

A New Preprocessing Approach for Cell Recognition

Xi Long, *Student Member, IEEE*, W. Louis Cleveland, and Y. Lawrence Yao

Abstract—In this paper, we describe a novel strategy for combining fisher’s linear discriminant (FLD) preprocessing with a feedforward neural network to classify cultured cells in bright field images. This technique was applied to various experimental scenarios utilizing different imaging environments, and the results were compared with those for the traditional principal component analysis (PCA) preprocessing. Our FLD preprocessing was shown to be more effective than PCA due in large part to the fact that FLD maximizes the ratio of between-class to within-class scatter. The new cell recognition algorithm with FLD preprocessing improves accuracy while the speed is suitable for practical applications.

Index Terms—Cell recognition, fisher’s linear discriminant, neural networks, principal component analysis.

I. INTRODUCTION

RECENT progress in the development of methods for molecular genetic analysis (e.g., RT-PCR, microarrays) has brought sensitivities to the level where single cells can be analyzed [1]. To carry out single-cell-level assays on a significant numbers of cells, high-throughput robotic systems are needed. These systems require identification of cultured cells (often in bright field images) for micromanipulation and subsequent molecular analysis. Identification of cultured cells in bright field images is a difficult task. The difficulty arises from multiple factors such as the variability of cell size and morphology, the presence of “trash,” as well as variations in microscope parameters. Traditionally, identification of cells is done by experienced human observers. However, the use of human observers represents a severe impediment to the development of high-throughput robotic systems. Therefore, there is a major need for algorithms that permit automatic recognition of cells in bright field images.

With regard to automatic cell recognition, the method of combining “pixel patch” decomposition and artificial neural network (ANN) classification has recently begun to be explored and is considered advantageous over traditional image processing techniques because it can largely eliminate end-user programming and can accommodate variations in cell type and other conditions with training [2]–[5]. This is especially important for biologists, who may have minimal image processing and programming skills. Powerful statistical data processing

techniques have frequently been used in preprocessing to generate abstract representations that are better suited for subsequent neural network analysis. For example, Nattkemper *et al.* applied principal component analysis (PCA) in evaluation of fluorescence microscopy images of cells [3]–[5]. However, most existing applications are successful in dealing with images in which nuclei, stained with a fluorescent probe, have a characteristic color and are relatively uniform in size and shape. Bright field images of unstained cells represent a more difficult challenge which requires a more sophisticated technique.

In this study, a novel cell recognition approach based on feedforward neural networks has been developed. We exploited the advantages of fisher’s linear discriminant (FLD) over PCA as a preprocessing step for a feedforward neural network and adapted a novel strategy to make possible the use of FLD preprocessing. It was found that FLD in combination with feedforward neural networks was more effective than PCA and gave performance that was adequate for use in a practical robotic system for single-cell-level molecular analysis.

II. AUTOMATIC RECOGNITION OF CELLS IN IMAGES

A. Recognition Process

We used a “pixel patch” strategy similar to that described in [3]–[5]. The process involves two stages: preprocessing and classification. The major task of preprocessing is to derive a representation of cells which makes subsequent classification computationally effective and insensitive to environmental changes by providing the classifier only with information essential for recognition. In the classification stage, a neural network is trained to determine if a pixel patch contains a centered cell body. This is done with pixel patches represented by feature vectors derived in preprocessing.

As illustrated in Fig. 1, cells in a digitized microscope image are detected in the following steps: First, for each pixel p in the microscopic image, a pixel patch which consists of its $m \times m$ neighborhood centered at that pixel is extracted and mapped to a confidence value $C[p] \in [-1, 1]$ by the classifier, where the size m can be adjusted to accommodate cell size ($m = 25$ was used in our experiments). After all the pixels are processed, a new image, called a confidence map, is created. Pixels in the confidence map are the confidence values of their corresponding pixel patch in the original microscope image and form “mountains” with large peaks that represent cell positions. The cell positions can then be easily found by identifying local maxima in “mountains.” To increase speed (this procedure is optional), patches clearly indicating background, i.e., patches with average pixel intensities below a user-defined value (often determined by a simple histogram), are not analyzed.

