A hybrid system for cell detection in digital micrographs

Tim W. Nattkemper
Applied Neuroinformatics Group,
Faculty of Technology
Bielefeld University
PO Box 100131, Bielefeld, Germany
email: tnattkem@techfak.uni-bielefeld.de

Walter Schubert
Meltec Ltd., Magdeburg, Germany
&
Institute of Medical Neurobiology
Magdeburg University, Germany

Thomas Hermann, Helge Ritter
Neuroinformatics Group,
Faculty of Technology
Bielefeld University, Germany

ABSTRACT

To analyze large sets of digital micrographs from high-throughput screening studies with constant accuracy, advanced image processing algorithms are necessary. In the literature, systems have been proposed applying model-based fitting algorithms, morphological operators and artificial neural networks (ANN). Because single approaches show limited performance, we propose a hybrid system that combines the Hough transform with a multi-layer perceptron (MLP) network. Our results show, that the combination of both approaches improves the performance and the positions of cell bodies are obtained with increased sensitivity and positive predictive value.

KEY WORDS

Molecular and Cellular Engineering, Optical Imaging, Fluorescence Microscopy, Medical Image Processing, Artificial Neural Networks

1 Introduction

In recent years, substantial progress could be observed for the application of optical microscopy in pharmaceutical research, particularly in molecular and cellular biology [1, 2]. Advances in molecular staining together with improved imaging technology now enables visualization of macromolecules in cells (e. g. [3, 4]). This paves the way to getting a broader view on cell functions [5] because in microscopy based approaches the information about *in vivo* spatial correlations in the molecular data is preserved, which is a feature not owned by many other methods (as e. g. gel-electrophoresis or microarrays). The demand for topological information is strongly motivated by the new established research field, called *systems biology* [6], where measurements of different cellular parameters, i. e. genome, transcriptome, proteome and metabolome, are integrated into cellular regulation models. In this work, images from immunofluorescence experiments, referred to as fluorescence micrographs are to be evaluated. There, macromolecules are visually proven by staining with fluorescent monoclonal antibody markers (mAb). This field is of special interest in the context of validation of target structures in drug discovery, where an advanced understanding of the functional and topological relationships of proteins is regarded as a prerequisite to future breakthroughs.

After image post-processing and registration, the data usually has to be analyzed by (a) direct manual exploration (generally by visual inspection) and/or (b) statistical datamining methods. To perform (b), quantitative data has to be extracted from the images. To this end, it is inevitable to perform an image segmentation interactively or full automatically. But the segmentation of micrographs has to deal with certain problems that are specific to this particular domain [7]. Those are inhomogeneous illumination, occlusion, considerable variation of shape, size and signal intensity to name a few.

Most of the work in biomedical image analysis is concerned with the segmentation and visualization of macrobiological structures [8]. Compared to this, approaches for the analysis of microbiological image domains are less published. One reason for this is the large diversity in the image domains of microbiological studies, which is caused by the complexity of cell functions, resulting from the overwhelming diversity of expressed molecules and their relationships.

In the past, algorithms for automatic micrograph analysis (or related tasks) had been proposed based on watershed algorithm [9], thresholding and morphological operators [10, 11], deformable models [12], artificial neural networks (ANN) [13], advanced version of the Hough transform [14, 15, 16]. The performance of the approaches is often evaluated using receiver operator characteristic (ROC) analysis, where complementary pairs of terms (for example *sensitivity* against *specifity*) are plotted [17, 18, 19].

The works above aim at the application, further development and evaluation of single approaches. Indeed, the different approaches seem to have their individual strength and weaknesses. Thus, for a first time we present a hybrid system that aims to combine the strengths of two approaches: the Hough transform and an artificial neural network (ANN) classifier. The Hough transform computes cell position candidates, which are post-processed by a trained multi-layer perceptron (MLP) network. From the counts of *true positive* (TP), *false positive* (FP) and *true negative* (TN) positions the acuuracy of the detection is as-

sessed in ROC terms of *sensitivity* (SE=TP/(TP+FN)) and *positive predictive value* (PPV=TP/(TP+FP)).

2 Material

Basis of our study is a set of fluorescence micrographs from a collaborative study. The micrographs show lymphocyte cells in a blood sample. The intensity values identify the presence of a molecule via immunofluorescence [20]. In the visual field, the cell bodies appear round, but have an irregular fluorescence signal. For a following statistical analysis, it is necessary to locate fluorescent cell bodies under the applied dye. In principle, many more protein signals than illustrated here can be captured in one biological sample by using new robotic imaging approaches. The images are of size 658×517 pixels, the average diameter of a cell is 10 pixels.

