Multiple Event Detection and Recognition for Large-scale Power Systems through Cluster-based Sparse Coding

Yang Song, *Student Member, IEEE,* Wei Wang, *Student Member, IEEE,* Zhifei Zhang, *Student Member, IEEE,* Hairong Qi, *Senior Member, IEEE* and Yilu Liu, *Fellow, IEEE*

***Abstract*—Accurate event analysis in real time is of paramount importance for high-fidelity situational awareness such that proper actions can take place before any isolated faults escalate to cascading blackouts. Existing approaches are limited to detecting only single or double events or a specified event type. Although some previous works can well distinguish multiple events in small scale systems, the performance tends to degrade dramatically in large-scale systems. In this paper, we focus on multiple event detection, recognition, and temporal localization in large-scale power systems. We discover that there always exist "regions" where the reaction of all buses to certain event within each region demonstrates high degree similarity, and that the boundary of the "regions" generally remains the same regardless of the type of event(s). We further verify that, within each region, this reaction to multiple events can be approximated as a linear combination of reactions to each constituent event. Based on these findings, we propose a novel method, referred to as *cluster-based sparse coding* (CSC), to extract all the underlying single events involved in a multi-event scenario. Multiple events of three typical disturbances (e.g., generator trip, line trip, and load shedding) can be detected and recognized. Specifically, the CSC algorithm can effectively distinguish line trip events from oscillation which has been a very challenging task for event analysis. Experimental results based on simulated large-scale system model (i.e., NPCC) show that the proposed CSC algorithm presents high detection and recognition rate with low false alarms.**

***Index Terms*—Multi-event analysis, cluster-based sparse coding, frequency measurement, large-scale power systems, classification-based method.**

I. Introduction

CONVENTIONAL US power system planning is typically performed with the consideration of possible N-1 contingency events of intrinsic equipment failure, where N-1 contingency means one component (or one set of closely related components) fails. Occasionally, some regions of the US power system are designed to handle a few critical N-2 or even N-3 contingencies (i.e., simultaneous failures of 2 or 3 components) with the assistance of post-contingency remedial actions. Nevertheless, recent system disturbance reports [1] from the North American Electric Reliability Corporation (NERC) have made it obvious that major disturbances typically involve a number of unlikely, unplanned events, which will lead to N-X operations under emergency, making multiple event analysis of paramount importance.

In power systems, cascading or simultaneous faults/events are common problems that may lead to large area blackout. Recent large-scale failures of the grid, such as the August 2003 U.S. northeastern blackouts and the July 2012 India blackouts, indicate that wide-spread blackouts are always initiated from sporadic fault events, but can cause devastating effect on our everyday life [2]. In order to ensure safe and healthy operations of the power system, the wide-area situational awareness system (WASA) can essentially provide high-resolution understanding of the power system dynamics, such that disturbance can be better mitigated in time before it escalates to any unexpected cascading blackouts [3] through the butterfly effect.

In recent decades, many works related to disturbance or event detection and recognition have been reported. Various types of signals have been used for analysis purpose, including frequency [4]–[9], power and voltage [10], and phasor angle [11], [12]. Generally, event analysis approaches fall into two categories: model-based and classification-based. Model-based methods (e.g., [10], [13], [14]) use many power system parameters to model the power grid, making it inflexible to structure changes and the algorithm might be only applicable to one type of signal since different events cause different effects on voltage, current, frequency and power angle readings. For these reasons, classification-based approaches are more popular [6], [15]–[21]. Nonetheless, they are only suitable for single event classification as the amount of classes may increase exponentially in multi-event problems. In fact, although any single event can be treated as a distinctive pattern, the combination of two or more single events can theoretically generate infinite number of patterns.

Existing research on multiple event analysis is limited from several perspectives. For example, Ref. [2] provided only the warning of line failures in the transmission system of the power grid. Ref. [22] only focused on line trip disturbance. Similarly, Ref. [11] could handle multiple events but only line outages. Ref. [6] only successfully distinguished sequential events if their occurrence time is sufficiently apart such that the event signal can be segmented into ones from multiple single events. Ref. [23] proposed the nonnegative sparse event

Manuscript received January 12, 2016; revised June 14, 2016, September 24, 2016 and December 21, 2017; accepted January 15, 2017. Date of current version is December 21, 2016. This work was supported in part by the Engineering Research Center Program of the National Science Foundation and the Department of Energy under NSF Award Number EEC-1041877, in part by the CURENT Industry Partnership Program, and in part by NSF-CNS-1239478. Paper no. TPWRS-00079-2016. The author is with the Department of Electrical and Computer Engineering, University of Tennessee, Knoxville, TN, 37996, USA e-mail: (ysong18@vols.utk.edu; wwang34@utk.edu; zzhang61@vols.utk.edu; hqi@utk.edu; liu@utk.edu).

unmixing (NSEU) approach which can extract each single event from the mixed multi-event, but it only performed well in small power systems.

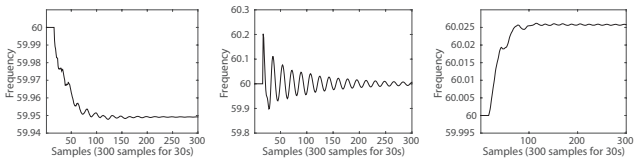
In this paper, we focus on multi-event detection, recognition and temporal localization for large-scale power systems. We choose to analyze the frequency signal because events like generator trips, line trips, or load shedding immediately cause sudden frequency changes within the power system. The proposed approach is fundamentally a classification-based method.

The main contribution of the paper is three-fold: 1) we provide a comprehensive analysis of the challenges for multiple events analysis in large-scale power systems; 2) we propose a new scheme—cluster-based sparse coding (CSC) that takes advantage of the characteristic of the power system having similar reaction to events within a local region and incorporates it into the problem formulation to effectively perform multiple event analysis; and 3) we conduct thorough evaluation of the CSC method using the simulated data of the Northeast Power Coordinating Council (NPCC) system.

The rest of the paper is organized as follows: Section II analyses the challenge of multiple events analysis in large-scale power systems. Section III briefly reviews the basic theory of sparse coding. Section IV elaborates on the proposed cluster-based sparse coding algorithm. Section V demonstrates the superior performance of this algorithm based on simulated data of the NPCC system. Finally, summary and conclusion, as well as future work, are discussed in Section VI.

II. Challenge

When multiple events occur in cascading fashion, the electromechanical waves generated will interfere with each other, and the frequency measurement taken at a certain bus would more than likely be a "mixture" of multiple frequency signals. Thus, how to determine the number of events that occurred and identify the types of events with precise estimation of event occurrence time using simply the observed mixture is a very challenging problem. In the following, we describe the challenges from the perspectives of oscillation, intra-class variance, inter-class similarity of the power grid, and the large dynamic range of event amplitudes.

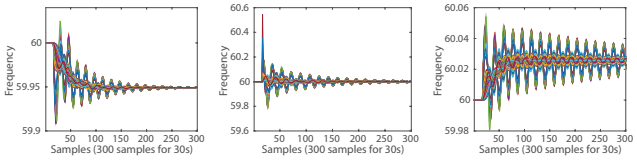


(a) GT (b) LT (c) LS

Fig. 1. Frequency signals on one bus corresponding to three typical event types. GT: generator trip, LT: line trip, LS: load shedding.

A. Oscillation

According to the disturbance events detected by the FNET/GridEye system [24] (a world-wide frequency network jointly developed at the University of Tennessee and Oak Ridge National Laboratory that monitors the power grid at the distribution level), a typical power system event falls into one of the three categories—generator trip (GT), load shedding (LS), and line trip (LT), as shown in Fig. 1. Generally speaking, these three types of events also come with strong oscillation [25]. For GT and LS, the oscillation can be treated as noise; for LT, however, not only the frequency pattern of oscillation resembles that of LT, the scale of oscillation is also comparable to that of LT. Therefore, oscillation presents great challenge to the detection of LT. In addition, oscillation follows various decay rates, making it even harder to distinguish it from an actual LT event. Note that, in general, oscillations can also be counted as a type of event. However, it carries different characteristics from the other three events. While the frequency measure of GT, LT, and LS have distinctive signatures (patterns), oscillation does not. Since the proposed approach is pattern-based, we have excluded oscillation from the list of detected events. In practical situational awareness systems, e.g., FNET/GridEye, a separate oscillation trigger [26] is used independent of the event trigger which mainly detects the three major disturbances, i.e., GT, LT, and LS.



