

BACKGROUND ESTIMATION FOR MICROSCOPIC CELLULAR IMAGES

Nezamoddin N. Kachouie and Paul Fieguth

Department of Systems Design Engineering
University of Waterloo, 200 University Ave. West, Waterloo, Canada

ABSTRACT

Background modelling is a key task in tracking applications. Our interest in this paper is the accurate estimation of static backgrounds in scientific imaging, such as those in automated stem cell tracking. In this paper, an effective background estimation method is proposed. First, the segmentation results are used to remove the foreground objects, then the background is robustly estimated over the resultant 3-Dimensional residual image sequence. We do spatio-temporal background estimation over a local neighbourhood with a robust trimmed mean. The experimental results generated by the proposed method are quite promising.

Index Terms— Background Estimation, Spatio-Temporal Analysis, Statistical Detection, Tracking, Biomedical Imaging.

1. INTRODUCTION

In most video analysis applications, the goal is tracking one or multiple moving objects over the data stream such as human tracking, traffic control, and medical imaging. Although most of the televised videos involve frequent scene cuts and camera motion, a great deal of imaging such as medical and biological imaging are based on a fixed camera which yields a static background and a dynamic foreground. Moreover, in most tracking problems the dynamic foreground is of interest, hence an accurate removal of the background is desired. Removing the estimated background leaves us with foreground on a plain background. The estimated background might be composed of random temporal noise, temporal illumination variations, spatial distortions caused by CCD camera pixel non-uniformities, and stationary or quasi-stationary clutter and background structures.

There are different methods for background estimation using different image features at each pixel location. In most of them spectral features of each pixel representing gray level intensity or colour information of the pixel have been used to model the background [1, 2]. Some of them have used spatial features to model the local structures of the background image [3, 4]. Methods which employ spatial and spectral

This research has been funded by the Natural Science and Engineering Research Council of Canada (NSERC).

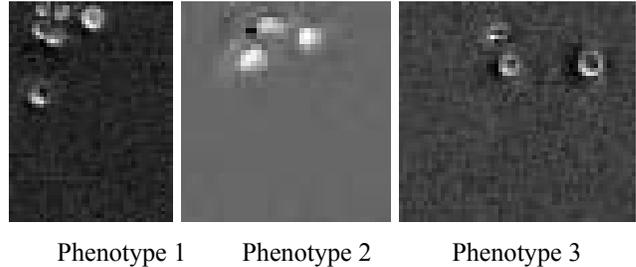


Fig. 1. Different stem cell phenotypes: (Phenotype 1) Bright boundary and dark interior. (Phenotype 2) Uniformly bright. (Phenotype 3) Poor contrast.

features have a good performance when the background image consists of stationary objects with static pixels but they demonstrate a poor performance when the background image consists of non-stationary objects with dynamic pixels. A robust vision system can accurately model the non-stationary elements of the background if it could effectively use the temporal features [1, 5, 6]. Among the methods which use temporal features, Gaussian mixture model has been widely used and performed well to estimate non-stationary temporal background pixel distributions [5]. Different extensions of Gaussian mixture models have been introduced to improve its performance and reduce the running time [7, 8].

2. PROPOSED METHOD

Most tracking problems have an implicit, nonparametric model of the background; by developing a model for the background it is possible to find a classifier that labels each image pixel as background/not background. That is, the foreground is identified as that which is not background. In contrast, our cell tracking problem admits an explicit model of the foreground. We have developed a cell localizing model [9], however because of the low SNR of our problem, it is desired to remove all deterministic non-cell variations in the image (i.e. the background) before localizing the cells. Although cell localization appears to be a foreground/background classifier, there is a difference; we do not need to actually segment the image, only to identify the cell locations. Therefore we do not need to reliably classify each pixel definitively as fore-

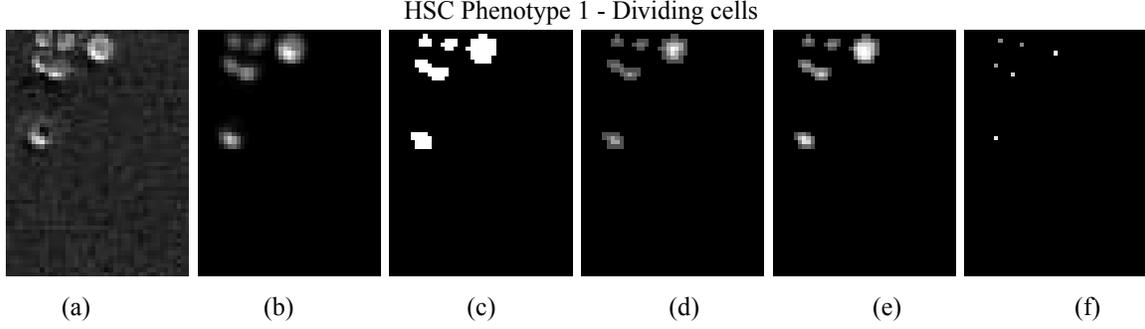


Fig. 2. (a) Original HSC image. (b) Application of a circular mean square template. (c) Classification of circular mean square to cell and background classes by minimizing the inter-class variance. (d) Euclidean distance of cell pixels from the background. (e) Product of Circular mean square and Euclidean distance. (f) Cell center locations after thresholding the maxima of (e).

