Automated Analysis of Mitotic Phenotypes in Fluorescence Microscopy Images of Human Cells

Nathalie Harder[1], Beate Neumann[2], Michael Held[2], Urban Liebel[2], Holger Erfle[2], Jan Ellenberg[2], Roland Eils[1] and Karl Rohr[1]

[1]German Cancer Research Centre (DKFZ), Dept. Theoretical Bioinformatics, and University of Heidelberg, IPMB, Dept. Bioinformatics and Functional Genomics, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
[2]European Molecular Biology Laboratory (EMBL), Gene Expression and Cell Biology/Biophysics Programmes, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
Email: n.harder@dkfz-heidelberg.de

Abstract. High-throughput screens of the gene function provide rapidly increasing amounts of data. In particular, the analysis of image data acquired in genome-wide cell phenotype screens constitutes a substantial bottleneck in the evaluation process and motivates the development of automated image analysis tools for large-scale experiments. Here we introduce a computational scheme to process multi-cell time-lapse images as they are produced in high-throughput screens. We describe an approach to automatically segment and classify cell nuclei into different mitotic phenotypes. This enables automated identification of cell cultures that show an abnormal mitotic behaviour. Our scheme proves a high classification accuracy, suggesting a promising future for automating the evaluation of high-throughput experiments.

1 Introduction

The technology of *RNA interference* (RNAi) is an effective method to identify the biological function of genes in the field of functional genomics. Together with the availability of complete genome sequences from several organisms, RNAi enables genome-wide high-throughput screening of gene function. In such screens all known genes of a considered organism are systematically silenced one after the other by inhibiting their expression, and then the resulting morphological changes are analysed. However, such large-scale knockdown screens produce an enormous amount of data which requires tools for automated image analysis.

Our work is carried out within the EU project *MitoCheck*, which aims to elucidate the coordination of mitotic processes in human cells at a molecular level. The goal is to obtain a better understanding of the mechanisms of cancer development. To identify the genes that are involved in cell division (mitosis), genome-wide high-throughput RNAi screens are performed. The effect of a silenced gene on mitosis is studied based on fluorescence microscopy time-lapse images of the treated cell culture. An automated evaluation of the resulting image sequences requires classification and quantification of different mitotic patterns.

Previous work on automated analysis of cell images has been done in different application fields. Based on fluorescence microscopy complete cells as well as single subcellular structures have been studied. Classification of complete cells has been performed, for example, to investigate the influence of drugs on cellular proteins [1]. There, cells are classified based on morphological characteristics of the plasma membrane that depend on the level of protein activation. Another application of complete cell classification has been considered in [2], where the detection of mitotic cells in images from automated microscope systems has been studied. Furthermore, work has been done on automated recognition of subcellular structures, which is a major task in location proteomics (e.g., [3, 4, 5]). In this field, the subcellular location of proteins is investigated in order to understand their function. Automated analysis of cell images also plays an increasing role in cytopathology, where computational methods have been developed to segment and classify different cell types in brightfield microscopy images of cell smears for early cancer detection [6].

We have developed an approach to analyse multi-cell images from large-scale RNAi screens. In comparison to previous work we analyse cell array images that include a large number of cell nuclei in different mitotic stages. We classify the nuclei into four different phases of the cell life cycle (*interphase, mitosis, apoptosis*, and *shape*). This enables an evaluation of the mitotic behaviour of a considered cell culture over time. Our automated image analysis work flow covers three main steps: Segmentation of multi-cell images, image feature extraction, and classification. In order to find the most appropriate algorithm for fast and accurate segmentation, we compared three different thresholding techniques according to their segmentation quality and computation time. Based on a quantitative evaluation we found that a region adaptive thresholding approach yields the best results. Using this approach we segment the images, compute a large set of different types of image features, and then apply a Support Vector Machine classifier. Our approach has been successfully applied using multi-cell images from genome-wide high-throughput screens. We obtained a high classification accuracy of 96% that has been verified by ten-fold cross-validation.

