Multiclass Detection Of Cells In Multicontrast Composite Images

Xi Long[a], **W. Louis Cleveland**[b], and **Y. Lawrence Yao**[a]

[a]Mechanical Engineering Department, Columbia University, New York, NY 10027 USA

[b]Department of Medicine at St. Luke's-Roosevelt Hospital Center and Columbia University, New York, NY 10019 USA

Abstract

In this paper, we describe a framework for multiclass cell detection in composite images consisting of images obtained with three different contrast methods for transmitted light illumination (referred to as multicontrast composite images). Compared to previous multiclass cell detection results (Long *et al.*, 2008), the use of multicontrast composite images was found to improve the detection accuracy by introducing more discriminatory information into the system. Preprocessing multicontrast composite images with Kernel PCA was found to be superior to traditional linear PCA preprocessing, especially in difficult classification scenarios where high-order nonlinear correlations are expected to be important. Systematic study of our approach under different overlap conditions suggests that it possesses sufficient speed and accuracy for use in some practical systems.

Keywords

Cell detection; Multicontrast composite images; Kernel methods; Support Vector Machines

1. Introduction

To achieve high-throughput with robotic systems based on optical microscopy, it is essential to replace the human observer with computer vision algorithms that can identify and localize individual cells as well as carry out additional studies on these cells in relation to biochemical parameters (Huang, 2003, 2004; Zhao *et al.*, 2004; Murphy, 2004, 2005). Considering the fact that the latter task is best achieved with the use of fluorescent probes and the number of fluorescence channels is limited, it is highly desirable to accomplish the cell identification and localization task with transmitted light microscopy, e.g. brightfield illumination. As a first step towards the goal of fully automated microscopy, algorithms for detection of unstained cells of a single type in brightfield images have been developed (Long *et al.*, 2005 a, b, 2006). To proceed further, it is of critical importance to develop algorithms that can detect cells in mixtures of multiple cell types and sort them into subtypes, which we refer to as "multiclass cell detection".

In our first attempt at multiclass cell detection, we previously formulated the task as a supervised, multiclass pattern recognition problem and solved it by extension of the Error Correcting Output Coding (ECOC) method (Dietterich and Bakiri, 1995; Allwein *et al.*, 2000) to enable probability estimation (Long *et al.*, 2008). The use of ECOC introduces redundancy into the ensemble of classifiers to increase detection accuracy. The probability estimation provided both cell-type identification as well as cell localization relative to pixel coordinates. In one implementation using this new algorithm with Support Vector Machines (SVMs) (Vapnik, 1998; Burges, 1998) as base binary classifiers, we were able to subtype and localize cells in brightfield images of cell mixtures prepared by mixing cells from three different cell lines. However, in this application, the detection accuracy was only about 80%, which is judged to be inadequate for many cases. In order to generate a more practical tool for automatic cell detection, more sophisticated algorithms are needed.

It should be pointed out that, with respect to microscopy, brightfield images represent a worst-case scenario and may not provide enough information for classification. To further explore the task of multiclass cell detection, it is therefore natural to also introduce additional discriminatory information into input images. A possible way to achieve this goal is to perform cell detection in multidimensional images, since they contain much more discriminatory information that derives from the use of different imaging techniques (channels) and conditions. In the field of transmitted light microscopy, commonly used techniques include phase contrast and Hoffman modulation contrast. Image sets that contain multiple images of the same specimen obtained with different microscopy techniques or different conditions with the same technique are called multidimensional or multivariate images. In many cases, they can provide information well beyond the limits achievable with individual techniques (Nattkemper, 2004). Therefore, cell detection in multidimensional images has a great potential to improve the accuracy over that achieved with single channels.

A major task for object recognition in multidimensional images is to extract essential information contained in the image stack (Bonnet *et al.*, 2000). Since a multidimensional data set always contains some redundant information that is not essential for recognition, it is therefore often advantageous to reduce the dimensionality of the vectors used to represent the objects studied. This process can help to define object representations that are more suitable for classification.

Currently existing approaches for automatic analysis of multidimensional cell images are dominated by techniques that linearly combine information from different images. For example, Nattkemper *et al.*, first performed automatic lymphocyte detection on all individual images in a fluorescence image stack and then linearly combined the detection results with heuristic rules (Nattkemper *et al.*, 2001, 2003, 2004). Wuringer *et al.* achieved automatic coregistration, segmentation, and classification of cell nuclei in a similar fashion (Wuringer *et al.*, 2004). Furthermore, linear Principal Component Analysis (PCA) has also been directly applied on fluorescence image stacks to reduce the dimensionality of input data (Bonnet *et al.*, 2000). The results reported in these examples are very promising. However, these applications concentrated on image stacks in which different parameters of the same imaging technique were applied to record a set of images. Problems of this category are called intramodular problems (Nattkemper *et al.*, 2004). They usually have identical spatial resolutions and similar pixel-value scales. In this paper, we explore the possibility of combining different contrast methods under transmitted light illumination. We refer to image stacks of this type as "multicontrast composite images". It should be pointed out that, by definition, multicontrast composite images are also multidimensional images. However, the analysis of these images represents an intermodular problem as defined by (Nattkemper *et al.*, 2004).

Compared to intramodular problems, intermodular problems represent a much more difficult challenge. This is because the imaging techniques used in an intermodular problem are typically based on different physical effects which may generate image stacks with different spatial resolutions and pixel-value scales. Consequently, the correlation of the pixel information in the intermodular images is much more complex than that of the intramodular images. In this case, a linear feature extraction as described above is less likely to be adequate. This is due to the fact that linear feature extraction decomposes data set into orthogonal components. However, the real sources of information in intermodular images have very little chance to be orthogonal (Bonnet *et al.*, 1998). To capture the basic sources of information, nonlinear feature extraction algorithms are needed.

Recently, nonlinear statistical analysis techniques have started to be used in the field of automatic multidimensional image analysis. Among these techniques, Kernel PCA (Schölkopf *et al.*, 1998) is the most popular one. Kernel PCA is a nonlinear generalization of the conventional linear PCA using kernel based methods (Müller *et al.*, 2001). It has been successfully used in applications ranging from image fusion (Twellmann, 2004) to multiview feature tracking (Meltzer *et al.*, 2004) to object recognition in 2D images (Schölkopf *et al.*, 1998) and has been shown to extract features with improved efficiency. In the study reported here, a multiclass cell detection algorithm in composite images obtained with three contrast methods in transmitted light illumination (brightfield, phase contrast and Hoffman modulation contrast) is, to our knowledge, proposed for the first time. This paper is an extension of our previous work. However, instead of using linear PCA to extract features from brightfield images, we used Kernel PCA to extract features from the composite images. The experimental results suggest that: 1) the use of multiple contrast methods provides more discriminatory information about the cells and therefore improves the detection accuracy; 2) for multicontrast composite images of cells, Kernel PCA extracts features that are more suitable for classification compared to linear PCA.

2. Materials and Experimental Conditions

The cell lines used in this work were K562 (human chronic myelogenous leukemic cells, ATCC; Cat. No.CCL-243), A20.2J (murine B-cell lymphoma cells, ATCC; CAT. NO. HB-97) and EAT cells (Ehrlich Ascites Tumor cells, ATCC; Cat. No.CCL-77). All cells were grown at 37.0 °C in BM+1/2 TE1+TE2 +10% fetal calf serum (FCS) (Cleveland *et al.*, 1983). For microscope observation, cells in culture medium were dispensed into polystyrene 96-well microplates, which have glass bottoms that are 0.175mm thick. Cell viability was determined by nigrosine staining (Mishell *et al.*, 1980) before and after microscope observation and was greater than 95%.

To obtain an accurate and objective training and testing standard, the fluorescent probes for living cells (CellTracker™ CAT. No. C2110, C2925 and C34552, Molecular Probes) were used to label different cell lines. K562 cells were labeled with blue fluorescent probe (C2110) and observed with a standard DAPI filter set (31000v2, Chroma). CR10 cells were labeled with green fluorescent probe (C2925) and observed with a standard FITC filter set (41001, Chroma). EAT cells were labeled with red fluorescent probe (C34552) and observed with a standard Texas Red filter set (41004, Chroma).

An inverted microscope equipped with a 20× Nikon Hoffman modulation contrast objective (Numerical Aperture: 0.45), a 20× Nikon phase contrast objective (Numerical Aperture: 0.45) and an Andor iXon DV 887-BI EMCCD camera was used to obtain digitized images. Brightfield images were obtained with the Hoffman modulation contrast objective using the Hoffman condenser in the brightfield position. For each microscope field, a set of six images was acquired (Fig. 1). Three images were acquired with three different contrast methods in

transmitted light illumination (brightfield, phase contrast and Hoffman modulation contrast) and were used for SVM training or testing. Three auxiliary fluorescence images were also acquired to distinguish different cell lines, which were labeled blue, red or green.

Twenty sets of microscope images were acquired and used in our cell detection experiments. In each experiment, two subsets were extracted: one exclusively for training and another exclusively for testing. To avoid introducing false training samples, ambiguous objects showing more than one fluorescence color were manually deleted from the training images. The total number of deleted objects was less than 1% of the total cell count.

The computer programs were written in MATLAB and C++. Our algorithm was implemented with the LIBSVM version 2.5 (Lin *et al.*, 2004), which was compiled as a dynamic link library for MATLAB. All experiments were implemented in the environment of MATLAB Version 6.5.0.180913a (R13) supplemented with Image Processing Toolbox Version 3.2. A standard PC equipped with an Intel Pentium 4/2.8G processor and 256-MB RAM was used.

3. Overall Framework for Cell Detection in Multicontrast Composite Images

In this section, a multiclass cell detection framework for multicontrast composite images of cultured cells is presented. The framework uses pixel patches as the primary input data elements and employs Kernel PCA in the preprocessing step to nonlinearly reduce the dimensionality of the input data. It also employs the multiclass classification and probability estimation (Long *et al.*, 2008) for image analysis, which permits not only the identification of the desired cells but also gives their locations relative to the pixel coordinates of the primary image. Essentially, the software is taught to classify pixel patches into different classes. Each class corresponds to a single cell-type, except for the larger class containing all undesired objects (e.g. background, fragments of cells, trash, and off-center cells), denoted as "Non-cell". For cell classes, the localization information of the classified pixel patches in the image is used as cell locations.

The essential aspects of this framework are illustrated in Fig. 2. Basically, we first train an ensemble of SVM classifiers with ECOC. This is done with input vectors that are derived from manually-extracted training patches and are represented as nonlinear combinations of feature vectors derived in Kernel PCA preprocessing (Schölkopf *et al.*, 1998).

For each position *p* (specified by pixel coordinates) in the testing images (excluding positions at the edges which do not permit complete pixel patches) three pixel patches centered at that position are extracted and represented in the same way as that in the training process, one patch from each channel. The probability that the input vector derived from these extracted patches belongs to each class is calculated by ECOC probability estimation (Long *et al.*, 2008). For each class corresponding to a cell type, this probability is then used as a "confidence value" C [*p*] ∈ [0,1] in a "confidence map" for that cell type. Pixels in each confidence map are the confidence values of their corresponding patches in the original images and form "mountains" with large peaks representing a high probability of presence of the corresponding cell type. A given peak in a confidence map is compared with the corresponding peaks in the other confidence maps. The confidence map with the highest peak at that location gives the assignment of class membership. Localization is provided by the pixel coordinates of the highest peak. It should be pointed out that generating a confidence map for the "Non-cell" class is unnecessary in our case since localization of the non-cell objects is not important for us.

In the ECOC approach, binary classifiers have to be trained as the base classifiers. The choice of base classifier can be arbitrary (Long *et al.*, 2008). In this work, we used Support Vector Machines (SVM) (Vapnik, 1998; Burges, 1998) with the RBF kernel $K(x,y) = e^{-\gamma\|x-y\|^2}$ The SVM classifier in our experiment is implemented by modifying LibSVM. SVMs using the

RBF kernel require empirical optimization of two parameters, C and γ. The regularization parameter C controls the trade-off between errors in training data and the SVM margin maximization. It creates a soft margin that permits some misclassifications in training. Increasing the value of C increases the cost of errors at training time and forces the creation of a more accurate model on the training data, which may or may not generalize well on the testing data. C and the RBF kernel parameter γ are optimized using a two-step "grid-search" method for each classifier (Lin *et al.*, 2004). In the first step, a coarse grid-search with a grid size of 1 was used to localize a Region of Interest (ROI) containing the optimal values (shown in Fig. 3). In the second step, a fine grid-search over the ROI with a grid size of 0.25 was used to give more precise values for C and γ. The result is shown in Fig. 4.

