A Statistical Framework for Estimation of Cell Migration Velocity

Guangming Huang Jiwoong Kim Xinyu Huang Gaolin Zheng Alade Tokuta
Department of Mathematics and Computer Science
North Carolina Central University
1801 Fayetteville Street
Durham, NC 27707, USA
{ghuang, jkim7, huangx, gzheng, atokuta}@nccu.edu

ABSTRACT

Migration velocity of cell populations *in vitro* is one of important measurements of cell behaviors. As there are massive amount of cells in one image that share similar characteristics and are highly deformable, it is often computational expensive to track every individual cell. It is also difficult to track cells over a long period of time due to propagation of segmentation and tracking errors. This paper presents an algorithm to estimate migration velocity of cell populations observed by time-lapse microscopy. Instead of tracking cells individually, our proposed algorithm computes mutual information between image blocks of consecutive frames. The migration velocity is then estimated by a linear regression, with mutual information and foreground area ratio as input. Experiments on a variety of image sequences verified that our algorithm can give accurate and robust estimation under different situations in real-time.

Keywords

Cell Migration, Mutual Information, Linear Regression.

1. INTRODUCTION

It is important to measure cell migration velocity in many biomedical applications, such as wound healing assay of cell monolayers [YPW+04] and analysis of red blood cell in microcirculation [WZH+09]. For cell populations, there are mainly two obstacles to estimate the migration velocity accurately and robustly. First, all the cells in a population have very similar characteristic, such as shape and intensity. Second, cells are often highly deformable. For example, two cells could merge into one cell, and one cell could divide to two or more cells. As a result, it could be difficult and computational expensive to track every cell in image sequences. Figure 1 shows three examples of cell populations. Cells in some types of images even have very similar intensities to the background as shown in figure 1(b). Thus, the segmentation algorithms based on intensity values would easily fail.

In this paper, we present an efficient and novel algorithm to estimate cell migration velocity. Our algorithm first computes mutual information and foreground ratio between image blocks of two consecutive frames. A linear regressor is then trained and applied to estimate migration velocity.

Mutual information has been widely used to align two images in many medical applications [PMV03] in order to reduce the error during the image acquisition (*e.g.,* finger jigging). However, to our knowledge, there is no work that uses mutual information as an input variable of regression for the estimation of cell migration velocity. As there are no individual cells involved in the estimation process, our algorithm contains no accumulated segmentation and tracking errors.

Therefore, the proposed algorithm has several advantages. First, accurate segmentation and data association of cell contours are not required. Second, it can be performed in real-time without using motion trackers for all the cells. Third, since tracking accuracy is not an issue (*e.g.,* there are no accumulation errors), it can be used to estimate migration velocity for a long period.

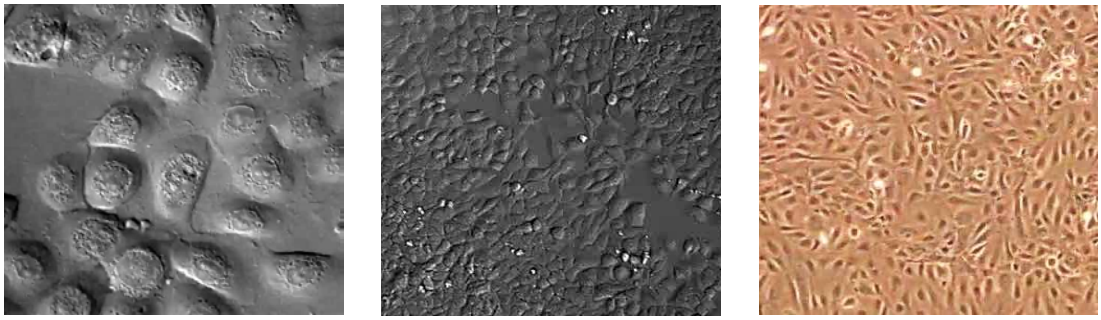


Figure 1. Examples of microcopy images of cell populations: (*left*) primary keratinocytes [Cel09]. (*middle*) cancer cells captured spinning disc confocal microscope [Mar09]. (*right*) human umbilical vein endothelial cells (HUVEC) [Yam09].