ground or background, with unavoidable error around the cell margins. Rather to accurately estimate the background we only need to identify most background pixels, most of the time.

3. CELL LOCALIZATION

Let $\mathbf{I} = (I_1, I_2, \dots, I_K)$ be a set of K images, each frame having $N \times L$ pixels. As the foreground cells are essentially outliers relative to the background statistics, to estimate the background we need to identify and remove the foreground, thus we remove the pixels which are associated to the located cells in each frame, specifically all pixels inside a rectangular box with side length $2r_k^m$, centred at (x_k^m, y_k^m) . The image after removing the objects for a typical frame is depicted in Figs. 3(c).

Except for Phenotype 1, the other HSC phenotypes cannot be modelled as an object with dark interior and bright boundary, therefore the proposed method in [9] performs poorly to detect HSCs of Phenotype 2 and Phenotype 3. To design a general method that could be applicable for detecting different HSC phenotypes investigated in this research, some common features among different HSC phenotypes must be extracted. All HSC phenotypes in this work can be characterized as an approximately circular object. Cell pixels have also high intensity variations against a uniform background.

HSCs are modelled as a circular anomaly which is represented by a set of pixels with significant intensity variations against the uniform background. Assuming (x, y) and r as center coordinates and radius of a cell respectively, we construct the set $G(z_k^m, I_k)$, which returns the inside cell pixels

$$G(z, I) = \{I_{ij} | (x - i)^2 + (y - j)^2 \leq (r)^2\} \quad (1)$$

from which we extract the sample mean of square intensities

$$\bar{G} = \frac{\sum_{g \in G} g^2}{|G|} \quad (2)$$

To recognize cells from the uniform background, first (2) is applied to the cell image and \bar{G} is computed. The variance of the pixels $\{g | g \in G\}$ located inside a ring with radius r is

$$\sigma^2 = \frac{\sum_{g \in G} (g - \mu)^2}{|G|} = \frac{\sum_{g \in G} g^2 - \sum_{[1, |G|]} \mu^2}{|G|} \quad (3)$$

after simplification we have

$$\bar{G} = \frac{\sum_{g \in G} g^2}{|G|} = \sigma^2 + \mu^2 \quad (4)$$

Thus for G located in the uniform background we find $\bar{G}_{bkg} = \sigma_{bkg}^2$ whereas for G located inside a cell we have $\bar{G}_{cell} = \sigma_{cell}^2 + \mu_{cell}^2$. For all of the different cell phenotypes one or both of the σ_{cell}^2 , μ_{cell}^2 are significantly higher than those of the background, therefore $\bar{G}_{bkg} \ll \bar{G}_{cell}$ and as a result \bar{G} can be used to detect HSCs in the uniform background by classifying to two classes, cell and background, by minimizing the inter-class variance

$$\sigma^2(T) = l_{cell}(T) \cdot (\bar{G}_{cell}(T) + \mu_{cell}^2(T)) + l_{bkg}(T) \cdot \bar{G}_{bkg}^2(T) \quad (5)$$

where $l_{bkg}(T)$ and $l_{cell}(T)$ are the number of pixels in the background and the cell classes, $\sigma_{bkg}^2(T)$, $\sigma_{cell}^2(T)$, and $\sigma^2(T)$ are variance of background, variance of cell class and inter-class variance considering the threshold (T).

Considering the HSC as a circular anomaly in the proposed method it can be concluded that the cell center has the maximum distance to the cell boundary in comparison with any other pixel in the cell area. Thus to fit a circular shape to the classified anomalous regions, we compute $D(cell_p, bkg_p)$, the Euclidean distance of each anomalous pixel $cell_p = (x_{cell}, y_{cell})$ from its closest background pixel $bkg_p = (x_{bkg}, y_{bkg})$

$$D(cell_p, bkg_p) = \sqrt{(x_{cell} - x_{bkg})^2 + (y_{cell} - y_{bkg})^2} \quad (6)$$