3 Methods

3.1 Hough transform based cell detection

The general idea of the Hough transform [21] is to define a mapping from the pre-processed feature space to a parameter space, the so called *accumulator*, where each pixel that belongs to the contour of an object in the image space is mapped to the same point in the accumulator. The more regular the object contour is, the more pixel are mapped to the same point in the array. The objects in the image are detected by post-processing local maxima in the accumulator.

Since the cell have circle-like shape we apply the Hough transform for circle detection. The accumulator is a three dimensional array (u, v, r) of circle origins (u, v) and radii $r = 10 \pm 3$. To realize the circle detection different strategies have been proposed [14, 15, 22, 23]. Our system employs the approach proposed by [22]. In the three dimensional array, all entries are incremented by one, that represent a possible cell center at (x, y) for different radii r (see Fig. 2 (a)). Each point is represented as a cone in the array and a circle with radius r' at image position (x', y') is represented by a local maximum in the array at $u = x', v = y', r = r'$.

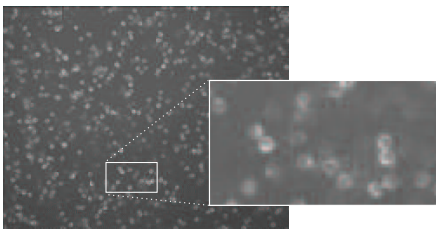

Figure 1. Example fluorescence micrograph of lymphocyte cells in a blood sample

To extract the cell positions from the computed accumulator, the radius cross sections are projected onto one plane (see Fig. 2 (b)): $\bar{F}(u, v) = \sum_r F(u, v, r)$. In this projection image, cell positions are represented by large local maximum values. Cell position candidates are computed applying a maximum filter followed by the application of a threshold t_h that eliminates false candidates caused by low local maxima.

3.2 MLP based cell classification

Artificial neural networks (ANN) are applied to image patch classification. A supervised learning algorithm is provided a training set $\Gamma_{\text{train}} = \{(\mathbf{v}^{(\lambda)}, y^{(\lambda)})\}_{\lambda=1,\ldots,n_t}$ of image regions $\mathbf{v}^{(\lambda)}$ (e. g. a 16×16-neighborhood) and class labels $y^{(\lambda)}$. The label is set to $y^{(\lambda)} = 1$ for regions $\mathbf{x}^{(\lambda)}$ that contain a particular biological structure (in this case a cell) and $y^{(\lambda)} = -1$ for those who do not. After training the ANN, the network realizes a mapping $C : \mathbf{v} \mapsto y$, i. e. a classification of image regions the degree of belief, that the region shows a fluorescent cell body. We apply the widely used multi-layer perceptron[1] (MLP) trained with the back-propagation rule [26]. To improve the performance of the MLP the training algorithm is modified. First, a regularization is added to the energy function to avoid large weight vectors, which is called *weight decay*. Second, we applied *resilient* back-propagation (RPROP) [27] to speed up the network learning, which is achieved by local adaption of the weight learn rates according to the present behavior of the error function. Best results are obtained with a MLP of 256 input neurons, 2 hidden layers of 32 and 8 neurons and 1 output neuron, trained in 150 epochs on the training set.

3.3 Hybrid system

The hybrid system for cell detection computes the positions of cells in two steps. In a first step a circle Hough transform with a low threshold value t_h computes candidate cell positions. The 16×16 sized image patches $\mathbf{v}_{(x,y)} \in [0; 1]^{256}$ at these positions are fed into the MLP classifier $C(\mathbf{v}_{(x,y)})$, with

$$C(\mathbf{v}_{(x,y)}) = \begin{cases} \text{fluorescent cell at } (x,y) & \text{if } C(\mathbf{v}_{(x,y)}) > t_o \\ \text{no fluorescent cell at } (x,y) & \text{else.} \end{cases} \quad (1)$$

4 Results

The system performance is evaluated by plots of ROC terms SE and PPV. Different variants of the system are ap-

[1] In recent years, kernel based methods have been object of much research effort and gained remarkable popularity in the field supervised learning. The most prominent algorithm among these is the *Support Vector Machine* (SVM) proposed by V.Vapnik [24] for binary classification. Nevertheless, SVMs are not considered here, because in the field of medical informatics, MLPs are still the most widely used supervised learning architecture [25] and is reported to reach performances comparable to the SVM.