The remainder of this paper is organized as follows. Section 2 describes related work. Our proposed algorithm is given in section 3. Section 4 shows the experiments using three different datasets. The conclusion is given in section 5.

2. RELATED WORK

Wu *et al.* proposed an approach to measure velocity of red blood cell from capillary video using the optical flow technique [WZH+09]. In order to apply the optical flow technique, the skeleton of vessel needs to be extracted first based on a set of pre-processing steps, such as connected component labeling, thinning, and length pruning. These pre-processing processes may not be applied to other types of cells in general due to occlusions, deformations, and even varying illuminations. More importantly, the velocity determination in this approach is based on two assumptions: a) intensity of each cell does not change over time; b) the surrounding area of the cell move in a similar manner. These two assumptions are fundamental to apply the optical flow on the skeletons. They however, are too restricted and cannot be extended in general. Other approaches [DSA+08, LGM04] that based on detection and graph extraction of vessel shapes also have similar problems.

In [LMC+08], Li *et al.* proposed an automated tracking algorithm to track hundreds to thousands of cells and construct lineages simultaneously. This tracking system first segments candidate cell regions and tracks them over frames by forming a minimization problem with a topological constraint. Then this system predicts and filters the cell motion dynamics using interacting multiple model filters, and construct lineages by checking the entire tracking history. The proposed system can be used to analyze a number of cell behaviors including migration, division, death, and so on.

According to [MDI+09], tracking in cell can be divided into two stages, segmenting individual cells and connecting cells over time. However, since each possible candidate cell needs to be considered, the whole process could be computational expensive. For the purpose of estimating migration velocity, the algorithms based on tracking individual cells could be complicated and hence may not be the best choice. Moreover, common used segmentation algorithms based on intensity values could easily fail to distinguish between background areas and candidate cells.

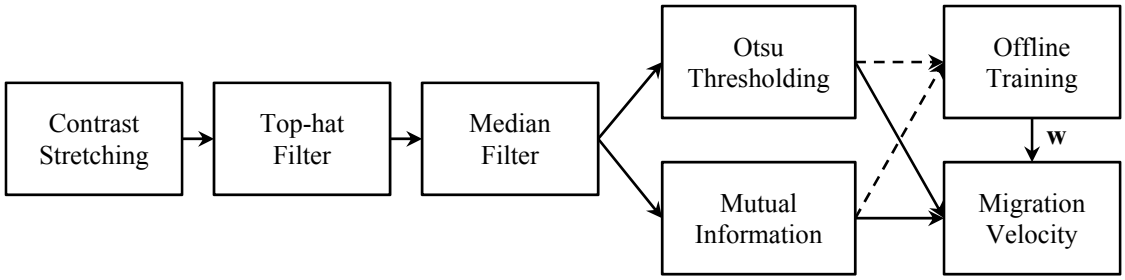


Figure 2. Overview of our methodology

3. METHODOLOGY

Our algorithm can be divided into three modules, 1) image enhancement and foreground detection; 2) computation of mutual information between image blocks; 3) velocity estimation using linear regression. Figure 2 gives an overview of the algorithm.

3.1 Image Enhancement and Foreground Detection

The purpose of image enhancement is to reduce noise and enhance the image contrast. For simplicity, we assume two image frames has been aligned. This can be achieved by an affine transformation based on an estimated homography [HZ04], or a non-linear transformation using mutual information [PMV03].

First, we enhance the image by a linear contrast stretch that enlarges the range of intensity values to the entire available range. If the cells appear darker than the background, the images are inverted so that the cell interior areas are brighter. Next, the enhanced image is convolved with a 7×7 top-hat filter, which is often used to detect bright features on a dark background. A 3×3 median filter is then applied twice to remove small grey objects that could be noise or artifacts caused by the top-hat filter. The output image after the median filter is the input image for the computation of mutual information. We further detect the foreground by using Otsu thresholding [Otu79] to obtain a binary image where the white pixels indicate the foreground. Figure 3 visualizes these image processing steps.

Unlike many algorithms that require fairly accurate foreground and cell detections, we only need to detect a rough foreground as shown in Figure 3(f). Moreover, there are no accumulated segmentation errors.

