Detection of tumor cell spheroids from co-cultures using phase contrast images and machine learning approach

Neslihan Bayramoglu*, Mika Kaakinen†, Lauri Eklund†§, Malin Åkerfelt¶, Matthias Nees‡,
Juho Kannala*, Janne Heikkilä*

{neslihan.bayramoglu, mika.kaakinen, lauri.eklund, juho.kannala, janne.heikkila}@oulu.fi
{malin.akerfelt, matthias.nees}@vtt.fi
{*Center for Machine Vision Research, †Biocenter Oulu, §Oulu Center for Cell-Matrix Research
and Faculty of Biochemistry and Molecular Medicine}, University of Oulu, Oulu, Finland
¶Centre for Biotechnology, University of Turku, Turku, Finland
‡VTT Technical Research Centre of Finland, Turku, Finland

Abstract—Automated image analysis is demanded in cell biology and drug development research. The type of microscopy is one of the considerations in the trade-offs between experimental setup, image acquisition speed, molecular labelling, resolution and quality of images. In many cases, phase contrast imaging gets higher weights in this optimization. And it comes at the price of reduced image quality in imaging 3D cell cultures. For such data, the existing state-of-the-art computer vision methods perform poorly in segmenting specific cell type. Low SNR, clutter and occlusions are basic challenges for blind segmentation approaches.

In this study we propose an automated method, based on a learning framework, for detecting particular cell type in cluttered 2D phase contrast images of 3D cell cultures that overcomes those challenges. It depends on local features defined over superpixels. The method learns appearance based features, statistical features, textural features and their combinations. Also, the importance of each feature is measured by employing Random Forest classifier. Experiments show that our approach does not depend on training data and the parameters.

I. INTRODUCTION

Cell segmentation is a fundamental and a critical problem in microscopic image analysis. Compared to other domains, microscopic image segmentation has its own requirements and challenges such as handling deformable models and dynamic appearance in noisy images. Therefore, it is not applicable to utilize a general segmentation framework to address those challenges and needs. In addition, differences in cell imaging techniques as well as varieties in cell types and experiments increase the problem diversity in microscopic image segmentation research. In this work, we propose a method for detecting tumour cell spheroids in *phase contrast* (PC) images of co-cultures of fibroblasts and tumor cells in an experimental 3D model.

A. Phase contrast microscopy & motivation

Phase contrast imaging is the most commonly used [1], label free microscopic method for visualising cells. The principle of PC imaging is based on the phase difference between a transmitted light and the received wave that passes through objects. PC imaging has been shown to work very well for thin and transparent objects. Compared to bright field imaging, it exposes more fine details and time consuming cell labelling, and phototoxic effects of high energy light, necessary for fluorescence microscopy are omitted.

Most often previous studies utilizing PC images are from thin monolayer (2D) samples. However, monolayer screening is not suitable for every application. For example, for cancer research and drug discovery, 3D model systems mimic the complexity of the human tumour environment more accurately than monolayer cell cultures [4]. Although PC is not ideal for acquiring high quality in-focus images from 3D cultures, it is very convenient technique for long term living cell imaging due to the ease of use, and lack toxic effect to the cells. In this work we aim to develop computer vision based methods for analysis of PC images from 3D cultures and to distinguish different cell types in co-cultures of fibroblasts and tumor cells.

B. Dynamic cell morphology

The relationship between cancer progress, i.e. cell ability to sense its environment and adapt to it, and its morphologic appearance are crucial elements in the tumor progression [2]. These morphology changes in three-dimensional environments compared to two dimensional models resembles more closely tumor environment. Therefore, 3D models in cancer investigations are useful as an intermediate step between primary drug screening and animal experimentation [4]. To monitor and quantify the cell dynamics in cancer biology in live cell microscopy which is comprised of long image sequences, automated image analysis solutions are needed. The first step in analysing morphological cell changes and other higher level processing tasks is to detect cells.
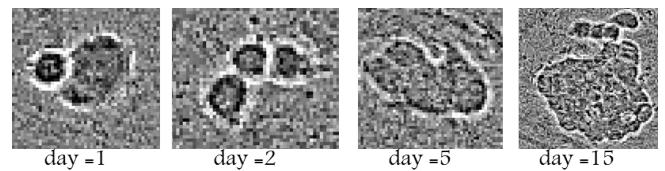

day =1   day =2   day =5   day =15

Fig. 1.   Growth of a tumor cell spheroid over time. At day 1 few cells form a small cluster. Cell number increase over time. Images are not in scale.