(a) GT (b) LT (c) LS

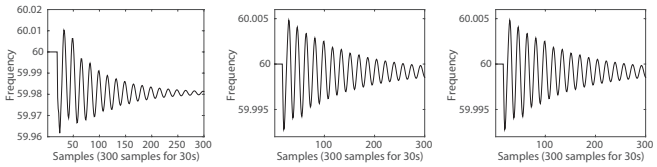
Fig. 2. Frequency signals on several buses (plotted in different colors) responding to the same single event present large intra-class variance.

B. Intra-class Variance

In a small power system, frequency signals of all buses responding to the same event always appear similar. When it comes to a larger and more complicated power system, however, reaction caused by an event can vary largely at different buses. This phenomenon is referred to as the large *intra-class variance*, that is, event signals belong to the same class show substantial variations, as shown in Fig. 2. This feature of large power systems raises a challenging issue—which signal pattern(s) should be used to represent the behavior of an event? In addition, the same event type caused by different devices at different physical locations may also yield different frequency patterns, adding more difficulty to the event identification problem.

C. Inter-class Similarity

Quite opposite to the intra-class variance, different types of events may result in similar response on certain buses, which is illustrated in Fig. 3. This is referred to as the large *inter-class similarity*, presenting yet another level of difficulty for event identification.

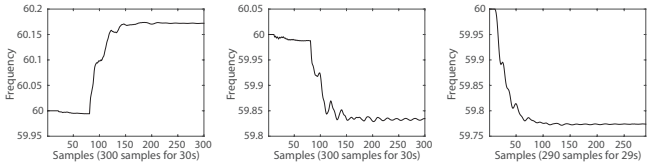


(a) GT (b) LT (c) LS

Fig. 3. Frequency signals display similar patterns responding to different single events.

D. Unbalanced Events

In multi-event cases, a small disturbance accompanied with large ones is referred to as an unbalanced multi-event, which usually results in missed detections since those small disturbances are easy to be considered as noise. Here, "small" and "large" refer to the amplitude of the frequency disturbance caused by an event. From the three scenarios shown in Fig. 4, we observe that the small scale disturbance is submerged by the large scale disturbance.



(a) GT + LS (b) GT + GT (c) GT + LT

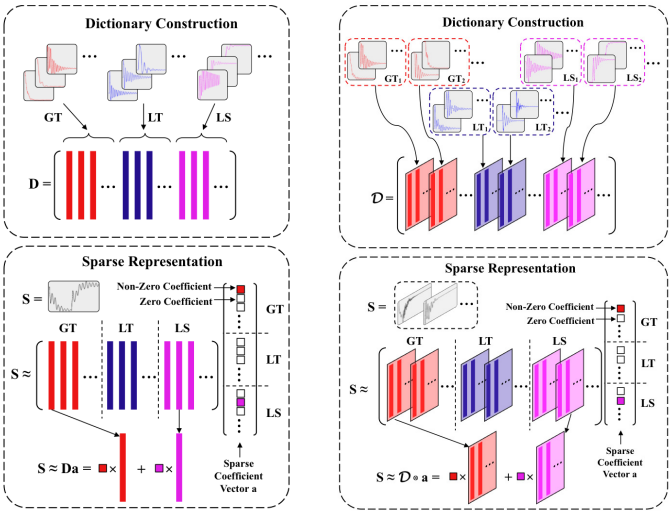
Fig. 4. Three cases with unbalanced concatenated events: (a) a small GT occurred at the 2nd second and a large LS at the 8th second; (b) a small GT occurred at the 1st second and a large GT at the 8th second; (c) a large GT occurred at the 1st second and a small LT at the 7th second.

III. Related Work

Wang et al. [8], [23] proposed the nonnegative sparse event unmixing (NSEU) method, which are the first few works on multiple event analysis. NSEU uses the frequency signal obtained at the distribution level for event unmixing where they consider the frequency signal caused by a multi-event case as a linear combination of signals from each individual event, as modeled in Eq. (1). In this linear mixing model, $\mathbf{s}$ refers to the multi-event mixed signal, $\mathbf{D} = [\mathbf{d_1}, \mathbf{d_2}, \cdots, \mathbf{d_k}]$ is an overcomplete dictionary whose columns, $\mathbf{d_i}$ $(1 \leq i \leq k)$, denote the representative single event signals, or the "root patterns", making up the observed mixture, $\mathbf{s}$, and $\mathbf{a}$ is the coefficient vector, indicating the weight of each single event signal. For example, the three typical power system events–GT, LS, LT–can be referred to as the "root faults/patterns". In addition, for the purpose of locating the event occurrence time, each root pattern in the dictionary is extended to a series of "time-shifted" version of the root patterns [8], forming a group of root faults with exactly the same shape but different starting time. The diagram of constructing the dictionary can be found in the top of Fig. 5(a).

$$\mathbf{s} = \mathbf{Da} \tag{1}$$

$$\begin{aligned} &\arg\min_{\mathbf{a}} \left\{ \|\mathbf{s} - \mathbf{Da}\|^2 + \lambda \|\mathbf{a}\|_1 \right\} \\ &\text{s.t. } \mathbf{a} \succeq 0 \end{aligned} \tag{2}$$



(a) Basic sparse coding (b) Cluster-based sparse coding

Fig. 5. Illustration of the dictionary construction (top row) and sparse representation (bottom row) processes of two methods: (a) the NSEU method in [23], and (b) the proposed cluster-based sparse coding.

In a real-world multi-event scenario, the number of the constituent single event should be very few, e.g., two generator trips concatenated by three line trips [8]. Therefore, compared to the dimension of the dictionary, which is around 10,800 as used in NSEU, the number of non-zero elements in the mixing coefficient $\mathbf{a}$ should be very small, or "sparse".

Sparse coding, which models data vectors as sparse linear combinations of basis elements, has been widely used to solve decomposition problems in many areas, such as machine learning, signal processing, and statistics [27]. For multiple events analysis, sparse coding aims to solve the optimization problem in Eq. (2), where the sparse coefficient $\mathbf{a}$ is sought in order to minimize the reconstruction error, $||\mathbf{s}-\mathbf{Da}||^2$. Usually, L_1 norm, $||\mathbf{a}||_1$, is employed as the penalty term to enforce a sparse $\mathbf{a}$. Many existing sparse coding algorithms like [28]–[33] can be employed to solve the sparse decomposition problem. In addition, based on the physical meaning of this application, the coefficient vector $\mathbf{a}$ should be non-negative. The overview of the basic sparse coding scheme for multi-event detection is shown in Fig. 5(a). The non-zero coefficients in $\mathbf{a}$ indicate which single events in the dictionary are involved in making up this multi-event case. For example, in the bottom of Fig. 5(a), the first non-zero coefficient in $\mathbf{a}$ indicates there is a GT happened and the frequency profile of the event is similar to the first root pattern in the dictionary $\mathbf{D}$. This algorithm has been successfully applied to small power systems, but less effective to large-scale power systems.

IV. Cluster-based Sparse Coding

A new approach, CSC, is thus proposed here to solve the challenging issues present in large-scale power systems that usually consist of hundreds or even thousands of buses. Assume there are n buses in a large-scale power system, and m single event tests (e.g., GT, LS, LT) are implemented to obtain the training data of single disturbances. In each individual test,

the time-series frequency signals from n buses are recorded in the format of Eq. (3), where $\mathbf{S}_i \in \mathbb{R}^{l\times n}$ $(i = 1, 2, \cdots, m)$ is a matrix that denotes the signals of all buses collected in the ith test, and $\mathbf{s}_{ij}$ $(j = 1, 2, \cdots, n)$ is a column vector with l elements that denotes the signal sequence on the jth bus in the ith test.