3.2 Computation of Mutual Information

Mutual information is usually used to test the independence between two random variables **x** and **y**. If two random variables are independent, the joint probability $p(\mathbf{x}, \mathbf{y})$ can be factorized into the product of their marginals $p(\mathbf{x})p(\mathbf{y})$. In our application, the random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are two same image blocks from two consecutive image frames.

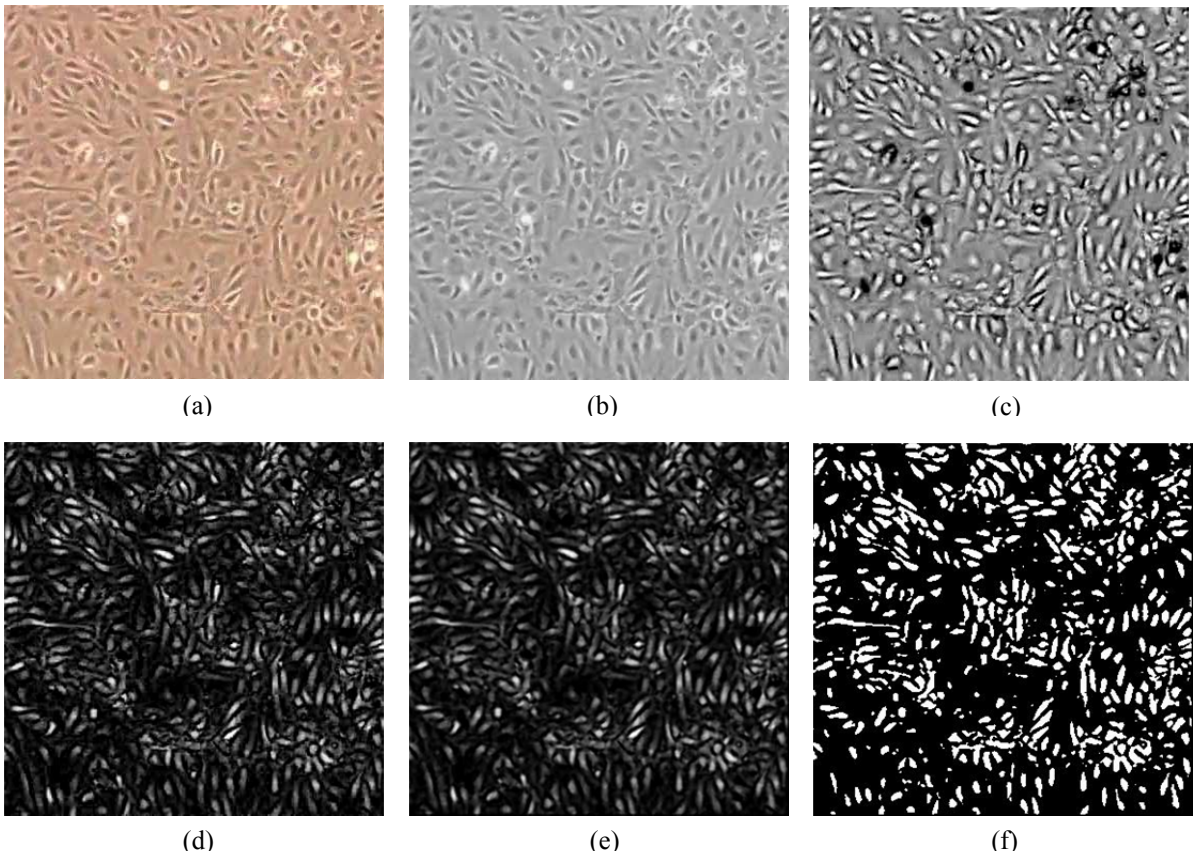


Figure 3. Image enhancement and foreground detection. (a) Original color image. (b) Grayscale image. (c) Result after intensity inversion and contrast stretching. (d) Result of top-hat filtering. (e) Output of median filtering. (f) Foreground detection by Otsu thresholding.

We do not use the entire image frames as the random variables. This is mainly because that the majority of an image could be the background, in which case, when the entire images are used, the mutual information between them could mainly reflect the differences between backgrounds. Furthermore, when the captured image has a high resolution, it also could be inefficient since the memory storage is increased. Therefore, we divide each frame into a set of image blocks with overlapping areas. The size of the image block is determined by the maximum of cell migration velocity, which can be easily estimated by the visual inspection. For many types of cells, the size of an image block is often around a few times that of a single cell.

