AUTOMATIC CELL REGION DETECTION BY K-MEANS WITH WEIGHTED ENTROPY

Benjamin X. Guan, Bir Bhanu, Ninad S. Thakoor

Center for Research in Intelligent Systems
University of California, Riverside
Riverside, USA
{xguan001,bhanu,nthakoor}@cris.ucr.edu

Prue Talbot, Sabrina Lin

Stem Cell Center
University of California, Riverside
Riverside, USA
{prudence.talbot, sabrina.lin}@ucr.edu

ABSTRACT

In this paper, we propose an automatic method to detect human embryonic stem cell regions. The proposed method utilizes the K-means algorithm with weighted entropy. As in phase contrast images the cell regions have high intensity variation, they usually yield higher entropy values than the substrate regions which have less intensity variation. Thus, the entropy can be used as an important feature for the detection of stem cells. However, homogeneity in intensity within some of the cell bodies and halos surrounding the cell bodies also gives low entropy values. Therefore, we introduce a weighted entropy formulation which fuses entropy and image intensity information to detect the entire cell regions.

Index Terms— K-means, Weighted entropy

1. INTRODUCTION

Time lapse imaging of embryonic stem cells has become an extremely useful method for studying the characteristic and behavior of these cells. Many biological assays in the field of prenatal toxicology and stem cell differentiation have been and can be further developed with collection and analysis of video [1]. However, the large number of image data means that analyzing them can be time consuming. In order to reduce the time required for analyzing massive image data of stem cells, we need to analyze only the essential image information pertaining to the stem cells. Therefore, an automatic, fast, and accurate method for stem cell detection is important for stem cell research.

K-means algorithm is conventionally used [2, 3] to segment the image into k clusters based on the pixel intensity similarity alone. The algorithm is a good candidate for automated analysis of stem cell images. In K-means, each pixel is grouped into the cluster which has the nearest mean intensity value. Parts of the cell have intensities lower than the substrate (background) intensities while other parts have higher intensities (See Figure 1). This means that, K-means on intensity values would fail to extract the stem cell regions in their entirety. In this paper, we propose a weighted entropy scheme in which the clustering of each pixel depends on each pixel's weighted entropy value instead of the pixel's

intensity value alone. We demonstrate improvements in the stem cell detection accuracy with the proposed method.
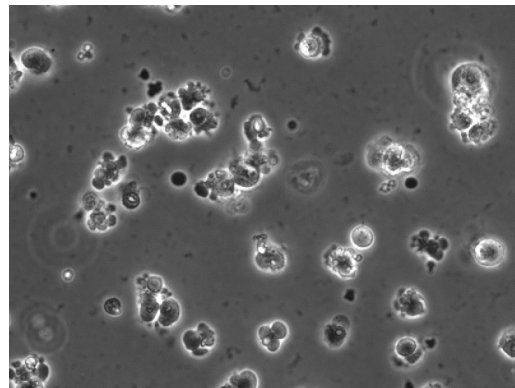

Fig. 1: Example phase contrast image with $20\times$ objective

2. TECHNICAL APPROACH

The cell regions in the phase contrast image hold high significance for analyzing cell behavior over time. Due to the internal cell structure, the cell regions have higher intensity variation than the substrate region. Consequentially, the cell regions yield higher local entropy values than the substrate region. However, parts of cell such as cytoplasm as well as halo surrounding the cell also exhibit less variation in the intensity. Thus the local entropy information alone is not sufficient to distinguish the cell regions. Table 1 summarizes the various image region properties. Using these properties, We

Table 1: Image regions and properties

Image region	Intensities	Entropy
Substrate	Mid	Low
Halo	High	Low
Cell regions	Low/Mid/High	High
Cytoplasm	Low	Low

develop a weighted entropy approach for detection of stem cell region.

2.1. Weighted Entropy

As cell regions are nothing but non-substrate regions, our aim is to detect the non-substrate parts of the image. We first define a substrate weight as,

$$W_s(i) = \frac{1}{\sqrt{2\pi\omega^2}} \exp -\frac{1}{2} \left(\frac{i-p}{\omega}\right)^2 \qquad (1)$$

where W_s is a gaussian shaped weighting function centered at p and with width ω, $i \in [0, L]$ is the the image intensity. For an 8-bit image, $L = 255$. The substrate weight is expected to have higher values in the substrate regions, i.e., middle intensities of the image. A complementary cell weight is given by,

$$W_c(i) = \max(W_s) - W_s(i) \qquad (2)$$

The cell weight is expected to be high in most of the cell regions.

