International Journal of Information Technology and Knowledge Management
January-June 2012, Volume 5, No. 1, pp. 59-64

ADVANCED IMAGE ANALYSIS BASED SYSTEM FOR AUTOMATIC DETECTION AND CLASSIFICATION OF MALARIAL PARASITE IN BLOOD IMAGES

Neetu Ahirwar[1], Sapnojit. Pattnaik[1], and Bibhudendra Acharya[2]

ABSTRACT: This paper investigates the possibility of rapid and accurate automated diagnosis of red blood cell disorders and describes a method to detect and classify malarial parasites in blood sample images acquired from light microscopes. As malaria is an infectious disease which is mainly diagnosed by visual microscopical evaluation of Giemsa stained blood smears. As it poses a serious global health problem, automation of the evaluation process is of high importance. The image classification system is designed to positively identify malaria parasite in thin blood smears. Morphological and novel threshold selection techniques are used to identify erythrocytes (red blood cell) and possible parasites present on microscopic slides. Image features based on colour, texture and the geometry of the cells and parasites are generated, as well as features that make use of a priori knowledge of the classification problem and mimic features used by human technicians. Classifier using back propagation feed forward neural network distinguishes between parasite infected and non-infected blood images.

Keywords: Malaria, Parasite, Neural Network, Erythrocytes.

1. INTRODUCTION

Malaria is a serious global health disease caused by a blood parasite named plasmodium spp. The World Health Organization estimates 300-500 million malaria cases and more than 1 million deaths per year[1]. It is caused by any of the four different species of Plasmodium parasite, vivax, ovale, malariae and falciparum. Disease is transmitted via the bite of an infected female of the Anopheles mosquito. The most widely used technique for determining the development stage of the malaria disease is visual microscopical evaluation of Giemsa stained blood smears[3]. However this is a routine and time consuming task and requires a trained operator. Besides a recent study on the field shows the agreement rates among the clinical experts for the diagnosis are surprisingly low.

In peripheral blood sample, visual detection and recognition of Plasmodium spp is possible and effect via a chemical process called (Giemsa) staining[4]. The staining process slightly colorizes the red blood cells (RBCs) but highlights Plasmodium spp parasites, white blood cells (WBC), and platelets or artifacts. Detection of Plasmodium spp requires detection of stained objects. However, to prevent false diagnosis, the stained objects have to be analyzed further to determine if they are parasite or not.

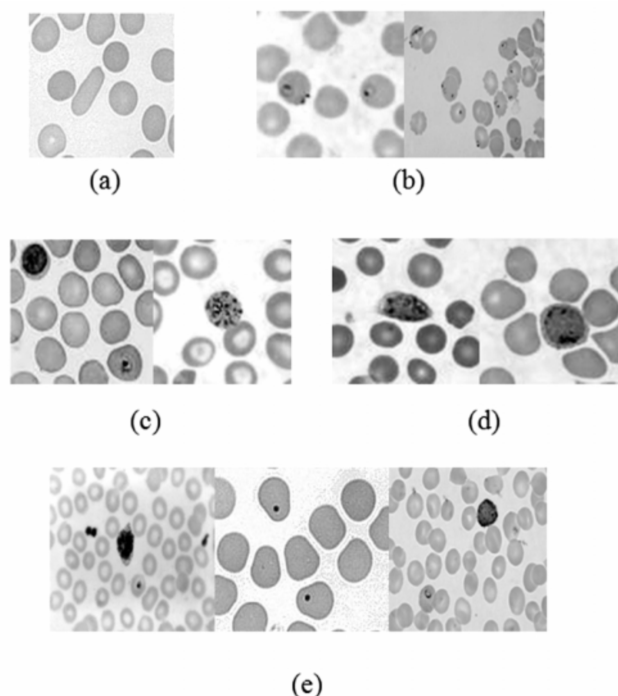

Figure 1: (a) Normal RBC, (b) Plasmodium Falciparam, (c) P. Malarie , (d) P.Ovale, (e) P. Vivex

In this paper, we present an automatic method for detection and classification of malaria parasites present in thin blood smears. Our approach can be divided into several well-defined stages, presented in Fig. 2.

