31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Energy Based Evolving Mean Shift Algorithm for Neural Spike Classification

Zhi Yang, Qi Zhao and Wentai Liu
School of Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064
{yangzhi, zhaoqi, wentai}@soe.ucsc.edu

***Abstract*— This paper presents a novel nonparametric clustering algorithm, called energy based evolving mean shift (EMS) clustering. It defines an energy function to characterize the compactness of the underlying data set and proves the clustering procedure converges. Through iterations, the data points collapse into well formed clusters and the associated energy approaches zero. Although as a general algorithm, the EMS is designed for resolving neural spikes to individual sources which is usually called "spike sorting".**

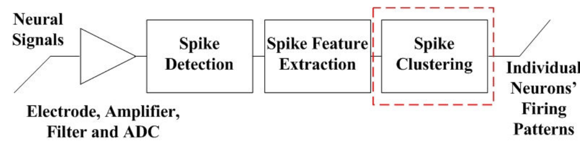

Fig. 1. Diagram of spike sorting

I. INTRODUCTION

Most neurons in the brain communicate by firing action potential which is a "spike" of positive and negative ionic discharge that travels along the membrane. The brief voltage spike can be recorded by microelectrodes and used to decode the information generated by the recorded neurons. Very often one electrode is surrounded by multiple firing neurons, and their recorded activities become superimposed. To correctly understand the information in the biological neural network, it is a critical step to resolve spikes to individual neuronal sources [1], [2], which is called spike sorting.

Spike sorting is essentially a high dimensional clustering problem. Directly classifying the recorded waveforms in high dimensional space is not preferred partially because data points in high dimensional space are very sparse, and clustering algorithms tend to be imprecise. Algorithms which can reduce the data to a few significant features are normally applied before further clustering [3], [4].

Clustering the extracted neural features, however, is still challenging due to the following factors. First, the shapes of the clusters can be irregular and unpredictable. Second, the density and size of each cluster varies significantly, which imposes difficulty for many fixed bandwidth clustering algorithms [5]. Third, the number of data obtained is limited due to collecting and processing expenses and the estimated density distribution is not accurate. In addition, it is likely that noise is mis-identified as spike events, which contaminates the density distribution. Fourth, efficiency is more of a concern since the algorithm needs to be realized with hardware subjected to power and size limitations. Algorithms which need intensive computation and large memory are less preferred. To approach these challenges, we propose a novel energy based evolving mean shift (EMS) clustering algorithm.

Figure 1 depicts the system that the proposed algorithm is tested on [6]. In this work we employ the recently proposed spike feature extraction algorithm for the functional blocks proceeding the clustering block [7], [8].

The rest of paper is organized as follows. Section II presents the EMS clustering algorithm for spike sorting. Section III shows experimental results and section IV concludes the work.

II. EVOLVING MEAN SHIFT SPIKE CLUSTERING

In this section, we introduce a new clustering algorithm that is designed to group the extracted spike features. The algorithm is based on the well-known mean shift algorithm, which is a tool invented in the 1970s [9], and successfully applied to areas such as visual tracking, image segmentation and clustering. For clustering, the mean shift based algorithm is nonparametric, which does not require prior knowledge of the number of clusters, and does not constrain the shapes of the clusters.

The reported clustering algorithm in this work is named as evolving mean shift (EMS). The main novelty of our algorithm and the advantages are described as follows. First, the proposed EMS clustering algorithm inherits the advantages from the mean shift algorithm, e.g., it is nonparametric and robust to various cluster geometry and density variation. Compared with the mean shift algorithm, EMS is faster and more capable of handling dataset with large portion of plateau regions. Compared with the blurring mean shift algorithm [10], which is a variant of mean shift and proposed to accelerate the convergence, EMS has further improved speed while not requiring a post combining procedure to perform partition. In addition, EMS is insensitive to noisy events, as it favorably handles those noisy events at an early stage so that thy are less likely to mislead the classification of other events. This feature is important to spike sorting, where noise and recording artifacts are frequently observed. In the rest of this section, the EMS clustering algorithm will be explained in detail.