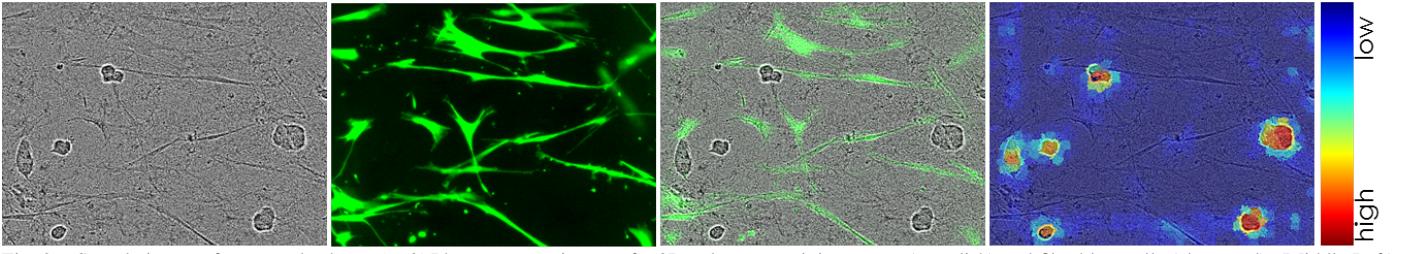The shape of cell spheroids vary in time due to proliferation

Fig. 2. Sample images from our database. *(Left)* Phase contrast image of a 3D culture containing tumor (roundish) and fibroblast cells (elongated). (Middle Left) Fluorescent image of the same culture. Green Fluorescent Protein (GFP) is used to label fibroblast cells. (Middle Right) Phase contrast and fluorescent images are superimposed and making the interpretation of PC image easier. (Right) Our learning based probabilistic detection result. Colormap indicates confidence level. Best viewed in color.

of cancers cells and changes in their extracellular environment (Figure 1). Such variety in shapes prevent image analysis methods to utilize prior information about cell shape. In addition, depending on the intervals between imaging times, cancer cells within the spheroids can change their shapes thoroughly and/or change their positions relative to the previous time step.

II. PREVIOUS WORK

Because of its simplicity the most commonly used method in cell detection applications is *thresholding* (i.e. Otsu's method [6]). However, it works only for high quality images in which cells have significantly different intensities from the background. Dynamic (adaptive) thresholding [7] provides better results compared to global thresholding but its power does not go beyond compensating uneven image illumination and some noise.

The *watershed* method is one of the most popular approach among region based segmentation methods [5]. The method is highly sensitive to local minima (false edges) or requires high precision markers which is hard to obtain automatically. Although watershed segmentation has a widespread usage in medical image analysis with many variants, high noise levels greatly effect the segmentation result. Besides, if there is non-uniform intensity profile within cell or cell clusters and also along the background as in our situation, watershed method results over segmentation.

Deformable models such as active contours (Snakes [15] and Level Sets [18]) are well-know methods for extracting the boundaries of objects in biomedical images. The basic idea is to evolve a curve in a constrained way for detecting objects. The segmentation task is formulated as an energy minimization problem based on several functions such as gradient flow, edge, intensity, and curvature information. The major disadvantage of Snakes is the necessity of initializing a contour nearby the objects edges. Moreover, it is sensitive to local minima and sensitive to parameter selection. Although many improvements on active contours approaches are introduced [16], [17] such methods are always suffering from missing open contours or sticking at false edges.

