ENHANCED DETECTION OF CHEMICAL PLUMES IN HYPERSPECTRAL IMAGES AND MOVIES THROUGH IMPROVED BACKGROUND MODELING

Yi (Grace) Wang[1], *Mauro Maggioni*[1,2,3], *and Guangliang Chen*[4]

Departments of [1]Mathematics, [2]ECE, and [3]Computer Science, Duke University
[4]Department of Mathematics and Statistics, San Jose State University

ABSTRACT

We extend recent work that models the background in hyperspectral images by a single distribution (Gaussian or subspace) to use a mixture of such distributions. This seems to better capture the complexity of the background, which often consists of heterogeneous regions (e.g., sky, mountain and ground). We derive mixture versions of the previous estimators and apply them to benchmark data sets for detecting chemical plumes of known chemicals in hyperspectral images and movies. Our experiments show that the mixture background models consistently outperform their counterparts with a single distribution.

1. INTRODUCTION

Hyperspectral imaging technology has been increasingly used in many civilian and military applications to remotely sense chemical clouds that are often toxic and come from various sources (e.g., due to natural disasters or terrorism attack). It collects radiance data from the physical scene through specially-designed hyperspectral imaging sensors and stores them in the form of $m \times n \times p$ arrays, where m, n correspond to the spatial dimensions and p the spectral dimension. Under physically reasonable assumptions and simplifications, the following linear mixing model is commonly adopted for radiance data at all pixels $\mathbf{x} \in \mathbb{R}^p$ of the scene:

$$\mathbf{x} = \sum_{1 \le i \le N_G} g_i \mathbf{s}_i + \mathbf{v}.$$

Here, N_G is the number of chemicals, $\mathbf{s}_i$ the radiance spectrum for the ith chemical, g_i the coefficient, and $\mathbf{v}$ the radiance spectrum of the background. In practice, it is usually assumed that $N_G \le 3$. In this paper, due to the data under consideration, we focus on the case $N_G = 1$ (thus, $\mathbf{x} = g\mathbf{s} + \mathbf{v}$).

To effectively separate the chemical plume from the background clutter, one needs to choose a proper model for the background radiation $\mathbf{v}$. Current approaches represent the background by either a single Gaussian cloud

$$\mathbf{v} \sim N(\mathbf{m}_b, \Sigma_b)$$

or a single subspace

$$\mathbf{v} = \mathbf{B}\alpha + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma_b^2 \mathbf{I})$$

and then derive corresponding statistical estimators which assign detection scores to the pixels (see the review paper [2]). These two considerations have lead to, respectively, the Normalized Matching Filter (NMF) detector (also known as the Adaptive Cosine Estimator or, in short, ACE) [4]:

$$T_{\text{NMF}}(\mathbf{x}; \Sigma_b, \mathbf{s}) = \frac{(\mathbf{s}^T \Sigma_b^{-1} \mathbf{x})^2}{(\mathbf{s}^T \Sigma_b^{-1} \mathbf{s})(\mathbf{x}^T \Sigma_b^{-1} \mathbf{x})},$$

and the Normalized SubSpace (NSS) detector [3]:

$$T_{\text{NSS}}(\mathbf{x}; \mathbf{B}, \mathbf{s}) = \frac{\|(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{x}\|^2}{\|(\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{x}\|^2},$$

where $\mathbf{A} = [\mathbf{s}, \mathbf{B}]$ is the concatenated matrix.

While these detection algorithms have shown their effectiveness in many applications, there is still room for improvement. In particular, the background often consists of heterogeneous regions (such as sky, mountain, desert) which may require a separate Gaussian cloud/subspace per region to better capture the complexity of the background. Here, we address this generalization to enhance the detectability of the chemical gas region.

We consider two different scenarios: (I) only a single hyperspectral cube is available to us and (II) we have access to a time series of hyperspectral images. In the former setting, we use the available cube for both background learning and chemical detection, while in the latter one, we use the first few frames (assumed to have no gas) for background modeling and any subsequent frame for testing based on the learned background model. While the temporal dimension provides additional information, we make no use of it in the approach presented in this paper. We comment that this makes the problem much harder (as the temporal information is abandoned), but the solution more general (e.g. it applies to still images).

The authors thank Dimitris G. Manolakis for sharing with them a manuscript [2] and for his correspondence answering several questions. This research was supported by the ATD program funded by NSF ATD DMS 1222567, NGA and DTRA.