A. EMS Energy Function

Here we describe an energy function for EMS to evaluate the compactness of the underlying dataset. Formally, given a dataset $X = \{x_i\}_{i=1...N}$ of N points, the energy of X is defined as the sum of energy from individual point $x_i|_{i=1...N}$ as

$$E(X) = \sum_{i=1}^{N} E_{x_i}, \quad (1)$$

where

$$E_{x_i} = \sum_{j=1, j\neq i}^{N} (E_{x_j.x_i} + E_{x_i.x_j}). \quad (2)$$

In Eq. 2, $E_{x_j.x_i}$ is the energy contributed by point x_j to point x_i with kernel $K(x)$ and bandwidth h_{x_i},

$$E_{x_j.x_i} = f(h_{x_i})(K(0) - K(\frac{x_i - x_j}{h_{x_i}})), \quad (3)$$

where $K(x)$ is an arbitrary isotropic kernel with a convex profile $k(x)$, i.e., it satisfies $K(x) = k(|x|^2)$ and $k(x_1) - k(x_2) \geq k'(x_2)(x_1 - x_2)$. Without loss of generality, we set $k(0) = 1$ and Eq. 3 reduces to $E_{x_j.x_i} = f(h_{x_i})(1 - K(\frac{x_i - x_j}{h_{x_i}}))$. The intuition of $(1 - K(\frac{x_i - x_j}{h_{x_i}}))$ is that the closer the point x_j to x_i, the less energy it contributes to x_i. $f(h_{x_i})$ is a shrinking factor that is designed to be a monotonically increasing function of bandwidth h_{x_i}, as will be discussed in section II.E. It is worthy mentioning that after assigning an initial global bandwidth h_0, bandwidth h becomes independent to the user and is trained through iterations.

Let $f(0) = 0$ and it is straightforward to verify that the energy definition satisfies

(1) $E(X) \geq 0$;

(2) $E(X) = 0$ when fully clustered.

B. The Evolving Mean Shift (EMS) Clustering Algorithm

We outline the evolving mean shift clustering algorithm as follows:

Algorithm 1 The EMS Clustering Algorithm

Input: A set of data points X^k, where k is an iteration index and is initialized to 0

Output: A clustered set of data points X_{EMS}

- Select one data point $x_i^k \in X^k$ whose movement could substantially reduce the energy as defined in Eq. 1. The point selection scheme is discussed in section II.C.
- Move x_i^k according to the EMS vector defined in Eq. 5, specifically $x_i^{k+1} = x_i^k + \overrightarrow{EMS_x^k}$.
- Compute the updated bandwidth $h_{x_i}^{k+1}$ for point x_i^{k+1} according to Algorithm 2, and adjust the EMS vectors for all points using Eq. 5.
- If $E(X^k)$ satisfies the stopping criterion, stop; otherwise, set $k \leftarrow k + 1$ and go to the 1^{st} step.

C. Point Selection

Selecting a point with the largest energy reduction for moving has several important benefits. First, it avoids operations of data that lead to small energy reduction (e.g. data points in plateau regions); therefore, requires less iterations compared with the mean shift or blurring mean shift algorithm. Second, it efficiently pushes loosely distributed points toward a localized peak, which prevents them being absorbed into nearby clusters with larger densities. As a result, poorly separated clusters with different densities could be handled appropriately.

To select a data point with the largest energy reduction, at the initialization stage, the EMS vector is computed for each data point. Each following iteration moves a selected point to a new location according to the EMS vector, updates its bandwidth according to Algorithm 2 (section II.E) and adjusts the EMS vectors for all the data points. Based on the adjusted EMS vectors, a new point corresponding to the largest energy reduction is selected for the next iteration.

Besides the speed enhancement, the point selection scheme offers EMS additional advantages compared with its competing algorithms. A notorious drawback of blurring mean shift is that a direction of larger variance converges more slowly rather than the reverse; as a result, blurring mean shift frequently collapses a cluster into a "line" by taking a number of iterations. After that, the blurring mean shift algorithm converges data slowly to the final state and may break the "line" into many segments. These segments are over-partitioned sub-clusters, which have to be post combined hierarchically. As a result, the final clustering results from blurring mean shift is heavily influenced by the post combining procedure, which makes it unsuitable to the spike sorting applications due to the difficulty of correctly guessing the number of neurons. Because of incorporating the point selection scheme and a bandwidth updating scheme as explained in Section II.E, as a comparison to blurring mean shift, EMS does not have convergence bias to directions and avoids over-partitioning a cluster into many small segments as blurring mean shift does. Consequentially, EMS can work alone without using a post combining procedure. A convergence illustration of EMS is presented in Figure 2.

D. EMS Vector and Energy Convergence

The proof of the convergence of the EMS algorithm is given in this subsection.