[1] Department of Electrical Engineering, N.I.T, Raipur, India, E-mail: nahirwar13@gmail.com, E-mail: swapnajit.pattnaik@gmail.com,
[2] Department of Electronics & Telecommunication Engineering, N.I.T, Raipur, India, E-mail: bacharya.etc@nitrr.ac.in

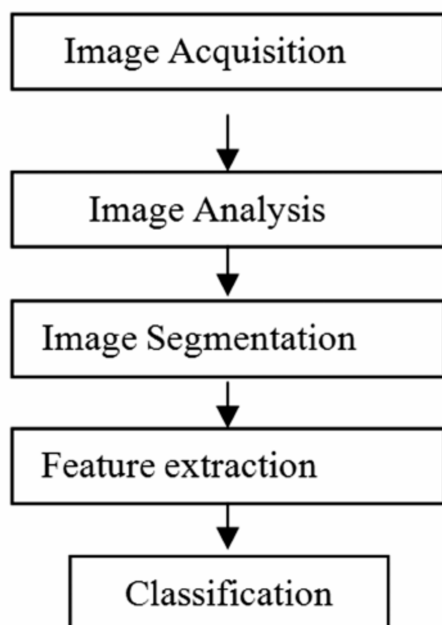

Figure 2: General Structure of the Method

Artificial Neural Network (ANN) has been employed together with image processing techniques to automate the diagnosis of malaria using the morphological features of erythrocytes in blood images. Artificial Neural Networks (ANN's) have proven to be a promising paradigm for Intelligent Systems. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems.

ANN's have the advantage of learning by example and the ability to generalize from their training data to other data. They are fault tolerant in the sense; they can produce correct outputs from noisy and incomplete data. ANN's are relatively inexpensive to build and train [23].

After acquisition, the image is first analyzed in order to reduce noise present in an image. The SUSAN filter is often used to remove noise from the images[16]. After that, the image is segmented in order to isolate the interesting parts and remove noise and undesired components. Next the feature extraction process is applied, to extract the useful information from the segmented objects, and finally the classification between parasite infected and non infected images can be operated according to the characteristics extracted by the previous stage. Each of these steps will be presented in the next sections.

2. PRESENT WORK

The various steps used in the present MALARIA parasite detection and classification system are discussed below.

2.1. Image Acquisition and Database Collection

Images of Giemsa stained blood smears were selected from public health image library. Oil immersion views (10 × 1000) of Giemsa stained blood films were captured using a binocular microscope mounted with a digital camera. Captured images were 460 pixels × 307 pixels bitmap images. The database consists of 90 images

2.2. Pre-processing

The purpose of the pre-processing stage is to remove unwanted effects such as noise from the image, and transform or adjust the image as necessary for further processing. I have implemented a new approach to low-level image processing-SUSAN (Smallest Univalue Segment assimilating Nucleus) Principle [16], which performs Edge and Corner Detection and Structure Preserving Noise Reduction. Canny edge detector, which has become one of the most widely used edge finding algorithms, is found to be ten times slower than this SUSAN approach. The results are stable for Canny but the edge connectivity at junction is poor and corners are rounded. The fact that SUSAN edge and corner enhancement uses no image derivative, explains why the performance in the presence of noise is good. The integrating effect of the principal together with its non-linear response gives strong noise rejection.

2.4. Image Segmentation

Techniques have been proposed earlier that make use of thresholding or morphology to segment an image. In this section we have presented a technique that takes benefit of morphological operation and thresholding at appropriate positions in the whole process to maximize the productivity of the algorithm. In order to use morphological methods for image segmentation, the shape and size of the objects in the image must be known. The most commonly used morphological procedure for estimating size distribution of image components is the Granulometry. The size and eccentricity of the erythrocytes are also required for the calculation of some feature values (as these can be indicative of infection). The shape of the objects (circular erythrocytes) is known a priori, but the image must be analyzed to determine the size distribution of objects in the image and to find the average eccentricity of erythrocytes present.

The next stage of the process identifies and segments potential parasites and erythrocytes from the image background. To extract the infected erythrocytes, it is first necessary to identify them from the combination of parasites and erythrocytes in the image, and then segment them from the background.

