Joint Identification and Tracking of Multiple CBRNE Clouds Based on Sparsity Pursuit [*]

Huimin Chen
Department of Electrical Engineering
University of New Orleans
New Orleans, LA, 70418, U.S.A.
hchen2@uno.edu

X. Rong Li
Department of Electrical Engineering
University of New Orleans
New Orleans, LA, 70418, U.S.A.
xli@uno.edu

Abstract – *The evolution of chemical, biological, radiological, nuclear and explosive (CBRNE) clouds depends considerably on its composition. For example, cloud tracking usually relies on a diffusion model of the average atmospheric concentration of the CBRNE material; identification of its composition can benefit greatly from knowledge about the propagation of the compounds. As a result, substance classification and cloud tracking help each other significantly. However, few research efforts consider* joint *identification and tracking of CBRNE clouds using a network of possibly heterogeneous sensors. This paper deals with such joint identification and tracking. We assume that the chemical composition has a sparse representation in the Raman spectra with a reference library and apply a sparsity pursuit algorithm to adaptively refine the cloud propagation model based on the estimated composition. We demonstrate the benefit of joint identification and tracking of the aggregated clouds when individual substance has a different diffusion coefficient. The results also provide guidelines for selecting an appropriate sensor combination to accurately predict the cloud boundary.*

Keywords: Joint identification and tracking, contaminant cloud, compound classification, sparsity pursuit, sensor management.

1 Introduction

The threat of pollution due to the release of chemical, biological, radiological, nuclear and explosive (CBRNE) materials, either accidentally or deliberately, has drawn significant attention to national defense and homeland security. Early detection and classification of CBRNE agents are vital for rapid responsive action to reduce the extent of human exposure. Accurate tracking and prediction of the contaminant clouds also help in the post-accident management to minimize disastrous impact. Many research results are available on detecting and localizing single or multiple plume sources with autonomous vehicles or sensor networks such as [15] for a vapor-emitting source, [2] for a nuclear source, and [10, 11] for a chemical source. In [9] the plume detection and localization problem is formulated as abrupt change detection using sparse sensor measurements. Bayesian inference methods have also been applied to plume detection [3, 16, 14], localization [8, 21, 22, 4] and tracking [18]. The development of a large scale testbed was reported in [5] for the study of contaminate materials. Classification of chemical or biological agents has been studied mainly using the sensing devices based on Raman spectroscopy [7]. Raman spectrum represents the intensity of the scattered photon as a function of frequency shift, which can be used to identify the chemical composition of a substance. To represent the target spectrum as a linear combination of the reference spectra, one needs to build a library of the spectra of known chemicals as well as the background spectrum. Various learning algorithms have been studied within this supervised regime [17, 13]. Other detection techniques have also been proposed to handle the cases without exact knowledge of the background and reference spectra [12, 20]. So far few research efforts consider *joint* identification and tracking of CBRNE clouds using a network of possibly heterogeneous sensors. The difficulty lies in the lack of accurate predictive model for mean transport and turbulent diffusion of the aggregated CBRNE compounds. In reality, the CBRNE composition can only contain a few known substances in the Raman spectra with a reference library. Thus we propose a sparsity pursuit algorithm to adaptively refine the cloud propagation model based on the estimated composition. We demonstrate the benefit of joint identification and tracking of the aggregated clouds when individual substance has a different diffusion coefficient. The results also provide guidelines for selecting an appropriate sensor combination to accurately predict the cloud bound-

---

[*] Research supported in part by ARO through grant W911NF-08-1-0409, ONR-DEPSCoR through grant N00014-09-1-1169, NAVO through Contract # N62306-09-P-3S01, and Louisiana Board of Regents through NASA EPSCOR DART-2.

ary.

The rest of the paper is organized as follows. Section 2 presents the formulation of the joint classification and tracking problem using both concentration and Raman spectrum measurements. Section 3 describes the sparsity pursuit for Raman spectra estimation. Section 4 discusses the benefit of Raman spectra for source localization. Section 5 presents the benefit of source localization for the identification of source composition. Section 6 extends the joint classification and localization problem to joint classification and tracking of multiple contaminant clouds. Simulation results are provided in Section 7 and conclusions are in Section 8.

