Unsupervised Unstained Cell Detection by SIFT Keypoint Clustering and Self-labeling Algorithm

Firas Mualla[1], Simon Schöll[1,3,4], Björn Sommerfeldt[2], Andreas Maier[1,3], Stefan Steidl[1], Rainer Buchholz[2], and Joachim Hornegger[1,3]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Institute of Bioprocess Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg
[3] SAOT Graduate School in Advanced Optical Technologies
[4] ASTRUM IT GmbH

Abstract. We propose a novel unstained cell detection algorithm based on unsupervised learning. The algorithm utilizes the scale invariant feature transform (SIFT), a self-labeling algorithm, and two clustering steps in order to achieve high performance in terms of time and detection accuracy. Unstained cell imaging is dominated by phase contrast and bright field microscopy. Therefore, the algorithm was assessed on images acquired using these two modalities. Five cell lines having in total 37 images and 7250 cells were considered for the evaluation: CHO, L929, Sf21, HeLa, and Bovine cells. The obtained F-measures were between 85.1 and 89.5. Compared to the state-of-the-art, the algorithm achieves very close F-measure to the supervised approaches in much less time.

1 Introduction

Cell detection plays a vital role in biomedical image analysis. Automatic image-based cell detection approaches can be used for estimating the number of cells [1,2], initializing cell segmentation algorithms [3], cell tracking [4], and for extracting features which can be used for further analysis such as cell viability determination [5].

In fluorescence microscopy, cells are stained using a fluorescent dye. This reshapes the cell detection problem as a relatively easy task due to the high contrast obtained by staining. On the other hand, in some biological applications [6], it is desired to avoid staining because of its side effects on cells. In this case, cell detection is more challenging and sometimes very difficult [7,8,9].

We know from bioprocess engineers that, for unstained cell imaging, bright field and phase contrast are the most widely used microscopic modalities. Therefore, they form together a very appropriate choice for the evaluation of unstained cell detection.

Several machine learning approaches have been proposed in the literature in order to cope with the difficulty of the problem. Some approaches follow a pixel-wise classification strategy [5,8,10,11]. Others perform the classification at the level of image interest points [9,12,13,14]. The latter have some useful properties. First, the problem is sparse compared to the pixel-wise classification. Second, the interest points are characterized by features and/or descriptors which can be utilized for detection as in [9, 12]. Third, they can be employed to achieve scale- and orientation-invariant training as in [9].

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 377–384, 2014.

All the previous approaches are dependent on supervised learning. The latter transfers part of its inductive bias to the training data which makes the approach adaptable by simply changing the training set. This has the advantage that it can model very complicated situations and provide reliable results as long as the training set is representative. On the other hand, its drawback is that it requires labeled ground truth. In many cases, the users of cell image analysis software would sacrifice some detection accuracy in favor of having a labeling-free system. This preference becomes more serious when the system has to be trained for each new cell line.

In this paper, we advocate an unsupervised machine learning approach for unstained cell detection. Technically speaking, we also employ supervised learning, but with ground truth learned automatically from the input image. The proposed approach was tested on five cell lines with diverse visual appearance. Our results show that we are very close in terms of detection rate to the state-of-the-art supervised learning approaches. However, our approach has a much faster runtime and does not require manually-labeled ground truth.

Related Work. In [12] and [14] on phase contrast microscopy and [9] on bright field microscopy, the training was done at two levels: First, cells and background are separated by machine learning techniques. Second, the difference between interest points belonging to the same cell and interest points belonging to neighboring cells is learned. The support vector machine (SVM) classifier and interest points detected by a set of Laplacian filters were utilized in [12] and [14] while the random forest classifier and the scale invariant feature transform (SIFT) keypoints were used in [9]. The previous approaches require ground truth of segmented cells. In other words, cell borders should be delineated and each cell should have a distinguishing identifier in the ground-truth mask. In [13], maximally stable extremal regions (MSER) keypoints were utilized and a structured SVM was used to learn a bijective mapping between the MSER regions and the ground-truth cell centers. Compared to [9], [12], and [14], this approach has the advantage that it is easier to train because only cell centers are required as ground truth. The closest to our approach is [9], but ours uses unsupervised learning and it is thus labeling-free.