Since phase contrast images of 3D cultures are extremely degraded, image gradient, even after filtering, does not provide useful information for segmenting objects. This is demonstrated in Figure 3. In this particular image patch, tumour cells (outlined with red) and fibroblast cells (outlined with green) are residing with some occlusions (outlined with blue). Like in most of our cases, gradient magnitude cannot resolve

cell boundaries clearly (Figure 3, right). In addition to image degradation, edges that are belonging to other cell types (fibroblasts) are inevitably contributing to the gradient image. Therefore, sole gradient/edge based segmentation approaches are not suitable in such circumstances. Fibroblast cells and cell interactions in our cell culture can be observed better in Figure 2. For a detailed information on significant past efforts in cell segmentation research, the reader can refer to a recent review given in [3].
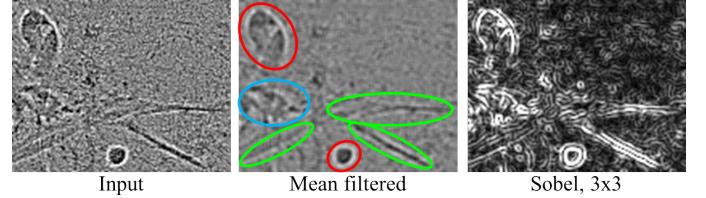

Input   Mean filtered   Sobel, 3x3

Fig. 3. Markings in mean filtered image are as follows. Red: tumour cells, green: fibroblast cells, blue: fibroblast cells and tumour cells are heavily occluded.

While simple intensity thresholding mechanisms are enough to segment and detect cells in some cell research the state-of-the-art segmentation methods produce unsatisfactory results for complex experiments such as in our case. Apart from the segmentation performance, segmentation frameworks are not capable of discriminating different cell types in mixed populations. Therefore, in such situations, instead of utilizing blind segmentation schemes learning based schemes are more appropriate. By this way, image features that are specific to particular cell type that we want to detect could be extracted. Yin et al. [20] adopt pixel-based classification scheme employing local histograms with a sliding window and train a set of Bayesian classifiers to segment cells in monolayer PC images containing one cell type. Arteta et al. [19] use features over *Maximally stable extremal regions* (MSER) [23] with *Support Vector Machines* (SVM) classifier to detect cells in cultures consisting of single cell type. Features basically consist of several concatenated histograms of intensities in various contexts. MSER initialization in [19] reduces the computational complexity of the sliding window approach adopted in [20]. Another related technique [21] uses prob-abilistic SVM classification results of superpixel features as an input to graph-cut algorithm to segment cells. In [21] Ray descriptors for capturing object shape features are incorporated with texture based rotation invariant features. Ray features are first employed in another learning based work [22] for detecting mitochondria and neuron nuclei.

In this study, we employed PC images taken from 3D models in which tumor cells reside and interact dynamically
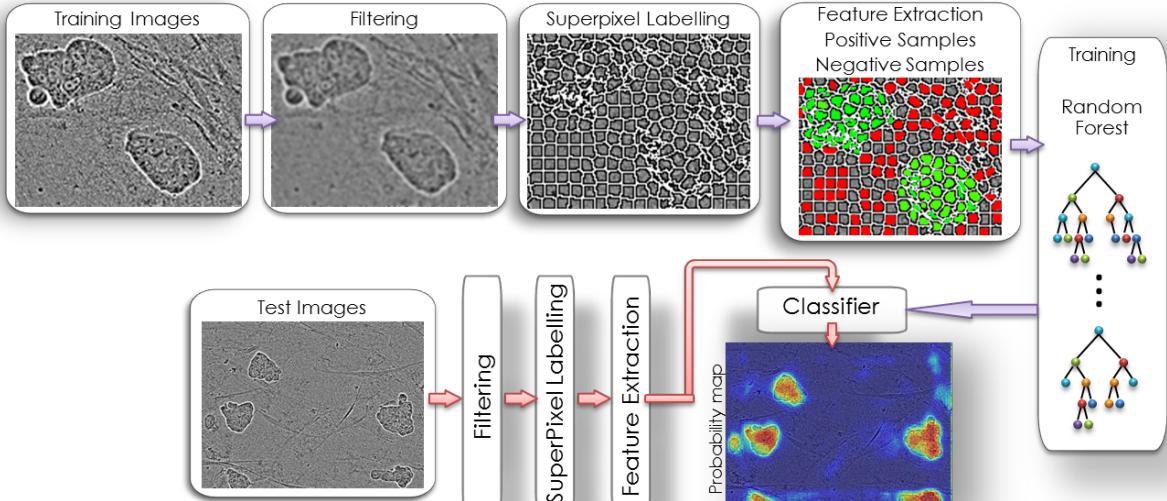
Fig. 4. Flowchart of the proposed method. PC images are first filtered to remove noise. Superpixel segmentation is performed to reduce complexity and to work on compact regions that follow image boundaries. A Random Forest classifier is trained on features that are extracted from both positive and negative samples. During test time similar process is followed and superpixels are assigned to probabilistic outputs based on the average decision generated by the trees. Final decision is made by simple thresholding.