For an image I with dimensions $M \times N$, weighted local entropy E_W at the location (x, y) is defined as

$$E_W(x, y) = \log \left(1 + E(x, y) \sum_{\forall (x', y') \in \mathcal{N}} W_c(I(x', y'))^2\right) \qquad (3)$$

where $E(x, y)$ is the local entropy of the image gradient magnitude at location (x, y). We use gradient magnitude in place of intensities as it can further distinguish the cell regions from the substrate region [4]. The summation term is the local spatial energy [5] of the cell weights over a spatial neighborhood $\mathcal{N}$. Equation (3) enhances the separation between cell and substrate regions by combining intensity information W_c and local entropy E.

2.2. Optimization Metric

The substrate and cell region weights depend on the parameters p and ω as seen in eq. (1). As substrate constitutes the majority of the image, center p of the weighting function is set to the mode of the image histogram. To determine the width parameter ω, we propose an iterative algorithm which maximizes the following metric.

$$M(m_c, m_s) = (m_c - m_s)^2 \qquad (4)$$

where m_f and m_b are mean intensities from cell region C and substrate region S of the image.

$$m_c = \frac{1}{|C|} \sum_{\forall (x,y) \in C} I(x, y) \qquad (5)$$

$$m_s = \frac{1}{|S|} \sum_{\forall (x,y) \in S} I(x, y) \qquad (6)$$

At each iteration, K-means algorithm is used for segmenting the weighted entropy E_W into a cell region C and

substrate region S by varying ω. The cell region C is the the one with higher mean entropy and the other one is the substrate region S. Algorithm 1 outlines the complete optimization process.

Algorithm 1: WeightedEntropyCellRegionDetection

Input : $I \leftarrow$ Phase contrast image with stem cells.
Output: $C \leftarrow$ Cell region. $S \leftarrow$ Substrate region.
1. Set center $p =$Mode(Histogram(I)).
2. Set initial width $\omega = 1$.
3. Set the initial optimization metric $M_0 = 0$.
4. Set iteration $n = 1$.
5. Calculate weighted entropy E_W from eqs. (1) to (3).
6. Apply K-means to cluster E_W in to two clusters.
7. Determine cell region C and substrate region S.
8. Calculate m_c and m_s with eqs. (5) and (6).
9. Compute metric M_n for the current iteration from eq. (4).
10. If $M_n > M_{n-1}$, set $\omega = \omega + 1$, $n = n + 1$ and go to step 5; else terminate the algorithm.

---

3. EXPERIMENTAL RESULTS

Phase contrast stem cell images used in the experiments were acquired under objectives $10\times$, $20\times$, and $40\times$ with a BioStation IM [6] system at 800×600 resolution.

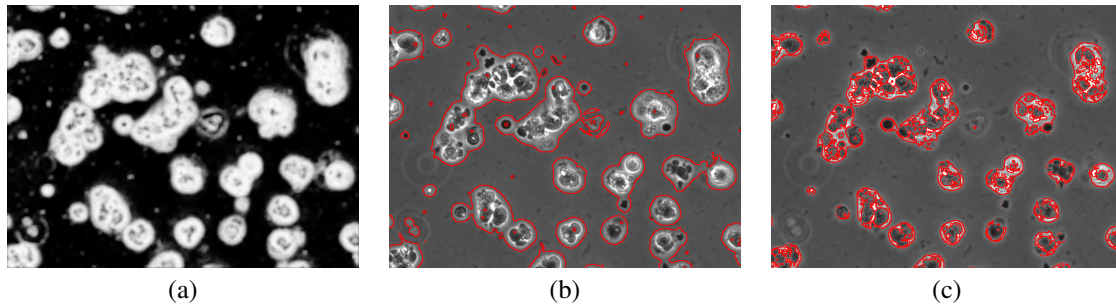### 3.1. Phase Contrast Image Taken with $10\times$ Objective

Figure 2 shows the results for one of the images captured with $10\times$ objective. Figures 2(b) and (c) show cell weights and local entropy respectively of the image under $10\times$. It can be seen that these images contain complementary cell region information as cell weights capture halo regions better while local entropy captures the cell interior. Figure 2(d) shows the optimal weighted entropy, and it clearly separates the cell regions from the substrate when compared to cell weights and local entropy alone. For comparison, along with outcome of the proposed method (fig. 2(e)), result of conventional K-means with intensities is alo shown in fig. 2(f). The proposed method is able to detect the cells in their entirety. Intensity similarity between cell bodies and the substrate regions that hinders the conventional K-means segmentation does not affect the proposed method.