4. A Brief Summary of Kernel PCA

As noted above, Kernel PCA is used in the preprocessing stage to reduce the dimensionality of input data (raw pixel patches). In this section, some fundamental aspects of Kernel PCA are briefly described. A detailed introduction to Kernel PCA can be found in (Schölkopf *et al.*, 1998). Principal component analysis (PCA) is one of the most popular dimension reduction methods. Given a set of M centered samples $\mathbf{x}_k : k = 1,2,\ldots,M$, $\mathbf{x}_k \in \mathbf{R}^N, \sum_{k=1}^{M}\mathbf{x}_k=0$, PCA diagonalizes the covariance matrix:

$$\mathbf{C}=\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i\mathbf{x}_i^{T}$$

(1)

by solving the eigenvalue equation

$$\lambda\mathbf{v}=\mathbf{C}\mathbf{v}$$

(2)

for $\lambda \geq 0$ and $\mathbf{v} \in \mathbf{R}^{N} \setminus \{0\}$. The eigenvectors $\mathbf{v}$ corresponding to the bigger eigenvalues λ capture more variance. Therefore, the set of the first $n \leq N$ eigenvectors or Principal Directions (PDs) carry more variance than any other n orthogonal projections.

In recent years, kernel based methods have been more and more widely accepted as a new revolution in the machine learning area. The concept of kernel based methods is based on the combination of well-known linear algorithms such as PCA with nonlinear kernel functions (Müller *et al.*, 2001), which allow more powerful nonlinear solutions.

Consider a nonlinear mapping

$$\begin{aligned}\mathbf{\Phi}:&\mathbf{R}^n \to \mathbf{H}\\ &\mathbf{x} \mapsto \mathbf{x}'=\mathbf{\Phi}(\mathbf{x})\end{aligned}$$

(3)

which maps the examples $\mathbf{x}_k \in \mathbf{R}^N$ to some feature space $\mathbf{H}$. Assuming that the mapped data are centered in $\mathbf{H}$ (if not, they can be easily centered. See Schölkopf *et al.*, 1998), one can perform PCA in $\mathbf{H}$. This is equivalent to find the eigenvectors λ of the covariance matrix in $\mathbf{H}$ (Müller *et al.*, 2001)

$$\tilde{\mathbf{C}}=\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}'_i\mathbf{x}'^{T}_i=\frac{1}{M}\sum_{i=1}^{M}\mathbf{\Phi}(\mathbf{x}_i)\mathbf{\Phi}(\mathbf{x}_i)^{T}$$

(4)

It can be easily shown that the eigenvectors $\mathbf{v}'$ in $\mathbf{H}$ lie in the span of $\mathbf{\Phi}(\mathbf{x}_1),\ldots, \mathbf{\Phi}(\mathbf{x}_M)$. Therefore, the above problem is equivalent to the following eigenvalue problem (Schölkopf *et al.*, 1998)

$$M\lambda\boldsymbol{\alpha}=\mathbf{K}\boldsymbol{\alpha} \qquad (5)$$

where $\boldsymbol{\alpha}$ is a column vector of coefficients $[\alpha_1,\ldots, \alpha_M]^T$, whose elements describe the dual form of the eigenvector by

$$\mathbf{v'}=\sum_{i=1}^{M}\alpha_i\mathbf{\Phi}(\mathbf{x}_i) \qquad (6)$$

and $\mathbf{K}$ is a symmetric matrix, called Gram matrix, with elements

$$\mathbf{K}_{ij}=\mathbf{x'}_i \cdot \mathbf{x'}_j=\mathbf{\Phi}(\mathbf{x}_i) \cdot \mathbf{\Phi}(\mathbf{x}_j), \quad i, j=1,\ldots, M \qquad (7)$$

By normalizing $\boldsymbol{\alpha}^k$ corresponding to the kth eigenvalue λ_k of $\mathbf{K}$ to ensure $\lambda_k(\boldsymbol{\alpha}^k \cdot \boldsymbol{\alpha}^k) = 1$, principal components in $\mathbf{H}$ can be extracted by projecting example $\mathbf{x}$ on $\mathbf{v}'^k$

$$\tilde{x}=\mathbf{v'}^k \cdot \mathbf{x'}=\mathbf{v'}^k \cdot \mathbf{\Phi}(\mathbf{x})=\sum_{k=1}^{M}\alpha_i^k(\mathbf{\Phi}(\mathbf{x}_i) \cdot \mathbf{\Phi}(\mathbf{x})) \qquad (8)$$

It should be noted that, in equation (8), all vectors appear in the form of the inner product of the mapping $(\mathbf{\Phi}(\mathbf{x}_i) \cdot \mathbf{\Phi}(\mathbf{x}))$ instead of the explicit mapping alone. Consequently, kernel functions can be used to calculate the mapping implicitly. This results in

$$\tilde{x}=\sum_{k=1}^{M}\alpha_i^k k(\mathbf{x}_i, \mathbf{x}) \qquad (9)$$

where the kernel function $k(\mathbf{x}_i,\mathbf{x}_j) = \mathbf{\Phi}(\mathbf{x}_i)\cdot\mathbf{\Phi}(\mathbf{x}_j)$ returns the result of dot product of two feature vectors in the feature space $\mathbf{H}$.

The basic principle of Kernel PCA is illustrated in Fig. 5. With Kernel PCA, one is actually performing linear PCA in some high dimensional feature space $\mathbf{H}$, analogous to linear PCA in the input space. Since $\mathbf{H}$ is nonlinearly related to the input space, the linear relationship (dashed line) that is generated in the feature space $\mathbf{H}$ is nonlinear in input space (dashed curve). Kernel PCA can therefore obtain more powerful nonlinear solutions while retaining most of the properties of linear PCA.

5. Results and Discussion

5.1 Control experiments

As has been mentioned above, to obtain an accurate and objective training and testing standard, fluorescent probes with different characteristic colors were used in our experiments to stain cells from different cell lines. Ideally, the fluorescent probes should only be detectable under fluorescence and should be completely invisible to the classifiers in images obtained with transmitted light illumination. Although the concentrations of the fluorescent probes were

carefully controlled to the lowest possible level as determined experimentally, there is still the possibility that the fluorescent light from the probes could alter the transmitted light images in a way that could be detected by the classifiers. This would, of course, affect the cell detection accuracy. To check if the use of fluorescent probes affects cell detection ability of the classifier under transmitted light illumination, a series of control experiments were performed.

The control experiments were done as follows

1. stained and unstained cells of the same type were first mixed in equal proportions and the images were taken.

2. a 3-class classifier was trained to classify objects in the following categories: a) stained cells; b) unstained cells and c) non-cells.

3. Experiments were then done to determine if the classifiers trained as described in 2) could distinguish between fluorophore-labeled and unlabeled cells. If the fluorescent probes do not affect cell detection, then the distinction between stained and unstained cells should be impossible. Since the distinction between cell (categories a & b) and non-cell objects (category c) is much easier than that between stained (category a) and unstained cells (category b), the detection accuracy of both stained and unstained cells should be close to 50%, as expected for random classification.

   As a metric of randomness, P-values for the problem of distinguishing between stained and unstained cells were also calculated. The P-value is defined as the probability of getting something more extreme than the result, when there is no effect in the population. If the fluorescent probes do not affect cell detection, P-values should be close to 1.

4. Repeat the experiment for all three cell types (A20.2J, EAT and K562) and for all transmitted light channels (brightfield, Hoffman modulation contrast and phase contrast).

Shown in Fig. 6 are some typical sample images used in the control experiments. Fig. 6 (a) shows a brightfield image of mixture of stained and unstained A20.2J cells. Its corresponding fluorescence image is shown in Fig. 6 (b). The detection result of Fig. 6 (a) is shown in Fig. 7, in which the stained and unstained cells detected are denoted by squares and crosses respectively.

A total of 9 control experiments have been done to cover all possible combinations of three cell types and three transmitted light channels. The detection accuracies of stained (category a) and unstained cells (category b) are summarized in Fig. 8. One can see from the figure that, in all cases, the detection accuracies of both stained and unstained cells are very close to 50%. Considering the fact that the detection accuracies are very high (>95%, shown in Fig. 9) for the entire set of cell objects (stained & unstained), the P-values suggest a random classification of stained versus unstained cells. To show the randomness of classification in above experiments, the P-values for distinguishing between stained and unstained cells from different cell lines under different imaging channels are plotted in Fig. 10. As expected, all P-values are very close to 1. Thus, the results of the control experiments suggest that the use of fluorescent probes does not affect cell detection accuracy of any cell type in any transmitted light channel.

5.2 Automatic cell detection in multicontrast composite images

In this section, we evaluate quantitatively multiclass cell detection in multicontrast composite images of cell mixtures prepared by mixing cells from three different cell lines (A20.2J, EAT and K562). The composite images consist of images obtained with three contrast methods in transmitted light illumination: brightfield, Hoffman modulation contrast and phase contrast.

The overall framework of this approach has been described in Section 3. In what follows, the detailed experiment is described in steps. The experimental result is also quantitatively analyzed.

1. Pixel patch extraction and generation of primary input vectors—Since individual cells typically occupy only a small percentage of total image area, it is advantageous to decompose images using pixel patches that just large enough to contain the largest cells in the image. In actual experiments, images obtained with different contrast methods were first manually coregistered. Then, for a possible location in the coregistered microscope images (except in the 15-pixel margin around the edges), three 29×29 pixel patches centered at that location were extracted. These pixel patches can be viewed as a 29×29×3 dimensional input vector (see Fig. 11). Our experiments indicate that performance is not very sensitive to small variations in patch size, e.g. a patch size of 31×31 produced similar results (data not shown). Since many locations in the image are uniform background, a "mask" was created to exclude these patches. Essentially, the "mask" eliminated all pixel patches whose average pixel intensities were below a user-chosen threshold.

A training set was created with the aid of an interactive program that displays the digitized microscope images and allows a user to select the locations of cell centers with a mouse cursor after manual comparison of transmitted light and fluorescence images. For each cell type, the pixel patches extracted from the selected cell locations were used as input vectors of that class. The input vectors in the "Non-cell" class were then generated automatically by extracting all the pixel patches whose centers were $r \geq 7$ pixels away from any of the manually selected cell locations. The value of r was empirically chosen in relation to the sizes of cells and pixel patches.

2. Kernel PCA preprocessing and construction of preclassified training set—
The primary input vectors generated above are 29*29*3=2523 dimensional. To define a new representation space that is more suitable for classification, Kernel PCA preprocessing was used to reduce dimensionality to n=10 for all input vectors (Fig. 11). The dimensionality of 10 is experimentally chosen (data not shown). We followed the method described in (Schölkopf *et al.*, 1998) and used a polynomial kernel function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$$

(10)

which is parameterized by the degree of polynomial d. This parameter is chosen by testing the range 1 through 6 using a manually extracted testing set of 1000 samples (250 for each class). To simplify the optimization process, we also followed Schölkopf *et al.*'s strategy (Schölkopf *et al.*, 1998) and used linear SVM classifiers. The experimental results show that the classification accuracy gets improved as the degree goes higher. However, increasing from the degree 5 to 6 only gives very limited improvement (0.1%). We therefore chose d=5 for our actual cell detection experiments.

After all input vectors are preprocessed, each attribute of the Kernel PCA-preprocessed vectors was linearly scaled to the range [−1, +1]. The main advantage of scaling is to avoid computational difficulties and to avoid the dominance of attributes with greater numeric ranges over those with smaller numeric ranges (Lin *et al.*, 2004). Finally, the classes were labeled with ordinal numbers.

3. ECOC training—The same procedure in our previous work (Long *et al.*, 2008) was used in this study. Basically, we formulate the detection of multiple cell types in mixtures as a supervised, multiclass pattern recognition problem and solve it by training an ensemble of

binary classifiers based on Error Correcting Output Coding (ECOC). The ECOC procedure creates teams of binary classifiers that "vote" for classification. The ECOC was implemented with a sparse matrix that was selected from 10000 randomly generated 4×10 matrices. To select the optimum matrix in the set of 10000, we calculated the minimum Hamming distance between all pairs of the rows for each matrix. The matrix with the biggest minimum distance was chosen (Lin *et al.*, 2004). For each binary SVM classifier, the parameters are independently optimized following the aforementioned two-step grid search procedure. During the process of binary classifier training, the Compensatory Iterative Sample Selection (CISS) algorithm (Long *et al.*, 2006) was employed to address the imbalance problem caused by the large "Non-cell" sample set. This algorithm maintains a fixed-size "working set", in which the training samples are kept balanced by iteratively choosing the most representative training samples for the SVM. These samples are close to the boundary and are therefore more difficult to classify. This scheme can make the decision boundary more accurate, especially when applied to difficult scenarios.

4. Identification and localization of living cells in multicontrast composite images—In order to examine the effect of Kernel PCA preprocessing on images with different levels of complexity, three different scenarios were created. In Scenario 1, only images obtained with brightfield channel are used. Since there is no difference in spatial resolutions and pixel value scales in a single channel, this scenario represents a very simple case. Scenario 2 is more complex since it uses the combination of brightfield and Hoffman modulation contrast images. Scenario 3 represents the most complex case where images obtained with all three contrast methods are combined for cell detection. For each scenario, three image groups representing different levels of cell density and overlapping conditions were tested. Images from Group 1 have very low cell density and very little cell overlap. The cell density in Group 2 is moderate and represents an ideal situation for automatic cell micromanipulation. Cell density in Group 3 is very high, making images from this group very difficult for automatic detection.