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{s}_{i1} & \mathbf{s}_{i2} & \cdots & \mathbf{s}_{in} \end{bmatrix} \tag{3}$$

After m tests, all signals are collected and organized as Eq. (4), where $\mathbf{s}_{*j} = \{\mathbf{s}_{ij}\}_{i=1}^{m} = \{s_{1j}, \cdots, s_{ij}, \cdots s_{mj}\}^T, (i = 1, \cdots, m)$. Obviously, $\mathbf{s}_{*j}$ represents all the signals collected from the m tests on the jth bus. $\mathbf{s}_{*j}$ is still a column vector formed by concatenating all $\{\mathbf{s}_{ij}\}_{i=1}^{m}$. We denote the signals in this way in order to facilitate the hypothesis test where we would discover the existence of groups/clusters of buses that react similarly regardless of the event type.

$$\begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_m \end{bmatrix} = \begin{bmatrix} \mathbf{s}_{11} & \mathbf{s}_{12} & \cdots & \mathbf{s}_{1n} \\ \mathbf{s}_{21} & \mathbf{s}_{22} & \cdots & \mathbf{s}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_{m1} & \mathbf{s}_{m2} & \cdots & \mathbf{s}_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_{*1} & \mathbf{s}_{*2} & \cdots & \mathbf{s}_{*n} \end{bmatrix} \tag{4}$$

A. Hypothesis

In order to reduce the large intra-class variance as well as the inter-class similarity so as to improve the recognition accuracy, certain smoothing is needed such that the characteristics of events of different classes can be more distinctive and those of the same classes can be more similar. We hypothesize that buses within close vicinity should react similarly to different events. We have conducted comprehensive experiments to validate our hypothesis—as illustrated in Fig. 6—where a clustering algorithm, k-means, is adopted to group the frequency responses of the 140 buses in the NPCC system. We observe from Fig. 6 that for certain event, buses within certain vicinity do belong to the same cluster, and that this membership generally remains the same regardless of the different events occurred. This experiment shows strong evidence that well validates the hypothesis of the existence of clusters and the possibility of performing the smoothing within each cluster in order to suppress the large intra-class variance while amplifying the inter-class difference.

B. Selection of Cluster Number

The cluster number k is chosen based on the correlation of frequency signal from all buses. The correlation is based on the Pearson correlation coefficient as Eq. (5). A and B are two variables with N scalar observations. And μ and σ are the mean and standard derivation.

$$\rho(A, B) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right) \tag{5}$$

From the training set, based on each case study (i.e., each event case data), we can calculate the correlations between buses which form a matrix of 140×140 dimension. We randomly select 100 cases to compute the correlation between buses. Fig. 7 is generated by averaging all 100 correlation matrices. We observe from the figure that there are roughly 5 or 6 highlighted (yellow) blocks along the diagonal, which indicate high correlation. We perform empirical study by assuming k equals to 4, 5, and 6. The performance when $k = 5$ is better than that when $k = 4$. The performance is similar between $k = 5$ and $k = 6$. However, since $k = 6$ increases the computational complexity, we choose $k = 5$ to be the final number of clusters.

C. Preprocessing

Based on the hypothesis analysis in Sec. IV-A, as well as the method of selecting an appropriate cluster number k in Sec. IV-B, we conduct a critical preprocessing procedure—smoothing within clusters. We first group the column vectors, i.e., the frequency responses from each individual bus, $\{\mathbf{s}_{*j}\} = \{\mathbf{s}_{*1}, \mathbf{s}_{*2}, \cdots, \mathbf{s}_{*n}\}$, using the k-means clustering method, where we assume k $(k < n)$ clusters are formed. Then a simple smoothing operation is conducted by averaging the column vectors that belong to the same cluster. Mathematically, we design a so-called "indicator matrix", $\mathbf{X}$ (Eq. (6)), to record the cluster assignment information of each bus.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}_{n\times k} \tag{6}$$

The indicator matrix, $\mathbf{X}$, is of $n \times k$ dimension, where n is the number of buses and k is the number of clusters. Each element in the indicator matrix, x_{ij} can only take on a binary value with 1 indicating the ith bus belonging to the jth cluster and 0 indicating no-membership. Therefore, one and only one element in each row of $\mathbf{X}$ is 1 (assuming no overlapping between cluster memberships) and that all the column vectors $\mathbf{x}_j$'s $(j = 1, 2, \cdots, k)$ are orthogonal to each other.

Based on the indicator matrix, X, the smoothing within cluster operation can be expressed in Eq. (7), where $\mathbf{e}_{1\times n}$ denotes a $1\times n$ all-one vector. The function "diag()" constructs an $n \times n$ diagonal matrix with elements of the input vector on the main diagonal.

$$\mathbf{S}'_i = \mathbf{S}_i\mathbf{X}\text{diag}^{-1}\left(\mathbf{e}_{1\times n}\mathbf{X}\right) \tag{7}$$

In Eq. (7), the first component $\mathbf{S}_i\mathbf{X}$ returns the accumulated frequency data from buses within the same group. And by multiplying with $\text{diag}^{-1}\left(\mathbf{e}_{1\times n}\mathbf{X}\right)$, $\mathbf{S}'_i$ returns the average frequency reading of all signals within their groups. Eq. (7) reduces or projects the number of signals from n to k (i.e., from $\mathbf{S}_i$ to $\mathbf{S}'_i$) which is much smaller than the original number of buses. $\mathbf{S}'_i$ will then replace $\mathbf{S}_i$ as the representative reaction of that cluster of buses to certain events. This largely reduces computational complexity and at the same time preserves intra-class characteristics while amplifying inter-class differences, leading to potentially higher recognition accuracy.

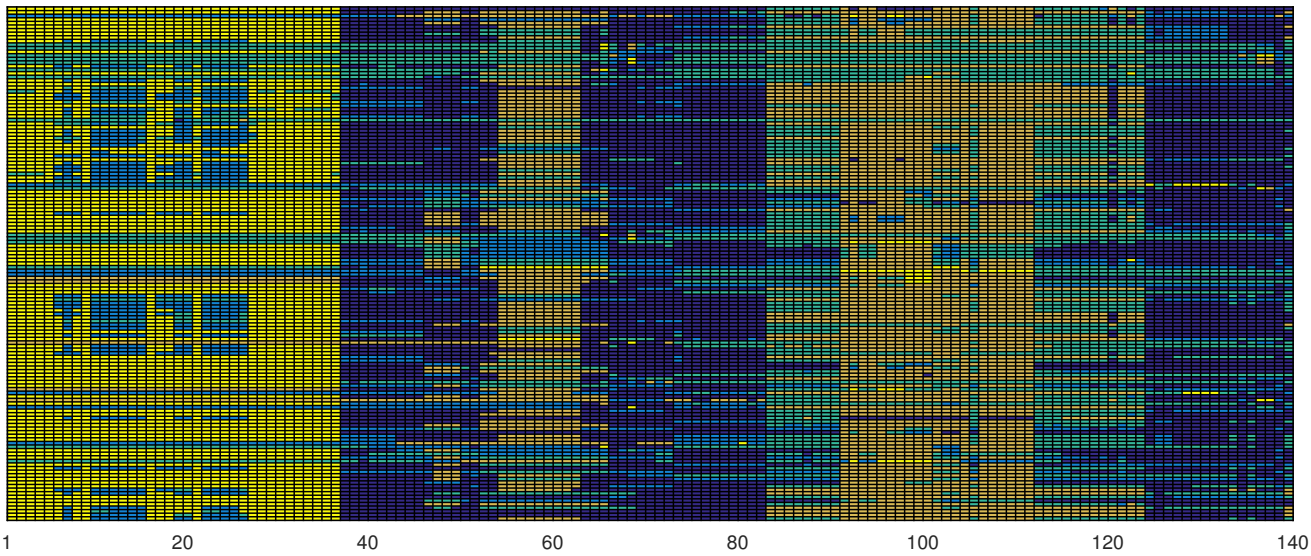


Fig. 6. Validating the hypothesis that some groups of buses show similar response to all types of events. The horizontal axis indicates the 140 buses in the NPCC system, and the vertical axis denotes the 144 tests that are GT, LT or LS. For each event, the reaction of all buses are separated into five clusters through k-means. The colors demonstrate the membership of the buses. Generally, the buses show stable membership for different types of events.

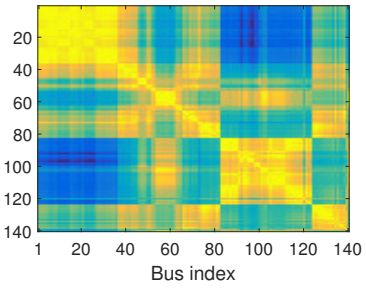


Fig. 7. The averaged correlation matrix of 140 buses from 100 randomly selected training cases. The x- and y-axis are indices of buses.

