

PRECISION CELL BOUNDARY TRACKING ON DIC MICROSCOPY VIDEO FOR PATCH CLAMPING

John Lee and Christopher J. Rozell

Georgia Institute of Technology
{john.lee,crozell}@gatech.edu

ABSTRACT

One method of patch clamping on brain tissue slices *in vitro* requires a human operator to visually track a cell's boundary and delicately make contact with a cell's membrane using a micropipette's tip. This type of patch clamping may be automated with computer vision methods; yet this is challenging since it requires precision cell-boundary tracking in the presence of heavy noise and interference. In this work, we present a cell-boundary tracking computer vision system which employs a novel deconvolution algorithm specifically created for this application. The deconvolution algorithm was designed to exploit static and dynamic structure in the cell's edges using a reweighted edge-sparsity prior. Quantitative results on simulated data demonstrate the superiority of the proposed algorithm against previous state-of-the-art algorithms. Lastly, the algorithm is applied on real patch clamping video data and qualitative results are discussed.

Index Terms— Differential interference contrast (DIC) microscopy, cell segmentation and tracking, automated patch clamping, deconvolution, sparse dynamical signal estimation

1. INTRODUCTION

Patch clamping is an important experimental technique of taking high-fidelity electrophysiological measurements from single cells; not only is it considered the "gold" standard for recordings of electrical potentials, it also allows for intercellular access (i.e., extracting or inserting chemical and biological material from it). One method of patch clamping [1] requires a micropipette to be positioned adjacent to a cell's membrane, for a tight seal to be formed using suction between its tip and the cell's membrane. The sheer laboriousness of such manual patch clamping makes automating this process highly advantageous. Recent work on *in vivo* patch clamping has had some success in using a robotic actuator to maneuver the pipette to form the seal [1]. We seek to extend this method into the important domain of *in vitro* patch clamping of brain slices by enhancing it with visual guidance, since microscopy imagery is typically available here.

In this work, we propose a computer vision (CV) system for this application. The patch clamping process is extremely delicate and the micropipette has to arrive just at the cell's boundary, and within tolerances of the robotic actuator ($1 - 2\mu m$). The objective of the CV system is therefore to precisely localize the cell's membrane at each frame of differential interference contrast (DIC) microscopy streaming video, throughout the entire patch clamping process. In

particular, the CV system was developed using the *in vitro* patch clamping of cells found in acute brain slices that were prepared from adult (P50-P180) mice as described in [1]. The two major challenges to the CV system on such data were identified: precision cell segmentation in the presence of heavy noise and interference due to the inherent organic material in such tissue slices, and segmentation interference due to the presence of the micropipette when it is in close proximity to the targeted cell.

It is typical for *in vitro* brain slice electrophysiology [1] to use DIC microscopy since it is an intrinsic contrast microscopy that does not cause photo-bleaching and phototoxicity, in contrast to exogenous contrast methods like fluorescence microscopy [2]. Cell segmentation on DIC microscopy is especially difficult due to the presence of DIC optical artifacts. Prior work in the DIC cell segmentation literature were found unsuitable in meeting the needs of this particular application. For example, traditional image processing methods for segmentation were applicable only on imagery with extremely low noise [3, 4, 5]. The more sophisticated method of deconvolution was sought since it was able to handle high-noise application such as ours. Three promising and recent variants of DIC deconvolution algorithms found in [6, 7] exploited structures of smoothness, sparsity, and dynamics for non-linear deconvolution, but they too were not capable of handling the heavy organic interference experienced in this data (this evidenced in section 3).

A main contribution of this paper is a novel cell segmentation and boundary tracking algorithm (in the pixel domain), which extends the framework provided by [6] to meet the specific demands of this setup. The proposed algorithm is formulated as an iterative alternating algorithm that incorporates three key features: a filtering mechanism to handle organic tissue interference, a regularized least-squares optimization with a robust edge-sparsity regularizer that integrates dynamic edge-tracking capabilities, and an inpainting framework that facilitates concurrent removal of the pipette with cell reconstruction.