Manuscript received February 18, 2004; revised November 5, 2004 and December 20, 2004. This work was supported by the National Institutes of Health under Grant CA89841.

X. Long and Y. L. Yao are with the Mechanical Engineering Department, Columbia University, New York, NY 10027 USA (e-mail: xl2002@columbia.edu; yly1@columbia.edu).

W. L. Cleveland is with the Department of Medicine, St. Luke’s Roosevelt Hospital Center and Columbia University, New York, NY 10019 USA (e-mail: wlc1@columbia.edu).

Digital Object Identifier 10.1109/TITB.2005.847502

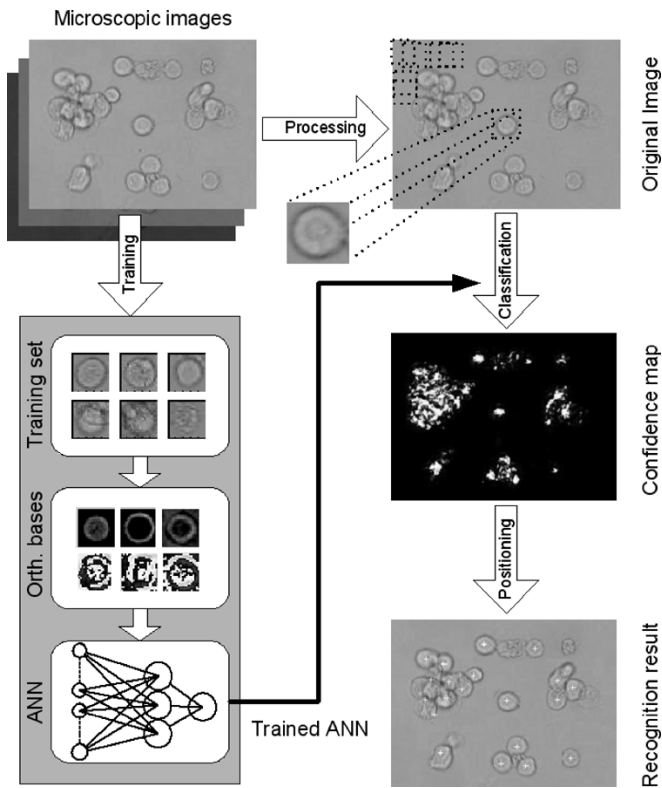


Fig. 1. Illustration of the recognition process.

B. A Novel FLD Preprocessing Strategy

Fig. 2(a) shows a typical learning set used for deriving the representation of cells. This set was manually selected from the digitized microscope images. The learning set Ω is composed of two subsets ($\Omega = \Omega^{\text{pos}} + \Omega^{\text{neg}}$). Ω^{pos} contains patches of centered cells and is labeled “Cell.” All patches in Ω^{pos} belong to a single class. Ω^{neg} is labeled “Non-cell.”

Each pixel patch in the learning set is treated as a point(vector) in an $m \times m$ dimensional image space. Considering the complexity and vastness of this image space, statistical data processing techniques, such as PCA and FLD, are employed to reduce it to a subspace with much lower dimensionality.

PCA has been widely used as preprocessing technique in combination with ANNs [3]–[5]. However, it is known that PCA preprocessing only considers the total variation of the whole learning set and therefore maximizes both the between-class variation, which is useful for classification, and the within-class variation, which ideally should be minimized. An alternative technique is FLD, which considers not only between-class variation but also within-class variation, and maximizes the ratio of between-class variation to within-class variation [7].