D. Dictionary Construction

After the preprocessing step that generates the representative signal, $\mathbf{S}'_i$, from each cluster, the next critical component in linear unmixing is to construct the dictionary, $\mathbf{D}$, as shown in Eq. (1). As described in Sec. II, events of the same type generally share similar frequency characteristics. For example, the GT event usually starts with a frequency drop, the LS event starts with a frequency increase, and LT follows a well-damped frequency oscillation. Meanwhile, due to the intra-class variance, the same type of events also possess certain degree of differences. In order to learn the most representative frequency patterns for each category of event, we apply the "mean shift clustering" [34] algorithm on $\mathbf{S}'_i$ to automatically learn r discriminative patterns. Mean shift clustering is a non-parametric clustering algorithm where the number of clusters can be automatically determined based on an input parameter (i.e., the bandwidth) that specifies the minimum distance between two clusters. We refer to these r discriminative patterns as the "root patterns". For the basic sparse coding approach as used in NSEU [8], each root pattern is denoted as a column vector, and a dictionary is built by stacking all root patterns into a two-dimensional matrix $\mathbf{D}$, as shown in the top of Fig. 5(a). However, in the proposed CSC algorithm as shown in the top of Fig. 5(b), each root pattern is denoted as matrix $\mathbf{D}_i \in \mathbb{R}^{l\times k}$ $(i = 1, 2, \cdots, r)$. Since the dictionary is stacked by two-dimensional matrices $\mathbf{D}_i$, it can be represented as shown in Eq. (8) which is a three-dimensional matrix or so called "tensor matrix".

$$\mathcal{D} = \{\mathbf{D}_1, \cdots, \mathbf{D}_i, \cdots, \mathbf{D}_r\} = \{\mathbf{D}_i\}_{i=1}^{r} \in \mathbb{R}^{l\times k\times r} \quad (8)$$

In other word, in the basic sparse coding approach, only one layer of grouping is conducted where the power grid is treated as a flat system, i.e., each bus in the system is considered an independent individual participating in the recognition task. In the proposed CSC approach, two layers of grouping are conducted where the power grid is treated as a hierarchy with different buses grouped into clusters first and the representative group response is used to participate into the recognition task.

E. Cluster-based Sparse Representation

The cluster-based sparse coding is developed, aiming at overcoming the challenges mentioned in Sec. II. Different from the basic sparse coding scheme as shown in the bottom of Fig. 5(a) and Eq. (2), each root pattern is a group of frequency patterns in cluster-based sparse coding, whose scheme is illustrated in the bottom of Fig. 5(b).

Given $\mathbf{S}_{test} = [\mathbf{s}_{test,1}, \mathbf{s}_{test,2}, \cdots, \mathbf{s}_{test,n}] \in \mathbb{R}^{l\times n}$ that denotes the raw signals of n buses from a new test, with each column vector representing the l-element frequency signal sequence on one bus. The preprocessing procedure as shown in Eq. (7) can be applied to extract signals $\mathbf{S}'_{test} = \left[\mathbf{s}'_{test,1}, \mathbf{s}'_{test,2}, \cdots, \mathbf{s}'_{test,k}\right] \in \mathbb{R}^{l\times k}$, where $k < n$ and k is the number of clusters that groups the power grid into subsystems showing strong group behavior. Then, the sparse coding algorithm can be employed to fit $\mathbf{S}'_{test}$ using root patterns in the dictionary.

The signal needed to be analyzed is $\mathbf{S}'_{test} \in \mathbb{R}^{l\times k}$. The final goal is to find the sparse coefficients $\mathbf{a} \in \mathbb{R}^r_+$ that can well reconstruct $\mathbf{S}'_{test}$ as shown in Eq. (9).

$$\mathbf{S}'_{test} \approx \hat{\mathbf{S}} = a_1\mathbf{D}_1 + a_2\mathbf{D}_2 + \cdots + a_r\mathbf{D}_r \tag{9}$$

where $\hat{\mathbf{S}}$ is the reconstructed signal matrix, and a_i $(i = 1, 2, \cdots, r)$ is the ith element of the sparse coefficient vector $\mathbf{a}$. We need to minimize $\|\mathbf{S}'_{test} - \hat{\mathbf{S}}\|_F$, where $\|.\|_F$ is the Frobenius norm. Denote $\sum_{i=1}^{r} a_i\mathbf{D}_i$ as $\mathcal{D} \otimes \mathbf{a}$, we can write the objective function of cluster-based sparse coding as in Eq. (10),

$$\begin{aligned} &\arg\min_{\mathbf{a}} \left\{ \|\mathbf{S}'_{test} - \mathcal{D} \otimes \mathbf{a}\|_F^2 + \lambda\|\mathbf{a}\|_1 \right\} \\ &\text{s.t.} \quad \mathbf{a} \succeq 0 \end{aligned} \tag{10}$$

where $\lambda > 0$ is the sparse parameter to adjust the weight between the sparsity of $\mathbf{a}$ and reconstruction error. λ strongly impacts the solution of sparse coding. If λ is too large, it will be difficult to fit those signals with strong oscillation; if λ is too small, $\mathbf{a}$ will not be sparse enough to guarantee the identification accuracy. In the experiments, we start with a relatively large λ, which is then reduced until the reconstruction error is below a tolerance.

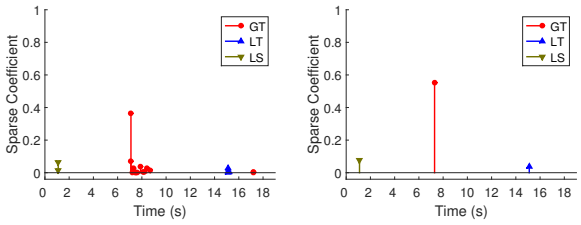
Eq. (10) can be solved by many methods, such as Basis Pursuit (BP) [33], Matching Pursuit (MP) [32], Lasso [28], etc. Based on the feature-sign search algorithm [29], we use a nonnegative feature-sign search method [35] to ensure a nonnegative sparse coefficient vector $\mathbf{a}$. The computational complexity of this algorithm is $O(r)$ where r is the dictionary size. Vectorization method is used to significantly simplify this tensor sparse coding procedure. The vectorization of $\mathcal{D} \in \mathbb{R}^{l\times k\times r}$ is to vectorize each root pattern matrix into a vector by concatenating the columns together, thus converting the three-dimensional matrix $\mathcal{D}$ to a two-dimensional matrix $\tilde{\mathbf{D}} \in \mathbb{R}^{lk\times r}$. Similarly, matrix $\mathbf{S}'_{test} \in \mathbb{R}^{l\times k}$ is also vectorized into a vector, $\tilde{\mathbf{S}} \in \mathbb{R}^{lk}$. Then Eq. (10) is converted to Eq. (11).

$$\begin{aligned} &\arg\min_{\mathbf{a}} \left\{ \frac{1}{2}\mathbf{a}^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\mathbf{a} - \tilde{\mathbf{S}}^T\tilde{\mathbf{D}}\mathbf{a} + \lambda\|\mathbf{a}\|_1 \right\} \\ &\text{s.t.} \quad \mathbf{a} \succeq 0 \end{aligned} \tag{11}$$

F. Post-processing

Theoretically, the sparse coefficients $\mathbf{a}$ obtained from Eq. (11) can accurately indicate the single events (and their occurrence time) involved in a multi-event case. In practice, however, $\mathbf{a}$ appears to be noisy — a large coefficient is always surrounded by some small coefficients of the same event type, which would mislead the analysis of the number of events, as well as the event type, generating more false alarms. This is illustrated in Fig. 8(a) where the sparse coefficients are shown from an unbalanced triple event case involving a small-scale LS happened at the 1st second, a GT at the 8th second, and a small-scale LT at the 15th second.

In order to solve this problem, we accumulate those small coefficients clustered around the large coefficient of the same event type. A time interval t in second is set to decide how close a small coefficient to the large one need to be in order for the coefficients to be combined together.



(a) Before post-processing (b) After post-processing

Fig. 8. Effect of post-processing on a multi-event unbalance case. (a) and (b) demonstrate sparse coefficients before and after post-processing. The horizontal and vertical axes indicate the time an event occurs and the amplitude of the coefficient, respectively. The colors denote different constituent single events. The small weights corresponding to LT and LS will be increased by clustering and the noise coefficient (The rightmost bar in (a)) will be removed.

Here, we apply the mean shift clustering algorithm again, but for a different purpose. Recall that we had used the same clustering algorithm in the dictionary construction process to derive the r representative root patterns in Sec. IV-D. In the post-processing stage, we apply the mean shift clustering algorithm on the sparse coefficients to automatically group the coefficients into a number of clusters based on a bandwidth, t. In Fig. 8(b), we set $t = 3.5$. However, it is possible that two events occur almost simultaneously, under which the time interval t could be set to be a smaller value, e.g., $t = 1$ or even $t = 0$, to avoid combining two coefficients within a very short time space but actually representing two simultaneous events. In general, a smaller t results in higher false alarm rate, and vice versa. The effect of t will be further evaluated in Sec.V-C3.