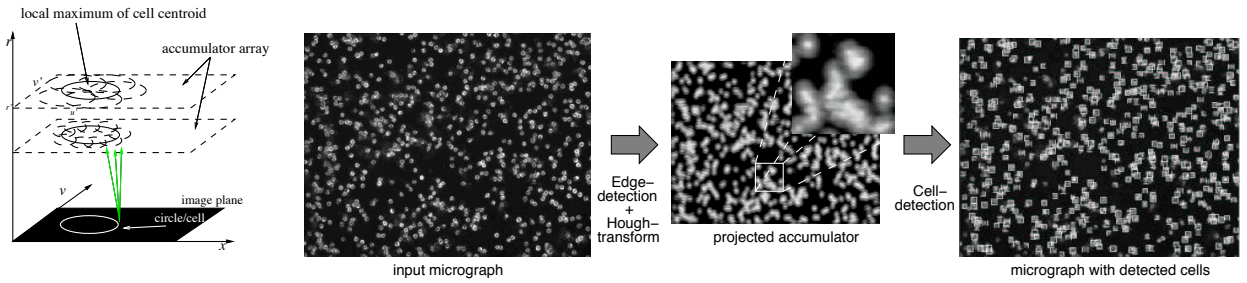

Figure 2. (a) Contour pixels are mapped to cells of potential centroid points in the accumulator. The circles in the image are represented by local maxima in the three dimensional array. (b) The accumulator of the micrograph (left) is computed based on a pre-processed feature map (an edge map computed using a Laplace filter), and projected onto one plane (middle). In the enlarged subregion one can see maximum accumulator values that identify the cells' centroids. A following post-processing in the projection plain results in a list of cell positions (right).

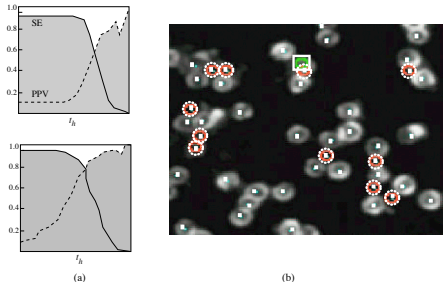

Figure 3. (a) Sensitivity (solid) and positive predictive value (dashed) for the Hough transform based cell detection are plotted for different accumulator maximum threshold t_h and different edge thresholds $t_e = 30$ (left) and $t_e = 60$ (right). (b) In an enlarged sub-region of the micrograph, the computed cellpositions are marked with white points. False positive cell positions are marked with a dashed circle, one false negative with a white square.

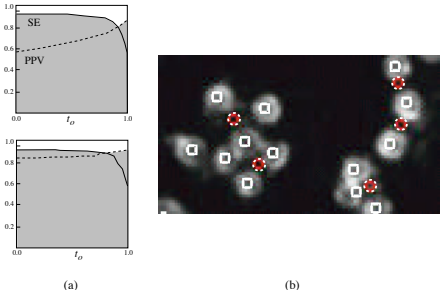

Figure 4. (a) The plots of sensitivity (solid) and positive predictive value (dashed) are shown for the hybrid cell detection system. A Hough transform with a large threshold value t_h (right plot) achieved better results than a less specific parameterized one (left plot). (b) The final cell position candidates are marked with white squares. False positive positions are eliminated by the MLP (dashed circles).

plied and tested. First, the cell positions are computed applying just the Hough transform (see Figure 3). We observe a strong growing number of FP cells for slightly decreasing the threshold value t_h, expressed by the steepness of the PPV decrease. A closer look at the positions of the FP reveals, that the Hough transform tends to produce FPs *between* cells (cf. 3 (b)). Since the cells use to be densely clustered in in-vivo samples a purely Hough transform based cell detection is not sufficient. Second, we apply the hybrid cell detection system to the data and compute SE/PPV plots for varying the threshold value t_o. The preceding Hough transform is applied with different values for t_h. The plots show the best system performance for larger t_h (see Figure 4 (a)). It can be shown, that the MLP eliminates FP positions considerably (see Figure 4 (b)). Especially FPs between cells are correctly rejected by the MLP.

5 Discussion

We presented a first application of a hybrid system of Hough transform and MLP to the problem of cell detection in fluorescence microscopy. It is shown, that post-processing the Hough transform result by a trained ANN classifier reduces the number of false positive positions considerably. The system, processes one micrograph in less than 1 minute on a standard PC and is easy to extend to new image data, condition to the regularity of the cells' shapes.