with the surrounding matrix and fibroblast cells (Figure 2) and we propose a supervised learning framework to detect tumour cells. Although fluorescent images of the experiments which can be used to distinguish fibroblast cells are available, we utilize only PC images for reasons mentioned above. In contrast to previous studies, our data contains different cell types that can change their appearance and shape significantly over time. Besides, the images have poorer qualities, lower *Signal to Noise Ratio (SNR)* due to underlying imaging technique (PC) and the cell culture thickness. Moreover, we employ both appearance and texture based descriptors together with statistical measures to extract as many distinctive features as possible. We measure the importance of each feature by training random forest classifier and obtain probabilistic detection results. The method comprises four main steps: filtering, superpixel segmentation, feature extraction, and classification (Figure 4). Working on PC images of thick samples comes at the price of reduced image quality. To increase the *SNR* we apply Gaussian filtering at the beginning of the training and the testing process.

III. SUPERPIXEL SEGMENTATION

Superpixel labelling is an over segmentation strategy for partitioning images into smaller patches that are spatially contiguous and similar in intensity. It is a common preprocessing step to work on superpixels as the basic unit instead of working on the pixel level. This greatly reduces the computational complexity of following image processing tasks. Superpixel clusters are convenient as they produce compact regions that follow image boundaries.

In this study we adopt *simple linear iterative clustering* (SLIC) [24] method because of its simplicity and flexibility in the compactness and number of the superpixels it generates. The algorithm adopts *k-means* clustering in intensity and spatial domain. Starting from *k* regularly spaced cluster centers, each pixel in the image is associated with the nearest cluster center. The process of associating pixels with the nearest cluster center and recomputing the cluster center is repeated

iteratively until convergence. At each iteration, superpixels are reassigned to the average color and position of the associated input pixels. The distance measure *(D)* in computing the distance between a pixel and cluster center is a simple Euclidean norm in the five-dimensional space (color(CIELAB) +spatial).

$$D = \sqrt{d_c{}^2 + \frac{d_s{}^2}{S} m^2}, \qquad (1)$$

where d_c is the color distance, d_s is the spatial distance, *m* is a parameter that weighs the relative importance between color similarity and spatial proximity, and *S* is a parameter indicating the size of superpixels. Compactness, *C*, (i.e.more compact superpixels have lower area to perimeter ratio) can be controlled by *m*.

IV. FEATURE EXTRACTION

We utilize well known local image descriptor scale invariant feature transform SIFT [25] to extract appearance based features. It has been shown to be successful in various image processing applications for detecting and describing interest points (keypoints) locally. SIFT descriptors are normally computed at local maxima and minima of difference-of-Gaussian (DoG) images. In this study we compute SIFT descriptors (f_{sift}) at points specified by superpixel centers at fixed scale. SIFT descriptor in this context are similar to Histograms of Oriented Gradients (HOG) descriptor. However, rotation invariance is ensured as it is done in [25] such that a gradient orientation histogram is computed in the neighbourhood of the keypoint in a weighted manner. The highest peak of the histogram is regarded as the "dominant" orientation and finally descriptor is evaluated relative to the orientation of the keypoint.

Secondly, histograms of intensity values are employed. Such intuitive features are already used in many learning based applications but their performances could significantly vary depending on the input characteristics. Therefore, this feature could also be interpreted as a baseline method for comparison

purposes. In this work, intensity values from pixels contained in superpixel k are extracted and represented with an *n-bin* histogram $f_{intensity}$.