After combining coefficients of the same event within the same cluster, the sparse coefficient vector becomes much more sparse than the one obtained directly from sparse coding. However, some tiny coefficients may still remain because they are at a distance greater than t from other clusters, and thus form their own clusters. Empirical study shows that a small event would yield a cluster of small coefficients that are close to each other which will be combined to increase the coefficient corresponding to the small event; whereas the noise would yield isolated coefficients that will not be combined with others and thus the coefficient would remain small. We set a threshold to filter out these small isolated coefficients in the post-processing step. In all experiments conducted in this paper, we set the threshold to be 5% of the largest coefficient, For the case study shown in Fig. 8(a), the coefficients corresponding to two small events (i.e., LS and LT) are increased by combining the close-by coefficients. However, the small GT coefficient around the 17th second is not combined with others due to the isolated nature (i.e., further distance from the other clusters) and thus would remain small. This coefficient is eventually filtered out by the thresholding process. The final sparse coefficient vector is illustrated in Fig. 8(b). Note that we use Eq. (12) to calculate the final occurrence time of the combined coefficients,

$$\hat{T} = \frac{\sum_{a \in \mathcal{C}} a \times T(a)}{w}, \quad w = \sum_{a \in \mathcal{C}} a \tag{12}$$

where a denotes certain sparse coefficient, $\mathcal{C}$ denotes the set of sparse coefficients that occur within certain time span and belong to the same event type, and $T(a)$ is the occurrence time corresponding to the sparse coefficient. Note that the clustering is not based on the value of the sparse coefficient, but based on the occurrence time of the detected events. Therefore, if there are two large events (i.e., with high sparse coefficients) occur almost simultaneously (e.g., less than 0.1s), the clustering algorithm would still group them into one cluster. On the other hand, if there are two small events (i.e., with low sparse coefficients) occur very apart from each other (e.g., more than 8s), the clustering algorithm would put them into two clusters.

V. Experiment

In this section, we conduct a series of experiments to evaluate the performance of the proposed cluster-based sparse coding (CSC) algorithm for large-scale power systems. We use the "NPCC" testbed implemented by Power System Simulator for Engineering (PSS/E) to generate simulation data of various event types. The "NPCC" testbed is based on a 140-bus, 48-machine, 233-branch model and represents the backbone transmission of the northeast region of the the Eastern Interconnection, which was involved in the 2003 blackout event. The base load level of the model is 28GW. Approximately 27 generators are modeled with detailed models and the remaining units are reduced models or dynamic equivalents [36], [37].

A. Testing Scenarios

Three different test scenarios are generated, representing the three general types of event cases, S1C, M2C, and M3C, denoting the cases with one, two, and three concatenated single events, respectively, in a signal sequence. Note that the event types and their occurrence time in a test are randomly selected, and there are over 100 tests for each type of cases. All these simulation data are generated through PSS/E by controlling the type of disturbance and the occurrence time. Moreover, in order to well demonstrate the effectiveness of the proposed new scheme, especially the ability to handle the challenges mentioned in Sec. II, various combinations of different event types are generated, such as GT followed by LS, GT with LT, or GT followed by GT in M2C. In addition, for each kind of combination, for instance, GT with LS, disturbance with various scales are involved in order to evaluate the performance of CSC in handling unbalanced event cases. Note that the LT events contain those over both long (i.e., tie line trips) and short electrical distances. Tie line trips may excite inter-area oscillations are usually with poor damping ratio, and line trips that happen within a local area may cause the local oscillations model are usually with relative large damping ratio [38], [39].

B. Performance Metrics

We adopt the same performance metrics as used in [8] for a fair comparison.

- ***Detection accuracy*** (DA): ratio between the number of detected root events/faults that actually happened and the number of total root events/faults according to the ground truth.
- ***False alarm rate*** (FA): ratio between the number of detected root events/faults not actually happened and the total number of root events/faults according to the ground truth.
- ***Root-pattern recognition rate*** (RPR): ratio between the number of correctly identified events (i.e., events with correct type of root-pattern) and the number of correctly detected events.
- ***Occurrence time deviation*** (OTD): deviation between the detected occurrence time and the ground truth.

C. Algorithm Evaluation

The proposed CSC algorithm is evaluated through six experiments. The first experiment aims to conduct an overall comparison with state-of-the-art. The second experiment studies the effect of different signal preprocessing methods. The third experiment evaluates the importance of post processing. The fourth experiment analyzes how the challenges of intra-class variance and inter-class similarity are alleviated by CSC. The fifth experiment evaluates the potential of CSC in recognizing simultaneous events (i.e., events occur within 1, 2, or 3 seconds of time span). The last experiment investigates the effect of measurement error (i.e., noise) in the input data to the algorithm performance.

1) Comparison with state-of-the-art: This experiment compares the proposed CSC method with the NSEU method in [8] on a large-scale power system.

We compare these two methods on the NPCC test dataset which contains S1C, M2C and M3C. Roughly, over 100 testing samples are created for each type of cases. The experimental results are shown in Table I. In each test, the regularization parameter λ of the CSC method is adaptively selected from $[10, 100]$ and the time span used in post-processing in order to group small sparse coefficients within a certain range is 3.5 seconds. As compared to the NSEU method, the CSC method improves the detection and recognition accuracy and decreases the false alarm for all three types of multi-event cases. Specifically, for single event detection, the CSC method achieves a perfect 100% DA and 0% FA. For multiple events detection, DA is above 90% with relatively low FA. Here, we constrain FA to be under 3% to ensure a more reliable detection, while at the same time keeping a relatively high DA and RPR. However, the NSEU method can only achieve 76% DA with 25% FA for M2C cases, and 58% DA with 17% FA for M3C.

The reasons that CSC performs much better than NSEU can be summarized from three aspects: First, in the preprocessing step (Sec. IV-C), CSC divides the buses into clusters and calculates the signal averages within the clusters to represent an event [Eq. (7)], which effectively reduces the

intra-class variation while improving the discrimination power against inter-class similarity. By performing average within each cluster (i.e., local average) instead of averaging over the entire system (i.e., global average), the dictionary atoms of CSC are much more discriminative than those of NSEU. See Sec. V-C2 for further comparison between different averaging schemes.

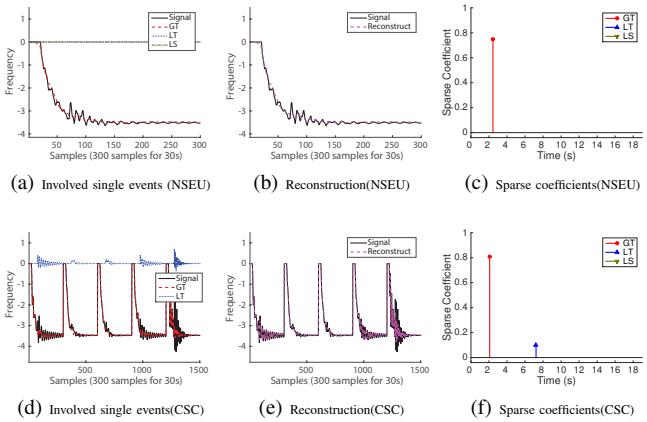
Second, in the dictionary construction and sparse representation step (Secs. IV-D and IV-E), each atom (or root pattern, $\mathbf{D}_i$) in the dictionary $\mathcal{D}$ [Eq. (8)] is a matrix instead of a column vector as in NSEU, and each column vector in $\mathbf{D}_i$ represents a cluster average of frequency readout of buses within one cluster. Since in the reconstruction process, the same sparse coefficient $\mathbf{a}_i$ is applied to $\mathbf{D}_i$, then the same coefficient would be applied to each column of $\mathbf{D}_i$, as shown in Eq. (9). That is, for small-scale events with weak patterns in some clusters but strong patterns in other clusters, the same sparse coefficient is applied in the reconstruction process, thus reducing the chance of missing weak events. This is further illustrated in the example shown in Fig. 9, where a double-event case is simulated with a GT and an LT occurred at the 2nd and the 7th second, respectively. In Fig. 9(a) and 9(b), NSEU performs the average over all buses and the averaged signal overshadows the representative feature of LT, which is why only a GT at 2.439 second is identified. On the other hand, the proposed CSC divides the signals of 140 buses into five clusters and calculates the averages of each cluster, as shown in Fig. 9(d), where we observe that the cluster-based average well preserves the discriminative feature of LT, especially in the 1st, 4th, and 5th clusters. Although the 2nd and 3rd clusters present weak LT pattern, the common sparse coefficient constraint forces these two clusters to include the LT component. As a result, CSC successfully detects a GT at 2.0836 second and an LT at 7.2986 second. The advantage of CSC effectively helps identify unbalanced events which would otherwise be missed using NSEU.