where

$$bkg_p = \arg \left\{ \min_{bkg_p} D(cell_p, bkg_p) \right\} \quad (7)$$

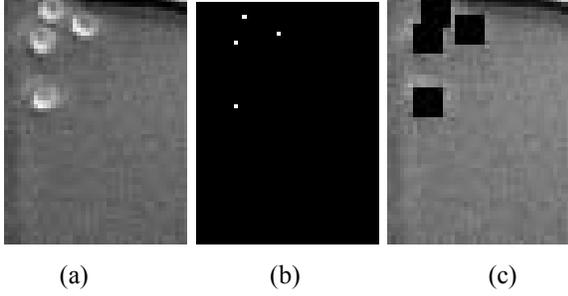


Fig. 3. (a) Original blood stem cell sequence. (b) Cell centre locations obtained by applying the cell model [9]. (c) Original frame after removing the foreground cells in which imperfect segmentation causes slight cell removal failure

We then compute the product of the Euclidean distance map and the circular mean square

$$P_{cell} = D_{cell} \cdot \bar{G} \quad (8)$$

Therefore from pixels with the same circular mean square value in \bar{G} , the one that is located closer to the centroid of a segmented cell region and so has a higher value in D_{cell} , will have higher value in P_{cell} and as a result will be more likely to be a cell center. Finally, to locate the cell centers, we find the local maxima in P_{cell} and then threshold the local maxima map.

4. BACKGROUND ESTIMATION

As the foreground cells are essentially outliers relative to the background statistics, to precisely estimate the background we need to identify and remove the foreground, thus we remove the pixels which are associated to the located cells in each frame, specifically all pixels inside a rectangular box with side length $2r_k^m$, centered at (x_k^m, y_k^m) as depicted in Fig. 3 for Phenotype 1. For each frame I_k we can write

$$I_k = F_k + B + n_k \cdot \mathbf{1} + V_k \quad (9)$$

where F_k is the dynamic foreground, B is the fixed background, n_k models the temporal variations in global lighting, and V_k is spatio-temporal random additive noise. The cells are localized by applying the proposed localization method in Sec. 3 to the original sequence \mathbf{I} , generating a set of located cell centres

$$Z_k = \{z_k^m \mid m \in [1, M_k]\} \quad (10)$$

where M_k is the number of located cells in frame k .

A residual sequence is obtained by removing the localized cells. The residual frame contains all pixels of original frame I_k except those belong to the foreground and which have been removed. The set of pixels to remove is

$$H_k = \{(p, q) \mid p \in [x_k^m - r_k^m, x_k^m + r_k^m], \\ q \in [y_k^m - r_k^m, y_k^m + r_k^m], 1 \leq m \leq M_k\} \quad (11)$$

Each remaining pixel represents a noisy sample of the background:

$$I_{ijk} = \{B_{ij} + n_k + V_{ijk} \mid (i, j) \in L \setminus H_k\} \quad (12)$$

where L is frame lattice of pixels. The temporal noise n_k can be estimated as

$$\hat{n}_k = \underset{\{(i,j) \in L \setminus H_k\}}{\text{mean}} (I_{ijk}) \quad (13)$$

For temporal correction, the estimated temporal noise is subtracted from the residual frame:

$$g_k = I_k - \hat{n}_k \cdot \mathbf{1} \quad (14)$$

The temporally corrected residual sequence $\mathbf{g} = [g_{ijk}]$ satisfies

$$g_{ijk} = \{B + \tilde{n} + V\} \simeq \{B_{ij} + V_{ijk} \mid (i, j) \in L \setminus H_k\} \quad (15)$$

where \tilde{n} is the estimation error. We then precisely estimate the background $B = [B_{ij}]$, consisting of spatially stationary distortions and illumination variations at each pixel location over K frames of temporally corrected residual sequence $\mathbf{g} = (g_1, g_2, \dots, g_K)$:

$$\hat{B}_{ij} = \underset{(p,q,r) \in Q(i,j,d)}{t_mean}^\Omega (g_{pqr}) \quad (16)$$

where t_mean^Ω is the trimmed mean, with trimming parameter Ω , calculated by sorting the values $g_{p,q,r}$, removing the first and last $\Omega\%$, and computing the sample mean over the remaining samples. $Q(i, j, d)$ contains all pixels in the residual sequence which fall inside 3-D spatio-temporal clique C :

$$Q(i, j, d) = \{(p, q, r) \mid (p, q) \in C(i, j, d), (p, q) \in L \setminus H_r\} \quad (17)$$

where $C(i, j, d)$ is

$$C(i, j, d) = \{(p, q) \mid p \in [i - d, i + d], q \in [j - d, j + d]\} \quad (18)$$

Here $d \geq 0$ is the window parameter, which can vary from pixel to pixel, controlling the spatial extent of pixel samples to be used in computing the trimmed mean. Clearly there is a tradeoff here between spatial resolution (smaller $d \implies$ less spatial blurring) and trimmed mean accuracy (larger $d \implies$ more samples). The chose of d is determined by the number of background samples needed to reliably compute the trimmed mean, therefore the selected d is computed as

$$\hat{d}_{i,j} = \min \{d \mid \|Q(i, j, d)\| > T\} \quad (19)$$

where $\|\cdot\|$ counts the number of elements in the set, and the threshold T is selected based on background statistics as

$$T \leq D < [\text{size of clique} = n^2 * K] \quad (20)$$

where a precise estimation of background would be possible. Eventually the estimated spatial illumination variations B is subtracted out from the temporal corrected sequence g as $\hat{F} = \mathbf{g} - \hat{B}$.