2. ALGORITHM

2.1. Mixture Models

To better represent the complex background, we propose to use *Gaussian mixture models* $\mathbf{v} \sim \sum_j \pi_j \mu_j$, where $\pi_j > 0$ are coefficients with $\sum \pi_j = 1$, and μ_j represent probability measures associated to normal distributions $\mathcal{N}(\mathbf{c}_j, \Sigma_j)$ with centers $\mathbf{c}_j$ and covariance matrices Σ_j. We are particularly interested in the case where the Σ_j are all rank-deficient, and therefore each $\mathcal{N}(\mathbf{c}_j, \Sigma_j)$ is supported on an affine subspace. More generally, we can assume a *subspace mixture model* $\mathbf{v} \sim \sum_j \pi_j \mu_j$ where each μ_j is a probability measure with support contained in a low dimensional subspace passing through $\mathbf{c}_j$ and spanned by basis $\mathbf{B}_j$. The model parameters Θ_j in each case, $(\mathbf{c}_j, \Sigma_j)$ or $(\mathbf{c}_j, \mathbf{B}_j)$, can be estimated by K-means, K-means-like subspace clustering algorithms (e.g. [5]), fast multiscale techniques [7], or Expectation-Maximization (EM) methods. We may then adapt the estimators reviewed in [2] for use with the mixture models: A signal $\mathbf{x}$ is assigned to the jth cluster by maximizing the probability $p(\mathbf{x}|\Theta_j)$. Then, given a target signature $\mathbf{s}$ for the chemical cloud, the mixture versions of the two estimators, NMF and NSS, are given by,

$$T_{\text{mixNMF}}(\mathbf{x}) = T_{\text{NMF}}(\mathbf{x}; \Sigma_j, \mathbf{s}),$$
$$T_{\text{mixNSS}}(\mathbf{x}) = T_{\text{NSS}}(\mathbf{x}; \mathbf{B}_j, \mathbf{s}).$$

Alternatively, for the subspace mixture model, we may solve for g by least squares: let $\hat{\beta} = (\mathbf{A}_j^T \mathbf{A}_j)^{-1} \mathbf{A}_j^T \mathbf{x}$, where $\mathbf{A}_j = [\mathbf{s}, \mathbf{B}_j]$, and let $\hat{g}$ be the first entry of $\hat{\beta}$. This yields the mixture Linear Coefficient estimator $T_{\text{mixLC}} = \max\{\hat{g}, 0\}$.

We summarize the above steps in Alg. 1.

Algorithm 1 Mixture-models based detectors

Input: $\{\mathbf{I}^{(\ell)}\}_{1 \le \ell \le L} \subset \mathbb{R}^{mn}$: L hyperspectral frame(s), $\mathbf{s} \in \mathbb{R}^p$: chemical gas signature, T_{mix}: mixture-estimator (one of the three proposed above)

Output: $\{T_{\text{mix}}(\mathbf{I}^{(\ell)})\}_{1 \le \ell \le L}$: pixelwise detection scores

1: Fit a corresponding mixture model $\{\Theta_j\}$ to the first frame $\mathbf{I}^{(1)}$.

2: **for** each frame $\ell = 1, \ldots, L$

- Assign the pixels of $\mathbf{I}^{(\ell)}$ to nearest component models Θ_j by maximizing $p(\mathbf{x}|\Theta_j)$
- Evaluate T_{mix} at all pixels (within corresponding clusters) to produce detection scores for frame ℓ

end for

2.2. Computational Complexity

In mixture models, a segmentation of the spatial pixels is involved. Apparently, its complexity depends on the method we choose for segmentation: with the algorithm in [5], it is of the order $O(c_1 \cdot m \cdot n \cdot (d \cdot p + c_2 + \log(m \cdot n)))$, where d is the intrinsic dimension of the subspaces, c_1 and c_2 are two parameters of the algorithm (typical choices for our case are e.g., 10 and 20 respectively). The number of operations for computing the estimators T_{mixNMF}, T_{mixNSS}, T_{mixLC} are respectively of the order $O(p^2)$, $O((p+d) \cdot d^2)$ and $O((p+d) \cdot d^2)$.

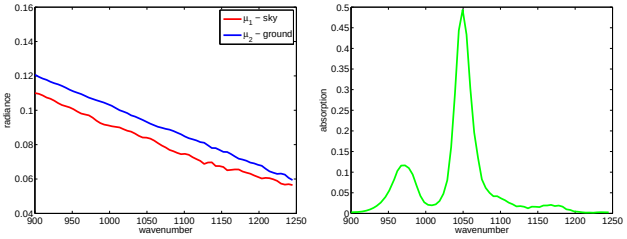

Fig. 1. Left: the mean radiance of sky and ground. Right: the absorption of the gas.