2 Methods

We make a heavy use of SIFT related concepts. Therefore, we introduce SIFT in Section 2.1 and we then describe our method in Section 2.2.

2.1 SIFT

SIFT [15] is a local image feature detector and descriptor. Each detected keypoint is characterized by its spatial coordinates, a scale, an orientation, a difference of Gaussians (DOG) value, and a principal curvatures ratio (PCR) value. The DOG value indicates the keypoint strength and its range is proportional to the dynamic image range. Its sign is positive for black-on-white blobs and negative for white-on-black blobs. The PCR value is defined as [15]:

$$\mathrm{PCR} = \frac{\mathrm{Tr}(\mathbf{H}_{DOG})^2}{\mathrm{Det}(\mathbf{H}_{DOG})} \tag{1}$$

where $\mathbf{H}_{DOG}$ is the Hessian of DOG, Tr is the trace, and Det is the determinant. PCR has a minimum of 4 for the isotropic blobs and its value increases theoretically until $+\infty$ by increased blob anisotropy.

2.2 Cell Detection by Keypoint Clustering and Self-labeling

2.2.1 Keypoint Extraction

Our algorithm starts by extracting SIFT keypoints of the input image $\mathbf{I}$. These keypoints are not thresholded using the PCR or the DOG values. In other words, all detected SIFT keypoints of all strength and anisotropy values are considered at this step.

2.2.2 Blob Type Detection

The keypoint blob type is determined by the DOG sign. As mentioned in Section 2.1, it is either black on white (+1), or white on black (-1). Based on [9], the blob type is computed using the following equation:

$$\beta = \text{sign}\left(\frac{\sum_{i=1}^{N} \omega_i \left|DOG(\mathbf{p}_i)\right| H(DOG(\mathbf{p}_i))}{\sum_{i=1}^{N} \omega_i \left|DOG(\mathbf{p}_i)\right|} - \frac{1}{2}\right) \tag{2}$$

$$\omega_i = \frac{s(\mathbf{p}_i)}{\text{PCR}(\mathbf{p}_i)} \tag{3}$$

where $\mathbf{p}_i$, $i = 1..N$ are the keypoints in the image as obtained in Step 2.2.1, N is their number, $s(\mathbf{p}_i)$ is the scale of $\mathbf{p}_i$, and H is the Heaviside step function. If $\beta = +1$, the positive DOG keypoints are considered while the negative DOG keypoints are discarded, and vice versa.

2.2.3 Scale Adaptive Smoothing

The image $\mathbf{I}$ is smoothed with a Gaussian kernel whose standard deviation is the mean keypoint scale. The latter is computed using the following equation:

$$\sigma = \frac{\sum_{i=1}^{M} \left|DOG(\mathbf{p}_i)\right| s(\mathbf{p}_i)}{\sum_{i=1}^{M} \left|DOG(\mathbf{p}_i)\right|} \tag{4}$$

where M is the number of the keypoints resulting from Step 2.2.2, i.e. only one blob type is considered.The smoothed image $\mathbf{I}_\sigma$ is saved for further processing.

2.2.4 Second Keypoint Extraction

Step 2.2.1 is applied on the smoothed image $\mathbf{I}_\sigma$ and the keypoints which conform to the previously computed β are considered while the others are discarded.

2.2.5 Cell/Background Keypoint Clustering

At this step, the keypoints are clustered into one of two categories: cells and background. K-medians clustering, i.e. intra-cluster $\ell 1$-norm minimization, is applied with $K = 2$. One-dimensional Otsu thresholding is applied on the DOG values of the keypoints and the two resulting clusters are used to initialize the Lloyd's iteration. The features are modality-specific. For bright field microscopy, at each keypoint $\mathbf{p}_i$, we employ $\text{DOG}(\mathbf{p}_i)$ and smoothed image intensity $\mathbf{I}_\sigma(\mathbf{p}_i)$ as features. For phase contrast

microscopy, we use $\mathrm{DOG}(\mathbf{p}_i)$ and $\mathrm{VAR}(\mathbf{I}, \mathbf{p}_i, \sigma)$. The latter is the local variance of the original image $\mathbf{I}$ within a square neighborhood centered at $\mathbf{p}_i$ with a half side-length equal to σ (up to an integer approximation). The features are normalized to $[0, 1]$ so that they contribute equally to the $\ell 1$-norm. After termination, the keypoints which belong to the background cluster are discarded.