2. PATCH CLAMPING CELL BOUNDARY TRACKER SYSTEM

2.1. System Design

Non-linear deconvolution processing is a bottleneck in terms of computational complexity and therefore the method of template matching (TM) was employed to effectively reduce the processing area from the entire video frame (e.g. 512×512) to a cell patch (e.g. 64×64); in addition, TM also provides a first order approximation of spatial coordinates of the cell for further refinement (during the deconvolution/segmentation stage). Next, pipette detection and removal are used as strategies to deal with pipette interference. Detection of the pipette was performed via TM, followed by an align-

This work was partially supported by NSF grant CCF-1409422, James S. McDonnell Foundation grant 220020399, and DSO National Laboratories of Singapore. DIC microscopy data courtesy of I. Kolb and C. Forest.

ment of its mask location with the cell patch. Finally, removal is performed via a deconvolutional inpainting method, which will be described in section 2.3. Figure 1 presents a sequential flow of the processes described in this section.

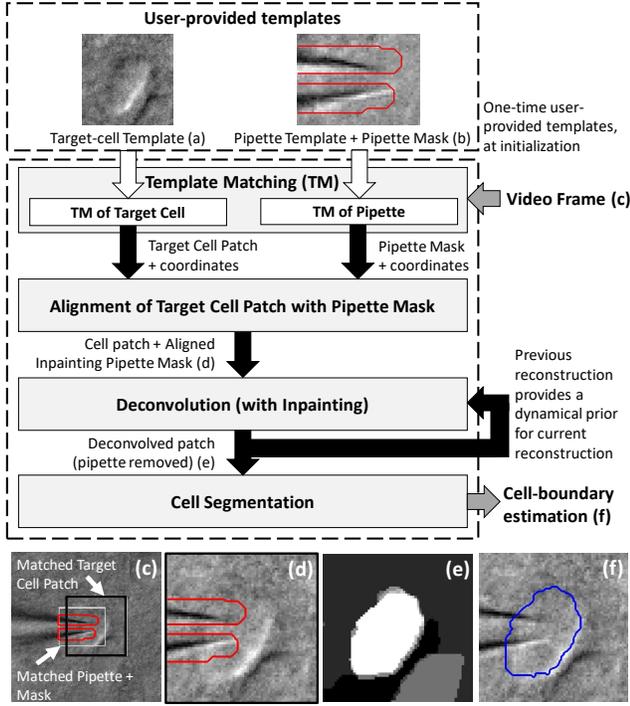


Fig. 1. A block diagram of a cell-boundary tracking system is presented along with examples from real data (a-f). To initialize, a user provides templates of the target-cell (a) and the pipette (b) from a still video frame. At each frame, template matching is applied to localize the cell and reduce a larger video frame (c) to a smaller patch. The matched cell image patch and pipette mask is also prepared for inpainting by lining them up in (d). Finally, deconvolution is applied (e) and segmented (f) using a global thresholding method such as Ostus method to produce a cell-boundary estimation.

2.2. Template Matching and Mask Alignment

At initialization, static templates of the target-cell and the pipette are identified and extracted from a still video frame. This may be done manually by having a user draw bounding boxes around the required objects from a graphical user interface. We note that the initialized templates will be unique to the given shear angle (set by the DIC microscope’s analog optical settings) and is assumed static. In this application, we also assume that the pipette is a rigid object that morphologically undergoes only translation with respect to the imaging plane. Therefore, we may exploit an initial pipette mask that is static relative to its template (Fig. 1(b)), to locate an updated mask of the pipette at subsequent frames using TM’s found coordinates (Fig. 1(c-d)).

