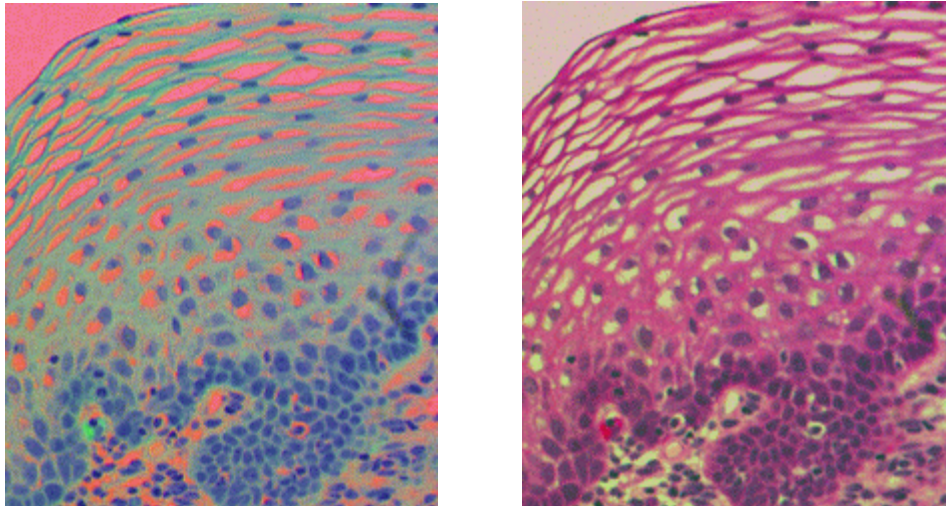

Figure 7: False and True Color Mapping

The falsely colored image achieves the goal of detecting, isolating, and highlighting cell structures. The resulting image can then be further analyzed to gather statistical data, such as number of nuclei per square millimeter, spatial distribution of nuclei, and ratio of nuclei to cytoplasm size, etc. These statistical properties can then be used to help pathologists determine whether the cell sample is benign or malignant [17]. The true color image is created from a series of high-resolution gray scale images in the visible spectrum and is of higher resolution than would be collected from an equivalent color camera.

4. PERFORMANCE

The concurrent PCT algorithm was used on the above data set using a varying number of processors on the Silicon Graphics PowerChallenge. Figure 8 shows the speed up gained as a function of the number of processors, and the ideal speed up. The graph demonstrates that the speed up gained is close to linear. However, the performance dropped approximately 20 % from ideal in some cases. This speed degradation was caused by the sequential code in step 3 and 4. The performance related with forming the covariance matrix is directly proportional to the number of processors, while the eigenvector calculation is related to the number of image frames used in the problem. This latter aspect dominates the sequential overheads. Although the eigenvector algorithm has a complexity of $O(n^3)$, at the typical problem size of 38 frames, the time used is not dominant. The algorithm was also tested on a larger data set, a hyper-spectral landscape images for camouflage detection, with 210 spectral bands. The performance drop was 15 % which suggested that using a larger number of spectral bands diminishes

rather than extenuates the impact of sequential code. This is because the performance of steps 1,2, and 5 are related to both the number of spectral bands and the image size.
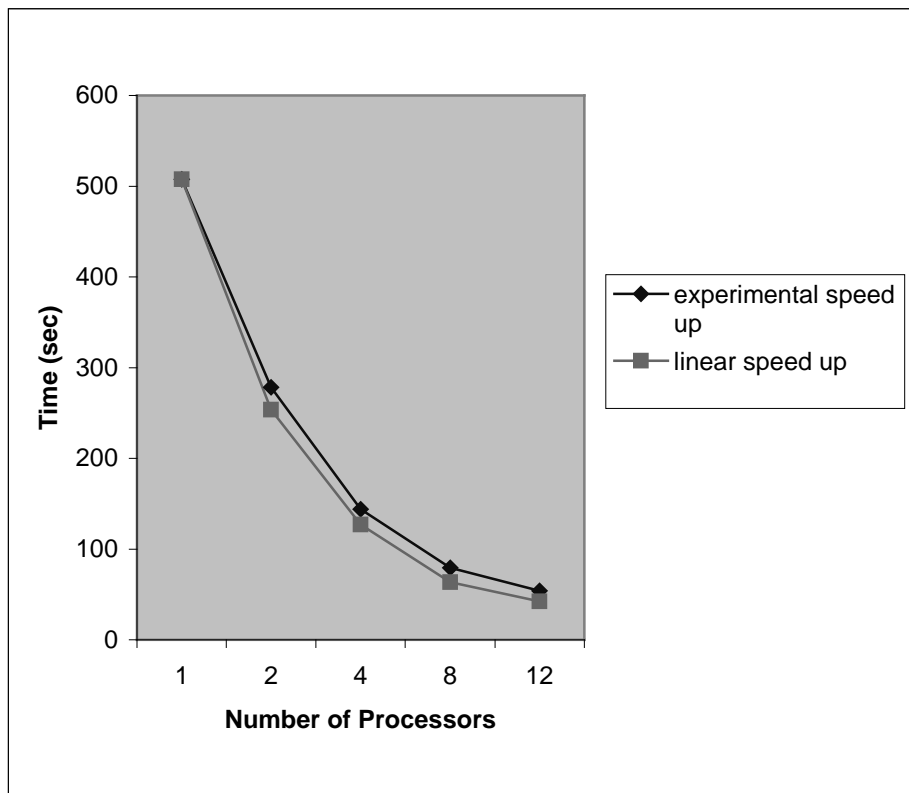