2 Problem Formulation

2.1 Atmospheric Dispersion Model

If any type of pollutant material is released into the atmosphere, the propagation of the contaminant cloud will eventually lead to a concentration distribution somewhere downwind of the source. For convenience, we assume that the contaminant cloud consists of possibly a combination of chemical substances which can be modeled by a collection of particles. We set a Cartesian coordinate system with x-axis oriented towards the mean wind direction, y-axis in the cross-wind direction and z-axis in the vertical direction. If the source of a pollutant is located at (x_0, y_0, z_0) with release rate $q(t)$, then at time t, the concentration of the pollutant at some down-stream location $(x, y, 0)$ can be written as [4]

$$C(x, y, 0, t) = \int_0^t K(t, \tau) q(\tau) d\tau \qquad (1)$$

where the kernel $K(t, \tau)$ is

$$K(t, \tau) = \frac{\exp\left[-\frac{[x-x_0-u(t-\tau)]^2}{4K_x(t-\tau)} - \frac{(y-y_0)^2}{4K_y(t-\tau)} - \frac{z_0^2}{4K_z(t-\tau)}\right]}{8\pi^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}(t-\tau)^{\frac{3}{2}}} \qquad (2)$$

with mean wind speed u and diffusion coefficients K_x, K_y, K_z along x, y and z directions, respectively.

2.2 Sensor Measurement Models

We assume that one type of sensor can measure the concentration of the contaminant source in the atmosphere such as the light distance and ranging sensor which emits pulses of ultra-violet light and measure the returning fluoresced light. Under this setting, there are J sensors deployed at fixed locations where sensor j is located at $(x_j, y_j, 0)$ and collects N concentration readings $\mathbf{c}_j = \{C(x_j, y_j, 0, t_n)\}_{n=1}^N$. Denote by $\mathbf{c} = \{\mathbf{c}_j\}_{j=1}^J$ the collection of all sensor readings. Denote by $\mathbf{q} = \{q(\tau_i)\}_{i=1}^M$ the discretized source release sequence where $q(\tau_i)$ is the release rate at time τ_i. Ideally, we have the following observation equation

$$\mathbf{c} = A(\mathbf{p})\mathbf{q} \qquad (3)$$

where $\mathbf{p} = (x_0, y_0, z_0)$ denotes the unknown source location. Note that for measurement $c_j(x_j, y_j, 0, t_n)$, the corresponding element $a_{(jn,k)}$ in $A(\mathbf{p})$ is given by

$$a_{(jn,k)} = K(t_n, \tau_k)\beta_{nk} \qquad (4)$$

where β_{nk} is a quadrature weight [4]. The estimation of source location $\mathbf{p}$ and release rate $\mathbf{q}$ using a collection of concentration readings can be formulated as the least squares problem given by

$$(\mathbf{p}^*, \mathbf{q}^*) = \arg\min_{\mathbf{p},\mathbf{q}} ||\mathbf{c} - A(\mathbf{p})\mathbf{q}||_2^2 \qquad (5)$$

Note that this formulation is valid only for a single source.

In order to identify the composition of the pollutant substance, we assume that another type of sensor can provide the Raman spectra of both background and polluted area. When one has L chemical substances with known Raman spectra $\{\mathbf{s}_i\}_{i=1}^L$, we can express the true spectra of the pollutant source as

$$\mathbf{s} = \sum_{i=0}^{L} \alpha_i \mathbf{s}_i \qquad (6)$$

where $\mathbf{s}_0$ is the background spectra. The coefficient vector $\mathbf{a} = [\alpha_0\ \alpha_1\ ...\ \alpha_L]^T$ has to be estimated from the measured spectrum at different bins with an approximate linear model given by [13]

$$\mathbf{z} = G\mathbf{a} + \mathbf{v} \qquad (7)$$

where $G = BPS$ and B is the calibration matrix taking flat-field response into account; P is the point spread function of the diffraction grating used to spread the spectral energy across measurement bins; $S = [\mathbf{s}_0\, \mathbf{s}_1\, ...\, \mathbf{s}_L]$ is the matrix of spectrum library. The measurement noise $\mathbf{v}$ absorbs the large signal approximation error, which in general is non-Gaussian. A popularly used method to estimate $\mathbf{a}$ is the so called nonnegative weighted least squares given by [17]

$$\min_{\mathbf{a}} (\mathbf{z} - G\mathbf{a})^T W (\mathbf{z} - G\mathbf{a}) \text{ s.t. } \mathbf{a} \geq 0 \qquad (8)$$

where W is the weight matrix. Note that the above formulation assumes that the number of measurement bins is at least $L + 1$.

2.3 Joint Source Identification and Localization

We assume that the concentration readings are the results of aggregation from multiple source releases. When there are l sources with unknown locations $\{\mathbf{p}(i)\}_{i=1}^l$ and release rate sequences $\{\mathbf{q}(i)\}_{i=1}^l$, we have the following ideal observation equation

$$\mathbf{c} = \sum_{i=1}^{l} A(\mathbf{p}(i))\mathbf{q}(i) \qquad (9)$$

The source localization problem becomes identifying the number of sources, the corresponding origins and the release sequences jointly based on the concentration readings from multiple sensors. Unfortunately, localization of multiple sources from concentration readings is often ill-posed when the released source contains multiple chemical substances since the sensors collecting the concentration readings can not identify the composition of the source material.