2.2.6 Cell/cell Keypoint Clustering

The goal of this step is to cluster the cell keypoints resulting from the previous step into N_c clusters where two keypoints belong to the same cluster if and only if they belong to the same cell. N_c is not known a priori. In order to achieve this goal, a classifier which ranks each pair of keypoints as belonging to the same cell or not is required [9], [14]. We propose to learn this classifier from the input image using a self-labeling algorithm instead of manually-labeled ground truth. Informally speaking, the algorithm trains a keypoint-pair classifier on *extreme* cases (for which ground truth labels can be assumed) and applies the resulting classifier on *intermediate* cases. This is achieved as follows:

1) Consider Ψ to be a set of keypoint pairs defined as: $(\mathbf{p}_i, \mathbf{p}_j) \in \Psi \Leftrightarrow \|\mathbf{p}_i - \mathbf{p}_j\|_2 \geq \rho$ where $\rho = \alpha \cdot \sigma$ and α is a constant. ρ must be larger than the maximum cell length. Due to the use of SIFT, *safe* values for α can be set easily regardless of the image resolution or cell type. We set it to 10 in our experiments.
2) Randomly choose N_1 elements, i.e. keypoint pairs, from Ψ. Label each of them as *cross* which means that the two corresponding keypoints belong to two different cells.
3) Randomly choose N_2 keypoints from the set of cell keypoints and form the set Ω. The probability of selecting a keypoint is proportional to its scale. Both N_1 and N_2 were set to 100 in our experiments.
4) Motivated by the intuition that *short* line segments are very unlikely to span two cells: For each element $\mathbf{p}_i$ in Ω, choose a random orientation θ_i and form the point $\mathbf{q}_i = \mathbf{p}_i + (s(\mathbf{p}_i)\cos(\theta_i), s(\mathbf{p}_i)\sin(\theta_i))$. Label each pair $(\mathbf{p}_i, \mathbf{q}_i)$ as *inner* which means that the two corresponding points belong to the same cell. The labels obtained by this step and by Step 2 are illustrated in Figure 1 (b).
5) For each *inner*/*cross* pair $(\mathbf{p}_i^*, \mathbf{p}_j^*)$, extract the following feature after [9]: $F_{ij} = \mathbf{I}_\sigma(\mathbf{p}_i^*) - 2\,\mathrm{extremum}_{ij} + \mathbf{I}_\sigma(\mathbf{p}_j^*)$. $\mathrm{extremum}_{ij}$ is, by definition, either the maximum (when $\beta = +1$) or the minimum ($\beta = -1$) intensity along the line segment between $\mathbf{p}_i^*$ and $\mathbf{p}_j^*$.
6) Estimate the two class conditional densities $\mathrm{P}(F|\textit{inner})$ and $\mathrm{P}(F|\textit{cross})$ assuming a Gaussian distribution.

So far, a keypoint-pair classifier was trained using the input image. The *posterior* probability $\mathrm{P}(\textit{cross}|F)$, assuming equal priors, is then used to rank each two nearby keypoints (cf. Figure 1 (c)). This ranking expresses the probability that they belong to two different cells. In order to reduce runtime, only the three nearest neighbors of each keypoint are considered. The resulting ranks are then used as input for an agglomerative hierarchical clustering with average linkage similar to [9]. The resulting clusters at a cut-off equal to 0.5 (cf. Figure 1 (d)) represent the detected cells. Inside each cluster, the arithmetic average of the keypoint coordinates identifies the center of a detected cell.

