# Object Tracking by Hierarchical Decomposition of Hyperspectral Video Sequences: Application to Chemical Gas Plume Tracking

Guillaume Tochon, *Member, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, Mauro Dalla Mura, *Member, IEEE*, and Andrea L. Bertozzi, *Member, IEEE*

***Abstract*— It is now possible to collect hyperspectral video sequences at a near real-time frame rate. The wealth of spectral, spatial, and temporal information of those sequences is appealing for various applications, but classical video processing techniques must be adapted to handle the high dimensionality and huge size of the data to process. In this paper, we introduce a novel method based on the hierarchical analysis of hyperspectral video sequences to perform object tracking. This latter operation is tackled as a sequential object detection process, conducted on the hierarchical representation of the hyperspectral video frames. We apply the proposed methodology to the chemical gas plume tracking scenario and compare its performances with state-of-the-art methods, for two real hyperspectral video sequences, and show that the proposed approach performs at least equally well.**

***Index Terms*— Binary partition tree, gas plume tracking, hyperspectral video sequence, object detection.**

## I. Introduction

HYPERSPECTRAL imaging (also called imaging spectroscopy) is the process of dividing the electromagnetic spectrum into several narrow and contiguous wavelengths, and simultaneously acquiring an image for each wavelength. All those single-band images are then stacked in a 3-D ($N_x \times N_y \times N_\lambda$) data cube to produce the resulting hyperspectral image (HSI), where $N_x$ and $N_y$ correspond to the number of rows and columns of the single-band images, respectively, and $N_\lambda$ is the number of wavelengths (also called spectral bands). To each pixel of the image is, therefore, associated an $N_\lambda-$dimensional vector (or spectrum), which depicts the way the pixel site has interacted with the incident light and can be viewed as a function of the spectral wavelength $\lambda$. This spectrum depends on the materials composing the pixels, since each physical material can be uniquely defined by its spectral signature [1]. When analyzing a scene in the visible and near-infrared domain, this signature is called *reflectance* and corresponds to the light that was reflected by the scene. When working in the longwave infrared (LWIR) domain, the signature is expressed in terms of *emissivity*, being the ratio of the energy emitted by the scene with respect to the incident energy. Hyperspectral imagery, by acquiring detailed spectral properties of the scene, finds an always-increasing number of real-life applications in various remote sensing fields, such as vegetation mapping [2], geological [3], and hydrological sciences [4], as well as food quality inspection [5], [6] and medical imagery [7], [8], among others. However, this wealth of spectral information comes with several drawbacks, such as the high-dimensional nature of the data to be handled or the computational burden due to the large amount of data to process, making hyperspectral imagery a very dynamic and quickly evolving field of research [9].

Thanks to the fast development of imaging sensors, it is now possible to acquire sequences of HSIs at near real-time rates. The combination of the high spectral resolution proper to HSIs with the ability of video sequences to record phenomena evolving with time is appealing for the time monitoring of objects based on their spatial and spectral properties. However, some additional efforts are required to extend traditional video processing techniques (such as block noise reduction, motion compensation, or object tracking) to the high-dimensional space structured by hyperspectral data. In addition, available benchmark hyperspectral video data sets are scarce, and the lack of ground-truth data makes the quantitative evaluation of any novel method very challenging.

In particular, we focus, in this paper, on object tracking in hyperspectral video sequences. Object tracking can be defined as the process of following the motion of points or regions of interest as they evolve with time within a video sequence. Object tracking finds numerous applications in everyday life, such as automated surveillance, motion-based recognition, visual servoing, or traffic monitoring, and has been widely studied in the area of computer vision [10], [11] within the framework of traditional video sequences. However, most existing algorithms poorly adapt to the high dimensionality inherent to hyperspectral data. To the best of our knowledge, the only existing tracking method specifically designed and evaluated on real-time hyperspectral video sequences is the one introduced in [12] and [13]. It makes use of the

Manuscript received June 30, 2016; revised January 9, 2017; accepted March 13, 2017. Date of publication May 5, 2017; date of current version July 20, 2017. This was supported in part by ANR under Grant ANR-16-ASTR-0027-01 and in part by the National Science Foundation under Grant DMS-1118971, Grant DMS-0914856, and Grant DMS-1417674. *(Corresponding author: Guillaume Tochon.)*

G. Tochon was with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, F-38000 Grenoble, France. He is now with the EPITA Research and Development Laboratory, Graduate School of Computer Science and Advanced Techniques, F-94276 Paris, France (e-mail: guillaume.tochon@lrde.epita.fr).

J. Chanussot and M. Dalla Mura are with the Grenoble Images Speech Signals and Automatics Laboratory, Centre National de la Recherche Scientifique, Université Grenoble Alpes, F-38000 Grenoble, France.

A. L. Bertozzi is with the Department of Mathematics, University of California at Los Angeles, Los Angeles, CA 90095-1555 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2694159

0196-2892 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

TABLE I
SUMMARY OF EXISTING WORKS RELATED TO THE TRACKING OF CHEMICAL PLUMES IN LWIR HYPERSPECTRAL SEQUENCES

| General approach / Related properties | Statistical | Clustering-based | Object tracking |
|---|---|---|---|
| Previous works | AMF [19], AMSD [21], CMF [16], ACE [24] | graph Laplacian [32], MBO [33], [34], RNMF [36] | Binary Partition Tree [17] |
| Use of temporal information | ✗ | ✓ | ✓ |
| On-the-fly detection/tracking | ✓ | ✗ | ✗ |
| No reference spectrum needed | ✗ | ✓ | ✓ |

mean shift tracker algorithm [14], and the tracked object is represented as a fixed primitive geometric shape and does not adapt well to applications where either the tracked object is nonrigid or where the precise shape of the object is required. The development of new algorithms able to face these challenges is necessary. Chemical gas plume tracking is a typical application that would surely benefit from the design of such new hyperspectral object tracking methods. As a matter of fact, such application is of great interest for several domains. In the environmental protection field, for example, gas plume tracking could be exploited to monitor pollutant gas clouds emitted by industrial sources [15], in order to minimize their impact on the environment and the potential harm they could cause on human population living nearby. In the defense and security area, a possible usage of such tracking method could be to detect the use of chemical gas weapons [16]. Most gases do not respond in the visible spectrum range, but only in a restrained portion of the LWIR domain, hence the need of a fine sampling of the electromagnetic spectrum and the incapacity of classical video techniques to detect (and, *a fortiori*, to track) them. In addition, a gas plume is a nonrigid object whose shape evolves unpredictably with time. The necessity of a fine spectral description of the scene over time makes hyperspectral video sequences the most suited tool for such detection and tracking application.