The template matching stage performs two brute force searches across the video frame to find the two patches \mathbf{Y}_{cell} , $\mathbf{Y}_{\text{pipette}}$, that are most similar to the target-cell template \mathbf{T}_{cell} and the pipette template $\mathbf{T}_{\text{pipette}}$ respectively. Formally, the template matching returns

the candidate patch $\mathbf{Y} = \mathbf{C}_{(x,y)}$ using the coordinates obtained via:

$$(x, y) = \arg \max_{(x,y) \in \Gamma} \text{sim}(\mathbf{t}, \mathbf{c}_{(x,y)}),$$

where $\mathbf{t} = \text{vec}(\mathbf{T})$ is the template vectorized, $\mathbf{c}_{(x,y)} = \text{vec}(\mathbf{C}_{(x,y)})$ is a candidate patch taken at spatial coordinates (x, y) , Γ is the search space over the image, and the similarity metric is given by normalized cross-correlation as

$$\text{sim}(\mathbf{t}, \mathbf{c}) = \left\langle \frac{\mathbf{t} - \mathbf{t}_\mu}{\|\mathbf{t} - \mathbf{t}_\mu\|_2}, \frac{\mathbf{c} - \mathbf{c}_\mu}{\|\mathbf{c} - \mathbf{c}_\mu\|_2} \right\rangle,$$

with the vector means denoted with subscript μ (i.e., $\mathbf{x}_\mu = \frac{1}{N} \sum_{i=1}^N x_i$). Experiments validated that TM (using normalized cross-correlation) is sufficiently robust for our purposes since neither the cell nor the pipette undergo radical morphological deformations.

To facilitate inpainting, a mask alignment is carried out to line up the pipette’s mask with respect to the target cell patch coordinates. An example alignment is demonstrated using figures 1(c) and 1(d).

2.3. Deconvolution and Inpainting Algorithm

An algorithm called Pre-filtered Reweighted Total Variation Dynamic Filtering (PF+RWTVD-DF) is proposed to tackle segmentation in patch clamping applications for DIC imagery of this type of brain slices specimens by simultaneously performing deconvolution and inpainting.

We begin by describing the deconvolution part. Deconvolution reconstructs the optical path length (OPL) image, defined as the product of the path distance of light through the system and the index of refraction of its medium [8], that has been blurred by the microscope’s point spread function. It is decomposed into two subparts: a linear pre-filtering (PF) mechanism to reduce interference, and a non-linear optimization to estimate the OPL image. Formally, it is given as an alternating iterative algorithm:

$$\begin{aligned} \hat{\mathbf{x}}_k^{(t)} &= \arg \min_{\mathbf{x}_k^{(t)}} \frac{1}{2} \|\mathbf{f}_k * \mathbf{y}_k - \mathbf{D}\mathbf{x}_k^{(t)}\|_2^2 + \\ &\gamma_0 \sum_i |\gamma_k^{(t)}[i] (\mathbf{T}\mathbf{x}_k^{(t)})[i]| \quad \text{s.t.} \quad \mathbf{x}_k^{(t)} \geq 0 \quad (1) \\ \gamma_k^{(t+1)}[i] &= \frac{\kappa + 1}{\kappa |(\mathbf{T}\hat{\mathbf{x}}_k^{(t)})[i]| + |g_k((\mathbf{T}\hat{\mathbf{x}}_{k-1})[i])| + \eta}, \end{aligned}$$

where $\mathbf{x}_k^{(t)} \in \mathbb{R}^N$ is a vectorized version of the OPL image that is to be approximated at frame k with t being the algorithmic iteration number, $\mathbf{y}_k \in \mathbb{R}^N$ is a vectorized form of the bias removed (as described in [6]) observation image from the microscope, \mathbf{f}_k is a pre-filter, $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a Toeplitz implementation of a convolution against a DIC microscopy point spread function (PSF) \mathbf{d} such that $\mathbf{D}\mathbf{x} \equiv \mathbf{d} * \mathbf{x}$, \mathbf{T} is a total variation operator, γ_k are individual weights on the pixels, $g(\cdot)$ is a Gaussian blurring operator, γ_0 and κ are user-tunable constants, and η is some small constant to prevent a division by zero.