It should be noted that when PCA is used in the preprocessing step to reduce the dimension of the initial feature space, there is no restriction on the dimension of the resulting subspace. Whereas, when FLD is used, one must reduce the dimension of the resulting subspace to no more than $c - 1$, where c is the number of classes [7]. In our case, where only one type of cell is present, there are only two classes: “Cell” and “Non-cell.” Therefore, if FLD preprocessing is directly used, the resulting

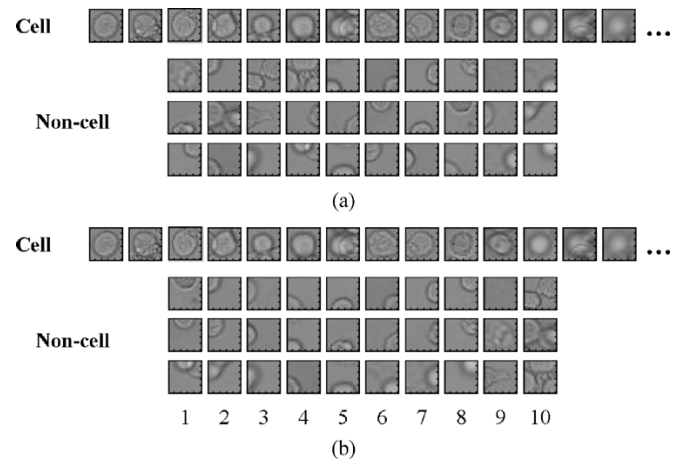


Fig. 2. Sample cell patches in the learning set. All patches are sized 25×25 pixels. (a) Original learning set. (b) A novel strategy that makes possible the use of FLD: factoring the “Non-cell” class into subclasses.

subspace can only be one-dimensional, which is inadequate for a complex pattern recognition problem like cell recognition.

We have developed a novel strategy to enable the use of FLD. To our knowledge, this strategy has never been used previously. The key idea of this strategy is to treat the cell recognition as a multiclass problem in the preprocessing stage. It can be done by dividing the “Non-cell” class into multiple subclasses. This is possible because the “Non-cell” class contains a diverse set of cell fragments and noncell objects. Factoring this class into n subclasses that are reasonably homogeneous makes the total number of classes equal to $n + 1$ (“Non-cell” classes + “Cell” class). The dimension of the subspace produced by FLD preprocessing can now be as high as $(n + 1) - 1 = n$. The above principle is illustrated by the example shown in Fig. 2(b), where the “Non-cell” class is divided into ten subclasses. Classes 1–8 contain a specific fraction of a cell. Pixel patches in class 9 are almost blank (i.e., background) and class 10 includes pixel patches with multiple fragments of different cells. This “factoring” of the “Non-cell” class into subclasses makes the use of FLD a practical possibility, since the total class number is now 11 and the dimension of the feature space after FLD preprocessing can therefore be as high as 10. An additional benefit of our strategy is that it reduces variation within individual “Non-cell” subclasses. This reduction of within-class variation is in addition to that which derives from the inherent ability of FLD to reduce within-class variation.

C. Classification

The classifier we used in this paper is a feedforward ANN. It has been widely used and proven to be a powerful tool for classification tasks. The ANN uses the input/output (I/O) mapping learned from a set of training samples to generalize to data not “seen” before. The training of an ANN involves gradually modifying the synaptic weights according to the back-propagated output error for each sample until the desired average responses are obtained on the entire set.

It is known that the capacity for generalization greatly depends on the structure of the neural network [8]. Generally, more neurons in the hidden layers give the system more capacity to

partition the data. However, if the network has too many neurons, it will learn insignificant aspects of the training set and lose its generalization ability, a phenomenon termed overfitting. A rule of thumb is to obtain a network with the fewest possible neurons in the hidden layer. We employed an empirical method [9] to optimize the neuron numbers in the hidden layers. The training results of our ANNs are shown in Section III-B.