With the help of Raman spectrum measurements, we can distinguish L chemical substances from the library. Ideally, for two different chemical sources i and j, we have

$$\frac{||\mathbf{q}(i)||_1}{||\mathbf{q}(j)||_1} = \frac{\alpha_i}{\alpha_j} \tag{10}$$

since the composition of the substance depends on the cumulative effect of the source release when the Raman spectroscopy is made close to the source origin. Given both concentration measurement $\mathbf{c}$ and Raman spectrum measurement $\mathbf{z}$, one may want to identify the composition of the pollutant sources first and then estimate the origin of each chemical material as well as the corresponding release rate sequence. Note that multiple sources can come from the same origin with the same or different release rate.

2.4 Joint Source Identification and Tracking

In practice, we have to collect concentration readings and Raman spectrum measurements sequentially and update the release rate estimate of the identified sources based on the measurements accumulated so far. Denote by $Q(i)_k$ the cumulative release from source i between time t_{k-1} and t_k. Let $Q_k^L = [Q(1)_k \ldots Q(L)_k]^T$ be the augmented release vector of all possible materials from the reference library. Let $\mathbf{p}_k^L = [\mathbf{p}(1)_k \ldots \mathbf{p}(L)_k]^T$ be the corresponding origins of the released sources. Note that $\mathbf{p}(i)_k$ will be undefined when $Q(i)_k$ is below a predetermined threshold, meaning that there is no release of source i between t_{k-1} and t_k. Denote by Z_k the concentration readings and Raman spectrum measurements collected between t_{k-1} and t_k. Let $Z^k = \{Z_1, \ldots, Z_k\}$ be the cumulative measurements collected up to t_k. The joint source identification and tracking problem is to develop a sequential algorithm for updating the distribution from $p(Q_{k-1}^L, \mathbf{p}_{k-1}^L | Z^{k-1})$ to $p(Q_k^L, \mathbf{p}_k^L | Z^k)$ with the newly obtained measurement Z_k. Note that Q^L is in general very sparse, meaning that the source only contains a combination of a few chemicals from the possibly very large reference library.

3 Raman Spectra Estimation via Sparsity Pursuit

When one has a large reference Raman spectrum library of pollutant chemicals while the number of bins in the measured Raman spectra is limited, estimating the mixing coefficient vector $\mathbf{a}$ becomes an ill-posed problem. Fortunately, most entries of $\mathbf{a}$ will be zero in practical situations leading to a sparsity constraint. In the noiseless case, we can pursue for the sparsest solution given by

$$\min_{\mathbf{a}} ||\mathbf{a}||_0 \text{ s.t. } G\mathbf{a} = \mathbf{z} \text{ and } \mathbf{a} \geq 0 \tag{11}$$

where $|| \cdot ||_0$ measures the number of nonzero entries of the vector. Let $\mathbf{g}_i$ be the i-th column of G. We define the coherence of G by

$$\rho(G) = \max_{i,j,j \neq i} \frac{|\mathbf{g}_i^T \mathbf{g}_j|}{||\mathbf{g}_i||_2^2} \tag{12}$$

If a solution $\mathbf{a}^*$ satisfies $G\mathbf{a}^* = \mathbf{z}$ and is sparse enough compared with the incoherence of G, i.e.,

$$||\mathbf{a}^*||_0 \leq \frac{1 + \rho(G)}{2\rho(G)} \tag{13}$$

then $\mathbf{a}^*$ is unique [6]. This indicates that the number of measurements required to recover $\mathbf{a}^*$ depends on its sparsity and the incoherence of the measurement matrix G. When $\mathbf{a}^*$ is highly sparse while G is incoherent enough, we can uniquely determine $\mathbf{a}^*$ from much less than L measurements. To reduce the coherence of G, we can multiply G by a nonsingular square matrix T_G so that $\rho(T_G G) < \rho(G)$. Since for any vector $\mathbf{d}$ in the null space of G, we have

$$||\mathbf{d}||_\infty \leq \frac{\rho(G)}{1 + \rho(G)} ||\mathbf{d}||_1 \tag{14}$$

it is intuitively appealing to normalize each column of the positive matrix G in order to improve its incoherence. This can be achieved by subtracting the mean of each column, i.e.,

$$T_G = I - \frac{1 - \delta}{L} \mathbf{1}\mathbf{1}^T \tag{15}$$

where δ is a small number to ensure the non-singularity of the transform.