An ensemble of SVM classifiers was trained and tested on each image group in each scenario (totally 9 cases). For each ensemble, testing samples were from the same image group as the training samples. However, none of the samples used for training were used for testing.

The classifier ensembles were applied to pixel patches obtained by automatic pixel patch decomposition of entire microscope images described in Section 3. Fig. 12 shows the confidence maps for images shown in Fig. 1 (Scenario 3-scheme, Kernel PCA preprocessing). The range of the confidence value ([0,1]) in the confidence maps has been linearly scaled to [0,255] for grayscale representation. The detection results of the same images with PCA and Kernel PCA preprocessing are shown in Fig. 13 and Fig 14 respectively. The detected cell positions are denoted by different symbols (diamond, square and cross, one for each class) in the image.

Statistical cell detection results for whole microscope images in Scenarios 1, 2 and 3 are summarized in Fig. 15, Fig 16 and Fig 17, respectively. For each group, five testing image sets were used. We employed a "Free-response Receiver Operating Characteristics" method (FROC) (Chakraborty, 1989) with the average false positive (FP) number of all cell types in each image and the average sensitivity (true positive percentage, i.e., the percentage of cells that are identified correctly) of all cell types as performance indexes. As described above, the cell positions are identified as "peaks" of the "mountains" in the confidence maps. This requires a user-defined threshold for the definition of "peak". The FROC curve plots the relationship of false positives and sensitivity as a function of the threshold (not explicitly represented in the plot). In a practical application, a suitable threshold can then be selected to achieve the required behavior. Generally speaking, the bigger the area under the curve, the better the result

is. The results are shown in the form of systematical comparison between Kernel PCA and PCA preprocessing.

Some interesting points are revealed in these experiments. Firstly, as one can see from the results, the detection accuracy increases when switching from Scenario 1 to Scenarios 2 & 3 for both PCA and Kernel PCA preprocessing. This is reasonable since the use of additional contrast methods introduces more discriminatory information into the system by observing cells from different perspectives. Secondly, in Scenario 1, both methods produce similar detection accuracies, with Kernel PCA preprocessing only slightly better than PCA preprocessing. This holds true for all three image groups. For Scenario 2, where two channels are combined for detection, Kernel PCA preprocessing starts to show more advantage. A much greater advantage of Kernel PCA can be seen when all three channels are combined in Scenario 3. For example, when applied to image Group 2 in Scenario 3, if the average false positive acceptance number in each image is set at 1, Kernel PCA preprocessing achieves a sensitivity of 87.7%, which is about 3.6 percentage points greater than that of PCA preprocessing (To facilitate capturing this trend, Group 2 testing results in each Scenario are plotted in thick lines). We believe the reason lies in that, in Scenario 1, since only a single channel is used, the correlation of the pixel information in the images is relatively simple. Kernel PCA has no advantages over PCA preprocessing in this case. However, as new channels are added, the correlation of the pixel information in the images becomes more and more complex in Scenario 2 & 3. Thus, kernel PCA becomes more and more advantageous, most likely because it can capture complex, nonlinear correlations in the high dimensional image space.

With regard to the processing speed, when our current method is used with a 29×29 pixel patch, a 492×453 image requires a processing time of 5–15 minutes, depending on the number of objects present in the image. However, as yet, optimization of speed has not been attempted.

6. Conclusions

A framework for multiclass cell detection in multicontrast composite images has been described. The use of multiple contrast methods improves the detection accuracy, primarily owing to its ability to provide discriminatory information well beyond the limits achievable with single contrast methods. Our experimental results also suggest that Kernel PCA preprocessing is superior to traditional linear PCA preprocessing. This is consistent with the expectation that Kernel PCA can more efficiently capture high-order, nonlinear correlations in the high dimensional image space. The accuracy of our multiclass cell detection framework suggests that it can be useful in some systems that require automatic subtyping and localization of cells in mixtures of multiple cell types. Finally, we note that the contrast-generating techniques used to produce multicontrast images in this study involve a large number of variables. Further optimization of these variables may provide additional accuracy.

Acknowledgments

References

Long X, Cleveland WL, Yao YL. "A New Preprocessing Approach for Cell Recognition". IEEE Transactions on Information Technology in Biomedicine 2005a;9(3):407–412. [PubMed: 16167695]

Long X, Cleveland W, Yao Y. "Effective Automatic Recognition of Cultured Cells in Brightfield Images Using Fisher's Linear Discriminant Preprocessing". Image and Vision Computing 2005b;23:1203–1213.

Long X, Cleveland WL, Yao YL. "Automatic Detection of Unstained Viable Cells in Brightfield Images Using a Support Vector Machine with an Improved Training Procedure". Computers in Biology and Medicine 2006a;36:339–362. [PubMed: 16488772]

Long X, Cleveland W, Yao Y. "Multiclass Detection of Cell Mixture in Brightfield Images with Multiclass ECOC Probability Estimation". Image and Vision Computing 2008;26:578–591.

Allwein E, Schapire R, Singer Y. Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research 2000;1:113–141.

Dietterich TG, Bakiri G. Solving Multiclass Learning Problems via Error- Correcting Output Codes. Journal of Artificial Intelligence Research 1995;2:263–286.

Vapnik, V. Statistical Learning Theory. Wiley; 1998.

Burges C. "A tutorial on Support Vector Machines for pattern recognition". Data Mining and Knowledge Discovery 1998;2:122–167.

Nattkemper T. "A Neural Network-Based System for High-Throughput Fluorescence Micrograph Evaluation". University of Bielefeld, Faculty of Technology. 2001 Feb;

Nattkemper TW. Multivariate image analysis in biomedicine: a methodological review. Journal of Biomedical Informatics 2004;37(5):380–391. [PubMed: 15488751]

Nattkemper, TW.; Saalbach, A.; Twellmann, T. Evaluation of Multiparameter Micrograph Analysis with Synthetical Benchmark Images. Cancun, Mexico, Proc. of EMBC2003; 25th Annual Int. Conf. of the IEEE Engineering in Med. and Biol. Soc.; 2003 Sep.

Wuringer T, et al. Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images. Comput Med Imaging Graph 2004;28(1–2):87–98. [PubMed: 15127753]

Bonnet N. Multivariate statistical methods for the analysis of microscope image series: applications in materials science. Journal of Microscopy 1998;190(1–2):2–18.

Bonnet, N., et al. "Dimensionality reduction, segmentation and quantification of multidimensional images: application to fluorescence microscopy". In: Bearman, Gregory H.; Cabib, Dario; Levenson, Richard M., editors. Spectral Imaging: Instrumentation, Applications, and Analysis; Proceedings of SPIE; 2000.

Murphy RF. Automated Interpretation of Subcellular Location Patterns. Proc 2004 IEEE Intl Symp Biomed Imaging (ISBI 2004) 2004:53–56.

Huang K, Velliste M, Murphy RF. Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. Proc. SPIE 2003;4962:307–318.

Huang K, Murphy RF. Automated Classification of Subcellular Patterns in Multicell images without Segmentation into Single Cells. Proc 2004 IEEE Intl Symp Biomed Imaging (ISBI 2004) 2004:1139–1142.

Zhao T, Velliste M, Boland MV, Murphy RF. Object Type Recognition for Automated Analysis of Protein Subcellular Location. IEEE Trans. Image Proc 2005;14:1351–1359.

Murphy RF. Cytomics and Location Proteomics: Automated Interpretation of Subcellular Patterns in Fluorescence Microscope Images. Cytometry 2005;67A:1–3.

Schölkopf B, et al. "Nonlinear component analysis as a kernel eigenvalue problem". Neural Computation 1998;10:1299–1319.

Twellmann T, et al. "Image fusion for dynamic contrast enhanced magnetic resonance imaging". Biomedical Engineering Online 2004;3:35. [PubMed: 15494072]

Meltzer, et al. "Multiple view feature descriptors from image sequences via kernel principal component analysis". Computer Vision - ECCV 2004;Pt 1(3021):215–227.

Müller K, et al. "An introduction to kernel-based learning algorithms". IEEE Neural Networks 2001;12(2):181–201.

Cleveland WL, et al. Routine large-scale production of monoclonal antibodies in a protein-free culture medium. Journal of Immunological Methods 1983;56:221–234. [PubMed: 6338115]

Mishell, BB., et al. Preparation of Mouse Cell Suspensions. In: Mishell, BB.; Shiigi, SM., editors. Selected methods in Cellular Immunology. New York: W.H. Freeman and company; 1980.

http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Chakraborty DP. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Medical Physics 1989;16:561–568. [PubMed: 2770630]
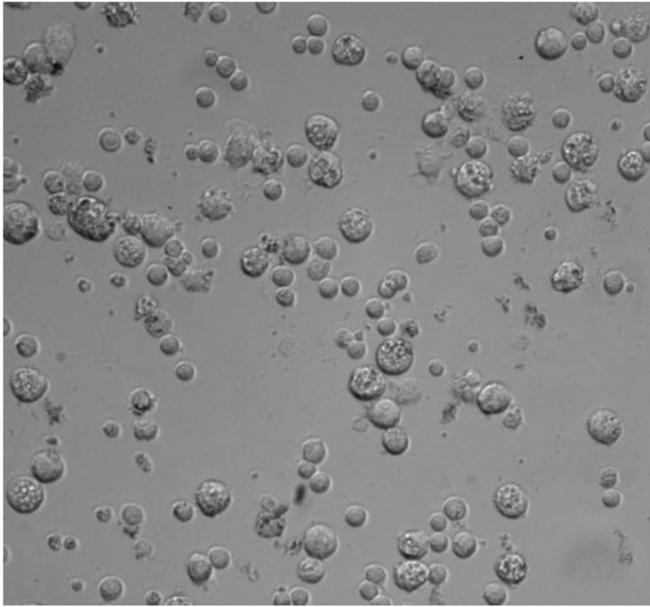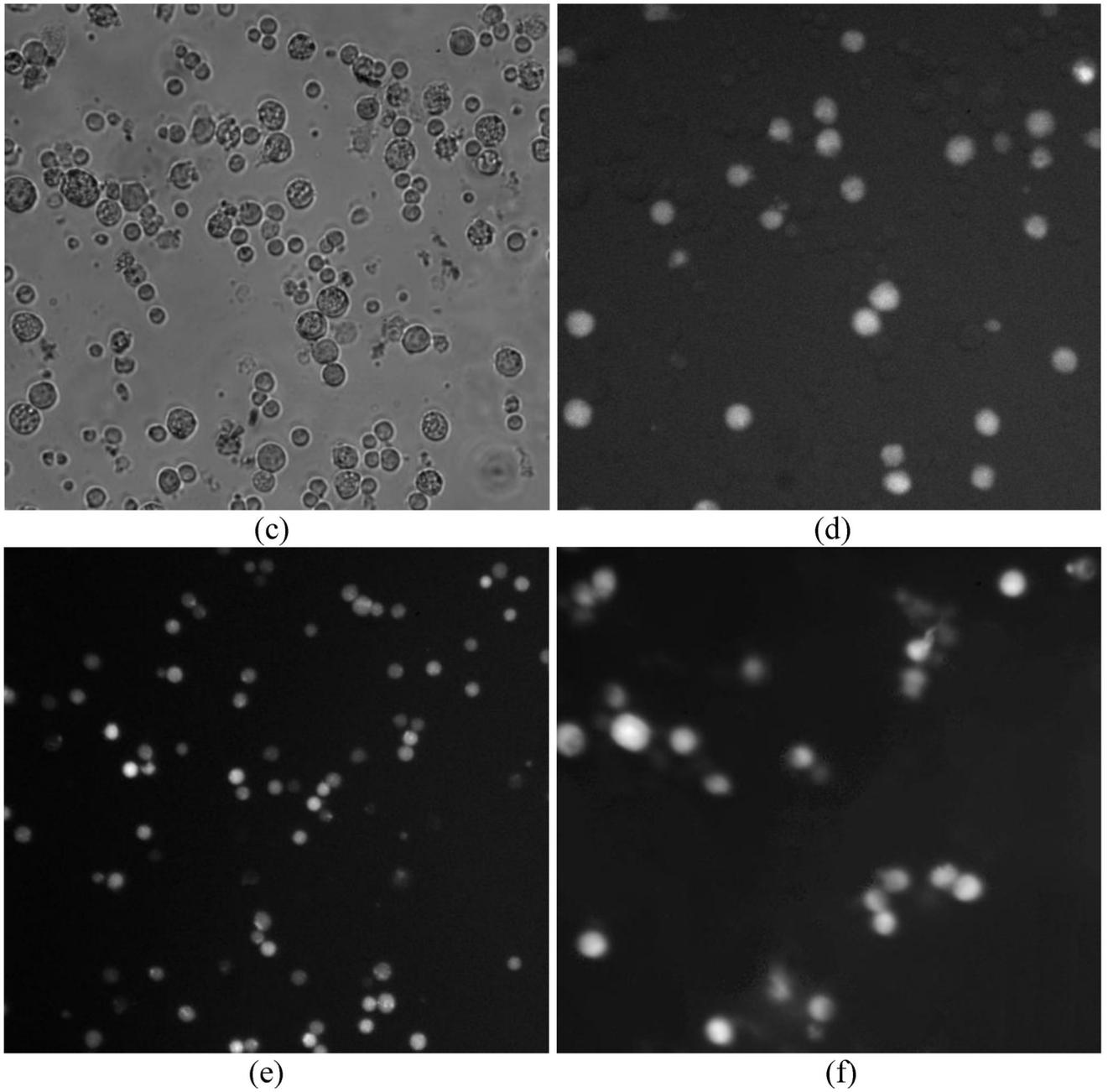
(a)
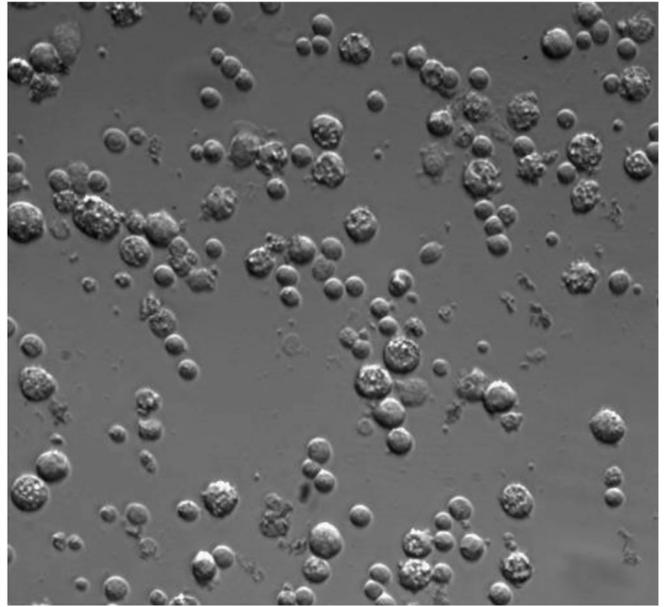
(b)

