Reconstruction of Ca^{2+} dynamics from low frame rate Ca^{2+} imaging data
CS229 final project. Submitted by: Limor Bursztyn

Introduction

Active neurons communicate by action potential firing (spikes), accompanied by transient increases in the intracellular Ca^{2+} concentration ([Ca^{2+}]$_i$). Fluorescent proteins that bind to Ca^{2+} allow to observe dynamic changes in [Ca^{2+}]$_i$ *in vivo* via fluorescent microscopy techniques, such as two-photon laser scanning microscopy (TPLSM). TPLSM is an optical sectioning method - simultaneous interaction with 2 photons is required to excite a fluorescent molecule and thus excitation effectively occurs only in the focal spot of the laser beam. As the laser beam is scanned over a plane within the tissue, emitted light from every spot within the plane is collected and used to reconstruct an image. Compared with electrophysiological measurement techniques, TPLSM offers the advantage of allowing to observe activity in large populations of identifiable neurons (Svoboda and Yasuda, 2006). However, the time resolution is limited by the requirement to collect sufficient fluorescence energy per pixel to obtain a high signal to noise ratio, given the properties of photon shot noise (Vogelstein et al., 2008). With this low time resolution it is difficult to determine exact spiking event times that are of interest (Ozden et al., 2008). Standard analysis of TPLSM images consists of marking a region of interest (ROI), corresponding to the location of a neuron, and then treating the average of pixel intensities in the ROI as a sample of the signal of interest. Thus, the signal is represented via samples taken at the low frame rate of acquisition. In this project, an attempt is made to extract a representation of the signal of interest with a higher time-resolution than that given by the frame rate, and thus facilitated a more accurate identification of spike times. In order to allow the use of every pixel within a frame for an accurate reconstruction of the signal, it is required to exclude redundant pixels from the ROI, which do not reflect the neuron activity but consist of noise alone. To achieve this goal, a mask refinement method, separating signal and noise pixels, was applied.

Methods

Data acquisition: Imaging was performed in anesthetized mice cerebella using a custom designed two-photon micropscope. MPscope software was used for data acquisition and control of the microscope (Nguyen et al., 2006). A single movie, acquired at a frame rate of 10fps and size 128*128 was used to test the refinement and interpolation methods. A movie acquired at a rate of 20fps was used to validate the simulated signal quality.

Simulation: The following characteristics were chosen for the simulated signal: a firing rate of 3Hz and exponential Ca^{2+} decay with time constants varying between 120 and 200 msec (slightly lower than the reported times of 280 ± 60 ms but typical for our dataset). A spike train corresponding to the chosen average firing rate was simulated as described in Dayan and Abbott, 2001, by dividing the time axis into constant interval bins and randomly assigning spikes to bins with a probability of Δt * (firing-rate). Then, the Ca^{2+} signal was constructed as a superposition of typical exponential decay responses to the fired action potentials, with amplitudes and decay time coefficients randomly chosen from the sets [1 0.9 0.8 0.7 0.6] and [0.2 0.18 0.16 0.14 0.12], respectively.
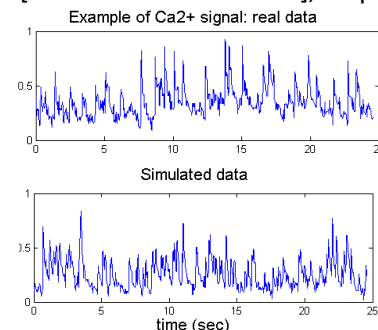

Figure 1. Comparison and simulated and real data. Top: excerpt from a real data set acquired at 20fps, Bottom: except from the simulated data set.

Gaussian random noise with 0.1 standard deviation as well as sinusoidal noise with frequencies 1/3 and 2/3 Hz and amplitudes 0.05 and 1 respectively, were added to make the signal waveform less 'ideal'. Figure 1 demonstrates the similarity between the real and simulated signals. Data acquisition with TPLSM was simulated for a neuron with realistic dimensions residing in a 128*128 or 256*256 frame (the mask used to define the region was formed using the real data set). Taking into account the scanning waveforms used by our microscope control software, a vector of samples that would have been acquired if the

simulated signal represented the activity of a neuron in the ROI was formed. 20fps was used as the highest acquisition rate, 5fps as the lowest. For mask refinement testing, pixels containing noise alone were added, and noise was added to the signal pixels as well when testing various noise conditions (see mask refinement results).