The gradient of E_{x_i} with respect to x_i can be obtained by exploring the linearity of kernel $K(x)$ as

$$\nabla E_{x_i} \quad (4)$$

$$= -2[\frac{\sum_{j\neq i}(\frac{x_j f(h_{x_i}) g(|\frac{x_i - x_j}{h_{x_i}}|^2)}{h_{x_i}^2} + \frac{x_j f(h_{x_j}) g(|\frac{x_j - x_i}{h_{x_j}}|^2)}{h_{x_j}^2})}{\sum_{j\neq i}(\frac{f(h_{x_i}) g(|\frac{x_i - x_j}{h_{x_i}}|^2)}{h_{x_i}^2} + \frac{f(h_{x_j}) g(|\frac{x_j - x_i}{h_{x_j}}|^2)}{h_{x_j}^2})} - x_i]$$

$$\times \sum_{j\neq i}[\frac{f(h_{x_i})}{h_{x_i}^2} g(|\frac{x_i - x_j}{h_{x_i}}|^2) + \frac{f(h_{x_j})}{h_{x_j}^2} g(|\frac{x_j - x_i}{h_{x_j}}|^2)].$$

The first bracket contains the evolving mean shift vector

$$\overrightarrow{EMS_{x_i}} = \frac{\sum_{j\neq i}(\frac{x_j f(h_{x_i}) g(|\frac{x_i - x_j}{h_{x_i}}|^2)}{h_{x_i}^2} + \frac{x_j f(h_{x_j}) g(|\frac{x_j - x_i}{h_{x_j}}|^2)}{h_{x_j}^2})}{\sum_{j\neq i}(\frac{f(h_{x_i}) g(|\frac{x_i - x_j}{h_{x_i}}|^2)}{h_{x_i}^2} + \frac{f(h_{x_j}) g(|\frac{x_j - x_i}{h_{x_j}}|^2)}{h_{x_j}^2})} - x_i. \quad (5)$$

As will be proven in Theorem 1, moving the data point along the EMS vector with length no larger than twice of the EMS vector magnitude, the energy strictly reduces.

Theorem 1 *Energy is reduced by moving the selected point according to the EMS vector.*

Proof. After the selected point x_i moves to x_i', the energy associated with x_i' is

$$E_{x_i'} = \sum_{j\neq i}(E_{x_i'.x_j} + E_{x_j.x_i'}). \quad (6)$$

In this proof, we assume that the bandwidths of all the data points remain static. The cases with adaptive bandwidth are validated in section II.E.

Without loss of generality, let $x_i = 0$. Applying the energy definition (Eq. 2) for x_i' and x_i, and considering the convexity of the kernel profile $k(x)$, the energy change of the dataset X is

$$\triangle E(X) = E_{x_i'} - E_{x_i} \quad (7)$$

$$\leq \sum_{j\neq i}(\frac{f(h_{x_i})}{h_{x_i}^2} g(|\frac{x_j}{h_{x_i}}|^2) + \frac{f(h_{x_j})}{h_{x_j}^2} g(|\frac{x_j}{h_{x_j}}|^2))(|x_i' - x_j|^2 - |x_j|^2)$$

$$= \sum_{j\neq i}(\frac{f(h_{x_i})}{h_{x_i}^2} g(|\frac{x_j}{h_{x_i}}|^2) + \frac{f(h_{x_j})}{h_{x_j}^2} g(|\frac{x_j}{h_{x_j}}|^2))(|x_i'|^2 - 2x_i' x_j)$$

Applying the definition of EMS vector for x_i (Eq. 5) and letting $G(x_i, x_j) = \frac{f(h_{x_i})}{h_{x_i}^2} g(|\frac{x_j}{h_{x_i}}|^2) + \frac{f(h_{x_j})}{h_{x_j}^2} g(|\frac{x_j}{h_{x_j}}|^2)$ results

$$\triangle E(X) = (\sum_{j\neq i} G(x_i, x_j))(|x_i'|^2 - 2x_i' \overrightarrow{EMS_{x_i}})$$

Since $\sum_{j\neq i} G(x_i, x_j)$ is strictly positive, to guarantee the energy reduction, it is required that

$$|x_i'|^2 - 2x_i' \overrightarrow{EMS_{x_i}} = |x_i' - \overrightarrow{EMS_{x_i}}|^2 - |\overrightarrow{EMS_{x_i}}|^2 \leq 0 \quad (8)$$

Particularly, $|x_i'|^2 - 2x_i' \overrightarrow{EMS_{x_i}}$ achieves the minimal value of $-|\overrightarrow{EMS_{x_i}}|^2$ when $x_i' = \overrightarrow{EMS_{x_i}}$. This completes the proof.