Last but not the least, the post-processing step (Sec. IV-F) in CSC combines coefficients of the same events occurred in close time span and removes isolated ones from the sparse coding results to effectively boost the DA rate and reduces the FA rate. See Sec. V-C3 for a further comparison of performance with or without the post-processing step.

TABLE I
COMPARISON RESULTS WITH THE NSEU METHOD AND THE CSC METHOD

	Number of tests	DA (%)	FA (%)	RPR (%)	OTD (s)
Method	The NSEU method [8]				
S1C	144	93.3	6.67	100	0.251
M2C	115	75.34	25.13	91.34	1.245
M3C	138	58.33	17.34	86.7	0.942
Method	The CSC method				
S1C	144	100	0	100	0.123
M2C	115	95.65	2.17	98.64	0.193
M3C	138	91.55	0.97	98.15	0.202

2) Effect of different preprocessing methods: This experiment studies the effect of different signal preprocessing methods. Three different preprocessing methods are compared,



(a) Involved single events (NSEU) (b) Reconstruction(NSEU) (c) Sparse coefficients(NSEU)

(d) Involved single events(CSC) (e) Reconstruction(CSC) (f) Sparse coefficients(CSC)

Fig. 9. A double-event case. (a) The black solid curve denotes the averaged frequency signal of the event, and the red dash curve indicates the detected GT event by NSEU. (b) The pink dash curve is the reconstruction result of NSEU. (c) The corresponding sparse coefficient and temporal location of NSEU: a GT at the 2.439 second. (d) The black solid curve is the concatenated cluster averages, and the red dash and blue dotted curves are detected GT and LT events by CSC. (e) The pink dash curve is the reconstruction result of CSC. (f) The corresponding sparse coefficients and temporal locations of CSC: a GT and an LT at the 2.0836 and the 7.2986 second, respectively.

the cluster-based averaging used in CSC (i.e., local average), the flat system-wide averaging used in NSEU (i.e., global average), and the manual selection of representative buses. Note that the manual selection is to select a bus which performs the best after comparing with the ground truth. Obviously, the manual selection approach would not be applicable in real-world problem solving. We include it here merely to provide a reference point. Table II demonstrates the experimental results. In Table II, parameters are adjusted to maintain the same DA while comparing FA, and the same FA while comparing DA, respectively.

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT PRE-PROCESSING METHODS

	Global Average	Manual Selection	Local Average
	Similar DA		
RPR (%)	99.54	99.09	98.64
FA (%)	9.13	6.09	2.17
	Similar FA		
RPR (%)	97.06	99.04	98.64
DA (%)	88.70	91.30	95.65

From Table II, we observe that the cluster-based average method always performs the best since it has lower FA and higher DA at the same time. The flat average method performs the worst because it cannot adapt to the intra-class variance of the power systems. Although the manual selection method is comparable with clustering method, manual selection is impractical in large-scale systems since it is difficult to recognize all features even for an expert.

3) On the importance of post-processing: To illustrate the effect of post-processing, we conduct further experiment by removing the post-processing part from CSC, and obtain the DA of M2C and M3C to be 91.26% (originally 95.65%) and 82.90% (originally 91.55%), respectively, as shown in

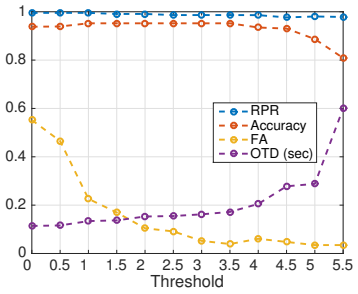


Fig. 10. The performance change based on the M2C case (with shortest time span between subsequent events being 3.5s) in the NPCC dataset with respect to different threshold value t. The vertical axis denotes RPR, accuracy, FA, and OTD (in sec), respectively, for corresponding curves.

Table III. Roughly, the post-processing increases DA by 4% and 9%, respectively, for the two multi-event cases. We also evaluate the effect of t to the algorithm performance. As shown in Fig. 10, when $t = 3.5$, the algorithm achieves the best overall performance.

TABLE III
THE DETECTION ACCURACY OF CSC WITH AND WITHOUT POST-PROCESSING

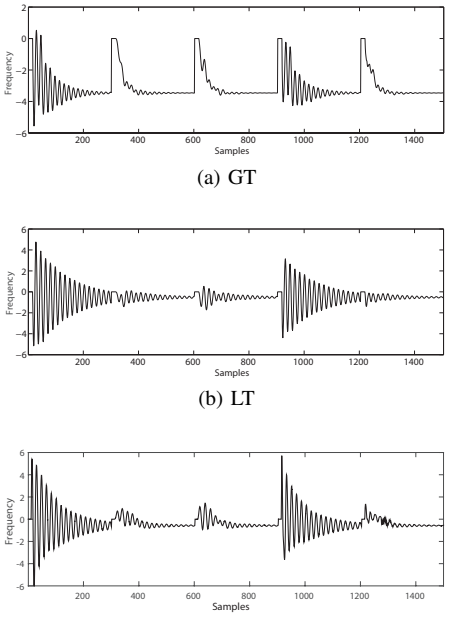
	M2C	M3C
Without Post Processing	91.26%	82.90%
With Post Processing	95.65%	91.55%

4) Handling intra-class variance and inter-class similarity: This experiment illustrates that the CSC method can deal with the challenges of intra-class variance and inter-class similarity. Fig. 11 shows some extracted signal sequences, each of which consists of five cluster averages obtained from the five clusters.

From Fig. 11(a) and (b) where S1C cases are displayed, it is easy to observe the intra-class variance since each of the five cluster averages corresponding to the same event significantly differ from each other. On the other hand, their first segments appear similar illustrating inter-class similarity. The multi-event case in Fig. 11(c) is also an unbalance case with a larger LT occurred at the 1st second, then followed by a small GT at the 8th second. Comparing the 5th (from 1200 to 1500) segment with the 1st (from 0 to 300) or the 4th (from 900 to 1200) segment, it is obvious that not all buses exhibit the unbalanced characteristic. The different frequency responses from different clusters actually help tackle the unbalanced event detection which has been elaborated in Sec. V-C1.

Even with the challenge of small inter-class variance and small intra-calss similarity, the proposed CSC method can still distinguish different types of events with high accuracy, as shown in Table IV. For GT and LS, the CSC method can achieve higher than 92.14% DA with lower than 5.88% FA. It is obvious that LT is the most difficult event to be distinguished. The main reason is that the oscillation yielded by LT varies in both frequency and amplitude. In addition, the patterns of oscillation and LT are similar.

5) Handling simultaneous multi-events: In this experiment, we demonstrate the potential of the proposed algorithm to solve N-X contingency. That is, simultaneous events occurred



(a) GT

(b) LT

(c) LT(1)+GT(8)

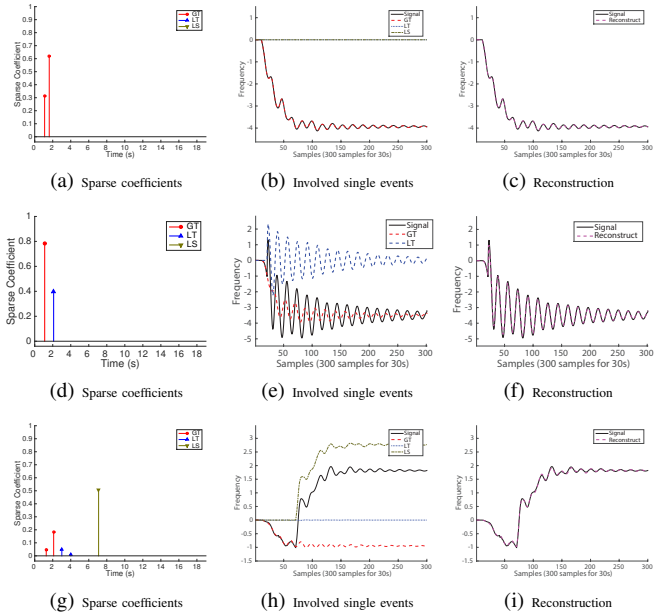
Fig. 11. Five cluster means concatenated together for three cases: (a) S1C case with 1 GT event; (b) S1C case with 1 LT event; and (c) an unbalanced M2C case with a large LT occurred at the 1st second, followed by a small-scale GT at the 8th second.