Interpolation: Locally-weighted logistic regression (LWLR) was used to predict values of the signal of interest at a rate of 20fps, based on samples of simulated acquisition or real data, acquired at 5fps, or at 10fps, respectively.

Spike detection: The correlation between the interpolated signal and 5 samples of the typical Ca^{2+} response to a spike – centered at the current time-point, was calculated at every time-point. The response used is given by:

(1) $\frac{1}{1+exp(-(t-t_0)/k)} exp\left(-\frac{(t-t_0)}{\tau}\right)$

Where k = 0.002 and t_0 = 0.15. The multiplication by the sigmoid is used to smooth the exponential rise and exclude values for t < t_0. To detect a spike, a threshold was applied to the result and maxima points above the threshold were chosen as detected spike times.

Mask refinement: Assuming that the intensity in every pixel is either a sum of signal and Gaussian noise or noise alone, a Gaussian Mixture Model consisting of two 2D, or 3D Gaussians was used to classify the pixels covered by the initial mask into 'signal' and 'noise' groups, based on their different characteristics. Model features were pixel intensities when a spike was fired, and at the next one or two frames. For signal pixels, the mean and variance is expected to be different at each frame, and the values are also expected to be correlated. For noise pixels, the means and variances are expected to be the same in all frames, and the values are expected to be independent. Accordingly, the following model equations were used:

(2) $Z \sim Bernoulli(\phi)$

(3) $p(x_1, x_2 | z = 1) =$

$\frac{1}{2\pi\sqrt{|\Sigma|}} exp\left(-\frac{1}{2}\binom{x_1 - \mu_1}{x_2 - \mu_2}^T \Sigma^{-1} \binom{x_1 - \mu_1}{x_2 - \mu_2}\right)$

(4) $p(x_1, x_2 | z = 0)$

$= \frac{1}{2\pi\sigma_N^2} exp\left(-\frac{1}{2\sigma_N^2}((x_1 - \mu_N)^2 + (x_2 - \mu_N)^2)\right)$

Where Z, the latent variable, is equal to 1 for a signal pixel, and to 0 for noise; x_1 is the intensity at the spike time and x_2 is the intensity in the same pixel at the next frame and x = $[x1\ x2]^T$. Model parameters were fit using the Expectation Maximization (EM) algorithm, with the following equations:

(5) $w_1 = p(z = 1 | x_1, x_2) =$

$\frac{p(x_1, x_2 | z=1)\phi}{p(x_1, x_2 | z=1)\phi + p(x_1, x_2 | z=0)(1-\phi)}$

(6) $\phi = \frac{1}{m}\sum_{i=1}^{m} w_1^{(i)}$

(7) $\mu_N = \frac{\frac{1}{m}\sum_{i=1}^{m} w_0^{(i)}\frac{1}{2}\left(x_1^{(i)} + x_2^{(i)}\right)}{\frac{1}{m}\sum_{i=1}^{m} w_0^{(i)}}$

(8) $\sigma_N^2 = \frac{\frac{1}{m}\sum_{i=1}^{m} w_0^{(i)}\frac{1}{2}\left(\left(x_1^{(i)} - \mu_N\right)^2 + \left(x_2^{(i)} - \mu_N\right)^2\right)}{\frac{1}{m}\sum_{i=1}^{m} w_0^{(i)}}$

(9) $\mu_{1,2} = \frac{\frac{1}{m}\sum_{i=1}^{m} w_1^{(i)} x^{(i)}}{\frac{1}{m}\sum_{i=1}^{m} w_1^{(i)}}$

(10) $\Sigma = \frac{\frac{1}{m}\sum_{i=1}^{m} w_1^{(i)}\left(x^{(i)} - \mu_{1,2}\right)^T\left(x^{(i)} - \mu_{1,2}\right)}{\frac{1}{m}\sum_{i=1}^{m} w_1^{(i)}}$

Similar equations were used for the 3D case.

After the EM algorithm converged, signal pixels were chosen as pixels having values of w_1 higher than 0.5 (i.e., pixels having a probability larger than 50% of being signal pixels), and the rest were chosen to be noise pixels, removed from the mask.