Let us denote an image block as **x** and the same image block in the next image frame as **y**. The mutual information between the variables **x** and **y** is defined as

$$I(\mathbf{x},\mathbf{y}) = \mathrm{KL}(p(\boldsymbol{x},\boldsymbol{y})||p(\mathbf{x})p(\mathbf{y})) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(x_i,y_j)\ln\left(\frac{p(x_i,y_j)}{p(x_i)p(y_j)}\right)$$

where KL(·) is known as the Kullback-Leibler divergence, and $p(\mathbf{x})$ and $p(\mathbf{y})$ are histogram distributions of two image blocks. We can see that $I(\mathbf{x},\mathbf{y}) = 0$ if and only if two image blocks are independent, which indicates a very large migration. The larger the $I(\mathbf{x},\mathbf{y})$, the more similar two image blocks are, which indicate a small migration.

3.3 Velocity Estimation using Linear Regression

For each pair of image blocks **x** and **y**, we describe the cell migration using two features: the mutual information *I* and the difference between foregrounds of **x** and **y**. The mutual information *I* measures independence between two image blocks including both foreground and background. The difference between two foregrounds is measured by the ratio of non-overlapping foreground to the union of the two foregrounds, which is defined by

$$D = \frac{|\mathbf{x}_F \cup \mathbf{y}_F - \mathbf{x}_F \cap \mathbf{y}_F|}{|\mathbf{x}_F \cup \mathbf{y}_F|}$$

where subscript *F* indicates the foreground. In general, the smaller the ratio, the more similar two image blocks are. Thus, the inputs for the regression are the mutual information *I* and area ratio *D*, and the output is the cell migration velocity *v*. Since there could exist multiple moving cells in one image block, the maximum velocity is chosen as output.

In order to avoid over-fitting problem, we only consider the second order of the inputs. Therefore, the possible variables include I, D, I^2, D^2, and ID. We adopt the *forward selection* to select a suitable model for the data. In this model selection approach, we add one variable that results in the largest reduction in sum squared errors (SSE), and then carry out a hypothesis test to determine whether this reduction in SSE is significant. If the reduction is significant, we continue the adding process and stop otherwise. The results show that the full model is the most suitable regression model when the area ratio can be robustly estimated. The regression model is given by

$$v(I,D,\mathbf{w}) = [1 \;\; I \;\; D \;\; I^2 \;\; D^2 \;\; ID]^T\mathbf{w}$$

where $\mathbf{w} = (w_0,\dots,w_5)^T$.

In order to estimate the parameters **w**, we first developed an interactive user interface to measure the velocities of a set of sample cells ($N \approx 30$) by manually clicking centroids of the same cells in two image frames. Then the parameters **w** can be estimated by the normal equations

$$\mathbf{w} = (\boldsymbol{\phi}^T\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\boldsymbol{v}$$

where $\boldsymbol{\phi}$ is $N \times 6$ design matrix given by

$$\boldsymbol{\phi} = \begin{pmatrix} 1 & I_1 & D_1 & I_1^2 & D_1^2 & I_1D_1 \\ 1 & I_2 & D_2 & I_2^2 & D_2^2 & I_2D_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & I_N & D_N & I_N^2 & D_N^2 & I_ND_N \end{pmatrix}$$

It could be very difficult to estimate the area ratio *D* for some types of microcopy images. For example, for the image type shown in figure 1(b), the intensities of both foreground and background are very similar and illumination conditions also change when cells are moving. Therefore, the most exiting cell segmentation algorithms based on intensity values would fail to detect foreground and moving cells accurately.

For these types of images, we discard the area ratio *D* in our regression model and only use the mutual information. Thus, the model could be changed to a polynomial regression with order 3, which is given by

$$v(I,\mathbf{w}) = [1 \;\; I \;\; I^2 \;\; I^3]^T\mathbf{w}$$

where $\mathbf{w} = (w_0,\dots,w_3)^T$. The normal equations remain same and the $N \times 4$ design matrix is defined by

$$\boldsymbol{\phi} = \begin{pmatrix} 1 & I_1 & I_1^2 & I_1^3 \\ 1 & I_2 & I_2^2 & I_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & I_N & I_N^2 & I_N^3 \end{pmatrix}$$

During the prediction stage, we divide each image frame into $n \times m$ blocks and estimate the migration velocity for each block using the linear regression. We can then plot the velocity distribution over time. The expectation of the migration velocity can also be estimated for each frame by

$$E(v) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}$$

where v_{ij} is the velocity in each block.