In the following, we propose a novel algorithm for hyperspectral object tracking. Our method, based on a hierarchical analysis of the frames of the hyperspectral sequence, is able to track a region of interest whose shape may evolve with time, without any prior knowledge about the materials constituting the region. The proposed work, sketched in [17], is based on a general assumption about the hyperspectral video sequence, namely, that only the object of interest is in motion with respect to a fixed background in the hyperspectral video sequence. It then uses the spectral, spatial, and temporal information derived from the sequence to perform a sequential object detection process over the hierarchical decomposition of each frame, finally producing the shape and extent of the tracked object without any prior knowledge on its shape or its spectral signature. We apply it to the chemical gas plume tracking challenge and compare its quantitative and qualitative performances against different state-of-the-art methods for two data sets.

The remainder of this paper is organized as follows. Section II presents the state of the art related to the tracking of chemical gas plume in LWIR sequences. Section III introduces some background notions, namely, the notations used throughout this paper as well as the binary partition tree (BPT) used to perform the hierarchical representation of the hyperspectral frames. Section IV presents the proposed hyperspectral object tracking algorithm from a methodological point of view. Its adaptation to the gas plume tracking is conducted in Section V. Results are presented and discussed in Section VI. Conclusion and future research avenues are drawn in Section VII.

## II. STATE OF THE ART

The detection of gas plumes has been already largely addressed in the literature [18], where most techniques can be categorized either as anomaly/target detectors or as clustering-based methods, as summarized in Table I. The most popular and natural approach is to consider the gas plume as anomalous with respect to the background, and thus make use of conventional anomaly and target detectors. The hypotheses made on the structure of the background clutter model, the nature, and the variability of the anomalous signature lead to different detectors, such as the adaptive matched filter (AMF) operated in [19], the adaptive matched subspace detector (AMSD) investigated in [20]–[22], the clutter matched filter (CMF) in [16] and [23], the adaptive cosine/coherence estimator (ACE), and the orthogonal subspace projection in [24]. Performance comparison of AMF and ACE detectors for gas plume detection can be found in [25]. For further details about previous anomaly detectors, the reader is referred to [26]–[28]. The major drawback of these methods is that they cannot be operated without a reference target spectrum, often estimated using spectral libraries [29], [30]. Moreover, they do not use any temporal information, since the target detection process is applied on each frame, independently of the previous results. Those methods nevertheless allow for frame-by-frame tracking, since the frames in the sequence are processed directly after their release. Moreover, the target detection process is often merely implemented as a statistical test, and is, in practice, computationally fast enough to be real time (we consider here a technique to be real time if the processing of a frame can be completed before the next frame is acquired, for frame rates in the order of a few seconds per frame).

A second popular approach that recently emerged is to address the plume detection problem as a clustering. In that case, it is assumed that the properties of the spectral signature of the plume are sufficiently different from those of the background, so it is possible to compose a cluster solely containing the plume. Clustering-based method notably

include [31], which performed spectral clustering (note that it also requires a reference target spectrum), [32], which investigated graph Laplacian-based spectral clustering, [33] and [34], where a semisupervised diffused interface clustering (called the Merriman-Bence-Osher scheme [35]) was implemented, and [36], where the clustering is handled through robust nonnegative matrix factorization (RNMF). Temporal information is only considered in [33], [34], and [36], where the proposed clustering is performed on several stacked consecutive frames (seven in the case of [33] and [34], and even the whole sequence for [36]). While it is necessary to differentiate the faint gas plume signature from the background, the main downside of stacking consecutive frames is that the underlying clustering methods become not suited for the on-the-fly detection and tracking of the chemical plume, which is an essential characteristic for passive surveillance scenarios.

In a preliminary study [17], we proposed an alternative approach to perform the tracking of a chemical plume in LWIR hyperspectral sequences. Relying on a hierarchical representation of the frames in the sequence, the tracking was performed by sequentially recomposing the followed object from the collection of regions provided by the hierarchical decomposition of each frame. However, a preprocessing stage (being a Midway equalization [37]) was conducted over the whole sequence, making the gas plume impossible to be instantly tracked. In addition, the frame featuring the release of the plume was supposedly known in advance, in order to launch the tracking method on time. Here, we relax those limitations.

1) The time where the plume appears in the sequence is no longer supposed to be known, and a change detection procedure is conducted on every frame in order to trigger the tracking.
2) The proposed method does not require any preprocessing step, such that the tracking can now be conducted on a frame-by-frame basis.
3) A generic *one object, one region* paradigm is adopted for the hierarchical object detection process. This simple rule actually allows for a greater flexibility. Indeed, the matching procedure (i.e., retrieving the tracked object among the set of candidate ones, provided by the hierarchical representation) becomes invariant of the application, from which only depend the defined features for the object detection procedure.

In addition, we provide here a quantitative and qualitative performance analysis and comparison with state-of-the-art methods, by notably delineating some reference data for two different LWIR hyperspectral sequences. To the best of our knowledge, this is the first time that ground-truth data are created for gas plume tracking applications.

## III. Background Notions

### A. General Notations

In the following, let $\mathfrak{I} = \{\mathcal{I}_t, t = 1 \ldots, N_t\}$ be a sequence of $N_t$ hyperspectral frames indexed by $t$ (supposed to correspond to a time index). Each hyperspectral frame can be viewed as a mapping $\mathcal{I}_t : E \rightarrow \mathbb{R}^{N_\lambda}$, where $E \subseteq \mathbb{Z}^2$ represents the spatial support of the frame and $N_\lambda$ is the number of spectral channels. The elements of $E$ are the pixels $x_i \in E$, and their associated values through $\mathcal{I}_t$ are $\mathcal{I}_t(x_i) \in \mathbb{R}^{N_\lambda}$ and will be denoted in short by $\mathbf{x}_i^t$. More generally, bold notations will refer to vector-valued elements. A region $\mathcal{R} \subseteq E$ is a collection of pixels sharing some properties, and will be equally handled either as a set or through its indicator function with respect to $E$, $\mathbb{1}_\mathcal{R} : E \rightarrow \{0, 1\}$ with $\mathbb{1}_\mathcal{R}(x_i) = 1$ if $x_i \in \mathcal{R}$, 0 otherwise, which is the binary representation of $\mathcal{R}$.

### B. Hierarchical Representations

By definition, a pixel is the smallest structuring element of an image and is the lowest scale, or level of details, at which the image can be represented. Many image processing low-level applications, such as denoising or filtering, operate on pixels. However, for some other applications (such as segmentation or object recognition), the pixel-based representation is not well suited, because the scale of interest is larger than pixels. Consequently, those applications are performed on region-based representations of images, which are high-level descriptions of such images, since each region, composed of a set of pixels, is assumed to contain some semantic meaning.

For HSIs, region-based representations allow to combine spectral and spatial properties of the scene, and several studies have shown the interest of such approaches with respect to traditional pixel-based approaches for classification [38], segmentation [39], and unmixing [40], notably. Images intrinsically contain several scales of interest, and the level of exploration to choose greatly depends on the underlying application. Hierarchical representations are a solution to tackle this issue, since they represent the image under all potential scales of interest. More specifically, a hierarchy, hereafter denoted by $H$, can be described as a collection of regions, which are either pairwise disjoints or nested

$$H = \{\mathcal{R} \subseteq E\} \quad \text{s.t } \forall \mathcal{R}_i,\ \mathcal{R}_j \in H,\ \mathcal{R}_i \cap \mathcal{R}_j \in \{\emptyset, \mathcal{R}_i, \mathcal{R}_j\}. \tag{1}$$

The hierarchy can be built regardless of the application, and the level at which it should be explored can be tuned afterward.

