A New Template Matching Method using Variance Estimation for Spike Sorting.

Hansang Cho[1], D. Corina[2], J. F. Brinkley[3], G. A. Ojemann[4], and L. G. Shapiro[5]
[1,5]Department of Electrical Engineering, [2]Department of Psychology, [3]Department of Biological Structure, [4]Department of Neurosurgery, [5]Department of Computer Science, University of Washington, Seattle, WA, USA

Abstract— The analysis of single unit recording data requires a spike sorting method to separate blended neuronal spikes into separate neuron classes. A new template matching method for spike sorting based on shape distributions and a weighted Euclidean metric is proposed. The data is first roughly clustered using a Euclidean distance metric. Then the Levenberg-Marquardt method is used to estimate the variances of the neuron classes using curve fitting on the clustered data. Finally, the weighted Euclidean distance method is applied to minimize errors caused by different variances. This method provides optimized template matching results when the neuron variances are considerably different.

I. Introduction

Classification of neural spikes is a prerequisite for studying neural activity and brain functions. In extracellular recordings, an electrode placed in the cortex usually records spike activity from a number of neurons proximal to the electrode. Each individual neuron's signal, therefore, has to be separated by a classification method, which is called neuron spike sorting. The goal of spike sorting is to find the firing positions in time from each neuron. Due to the various shapes of neuron spikes and the lack of a priori knowledge about the neuronal information of each patient, the template matching method is a reliable method of spike sorting. After extracting templates, a Euclidean metric is used to calculate the distance to the templates, and then neuron spikes are clustered into nearest-neighbor classes.

A Euclidean metric is a very efficient way to classify data into nearest-neighbor templates when the classes have similar variances [1]. However, if the classes to be clustered have a wide range of variances, data in a class with larger variance can be misclassified into an adjacent class with smaller variance, resulting in errors. Fig. 1(a) demonstrates how a Euclidean metric can cause classification errors on neuron spikes having different variances. Provided that the class variances are already known, the classification errors can be decreased, as shown in Fig. 1(b). However, it is very difficult to measure variances for neuron clusters, since they differ for each neuron and the heights of neuron spikes are sometimes too close to separate.

In this paper, we present a new method for measuring variances of neuron firings using shape-based estimation, and for using the estimated variances to improve the classification process. The proposed method uses the Levenberg-Marquardt algorithm as a shape-based estimation method, in conjunction with a weighted Euclidean metric for refined template matching.

II. Methodology

Electrophysiological recordings were obtained during neurosurgery performed by Dr. George A. Ojemann in the Neurosurgery Department, University of Washington. For each patient, four extracellular electrodes were placed in the temporal lobe to collect neuronal firings. The microelectrode recording was performed in the sections of the cortex that were subsequently resected as part of the surgical therapy for epilepsy.

A. Template extraction

Although a general band-pass filter can simultaneously remove both high and low frequency noise, it can cause shape distortion on spike signals, because an ideal filter cannot be implemented. Since the shapes of neuron firings are one of the most important factors in spike sorting, the general band-pass filter method may lead to large errors in spike classification.

In [4] the maximal overlapped discrete wavelet transformation (MODWT) is employed to extract the only

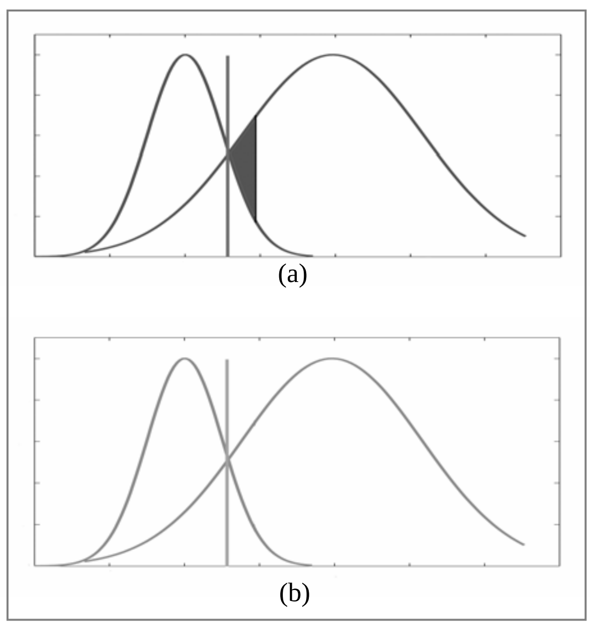

Fig. 1. a) errors caused by applying a Euclidean metric to neuron classes with different variances; b) improved classification using the known variances of the two classes