III. RESULTS AND DISCUSSION

A. Image Acquisition

Our experiment has used images of cultured cells as well as images of polymeric microspheres as models of cells. The microspheres were 25- μm -diameter, red Fluorescent Polymer Microspheres from Duke Scientific (Cat. no. 36-5). The cells used were K562 chronic myelogenous leukemic cells (ATCC; Cat. No.CCL-243) grown at 37.0 $^{\circ}\text{C}$ in BM+1/2 TE1+TE2+10% fetal calf serum (FCS) [6]. Cells and microspheres in culture medium were dispensed into polystyrene 96-well microplates. Images were taken with an Olympus Model-CK inverted microscope using a 20 \times planachromat objective.

B. ANN Optimization and Training

In order to train and optimize the neural classifier, a set Φ of 1700 I/O pairs ($\Phi = \{(I_i, O_i)\}, i = 1, 2, \dots, 1700$) was created by projecting the learning set Ω to linear subspaces using both PCA and FLD. Accordingly, the set was also composed of two subsets $\Phi = \Phi^{\text{pos}} + \Phi^{\text{neg}}$. The positive subset $\Phi^{\text{pos}} = \{(I_i^{\text{pos}}, 1)\}$ consisted of feature vectors I_i^{pos} computed from the pixel patches in Ω^{pos} , together with the target output classification value $O_i^{\text{pos}} = 1$. The other subset $\Phi^{\text{neg}} = \{(I_i^{\text{neg}}, -1)\}$ consisted of feature vectors I_i^{neg} computed from pixel patches in Ω^{neg} and the target output value $O_i^{\text{neg}} = -1$ of the classifier. This set was further split into a training set of 1400 samples and a validation set of 300 samples. The training set was exclusively used to modify the weights and the validation set was exclusively used to estimate the generalization ability.

Besides the input layer, three layers were used in our experiment: two hidden layers, and one output layer. The hyperbolic tangent sigmoid function was used as the transfer function throughout the network. The optimal number of neurons in the two hidden layers was estimated by independently changing the number of hidden neurons in each layer, and evaluating the generalization properties of the ANN on the validation set at each step as described in [9]. To avoid entrapment in a local error minimum, every training session was repeated five times and the best weights were used for each number of hidden neurons.

Fig. 3 illustrates the generalization properties of the ANN for different numbers of neurons in the first layer, while keeping the size of the second hidden layer constant at five neurons. The mean squared error was plotted versus the number of neurons. The error rate improved as the number of hidden neurons was increased, but leveled out at around 40 neurons when preprocessed by PCA and 37 neurons by FLD. The experiment was repeated with the number of neurons in the second layer changed from 1 to 10 and similar but worse results were obtained (not shown). Based on above results, we chose 40 neurons for PCA preprocessing and 37 for FLD preprocessing in the first hidden

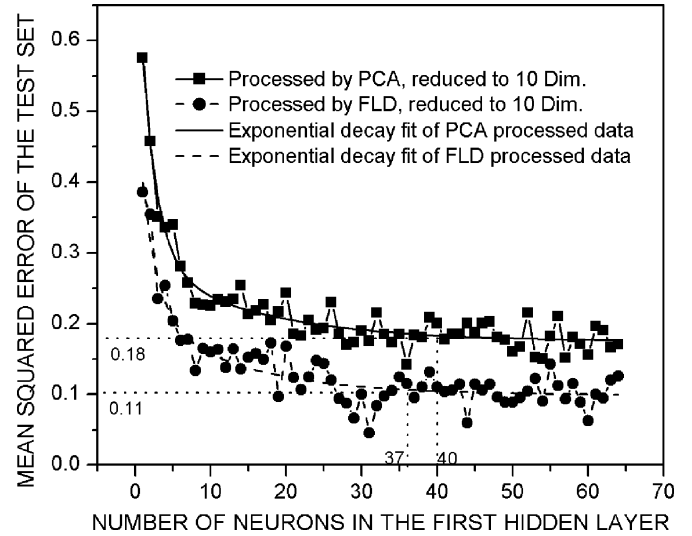


Fig. 3. Effect of neuron number in the first hidden layer on the generalization properties of the ANN. The size of the second hidden layer was kept constant at five neurons.

layer and five neurons in the second hidden layer for our subsequent studies. Similar procedures were also used during the microsphere experiments (data not shown).