Figure 8:Performance Chart

Finally, while useful for scaling tests, the Silicon Graphics PowerChallenge is relatively old processor technology. A two processors Pentium II machine was also used. The overhead of sequential code on two processors was reduced to insignificant and the algorithm achieved ideal performance speed up within the bound of measurement, while running 3.6 times faster.

5. CONCLUSION

A preliminary approach to the fusion of multi-spectral images for the analysis of cervical cancer has been described. The three technologies used are a multi-spectral microscope, the MathWeb image analysis tool, and a concurrent PCT algorithm. The algorithm was evaluated from both the perspectives of resulting image quality and performance. A more complete experimental and analytical study for distributed machines is in progress. We aim to define an analytical model from which performance predictions can be made. In addition, we are currently extending the range of the microscope outside the visible spectrum.

6. ACKNOWLEDGEMENTS

7. REFERENCES

1. Bernhard Flury, *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, Inc., 1988.
2. Charles F. Van Loan, *Introduction to Scientific Computing : A Matrix-Vector Approach Using Matlab*, Prentice-Hall, 1996.
3. David L. Hall, "An Introduction to Multisensor Data Fusion," Proceedings of The IEEE, Vol. 85, No. 1, January 1997, pp. 6-23.
4. David L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Boston, MA:Artech House, 1992.
5. David W. Fanning, *IDL Programming Techniques*, Fanning Software Consulting, 1997.
6. Ernest L. Hall, *Computer Image Processing and Recognition*, Academic Press, 1979.
7. Hartmut F. W. Hoft, Margret H. Hoft, *Computing With Mathematica*, Academic Press, 1998.
8. H. Li, B. S. Manjunath, S. K. Mitra, "Multisensor Image Fusion Using the Wavelet Transform," Graphical Models and Image Processing, Vol. 57, No. 3, May 1995, pp. 235-245.
9. Ioana M. Martin, Dan C. Marinescu, "Concurrent Computation and Data Visualization for Spherical-Virus Structure Determination," IEEE Computational Science & Engineering, pp. 40-52. October-December 1998.
10. J. Edward Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, Inc., 1991.
11. J. Watts, S. Taylor, M. Rieffel, and Palmer, "The Concurrent Graph: Basic Technology for Irregular Problems, " IEEE Parallel and Distributed Technology, 4(2):15-25, Summer 1996.
12. J. Watts, S. Taylor, "A Practical Approach to Dynamic Load Balancing," IEEE Transactions on Parallel and Distributed System, Vol. 9, pp. 235-248, 1998.
13. L. Brown, "A Survay of Image Registration Techniques," ACM Comput. Surv. 1992, pp. 325-376.
14. Rafael C. Gonzalez, Paul Wintz, *Digital Image Processing*, Addison-Wesley Publishing Company, Inc., 1977.
15. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, Inc., 1993.
16. R. Luo and M. Kay, "Data fusion and Sensor Integration: State of the Art in 1990s," *Data Fusion in Robotics and Machine Intelligence*, pp. 7-136, Academic Press, San Diego, 1992.
17. Stanley L. Robbins, Ramzi S. Cotran, Vinay Kumar, *Pathologic Basis of Disease*, W.B. Saunders Company, 1994.
18. Wandell and Marimont, "Linear Models of Surface and Illuminant Spectra," Journal of the Optical Society of America. Vol. 9, No. 11, pp. 1905, 1992.