In the noisy setting, we can solve the following constrained optimization problem

$$\min_{\mathbf{a}} ||\mathbf{a}||_1 \text{ s.t. } ||\mathbf{z} - G\mathbf{a}||_W^2 \leq \epsilon \text{ and } \mathbf{a} \geq 0 \tag{16}$$

where W is the weight matrix and ϵ depends on the desired noise level, which makes the tradeoff between the sparsity of the solution and the accuracy fitting to the measurements. By relaxing the sparsity pursuit to ℓ_1-norm, we can solve the above problem using standard convex programming package [1]. Alternatively, one can also apply any greedy type algorithm such as orthogonal matching pursuit [19] to find the sparse entries of $\mathbf{a}$ one at a time. It is more efficient than using the convex programming but does not guarantee to recover the correct nonzero entries of $\mathbf{a}$ under a low signal-to-noise ratio.

4 Raman Spectrum Aided Source Localization

Assume that l different chemical substances have been identified with the corresponding estimated intensities given by $\{\alpha_1, ..., \alpha_l\}$. The goal is to estimate the origin $\mathbf{p}(i)$ and release sequence $\mathbf{q}(i)$ of the i-th source for $i = 1, ..., l$ using concentration measurements $\mathbf{c}$. We propose to solve the regularized least squares problem given by

$$\min_{\{\mathbf{p}(i),\mathbf{q}(i)\}} \left\| \mathbf{c} - \sum_{i=1}^{l} A(\mathbf{p}(i))\mathbf{q}(i) \right\|_2^2 + \lambda \left(\sum_{i=1}^{l} | \, \|\mathbf{q}(i)\|_1 - Q_0\alpha_i \, | \right) + \mu \left(\sum_{i=1}^{l} \|\mathbf{q}(i)\|_1 \right)$$

where the first regularization term promotes the total release of each source to be proportional to its estimated coefficient from the Raman spectrum with the scaling factor Q_0 while the second term promotes the release sequence to be sparse, i.e., with few nonzero entries in favor of the instantaneous release type. The parameters λ and μ can be adjusted from small positive values to large ones to envision the full regularization path that leads to sparse solutions of $\{\mathbf{q}(i)\}_{i=1}^{l}$.

In practice, we solve the above nonlinear optimization problem iteratively. We first make an initial guess of the source locations $\{\mathbf{p}(i)^0\}$, which can be the sensor locations where the l largest concentration readings were reported. Then we find the release rate for each source by solving

$$\min_{\mathbf{q}(i)} \left|\left|\mathbf{c} - A(\mathbf{p}(i)^0)\mathbf{q}(i)\right|\right|_2^2 + \lambda \, ||| \mathbf{q}(i)||_1 - Q_0\alpha_i| + \mu||\mathbf{q}(i)||_1 \qquad (17)$$

and compute the corresponding residuals

$$r_i = ||\mathbf{c} - A(\mathbf{p}(i)^0)\mathbf{q}(i)^0||_2^2 \qquad (18)$$

We select source j if

$$j = \arg\min_i r_i \qquad (19)$$

Next, we update the concentration measurement $\mathbf{c}$ by subtracting $A(\mathbf{p}(i)^0)\mathbf{q}(i)$ from it and then apply the same procedure to the remaining $l-1$ sources. Note that we have applied the greedy selection of the release sequence that leads to the smallest residual in each step. The above algorithm can be easily modified to become an orthogonal matching pursuit [19] procedure with slight increase of the computational requirement. Once we obtained the release estimate $\{\mathbf{q}(i)^0\}$, then we solve the source localization problem by minimizing the least squares directly, i.e.,

$$\min_{\mathbf{p}(i)} \left|\left|\mathbf{c} - A(\mathbf{p}(i))\mathbf{q}(i)^0\right|\right|_2^2 \qquad (20)$$

The iterations between the refined estimates of the source locations and the release sequences continue until the improvement in terms of the overall regularized least squares function is below a predetermined threshold.

5 Localization Aided Source Identification

Once we have obtained fairly accurate estimate of the source origins, we may want to deploy the Raman spectrum sensors close to the estimates to achieve better distinguishable spectra from the background. Alternatively, when two chemical materials have different diffusion coefficients, they will have distinguishable concentration variations in the spatial and temporal domains even when they have identical instantaneous releases at the same location. Denote by $\mathbf{k}(i)$ the diffusion coefficient of source i. Assume that we have obtained concentration measurement $\mathbf{c}$ and applied the greedy method to estimate the location and release sequence of the source one at a time by solving

$$\min_{\mathbf{p},\mathbf{k},\mathbf{q}} ||\mathbf{c} - A(\mathbf{p}, \mathbf{k})\mathbf{q}||_2^2 + \mu||\mathbf{q}||_1 \qquad (21)$$

and subtracted the contribution of the estimated source from $\mathbf{c}$ in order to extract another possible source. When l sources have been identified with the estimated diffusion coefficients given by $\{\mathbf{k}(i)\}_{i=1}^{l}$, we can compare them with the reference library and find those chemical materials with the corresponding diffusion coefficients close to the initial estimates. By doing this, the size of the reference Raman spectra (number of columns of G) can be significantly reduced.