Fig. 1.
A typical sample image set. (a) brightfield. (b) Hoffman Modulation Contrast. (c) Phase Contrast. (d) Blue fluorescence channel (K562 cells). (e) Green fluorescence channel (A20.2J cells). (f) Red fluorescence channel (EAT cells).
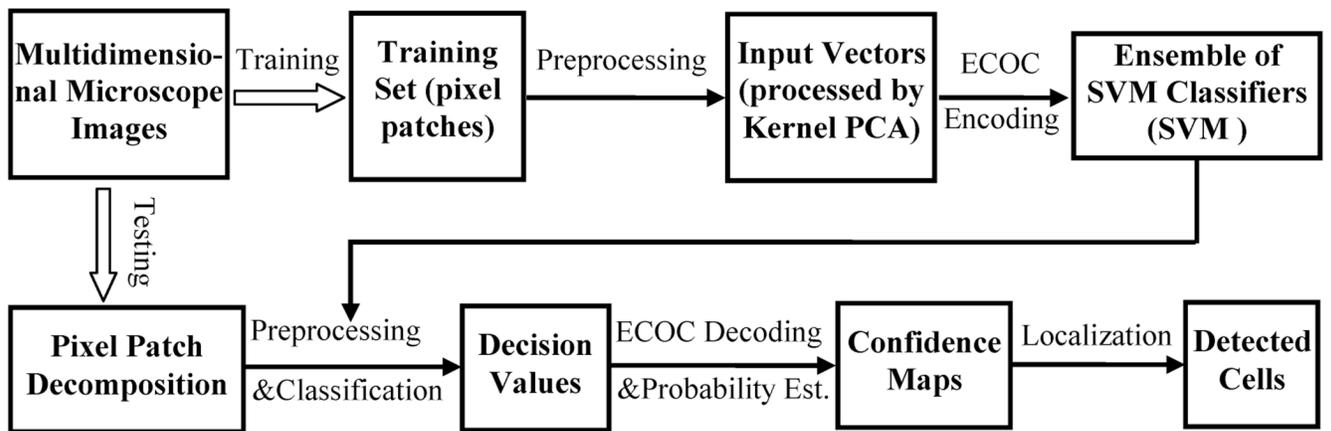
Fig. 2.
Illustration of the overall multiclass cell detection process in multicontrast composite images with ECOC probability estimation.
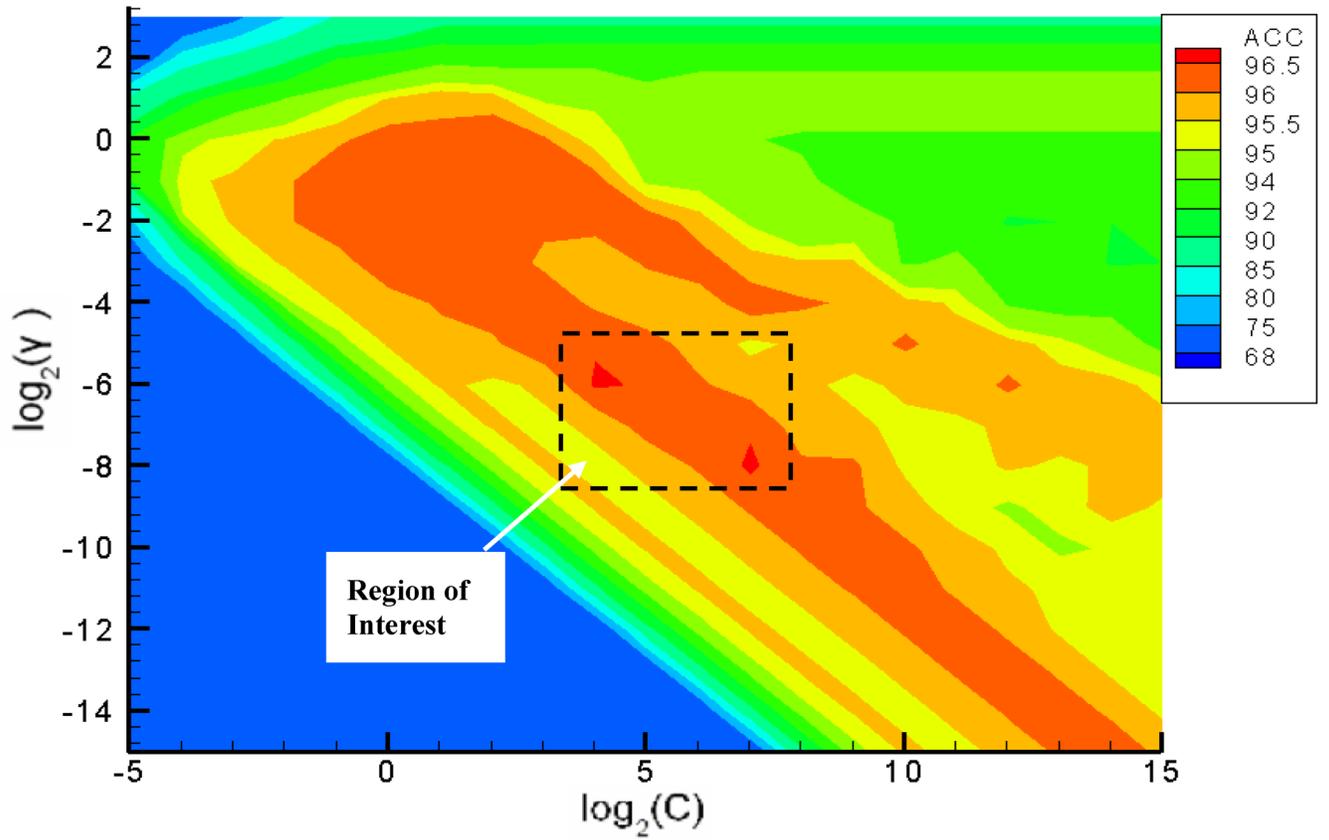
Fig. 3.
Coarse grid search, grid size = 1. *C* is the regularization parameter for SVMs; γ is the RBF
kernel parameter; ACC is the cross-validation accuracy.
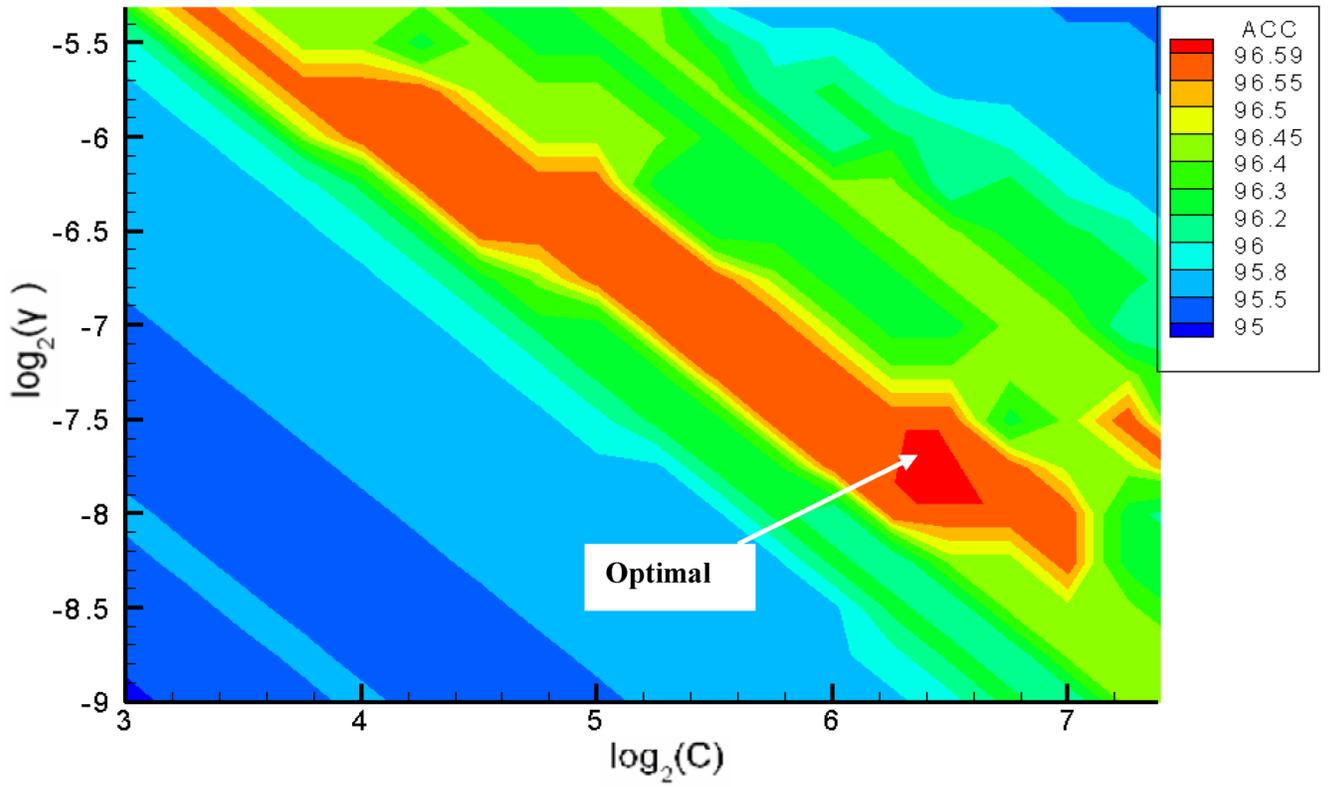
Fig. 4.
Fine grid search, grid size = 0.25. *C* is the regularization parameter for SVMs; γ is the RBF kernel parameter; ACC is the cross-validation accuracy.
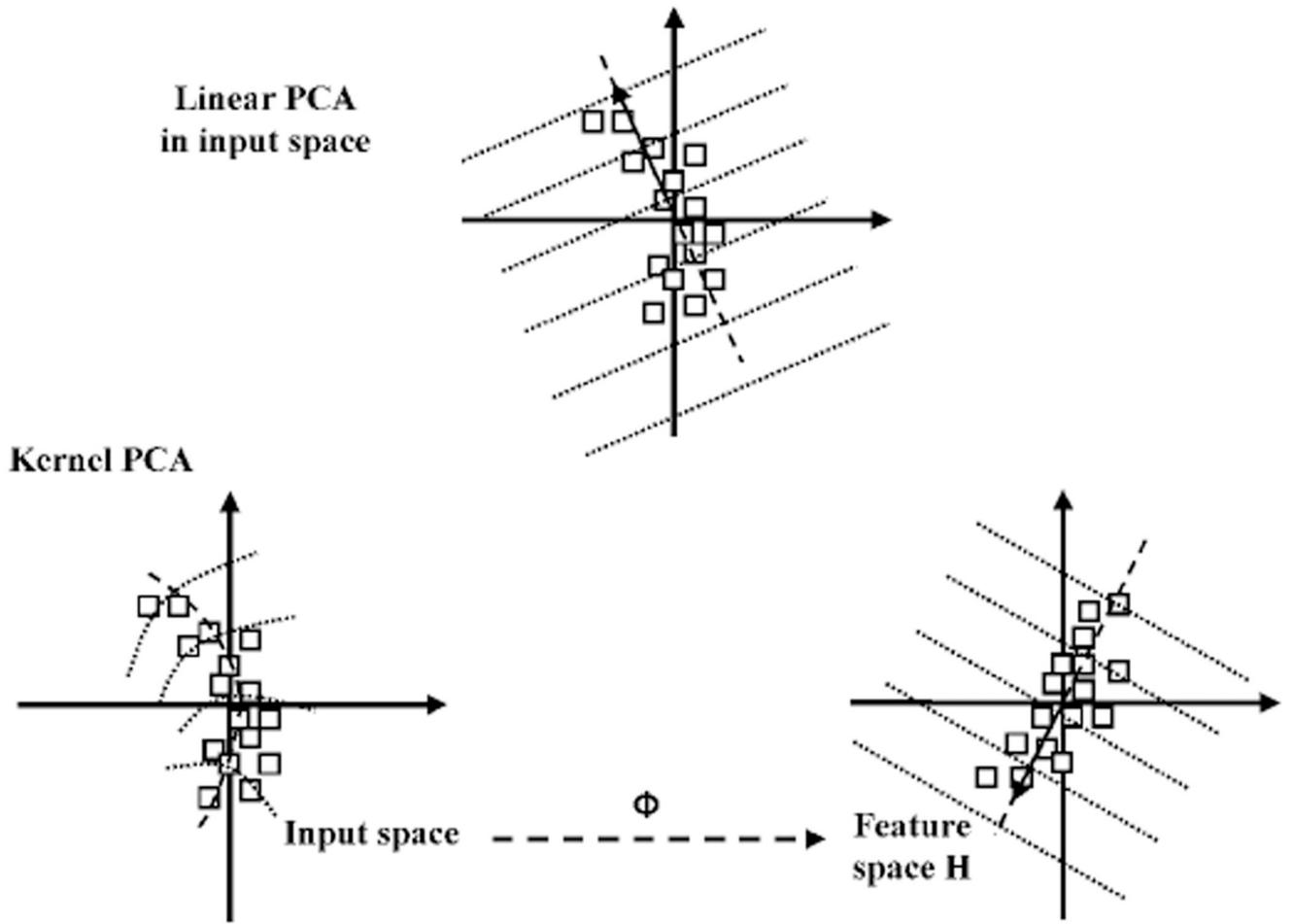
Fig. 5.
Illustration of the basic principle of Kernel PCA.
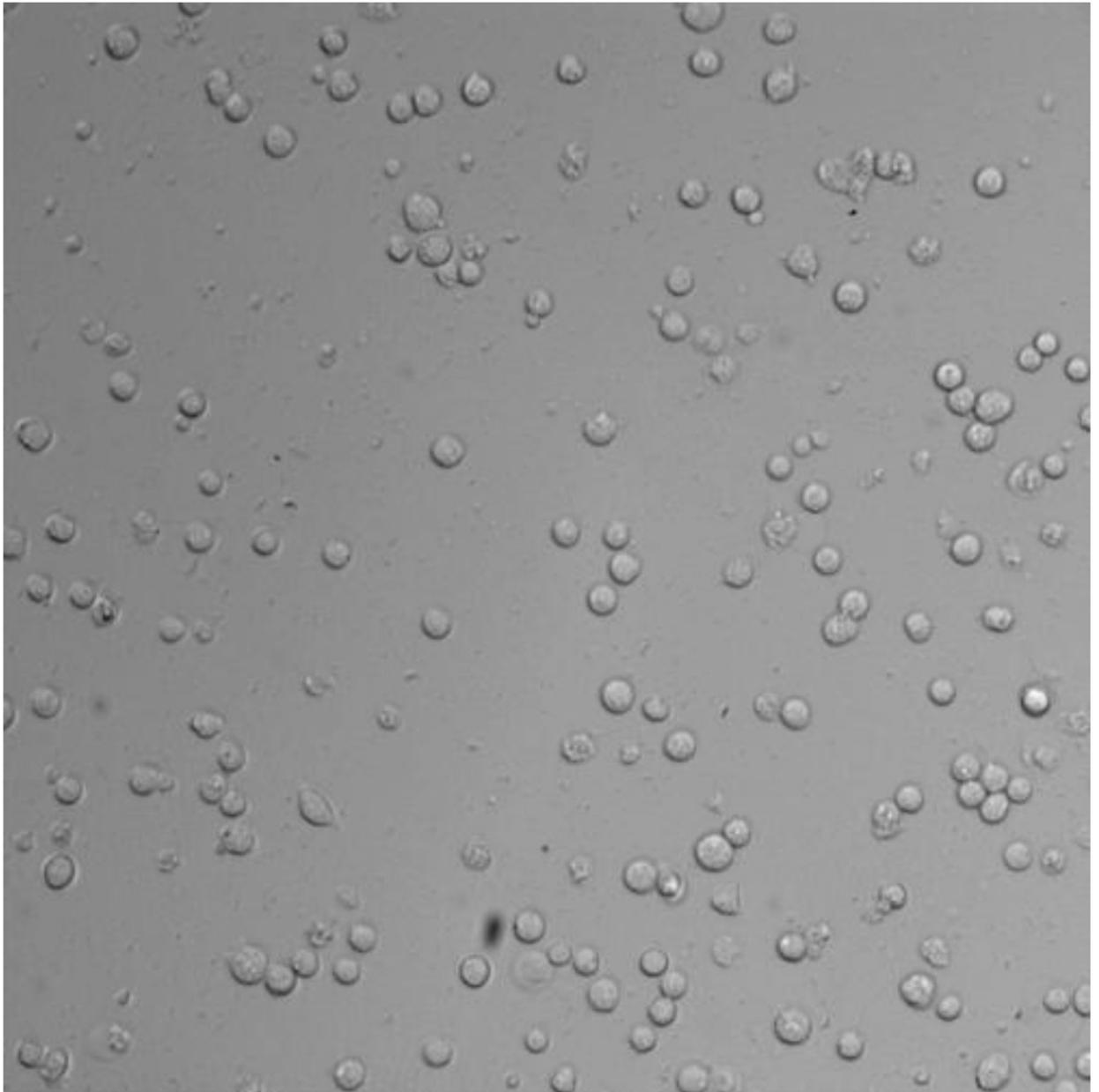
(a)
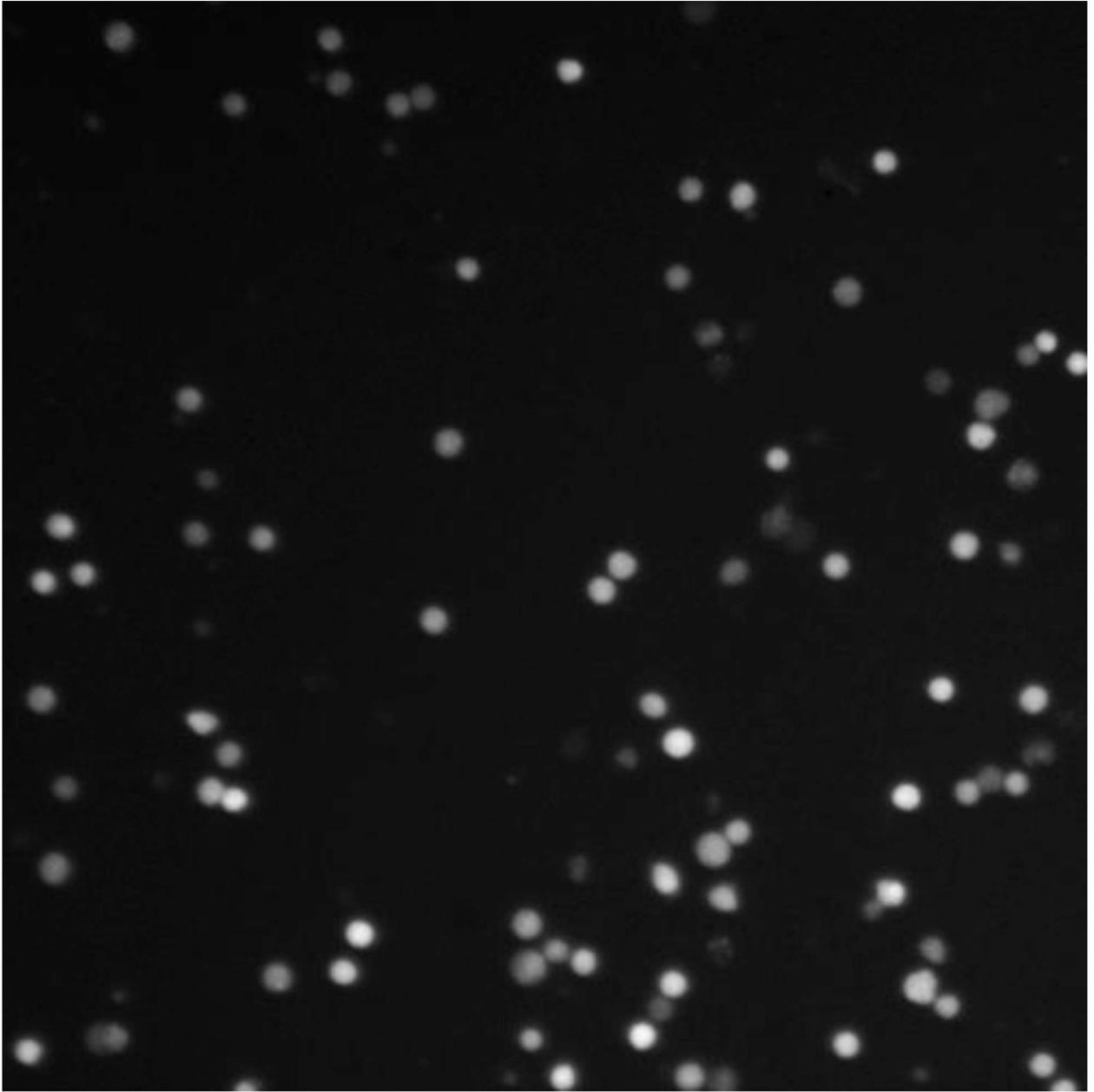
(b)