4. EXPERIEMNTS

4.1 Datasets

Our proposed algorithm is tested on three microcopy image sequences. A few frames of the same cell type are used for estimation of parameters **w**.

Dataset A has one sequence of human umbilical vein endothelial cells (HUVEC) that are isolated from normal umbilical vein [Cel09]. The images are cropped to 474 × 364.

Dataset B includes image sequences that show cell migration of primary keratinocytes before and after calcium switch [Yam09]. These images have a dimension cropped to 642 × 449 pixels.

Dataset C contains one image sequence taken overnight by using a spinning disc confocal microscope [Mar09]. It shows motility of cancer cells. This image sequence has 111 frames with 472 × 360 pixels/frame.

4.2 Estimation of Parameters

The parameters **w** for the linear regression are learned from the training data. For each dataset, the parameters are learned using a set of samples marked by our interactive graphic interface. The image block size is set to 64 × 64, which is around 2 times larger than a typical cell length The training sets include 20~40 samples from each dataset. Figure 4-7 show the scatterplot matrix of the 34 training samples from dataset A, and 25 training samples from dataset C. It is easy to see the strong correlation among mutual information, area ratio, and migration. The training samples from the dataset B have the distribution similar to the figure 4 and 5 from the dataset A.

The area ratio distribution in figure 7 is more dispersed than the distribution shown in figure 5. This is mainly because that the cell intensity is very close to the background intensity as shown figure 1(b). The segmentation based on intensity tends to fail. In this case, the regression with only mutual information can provide a more stable result. Table 1 gives a summary of parameters for dataset A and C.

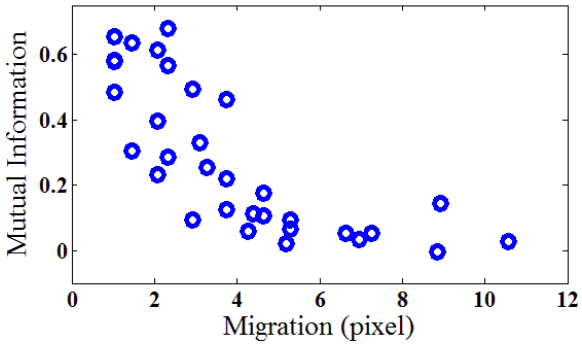


Figure 4. Mutual information *I* versus migration using 34 training samples from dataset A.

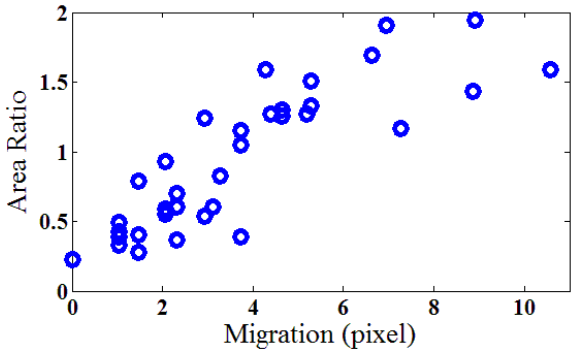


Figure 5. Area ratio *D* versus migration using 34 training samples from dataset A.

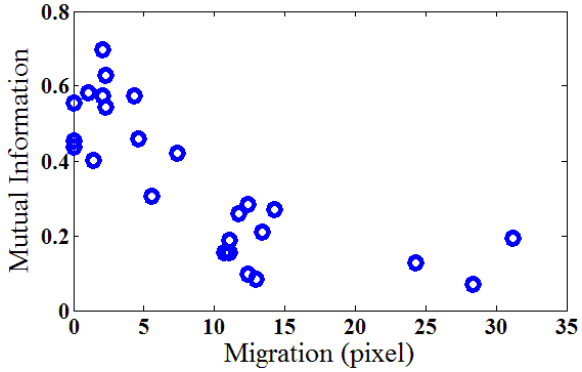


Figure 6. Mutual information *I* versus migration using 25 training samples from dataset C.