Lastly, we employ Center Symmetric Local Binary Patterns (CS-LBP) [8] which is developed for region description based on the Local Binary Pattern (LBP) [9]. It has been shown to be quite successful in many computer vision problems particularly in face analysis research [10]. CS-LBP produces smaller number of labels than LBP thus results shorter histograms which are better suited for region description. Intensity values of opposing pixels are compared and the differences are thresholded. For N equally spaced pixels on a circle of radius R centered at point p_i the (CS-LBP) operator is defined as :

$$CS\text{-}LBP_{p_i} = \sum_{i=0}^{N/2-1} s(I_i - I_{i+N/2})2^i, \quad s(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$
(2)

where I_i is the intensity at point p_i. We start with extracting rotation invariant CS-LBP signature for each pixel in the input image. Then *m-bin* histograms (f_{cslbp}) of CS-LBP labels over superpixels are constructed. CS-LBP features provide low level intensity and texture cues.

Finally, each superpixel is associated with a set of features $f_{sift}, f_{intensity}$, and f_{cslbp}. To cover all these features at a time, for each superpixel k, we make a composite feature vector f_k combining SIFT descriptors, intensity histograms, and CS-LBP features: $f_k = [f_{sift} \quad f_{intensity} \quad f_{cslbp}]$.

V. CLASSIFICATION

Random Forests (RF) is an ensemble of tree predictors introduced by Ho [13] in 1995 and later studied in depth by Breiman [11] in 2001. It is shown to be a fast and effective classification and regression method for many applications [12], [14]. Simply, RF classifier bring together weak learners to form a strong learner. Each tree in the forest consists of split and leaf nodes. Features at each node are selected randomly and nodes are split into two (binary partitioning) by calculating the best split based on these randomly selected features. Best split is evaluated by maximizing the information gain of the split. Leaf nodes are created when the maximum tree depth is reached or the number of training samples at the node is less than the predefined threshold. During the training each leaf node may store the empirical class distributions associated to the subset of training data that has reached that leaf node [12]. During the test time, the query is sent down starting from the root node to the leaves through all trees in the forest. Finally, an ensemble class posterior is obtained by averaging all tree posteriors. Final decisions for superpixels' class ids are made by simply thresholding the probabilistic output. Precision-Recall Curves are obtained by varying the threshold from 0 to 1. Detailed information on decision forests can be found in [12].

VI. EXPERIMENTS AND RESULTS

In this study three image sequences are employed. Each sequence has different experimental conditions and their details are beyond the scope of this work. However, due to varying conditions cell number, size, shape in time, cellular dynamics, and morphology may differ in each experiment.

	TABLE I. IMAGE SEQUENCES		
Sequence Name	Number of Images (Original / Annotated)		Train / Test
Collagen	330 / 185		19 / 166
Matrigel	331 / 27		0 / 27
Matrigel-Collagen	330 / 27		0 / 27

This is the only important information regarding the effects of experimental set-up. Sequences contain 330 images on the average with one image per hour frequency; i.e. two weeks of observations. For training and evaluation purposes we utilize only a subset of the whole dataset by sampling uniformly. Because, obtaining ground truth requires marking tumour regions manually which is a laborious process. Sequence names representing culture matrices *(Collagen, Matrigel, Matrigel-Collagen)* and corresponding number of images that have ground truth information are given in Table I. Number of images in training and testing sequence are shown in the last column of the table. To investigate the limitations of our methods we utilize images only from *Collagen* sequence for training. The other two sequences are used only for testing purposes. Collagen sequence is divided into two parts: one for training (19 images), other for testing (166 images). Images have sizes of 300×400 *(height$\times$width)* pixels.

For training Random Forest Classifier 1000 *trees* are employed. We test trees with 25 as the maximum *depth* level. At each node 5 *features* are *randomly* used for splitting. It has been found that the procedure is not overly sensitive to the number of features used to split at each node [14]. During the training time, 79 positive samples (number of superpixels) per image and 668 negative samples per image on the average are used. These numbers are given for an image segmentation of appropriately 1200 superpixels. During the test time class posterior of a superpixel feature (output of the RF classifier) is applied for each pixel belonging to that particular superpixel.