Fig. 6.
Typical sample images used in control experiments. (a) a brightfield image of mixture of stained and unstained A20.2J cells. (b) fluorescence image of the same field.
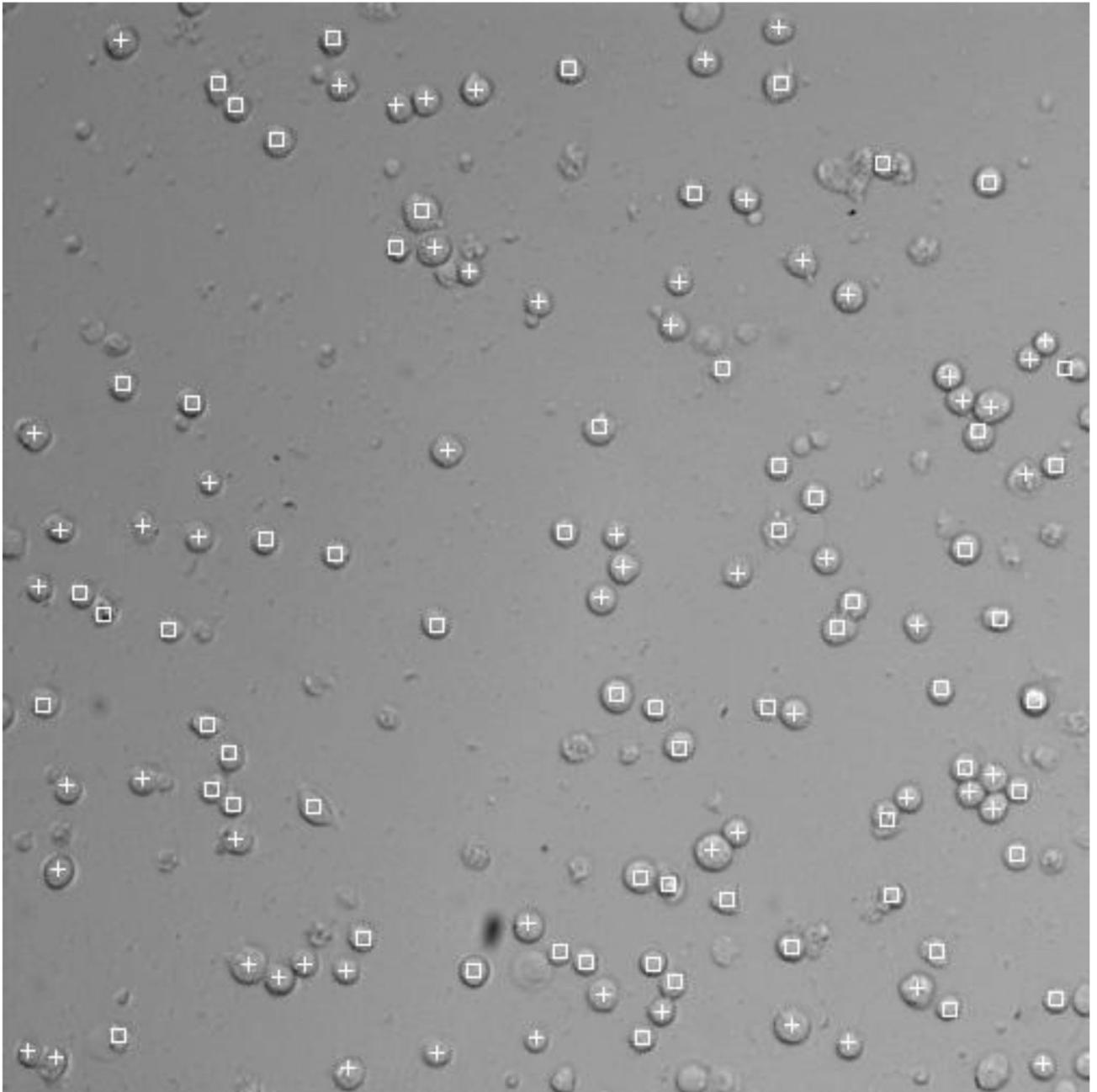
Fig. 7.
Classification results for the image shown in Fig. 6 (a). The cell positions detected are denoted by white symbols in the image. Squares: stained cells; Crosses: unstained cells.
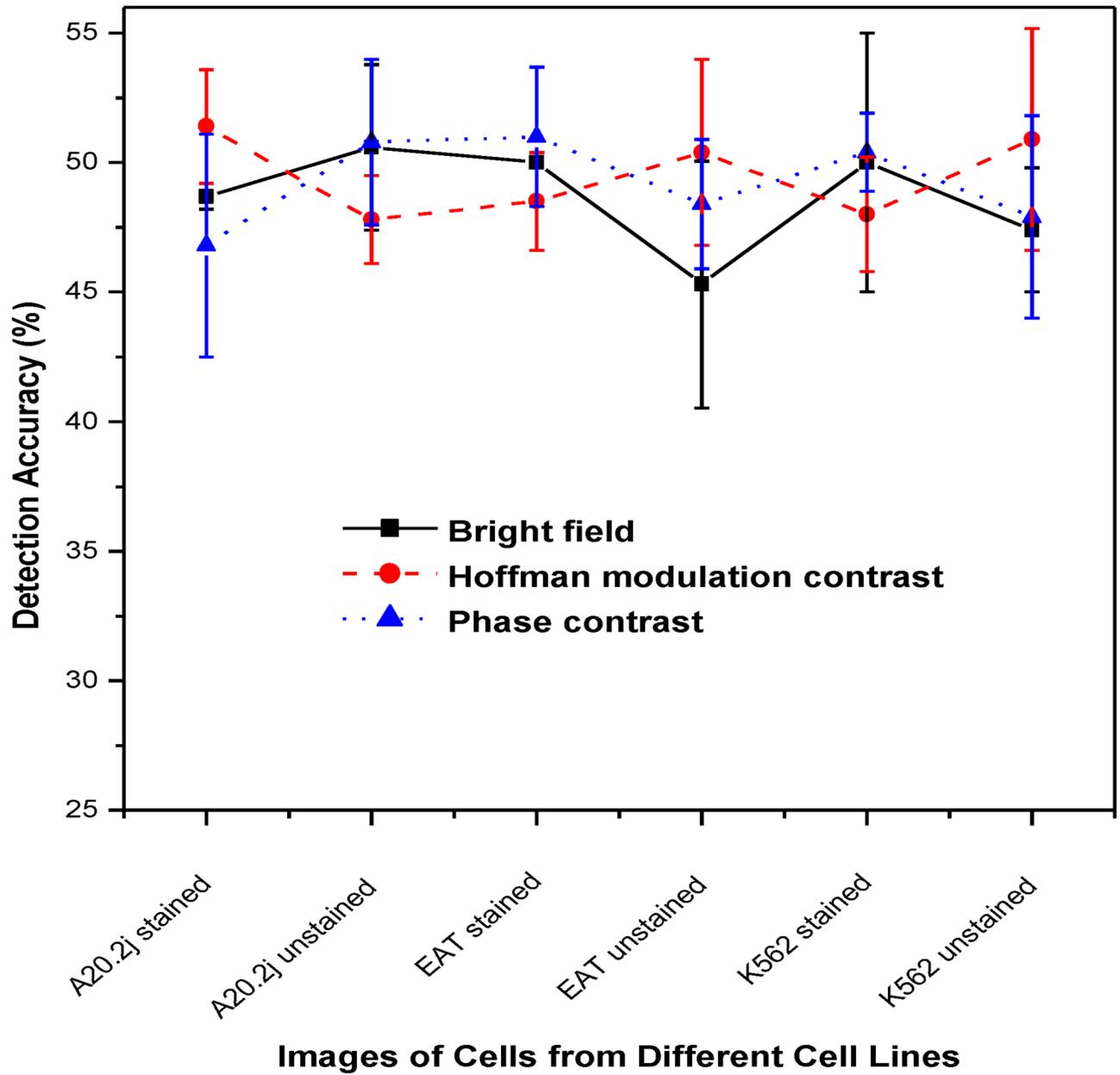
Fig. 8.
Detection accuracy of stained and unstained cells from different cell lines under different imaging channels.