bandwidth having information about the spike signal and to remove jitter and baseline noise, and the rotated principal component analysis (RPCA) is then employed to extract templates from the neuron signal. Since the goal of spike sorting is to find the positions of the firings in time for each neuron and since MODWT has the characteristic that the time-shift of details and approximation of multiresolutional analysis are consistent with the time-shift of input signals [2], the MODWT is appropriate for spike discrimination. As the wavelet transform kernel for the MODWT, the Haar wavelet is used, because it is better at preserving the fine details and has the most compact spatial support of all wavelets [3]. It is also able to represent the signal efficiently, with fewer coefficients than the other methods [4].

In the traditional PCA methodology, neuron firings are aligned by amplitudes to determine the principal components. However, it is impossible for the second or third principal components to have proper neural shapes, because they are required to be orthogonal to the first PC and the second PC, respectively. This problem is solved by the use of the RPCA. For the neuron signal, the most prominent characteristic of the spike shape is its amplitude. The firings having heights above a threshold are aligned by amplitude, and RPCA is run to extract the first PC, which is then removed from the original neuron signal. The reconstructed signal is realigned by amplitudes so that RPCA can again be applied to get the next PC. If the variance of the RPCA is above the threshold, it means the principal component is dominant in the input segments. PC extraction and reconstruction are repeated until the variance is below the threshold, and then the principal components become templates for each type of neuron signal [4].

B. Variance estimation

There is no a priori knowledge of the variances of neuron spikes, because each neuron has its own unique characteristics, such as mean values, variances, and firing patterns. Therefore, neuron spike variances have to be estimated using data dependent methods such as maximum likelihood estimation or the expectation maximization algorithm. However, previously classified data show that neuron spike classes are frequently too close to estimate variances using such statistical estimation methods. Fig. 2 shows the distribution of whole neuron spike heights, all from one patient, overlaid with the Gaussian distributions for each subclass. Since both the maximum likelihood estimation and expectation maximization algorithms try to find the proper number of Gaussian distributions, if the Gaussian mixture is too close, these methods do not effectively estimate variances

To solve this problem, we propose a shape-based algorithm for estimating the variances of neuron spikes. Since the statistical estimation methods are not effective for estimating variances based on the data itself when the classes are too close, we assess variances based on the distribution shapes. Even though the variances of each cluster are extremely close, we can at least get partial shapes of the variances, which are then used to estimate the complete Gaussian shape distributions. To acquire partial shapes of neuron spike classes, we apply a simple Euclidean distance