To validate the quality of the resulting mask, and choose between masks fitted to different frames, linear regression was used to fit 2D polynomials to the resulting samples acquired by averaging over lines in each frame. The mask yielding a minimal mean-square error (MSE) of fit was chosen as the optimal mask.

Preprocessing of real data: Frames were represented as normalized differences in fluorescence: $(F-F_0)/F_0$, where F is the current fluorescence intensity image and F_0 is the intensity averaged over 50 frames. A neuron residing within the field of interest was manually identified according to its typical activity pattern and spatial shape.

Performance analysis: For quantifying interpolation performance on simulated data sets, the MSE between estimated and true samples was calculated. Then, event detection was performed using both interpolated and uninterpolated data, showing that the rate of spike detection at a high time resolution was increased. Thus, a correct spike detection was defined as a difference smaller than 1/20sec between estimated and true spike times (TP), failure to detect a spike within a time window of 1/10sec around a true spike was counted as miss-detection (FN) and a

detected spike that is more than 1/20sec away from any real spike was defined as a false alarm (FP).

Results:

Interpolation of simulated data: The results of using interpolation with LWLR are presented in Fig. 2. Since using all the samples within a frame and using averages taken over lines yielded similar results, we used the latter, which is easier to implement. The lowest MSE for a small 128*128 frame, 0.047, with respect to the true signal, is similar to the MSE reached by averaging over frames 'acquired' at 20fps, 0.043 (the red line in the figure), which was calculated as a reference for performance quality. The bandwidth parameter used was 0.04, corresponding to averaging over 58 lines per prediction. Since there are 39 lines in a frame, this means information was extracted out of more than one, but less than two, frames. With a large 256*256 frame, the lowest MSE reached was 0.077. Unfortunately, this constitutes only a minor improvement compared to the lowest MSE achieved with interpolation based on averages taken over the entire 5fps frames, 0.087 (the black line in Fig. 2).
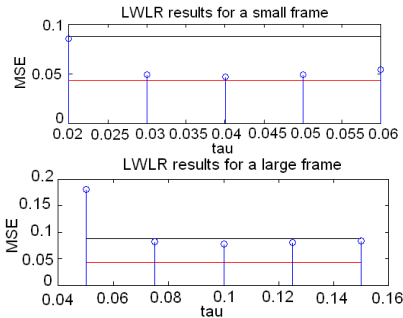

Figure 2. LWLR results (simulation) – MSE error between interpolated frames 'acquired' at a rate of 5fps, and the true signal at 20Hz. Top: small frame (128*128), Bottom: large frame (256*256). The red line indicates the MSE error for averaging over the ROI in frames 'acquired' at a rate of 20fps, the black line is the error for applying LWLR to averages over the entire ROI in the 5fps frame.

Spike Detection in Simulated data: The results of spike detection at a time resolution of 1/20sec are presented in Fig. 3.
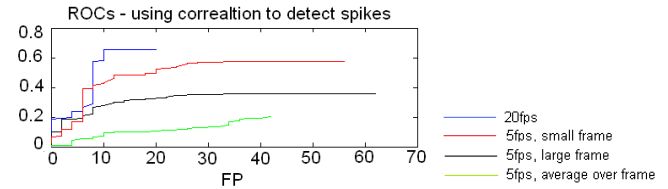

(a)
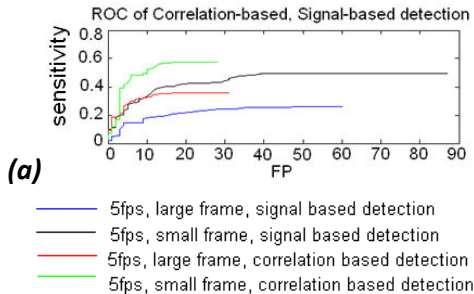
(b)

Figure 3. Spike detection results (simulation). (a) Comparison of signal-based detection and correlation-based detection. (b) Comparison of interpolated 5fps detection rates and uninterpolated 5fps, 20fps detection rates.