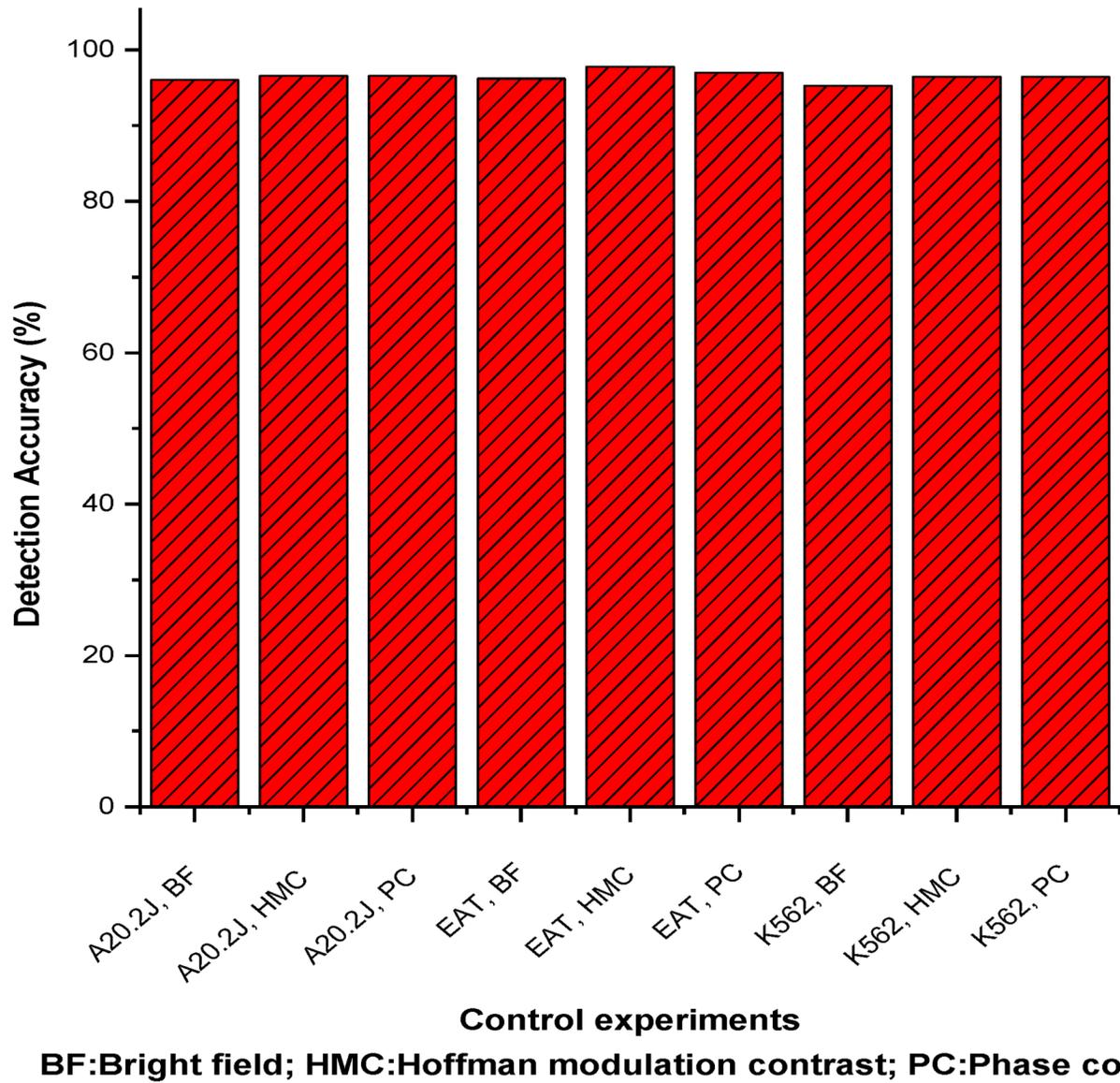
Fig. 9.
Detection accuracy of cell objects (stained & unstained) from different cell lines under different imaging channels.
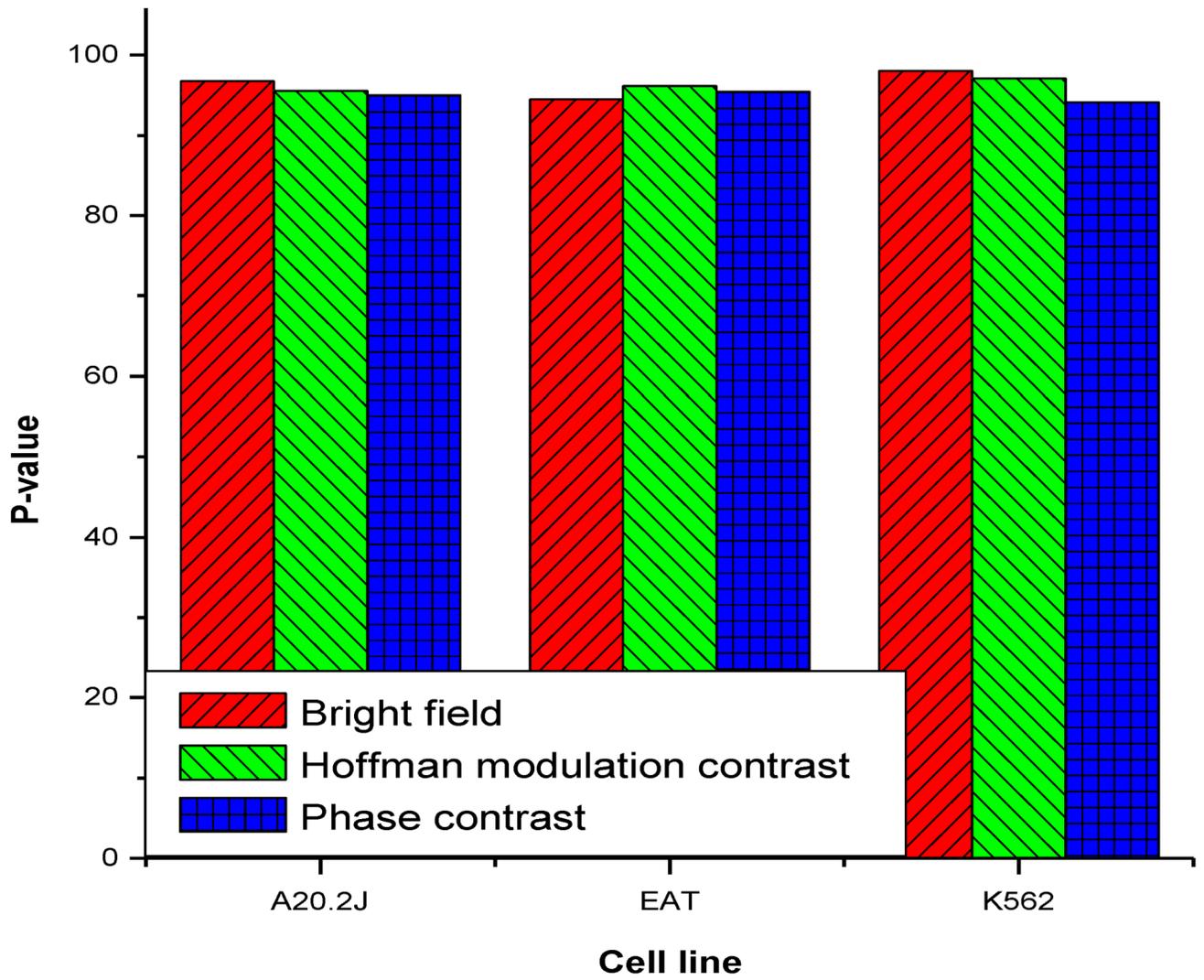
Fig. 10.
P-values for distinguishing between stained and unstained cells from different cell lines under different imaging channels.
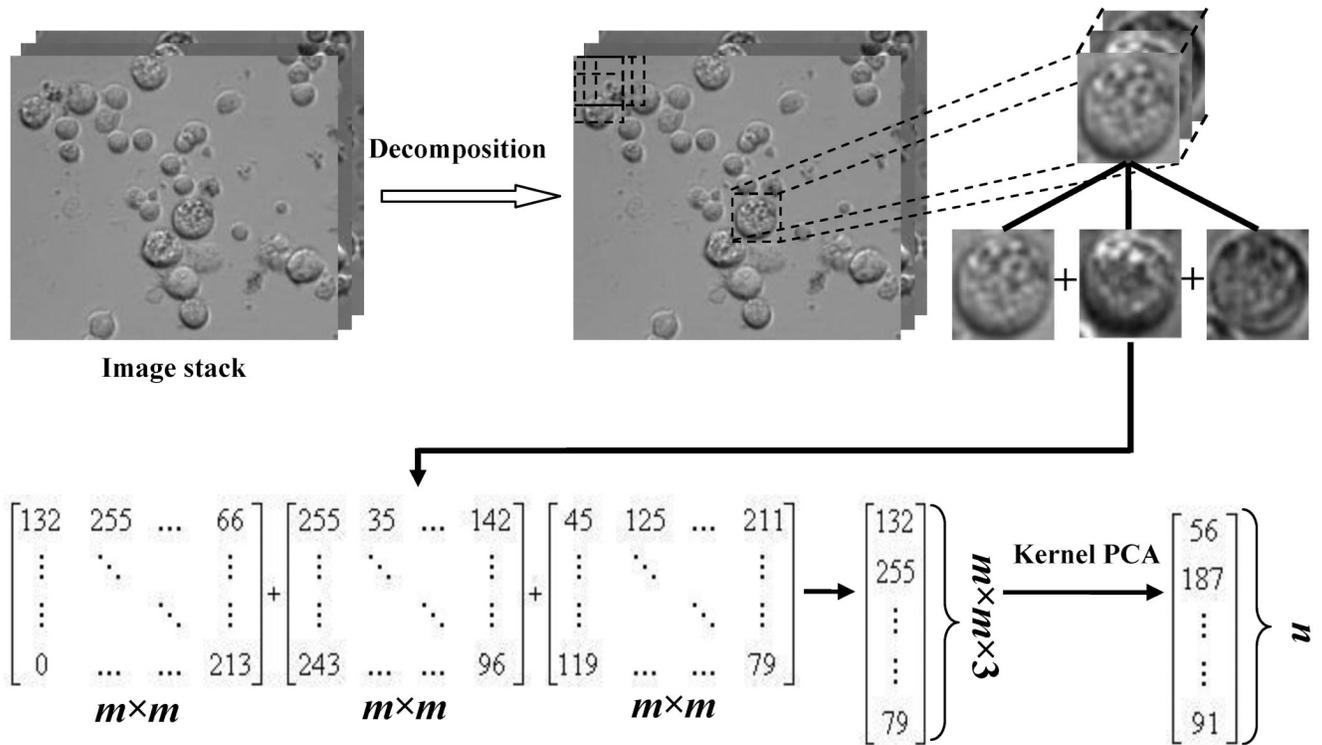
Fig. 11.
Illustration of pixel patch extraction and Kernel PCA preprocessing for cell detection in multicontrast composite images.
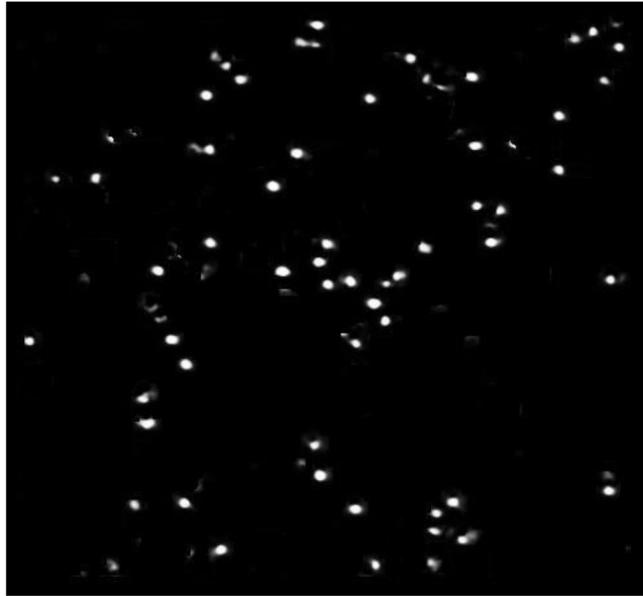
(a)
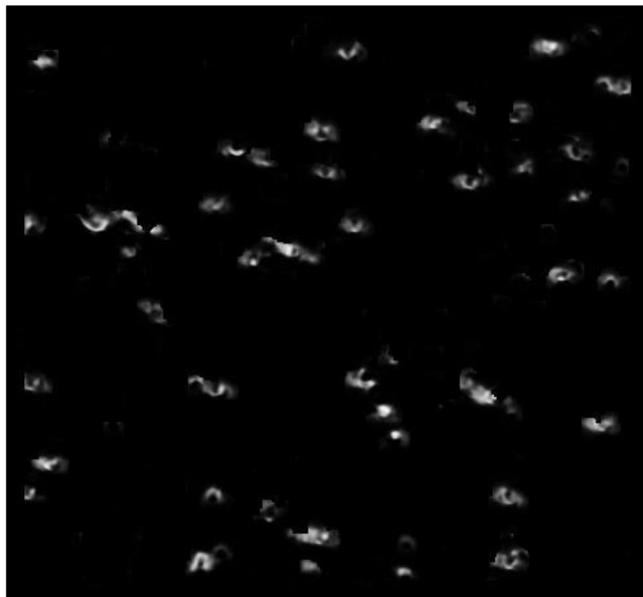
(b)