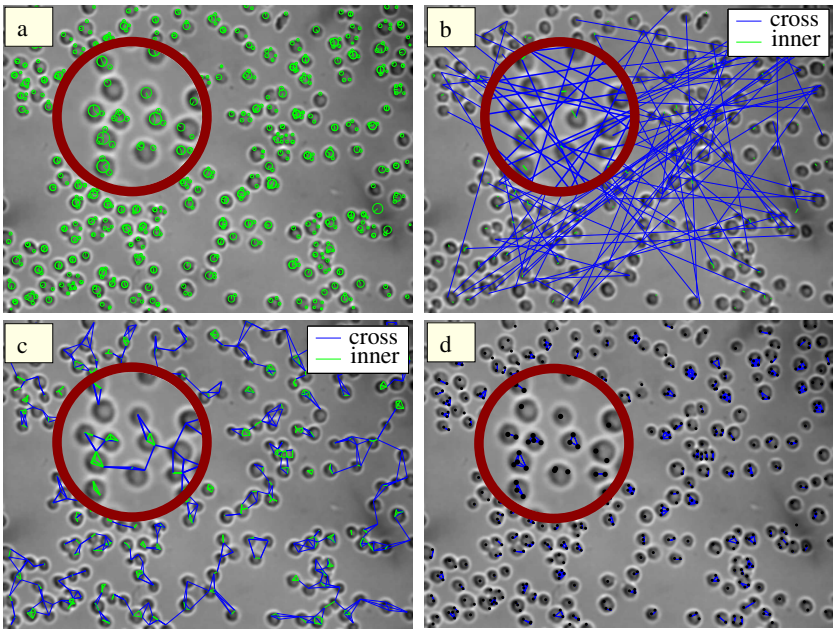


Fig. 1. Illustration of the cell/cell keypoint clustering. The circle inside each figure shows a magnified view. a) Cell keypoints resulting from the cell/background K-medians clustering. b) Point pairs chosen by the self-labeling algorithm for training a cell boundary potential. Each pair is indicated by a line segment. c) The learned boundary potential is employed to rank nearby keypoint pairs. The output is probabilistic, but only the binary classification result is shown. d) Result of hierarchical clustering using the ranks obtained from the previous step. Each cluster represents a detected cell.

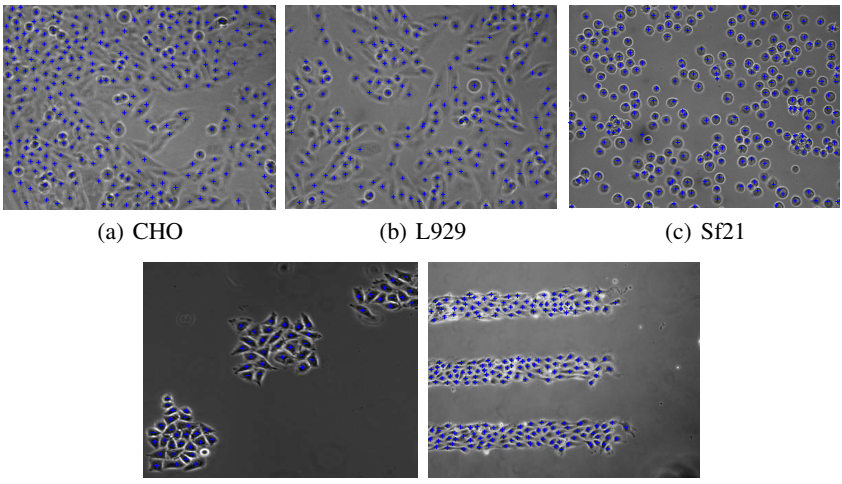
3 Evaluation

Table 1 contains a summary of the datasets used in the evaluation. The ground-truth type of all datasets except [13] is cell border delineation, while in the dataset of [13] a dot is marked at the center of each cell. This difference in ground-truth representation leads to a difference in the evaluation procedure. In all datasets except [13], a cell is considered detected if the hit point belongs to the cell mask and the *centeredness error* is used to assess the deviation from the cell center. Centeredness error is defined after [9] as the distance between the hit point and the cell's center of mass normalized by the cell major axis length and averaged over all correctly-detected cells in the considered image. In the dataset of [13], cell masks are not available. Therefore, a cell is considered detected if the distance to the ground-truth cell center is less than the minimum cell radius. The latter was set after [13] to 5 pixels. Figure 2 exemplifies detection results of our approach for all datasets. Quantitative evaluation and comparison with the state-of-the-art are described in the next paragraph. The evaluation results of all approaches except ours are given according to their corresponding papers: [9], [14], and [13].