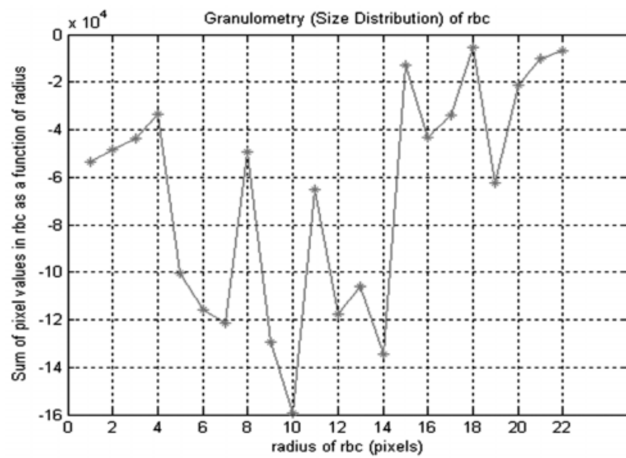

Figure 3: A Pattern Spectrum Showing the Size Distribution of Objects in a Sample can be Calculated using Granulometry [11]

This algorithm relies primarily on thresholding. The key to successfully segmenting an image using thresholding is threshold selection. The histogram of the complemented, green component of the sample image (Fig. 4) is a bimodal distribution typical of all the images considered.

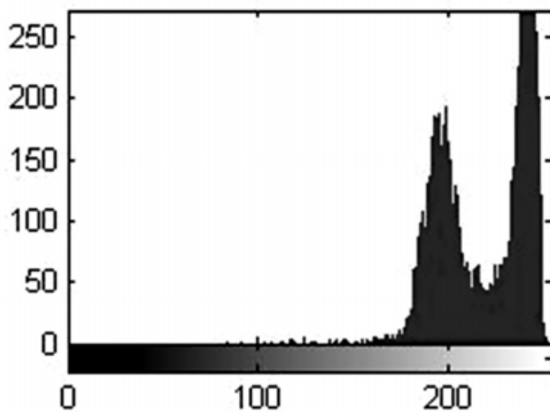

Figure 4: Bi-Modal Histogram of Plasmodium Falciparum

The threshold level is automatically selected using a modified version of the method that maximizes the separability of the resultant classes of the grey-level histogram [17]. The principle mode is due to the gray scale intensities of the image background, and the second mode is due to those of the erythrocytes in the image. Two threshold levels need to be determined from the histogram: one for erythrocytes, and one for parasites. The first threshold is selected to separate the erythrocytes from the background of the image. This essentially means separating the two modes of the image histogram. The first threshold is selected to separate the erythrocytes from the background of the image. This essentially means separating the two modes of the image histogram.

The next step is to select the second threshold to find parasites present in the image. A global threshold level, taking the threshold as the first local minimum in the histogram after the mode due to erythrocytes, is not sensitive enough. This is a common problem experienced with global threshold selection, caused by inconsistent intensities in the image. The solution is to find local threshold levels. The erythrocytes, having already been identified, provide excellent image regions in which to find these, especially since valid parasites are only found inside erythrocytes. The threshold is then found by taking the first minimum after the principal mode of the histogram incorporating only the erythrocytes. While this method has greater sensitivity, it is at the expense of a reduced specificity. There are also cases, particularly with P. ovale, where the global threshold is able to detect parasites that are missed by the local thresholds. This is due to colorization of the infected cells, which shifts the principle mode of the local histograms of the affected cells. For this reason, both local and global thresholds are used, and the union of the two binary images is used as the parasite marker image.

Invalid objects in the marker image (objects detected with the global threshold that lie outside any erythrocyte) are removed by taking the intersection of the parasite marker image with the binary mask of erythrocytes. The erythrocyte mask is dilated first, to ensure that blister'forms of the parasites, that appear to bulge out of the edge of the cells, are not removed. Other artifacts in the blood containing nucleic acid, particularly white blood cell nuclei, are also detected by this thresholding. They are removed by excluding all objects greater than an empirically determined size (chosen to exclude all objects greater than the largest trophozoite that one would expect to find.)