The pre-filter \mathbf{f}_k is designed to remove interference caused by organic substances in the brain slice; such interference was observed to display spectral characteristics, i.e., taking the form of spectral noise. Hence it is possible to (spectrally) estimate the observation by applying the classical Wiener filter:

$$|\mathbf{F}_k|^2 = \frac{|\mathcal{F}\{\mathbf{d}\} \cdot \hat{\mathbf{X}}_k|^2}{|\mathcal{F}\{\mathbf{d}\} \cdot \hat{\mathbf{X}}_k|^2 + |\hat{\mathbf{N}}_k|^2} \approx \frac{|\mathcal{F}\{\mathbf{d}\} \cdot \hat{\mathbf{X}}_k|^2}{|\hat{\mathbf{Y}}_k|^2},$$

where $\mathbf{F}_k = \mathcal{F}\{\mathbf{f}_k\}$ is the Fourier transform of the filter \mathbf{f}_k (at frame k), $\mathcal{F}\{\mathbf{d}\}$ is the Fourier transform of the DIC imaging function from Eq. (2), $\hat{\mathbf{N}}_k$ is the noise's spectrum estimate, $\hat{\mathbf{X}}_k$ is a signal's spectrum estimate, and $\hat{\mathbf{Y}}_k$ is the estimate of the signal plus noise's spectrum estimate. In practice, $\hat{\mathbf{X}}_k$ is estimated by averaging numerous radially averaged power spectrums of simulated cells, while $\hat{\mathbf{Y}}_k$ is estimated directly from the observation \mathbf{y}_k (from a least-squares polynomial fit of its radially averaged power spectrum).

After the filtering, the algorithm performs a least-squares reconstruction with an iteratively-reweighted edge-sparse regularizer to exploit the piecewise smooth structure found in OPL images of cells. The reweighting strategy was found to be ideal for cell segmentation because it enhances strong edges while simultaneously diminishing weak ones; in this regard we depart from the traditional notion of ℓ_2 reconstruction fidelity since we favor segmentation fidelity in this application. First order dynamics is incorporated into the reweighting stage by fusing the previous time-step's edge-estimations as a prior to form the current time step's maximum *a posteriori* estimate.

The remaining terms are explained briefly. The isotropic TV operator \mathbf{T} is defined on individual pixels (denoted with the subscript i) as $(\mathbf{T}\mathbf{x})[i] = ((\mathbf{H}_m\mathbf{x})[i]^2 + (\mathbf{H}_n\mathbf{x})[i]^2)^{1/2}$ where $\mathbf{H}_m, \mathbf{H}_n$ are matrix representations of the forward-difference operators in the vertical and horizontal directions respectively. The DIC PSF \mathbf{d} used here was originally proposed in [6] as a steerable first-derivative of Gaussian kernel:

$$\mathbf{d}[m, n] = -me^{-\frac{m^2+n^2}{\sigma_d^2}} \cos \theta_d - ne^{-\frac{m^2+n^2}{\sigma_d^2}} \sin \theta_d, \quad (2)$$

where σ_d refers to the Gaussian spread and θ_d refers to a steerable shear angle.

The inpainting part is easily incorporated by a minor modification of the deconvolution algorithm. Let the pipette mask's pixel-indices be denoted by the set $\bar{\Omega}$; conversely, the non-pipette pixel-indices are Ω . Eq. (1)'s reconstruction term $\|\mathbf{f}_k * \mathbf{y}_k - \mathbf{D}\mathbf{x}_k\|_2^2$ is now modified to become $\|(\mathbf{f}_k * \mathbf{y}_k)_\Omega - \mathbf{D}_\Omega\mathbf{x}_k\|_2^2$ with the subscript indicating its support. The pre-filtered observation has now been partially occluded by the presence of the mask pixels: $(\mathbf{f}_k * \mathbf{y}_k)_\Omega \in \mathbb{R}^M$ such that $M \leq N$. Similarly, the Toeplitz matrix has its support accordingly reduced: $\mathbf{D}_\Omega \in \mathbb{R}^{M \times N}$. We note that the mask should be judiciously applied with the remaining support being of reasonable size (e.g. $\|\Omega\|_0 \approx 0.9N$) and also that the unmasked regions retain sufficient structural detail of the original cell for a sensible reconstruction; i.e., large or important sections of the cell's edge should not be occluded. Figure 4 illustrates the deconvolution and inpainting capabilities of the algorithm on real data.