C. Microsphere Experiments

We first used polymeric microspheres as models of cells. The microspheres are very uniform in size, shape, and color and are stable over time. These properties facilitate experimental reproducibility and make it possible to create ideal scenes in which critical factors can be individually isolated and well controlled. Both PCA and FLD preprocessing were systematically studied under various conditions of focus, illumination, object size, and image noise. In all microsphere experiments, recognition was performed as described in Section II. For FLD preprocessing, the dimensionality was reduced to 10. For PCA preprocessing, results are shown when both 10 and 20 principal components were used since it has been suggested that more principle components can improve performance [7].

1) *Focus Variation*: In order to quantify the effects of focus variation, we created four image groups at different focal planes relative to the microsphere equatorial plane, with all other conditions unchanged: a) Focused: the focal plane is at the equator of the microsphere (i.e., 12.5 μm above the supporting surface); b) 12.5 μm : the focal plane is at the supporting surface; c) 25 μm : the focal plane is 25 μm below the equator; and d) 37.5 μm : the focal plane is 37.5 μm below the equator. Typical sample images are shown in Fig. 4. Two experimental schemes were performed on these images. In Scheme 1, each method was trained on the first group and then tested on all groups. In Scheme 2, each method was trained on the first and third group and then tested again on all groups: the test on the second group was an interpolation test and on the fourth group was an extrapolation test (results shown in Fig. 5).

2) *Illumination Variation*: To evaluate the effects of illumination variation, images were taken under five light intensity levels of the microscope: a) *Intensity level 3*: representing extremely weak illumination; b) *Intensity level 4*: representing

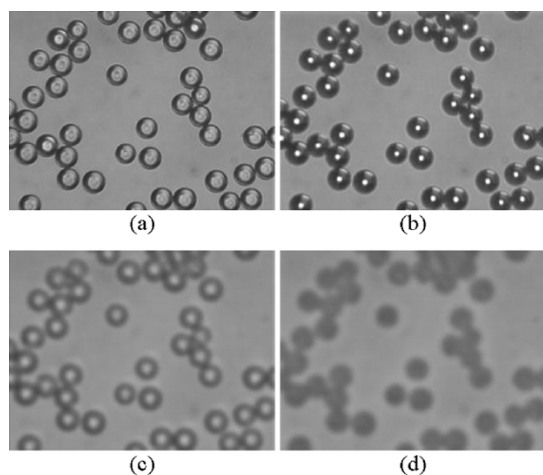


Fig. 4. Typical sample images for focus variation experiment. (a) Focused: the focal plane is at the equator of the microsphere. (b) $12.5 \mu\text{m}$: the focal plane is at the supporting surface. (c) $25 \mu\text{m}$: the focal plane is $25 \mu\text{m}$ below the equator. (d) $37.5 \mu\text{m}$: the focal plane is $37.5 \mu\text{m}$ below the equator.

weak illumination; c) *Intensity level 5*: representing normal illumination; d) *Intensity level 6*: representing strong illumination; and e) *Intensity level 7*: representing extremely strong illumination (image not shown). Two experimental schemes were performed using these images. To create the situation of small within-class variation, ANNs based on both PCA and FLD were trained with images only in Intensity level 3 and then tested with all levels in Scheme 1. In Scheme 2, within-class variation was purposely introduced by training the neural network with Intensity levels 4, 5, and 6 together and then tested again with all levels.

3) *Size (Scale) Variation*: In the size variation experiment, computer generated images of microspheres with 0%, 5%, 10%, 15%, and 20% variations in size were used (image not shown). Again, two schemes were used to examine the effect of size variation on both PCA and FLD methods. In Scheme 1, ANNs were trained with only microspheres having 0% size variation and tested to all sizes. In Scheme 2, they were trained using images with both 0% and 15% variation. The patch size used in both schemes was fixed to a value that was big enough to contain the biggest microspheres.