3. EXPERIMENTAL RESULTS

3.1. Simulations

We first use simulated data to demonstrate the performance of the proposed detection algorithms. First, radiance measurements at 68 different wavelength values are generated for $5,000$ pixels in the sky from $\mathcal{N}(\mathbf{c}_1, \Sigma_1)$ and for $5,000$ pixels on the ground from $\mathcal{N}(\mathbf{c}_2, \Sigma_2)$, where $\mathbf{c}_1$, $\mathbf{c}_2$ are taken from the MIT Lincoln Lab Challenge Data (see details in the next section) and are shown on the left of Figure 1, $\Sigma_1 = \text{diag}\{\sigma_{1,1}^2, \cdots, \sigma_{1,68}^2\}$, $\Sigma_2 = \text{diag}\{\sigma_{2,1}^2, \cdots, \sigma_{2,68}^2\}$, and all $\sigma_{i,j}$ are drawn i.i.d. from the uniform distribution on $[0.002a,\ 0.008a]$ with $a = \max(\mathbf{c}_2)$. Then, we take a chemical gas from the MIT Challenge Data and display its absorption signature s on the right of Figure 1. With this chemical, we generate radiance measurements for $1,000$ pixels on the ground with gas plume as $g\mathbf{s} + \mathbf{v}$, where $\mathbf{v} \sim \mathcal{N}(\mathbf{c}_2, \Sigma_2)$ and $g \sim \mathcal{N}(-0.01, 0.001)$. On the top left of Figure 2, radiance against wavenumber is plotted for 5 samples each from the groups of sky, ground and ground with plume. We compute the ACE statistic T_{mixNMF} using the truth values of $\mathbf{c}_i$ and Σ_i with $i = 1, 2$ as well as T_{NMF} using $\mathbf{c} = (\mathbf{c}_1 + \mathbf{c}_2)/2$ and $\Sigma = (\Sigma_1 + \Sigma_2)/2$ for all pixels. These ACE scores of 10 samples are displayed on the top right of Figure 2. At the bottom of Figure 2, we show the histograms of the ACE statistics for using 1 Gaussian on the left and 2 Gaussians on the right. The histograms are colored by groups of sky, ground and plume. From this figure, we see that the ACE scores of plumes are larger than those of only sky or ground when we use two Gaussians instead of only one Gaussian. This is consistent with our understanding of the ACE algorithm, i.e., radiance with plume is of a smaller angle from the gas signature and thus has a larger value of the ACE statistic.

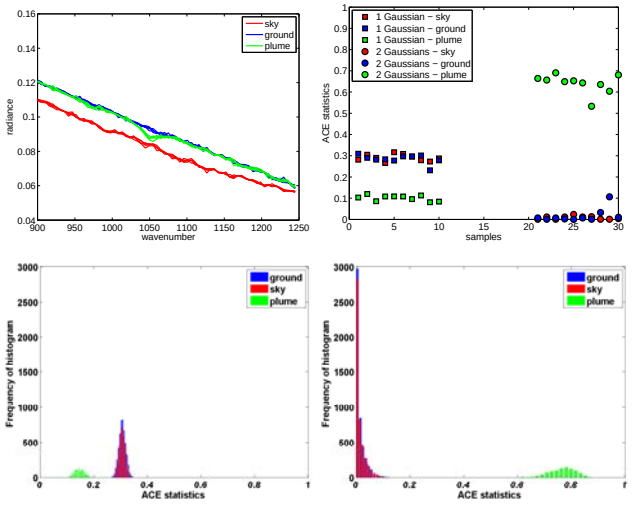

Fig. 2. Top left: radiance against wavenumber of 5 samples each from the groups of sky, ground and plume. Top right: the ACE scores of 10 samples computed by 1 Gaussian and 2 Gaussians respectively. Bottom left: histogram of the ACE scores of using 1 Gaussian. Bottom right: histogram of the ACE scores of using 2 Gaussians.

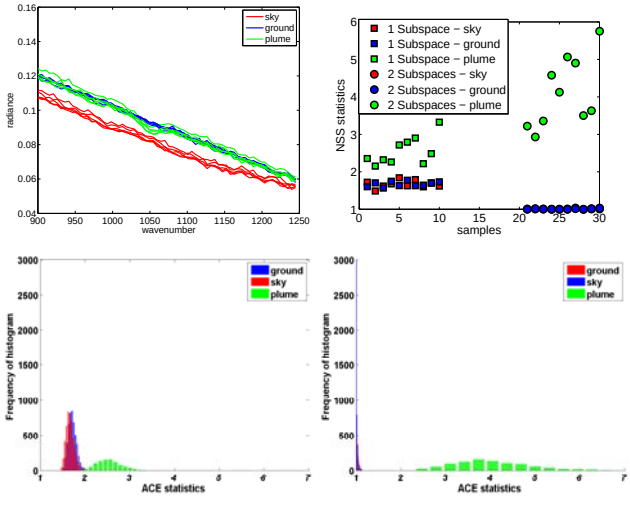

Fig. 3. Top left: radiance against wavenumber of 5 samples each from the groups of sky, ground and plume. Top right: the NSS scores of 10 samples computed by 1 subspace and 2 subspaces respectively. Bottom left: histogram of the NSS scores of using 1 subspace. Bottom right: histogram of the NSS scores of using 2 subspace.