The infected cells are identified by morphologically reconstructing the erythrocyte mask with the valid parasite marker. Binary reconstruction simply involves extracting the connected components of an image (the mask) that are marked by another image (the marker), where cells are clustered together, if an infected cell forms part of the group, then the entire aggregation is reconstructed. A morphological opening filter, using a disk-shaped SE with radius equal to the mean erythrocyte radius less the standard deviation, is applied to the grayscale, morphologically filtered green component of the image to remove any objects smaller than an erythrocyte. The morphological gradient-the difference between a dilation and erosion of the image-is then calculated using a diamond-shaped SE with unity length. The segmentation method is applied to each object in the reconstructed binary image of erythrocytes individually. Those objects that do not exceed the area of a circle with radius equal to the mean erythrocyte radius plus the standard deviation are regarded as being single cells, and are unmodified.

Unlike the method in Di Ruberto et al. [20], where the morphological gradients are used to generate marker images for the watershed algorithm, the objects deemed to be overlapping erythrocytes are segmented as follows. First, the intersection of the morphological gradient image and the dilated cell cluster is taken. This image is then transformed to a binary image by thresholding any value greater than zero. A series of morphological operations, namely a closing operation, thinning, and spur removal are then applied to generate a contour of the segmented erythrocytes. The contours are filled, and the segmented mask is again reconstructed with the valid parasite marker image to result in a segmented mask of infected cells. The erythrocytes that have been identified as possibly infected are then extracted from the image and passed to the next stage of the algorithm for feature generation. The binary mask of the erythrocyte, as well as a binary mask (obtained by local threshold selection based on the image histogram as detailed above) of parasite-like objects present in the cell, is also passed to the next stage.

2.5. Feature Generation and Classification

1. Feature Generation: Two sets of features are used for development. The first set will be based on image characteristics that have been used previously in biological cell classifiers, which include geometric features (shape and size), color attributes and grey-level textures.

   It will be advantageous to apply expert, a priori knowledge to a classification problem. This will be done with the second set of features, where measures of parasite and infected erythrocyte morphology that are commonly used by technicians for manual microscopic diagnosis are used. It's desirable to focus on these features, because it is already known that they are able to differentiate between species of malaria.

2. Feature Classification: The final classification of an erythrocyte as infected with malaria or not is accomplished by back propagation feed forward (BFF) neural network. Feed Forward Back propagation neural network architecture developed by Paul Werbos was chosen as it was a simple and one of the most commonly used ANN's [24]. In this type of ANN a new input could lead to correct output provided that the input being presented was similar to the inputs used in training the network. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs [24]. The features selected for the classifier are those that describe the colour and texture of the possible parasites. The training goal is to minimize squared errors, and training is stopped when the error of a validation set increased. This is done to avoid overtraining.

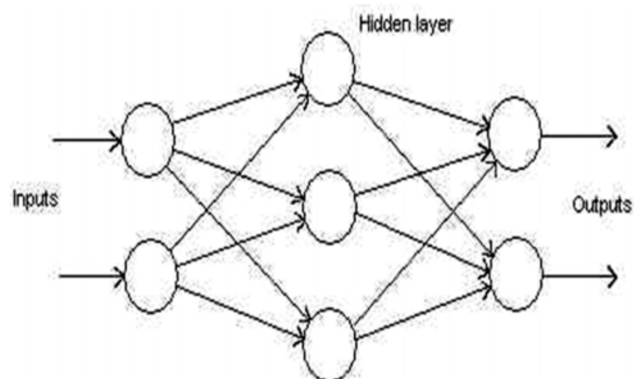

Figure 5: Simple Feed Forward Back Propagation Network

3. RESULTS

The performance and accuracy of the algorithm are analyzed using two measures: sensitivity, the ability of the algorithm to detect a parasite present; and positive predictive value (PPV), the success of the algorithm at excluding non-infected cells. These values are expressed in terms of true positives (TP), false positives (FP) and false negatives (FN):

$$\text{Sersitivity} = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

The algorithm has been tested on various malaria parasites. Table 1 gives the results of first order features of different parasite affected blood cells:

Table 1

S.No.	Image	Mean	Variance	Skewness	Kurtosis	Energy
1.	Simple RBC	6.8972	5.5766	–1.8530	5.0930	0.2983
2.	P.falciparam	7.6138	1.1776	–4.8723	29.5746	0.5935
3.	P.Malarie	7.0064	2.2837	–2.4107	10.0099	0.3026
4.	P.Ovale	7.2485	2.3814	–2.1044	9.3526	0.2518
5.	P.Vivex	7.1227	2.5648	–3.0788	12.0602	0.4277

The training data set for the back propagation feed forward NN had total 77 images contain both malaria parasites and nonparasite images. At the end of training the network achieved its performance Function When tested with a selected set of different images other than that used for training the NN was able to categorize it accordingly.