3. RESULTS

3.1. Synthetic Data (without inpainting)

The proposed algorithm is compared against other state-of-the-art deconvolution algorithms on its segmentation performance, using synthetic DIC microscopy videos of cells. A realistic DIC microscopy cell simulator (without the pipette) was created to generate realistic looking cells and replicate the heavy noise and interference found in the data (as per [1]). A sample of both simulated and real single-frame cell patches are displayed in Fig. 2(a-b) to visually validate its realism. For the experiment, synthetic videos were generated to simulate cell movement during the patch clamp process. A snapshot of frames from one such simulated video is shown in Fig. 2(c).

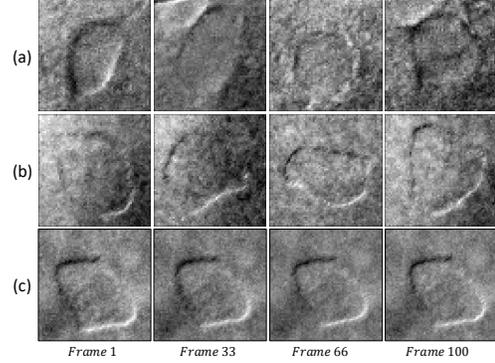


Fig. 2. The first and second rows reflects visual similarity between real (a) and simulated (b) DIC microscopy images of cells. The third row (c) shows a synthetically generated video of a cell experiencing translation and morphological contraction and dilation.

Two boundary error metrics were designed specifically to reflect errors relevant to this patch clamping application. The metrics *Average Boundary Error* (ABE) and *Maximum Boundary Error* (MBE) measures the average and maximum Euclidian distances respectively, between pixels along the contour of the segmentations of the approximated segment and the ground-truth segment:

$$\text{ABE}(A, B) = \frac{1}{\|A\|_0} \sum_{a \in A} \|p_{l,B} - p_a\|_2$$

$$\text{MBE}(A, B) = \max_{a \in A} \|p_{l,B} - p_a\|_2,$$

where $\{A\}, \{B\}$ are sets pertaining to contour pixels of the ground truth segment and the evaluated segment respectively, $p_a \in A, p_b \in B$, and $p_{l,B} = \arg \min_{p_b \in B, p_l \in l} \|p_b - p_l\|_2$ refers to the closest distance on the line segment l starting at the center of the ground truth segment projecting out in the positive direction towards p_a .

The state-of-the-art algorithms under consideration are least-squares regularized ℓ_1 and TV (L1+TV) [6], least-square regularized ℓ_1 and Laplacian Tikhonov (L1+Tik) [6], and least-square regularized re-weighted ℓ_1 , weighted Laplacian Tikhonov, and weighted dynamic filtering (RWL1+WTik+WDF) [7]. The deconvolution from each algorithm is segmented using a global threshold (using Ostu's method) to produce a binary image mask for evaluation.

100 unique videos of synthetic cells patches were generated with each video containing 64×64 pixels \times 100 frames. The DIC PSF's parameters of $\sigma_d = 0.5, \theta_d = 235^\circ$ are assumed to be precisely known by the user. The relationship between physical length and pixels is approximated to be 3.75 pixels per $1.0\mu\text{m}$, by comparing real cells with simulated cells. Exhaustive parameter sweeps on sparsity and smoothness were performed on the state-of-the-art algorithms (L1+TV, L1+Tik, and RWL1+WTik+WDF) using first frame of the video, by finding the closest reconstruction/deconvolution (in the ℓ_2 least squares sense) to the ground truth. For RWL1+WTik+WDF, the dynamics parameter was set as 1.0×10^{-3} , with a maximum of 80 reweighting steps. For PF+RWTV-DF, we fixed $\gamma_0 = 3.0 \times 10^{-3}, \kappa = 5$, with 4 reweighting steps across all videos.

Each video was reduced to a single ABE/MBE statistic by averaging all the ABE/MBE values over the individual frames, and the distributions of these values are reflected in Fig. 3. PF+RWTV-DF was clearly observed to be the superior algorithm as shown by the low overall ABE/MBE values.