3.2. Phase Contrast Image Taken with $20\times$ Objective

Figure 3(b) and (c) shows the results of the proposed method and the conventional K-means segmentation for the image in fig. 1. Similar to $10\times$ image, the proposed method is able to capture majority of the cell regions while the conventional K-means capture only parts of the cells.
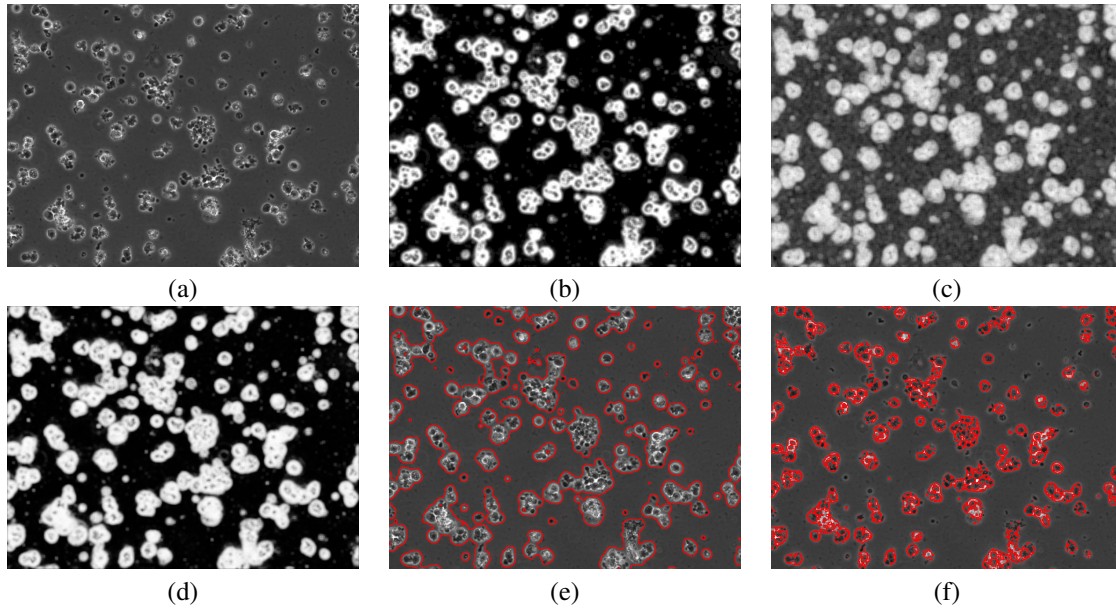
Fig. 2: (a) Original stem cell $10\times$ image, (b) Cell weights for the original image, (c) Entropy of the original stem cell image, (d) Optimal weighted entropy from the proposed algorithm, (e) Final result for the image from the proposed algorithm, (f) Final result for image by conventional K-means

Fig. 3: (a) Weighted entropy for $20\times$ image (fig. 1) from the proposed algorithm, (b) Final result for $20\times$ image from the proposed algorithm, (c) Final result for the image by conventional K-means

3.3. Phase Contrast Image Taken with $40\times$ Objective

The image shown in fig. 4(a) is a noisy image taken with $40\times$ objective. Figures 4(c) and (e) show comparison of the proposed method with/without pre-filtering with median filter. Pre-filtering is needed due to the significant noise which affects the local entropy computation. Figure 4(e) shows a significant improvement after a 5×5 median filtering.

For a comprehensive evaluation, we generated ground truth for 40 images from each video where cell regions were marked by an expert. Figure 5 shows the pixelwise true positive rate and false positive rate comparison of the three videos under different objectives. It shows that the proposed method has better stem cell detection rate while keeping the false positive rate low. The average number of iterations of the algorithm for the images was 11. The spatial neighborhood for the energy term in eq. (3) used were square windows of size 7×7, 9×9, 13×13 for images with objective $10\times$, $20\times$, and $40\times$ respectively. The result were obtained in about 35 seconds for each image with a Matlab imple-

mentation running on a 2.53GHz Intel(R) Core 2 Duo CPU.

4. CONCLUSIONS

This paper proposed a weighted entropy formulation for detection of stem cell regions in phase contrast images. The weighted entropy incorporated intensity characteristics and the local entropy for cell region detection. The proposed method provides an efficient way to solve a complicated problem and also has a simple implementation. It significantly improves the performance over the conventional intensity based K-Means clustering. The method attains higher cell region detection accuracy which is crucial for automated cell analysis on a large scale.