### C. Binary Partition Tree

Among all existing hierarchical region-based representations, the BPT, by its capacity of handling very high-dimensional data, such as HSIs, has been receiving increasing attention. The BPT representation was initially proposed in [41] for grayscale and color images, and was further extended to hyperspectral imagery in [42]. It is now used for classical hyperspectral remote sensing tasks, such as segmentation [43], [44], classification [45], unmixing [46], and object detection [47], [48].

*1) Construction of the BPT:* Starting from an initial partition, which can be the pixel level or a coarser segmentation of the image, an iterative bottom–up region merging algorithm fuses neighboring regions until there is only one region remaining (corresponding to the whole image support). In the corresponding tree representation, each region of the

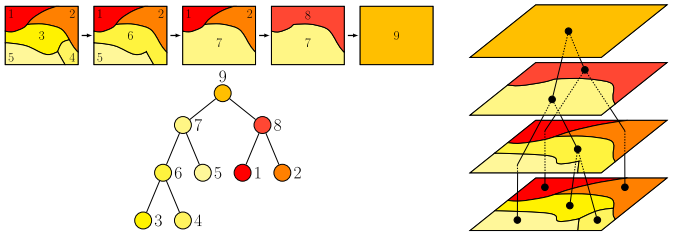


Fig. 1. Example of a BPT representation.

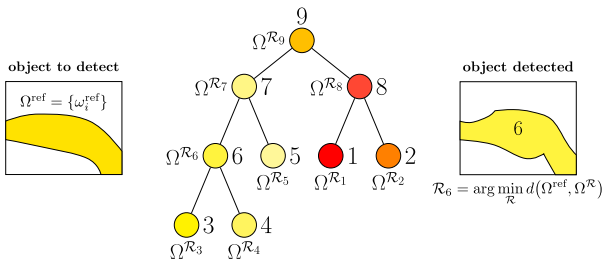



Fig. 2. Example of object detection using the BPT structure presented in Fig. 1.

initial partition is a *leaf* node, the whole image is the *root* node, and each node in-between corresponds to the merging of two children regions. An example of BPT representation can be observed in Fig. 1. A proper region merging algorithm is bound to the definition of three parameters of prime importance:

1) the initial partition of the image, specifying the initial regions on which the BPT is built;
2) the region model $\mathcal{M}_{\mathcal{R}}$, which defines how a region $\mathcal{R}$ is mathematically modeled, and how to handle the union of two regions;
3) the merging criterion $\mathcal{O}(\mathcal{R}_i, \mathcal{R}_j)$, which quantifies the similarity between two neighboring regions $\mathcal{R}_i$ and $\mathcal{R}_j$ by measuring the distance between their corresponding region models.

There exist several region models and their nonexhaustive list of adapted merging criteria in the BPT literature (see [44], [49]). It is also worth noting that a priority term is often added to the merging algorithm: at a current iteration, each region whose size is below a given percentage of the average region size for this iteration is given the merging priority, regardless of the distance with its neighbors. The threshold percentage is typically set to 15% [50], to avoid the unwanted presence of small and meaningless regions in the last iterations of the merging process.

*2) Object Detection With a BPT:* The BPT representation is naturally suited to perform object detection. As a matter of fact, the BPT can be viewed as a decomposition of the image in relevant regions across the image support and at various scales. It naturally provides a finite number of candidates regions represented by nodes in the tree structure, hence drastically reducing the search space with respect to a pixel-based representation. Plus, it is known from [51] and [52] that considering multiple segmentations of an image at various scales, hence more or less fine regions, greatly enhances the recognition of object positions and shapes.

The object detection process can be formalized as follows: given a set of $N$ reference features $\Omega^{\text{ref}} = \{\omega_i^{\text{ref}}\}_{i=1}^N$ corresponding to an object of interest, and given a BPT representation $H_{\text{BPT}}$ of an image, the object detection process retrieves, for each region $\mathcal{R} \in H_{\text{BPT}}$, its own set of features $\Omega^{\mathcal{R}} = \{\omega_i^{\mathcal{R}}\}_{i=1}^N$ in the image. It subsequently evaluates the similarity between $\Omega^{\text{ref}}$ and $\Omega^{\mathcal{R}}$ for some application-dependent distance function $d\ (\Omega^{\text{ref}}, \Omega^{\mathcal{R}})$ (also often called cost, objective, or energy function). The selected node is the one minimizing this distance function. Alternatively, all nodes below a set threshold can be retained. Fig. 2 shows an example of object detection conducted with a BPT.

The strength of the BPT-based object detection is the flexibility. As a matter of fact, the selection of relevant features and appropriate distance functions to compare them, combined with a proper BPT construction, ensures the method to be adaptable to almost any kind of application. Features classically used to perform BPT-based object detection for RGB images include mean CIE $L*a*b$ color and homogeneity within the region [53] and reference shape models [54], [55]. To perform road and building detection in HSIs acquired over urban environments, Valero *et al.* [47], [48] proposed spatial features, such as the area of the region and of the smallest oriented bounding box containing it, and spectral features, such as the correlation between the region mean spectrum and a reference spectrum (asphalt for roads, for instance), and some class membership homogeneity.

## IV. BPT-Based Hyperspectral Object Tracking

Object tracking algorithms are generally organized in two steps that are sequentially addressed.

1) The *motion prediction step*, whose goal is to estimate the position of the object in the next frame. This is usually conducted through an extrapolation from the current position with the estimations of the motion direction and velocity (plus some margin of error). Motion prediction allows to reduce the search space by defining an area where the object can be found with a high probability.
2) The *matching step*, which searches the object in the area predicted by the motion estimation step. It typically involves the definition of reference features for the sought object and their comparision with features derived from candidate objects located in the search space. The tracked object is declared to be the candidate whose features are the closest from the reference ones.

The formulation of the motion prediction and matching steps greatly depends on the nature of the tracker (punctual object, primitive shape, exact contour, or region). The reader is referred to [10] and [11] for complete and extensive reviews about classical object tracking algorithms.

The matching step of the tracking algorithm is equivalent to an object detection process and, therefore, falls in the scope of the BPT processing. Consequently, object tracking appears as an iterative object detection procedure, in which the features used to identify the tracked object in each frame of the sequence can be derived from the result of the previous motion prediction step. The proposed methodology, detailed

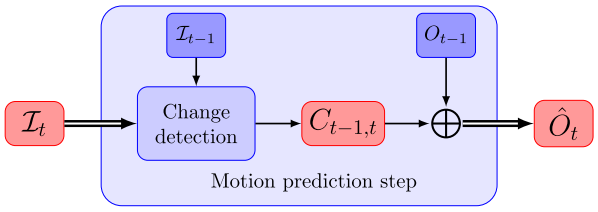



Fig. 3. Workflow of the proposed motion prediction step.

in the following, is based on the assumption that only the object of interest is in motion over a fixed background in the video sequence. This assumption may seem somewhat restrictive. For now, however, most spectrometer sensors providing hyperspectral video sequences are still sensors mounted on a tripod, therefore producing sequences of still images with a fixed background. Like classical object tracking algorithms, the proposed methodology is decomposed in a motion prediction step and a matching step.