For SIFT features we employ similar parameters as it is originally proposed in [25]. Image patches centered on keypoints with keypoint size of 16 in the OpenCV implementation are used. Patches are divided into into 4×4 pixel tiles and for each such tile a histogram of gradient orientations (relative to the dominant orientation) are computed with 8 bins resulting a 128 dimensional feature descriptor. Intensity histograms and CS-LBP histograms are calculated for 16 bins ($n = m = 16$). Finally, combined feature vector f_k for superpixel k becomes a vector of length 160.
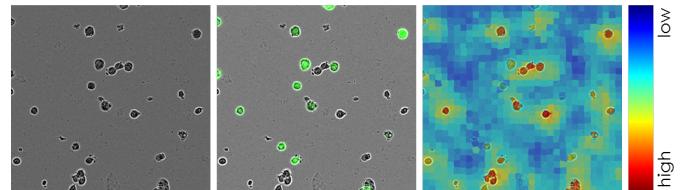

Fig. 6. Images are from the initial stages of the experiment. Fibroblast cells and tumour cells have very similar shapes and appearances in phase contrast images *(left)*. It is almost impossible to distinguish tumour and fibroblast cells without the help of fluorescence images *(middle)*. Classifier (right) in that stage results with both low precision and recall values when the probability threshold for making decision is set to a high value. Best viewed in color.

For evaluating performances we adopt Precision-Recall (PR) Curves where $precision = tp/(tp + fp)$ and $recall = tp/(tp + fn)$, *tp:* true positives, *fp:* false positives, and *fn:*
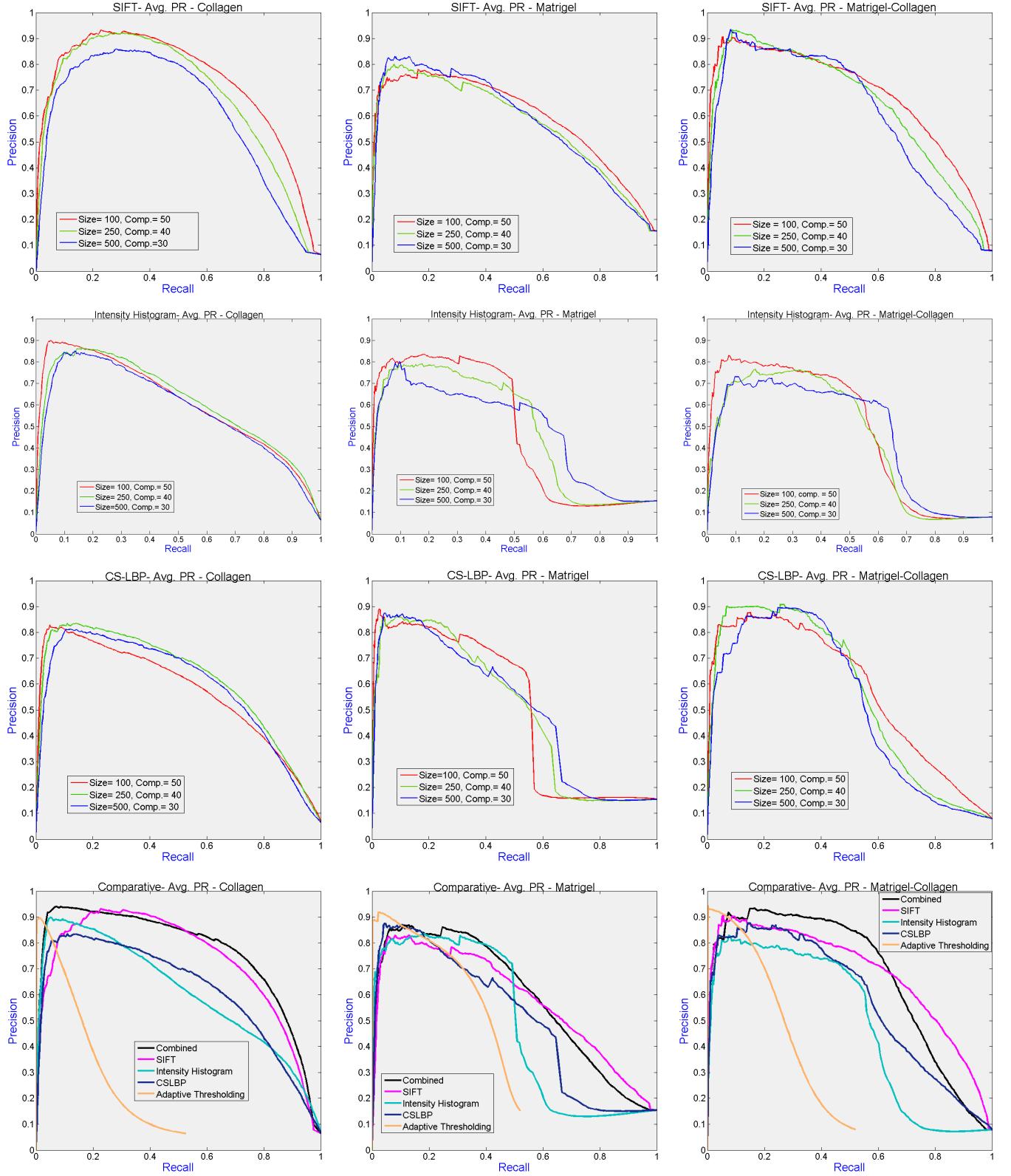
Fig. 5. Precision Recall curves. *(First Column)* Collagen experiments. *(Second Column)* Matrigel experiments. *(Third Column)* Matrigel-Collagen experiments. *(First Row)* SIFT features for varying segmentation parameters. *(Second Row)* Histogram of intensity features for varying segmentation parameters. *(Third Row)* CS-LBP features for varying segmentation parameters. *(Last Row)* Comparison of detection performance of combined features, SIFT features, histogram of intensity features, CS-LBP features and adaptive thresholding method. SIFT features yield higher precision especially at high recall values. Slight improvement is obtained by introducing the combined features. Varying parameters in superpixel segmentation process yield similar performance in classifying individual superpixels. Although training data is obtained only from Collagen sequence, our approach performs much better than the baseline method in other sequences as well. Best viewed in color.