Figure 3(a) shows that using the correlation with the typical response yields a more accurate detection of the spike time than using maxima points of the signal. This is probably due to the fact that the sampling by scanning might miss the exact peak time, and thus the maximum value in the sampled signal is not the actual maximum value of the true signal. However, by using a few more samples to better describe the shape of the response, a more accurate identification of the exact spike time can be reached. Figure 3(b) shows that interpolation improves the detection rate, compared with detection based on uninterpolated frame averages, but does not reach the level of accuracy which acquisition at a fast rate yields. Further optimization of the method (e.g., by adjusting the typical response parameters or changing the number of samples of the response used to calculate the correlation) may yield further improvement of the results.

Mask Refinement in Simulated data: Refinement quality was verified under various noise levels. In real data, since the most significant noise source is photon shot noise, it is expected that the variance of the noise for the higher mean signal pixels will be high as well, with respect to the lower mean noise pixels (in accordance with a Poisson distribution). However, since it is unclear what the exact statistical properties of the real data are (it is unknown which pixels correspond to signal and which to noise and thus is impossible to calculate the corresponding statistics), we concluded it will be useful to know what algorithm parameters work best at each of many possible noise levels added to the signal and noise pixels.

The parameter space which was explored to find the best configuration at each noise level, consisted of 3 initial mask sizes – including 320, 284 and 248 pixels (143 of which are signal pixels); 2 initialization methods (random, based on thresholding over x_1

values) and 2 or 3 features: intensity in a frame with a spike and in one or two consecutive frames. Specificity was chosen as the criterion for comparison, as the objective is to refine the mask and remove noise pixels from it. However, we also mention sensitivity values, as a drop in these means only a few signal pixels remain to be used for interpolation – and this should be avoided. The best results achieved in every noise condition (averaged over 3 trials) are shown in Table 1. Initialization using a threshold always yielded better performance than random initialization, thus all results reported are based on threshold initialization.

Init. Mask size	# of features	Specificity (Sensitivity) mean ± std.	Noise Std.	Signal Std.
320	3	1 ± 0 (0.99 ± 0.01)	0.3	0.1
320	2	0.97 ± 0.03 (0.99 ± 0.02)	0.5	0.1
320	3	1 ± 0.003 (1 ± 0.004)	0.2	0.3
320	3	1 ± 0 (0.99 ± 0.01)	0.1	0.4
320	2	0.98 ± 0.02 (0.87 ± 0.09)	0.3	0.4
320	2	1 ± 0 (0.61 ± 0.28)	0.1	0.6
284	2	0.94 ± 0.09 (0.65 ± 0.38)	0.5	0.6

It can be seen that a large initial mask size yielded better results than a small one in almost all cases. It could be that a large enough amount of noise pixels is required for model parameter estimation during training. For the choice of the number of features, however, it is not clear why the 3 feature model sometimes better performed than the 2 feature model and sometimes did not. One possibility is that the number of samples used for training is small for a 3 feature model and thus performance is inconsistent. In addition, using 3 features sometimes yielded an improvement compared with 2 features, but not consistently. For high noise levels (0.5, when keeping in mind signal amplitudes are no larger than 1), the sensitivity dropped significantly. It is possible that in such high-noise cases, it is better to use a threshold lower than 0.5 for assignment of signal pixels as such, even though this choice will inevitably cause a decrease in specificity. In addition, the method's performance is dependent on correct identification of events used for mask refinement – which is more difficult when the noise levels are high. Large events allow for a more accurate mask refinement, whereas refining a mask with a frame that was wrongly identified as consisting of an event yields very poor results. Running refinement on the same frame with different configurations showed that higher likelihood values correspond to a better refinement result. This is encouraging as the resulting likelihood can be used as a performance measure in real data as well.

Mask Refinement in Real data: In real data, it is impossible to quantify the performance of refinement as in simulation, since the true identity of the pixels is unknown. However, a correctly refined mask should yield line averages having a clear trend – being samples of the same signal taken at close, yet different, time points. Qualitatively, this result has been achieved for 2-3 out of every 10 frames over which the algorithm was run. Some qualitatively good refinement results are shown in Fig. 4.
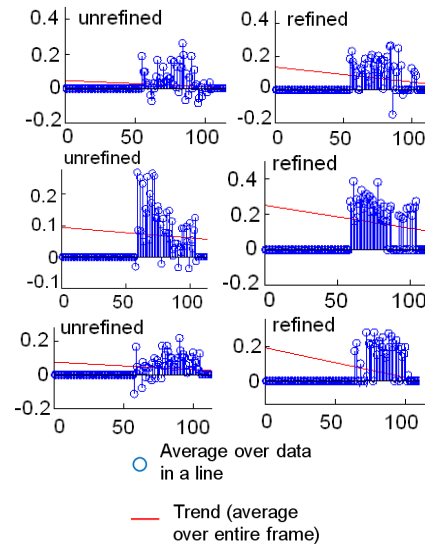

Figure 4. Line averages in real data frames before and after the mask was refined using GMM fitting.