### *A. Motion Prediction Step*

The purpose of the motion prediction step is to restrict, for each frame $\mathcal{I}^t$, the search space only to a neighborhood where the object is assumed to be found with a high probability. Here, we propose to go even one step further: being the object $O_t \subseteq E$ a region of $E$, the motion prediction step outputs an estimate region $\hat{O}_t$, such that the shape and position of $O_t$ and $\hat{O}_t$ globally coincide. This estimation is then used to steer the matching step to locate a candidate region that is similar to the estimate region both in terms of position and shape.

The method we propose to perform the motion prediction is decomposed in two inner steps, as shown in Fig. 3. First, the change mask $C_{t-1,t}$ between two consecutive frames $\mathcal{I}_{t-1}$ and $\mathcal{I}_t$ is estimated. This change mask features areas where significant change occurs between $t-1$ and $t$ due to the motion of the object. In a second step, the change mask is combined with the position of the object estimated at $t-1$, denoted $O_{t-1}$ to produce an estimation of position at $t$, named $\hat{O}_t$.

*1) Derivation of the Change Mask:* Recall that we position ourselves in the context of a video sequence where the depicted scene is a superposition of a still background and a moving object. We also assume that this object of interest can be represented by a (possibly unknown) fixed spectral response, denoted $\mathbf{o}$. Following these assumptions, it becomes possible to express each pixel value $\mathbf{x}_i^t$ as an additive combination of the object signature $\mathbf{o}$ and the background response $\mathbf{b}_i$ at location $i$ (which does not vary with time as the background is supposed to be still), plus some additive noise:

$$\mathbf{x}_i^t = \alpha_i^t \mathbf{o} + (1-\alpha_i^t)\mathbf{b}_i + \boldsymbol{\eta}^t \tag{2}$$

with $\alpha_i^t \in [0,1]$ being the fractional proportion of the object response in $\mathbf{x}_i^t$, and modeling the possible transparency of the object. Note that (2) can also be understood as a particular case of linear mixing model [56] where each background pixel as well as the object signature would be considered as endmembers. In that framework, $\alpha_i^t$ has the same meaning as classical spectral abundances. One can then write for the frame difference $\mathcal{I}_{\Delta t} = \mathcal{I}_t - \mathcal{I}_{t-1}$

$$\mathbf{x}_i^{\Delta t} = \mathbf{x}_i^t - \mathbf{x}_i^{t-1} = \alpha_i^{\Delta t}(\mathbf{o} - \mathbf{b}_i) + \boldsymbol{\eta}^{\Delta t} \tag{3}$$

with $\alpha_i^{\Delta t} = \alpha_i^t - \alpha_i^{t-1}$ is the temporal variation of the fractional proportion of object signature in the considered difference pixel $\mathbf{x}_i^{\Delta t}$, and $\boldsymbol{\eta}^{\Delta t} = \boldsymbol{\eta}^t - \boldsymbol{\eta}^{t-1}$. Consequently, a variation in the proportion of the object signature at pixel position $i$ between time instances $t-1$ and $t$ yields $\alpha_i^{\Delta t} \neq 0$, while $\alpha_i^{\Delta t} = 0$ when no change occurs. This observation naturally leads to formulate the following two-hypotheses test:

$$\begin{aligned} \mathbf{H}_0 &: \ \alpha_i^{\Delta t} = 0, \text{there is no change in } \mathbf{x}_i \text{ between } t-1 \text{ and } t \\ \mathbf{H}_1 &: \ \alpha_i^{\Delta t} \neq 0, \text{there is change in } \mathbf{x}_i \text{ between } t-1 \text{ and } t. \end{aligned} \tag{4}$$

For the purpose of hyperspectral anomaly or change detection, it is classically assumed that $\boldsymbol{\eta}^{\Delta t}$ follows a Gaussian distribution with zero mean and covariance $\Sigma$, $\boldsymbol{\eta}^{\Delta t} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ (see [26], [27]). In such case, the probability distribution function $f$ of $\mathbf{x}_i^{\Delta t}$ is known under both hypotheses

$$\begin{aligned} \mathbf{H}_0 &: \ f(\mathbf{x}_i^{\Delta t} | \alpha_i^{\Delta t} = 0) \sim \mathcal{N}(\mathbf{0}, \Sigma) \\ \mathbf{H}_1 &: \ f(\mathbf{x}_i^{\Delta t} | \alpha_i^{\Delta t} \neq 0) \sim \mathcal{N}\left(\boldsymbol{\mu}_i^{\Delta t}, \Sigma\right) \end{aligned} \tag{5}$$

with $\boldsymbol{\mu}_i^{\Delta t} = \alpha_i^{\Delta t}(\mathbf{o} - \mathbf{b}_i)$ unknown. Therefore, detecting a change in the time difference frame reduces to testing whether each pixel difference is drawn from a zero-mean Gaussian distribution or not. The classical solving of this two-hypotheses test involves a generalized likelihood ratio test (GLRT) whose expression for the pixel $\mathbf{x}_i^{\Delta t}$, $\Lambda(\mathbf{x}_i^{\Delta t})$ is the following:

$$\Lambda(\mathbf{x}_i^{\Delta t}) = \frac{\max_{\boldsymbol{\mu}_i^{\Delta t} \neq 0} f(\mathbf{x}_i^{\Delta t} | \mathbf{H}_1)}{f(\mathbf{x}_i^{\Delta t} | \mathbf{H}_0)} \underset{\mathbf{H}_0}{\overset{\mathbf{H}_1}{\gtrless}} \gamma. \tag{6}$$

The unknown mean $\boldsymbol{\mu}_i^{\Delta t}$ that maximizes the numerator of (6) is the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\mu}}_i^{\Delta t}$ of $\boldsymbol{\mu}_i^{\Delta t}$, which is defined as follows:

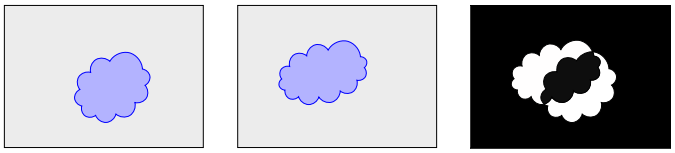
$$\hat{\boldsymbol{\mu}}_i^{\Delta t} = \frac{1}{S} \sum_{i=1}^{S} \mathbf{x}_i^{\Delta t} \tag{7}$$

where $S = S_{\text{width}} \times S_{\text{height}}$ is a predefined window of neighboring pixels of $\mathbf{x}_i^{\Delta t}$ (here set to $5 \times 5$, the rationale will be further detailed in Section IV-C).