TABLE IV
EXPERIMENTAL RESULTS OF THE CSC METHOD FOR DIFFERENT EVENT TYPE

	DA (%)			FA (%)		
	GT	LT	LS	GT	LT	LS
M2C	95.89	88.24	97.75	1.37	4.41	4.49
M3C	95.77	78.68	92.14	0.70	2.21	5.88

within very short time span, e.g., 1 or 2 or 3 seconds. Three multi-event simulation cases are shown in Fig. 12. In Fig. 12(a)-(c), the ground truth is a GT happened at the 1st second, then followed by another GT at the 1.4th second. Fig. 12(d)-(f) is a double-event case with a GT happened at the 1st second, then followed by an LT at the 1.6th second. Fig. 12(g)-(i) is a multi-event case with two GT happened at the 1st, 2nd second, respectively, followed by two LT happened at the 3rd, 4th second, and then an LS happened at the 7th second. In this experiment, we set t to be zero in the post-processing procedure. We observe that CSC is able to identify all individual events that form the multi-event cases. We also conduct further experiments to analyze the performance of the proposed algorithm with multiple events occurred within various scales of time span. The experimental result is shown in Table V. The results indicate with shorter time span, the performance will generally decrease. However, we can still obtain around 60% detection accuracy and at the same time with very low false alarm rate when the time span is less than 1 second. This result demonstrates the potential of the proposed method for multiple events analysis with short time span.

6) Sensitivity to sensor estimation error: In this experiment, we explore the robustness of CSC to different levels of sensor estimation error by manually adding Gaussian noise to the



(a) Sparse coefficients (b) Involved single events (c) Reconstruction

(d) Sparse coefficients (e) Involved single events (f) Reconstruction

(g) Sparse coefficients (h) Involved single events (i) Reconstruction

Fig. 12. Three simultaneous multi-event cases. Each row is the result corresponding to one case. (a)-(c): Case 1 (M2C) detection result: a GT at the 1.10 second, a GT at the 1.59 second; ground truth is a GT at the 1st second, and another GT at the 1.4th second. (d)-(f): Case 2 (M2C) detection result: a GT at the 1.25 second, an LT at the 2.1 second; ground truth is a GT at the 1st second, an LT at the 1.6th second. (g)-(i): Case 3 (M5C) detection result: a GT happened at the 1.10 second, a GT at the 2.11 second, an LT at the 3.02 second, an LT at the 4.01th second, and an LS at the 7.148th second, respectively; ground truth is a GT at 1st second, a GT at the 2nd second, an LT at the 3rd second, and an LT at the 4th second, an LS at the 7th second.

TABLE V
THE EXPERIMENTAL RESULT WITH CSC METHOD FOR MULTIPLE EVENT CASES WITH DIFFERENT TIME SPAN.

Time Span/Perf.	DA (%)	FA (%)	RPR (%)	OTD (s)
less than 1 sec	58.33/63.58	5.56/6.17	100/98.05	0.126/0.235
less than 2 sec	81.94/82.41	4.17/5.86	100/98.52	0.113/0.139
less than 3 sec	97.22/84.57	4.17/13.89	100/98.92	0.115/0.125

Note: XX/XX means the result for M2C/M3C cases respectively. For different time span scenarios, we randomly generated more than 100 event cases for M2C and M3C, respectively. We kept all the parameters to be the same for different scenarios, and specifically the time threshold t = 0. "Less than n second" means between $n-1$ and n second.

simulated data, and we use double-event cases as an example. The result is listed in Table VI. From the statistical study of FNET [40], the frequency estimation error is around $\pm$0.1mHz which is 57dB. Note that we convert the error margin from Hz to dB by $10\times\log10(\mu/\sigma)$, where μ is 60Hz, and σ is the sensor estimation error. Table VI demonstrates that even with larger noise (e.g., 0.5mHz), the proposed CSC still maintains the same level of performance. If the noise is 40dB, the CSC method can preserve the DA, RPR, and OTD, but the FA is substantially increased.

D. Complexity and Run-time Analysis

The most time-consuming component in the proposed algorithm is the solver for the sparse coefficient. We adopt the feature-sign algorithm as the solver [29] whose computational complexity is $O(r)$, where r is the dictionary size. Since in

TABLE VI
EXPERIMENTAL RESULTS OF THE CSC METHOD FOR DIFFERENT NOISY SIGNAL BASED ON M2C.

M2C	40dB (6mHz)	50dB (0.5mHz)	57dB (0.1mHz)	Clean data
DA (%)	95.65	95.22	95.22	95.65
FA (%)	29.13	5.77	3.12	2.17
RPR (%)	96.70	98.63	98.63	98.64
OTD (sec)	0.227	0.188	0.173	0.193

sparse coding, we need to build an overcomplete dictionary, meaning the number of columns is much larger than the number of rows (i.e., $r >> l$, where l is the length of signal or window size), the bottleneck here is r instead of l. Even with the current setting of r and l, and MATLAB running on a laptop with Intel Dual Core i7 CPU 2.4 GHz, the average run time of each case still achieves 97ms, which would satisfy the real-time requirement. Note that the sampling rate is 10 samples/second.

VI. CONCLUSIONS AND FUTURE WORKS

A novel cluster-based sparse coding algorithm was proposed to analyze multi-event problems in large-scale power systems. The challenges caused by the complexity of large-scale systems were thoroughly analyzed. Then, the cluster-based sparse coding method was elaborated that effectively resolves these challenging issues. Comprehensive and thoroughly designed experiments based on simulated data of the NPCC system revealed that the proposed scheme can provide an accurate, robust and high-resolution identification of multiple events.

Benefited by the accurate detection and recognition of events, an early alarm and proper action can be made in real time to avoid large blackouts. Future research will follow four directions. First, the scheme presented in this paper needs further validation with data from other large-scale real power systems and readings other than frequency. Second, more types of events need to be involved to conduct more comprehensive investigation for the multi-event problem. Third, a better dictionary construction method is needed since a succinct and effective dictionary would largely facilitate real-time detection. Finally, in current algorithm development, different values of t (the parameter used in post-processing) need to be used for the detection and recognition of multiple events that are further or more closely apart from each other. For example, to analyze events that are at least 3.5 seconds apart, a $t = 3.5$ needs to be used; and to analyze simultaneous events where two subsequent events need to be no more than 3.5 seconds apart, a $t = 0$ needs to be used. In real-world systems, we can adopt a double-threshold approach [41] to solve this problem. That is, for any given test sample, we apply two t values in the post-processing stage. When $t = 3.5$, we only keep the detected events that are at least 3.5 seconds apart; and when $t = 0$, we only keep the detected events that are at most 3.5 seconds apart. We then combine the kept events from using the two thresholds. More studies need to be performed in this aspect in order to handle various real-world multiple event mixing scenarios.

REFERENCES

[1] U.S.Canada Power System Outage Task Force: Final Report on the implementation of task force recommendations. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/Outage_Task_Force_-_DRAFT_Report_on_Implementation.pdf

[2] S. Soltan, D. Mazauric, and G. Zussman, "Cascading failures in power grids: analysis and algorithms," in *Proceedings of the 5th International Conference on Future Energy Systems*. ACM, 2014, pp. 195–206.

[3] H. Qi, Y. Liu, F. Li, J. Luo, L. He, K. Tomsovic, L. Tolbert, and Q. Cao, "Increasing the resolution of wide-area situational awareness of the power grid through event unmixing," in *2011 44th Hawaii International Conference on System Sciences (HICSS)*, Jan 2011, pp. 1–8.

[4] L. Liu, A. Rahimpour, A. Albright, J. Guo, H. Qi, and Y. Liu, "Multivariate empirical mode decomposition based signal analysis and efficient-storage in smart grid," in *IEEE Global Conference on Signal and Information Processing*, 2016.

[5] Y. Liu, L. Zhan, Y. Zhang, P. N. Markham, D. Zhou, J. Guo, Y. Lei, G. Kou, W. Yao, J. Chai, and Y. Liu, "Wide-area-measurement system development at the distribution level: An FNET/GridEye example," *IEEE Transactions on Power Delivery*, vol. 31, no. 2, pp. 721–731, April 2016.