(c)

Fig. 12.
Confidence maps for images shown in Fig. 1 using Scenario 3-scheme with preprocessing by Kernel PCA. (a) confidence map for A20.2J cells; (b) confidence map for EAT cells; (c) confidence map for K562 cells. The confidence values are linearly scaled to 0~255 for display.
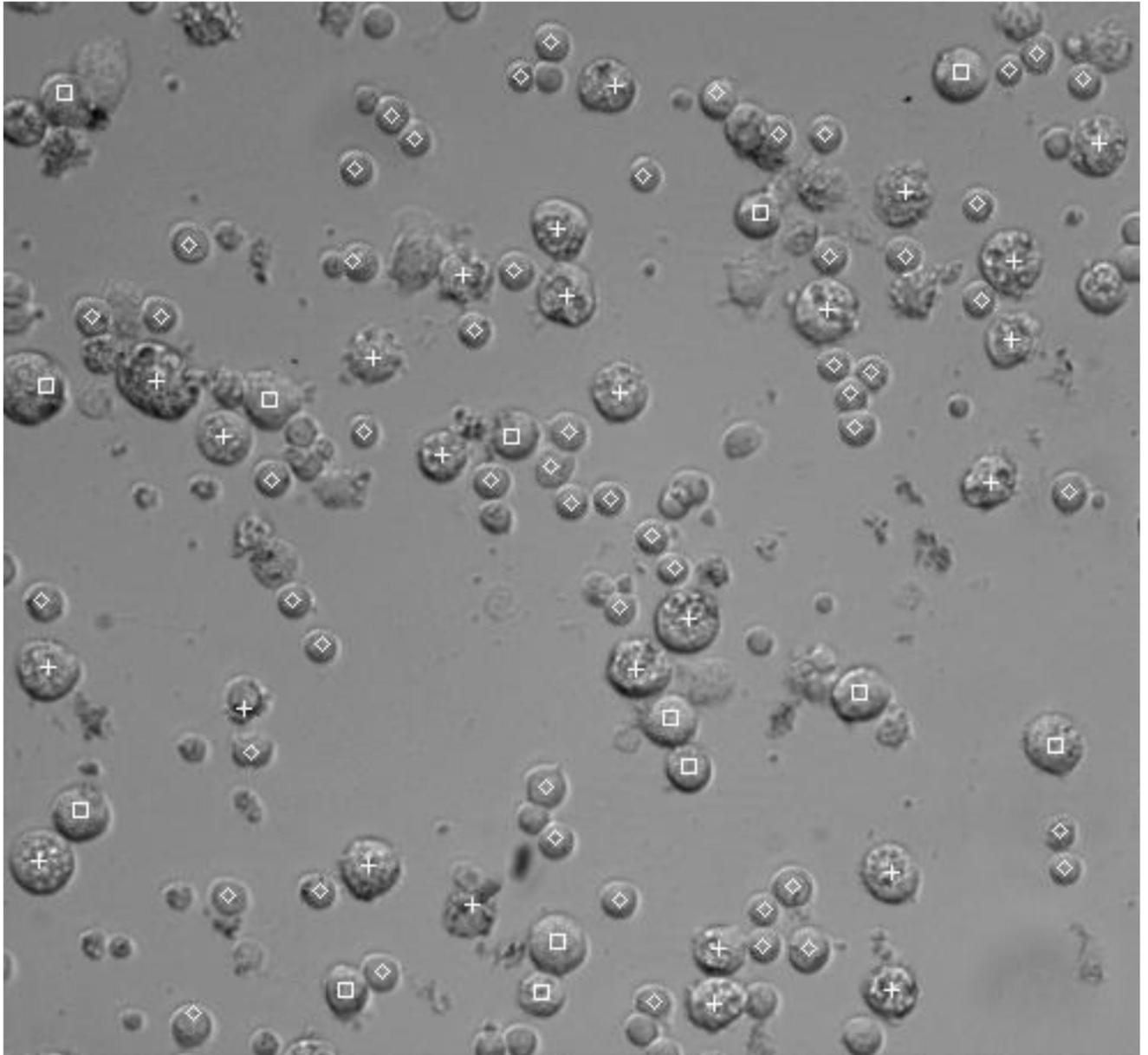
Fig. 13.
Classification results for the images shown in Fig. 1 using Scenario 3-scheme with preprocessing by PCA. The cell positions detected are denoted by white symbols in the image. Diamond: A20.2J cells; Square: EAT cells; Cross: K562 cells. Detection accuracy: A20.2J: 82.3%, EAT: 81.3 %, K562: 84% and Overall: 82.5%.
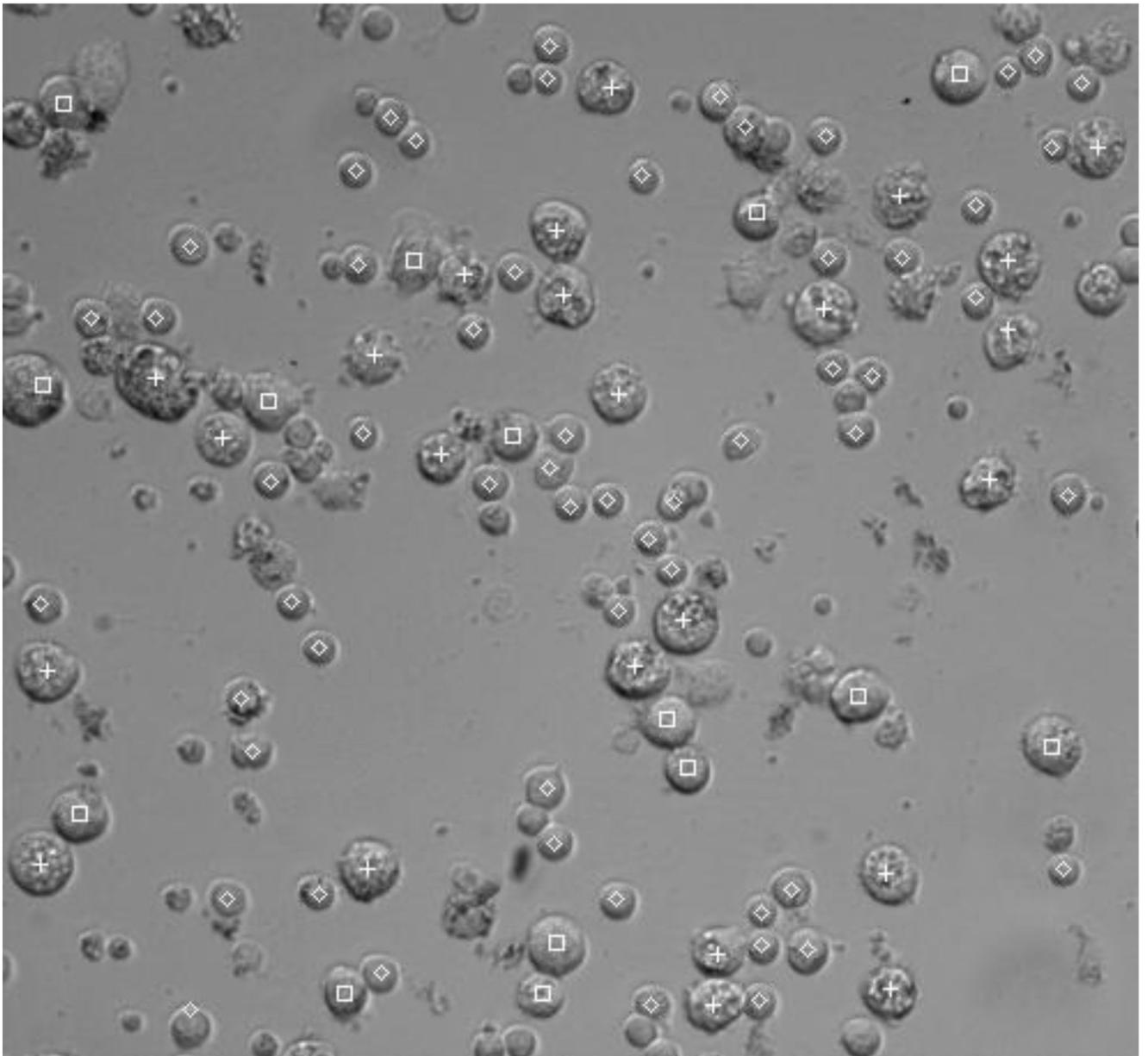
Fig. 14.
Classification results for the images shown in Fig. 1 using Scenario 3-scheme with
preprocessing by Kernel PCA. The cell positions detected are denoted by white symbols in the
image. Diamond: A20.2J cells; Square: EAT cells; Cross: K562 cells. Detection accuracy:
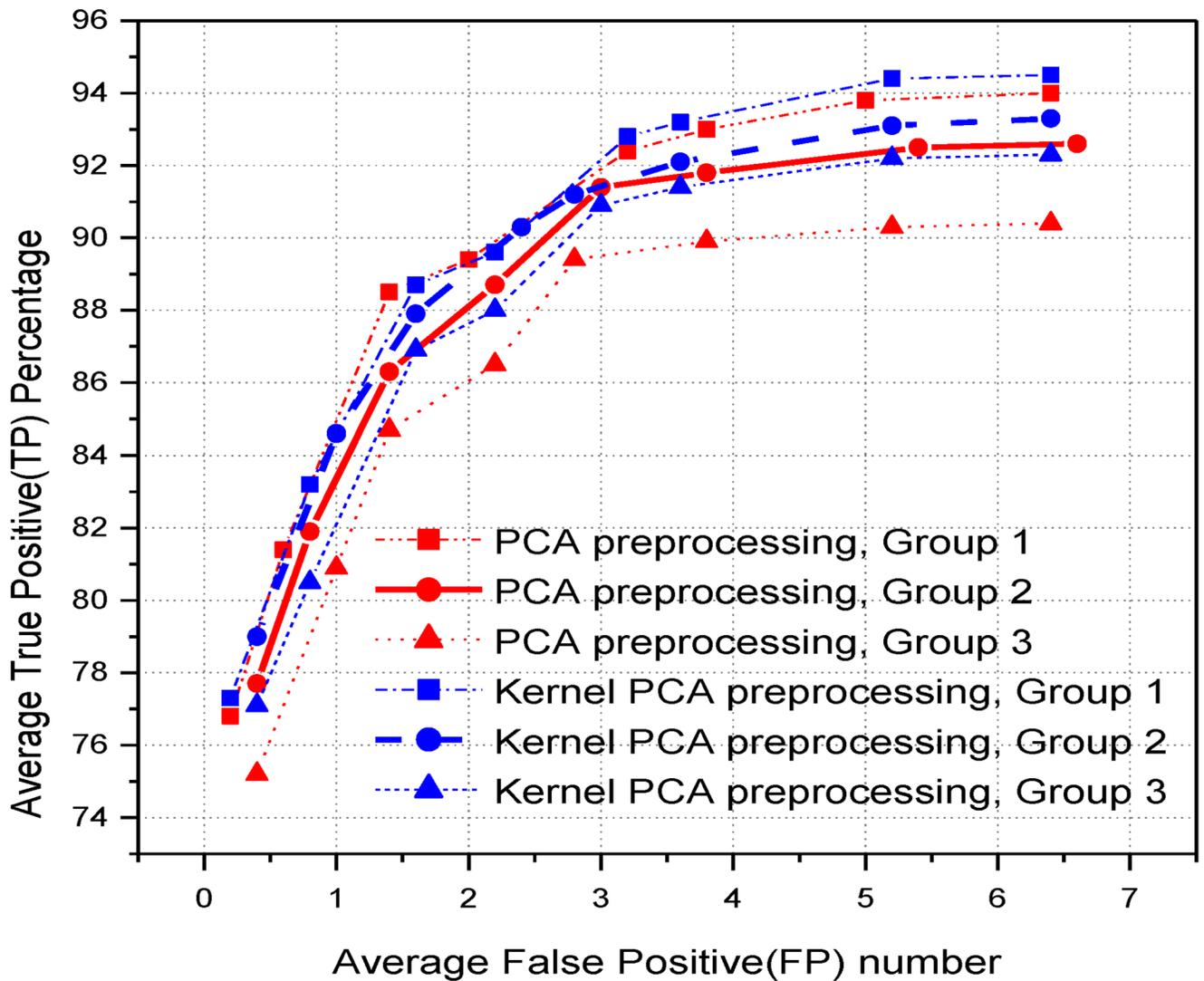A20.2J: 85.5%, EAT: 87.5 %, K562: 88% and Overall: 86.4%.