Plugging (7) into (6), and solving for $\mathbf{x}_i^{\Delta t}$ finally yields to

$$\Lambda(\mathbf{x}_i^{\Delta t}) = S(\hat{\boldsymbol{\mu}}_i^{\Delta t})^T \Sigma^{-1} \hat{\boldsymbol{\mu}}_i^{\Delta t} \underset{\mathbf{H}_0}{\overset{\mathbf{H}_1}{\gtrless}} \gamma_{\text{GLRT}} \tag{8}$$

with $(\cdot)^T$ denoting the transpose operation. The probability distribution of $\Lambda(\mathbf{x}_i^{\Delta t})$ under $\mathbf{H}_0$ and $\mathbf{H}_1$ is given in terms of $\chi^2_{N_\lambda}$ with $N_\lambda$ degrees of freedom (being the number of spectral channels in the hyperspectral frame) and noncentral $\chi^2_{N_\lambda,\phi}$ with $N_\lambda$ degrees of freedom and noncentrality parameter $\phi = \Lambda(\mathbf{x}_i^{\Delta t})$ [27], respectively. Knowing the distribution under both hypotheses allows one to set $\gamma_{\text{GLRT}}$ to achieve either a predefine probability of false alarm or a probability of detection. The binary change mask $C_{t-1,t}$ is finally obtained by thresholding according to $\gamma_{\text{GLRT}}$.



(a) (b) (c)

Fig. 4. (b) Object $O_t$ can be retrieved from (a) $O_{t-1}$ and (c) change mask $C_{t-1,t}$ by $O_t = O_{t-1} \bigoplus C_{t-1,t}$.

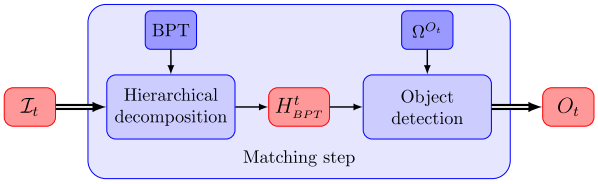



Fig. 5. Workflow of the proposed matching step.

*2) Estimation of the Position:* Under the assumption of a single moving object overlaying a fixed background, the change mask $C_{t-1,t}$ is composed of two categories of regions.

1) Regions being left by the object. Those are the areas that were occupied by the object in frame $\mathcal{I}_{t-1}$ ($\alpha_i^{t-1} > 0$) but no longer in $\mathcal{I}_t$ ($\alpha_i^t = 0$).
2) Regions invaded by the object. Those are the areas that are reached by the object in frame $\mathcal{I}_t$ (thus, $\alpha_i^{t-1} = 0$ and $\alpha_i^t > 0$).

Intuitively, the new position of the object, $O_t$, is composed of the previous position of the object, $O_{t-1}$, minus the regions that have been left by the object plus the regions that have been reached by it, as shown in Fig. 4. Mathematically, this can be formulated by

$$\hat{O}_t = O_{t-1} \bigoplus C_{t-1,t} \tag{9}$$

where $\bigoplus$ denotes the binary XOR operation. However, as both the previously known position of the object $O_{t-1}$ and the change mask $C_{t-1,t}$ may not be fully accurate, (9) is better to be used as a simple estimate of the new position and shape of the object. This estimate, output of the motion prediction step, is going to be further refined in the following matching step.

### B. Matching Step

*1) General Description:* The proposed matching step, shown in Fig. 5, involves the construction and processing of a BPT on the hyperspectral frame, and its use offers several advantages.

1) It drastically reduces the search space by representing the frame as a set of hierarchically nested regions. The set of candidate objects is only composed of regions that are supported by a node in the BPT.
2) It ensures to represent the frame at various scales, which is valuable as the size of the tracked object may evolve along the sequence.
3) It allows to benefit from all the efficient tree-based processing techniques for BPTs already available in the literature, and especially those related to object detection.

Therefore, the matching process aims at retrieving in the tree structure the node that represents the object and can be assimilated to a classical BPT-based object detection. In order to do so, a set of reference features for the tracked object is defined, and each candidate region has its own set of similarly defined features evaluated against the reference set. The region whose features match the reference the best is declared to correspond to the tracked object.

More specifically, let $\mathbf{\Omega}^{O_t} = \{\omega_i^{O_t}\}$ be the set of reference features for the tracked object $O_t$, where $\omega_i^{O_t}$ is an individual feature. For each frame, the matching process is done in three stages. First, each region $\mathcal{R}$ of the BPT $H^t_{\text{BPT}}$ has its similarly defined features collected in a set $\mathbf{\Omega}^{\mathcal{R}} = \{\omega_i^{\mathcal{R}}\}$. Then, each region has its features evaluated against the reference in order to find the one whose features match the reference the best. This implies the definition of a similarity measure $d_i$ for each pair of individual features $\omega_i^{\mathcal{R}}$ and $\omega_i^{O_t}$, so the overall matching distance $d(\mathcal{R}, O_t)$ can be formulated as follows:

$$d(\mathcal{R}, O_t) = \sum_{\omega_i \in \mathbf{\Omega}} \beta_i d_i\left(\omega_i^{\mathcal{R}}, \omega_i^{O_t}\right) \tag{10}$$

where the $\beta_i$ values are optional weights that can be set to stress the importance of some individual features against others. The last stage of the matching process consists of retrieving the region $\mathcal{R}^\star$ whose features are the closest from the reference ones

$$\mathcal{R}^\star = \underset{\mathcal{R} \in H^t_{\text{BPT}}}{\operatorname{argmin}}\, d(\mathcal{R}, O_t). \tag{11}$$

That region $\mathcal{R}^\star$ becomes the object representation in the current frame $\mathcal{I}_t$, i.e., $O_t \equiv \mathcal{R}^\star$, and is going to be used for the motion prediction step in the next frame $\mathcal{I}_{t+1}$.

*2) Definition of the Proposed Features:* Hyperspectral video sequences naturally provide a wealth of spectral, spatial, and temporal information, from which the set of reference features $\mathbf{\Omega}^{O_t}$ can be defined. More specifically, we propose, in the following, a spectral feature $\omega^{O_t}_{\text{spect}}$, a spatial feature $\omega^{O_t}_{\text{spat}}$, and a temporal feature $\omega^{O_t}_{\text{temp}}$ (i.e., $\mathbf{\Omega}^{O_t} = \{\omega^{O_t}_{\text{spect}}, \omega^{O_t}_{\text{spat}}, \omega^{O_t}_{\text{temp}}\}$) and their associated similarity functions to retrieve the tracked object $O_t$ in each frame $\mathcal{I}_t$.

*a) Proposed spectral feature:* The purpose of this feature is to provide some *a priori* knowledge regarding the spectral signature of the sought object. Class membership homogeneity and correlation with respect to a reference spectrum were, for instance, proposed in [47] and [48] for road and building extraction in hyperspectral urban scenes. However, those two features require the availability of some reference spectrum and are thus not applicable in a fully blind scenario. Therefore, assuming a moderate motion in the sequence (and thus an object overlaying similar backgrounds from one frame to the other), we propose the mean spectrum $\bar{\mathbf{o}}_{t-1}$ of the detected object $O_{t-1}$ in the previous frame as the spectral feature $\omega^{O_t}_{\text{spect}}$ of $\mathbf{\Omega}^{O_t}$.