[6] A. Bykhovsky and J. H. Chow, "Power system disturbance identification from recorded dynamic data at the northfield substation," *International Journal of Electrical Power & Energy Systems*, vol. 25, no. 10, pp. 787–795, 2003.

[7] D. Zhou, J. Guo, Y. Zhang, J. Chai, H. Liu, Y. Liu, C. Huang, X. Gui, and Y. Liu, "Distributed data analytics platform for wide-area synchrophasor measurement systems," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–9, 2016.

[8] W. Wang, L. He, P. Markham, H. Qi, Y. Liu, Q. C. Cao, and L. M. Tolbert, "Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1654–1664, 2014.

[9] J. Guo, Y. Ye, Y. Zhang, Y. Lei, and Y. Liu, "Events associated power system oscillations observation based on distribution-level phasor measurements," in *2014 IEEE PES Transmission and Distribution Conference and Exposition*, April 2014, pp. 1–5.

[10] M. J. Smith and K. Wedeward, "Event detection and location in electric power systems using constrained optimization," in *IEEE Power and Energy Society General Meeting*, 2009, pp. 1–6.

[11] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, 2012.

[12] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *IEEE Power and Energy Society General Meeting*, 2009, pp. 1–5.

[13] S. You, J. Guo, G. Kou, Y. Liu, and Y. Liu, "Oscillation mode identification based on wide-area ambient measurements using multivariate empirical mode decomposition," *Electric Power Systems Research*, vol. 134, no. 2016, pp. 158–166, 2016.

[14] S. You, J. Guo, W. Yao, S. Wang, Y. Liu, and Y. Liu, "Ring-down oscillation mode identification using multivariate empirical mode decomposition," in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[15] T. Bi, X. Song, J. Wu, and Q. Yang, "Novel method for disturbance identification in power systems," in *IEEE Power Engineering Society General Meeting*, 2006, pp. 5–pp.

[16] R. Agrawal and D. Thukaram, "Identification of fault location in power distribution system with distributed generation using support vector machines," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, Feb 2013, pp. 1–6.

[17] T. Ji, Q. Wu, L. Jiang, and W. Tang, "Disturbance detection, location and classification in phase space," *IET generation, transmission & distribution*, vol. 5, no. 2, pp. 257–265, 2011.

[18] N. Zhang and M. Kezunovic, "Improving real-time fault analysis and validating relay operations to prevent or mitigate cascading blackouts," in *IEEE Transmission and Distribution Conference and Exhibition*, 2006, pp. 847–852.

[19] Z. Zhang, Y. Song, H. Cui, J. Wu, F. Schwartz, and H. Qi, "Early mastitis diagnosis through topological analysis of biosignals from low-voltage alternate current electrokinetics," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 542–545.

[20] Z. Zhang, Y. Song, W. Wang, and H. Qi, "Derivative delay embedding: Online modeling of streaming time series," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 969–978.

[21] A. Rahimpour and H. Qi, "Non-intrusive load monitoring of hvac components using signal unmixing," in *The IEEE Global Conference on Signal and Information Processing (Global SIP 2015), Orlando, FL, USA*. IEEE, 2015.

[22] D. Zhou, Y. Liu, and J. Dong, "Frequency-based real-time line trip detection and alarm trigger development," in *2014 IEEE PES General Meeting Conference Exposition*, July 2014, pp. 1–5.

[23] W. Wang, L. He, P. Markham, H. Qi, and Y. Liu, "Detection, recognition, and localization of multiple attacks through event unmixing," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2013, pp. 73–78.

[24] W. Pentland, Can grideye save americas power grid? Forbes. [Online]. Available: http://blogs.forbes.com/williampentland/2011/02/02/can-grideye-americas-power-grid/

[25] R. M. Gardner, W. Li, J. West, J. Dong, Y. Liu, and G. Zhang, "Power system frequency oscillation characteristics," in *IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–7.

[26] Y. Zhang, "Frequency monitoring network (FNET) data center development and data analysis," Ph.D. dissertation, University of Tennessee, Knoxville, 2014.

[27] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[29] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *in Neural Information Processing Systems*, pp. 801–808, 2006.

[30] Y. Song, W. Wang, Z. Zhang, H. Qi, and Y. Liu, "Multiple event analysis for large-scale power systems through cluster-based sparse coding," in *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on*. IEEE, 2015, pp. 301–306.

[31] Y. Song, Z. Zhang, L. Liu, A. Rahimpour, and H. Qi, "Dictionary reduction: Automatic compact dictionary learning for classification," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 1–16.

[32] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[33] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[34] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[35] The non-negative feature sign sparse coding algorithm. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/36478-parelab/content/PaReLab/DbMultiFea/DbFeaStati/DbFeaDictionary/GetDict/L1QP_FeatureSign_yang.m

[36] Intrduction of large scale testbed at CURENT. [Online]. Available: http://curent.utk.edu/files/3713/6811/8508/Fact_sheet_tomsovic.pdf

[37] W. Ju, J. Qi, and K. Sun, "Simulation and analysis of cascading failures on an npcc power system test bed," in *IEEE Power & Energy Society General Meeting*, 2015, pp. 1–5.

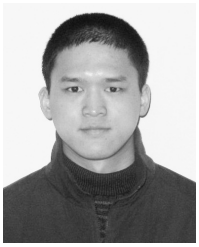
[38] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. McGraw-hill New York, 1994, vol. 7.

[39] G. Rogers, *Power system oscillations*. Springer Science & Business Media, 2012.

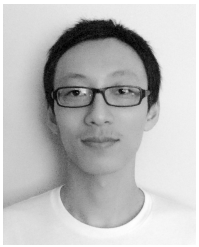
[40] Y. Liu, L. Zhan, Y. Zhang, P. N. Markham, D. Zhou, J. Guo, Y. Lei, G. Kou, W. Yao, J. Chai *et al.*, "Wide-area-measurement system development at the distribution level: An FNET/GridEye example," *IEEE Transactions on Power Delivery*, vol. 31, no. 2, pp. 721–731, 2016.

[41] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

Yang Song (S'15) is a Ph.D. candidate in the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. She received the B.S. and M.S. degrees in electrical engineering from the Northeastern University in 2010 and Zhejiang University in 2013, respectively. Her research interests include signal and image processing, pattern recognition and computer vision.



Wei Wang (S'12) received his Ph.D. in 2015 from the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. He also received his B.S. and M.S. degrees from University of Science and Technology Beijing in 2006 and Beihang University in 2009, respectively. His research interests include signal/image analysis, pattern recognition, machine learning and smart grid secure monitoring.



Zhifei Zhang (S'15) is a Ph.D. candidate in the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. He received the B.S. and M.S. degrees in electrical engineering from the Northeastern University in 2010 and Zhejiang University in 2013, respectively. His research interests are in signal and image processing, pattern recognition, machine learning and computer vision.

Hairong Qi (S'97-M'00-SM'05) received the B.S. and M.S. degrees in computer science from Northern JiaoTong University, Beijing, China in 1992 and 1995 respectively, and the Ph.D. degree in computer engineering from North Carolina State University, Raleigh, in 1999. She is currently the Gonzalez Family Professor with the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. Her research interests are in advanced imaging and collaborative processing, hyperspectral image analysis, and bioinformatics. Dr. Qi is the recipient of the NSF CAREER Award. She also received the Best Paper Awards at the 18th International Conference on Pattern Recognition (ICPR) in 2006, the 3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC) in 2009, and IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensor (WHISPERS) in 2015. She is awarded the Highest Impact Paper from the IEEE Geoscience and Remote Sensing Society in 2012.

Yilu Liu (S'88–M'89–SM'99–F'04) received the B.S. degree from XiAn Jiaotong University, China, and the M.S. and Ph.D. degrees from the Ohio State University, Columbus, OH, USA, in 1986 and 1989, respectively. She is currently a Governor Chair Professor at the University of Tennessee, Knoxville and Oak Ridge National Laboratory. Prior to joining UTK/ORNL, she was a professor at Virginia Tech. She led the effort to create the North American power grid monitoring network (FNET) at Virginia Tech which is now operated at UTK and ORNL as FNET/GridEye. Her current research interests include power system wide-area monitoring and control, large interconnection level dynamic simulations, electromagnetic transient analysis, and power transformer modeling and diagnosis.