false negatives. Evaluation is performed on the pixel level. Since there is no standard baseline method for comparing learning frameworks, we compare our method with adaptive thresholding which is one of the most widely used approach in biomedical image processing applications. Figure 5 presents PR curves for various test settings. Each column is dedicated to one image sequence *(Collagen, Matrigel, Matigel-Collagen)*. In the first row, PR Curves for only SIFT features are shown for varying parameters of superpixel segmentation. Similarly histogram of intensity feature performance is shown in the second column and PR curves for CS-LBP features are given in the third row. In the last row, in addition to the best performances of single features, PR curves of combined feature vector and adaptive thresholding are shown. Shapes of PR curves of our proposals especially at low recall values are different than accustomed PR curves (i.e. rising from low precision value instead of lowering from a higher value). The reason for that is explained with an example in Figure 6.

Three values for size (S) and compactness (C) parameters are tested: $(S, C) = \{(100, 50), (250, 40), (500, 30)\}$. Smaller region with less compact choice slightly improves performances for low recall values but does not have significant effects. Among three features, SIFT features performed better on the average. Combined features have higher PR values than SIFT's values for some operating regions. On the other hand, worst performance belongs to the adaptive thresholding for support region of a 15×15 window. In this method some regions in the PR curve cannot be reached. The most significant result in this study is to observe similar detection performances for two sequences (Matrigel and Matrigel-Collagen) which did not contribute to the training phase. Based on this, it can be concluded that the proposed method has no heavy dependency on the training data and the segmentation parameters, as long as the images have similar local features. Such similarity doesn't put constraints on global features like appearance, shape or size since our three sequences already have different cell dynamics.

VII. CONCLUSION

Learning based detection approaches are better suited than blind segmentation methods for cluttered and noisy microscopic image analysis. Learned features from single image sequence could be utilized to detect similar structures in other sequences that might have different dynamics. In addition, detection performance does not suffer from parameter tuning. Proposed method can be employed in different applications in biomedical image analysis where obtaining approximate regions with high precision and recall is more important than extracting exact cell boundaries.

REFERENCES

[1] Mika Kaakinen, Sami Huttunen, Lassi Paavolainen, Varpu Marjomäki, Janne Heikkilä, and Lauri Eklund. "Automatic detection and analysis of cell motility in phase contrast time lapse images using a combination of maximally stable extremal regions and Kalman filter approaches." Journal of Microscopy, January, 253(1):65-78.