6 Joint Source Identification and Tracking via Sparsity Pursuit

We assume that the pollutant source can come from a subset of L known chemical substances with known diffusion coefficient $\{\mathbf{k}(i)\}_{i=1}^{L}$ and Raman spectra G. Between t_{k-1} and t_k, we have the augmented measurement equation

$$Z_k = f(\mathbf{p}_k^l, \mathbf{k}_k^l, \mathbf{q}_k^l) + V_k \qquad (22)$$

where $f(\cdot)$ is a nonlinear function describing the concentration reading at any known sensor location and the Raman spectra from the aggregation of l active sources with locations $\mathbf{p}_k^l$, diffusion coefficients $\mathbf{k}_k^l$ and release rate $\mathbf{q}_k^l$. The noise vector V_k absorbs the measurement and unmodeled errors in the data collection process. We further assume that the existence of any source follows a discrete Markov chain independent of other sources. This assumption can be refined when one has prior knowledge on certain possible combinations of the chemical substances likely to be seen in the contaminant clouds. The transition probability between the birth and death state of each source is given by

$$P(b_k|b_{k-1}) = \begin{bmatrix} 1-p_b & p_d \\ p_b & 1-p_d \end{bmatrix} \tag{23}$$

When the i-th source is inactive from t_{k-1} to t_k, i.e., $b_k(i) = 0$, we mean that $\mathbf{q}_k(i) = 0$ while $\mathbf{p}_k(i)$ and $\mathbf{k}_k(i)$ are undefined. In order to sequentially update the probability of existence for each source, we need to know how to compute the likelihood function $\Lambda(Z_k|\mathbf{q}_k^L)$. This easily becomes exponentially explosive over time when we have to enumerate all combinations of the active and inactive source sequences. Alternatively, since active sources are sparse among all time instances, we make the following approximations to the likelihood function of source i

$$\Lambda(Z_k|\mathbf{q}_k(i)=0) \approx \max_j \Lambda(Z_k|\mathbf{q}_k(j)>0, \mathbf{q}_k(i)=0, i\neq j) \tag{24}$$

$$\Lambda(Z_k|\mathbf{q}_k(i)>0) \approx \Lambda(Z_k|\mathbf{q}_k(i)>0, \mathbf{q}_k(j)=0, j\neq i) \tag{25}$$

where maximum sparsity is used to simplify the computation for the existence of each source. If source i is the only active source during T consecutive updates from k to $k+T-1$, then we can find the maximum likelihood estimate of source location $\mathbf{p}(i)$ and diffusion coefficient $\mathbf{k}(i)$ directly by

$$\max_{\mathbf{p}(i),\mathbf{k}(i)} \Lambda(Z_k, ..., Z_{k+T-1}|\mathbf{p}(i), \mathbf{k}(i)) \tag{26}$$

assuming that the source location is fixed. In general, batch processing using least squares is needed to simultaneously estimate the locations and diffusion coefficients of multiple active sources. We can apply the greedy method described in Section 4 without the need of reestimating the release rate of each active source. Note that both concentration and Raman spectrum measurements enter the nonlinear least squares since we need Raman spectrum measurements to estimate the diffusion coefficients. The whole computational procedure to predict the propagation of the contaminant clouds is shown in Fig. 1. Note that (23) is used to predict the state of each source in future time so that the contaminated area is assessed via multi-step look ahead.

7 Numerical Simulation and Results

We present the simulation study of source classification, localization and release rate estimation using multiple concentration and Raman spectrum sensors. We are interested in both classification and parameter estimation accuracy.

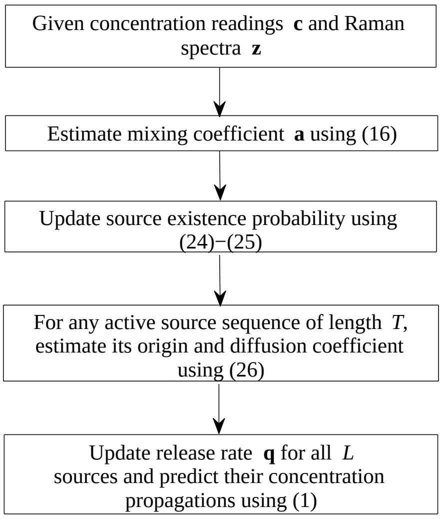

Figure 1: Computational procedure for joint source identification and tracking using concentration and Raman spectrum measurements