4) *Noise Variation*: The noise used in noise variation experiments was zero-mean Gaussian noise with different standard deviations. An image set with five groups of images, each have different noise levels was created by adding computer generated noise to original images. The original images (standard deviation equals zero) belonged to the first group. Groups 2, 3, 4, and 5 contained images in which the standard deviations equaled 15, 30, 45, and 60, respectively (image not shown). The two experimental schemes were: first, both PCA and FLD were applied to only Group 1 and then tested on all groups. Second, the training set was expanded to include both Groups 1 and 4.

In Fig. 5, we present only the results for focus variations. For illumination, size, and noise variations, similar results were obtained.

Some interesting points were revealed in the experiments. First, both PCA and FLD preprocessing performed well if presented with images in the test set which were selected from

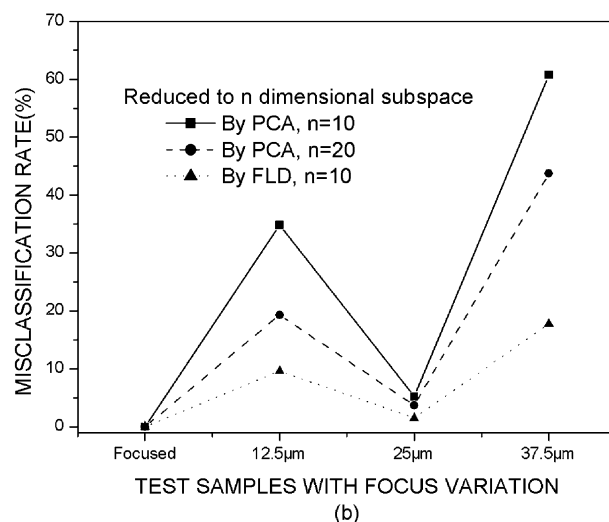
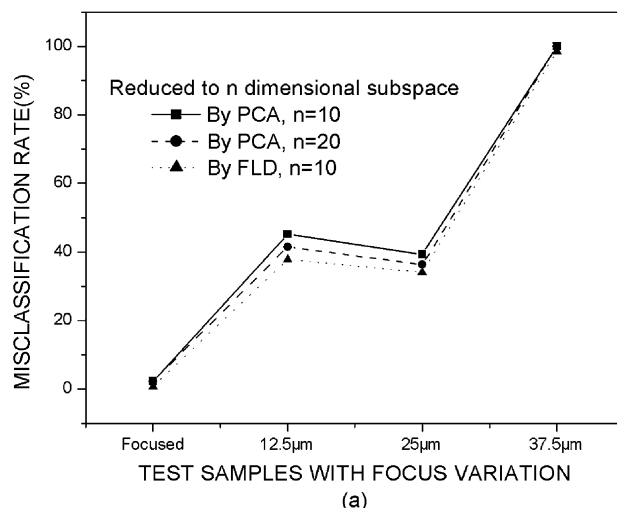


Fig. 5. Misclassification rates with different focus conditions and preprocessing methods. (a) Scheme 1: trained with only focused samples and applied to all samples. (b) Scheme 2: trained with focused and $25 \mu\text{m}$ focus variation samples and applied to all samples.

the group(s) used for training. This is reasonable because the classifiers have learned very similar data during the training. Second, increasing the number of principal components in PCA preprocessing did improve the performance, but it was still worse than that of FLD. Furthermore, both preprocessing methods performed similarly in Scheme 1 for each of the factors studied, but very differently in Scheme 2, with the error rate of FLD being much less than that of PCA in both interpolation and extrapolation tests. The reason lies in that, for Scheme 1, all images in the training set came exclusively from a single group, in which all microspheres had very homogeneous appearance. Therefore, when we extracted patches from these images and classified them into classes similar to those in Fig. 2, the within-class variations were very small. As expected, FLD was not superior to PCA in this case, since the variation was almost entirely between-class variation. Scheme 2, on the contrary, purposely introduced within-class variation into the training set by using images from different groups. In this case, the FLD method could learn the variation trend from the training set and choose projection directions that were