Recently, works have been published to learn Hough transform parameters for arbitrary shapes from a set of labeled data (e. g. [28]. Which is an interesting approach, although it contains some work load, because the algorithm has to be provided with some hand set landmarks in the images along the cell shapes. In contrast to this learning of a shape, our approach matches the image objects with a very generalized shape model, i. e. a circle. The lack of specificity is compensated by the application of an ANN, which is parameterized with high specificity. Besides the architecture of the MLP, the system contains only two critical parameters, that have to be tuned, which are the thresholds of the two modules t_h and t_o. From our point of view, this is more than acceptable, because both can be iteratively tuned full-automatically by evaluating SE and PPV.

Acknowledgments: This work was supported by the

BMBF grant BMBF-0312844. Special thanks go to Bernd Hafer, Christian Lange, Oliver Lichte, Wiebke Timm and Johanna Vompras for their contribitions to the implementation and evaluation of this system.

References

[1] D. L. Taylor, M. Nederlof, F. Lanni, and A. S. Waggoner. The new vision of light microscopy. *Amer. Sci.*, 1992.

[2] D. L. Taylor, E. S. Woo, and K. A. Giuliano. Real-time molecular and cellular analysis: the new frontier of drug discovery. *Curr. Op. Biotech.*, 2001.

[3] M. V. Boland. Automated classification of cellular protein localization patterns obtained via fluorescence microscopy. In *Proc. of IEEE EMBS*, 1997.

[4] M. P. Oksvold, E. Skarpen, J. Widerberg, and H. S. Huitfeldt. Fluorescent histochemical techniques for analysis of intracellular signaling. *J. of Histochem. & Cytochem.*, 2002.

[5] R. A. Wind. An integrated confocal and magnetic resonance microscope for cellular research. *J. Magn. Res.*, 2000.

[6] H. Kitano. Perspective in systems biology. *New Generation Computing*, 18:199–216, 2000.

[7] S. H. Ong, X. C. Jin, Jayasooriah, and R. Sinniah. Image analysis of tissue sections. *Comput. Biol. Med.*, 26:269–279, 1996.

[8] J. S. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Trans. PAMI*, 22(1):85–105, 2000.

[9] N. Malpica, C. O. de Solorzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. Garcia-Sagredo, and F. del Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, 28, 1997.

[10] D. Demandolx and J. Davoust. Multiparameter image cytometry: from confocal micrographs to subcellular fluorograms. *Bioimaging*, 4:159–169, 1997.

[11] A. Nedzved, S. Ablameyko, and I. Pitas. Morphological segmentation of histology cell images. In *Proc. of ICPR*, volume 1, 2000.

[12] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Med. Im. An.*, 1996.

[13] P.J. Sjöstrom, B.R. Frydel, and L.U. Wahlberg. Artificial neural network-aided image analysis system for cell counting. *Cytometry*, 36:18–26, 1999.

[14] R.O. Duda and P.E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15:11–15, 1972.

[15] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Patt. Rec.*, 13:111–121, 1981.

[16] P.R. Barber, B. Vojnovic, J. Kelly, C.R. Mayes, P. Boulton, M. Woodcock, and M.C. Joiner. Automated counting of mammalian cell colonies. *Phys Med Biol*, 2001.

[17] J.A. Swets. Measuring the acccuracy of diagnostic systems. *Science*, 240:1285–1293, 1988.

[18] H. M. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 1993.

[19] M. Greiner, D. Pfeiffer, and R. D. Smith. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.*, 45:23–41, 2000.

[20] W. Schubert. *Topological Proteomics, Toponomics, MELK-Technology*, chapter Proteomics of Microorganismus: Fundamental Aspects and Application, Advances in Biochemical Engineering/Biotechnology. 2003. In press.

[21] P.V.C. Hough. A method and means for recognizing complex pattern. US Pat. Appl. No. 3069654, 1962.

[22] G. Gerig and F. Klein. Fast contour identification through efficient Hough transform and simplified interpretation strategy. *Proc. of ICPR*, 8:498–500, 1986.

[23] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler. Comparative study of hough transform methods for circle finding. *Im. and Vis. Comp.*, 1990.

[24] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[25] I. Kononenko. Machine learning for med ical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23:89–109, 2001.

[26] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1986.

[27] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *Proc. of IEEE Int. Conf. on N. N.*, 1993.

[28] M. Brejl and M. Sonka. Object localization and border detection criteria design in edge based image segmentation: Automated learning from examples. *IEEE TMI*, 2000.