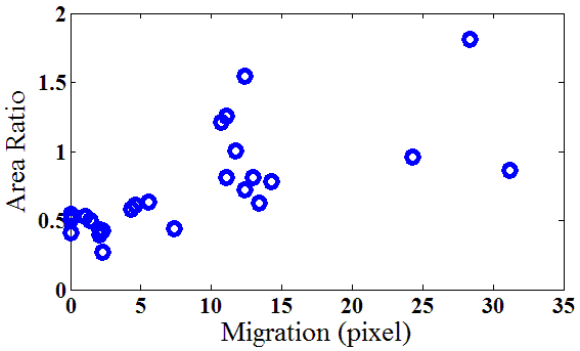


Figure 7. Area ratio D versus migration using 25 training samples from dataset C.

Table 1. Summary of parameters for dataset A and database C.

DB	w_0	w_1	w_2	w_3	w_4	w_5	Block Size
A	0.202	0.246	0.039	0.349	0.016	0.031	64×64
C	0.114	0.041	0.019	0.010	N/A	N/A	64×64

4.3 Estimation of Migration Velocity

After the training stages, we compute the migration velocity over the whole range of each dataset.

Figure 8 shows the result from the dataset A. We computed migration velocity using our algorithm for every 50 frames over 4271 image frames captured over 70 hours and 37 minutes. This velocity distribution with the bell shape is same as our expectation. At the beginning of the captured image sequences, the number of cells is relatively small and the cell growth rate is high. After the growth rate reaches its peak, the cells are very crowded and touch each other in the limited space. As a result, the growth rate decreases. Figure 9 shows two image frames at the beginning and at the end of the dataset A.

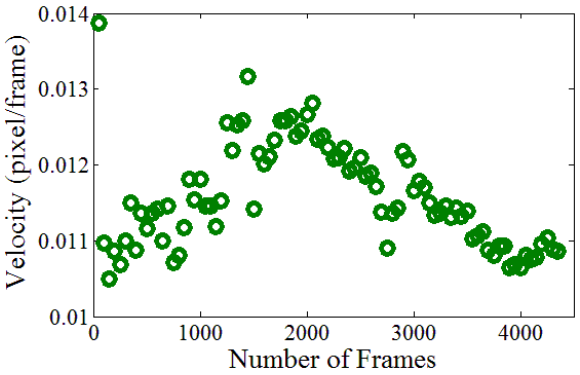
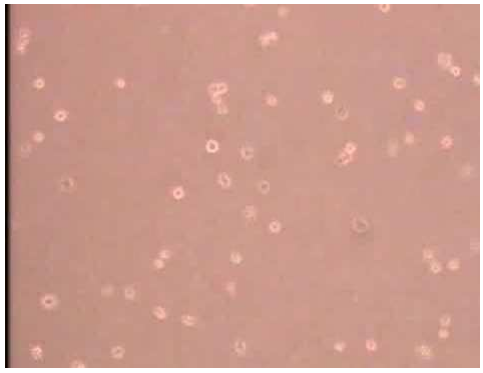
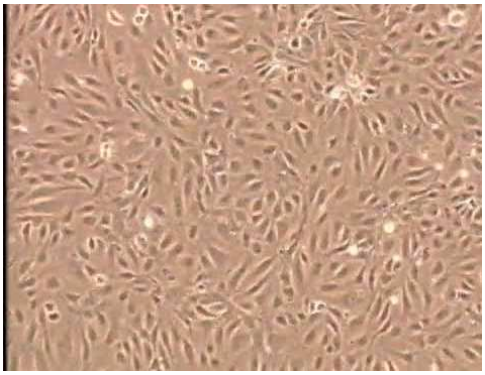


Figure 8. Velocity distribution for the dataset A.



(a) 92th Frame at 01m00s



(b) 4,250th Frame at 68h46m30s

Figure 9. Two image frames from the dataset A that could be used to further verify the cell growth rate.

For the dataset B, as the concentration of calcium was increased from low to high, cell migration velocity is prevented by the maturation of cell-cell adhesion after the calcium switch. Therefore, the velocity decreases shown in figure 10 is also same as our expectation. Here we take every 3 frames to compute velocity over 73 frames that are captured using 6 hours.