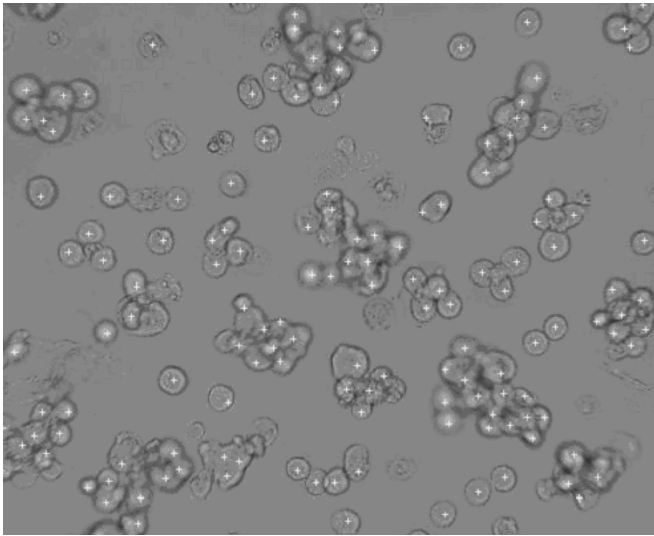


Fig. 6. Detection results of Scenario 3. The cell positions calculated were denoted by “+” in the images. Classifier: ANN; Preprocessing: FLD. Good agreement between calculated and actual positions as seen (Sensitivity: 94.38%, Positive predictive value: 91.52%).

nearly orthogonal to the within-class scatter, projecting away variations in focus, illumination, size, and noise; the PCA method could not. Consequently, the generalization ability of the neural network with FLD preprocessing was greatly improved and substantially better than a similar neural network with PCA preprocessing in Scheme 2-type experiments.

D. Living Cell Experiments

The method introduced in this paper was also systematically studied using bright field images of living cells. The testing images were divided into three groups denoting three different scenarios. Scenario 1 is the simplest case where cells are almost completely separate and the background is clean. Scenario 2 is more complex where most cells are attached to each other and there are trash and debris. Scenario 3 represents the most complex case where most cells are aggregated together and there is more trash and debris. These images show considerable out of focus blur, cells in clumps occupying multiple focal planes, as well as size variations.

The cell positions calculated by our system are denoted by white “+” in the images (see Fig. 6 for Scenario 3 result—Scenarios 1 and 2 are not shown). The detected cells were carefully examined and compared to the reference standard created by three experienced observers. Sensitivity (SE) and positive predictive value (PPV) [3] results are shown in comparison with the traditional PCA preprocessing method in Table I. The SE is defined as the percentage of cells in the reference standard, which are identified by the classifier. The PPV is the percentage of classifier detected cells which are also listed in the reference standard.

The results show that for Scenario 1, which presents a relatively simple case, both PCA and FLD produced very good results. For example, they both achieved SE values of 97.73% and PPV values of 100%. For Scenario 2, where the image is more complex, the SEs of PCA and FLD dropped to 87.76% and 95.92%, respectively, and PPVs dropped to 89.58% and

TABLE I
SE AND PPV RESULTS FOR LIVING CELL EXPERIMENTS

		Scenario 1	Scenario 2	Scenario 3
PCA	SE	97.73%	87.76%	82.5%
	PPV	100%	89.58%	83.02%
FLD	SE	97.73%	95.92%	94.38%
	PPV	100%	95.92%	91.52%

95.92%, respectively, which indicates that the FLD is superior to PCA when the image becomes more complex. This can be seen even more clearly in the very complex case represented by Scenario 3. Here, the SE percentage for FLD is 11.88 greater than that for PCA and the PPV percentage is 8.5 greater.