A comparison with [9] on bright field microscopy is shown in Table 2. The figures of [9] in Table 2 were obtained by image-wise cross-validation in each cell line: One

image per cell line is used for training and the other images of the same cell line are used for testing. The results of our approach were obtained by averaging each of the F-measure, time, and centeredness error over images per cell line. A comparison with [13] and [14] on phase contrast microscopy is shown in Table 3. The shown results of the approaches [13] and [14] in Table 3 were generated by the hold-out method: [13] was trained using 11 images and tested on other 11 images. Similarly, [14] was trained using 10 images and tested on other 10 images. We evaluated our approach on the same images which were used for *testing* each of them (the images described in Table 1). Tables 2 and 3 show that the proposed approach is very close in terms of F-measure and centeredness error (when available) to the supervised approaches. However, our approach is much faster especially when compared with the phase contrast approaches where it is one or two orders of magnitude faster.

The blob type was correctly picked for all images by Eq. 2. As can be seen in Eq. 2, this blob type is decided by the sign function. Therefore, the reliability of the decision is proportional to the absolute value of the sign operand. We observed a little improvement (data not shown) of this reliability when both PCR and scale are used for weighting (as in Eq. 3) compared to the case when only the scale is used.



(a) CHO (b) L929 (c) Sf21

(d) HeLa (e) Bovine

Fig. 2. Samples of the detection results. Each plus sign marks a detected cell.

Table 1. Datasets used in the evaluation

Cell line	Modality	Resolution	#Images	#Cells	Ground truth	Source
CHO	bright field	1280×960	6	1431	border delineation	[9]
L929	bright field	1280×960	5	1078	border delineation	[9]
Sf21	bright field	1280×960	5	1001	border delineation	[9]
HeLa	phase contrast	400×400	11	1156	center dots	[13]
Bovine	phase contrast	680×512	10	2584	border delineation	[14]

Table 2. Comparison with the state-of-the-art on bright field microscopy

	F-measure (%)			Time (seconds)			Centeredness error		
	CHO	L929	Sf21	CHO	L929	Sf21	CHO	L929	Sf21
Supervised Mualla et al. [9]	84.2	86.5	97.0	45.9	36.7	40.7	0.48	0.38	0.16
Proposed approach	85.1	88.3	89.5	10.5	10.9	14.4	0.40	0.42	0.23

Table 3. Comparison with the state-of-the-art on phase contrast microscopy

	F-measure (%)		Time (seconds)		Centeredness error	
	HeLa	Bovine	HeLa	Bovine	HeLa	Bovine
Supervised Pan et al. [14]	-	94.6	-	900.0	-	-
Supervised Arteta et al. [13]	88.0	-	30.0	-	-	-
Proposed approach	88.7	86.0	1.5	3.5	-	0.11

4 Discussion and Conclusion

Both blob type detection and scale adaptive smoothing were proposed in the supervised approach of [9]. In contrast to [9], where only keypoints which belong to cells (known from ground truth) are considered, we compute the blob type β in an unsupervised manner by considering all keypoints. In addition, we use both scale and PCR to weigh the keypoint contribution to β whereas only scale is used in [9]. For the scale adaptive smoothing, we use a weighted average instead of the simple arithmetic average used in [9]. In general, we can conclude that SIFT can be successfully employed for *unsupervised* structure-of-interest measurements such as mean scale and dominant curvature direction.