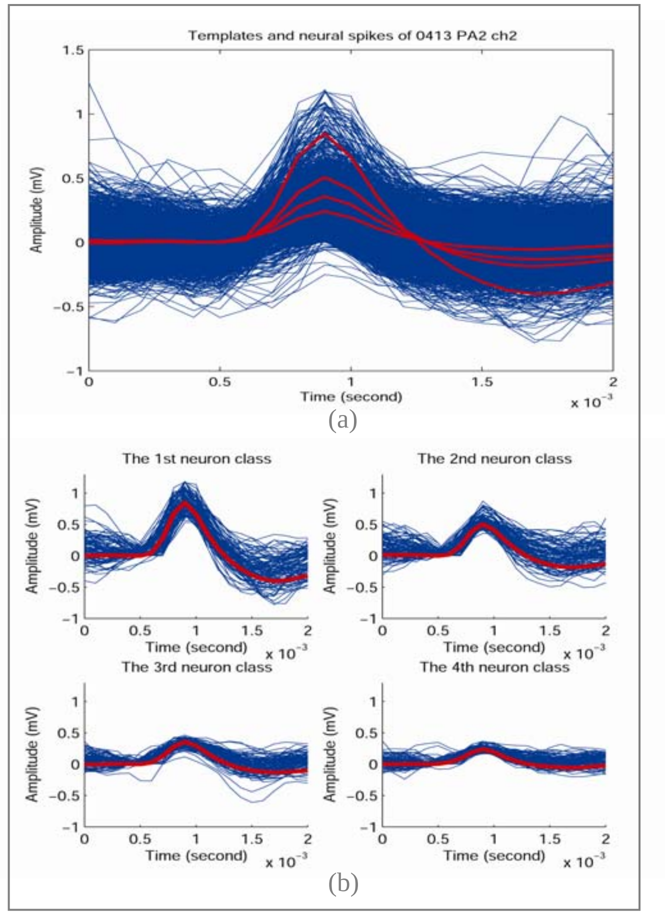

Fig. 3. (a) Neuron spikes aligned with amplitudes (blue) and templates extracted by the rotated PCA (red). (b) Clustered spikes to nearest neighbor templates (blue) and corresponding templates (red)

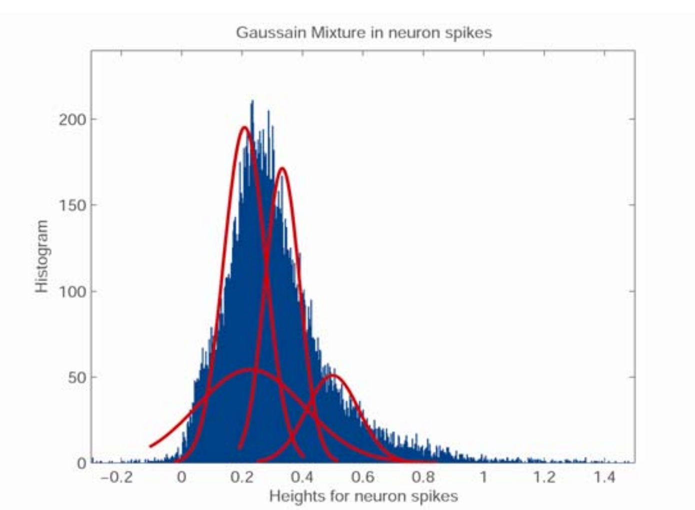

Fig. 2. Histogram of heights for entire neuron spikes. Red lines represent Gaussian distributions for each class inside the whole histogram

method by which neuron firings are classified into nearest neighbor templates with minimum distances. Fig. 3 shows whole neuron spike segments classified into templates by the Euclidean distance.

To obtain the variances of each class, the heights of neuron spikes in a class are collected and the corresponding histograms are plotted. Furthermore, outliers, defined as points outside three times the standard deviation of the collected data, are removed to reduce distortion of the distribution shapes. The Levenberg-Marquardt method is used as a shape-based algorithm to estimate variances of neuron spikes. The Levenberg-Marquardt algorithm is for solving nonlinear least squares problems. It is a line search method whose search direction is a cross between the Gauss-Newton and steepest descent directions. The Levenberg-Marquardt finds the minimum of a function $G(x)$ that is a sum of squares of nonlinear functions:

$$G(x) = \frac{1}{2}\sum_{i=1}^{m}[g_i(x)]^2 \tag{1}$$

Let $J_i(x)$ be the Jacobian of $g_i(x)$; the Levenberg-Marquardt searches in the direction given by the solution p to the equations

$$(J_k^T J + \lambda_k I)p_k = -J_k^T g_k \tag{2}$$

where λ_k are nonnegative scalars and I is the identity matrix. [6] [7]. In this paper, we used the curve fitting toolbox in Matlab for shape-based variance estimation.

C. Modified template matching

To obtain optimized classification results, a weighted Euclidean distance method is applied. The distance d_k between the input neuron spikes and templates is defined as

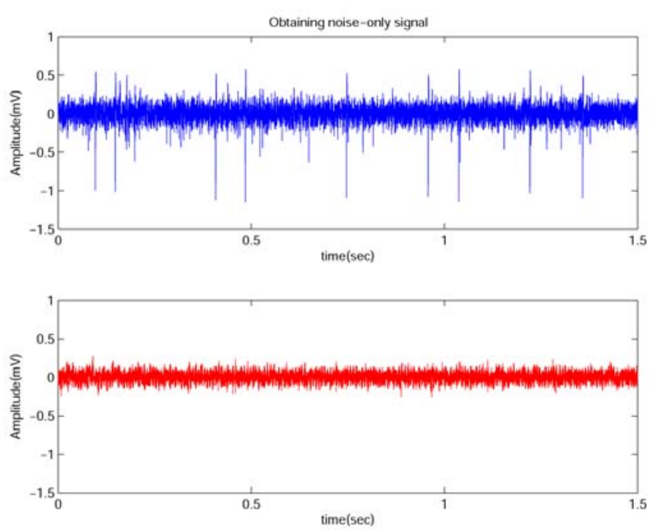

Fig. 4. Extraction of background noise from a real neuron signal. The upper figure shows the original neuron signal and the lower figure shows the background noise.