Next, we conduct a similar experiment to illustrate the performance of the mixture subspace model. We generate radiance measurements at the same 68 wavenumbers for $5,000$ pixels in the sky from $h\mathbf{c}_1+\epsilon$ and $5,000$ pixels on the ground from $h\mathbf{c}_2+\epsilon$, where $h \sim \mathcal{N}(1, 0.01)$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$, $\Sigma_\epsilon = \text{diag}\{\sigma^2_{\epsilon,1}, \cdots, \sigma^2_{\epsilon,68}\}$ with the $\sigma_{\epsilon,j}$ being drawn i.i.d. from $\mathcal{N}(0, 0.005\max(\mathbf{c}_2))$. The radiance measurements of $1,000$ pixels on the ground with plume are generated as $g\mathbf{s}+h\mathbf{c}_2+\epsilon$ with $g \sim \mathcal{N}(-0.01, 0.001)$. Figure 3 shows 5 samples each from the groups of sky, ground and plume on the top left. We then compute the statistic for the algorithm NSS and mixture NSS by using the ground truth. More precisely, we let $\mathbf{B}_i = \mathbf{c}_i$ for $i = 1, 2$, and $\mathbf{B} = (\mathbf{c}_1 + \mathbf{c}_2)/2$. These statistic scores are displayed on the top right of Figure 3. At the bottom of Figure 3, we show the histograms of the NSS statistics for using 1 subspace on the left and 2 subspaces on the right. The histograms are colored by groups of sky, ground and plume. Using the mixture subspace model, we obtain larger NSS scores for the plume pixels; the separation between the plume and other regions is also much more obvious.

3.2. Real data

Real data sets are provided within the Algorithms for Threat Detection (ATD) program, funded by NSF, DTRA and NGA.

3.2.1. MIT Lincoln Lab Challenge Data

MIT Lincoln Lab provides a package of four data sets, two of which are for released gas and two for embedded gas. In each of the four, available measures include a radiance data cube, its matrix form, the absorption coefficient spectrum for the gas of interest, plume present mask and some other quantities. The chemical needs to be detected based on a single cube, which is roughly of the size $200 \times 300 \times 100$, where 200×300 is the spatial size and 100 is the sensor spectrum size.

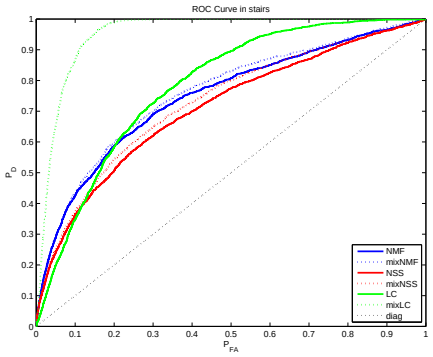

Fig. 4. Comparison between single and mixture models using ROC curves and detection maps.

We evaluate the performance of the three detectors (NMF, NSS, LC) and their corresponding mixture models (mixNMF, mixNSS, mixLC). The receiver operating characteristic (ROC) curves and the detection maps are shown in Figure 4 for one of the embedded data cubes (also the most difficult one). From the figure we see that using mixture models for background can improve the detection results. Among the three algorithms, mixture LC works the best.

3.2.2. *Fabry-Perot Interferometer Sensor Data*

This data set contains five time series of hyperspectral data, corresponding to different combinations of the kind of chemical material, the release amount, and the sensor used. We first checked the time derivatives of the spectra (along the sequence) but did not observe anything useful. We then applied our algorithm without utilizing the temporal information and compared our results with ACE [4] in Figure 5. We see that our detection map is consistently clearer than ACE. In this experiment, we used the first frame for background modeling and the 30th frame for testing (for all data sets); future work will utilize multiple clean frames for enhanced background modeling.