For the training of the back propagation ANN "Gradient Descent" algorithm with a performance function (error) goal of 10-5 and a learning rate of 0.2 was used in a batch processing mode.

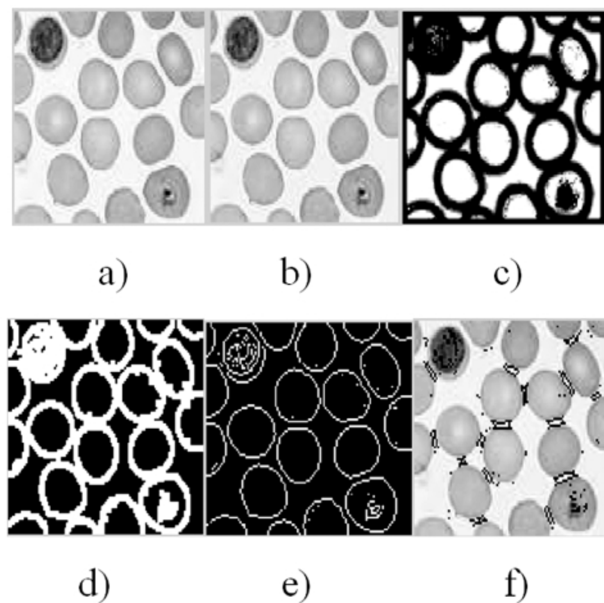

Figure 6: (a) Original Image, (b) Gray Scale Image, (c) Susan Output, (d) Dilated Gradient Mask, (e) Binary Gradient Mask, (f) Final Detected Cell

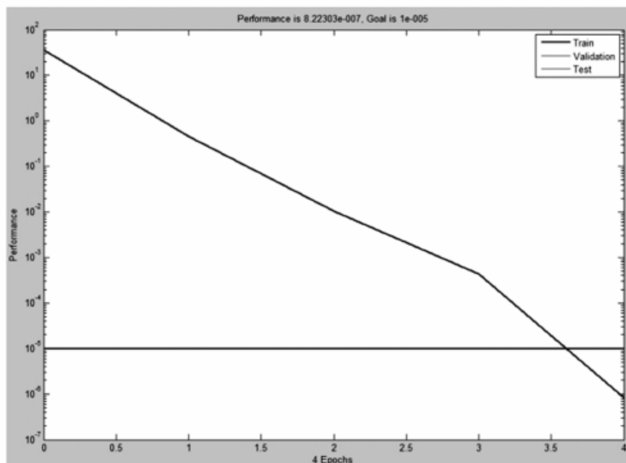

Figure 7: Performance of Feed Forward NN with Back Propagation Algorithm

4. CONCLUSION

There are different methods for malaria parasite detection. The proposed automated parasite detection algorithm has many advantages compared to other diagnostic techniques. It avoids the problems associated with rapid methods, such as being species-specific and having high per-test costs, while retaining many of the traditional advantages of microscopy, viz. species differentiation. Among the tested algorithms, SUSAN edge detection technique 'gave good localization of edges but formed a thick border making cell separation difficult. Otsu's algorithm' gave accurate separation of RBCs where as local and global thresholding segmented the parasites. Granulometry provides the size distribution of object in image. The BFF neural network which has been trained with the back propagation algorithm produced the highest performance.

REFERENCES

[1] World Health Organization. What is malaria? Facts Sheet no 94. http://www.who.int/mediacentre/factsheets/fs094/en/

[2] Foster S, Phillips M, Economics and its Contribution to the Fight Against Malaria. Ann Trop MedParasitol 92: 391-398, 1998.