The proposed spectral feature distance is derived from the two-sample Hotelling's T-square statistic, classically used to

test the equality of the mean vectors of two populations [57]. More specifically, let $\hat{\Sigma}_{\mathcal{R}}$ and $\hat{\Sigma}_{O_{t-1}}$ be the respective sample covariance matrices of $\mathcal{R}$ and $O_{t-1}$. The Hotelling's T-square statistic between $\mathcal{R}$ and $O_{t-1}$ has the following expression:

$$T^2(\mathcal{R}, O_{t-1}) = \frac{|\mathcal{R}||O_{t-1}|}{|\mathcal{R}|+|O_{t-1}|}(\bar{\mathbf{r}}-\bar{\mathbf{o}}_{t-1})^T \Sigma_{\text{pool}}^{-1}(\bar{\mathbf{r}}-\bar{\mathbf{o}}_{t-1}) \quad (12)$$

with $\Sigma_{\text{pool}}$ being the pooled covariance matrix

$$\Sigma_{\text{pool}} = \frac{(|\mathcal{R}|-1)\hat{\Sigma}_{\mathcal{R}} + (|O_{t-1}|-1)\hat{\Sigma}_{O_{t-1}}}{|\mathcal{R}|+|O_{t-1}|-2}. \quad (13)$$

The normalized $T^2$ statistic

$$F(\mathcal{R}, O_{t-1}) = \frac{|\mathcal{R}|+|O_{t-1}|-N_\lambda-1}{N_\lambda(|\mathcal{R}|+|O_{t-1}|-2)} T^2(\mathcal{R}, O_{t-1}) \quad (14)$$

follows an F-distribution with $N_\lambda$ numerator degrees of freedom and $|\mathcal{R}| + |O_{t-1}| - N_\lambda - 1$ denominator degrees of freedom. The final spectral feature distance is finally obtained by normalizing the F-statistic between $\mathcal{R}$ and $O_{t-1}$ with the F-statistic between $\hat{O}_t$ and $O_{t-1}$

$$d_{\text{spect}}\big(\omega_{\text{spect}}^{\mathcal{R}}, \omega_{\text{spect}}^{O_t}\big) = \frac{F(\mathcal{R}, O_{t-1})}{F(\hat{O}_t, O_{t-1})} \quad (15)$$

as it is desirable for a good candidate region to be closer from the target spectrum $\bar{\mathbf{o}}_{t-1}$ than the estimate region $\hat{O}_t$ (thus a distance value being less than 1).

*b) Proposed spatial feature:* The goal of this feature is to specify some descriptors related to the shape of the object to detect. For basic geometric shapes, a suited spatial feature can merely be defined as a reference template shape (for instance, an elliptical template shape in the case of face detection, as proposed in [55]), or the compactness and elongation of the smallest fitting bounding box (as used in [47] for road and building extraction). In this paper, the shape of the object to detect is, however, supposed to be unknown and is likely to evolve over time. Therefore, we propose here to take advantage of the estimate region $\hat{O}_t$ that should provide a good initial guess for the spatial position and shape of the object. Thus, the proposed spatial feature $\omega_{\text{spat}}^{O_t}$ is simply the output $\hat{O}_t$ of the motion estimation step.

Consequently, the proposed spatial feature distance evaluates how similar from $\hat{O}_t$ is any candidate region $\mathcal{R}$

$$d_{\text{spat}}\big(\omega_{\text{spat}}^{\mathcal{R}}, \omega_{\text{spat}}^{O_t}\big) = \frac{|\mathcal{R}\Delta\hat{O}_t|}{|\mathcal{R}|} \quad (16)$$

where $|\mathcal{R}\Delta\hat{O}_t|$ is the number of pixels in the symmetric difference between $\mathcal{R}$ and $\hat{O}_t$ (pixels either in $\mathcal{R}$ or in $\hat{O}_t$, but not in both). It corresponds to the percentage of error pixels with respect to the candidate region size. Good candidate regions are, therefore, expected to have a spatial distance less than 1.

*c) Proposed temporal feature:* The objective of this feature is to incorporate some information related to the motion in the sequence of the object to detect. Here, we define the temporal feature as a confidence area where the tracked object is expected to be found with a very high probability. This confidence area is derived from the estimate position $\hat{O}_t$ by a morphological dilation with a structuring element (SE), $\delta_{\text{SE}}(\hat{O}_t)$, and the percentage of inclusion of every candidate region $\mathcal{R}$ in the confidence area is evaluated

$$\mathcal{R}_{\%\hat{O}_t} = \frac{|\mathcal{R}\cap\delta_{\text{SE}}(\hat{O}_t)|}{|\mathcal{R}|}. \quad (17)$$

The proposed temporal feature distance is a hard thresholding of this percentage of inclusion

$$d_{\text{temp}}\big(\omega_{\text{temp}}^{\mathcal{R}}, \omega_{\text{temp}}^{O_t}\big) = \begin{cases} 0 & \text{if } \mathcal{R}_{\%\hat{O}_t} \geq \tau \\ +\infty & \text{otherwise.} \end{cases} \quad (18)$$

This distance allows to consider only regions in the BPT that have at least $\tau\%$ of their pixels in the confidence area as possible candidates, dismissing all other regions. It is possible to be more or less selective by varying the SE and the threshold $\tau$.

*d) Overall matching distance:* For each region $\mathcal{R}$ of the BPT, the overall matching distance to the set of reference features is obtained by adding the three feature distances

$$d(\mathcal{R}, O_t) = d_{\text{spect}}\big(\omega_{\text{spect}}^{\mathcal{R}}, \omega_{\text{spect}}^{O_t}\big) + d_{\text{spat}}\big(\omega_{\text{spat}}^{\mathcal{R}}, \omega_{\text{spat}}^{O_t}\big) + d_{\text{temp}}\big(\omega_{\text{temp}}^{\mathcal{R}}, \omega_{\text{temp}}^{O_t}\big). \quad (19)$$

Note that this is equivalent to only considering the spectral and spatial feature distances for regions whose temporal feature distance is equal to 0. The optional weights of (10), which trade off the influence of a distance with respect to the other, are all set to 1, since the spectral and spatial feature distances were designed to have a similar range of values for good candidate regions. The optimal region of the BPT is finally the one achieving the smallest global distance to the set of reference features, and is retrieved with an exhaustive search.