As noted previously, our results with microspheres suggest that FLD can better generalize from training sets with a single type of confounding factor. The experiments with living cells described in this section clearly show that FLD gives superior generalization even when multiple types of confounding factors are present simultaneously.

With regard to speed, when our current system is running with Matlab version 6.5.0.180913a (R13) on a standard PC with an Intel Pentium 4/1.6G processor and 256-MB RAM, a single training process usually takes about 30–40 min; the optimization process described in this paper usually takes 4–5 h. After the ANN is trained, a 640×480 sized image requires a processing time of 1–8 min, depending on the number of cells present. This is judged to be acceptable for many applications.

IV. CONCLUSION

An effective algorithm for cell recognition in bright field microscopy has been introduced in this paper. We developed a novel strategy to exploit the advantages of FLD preprocessing over PCA preprocessing. In addressing the variability of cell size and morphology, as well as variations in microscope parameters, such as focus and illumination, FLD preprocessing was clearly superior to PCA in all four types of variations studied. This holds true for both microsphere and actual cell recognition. The primary reason is that FLD optimizes the solution by maximizing the ratio of between-class scatter to within-class scatter, while PCA maximizes only the effect of total scatter.

REFERENCES

- [1] M. B. Kelz *et al.*, “Single-cell antisense RNA amplification and microarray analysis as a tool for studying neurological degeneration and restoration,” *Sci. Aging Knowl. Environ.*, 2002.
- [2] P. J. Sjöström, B. R. Frydel, and L. U. Wahlberg, “Artificial neural network-aided image analysis system for cell counting,” *Cytometry*, vol. 36, pp. 18–26, 1999.
- [3] T. W. Nattkemper, T. Twellmann, H. Ritter, and W. Schubert, “Human vs. machine: Evaluation of fluorescence micrographs,” *Comput. Biol. Med.*, vol. 33, pp. 31–43, 2003.
- [4] T. W. Nattkemper, H. Ritter, and W. Schubert, “A neural classifier enabling high-throughput topological analysis of lymphocytes in tissue sections,” *IEEE Trans Inf. Technol. Biomed.*, vol. 5, no. 2, pp. 138–149, Jun. 2001.
- [5] T. W. Nattkemper, H. Wersing, W. Schubert, and H. Ritter, “A neural network architecture for automatic segmentation of fluorescence micrographs,” *Neurocomputing*, vol. 48, pp. 357–367, 2002.
- [6] W. L. Cleveland, I. Wood, and B. F. Erlanger, “Routine large-scale production of monoclonal antibodies in a protein-free culture medium,” *J. Immunol. Meth.*, vol. 56, pp. 221–234, 1983.

- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [8] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.
- [9] A. S. Pandya and R. B. Macy, *Pattern Recognition with Neural Networks in C++*. Boca Raton, FL: CRC, 1996.



W. Louis Cleveland received the B.S. degree in physics from Columbia University, New York, NY, and the Ph.D. degree in chemistry from Rutgers University.

His postdoctoral training in basic immunology was in the Department of Microbiology, Columbia University. He is currently Director of the Laboratory for Molecular Mechanisms in Human Diseases at St. Luke's Roosevelt Hospital Center and is a Research Scientist in medicine at Columbia University. His current research interests include molecular mecha-

nisms in cancer and schizophrenia and the development of technology to study these mechanisms.



Xi Long (S'04) received the M.S. and B.S. degrees in mechanical engineering from Tsinghua University, China, in 2001 and 1998, respectively. He is working toward the Ph.D. degree in the Department of Mechanical Engineering at Columbia University, New York, NY.

His research interests are in computer vision, medical image processing, statistical learning, and pattern recognition.



Y. Lawrence Yao received the Ph.D. degree from the University of Wisconsin, Madison, in 1988.

He is a Professor in the Department of Mechanical Engineering at Columbia University, New York, NY. He is interested in multidisciplinary research in manufacturing and design, nontraditional manufacturing processes, laser materials processing, industrial manipulators, and monitoring and fault diagnosis of manufacturing processes and machinery.