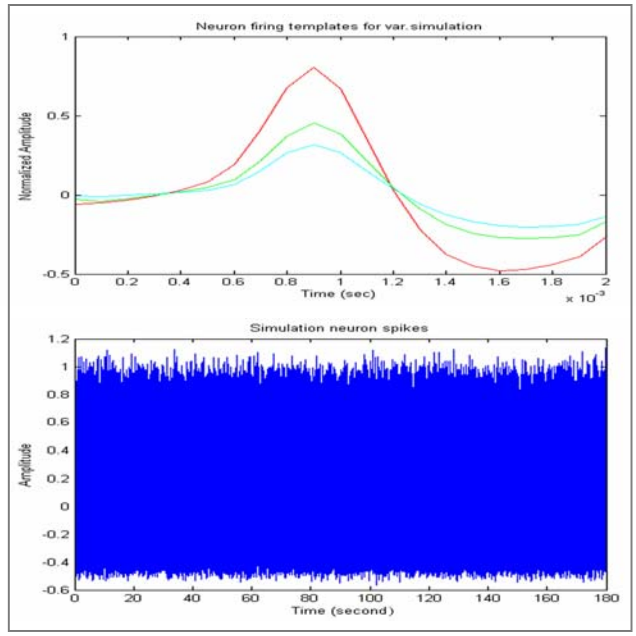

Fig. 5. (top) three neuron spike templates embedded into simulation data.; (bottom) simulation neuron signal ready to be analyzed.

$$d_k^2 = \frac{\sum_k w_{ik}(x_{ik} - t_{ik})^2}{\sum_k w_{ik}} \tag{3}$$

where x_{ik} is the value of variable k in neuron spikes i, and w_k is a weight of 1 or 0, depending upon the inverse of standard deviations of neuron class k.

III. Results

A. Simulation data

A simulated neuron signal was used to verify the proposed method under known conditions. The simulated signal was designed with three spike templates from one actual neuron of a patient and background noise. The spike train lasted for a 12 second duration (120,000 samples with 10 kHz sampling rate), embedded with 2000 neuron spikes from each template into the background noise. The background noise was extracted from a real neuron signal by setting zeros into the 2nd through the 5th details in the MODWT decomposition and applying the inverse MODWT. The signal associated with neuron spikes was removed, and only the noise part with baseline remained. Fig. 4 shows the background noise extracted from an actual neuron firing, and Fig. 5 shows the embedded templates and the whole simulated data set.

Our new classification method, using the Levenberg-Marquardt algorithm and the weighted Euclidean distance, was applied to four kinds of simulation data with diverse

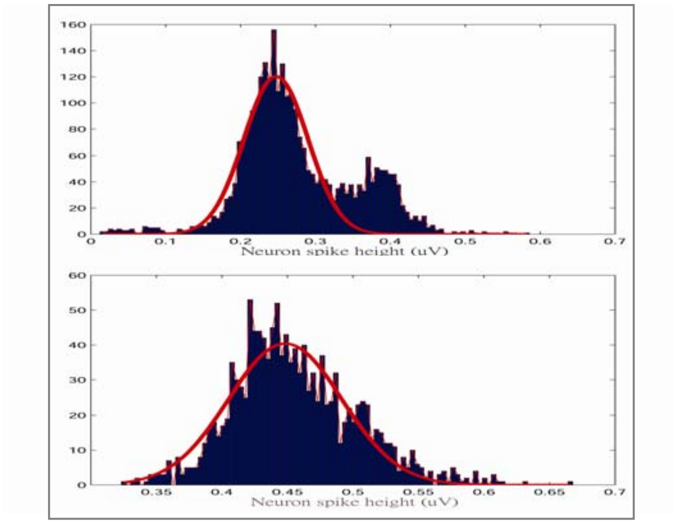

Fig. 6. Variance estimation on real data by applying the Levenberg-Marquartdt method (below) to data previously classified by a Euclidean metric (above).

variances. When the variances of three neuron spike classes are equal, the result obtained by our new method is the same as that obtained using the Euclidean distance alone. However, when the variances of three classes are significantly larger than the others, our method combining the Levenberg-Marquardt algorithm and the weighted Euclidean distance improves classification efficiency. Table I shows a comparison of our new method with the Euclidean metric, in which the numbers represent accurate classifications out of 2000 neuron spikes per class.