7.1 Raman Spectra Estimation

We assume that the background spectrum spreads across all bins of a Raman spectrum detector following a Poisson distribution with expected number of counts being 10. There are 1000 reference Raman spectra with randomly generated patterns across 256 bins. It is used as the reference library for identifying the active sources. In the simulation, we randomly select three sources and set the coefficient of each source being 1 to generate synthetic Raman spectrum measurements. We compute the probability of correctly identifying all active sources and the root mean square error (RMSE) of the estimated coefficients among those cases where all sources have been correctly identified. We compare convex program (CP) using (16) [1], orthogonal matching pursuit (OMP) [19] and nonnegative weighted least squares (NWLS) [13] on the top 20 matched source candidates to the measurement spectra. Note that NWLS is not directly applicable to the case that the number of reference sources is greater than the number of bins in the spectrum measurement. We modified the NWLS algorithm by selecting only 20 candidates ranked with respect to the residual of each source and then applying NWLS to the selected candidates. 5000 Monte Carlo trials were used to evaluate each algorithm and the results were summarized in Table 1. We can see that CP recovers the sparse source signal in all testing cases while OMP performs slightly worse due to the unmodeled background noise. In contrast, NWLS fails in most cases owing to not taking the sparsity of the signal into account.

Table 1: Comparison of Source Identification and Parameter Estimation Accuracy from Raman Spectra

Algorithm	CP	OMP	NWLS
Correct Prob.	1.0	0.98	0.32
RMSE	0.22	0.27	0.19

Table 2: Comparison of Localization Accuracy Among Three Scenarios

scenario	RMSE of source 1	RMSE of source 2
(i)	8.5	6.4
(ii)	12.3	9.5
(iii)	3.2	1.7

7.2 Joint Source Identification and Localization

Consider one source located at $(-40, 35, 12)$ with instantaneous release of $q(1) = 10^5$. Another source located at $(-30, 15, 0)$ has continuous release with the release rate

$$q(t) = [e^{-0.1(t-10)}u(t-10) + 2e^{-0.5(t-50)}u(t-50)] \cdot 10^5$$

The first source has the diffusion coefficients $K_x(1) = K_y(1) = 12$, $K_z(1) = 0.2113$. The second source has the diffusion coefficients $K_x(2) = K_y(2) = 5$, $K_z(2) = 0.1032$. We assume that the wind speed $u = 1.8$ along x-axis. Five sensors, located at $(0, 0)$, $(15, 15)$, $(30, 30)$, $(45, 45)$, $(60, 60)$, respectively, collect concentration readings synchronously with 100 samples per sensor. All sensors are on the ground with zero elevation. We added white Gaussian noise to the sensor readings with standard deviation $4 \cdot 10^{-3}$. Each sensor will have a detection of the possibly aggregated sources when the concentration reading exceeds 0.01. In addition, Raman spectrum sensor can be used at any designated location to measure the source spectra with the same Poisson distributed background having 10 counts in average. The coefficient of each source appeared in the Raman spectra depends on the location of the sensor with the scaling factor being 0.025 to the cumulative concentration of the source during the Raman spectroscopy.

We apply the proposed iterative algorithm to solve regularized least squares for source localization and release rate estimation. We consider three scenarios: (i) finding source locations using concentration measurements only (with known diffusion coefficients of each source), (ii) finding source release sequences using Raman spectrum sensor only (located at the origin), (iii) finding source locations and release sequences using both concentration and Raman spectrum measurements. Note that in (iii) the diffusion coefficients of each source are obtained from the reference library when the existence of the source has been identified from the Raman spectra. For one realization, we have the true release sequences of the two sources and the estimated ones shown in Fig. 2 based on the scenarios (i)–(iii). We can see that the estimated release sequences are sparse in nature: they contain more zeros than the true release sequence of the second source. Clearly, scenario (ii) yields the largest estimation error using Raman spectrum measurements only. Scenario (i) yields smaller error than (ii) while scenario (iii) yields the most accurate estimate of the release sequence, i.e., all nonzero entries of the release signals with large enough magnitude have been correctly identified in the estimates.

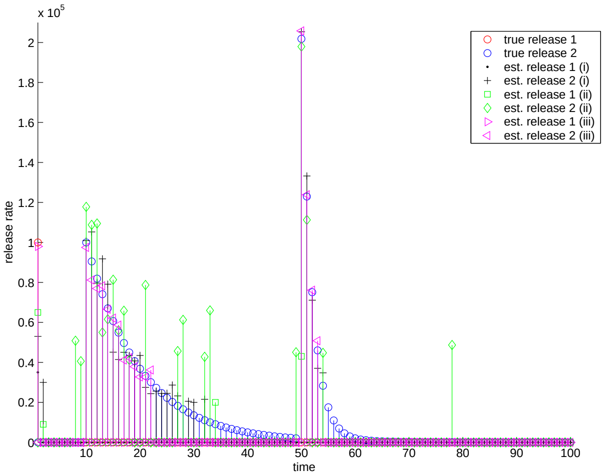

Figure 2: Comparison of release rate estimation of two sources

We conducted 100 Monte Carlo trials and computed the localization RMSE of each source. The results are listed in Table 2. It can be seen that the localization accuracy is significantly improved by exploiting both concentration and Raman spectrum measurements than either of the cases using only one type of the measurements.