### C. Initialization of the Object Tracking Procedure

The motion prediction and matching steps developed earlier in Sections IV-A and IV-B are sequentially addressed in order to track the object of interest. In order to trigger the tracking process, however, an initial detection needs to be performed to identify the object to track. This is the matter of the initialization phase. In the following, we assume that the frame in which the object starts moving (and thus where the object tracking process must be launched) is unknown. However, we presume that a few (at least two) still frames $\mathcal{I}_{t_1}, \ldots, \mathcal{I}_{t_{N_s}}$ are available prior to the object being in motion. This assumption seems, however, reasonable in a context of surveillance, where nothing is moving in most of the frames of the sequence.

To determine the point at which some motion appears in the sequence, the change detection procedure described in Section IV-A1 is applied for each new incoming frame $\mathcal{I}_t$, and a change mask $C_{t-1,t}$ is generated. If this change mask remains empty (all pixels of the frame difference $\mathcal{I}_{\Delta t}$ have been found not to feature any change), then it is stated that no motion is occurring in the sequence yet. Conversely, if at least one pixel was found to be changing between $t-1$ and $t$ (appearing as a 1 in the change mask $C_{t-1,t}$), then it is assumed that the object has started moving, and the object tracking process is triggered. Deriving the change mask requires the knowledge of the covariance matrix $\Sigma$ of the

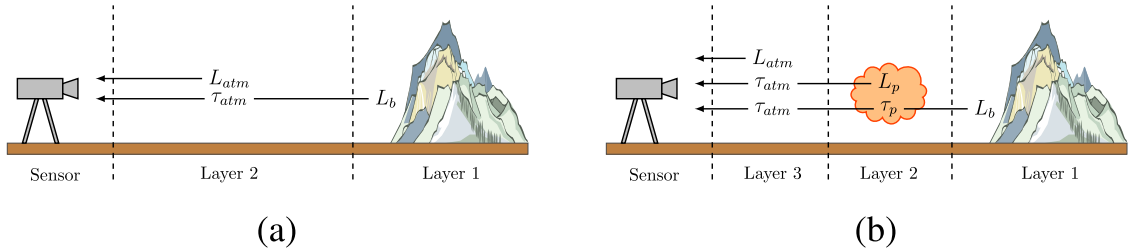



Fig. 6. Illustration of (a) two layers model and (b) three layers model for the at-sensor radiance.

noise $\boldsymbol{\eta}^{\Delta t}$, which is unknown in practice. However, due to the assumptions that several (say $N_s$) still frames are available, one can compute $N_s - 1$ frame differences featuring only instances of $\boldsymbol{\eta}^{\Delta t}$, which can be used to derive the sample covariance matrix $\hat{\Sigma}$ of $\Sigma$. In addition, due to the (ideally) large number of samples on which $\hat{\Sigma}$ is computed, it seems fair to state that the sample covariance $\hat{\Sigma}$ is a very good approximation of $\Sigma$, which is the reason why this latter was used in the derivation of the change mask instead of $\hat{\Sigma}$.

The tracking process being triggered once at least one pixel has been stated to change between $t-1$ and $t$, the change detection test needs not to suffer from any false alarm at all, as the tracking would be engaged too early otherwise. Therefore, instead of testing each pixel of the frame difference $\mathbf{x}_i^{\Delta t}$ individually in the change detection process, it is tested along with its $5 \times 5$ neighbors. Therefore, the pixel will be marked as changing if and only if some change is also happening in its direct neighborhood, decreasing the risk of false alarms with respect to the individual testing case (or conversely, guaranteeing that all pixels marked as change are really changing). In the derivation of the GLRT, this translates as the MLE $\hat{\boldsymbol{\mu}}_i^{\Delta t}$ being equal to $1/S \sum_{\mathbf{x}_i^{\Delta t} \in S} \mathbf{x}_i^{\Delta t}$ instead of simply $\mathbf{x}_i^{\Delta t}$.

The other main consequence of this no false alarm policy is related to the setting of the $\gamma_{\text{GLRT}}$ threshold. Usually, this threshold is derived using the distribution of the GLRT under hypothesis $\mathbf{H}_0$ in order to achieve a given probability of false alarm $p_{\text{FA}}$. In particular, it requires to invert the cumulative distribution function of the GLRT under $\mathbf{H}_0$. In our case, however, we wish to have $p_{\text{FA}} = 0$, and thus needs to invert the distribution of the GLRT under $\mathcal{H}_1$ to achieve a given probability of detection $p_D$. It is known from [27] that the GLRT $\Lambda(\mathbf{x}_i^{\Delta t})$ follows a noncentral $\chi^2_{N_\lambda,\phi}$ distribution under $\mathbf{H}_1$, with $N_\lambda$ degrees of freedom being the number of spectral bands in the frame, and $\phi = \Lambda(\mathbf{x}_i^{\Delta t})$ being the noncentrality parameter. The threshold $\gamma_{\text{GLRT}}$ and the probability of detection $p_D$ are linked with the following relationship:

$$\gamma_{\text{GLRT}} = (\mathbf{X}^2_{N_\lambda,\phi})^{-1}(1 - p_D) \tag{20}$$

where $\mathbf{X}^2_{N_\lambda,\phi}$ is the cumulative distribution function of $\chi^2_{N_\lambda,\phi}$. Therefore, using (20), one can set the value of $\gamma_{\text{GLRT}}$ to achieve a given probability of detection $p_D$. In practice, however, the cumulative distribution function $\mathbf{X}^2_{N,\phi}$ has no closed form expression, and is computationally slow to invert. However, it was shown [58, pp. 22–24] that if the random variable $\mathbf{Y}$ follows a noncentral $\chi^2_{N_\lambda,\phi}$ distribution with $N_\lambda$ degrees of freedom and noncentrality parameter $\phi$, then:

$$\frac{\mathbf{Y} - (N_\lambda + \phi)}{\sqrt{2(N_\lambda + 2\phi)}} \xrightarrow{p} \mathcal{N}(0,1) \quad \text{when } N_\lambda \to +\infty \text{ or } \phi \to +\infty \tag{21}$$

where $\xrightarrow{p}$ denotes the convergence in probability. Here, $N_\lambda$ being equal to the number of spectral channel in the hyperspectral frame (which is typically several hundreds), it can be considered high enough for the approximation (21) to hold. Therefore, the value of $\gamma_{\text{GLRT}}$ can be derived by inverting the cumulative distribution function of a standard normal distribution provided that the proper shift and scaling described in (21) is applied instead of this of the noncentral $\chi^2$ distribution as prescribed in (20) to achieve a given probability of detection $p_D$. Finally, note that $\gamma_{\text{GLRT}}$ varies from one pixel to the other, as it is linked to the value of $\phi = \Lambda(\mathbf{x}_i^{\Delta t})$.

In Section V, we apply the proposed methodology to the tracking of chemical gas plumes in hyperspectral video sequences.

## V. Experiments

Hyperspectral sensors measure the radiance, amount of electromagnetic energy emitted by the scene. The physical nature of this energy depends on the scanned spectral range. In the LWIR range, this emitted energy is governed by the radiative transfer theory. Therefore, in order to understand the physical nature of the used data sets, it is worth briefly describing the radiative transfer theory.