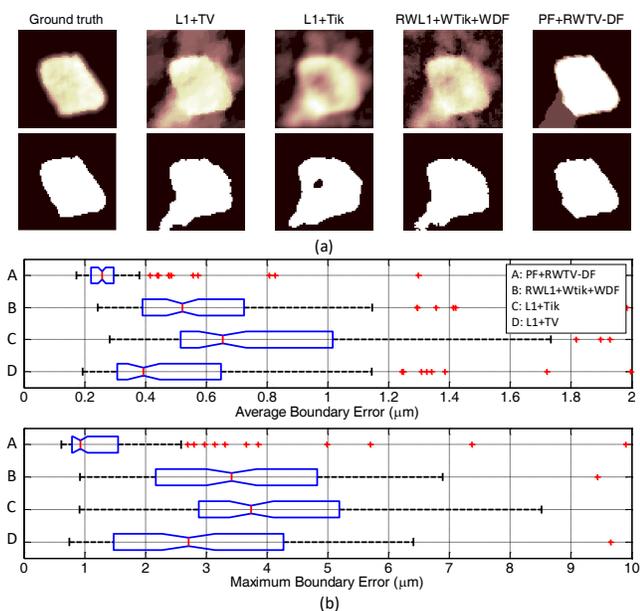


Fig. 3. The proposed algorithm (without inpainting) is tested for its segmentation accuracy, alongside other state-of-the-art algorithms [6, 7], on DIC videos of synthetically generated cells (100 videos with 100 frames per video). In (a), a sample snapshot of the deconvolution (top row) and segmentation (bottom row) from a single frame (of one video) is shown. Two metrics, average boundary error (ABE) and maximum boundary error (MBE), were devised to reflect segmentation errors that relate to the patch clamping application. The box-plots in (b) show the distribution of averaged ABE/MBE values for each of the 100 videos. The proposed algorithm demonstrably outperforms the other algorithms in both metrics.

3.2. Real Video Sequence (with Inpainting)

The proposed algorithm was applied on a real patch clamping video to showcase the cell segmentation and inpainting capability. Although lacking in ground truth, such a demonstration qualitatively validates the potential of the system. In Fig. 4, two particular frames were selected to highlight deconvolution and inpainting. In the first frame (Fig. 4(a)), the targeted cell is segmented in the absence of the pipette, validating the deconvolution performance of the algorithm on real data; this may be used as a reference of how the deconvolution/segmentation should appear with respect to the inpainting scenario. In the second frame, we notice that sans inpainting, the pipette severely interferes with the deconvolution and subsequently the segmentation. As shown in Fig. 4(b), simultaneous reconstruction of two distinctly different shapes (i.e., the pipette is long/thin while the cell is circular) with intensity differences (i.e., the pipette’s pixel intensities are stronger than the cells’) results in an obfuscation of the two objects during the reconstruction. The inpainting strategy is therefore an appropriate method to overcome this problem, as illustrated in Fig. 4(c).

4. CONCLUSION

This work presents a cell boundary tracking system for the patch clamping application on DIC microscopy video. The system’s goal was driven by this particular patch clamping’s application: to pre-

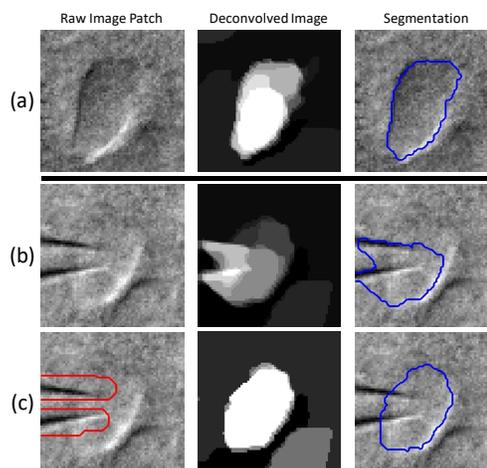


Fig. 4. The segmentation and inpainting capability of the proposed algorithm is highlighted. Row (a) demonstrate a snapshot in time of the segmentation process when no pipette interferes with the deconvolution. Later on, the pipette is in close proximity with the same cell (b) and the segmentation becomes erroneous. In (c), the same image is deconvolved with inpainting and its segmentation more accurately depicts the cell’s boundaries than in (b).