7.3 Joint Source Identification and Tracking

We consider that two locations, namely, $(-40, 35, 12)$ and $(-30, 15, 0)$, are the origins of potential chemical release. There are 100 reference chemical substances with different Raman spectra as well as diffusion coefficients. Two of the chemical substances are used to generate the contaminate cloud with instantaneous release $q = 10^5$ at each time if the source is active. For source 1, the diffusion coefficients are

$K_x(1) = K_y(1) = 12$, $K_z(1) = 0.2113$. The transition between active and inactive state follows a Markov chain with birth probability $p_b(1) = 0.02$ and death probability $p_d(1) = 0.99$. Source 2 has the diffusion coefficients $K_x(2) = K_y(2) = 5$, $K_z(2) = 0.1032$ and birth/death probabilities $p_b(2) = 0.05$, $p_d(2) = 0.95$. Initially, source 1 is active at $(-40, 35, 12)$ while source 2 is active at $(-30, 15, 0)$. We consider two configurations of sensor deployment for joint source identification and tracking. Configuration (i) fixes the concentration sensors at $(0, 0)$, $(15, 15)$, $(30, 30)$, $(45, 45)$, $(60, 60)$ and the Raman spectrum sensor at $(0, 0)$ throughout 100 updates of the source state. Configuration (ii) puts concentration sensors at (x_s, y_s), (x_s+15, y_s), (x_s+30, y_s), $(x_s + 45, y_s)$, $(x_s + 60, y_s)$ and the Raman spectrum sensor at (x_s, y_s) where (x_s, y_s) is the x-y coordinates of the centroid among the estimated source locations when sensor measurements allow one to localize multiple sources. The tracking algorithm keeps updating the existence probability of each source using both concentration and Raman spectrum measurements. When the existence probability of any source is above 0.5, the source is declared to be active. For a source being active in three time instances, we apply a batch of the concentration and Raman spectrum measurements throughout this interval to estimate the location and release rate using regularized least squares. The previously estimated source location, if available, is used to initialize the iterative algorithm. Note that the number of active sources is known in this batch processing step so that the regularization only promotes the sparsity of the source mixing. With each active source being identified, we find its diffusion coefficients from the reference library. When the active source's origin is localized with the estimated release sequence, we can apply the dispersion model to propagate the impact area of this contaminate cloud over time. To assess the prediction accuracy of the cloud tracking algorithm, we compare the predicted impact area at $t = 120$ with the true impact area calculated from the simulated source releases. We define two types of errors in the prediction of the impact area: False positive e_1 is the predicted area outside and true impact area over the whole true impact area; false negative e_2 is the true impact area not being covered by the prediction over the whole true impact area.

We conducted 100 Monte Carlo trials for the tracking algorithm to perform 100 updates and then assess the prediction accuracy of the impact area at $t = 120$ based on two sensor configurations. The results are listed in Table 3. We can see that both false positive and false negative in the impact area prediction are significantly reduced when we can deploy the sensors close to the estimated source origins and along the downwind directions with adequate spacing. Thus it is worth investigating the adaptive sensor placement problem to best characterize the contaminate clouds. Here we only addressed the issue of source identification, localization and release sequence estimation. When the wind field is not homogeneous, more sophisticated dispersion model is required to make long time prediction of the impact area.

Table 3: Comparison of Prediction Accuracy of the Impact Area

configuration	e_1	e_2
(i)	0.12	0.07
(ii)	0.06	0.03

8 Conclusions and Future Work

We considered *joint* identification and tracking of CBRNE clouds using a network of possibly heterogeneous sensors that can provide both concentration readings and Raman spectra. We assume that the chemical composition has a sparse representation in the Raman spectra with respect to a reference library possibly of large size. We applied a sparsity pursuit algorithm to adaptively refine the cloud propagation model based on the estimated composition. We conducted simulation study using several different source release scenarios and demonstrated the benefit of joint identification and tracking of the aggregated clouds when individual substance has a different diffusion coefficient. We also found that adaptive sensor deployment can provide better tracking accuracy in terms of predicting the impact area of the contaminate clouds than that using the same sensor suites but at fixed locations. One possible future work is to conduct realistic experimental study and compare with available Raman spectrum based classification and tracking benchmarks. Another avenue is to study the best configuration of heterogeneous sensors to consistently monitor a large surveillance area and quantify the benefit of dynamic sensor deployment compared with the static case where sensor locations are fixed.