E. Bandwidth Updating

To calculate the local bandwidth, a pilot density estimate is first calculated as

$$p(x_i) = \frac{1}{h_0^d} \sum_{j\neq i} K(\frac{x_i - x_j}{h_0}), \quad (9)$$

where h_0 is a manually specified global bandwidth and d is the dimension of the data space. Based on the pilot density estimate, local bandwidths are updated as [11]

$$h_{x_i} = h_0[\frac{\lambda}{p(x_i)}]^{0.5}, \quad (10)$$

where $p(x_i)$ is the estimated density at point x_i, λ is a constant that is by default assigned to be the mean of $\{p(x_i)\}$.

In each EMS iteration, the density estimate associated with the selected point is updated using a sample point density with bandwidth estimated from Eq. 10 as

$$p(x_i') = \sum_{j\neq i} \frac{1}{h_{x_j}^d} K(\frac{x_i' - x_j}{h_{x_j}}). \quad (11)$$

The procedure of updating the bandwidth is summarized as follows:

Algorithm 2 Adaptive bandwidth updating using sample point estimator

---

Input: The data point x_i^k that is selected to move in the k^{th} iteration and its corresponding bandwidth $h_{x_i}^k$

Output: An updated bandwidth $h_{x_i}^{k+1}$ for the selected point

- Calculate the updated density estimate $p(x_i^{k+1})$ for the selected point according to Eq. 11.
- Calculate the updated bandwidth $h_{x_i}^{k+1}$ for the selected point using Eq. 10 with the updated pilot density estimate $p(x_i^{k+1})$. If $h_{x_i}^{k+1} < h_{x_i}^k$, update the bandwidth with $h_{x_i}^{k+1}$; otherwise, set $h_{x_i}^{k+1} \leftarrow h_{x_i}^k$

During iterations, the bandwidth of each data point adapts to the local density. Though Algorithm 2 only updates bandwidth when it becomes smaller, experiments show that the bandwidth associated with the selected point frequently reduces after the movement. This phenomenon is intuitive, as the EMS iteration compacts a dataset, which leads to a smaller bandwidth according to Eq. 10. To satisfy that $E_{x_j.x_i}$ is a monotonically increasing function of $h(x)$, we have $\frac{\partial E_{x_j.x_i}}{\partial h_{x_i}} \geq 0$. For both Gaussian and Epanechnikov kernels, the requirements on $f(h_x)$ are the same

$$f(h_x) \sim O(h_x^\alpha),\ \alpha \geq 2 \quad (12)$$

F. Stopping Criterion

In this work, we use a broad truncated kernel with an adaptive bandwidth, based on which a reliable stopping criterion using the EMS vector or the total energy can be given. A broad truncated kernel $K_B(x)$ is defined as

$$K_B(x) = \begin{cases} K(x), & x < Mh_x \\ 0, & x \geq Mh_x, \end{cases} \quad (13)$$

where M is a positive constant satisfying that Mh_x can cover a large portion or the whole feature space. At an early stage of EMS iterations where a clear configuration of clusters has not been formed, the kernel K_B is similar to a broad kernel. Through iterations, the bandwidth reduces and converges to zero. As a result, K_B becomes a truncated kernel, which only covers a small region in the feature space and prevents the attraction of different clusters.

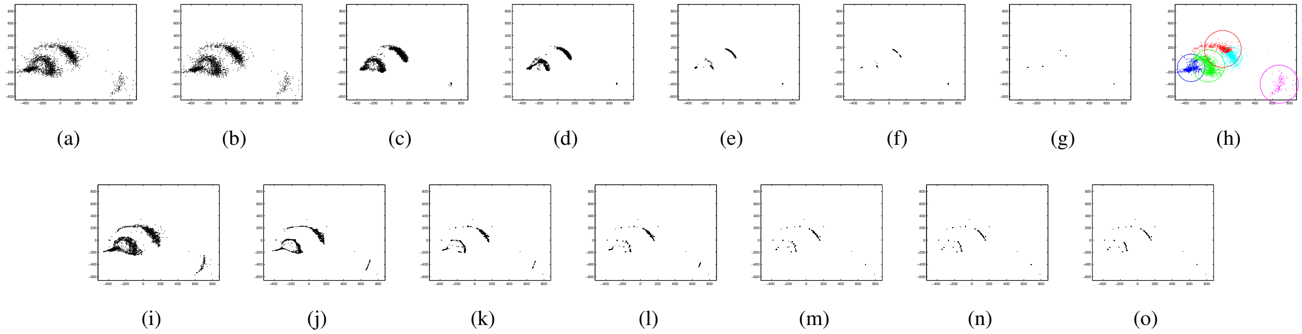

Fig. 2. (a) - (g) display the snapshots of EMS at $0, 0.5, 1, ..., 3$ iterations per point. The grouping results shown in Figure 2 (h) are obtained through 5 isolated modes in Figure 2 (g). (i) - (o) display the snapshot of BMS at $2, 4, ..., 14$ iterations, as a comparison.