Our approach allows to minimize the manual evaluation effort. Furthermore, the results of an automated evaluation are objective and reproducible.

2 Methods

Since we have to cope with a huge amount of multi-cell images, fast and reliable segmentation and labeling of single objects is crucial. Various advanced segmentation algorithms have been described in the literature, but as computation time plays an important role when dealing with huge amounts of data, segmentation accuracy as well as speed of the algorithm are decisive criteria. We investigated three different thresholding techniques and evaluated the segmentation accuracy by manually counting correctly and incorrectly segmented cells. This manual evaluation has been performed for each tested algorithm for four different images that included in total 761 cell nuclei. The first technique

is a global thresholding scheme, where the threshold is automatically calculated using Otsu's threshold selection method [7]. However, a global threshold did not yield satisfying results, even after applying background correction, due to overlapping grey values of background and objects. The evaluation of the segmentation accuracy resulted in only 55.9% correctly segmented cell nuclei.

The second and the third segmentation technique are two versions of an adaptive thresholding algorithm. This algorithm uses a quadratic sliding window to calculate local thresholds for different regions of an image. A local threshold is only calculated if the variance within the window reaches a user-defined threshold, else a global threshold is used [8]. This ensures that only for regions that contain a certain amount of information (e.g., regions including an object and background) the local threshold is calculated, which reduces the computation time. Global as well as local grey value thresholds are automatically calculated using Otsu's threshold selection method. In the first version of the algorithm we apply the grey value threshold only to the central pixel of the sliding window and shift the window by one pixel, which leads to a high segmentation accuracy of 98.0%. Since here the variance calculation has to be carried out for each pixel of the image, the computation time is rather high. In the second version of the algorithm, the threshold is applied to every pixel of the sliding window and the window is shifted by one window width. Thus, the number of calculations for determining the variance and the local threshold is decreased enormously. In our tests this strategy reduced the computation time for the segmentation by a factor of about 330 and still yielded 92.1% correctly segmented nuclei. Here, we used a window width in accordance with the average nucleus diameter (e.g., 30 pixels). Since the speed of the segmentation algorithm is a major criterion in our application and an accuracy of 92.1% in this context is acceptable, we chose to adopt the second version.

For each segmented and labeled cell nucleus we calculate a set of image features that has been previously used for the classification of subcellular phenotypes [5]. This set includes granularity features, object- and edge-related features, tree-structured wavelet features, Haralick texture features, grey scale invariants, and Zernike moments. In total we compute 353 features per cell object. For the training set, we standardize each feature w.r.t. a mean value of zero and a standard deviation of one. In the test set, the feature values are linearly transformed based on the transformation parameters from the training set.

To classify cell nuclei into the four classes *interphase*, *mitosis*, *apoptosis*, and *shape*, we apply Support Vector Machines with a Radial Basis Function (RBF) as kernel function. SVMs are mathematically well-founded and their complexity is independent of the dimension in feature space compared to other classification methods. This property allows us to work with a high number of features and we can skip the feature selection step as it is not crucial. We solve the multi-class classification problem with a "one-against-one"approach. To optimize the penalty parameter C and the kernel parameter γ for the Radial Basis Function, we perform a three-fold cross-validation with varying values of C and γ on the training set (model selection) prior to the actual training of the classifier.

Table 1. Confusion matrix, 510 training and 127 test samples; accuracy: 96.9%.