References

[1] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[2] S. M. Brennan, A. M. Mielke, and D. C. Torney, "Radiation Detection with Distributed Sensor Networks," *IEEE Computer*, pp. 57–59, August 2004.

[3] S. M. Brennan, A. M. Mielke, and D. C. Torney, "Radioactive Source Detection by Sensor Networks," *IEEE Nuclear Science*, 52(3), pp. 813–819, 2005.

[4] H. Chen, X. R. Li, “Source Parameter Estimation of Atmospheric Pollution Using Regularized Least Squares,” Proc. of *11th International Conference on Information Fusion*, Cologne, Germany, July 2008.

[5] J.-C. Chin, L.-H. Hou, J.-C. Hou, C. Ma, N.S. Rao, M. Saxena, M. Shankar, Y. Yong, and D.K.Y. Yau, “A Sensor-Cyber Network Testbed for Plume Detection, Identification, and Tracking,” *6th International Symposium on Information Processing in Sensor Networks*, pp. 541–542, 2007.

[6] D. L. Donoho, “Compressed Sensing,” *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] J. R. Ferraro, and K. Nakamoto, *Introductory Raman Spectroscopy*, Academic Press, 1994.

[8] J. A. Farrell, S. Pang, and W. Li, “Plume Mapping via Hidden Markov Methods,” *IEEE Trans. SMC-B*, 33(6), pp. 850–863, 2003.

[9] E. B. Fox, J. W. Fisher, and A. S. Willsky, “Detection and Localization of Material Releases with Sparse Sensor Configurations,” *IEEE Trans. on Signal Processing*, 55(5), pp. 1886–1898, May 2007.

[10] H. Ishida, T. Nakamoto, T. Moriizumi, T. Kikas, and J. Janata, “Plume-Tracking Robots: A New Application of Chemical Sensors,” *Biological Bulletin*, 200, pp. 222-226, 2001.

[11] H. Ishida, G. Nakayama, T. Nakamoto, and T. Moriizumi, “ Controlling A Gas/Odor Plume-Tracking Robot based on Transient Responses of Gas Sensors,” *IEEE Sensors Journal*, 5(3), pp. 537–545, 2005.

[12] S. Kay, C. Xu, and D. Emge, “Chemical Detection and Classification in Raman Spectra,” Proc. of *SPIE Conf. Signal and Data Processing of Small Targets*, 6969, Orlando, FL, USA, Apr. 2008.

[13] M. Mallick, B. Drake, H. Park, A. Register, D. Blair, P. West, and D. Emge, “Comparison of Raman Spectra Estimation Algorithms,” Proc. of *12th International Conference on Information Fusion*, Seattle, WA, Jul. 2009.

[14] M. Ortner, and A. Nehorai, “A Sequential Detector for Biochemical Release in Realistic Environments,” *IEEE Transactions on Signal Processing*, 55(8), pp. 4173–4182, 2007.

[15] A. Nehorai, B. Porat, and E. Paldi, “Detection and Localization of Vapor-Emitting Sources,” *IEEE Transactions on Signal Processing*, 43(1), pp. 243–253, 1995.

[16] G. Nofsinger, and G. Cybenko, “Distributed Chemical Plume Process Detection,” *IEEE MILCOM*, Atlantic City, NJ, USA, pp. 1076–1082, 2005.

[17] R. D. Palkki, and A. D. Lanterman, “Algorithms and Performance Bounds for Chemical Identification under a Poisson Model for Raman Spectroscopy,” Proc. of *12th International Conference on Information Fusion*, Seattle, WA, Jul. 2009.

[18] F. Septier, A. Carmi, and S. Godsill, “Tracking of Contaminant Clouds,” *12th International Conference on Information Fusion*, Seattle, WA, Jul. 2009.

[19] J. A. Tropp, “Greedy is Good: Algorithmic Results for Sparse Approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[20] W. Wang, T. Adali, and D. Emge, “Subspace Partitioning for Target Detection and Identification,” *IEEE Trans. on Signal Processing*, 57(4), pp. 1250–1259, 2009.

[21] T. Zhao and A. Nehorai, “Detecting and Estimating Biochemical Dispersion of A Moving Source in A Semi-Infinite Medium,” *IEEE Transactions on Signal Processing*, 54(6), pp. 2213–2225, 2006.

[22] T. Zhao and A. Nehorai, “Distributed Sequential Bayesian Estimation of a Diffusive Source in Wireless Sensor Networks,” *IEEE Transactions on Signal Processing*, 55(4), pp. 1511–1524, 2007.