B. Real Data

The proposed method was applied to actual neuron signals recorded from the temporal lobes of human subjects. The shapes of the spikes were also verified by experts. Furthermore, the new classification method was found to decrease the variances of the heights of neuron classes by around 10% compared to the Euclidean metric. Fig. 6 shows histograms of heights with variance estimations by the Levenberg-Marquardt method. The smaller hump on the right side of the upper figure demonstrates falsely detected data caused by applying the simple Euclidean distance.

IV. Discusssion

TABLE I
CLASSIFICATION RESULTS OF SIMULATION

	$\sigma_1 = 0.08, \sigma_2 = 0.08, \sigma_3 = 0.08$		$\sigma_1 = 0.12, \sigma_2 = 0.06, \sigma_3 = 0.06$	
	Euclidean	Weighted	Euclidean	Weighted
Neuron 1	1997	1998	1999	2008
Neuron 2	1989	1992	1826	1963
Neuron 3	2223	2219	2251	2105
	$\sigma_1 = 0.08, \sigma_2 = 0.12, \sigma_3 = 0.08$		$\sigma_1 = 0.07, \sigma_2 = 0.06, \sigma_3 = 0.12$	
	Euclidean	Weighted	Euclidean	Weighted
Neuron 1	1982	1992	1997	1997
Neuron 2	1726	2019	1201	1743
Neuron 3	2372	2009	2902	2360

As we can see in Table I, the results given by our new method are very similar to the those generated by the Euclidean distance method, because the two methods are identical if the variances of three neuron spike classes are the same. If the heights of adjacent neurons are very close but they have different variances, the proposed method can decrease false classifications. This is because errors of two close neurons with diverse variances by the Euclidean distance are greater than those of neurons with distinguishable distance between Gaussian center points..

The total detected number of neuron spikes sometimes exceeds the originally embedded neuron firings in simulation, because background noise with high amplitudes is sometimes falsely detected. Therefore, in Table I numbers exceeding 2000 (the actual number of neuron spikes) indicate the erroneous classification of background noise.

V. Conclusion

Although neuron spike templates can be accurately extracted using methods such as the rotated PCA, template matching by a Euclidean metric can cause classification errors due to widely differing variances. The proposed method uses the Levenberg-Marquardt algorithm to estimate the variances of each cluster based on distribution shape, and a weighted Euclidean metric to refine the classification results. The new method improves clustering, even when the heights of the neuron classes are too close to measure variances by existing statistical estimation methods.

Acknowledgement

This work was partially funded by NIH grants NS 36527 (Neural Correlates) and DC02310 (a Human Brain Project grant).

References

[1] Michael S Lewicki, “A review of methods for spike sorting: the detection and classification of neural action potentials”, *Network*, Nov;9(4). pp. R53-78, 1998.

[2] Donald B. Percival, and Andrew T. Walden, *Wavelet methods for time series analysis*, Campridge, 2000, ch. 5, pp. 159-179

[3] Woon-Young So, and Dong-Seok Jeong, “A study on the postprocessing of functional MRI using overcomplete wavelet transform”, *IEEE Tencon*, pp. 1174-1177, 1999

[4] Hansang Cho, D. Corina, G. A. Ojemann, J. Schoenfield, Leona Zamora, and Linda Shapiro, “A new neuron spike sorting method using maximal overlap discrete wavelet transform and rotated principal component analysis”, *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE* , Vol.3, pp. 2921-2924

[5] M. Barlow, S. Nigam, and E. H. Berbery, “ENSO, pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow”, *American meteorological society*, vol. 14, pp. 2105-2128, 2001

[6] Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math*. Vol. 11, pp 431-441, 1963.

[7] Press W., Flannery B., Teukolsky S., and Vetterling W., *Numerical recipes in C*, Cambridge University Press, 1988. pp. 542-7