5. ACKNOWLEDGEMENT
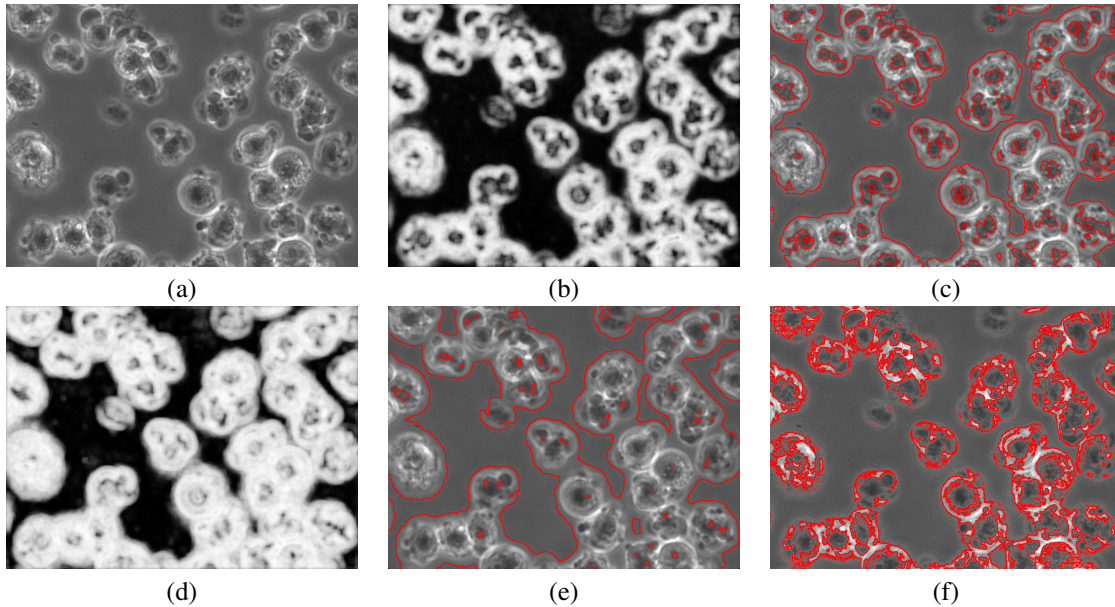
Fig. 4: (a) Original $40\times$ stem cell image, (b) Weighted entropy, (c) Final result for the image by our method, (d) Weighted entropy with 5x5 median pre-filtering, (e) Final result for the image from the proposed algorithm with 5x5 median pre-filtering, (f) Final result for the image by conventional K-means

6. REFERENCES

[1] P. Talbot and S. Lin, "Mouse and human embryonic stem cells: Can they improve human health by preventing disease?," *Current Topics in Medicinal Chemistry*, vol. 11, pp. 1638–1652, 2011.

[2] S. Tatiraju and A. Mehta, "Image segmentation using k-means clustering, EM and normalized cuts," Tech. Rep., UC Irvine, 2008.

[3] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions PAMI*, vol. 24, no. 7, pp. 881 –892, jul 2002.

[4] A. L. Barbieri, G.F. de Arruda, F. A. Rodrigues, O.M. Bruno, and L. da F. Costa, "An entropy-based approach to automatic image segmentation of satellite images," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 3, pp. 512 – 518, 2011.

[5] R. C. Gonzalez, R.E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Prentice hall, 2003.

[6] "Biostation-IM," http://www.nikoninstruments.com/Vyrobky/Cell-Incubator-Observation/BioStation-IM.
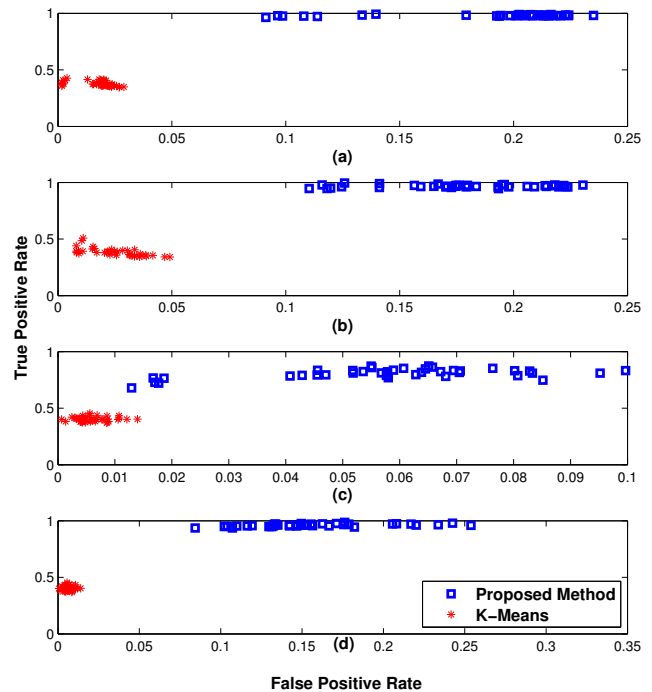
Fig. 5: (a) The results of the video under $10\times$ objective; (b) results of the video under $20\times$ objective; (c) results of the video under $40\times$ objective; (d) result of the video under $40\times$ objective after 5×5 median filter.