cisely identify and track the boundary of a single targeted cell. The main challenge faced is heavy noise and interference in the data, one part due to biological material in the specimen and another part due to the physical presence of the pipette. The highlight of the system is a novel deconvolution and inpainting algorithm, designed for cell-segmentation by exploiting inherent structures of edge sparsity and temporal dynamics. Simulations demonstrated that our algorithm was superior (at cell-segmentation) compared to previous state-of-the-art algorithms in heavy-noise-and-interference-type DIC imagery. Several video frames of pipette removal via inpainting were also demonstrated, highlighting the potential of such a system.

The contributions of the the proposed deconvolution algorithm are: (1) the pre-filter is an effective and computationally efficient way of tackling organic interference, (2) the strategy of edge-reweighting is excellent for cell segmentation, and (3) compared to a static TV-reweighting approach, our method exploits dynamical structure at negligible additional computational cost. While superior in segmentation performance, this algorithm relies on an iterative approach which is slow if not efficiently optimized. Future work aims for real-time implementation of the proposed algorithm, and this requires a parallelization method like alternating direction method of multipliers (ADMM), similar to work found in [9, 10].

While this system has promising potential, a main shortfall is that the inpainting approach is heavily dependent on the accuracy of the pipette mask. The current method (of template matching) is limiting because it assumes that pipette movement can only be a translation (in the image plane). Future work will aim towards more robust methods of pipette detection and tracking by: directly integrating into the system the actuator signals that control the pipette’s movement, for improved pipette tracking, and building a parametric 3D model of the pipette to generate virtually any template, at any given pipette orientation.

5. REFERENCES

- [1] Qiuyu Wu, Ilya Kolb, Brendan M Callahan, Zhaolun Su, William Stoy, Suhasa B Kodandaramaiah, Rachael L Neve, Hongkui Zeng, Ed S Boyden, Craig R Forest, et al., “Integration of autpatching with automated pipette and cell detection in vitro,” *Journal of Neurophysiology*, pp. jn-00386, 2016.
- [2] RA Hoebe, CH Van Oven, Th WJ Gadella, PB Dhonukshe, CJF Van Noorden, and EMM Manders, “Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging,” *Nature biotechnology*, vol. 25, no. 2, pp. 249–253, 2007.
- [3] Inpakala Simon, Charles R Pound, Alan W Partin, James Q Clemens, and William A Christens-Barry, “Automated image analysis system for detecting boundaries of live prostate cancer cells,” *Cytometry*, vol. 31, no. 4, pp. 287–294, 1998.
- [4] Nezamoddin N Kachouie, Paul Fieguth, and Eric Jervis, “Watershed deconvolution for cell segmentation,” in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 375–378.
- [5] Arjan Kuijper and Bettina Heise, “An automatic cell segmentation method for differential interference contrast microscopy,” in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.
- [6] Kang Li and Takeo Kanade, “Nonnegative mixed-norm preconditioning for microscopy image segmentation,” in *Information Processing in Medical Imaging*. Springer, 2009, pp. 362–373.
- [7] Zhaozheng Yin, Takeo Kanade, et al., “Restoring dic microscopy images from multiple shear directions,” in *Information Processing in Medical Imaging*. Springer, 2011, pp. 384–397.
- [8] United States General Services Administration Federal Standard, “Telecommunications: Glossary of telecommunication terms,” Aug. 7 1996.
- [9] Stanley H Chan, Ramsin Khoshabeh, Kristofor B Gibson, Philip E Gill, and Truong Q Nguyen, “An augmented lagrangian method for total variation video restoration,” *Image Processing, IEEE Transactions on*, vol. 20, no. 11, pp. 3097–3111, 2011.
- [10] Min Tao, Junfeng Yang, and Bingsheng He, “Alternating direction algorithms for total variation deconvolution in image reconstruction,” *TR0918, Department of Mathematics, Nanjing University*, 2009.