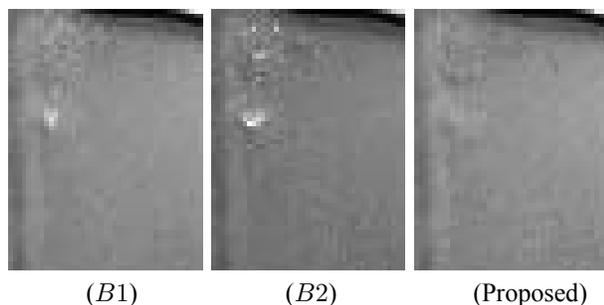


Fig. 4. The estimated background by applying: *B1*, *B2* (The method proposed in [6]), and Our proposed method. The proposed method estimated a uniform background in which cell boundary pixels are not presented, hence restoring the dynamic foreground robustly.

5. RESULTS

We have applied the proposed cell detection and background estimation method to different sequences and different phenotypes of phase contrast HSC images. The results obtained by the proposed method is compared with: i) Spatio-temporal pixel-wise version of [1] that we call it *B1*. ii) The proposed method by *Heikkila and Pietikainen* [6] as the most recent background modelling method with very promising results that we call it *B2*. iii) Frame-difference segmentation method. iv) Morphological averaging background estimation method in [10]. The proposed method outperforms the present background estimation methods including [1, 10], and the most recent background modelling approach based on texture information [6].

The original frames and located cells are depicted in Figs. 2 and 3 for dividing and non-dividing stem cells respectively. As we can observe nondividing and more challenging dividing cells are localized perfectly applying the proposed method. The estimated background images applying the proposed method, *B1*, and *B2* are depicted in Figs. 4(a), (b) and (c). As it can be observed in Figs. 4(a) and (b), *B1* and *B2* methods fail to precisely estimate the background in the spatial locations where cells have slow motion dynamics. As a result cell boundary pixels are visible in the estimated background by these methods. In contrast, as we can observe in the estimated background by the proposed method, not only the well boundaries are precisely estimated, but there are very smooth variations over the background image. The proposed method also estimates background in locations where cells have slow motion dynamics precisely as we can observe in 4(c).

6. CONCLUSIONS

A novel algorithm for cell detection/background estimation is proposed. The proposed method employs the detection re-

sults to remove the foreground objects and estimates the background over 3-D residual sequence. The proposed method is applied to different HSC image sequences and generated promising results. Using temporal detection information, the proposed method outperforms the present background estimation methods including the most recent background modelling based on texture information [6]. The future work focused to design a recursive foreground segmentation-background estimation version of the proposed method in which the segmentation and estimation results will recursively be used to improve the performance of each other.

7. REFERENCES

- [1] C.R. Wren, A. Azarbajejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Tran. on PAMI*, vol. 19, no. 7, pp. 780–785, 1997.
- [2] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction in proc. eur. conf. , 2000.," in *ECCV*, 2000.
- [3] J. Kato, T. Watanabe, S. Joga, and A. Blake, "An hmm-based segmentation method for traffic monitoring movies," *IEEE Tran. on PAMI*, vol. 24, no. 9, pp. 1291–1296, 2002.
- [4] L. Li and M. Leung, "Integrating intensity and texture differences for robust change detection," *IEEE Transactions on Image Processing*, vol. 11, pp. 105–112, 2002.
- [5] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE CVPR*, 1999, pp. 246–252.
- [6] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Tran. on PAMI*, vol. 28, no. 4, pp. 657–662, 2006.
- [7] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *IEEE ICPR*, 2004, pp. 28–31.
- [8] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," *IEEE Tran. on PAMI*, vol. 27, no. 5, pp. 809830, 2005.
- [9] N. N. Kachouie, P. Fieguth, J. Ramunas, and E. Jervis, "Probabilistic model-based cell tracking," *International Journal of BioImaging*, vol. 2006, pp. 1–10, 2006.
- [10] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel, "Toward robust automatic traffic scene analysis in real-time," in *International Conference Pattern Recognition (ICPR)*, 1994, pp. 126–131.