In the cell/cell clustering step, a self-labeling algorithm was employed to train a ranking classifier. This classifier learns from extreme cases and applies the learned model on intermediate ones. In other words, training and testing feature vectors are drawn from different distributions. Therefore, the features should be chosen carefully so that they do not overfit the training samples. With this in mind, we confined ourselves to use a one-dimensional feature space and a simple generative model. In future, we plan to improve the cell/cell clustering by applying transductive transfer learning techniques. On the other hand, for the possibly less-reliable cell/background clustering, we think that applying transductive learning methods may alleviate the limitations of K-medians. In the self-labeling algorithm, due to the use of SIFT, it was possible to define a scale-invariant notion of the *extreme* cases. Consequently, the algorithm could successfully detect cells in images of different resolutions and/or cell types without any change in the parameter values.

The proposed approach achieves detection accuracy which is very close to three state-of-the-art supervised cell detection approaches in much less time, without training data, and without manual parameter-tuning. We thus believe that the cell detection problem is, to a large extent, solvable by self-supervised techniques which learn from the input image itself.

Acknowledgment. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research

Foundation (DFG) in the framework of the German excellence initiative. Special thanks go to Mr. Arteta and Dr. Pan for providing us with the two phase contrast datasets.

References

1. Sjöström, P.J., Frydel, B.R., Wahlberg, L.U.: Artificial neural network-aided image analysis system for cell counting. Cytometry 36(1), 18–26 (1999)
2. Loukas, C.G., Wilson, G.D., Vojnovic, B., Linney, A.: An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. Cytometry Part A 55A(1), 30–42 (2003)
3. Ali, R., Gooding, M., Szilágyi, T., Vojnovic, B., Christlieb, M., Brady, M.: Automatic segmentation of adherent biological cell boundaries and nuclei from brightfield microscopy images. Machine Vision and Applications 23(4), 607–621 (2012)
4. Li, K., Chen, M., Kanade, T., Miller, E., Weiss, L., Campbell, P.: Cell population tracking and lineage construction with spatiotemporal context. Medical Image Analysis 12(5), 546–566 (2008)
5. Long, X., Cleveland, W., Yao, Y.: Automatic detection of unstained viable cells in bright field images using a support vector machine with an improved training procedure. Computers in Biology and Medicine 36(4), 339–362 (2006)
6. Lulevich, V., Shih, Y.P., Lo, S.H., Liu, G.Y.: Cell tracing dyes significantly change single cell mechanics. The Journal of Physical Chemistry B 113(18), 6511–6519 (2009)
7. van Opstal, W., Ranger, C., Lejeune, O., Forgez, P., Boudin, H., Bisconte, J., Rostene, W.: Automated image analyzing system for the quantitative study of living cells in culture. Microscopy Research and Technique 28(5), 440–447 (1994)
8. Long, X., Cleveland, W., Yao, Y.: A new preprocessing approach for cell recognition. IEEE Transactions on Information Technology in Biomedicine 9(3), 407–412 (2005)
9. Mualla, F., Schöll, S., Sommerfeldt, B., Maier, A., Hornegger, J.: Automatic cell detection in bright-field microscope images using SIFT, random forests, and hierarchical clustering. IEEE Transactions on Medical Imaging 32(12), 2274–2286 (2013)
10. Nattkemper, T., Ritter, H., Schubert, W.: Extracting patterns of lymphocyte fluorescence from digital microscope images. In: Intelligent Data Analysis in Medicine and Pharmacology, pp. 79–88 (1999)
11. Mualla, F., Schöll, S., Sommerfeldt, B., Maier, A., Steidl, S., Buchholz, R., Hornegger, J.: Using the low-pass monogenic signal framework for cell/background classification on multiple cell lines in bright-field microscope images. International Journal of Computer Assisted Radiology and Surgery, 1–8 (2013)
12. Pan, J., Kanade, T., Chen, M.: Learning to detect different types of cells under phase contrast microscopy. In: Microscopic Image Analysis with Applications in Biology (2009)
13. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Learning to detect cells using non-overlapping extremal regions. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 348–356. Springer, Heidelberg (2012)
14. Pan, J., Kanade, T., Chen, M.: Heterogeneous conditional random field: Realizing joint detection and segmentation of cell regions in microscopic images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2940–2947 (2010)
15. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)