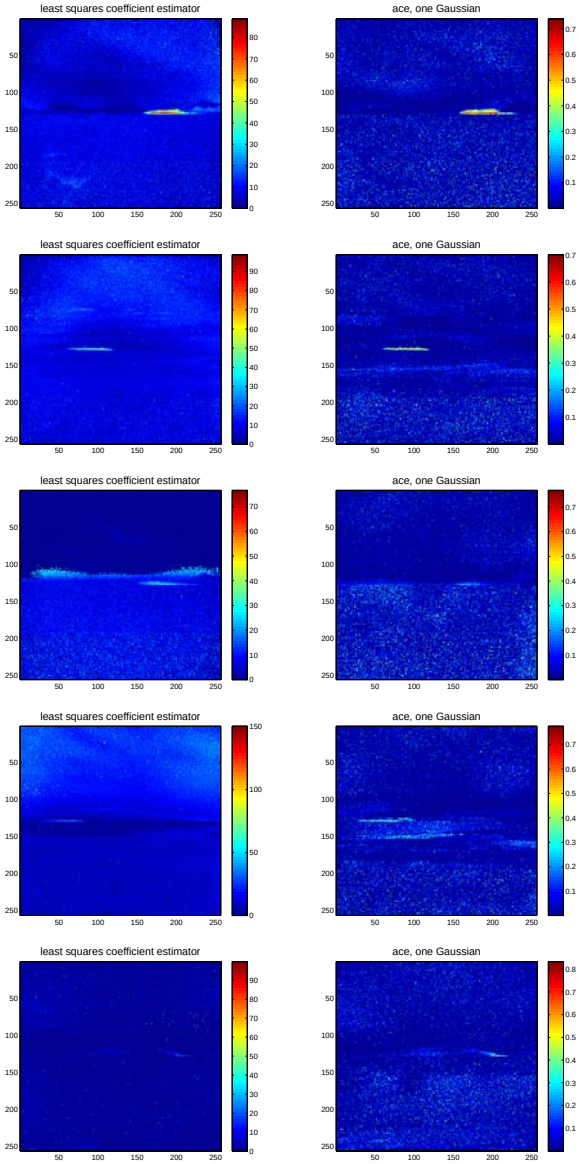

Fig. 5. Detection results by our estimator T_{mixLC} (left column) and ACE (right column) on the five hyperspectral movies of the Fabry-Perot data set. Note that our detection map is consistently cleaner than ACE.

4. REFERENCES

[1] A. Ihler and M. Mandel, Kernel density estimation toolbox. Available at *http://www.ics.uci.edu/ ihler/code/kde.html*, 2003.

[2] D. Manolakis, S. Golowich and R. DiPietro. Signal Processing for Hyperspectral Remote Sensing of Chemical Clouds *IEEE Signal Processing Magazine*, under review.

[3] D. Manolakis, C. Siracusa and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, 2001.

[4] S. Kraut, L. Scharf, and L. T. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*. vol. 49, no. 1, pp. 1–16, 2001.

[5] T. Zhang, A. Szlam, Y. Wang and G. Lerman. Hybrid Linear Modeling via Local Best-fit Flats. *CVPR*, 2010.

[6] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference and prediction. *Springer*, pp. 66–68, 2009.

[7] G. Chen, M. Iwen, S. Chin and M. Maggioni. A Fast Multiscale Framework for Data in High Dimensions: Measure Estimation, Anomaly Detection, and Compressive Measurements. Visual Communications and Image Processing (VCIP), 2012 IEEE.

[8] W.K. Allard, G. Chen, and M. Maggioni. "Multiscale geometric methods for data sets II: Geometric multiscale analysis". *Appl. Comput. Harmon. Anal.*, Available online Sep. 2011.

[9] M. Maggioni. "Geometric Estimation of Probability Measures in High Dimensions". *IEEE Asilomar Conference on Signals, Systems and Computers*, 2013.

[10] M. Maggioni. "Multiscale Geometric Measure Estimation". In preparation, 2014.

[11] M. Maggioni, S. Minsker and N. Strawn. "Dictionary Learning and Non-Asymptotic Bounds for the Geometric Multi-Resolution Analysis". *arXiv*, 2014.

[12] A. V. Little, Y.-M. Jung, and M. Maggioni. "Multiscale Estimation of Intrinsic Dimensionality of Data Sets". *Proc. AAAI*, 2009

[13] A. V. Little, M. Maggioni, and L. Rosasco. "Multiscale Geometric Methods for Data Sets I: Multiscale SVD, Noise and Curvature". MIT Report, Sep. 2012.

[14] A. V. Little. "Estimating the Intrinsic Dimension of High-Dimensional Data Sets: A Multiscale, Geometric Approach". *PhD Thesis, Duke University*, Apr. 2011.