Fig. 15.
FROC plots of different preprocessing methods when applied to different image groups from Scenario 1: 1) PCA preprocessing; 2) Kernel PCA preprocessing. For both methods, the dimensionality was reduced to 10.
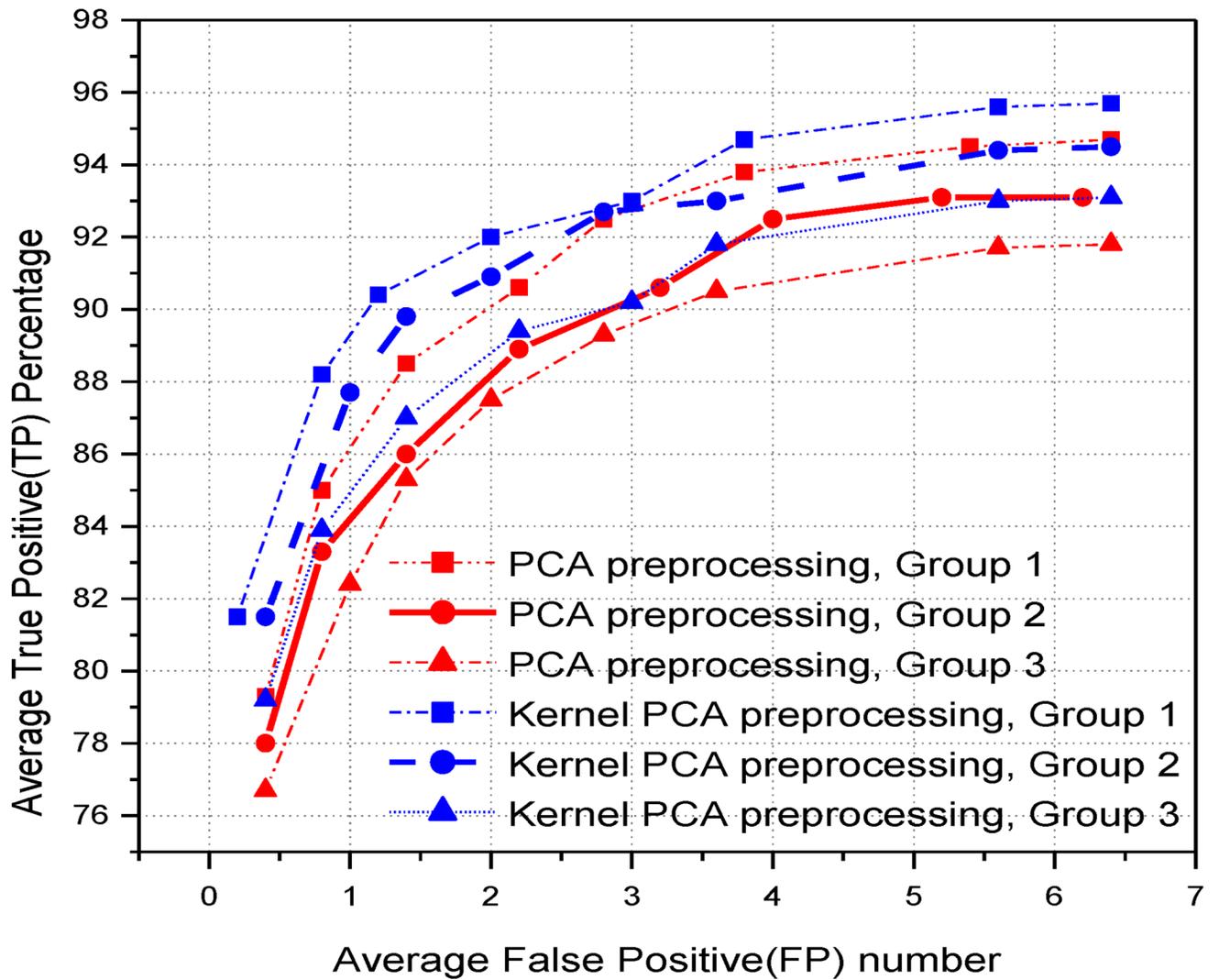
Fig. 16.
FROC plots of different preprocessing methods when applied to different image groups from Scenario 2: 1) PCA preprocessing; 2) Kernel PCA preprocessing. For both methods, the dimensionality was reduced to 10.
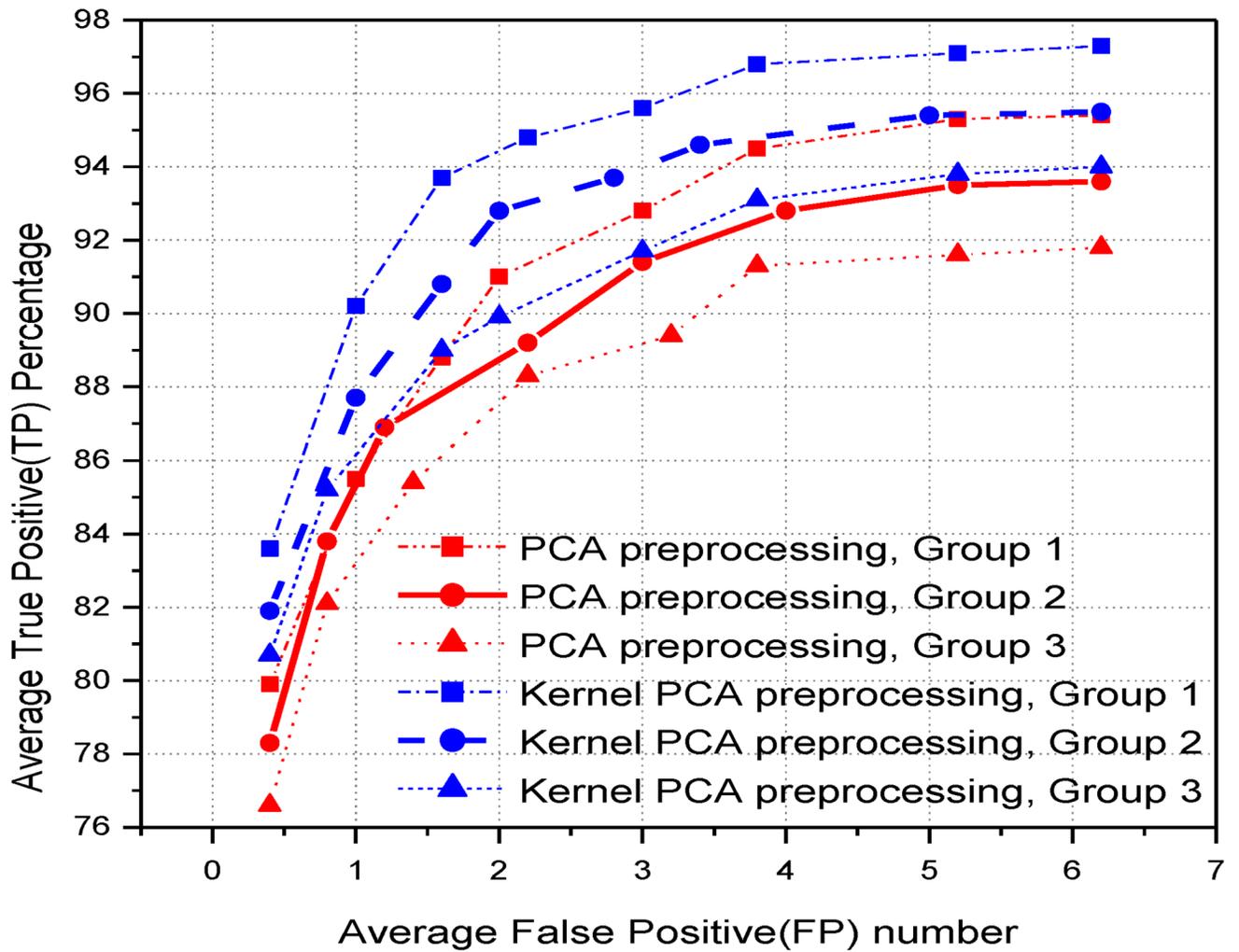
Fig. 17.
FROC plots of different preprocessing methods when applied to different image groups from Scenario 3: 1) PCA preprocessing; 2) Kernel PCA preprocessing. For both methods, the dimensionality was reduced to 10.