[3] Makler MT, Palmer CJ, Alger AL, A Review of Practical Techniques for the Diagnosis of Malaria. Ann Trop Med Parasitol 92(4): 419-433, 1998.

[4] Bloland PB (2001) Drug resistance in Malaria, WHO/CDS/CSR/DRS/ 2001.4. World Health Organization, Switzerland, 2001.

[5] Gilles H.M. "The Differential Diagnosis of Malaria". Malaria. Principles and Practice of Malariology (Wernsdorfer W.H., McGregor I Eds), 769-779, 1998.

[6] F. Castelli, G.Carosi, "Diagnosis of Malaria, Chapter 9, Institute of Infectious and Tropical Diseases", University of Brescia (Italy).

[7] Baird J.K., Purnomo, Jones T.R. "Diagnosis of Malaria in the Field by Fluorescence Microscopy of QBC ® Capillary Tubes". Transactions of the Royal Society of Tropical Medicine and Hygiene; 86: 3-5, 1992.

[8] Anthony Moody, Rapid Diagnostic Tests for Malaria Parasites, Clinical Microbiology Reviews,0893-8512/02/ $04.00_0 DOI: 10.1128/CMR.15.1.66-78.2002, pp. 66-78, Jan. 2002.

[9] Brown A.E., Kain K.C., Pipithkul J., Webster H.K. Demonstration by the Polymerase Chain Reaction of Mixed Plasmodium Falciparum and P. Vivax Infections Undetected by conventional microscopy. Transactions of the Royal Society of Tropical Medicine and Hygiene; 86: 609-612, 1992.

[10] Jean-Philippe Thiran, Benoit Macq, "Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissues". IEEE Transaction on Biomedical Engineering, 43, No. 10, October 1996.

[11] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra. "Automatic Thresholding of Infected Blood Images using Granulometry and Regional Extrema". In ICPR, pp. 3445-3448, 2000.

[12] Silvia Halim, Timo R. Bretschneider, Yikun Li, Estimating Malaria Parasitaemia from Blood Smear Images. 1-4244-0342-1/06/$20.00 ©IEEE, ICARCV 2006.

[13] Selena W.S. Sio, Malaria Count: "An Image Analysis-Based Program for the Accurate Determination of Parasitaemia", Laboratory of Molecular and Cellular Parasitology, Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore. May 2006.

[14] F. Boray Tek, Andrew G. Dempster and Izzet Kale, "Malaria Parasite Detection in Peripheral Blood Images", Applied DSP & VLSI Research Group, London, UK, Dec 2006.

[15] Rafeal C. Gonzalez, Richard E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall, 2006.

[16] S. M. Smith, J. M. Bardy, SUSAN-A New Approach to Low Level Image Processing", International Journal of Computer Vision, 23, and Issue 1 pp. 45 -78, may 1997.

[17] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms". IEEE Transactions on Systems, Man and Cybernetics, 9(1): 62.66, 1979

[18] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, "A New Method for Gray-Level Picture Thresholding using the Entropy of the Histogram". Graphical Models and Image Processing, 29: 273.285, 1985.

[19] T.W. Ridler and S. Calvard, Picture Thresholding using an Iterative Selection Method". IEEE Transactions on Systems, Man and Cybernetics, SMC-8: 630.632, 1978.

[20] Di Ruberto C, Dempster A, Khan S, Jarra B, "Analysis of Infected Blood Cell Images using Morphological Operators". Image Vis Comput 20(2):133-146, 2002.

[21] M.K. Hu, Visual Pattern Recognition by Moment Invariants", IRE Trans. Information Theory, 8, pp. 179-187, 1962.

[22] Mui JK, Fu K-S, "Automated Classification of Nucleated Blood Cells using a Binary Tree Classifier". IEEE Trans Pattern Anal Machine Intell 2(5): pp. 429-443, 1980

[23] Donna L Hudson, Maurice E Cohen; "Neural Networks & Artificial Intelligence for Biomedical Engineering", IEEE Press; 200. ISBN: 81-203-1842-0

[24] User Guide, Neural Network Toolbox for use with MATLAB [Monograph Available on CD Rom]. Howard Demuth, Mark Beale. The Math Works, Inc. Version 3.0; 1998.