[2] Debeir, Olivier, Ivan Adanja, Robert Kiss, and Christine Decaestecker. "Models of cancer cell migration and cellular imaging and analysis." The Motile Actin System in Health and Disease (2008): 123-156.

[3] Meijering, Erik. "Cell Segmentation: 50 Years Down the Road [Life Sciences]." Signal Processing Magazine, IEEE 29, no. 5 (2012): 140-145.

[4] Härmä, Ville, Johannes Virtanen, Rami Mäkelä, Antti Happonen, John-Patrick Mpindi, Matias Knuuttila, Pekka Kohonen, Jyrki Lötjönen, Olli Kallioniemi, and Matthias Nees. "A comprehensive panel of three-dimensional models for studies of prostate cancer growth, invasion and drug responses." PLoS One 5, no. 5 (2010): e10431.

[5] Beucher, Serge, and Christian Lantuejoul. "Use of watersheds in contour detection." (1979).

[6] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." Automatica 11, no. 285-296 (1975): 23-27.

[7] Sauvola, Jaakko, and Matti Pietikäinen. "Adaptive document image binarization." Pattern Recognition 33, no. 2 (2000): 225-236.

[8] Heikkilä, Marko, Matti Pietikäinen, and Cordelia Schmid. "Description of interest regions with local binary patterns." Pattern recognition 42, no. 3 (2009): 425-436.

[9] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, no. 7 (2002): 971-987.

[10] Bayramoglu, Neslihan, Guoying Zhao, and Matti Pietikäinen. "CS-3DLBP and geometry based person independent 3D facial action unit detection." Int. Conf. on Biometrics (ICB), 2013 , pp. 1-6. IEEE, 2013.

[11] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[12] Criminisi, A., J. Shotton, and E. Konukoglu. "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning." Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114 5, no. 6 (2011): 12.

[13] Ho, Tin Kam. "Random decision forests." In Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, vol. 1, pp. 278-282. IEEE, 1995.

[14] Bosch, Anna, Andrew Zisserman, and Xavier Muoz. "Image classification using random forests and ferns.", ICCV 2007, pp. 1-8. IEEE, 2007.

[15] Kass, Michael, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models." IJCV 1, no. 4 (1988): 321-331.

[16] Chan, Tony F., and Luminita A. Vese. "Active contours without edges." Image Processing, IEEE Transactions on 10, no. 2 (2001): 266-277.

[17] Caselles, Vicent, Ron Kimmel, and Guillermo Sapiro. "Geodesic active contours." IJCV 22, no. 1 (1997): 61-79.

[18] Osher, Stanley, and James A. Sethian. "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations." Journal of computational physics 79, no. 1 (1988): 12-49.

[19] Arteta, Carlos, Victor Lempitsky, J. Alison Noble, and Andrew Zisserman. "Learning to detect cells using non-overlapping extremal regions." MICCAI 2012, pp. 348-356. Springer Berlin Heidelberg, 2012.

[20] Yin, Zhaozheng, Ryoma Bise, Mei Chen, and Takeo Kanade. "Cell segmentation in microscopy imagery using a bag of local Bayesian classifiers." In Biomedical Imaging: From Nano to Macro, pp. 125-128. IEEE, 2010.

[21] Lucchi, Aurlien, Kevin Smith, Radhakrishna Achanta, Vincent Lepetit, and Pascal Fua. "A fully automated approach to segmentation of irregularly shaped cellular structures in EM images." MICCAI 2010, pp. 463-471. Springer Berlin Heidelberg, 2010.

[22] Smith, Kevin, Alan Carleton, and Vincent Lepetit. "Fast ray features for learning irregular shapes." In Computer Vision, 2009 IEEE 12th International Conference on, pp. 397-404. IEEE, 2009.

[23] Matas, Jiri, Ondrej Chum, Martin Urban, and Tomás Pajdla. "Robust wide-baseline stereo from maximally stable extremal regions." Image and vision computing 22, no. 10 (2004): 761-767.

[24] Achanta, Radhakrishna, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. "Slic superpixels." École Polytechnique Fédéral de Lausssanne (EPFL), Tech. Rep 149300 (2010).

[25] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60, no. 2 (2004): 91-110.