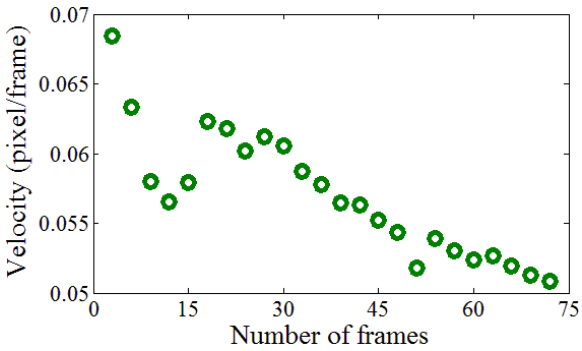


Figure 10. Velocity distribution for the dataset B.

Figure 11 shows the velocity distribution for the dataset C over 111 frames using more than 18 hours.

We sample every 10 frame. As the population of cancer cells gradually reaches its peak in the limited space, the velocity also decreases gradually. We further plot the mutual information distribution for every frame in order to verify our results. Figure 12 shows the distributions and the corresponding image frames. It is clear to see the distributions of mutual information are very similar to the distribution of the migration velocity.

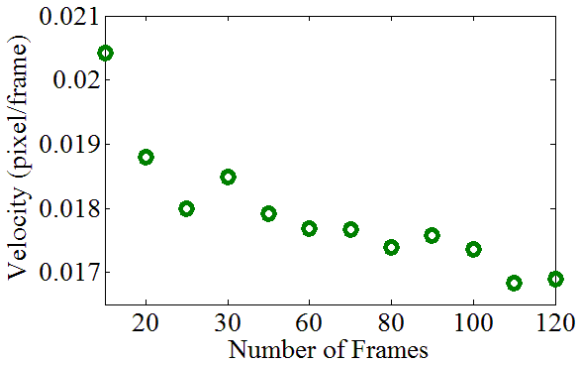
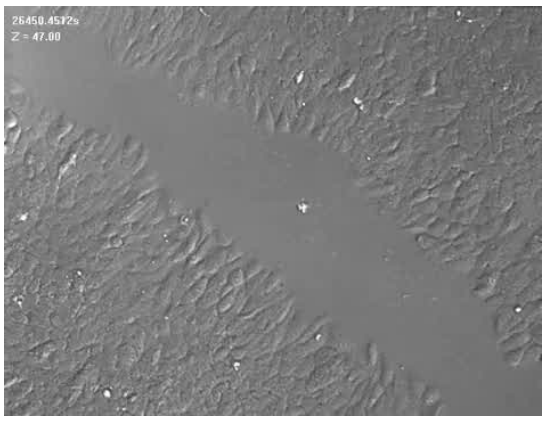
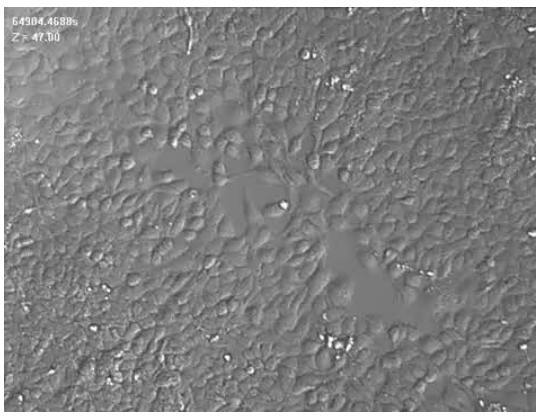


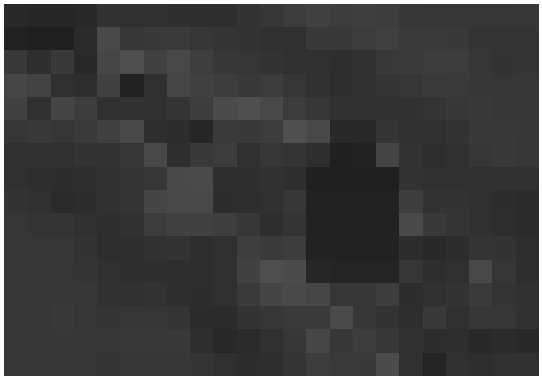
Figure 11. Velocity distribution for the dataset C.