Frames in which refinement clearly failed were easy to detect since they consisted of very few pixels identified as signal pixels. When fitting a 2D polynomial to the refinement result, the mask yielding a minimal MSE was also one that qualitatively seemed to have been successfully refined. When this optimal mask was used to interpolate the data – as shown in Fig. 5, the resulting signal seemed noisier than the original frame-averaged signal, but also consisted of

sharper peaks corresponding to spiking events. Spike event times detected based on the interpolated signal were not identical to those detected based on the averaged signal. Whether this indeed constitutes a better representation of the signal and more accurate determination of spike times cannot be verified at this point.
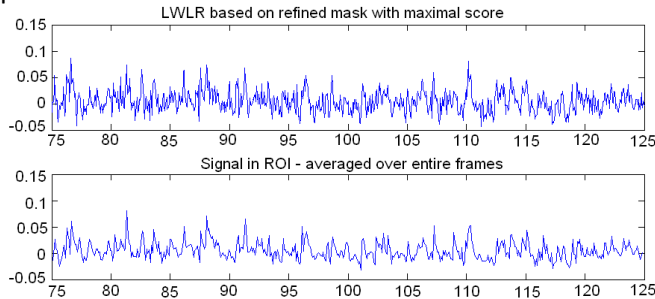

Figure 5. Result of applying LWLR to real data sampled at 10fps, interpolating it to a sampling rate of 20Hz, compared with averages taken over the entire ROI, at the frame rate.

Discussion:

Simulation results indicate that given the properties of the dynamically changing signal of interest and the method in which it is acquired, averaging over the ROI in each frame may cause loss of information that can be retrieved via a more careful signal interporlation. Using LWLR we in effect still perform averaging over information (thus, suppressing noise), however, since acquisition time is taken into account and the average is weighted accordingly, a better representation of how the signal changes over time is produced.

In addition, mask refinement results show that fitting a GMM to a manually selected mask can improve the identification of signal pixels and prevent noisy samples from affecting further analysis. Further testing is required to find an optimal way of applying this method to real data. Since the imaging is performed *in vivo* motion artifacts may affect the correct selection of pixels in various frames. Accordingly, for real data it may be required to fit a different mask for every event, or divide the signal into short time windows fitting a different mask for each. When doing so, it is possible that using a previously fit mask, from a previous event or window, to initialize the refinement procedure for the next

event or window, will yield better results than initializing using a threshold.

In order to further validate the results of mask refinement, interpolation, and spike detection on real data, as well as identify and optimize the most critical parts of the process, it is required to use a data set that consists of simultaneous electrophysiological measurements that allow accurate identification of spike times.

References

1. Dayan P, Abbott LF. 'Theoretical Neuroscience, Ch. 1 , MIT Press, Cambridge, MA, 2001.
2. Nguyen QT, Tsai PS, Kleinfeld D. "MPScope: a versatile software suite for multiphoton microscopy." J Neurosci Methods. 156(1-2): 351-9, 2006.
3. Ozden I, Lee HM, Sullivan MR, Wang SS. "Identification and clustering of event patterns from in vivo multiphoton optical recordings of neuronal ensembles." J Neurophysiol. 100(1): 495-503, 2008.
4. Svoboda K, Yasuda R. "Principles of two-photon excitation microscopy and its applications to neuroscience." Neuron. 50(6): 823-39, 2006.
5. Vogelstein, J., Watson, B., Packer, A., Yuste, R., Jedynak, B., and Paninski, L. "Spike inference from calcium imaging using sequential Monte Carlo methods." Under review, Biophysical Journal, 2008 (appears in Prof. Paninski's website - http://www.stat.columbia.edu/~liam/research/if.html