TABLE I

ACCURACY EVALUATION OF THE PROPOSED SPIKE SORTING METHOD

Sequence Number	1	2	3	4	5	6	7	8
Informative Samples	97.8%	97.8%	97.8%	97.0%	98.0%	99.2%	96.6%	92.0%

Note: Informative samples are harvested from both spikes and their derivatives. 3000 spikes each sequence from [3].

III. EXPERIMENTS

A. Spike Clustering with Animal Data

Extracted spike features from animal recordings that can deliver typical challenges of the application (irregular cluster geometry, density variation, sparse region, noise events, etc.) are used to compare the performance of EMS and burring mean shift. The results are shown in Figure 2. In Figure 2 (a) - (g), snapshots of EMS iterations at $0, 0.5, 1, ..., 3$ iterations per point are displayed. In Figure 2 (h), the grouping results based on 5 isolated modes in Figure 2 (g) are plotted.

As a comparison, the snapshots of blurring mean shift at $2, 4, ..., 14$ iterations are presented in Figure 2 (i) - (o). The first 4 - 6 iterations collapse the data into "lines". Afterwards, the convergence speed dramatically reduces. Not only that, collapsed "lines" are clearly broken into many segments through iterations. As a result, a post processing algorithm is critical to generate clustering results.

B. Spike Sorting with Synthesized Data

Synthesized spike sequences from waveclus [3] are used to test the performance our algorithms. Feature extraction method of using informative samples [7] has been applied. After the features have been extracted, clustering is done by the EMS clustering algorithm. The sorting accuracy tests are listed in Table I. A combination of our recently developed feature extraction algorithm [7] and EMS clustering algorithm gives high accuracy score when tested on the dataset.

IV. CONCLUSION

This paper presents a nonparametric EMS clustering algorithm for spike sorting application. It defines an energy function to quantify the compactness of the dataset and iteratively collapses the data to isolated clusters in a couple of iterations. The single parameter of bandwidth is initialized based on sample point estimators and updated to accommodate the evolving procedure. The main theoretical contributions in this work are the validations of two theorems stating that the evolving mean shift procedure converges, and further, the energy reduction rate. The low computational cost and good performance makes it suitable to apply in many other practical tasks or subjects in additional to spike sorting.

REFERENCES

[1] Lewicki MS. A review of methods for spike sorting: the detection and classification of neural action potentials. Network Comput Neural Syst. 1998;9:53–78.

[2] Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature. 2006;442:164–171.

[3] Quian Quiroga R, Nadasdy Z, Ben-Shaul Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Computation. 2004 Aug;16(8):1661–1687.

[4] Yang Z, Chen T, Liu W. A neuron signature based spike feature extraction algorithm for on-chip implementation. Proc 30th Ann Int Conf IEEE EMBS. 2008 Aug;p. 4237–4240.

[5] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Analysis and Machine Intelligence. 2002 May;24(5):603–619.

[6] Chae M, Liu W, Yang Z, Chen T, Kim J, Sivaprakasam M, et al. A 128 channel 6mW wireless neural recording IC with on-the-fly spike sorting and UWB transmitter. IEEE ISSCC Dig Tech Papers. 2008 Feb;7(6):241–261.

[7] Yang Z, Zhao Q, Liu W. Spike feature extraction using informative samples. Advances in Neural Information Processing Systems 21. 2009;p. 1865–1872.

[8] Hoang L, Yang Z, Liu W. VLSI architecture of NEO spike detection with noise shaping filter and feature extraction using informative samples. Proc 31th Ann Int Conf IEEE EMBS. 2009 Sep;.

[9] Fukunaga K, Hostetler LD. The estimation of the gradient of a density function, with application in pattern recognition. IEEE Transactions on Information Theory. 1975;21:180–187.

[10] Cheng YZ. Mean shift, mode seeking, and clustering. IEEE Transactions Pattern Analysis and Machine Intelligence. 1995 August;17(8):790–799.

[11] Comaniciu D, Ramesh V, Meer P. The Variable Bandwidth Mean Shift and Data-Driven Scale Selection. IEEE International Conference on Computer Vision. 2001;p. I: 438–445.