	Classifier Output			
True Class	Interphase	Mitosis	Apoptosis	Shape
Interphase	**100**	0	0	0
Mitosis	0	**91**	9	0
Apoptosis	0	17	**83**	0
Shape	0	0	0	**100**

3 Experimental Results

We have applied our approach using a set of fluorescence microscopy multi-cell images, considering in total 637 cell nuclei. The images have a grey value depth of 12 bit and a resolution of 1344 × 1024 pixels. All nuclei had been classified by biologists into one of the four phases of the cell life cycle: (1) nuclei in the growing and resting phase (*interphase*), (2) mitotic nuclei (*mitosis*), (3) nuclei performing programmed cell death (*apoptosis*), and (4) a class which covers cases of clustered nuclei (*shape*). Using our region adaptive thresholding approach all nuclei were segmented automatically and the image features were calculated. We split the available samples for each class randomly in the training data and test data (ratio 4:1), resulting in a training set size of 510 nuclei and a test set size of 127 nuclei. The feature values were standardized and a Support Vector Machine classifier was trained as described above. An evaluation of the experimental results yielded an overall classification accuracy of 96.9%.

Since our test set was relatively large, the result obtained from this classification can already be considered to be significant. In order to check the reliability of the result we repeated the classification step, applying a ten-fold outer cross-validation on the whole data set of 637 cell nuclei. This classification yielded an average accuracy of 96.0%. Thus, both classification results correspond very well and we can draw the conclusion that we can rely on an overall classification accuracy of around 96%.

We also determined the confusion matrices (see Tables 1 and 2) which revealed that misclassifications appeared mainly between the classes mitosis and apoptosis. This is due to the fact that samples of these classes are sometimes very similar. Even for a human observer it can be hard to distinguish these two classes based on a single image. The lower classification accuracy of the *apoptosis* class can be ascribed to the comparatively small number of 60 samples in this class compared to the other classes (*interphase*: 322, *mitosis*: 110, *shape*: 145).

4 Discussion

We have presented an approach for automated cell phenotype analysis which can handle multi-cell images and classifies the segmented cell nuclei with a high

Table 2. Averaged confusion matrix for ten-fold cross validation given 637 samples; average accuracy: 96.0%.

	Classifier Output			
True Class	Interphase	Mitosis	Apoptosis	Shape
Interphase	**99.38**	0.31	0	0.31
Mitosis	0	**90.91**	9.09	0
Apoptosis	1.67	20.00	**78.33**	0
Shape	0	0	0	**100**

accuracy into four classes. By comparing the percentage of cells per class over time for knockdown RNAi experiments with those from control experiments, it is possible to automatically detect if mitosis is affected. This then indicates that the silenced gene plays an important role in the process of mitosis. As certain patterns are hard to distinguish based on single images, the exploitation of temporal information by tracking of cells throughout an image sequence is one possibility to further improve the classification performance.

Acknowledgement. This work has been supported by the EU project MitoCheck.

References

1. Lindblad J, Wählby C, Bengtsson E, Zaltsman A. Image Analysis for Automatic Segmentation of Cytoplasms and Classification of Rac1 Activation. Cytometry Part A 2003;57A:22–33.
2. Gallardo G, Yang F, Ianzini F, Mackey MA, Sonka M. Mitotic Cell Recognition with Hidden Markov Models. In: RL Galloway Jr, editor. Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display, Proc SPIE. vol. 5367; 2004. p. 661–668.
3. Boland MV, Murphy RF. A Neural Network Classifier capable of Recognizing the Patterns of all Major Subcellular Structures in Fluorescence Microscope Images of HeLa cells. Bioinformatics 2001;17(12):1213–1223.
4. Danckaert A, Gonzalez-Couto E, Bollondi L, Thompson N, Hayes B. Automated Recognition of Intracellular Organelles in Confocal Microscope Images. Traffic 2002;3:66–73.
5. Conrad C, Erfle H, Warnat P, Daigle N, Lörch T, Ellenberg J, et al. Automatic Identification of Subcellular Phenotypes on Human Cell Arrays. Genome Research 2004;14:1130–1136.
6. Würflinger T, Stockhausen J, Meyer-Ebrecht D, Böcking A. Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images. Computerized Medical Imaging and Graphics 2004;28:87–98.
7. Otsu N. A threshold selection method from grey level histograms. IEEE Transactions on Systems, Man and Cybernetics 1979;9:62–66.
8. Gonzalez RC, Woods RE. Digital Image Processing. 2nd ed. Prentice Hall; 2002.