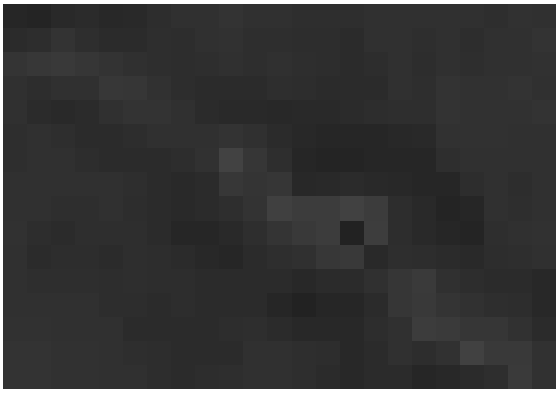
(a)



(b)



(c)



(d)

Figure 12. Images and distributions of mutual information. (a) 45th frame of dataset C. (b) 109th frame of dataset C. (c) distribution of mutual information corresponding to (a). (d) distribution of mutual information corresponding to (b). (c) and (d) are enhanced for the purpose of visualization. The brighter a region, the larger the mutual information is.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel algorithm to estimate cell migration velocity. As individual cell segmentation and tacking are avoided, this algorithm is efficient and robust. Our experiments show that this algorithm is also accurate and can be used to measure cell motility over a long time. In the future, we would like to extend our work to estimation of cell division, merging, and growth based on our regression framework.

6. ACKNOWLEDGMENTS

The authors acknowledge support of the National Science Foundation HRD 0833184.

7. REFERENCES

[Cel09] Cell Applications, Inc., Human Endothelial Cells, http://www.cellapplications.com/endothelial-cells.

[DSA+08] J. G. G. Dobbe, G. J. Streekstra, B. Atasever, R. van Zijderveld, and C. Ince, Measurement of functional microcirculatory geometry and velocity distributions using automated image analysis. Journal of Medical and Biological Engineering and Computing, Volume 46, Number 7, Pages 659-670, 2008.

[HZ04] Hartley, R. and Zisserman, A., Multiple View Geometry in Computer Vision, 2nd Edition, Cambridge University Press, ISBN: 0521540518, 2004

[LGM04] Lamberti, F., Gamba, A., and Montrucchio, B, Computer-assisted analysis of in-vitro vasculogenesis and angiogenesis processes. Journal of WSCG, Volume 12, Number 1-3, Pages 237-244, 2004.

[LMC+08] Kang Li, Eric D. Miller, Mei Chen, Takeo Kanade, Lee E. Weiss, and Phil G. Campbell, Cell population tracking and lineage construction with spatiotemporal context. Journal of Medical Image Analysis, Volume 12, Issue 5, Pages 546-566, 2008.

[Mar09] Charles Marcus, Cancer cell migration and movement, http://marcuslab.harvard.edu/index.shtml, Department of Physics, Harvard University.

[MDI+09] E. Meijering, O. Dzyubachyk, I. Smal, and W.A. van Cappellen, Tracking in Cell and Developmental Biology. Seminars in Cell and Developmental Biology, Volume 20, Issue 8, Pages 894-902, 2009.

[Otu79] Otsu, N., A threshold selection method from gray level histograms. IEEE Transactions on Systems, Man, and Cybernetics - Part B, Volume 9, pages 62–66, 1979.

[PMV03] Pluim, J.P.W., Maintz, J.B.A., and Viergever, M.A., Mutual-information-based registration of medical images: a survey, IEEE Transactions on Medical Imaging, Volume 22, Issue 8, Pages 986-1004, 2003.

[WZH+09] Chih-Chieh Wu, Geoffrey Zhang, Tzung-Chi Huang, and Kang-Ping Lin, Red blood cell velocity measurements of complete capillary in finger nail-fold using optical flow estimation. Journal of Microvascular Research, Volume 78, Issue 3, Pages 319-324, 2009.

[Yam09] Soichiro Yamada, Keratinocyte migration, http://yamadalab.ucdavis.edu/, Biomedical Engineering, University of California at Davis.

[YPW+04] Justin C Yarrow, Zachary E Perlman, Nicholas J Westwood, and Timothy J Mitchison, A high-throughput cell migration assay using scratch wound healing, a comparison of image-based readout methods, BMC Biotechnology, Volume 4, Number 1, 21, 2004