### A. Radiative Transfer Theory

The radiance of a material is defined as the amount of electromagnetic radiation, which passes through or is emitted from a particular unit area per solid angle, and is expressed in $\text{Wsr}^{-1}\text{m}^{-2}$. Consequently, the radiance of a material is expressed over the whole electromagnetic spectrum. However, the sensor cannot capture the radiance over the whole spectrum, but rather at some particular wavelengths. Thus, the *spectral* radiance is defined as the radiance of a material at a given wavelength, and is expressed in terms of $\text{Wsr}^{-1}\text{m}^{-3}$. When no plume is present in the scene, the spectral radiance $L(\lambda)$ (which is a function of the wavelength $\lambda$) reaching the sensor can be expressed according to the *two layers model* displayed in Fig. 6(a)

$$L(\lambda) = L_{\text{atm}}(\lambda) + \tau_{\text{atm}}(\lambda) L_b(\lambda) \tag{22}$$

where $L_{\text{atm}}(\lambda)$ and $L_b(\lambda)$ denote the atmosphere and background spectral radiances, respectively, and $\tau_{\text{atm}}(\lambda)$ is the

atmosphere transmittance, defined as the ratio of the quantity of light leaving the medium with respect to the quantity of light entering the medium. When the plume is present in the scene, (22) transforms into the so-called *three layers model*, shown in Fig. 6(b) and expressed as

$$L(\lambda) = L_{\text{atm}}(\lambda) + \tau_{\text{atm}}(\lambda)L_p(\lambda) + \tau_{\text{atm}}(\lambda)\tau_p(\lambda)L_b(\lambda) \quad (23)$$

where $L_p(\lambda)$ and $\tau_p(\lambda)$ stand for the plume spectral radiance and transmittance, respectively. It is often assumed that the contribution of the atmospheric radiance with respect to the plume and background radiances can be neglected. Also, the atmospheric transmittance can be approximated to 1 due to the short distance between the release point and the sensor, meaning that the atmosphere allows all the signal to pass through unaffected [21]. Under those assumptions, the at-sensor spectral radiance can be written

$$L(\lambda) = L_p(\lambda) + \tau_p(\lambda)L_b(\lambda). \quad (24)$$

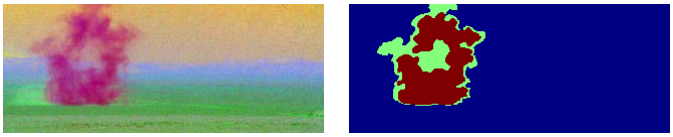
Spectral radiance is often converted into emissivity as the latter plays in the LWIR domain the same role as the reflectance does in the visible domain (in the sense that each material is uniquely defined by its spectral emissivity signature). The emissivity of a material, $\epsilon(\lambda)$, is the ratio of the energy radiated by this particular material to the energy radiated by a blackbody at the same temperature. While the former is the quantity acquired by the sensor, the latter is described using Planck's blackbody law

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{kT\lambda}\right) - 1} \quad (25)$$

where $T$ is the temperature of the surface in Kelvin, $h$ is Planck's constant, $c$ is the speed of light, and $k$ is Boltzmann's constant. The emissivity of each pixel is retrieved from the radiance through the use of some temperature emissivity separation (TES) algorithm [59], which operates in two steps: first, the apparent temperature of each pixel is estimated by inverting Planck's law. Each pixel radiance is then divided by its estimated blackbody curve to obtain its apparent emissivity.

### B. Data Sets

The data sets used to validate the proposed methodology were provided by the John Hopkins Applied Physics Laboratory (JHAPL). They were acquired in 2006 at the Dugway Proving Ground in Utah (USA) with a Field-Portable Imaging Radiometric Spectrometer Technology (FIRST) [60] LWIR sensor, located 2.82 km away from the release point. This sensor typically produces video sequences at a frame rate of 0.2 Hz, where each frame is an HSI composed of $128 \times 320$ pixels in the spatial domain and 129 spectral channels, spanning 7.81 to 11.97 $\mu$m in wavelength. The following experiments are conducted on two sequences featuring the explosive release of an acetic acid canister, denoted *aa12_Victory* and *aa13_Victory*. Despites each sequence initially contained hundreds of hyperspectral frames, only the small portion of the sequence featuring the plume release and diffusion was retained. The resulting sequences are both composed of 30 frames, and the gas release occurs at the 11th frame. Note that for the following experiments, it is assumed that only the first two frames of each sequence are known not to contain the gas plume, which is a reasonable assumption, since it should be easy in practical scenarios to obtain some reference, plume-free frames of some area of interest under monitoring. The radiance for each frame was converted into emissivity using the TES algorithm [21] provided by JHAPL along with the video sequences.



(a) (b)

Fig. 7. False color RGB composition of (a) 20th frame of aa12_Victory and (b) its corresponding ground-truth data.

Reference data were created for the two data sets in order to conduct some quantitative performance evaluations. To generate the ground-truth map for a given frame, a principal component analysis (PCA) was performed, and the three principal components (PCs) showing the highest contrast between the plume and the background were selected, creating a false color RGB composition of the scene. The PC selection was conducted by visually analyzing the 20 PCs (being most likely to feature some contrast between the gas plume and the background). The identity of the selected PCs varied between the two data sets, and even between two consecutive frames of a single data set, making this selection automated impossible. Two classes were carefully delineated from the RGB composition, the first corresponding to the strongly concentrated section of the plume (typically the central part), and the second to more diffused components. Fig. 7 displays an example of ground-truth data for the aa12_Victory sequence (more examples are further given in Figs. 8 and 10). Due to the inherent subjectivity of the ground-truth manual delineation task, it is advocated not to consider the created ground truth as a perfect gold standard, but rather as a support for the quantitative comparison of the performance of several methods.

### C. Experimental Setup

Gas plume tracking is a challenging task, since the gas plume is a nonrigid object with no real boundary and whose shape is evolving unpredictably. Moreover, the gas plume is an optically thin object whose concentration changes with time, as a natural consequence of the diffusion phenomenon, making it more and more difficult to detect, hence the need of appropriately tuning the proposed object tracking methodology. Here, we describe the experimental setup used in the following to conduct the experiments on the aa12_Victory and aa13_Victory sequences.

As it is assumed that the first two frames of both sequences only feature the background, the noise covariance matrix $\Sigma$ is computed from the difference between those two frames. From frame #3 on, the change detection procedure is conducted to detect the release of the plume, as described in Section IV-C, with a probability of detection $p_D = 0.99$. Due to the high

value of $p_D$, and in order not to suffer from the presence of any false alarm (the size of the plume immediately after the release being unknown, the tracking step is engaged if at least one pixel is flagged as changing), the change detection procedure is actually conducted on the PCA transform of the frame difference $\mathcal{I}_{\Delta t}$. As a matter of fact, the PCA transformation decorrelates the data, hence improving its separability with respect to the highly correlated hyperspectral data, and we noted that the performances of the statistical testing were greatly enhanced, performing the testing on the PCs instead, resulting in no false alarms at all despites the high value of $p_D$. Note that the use of a PCA does not modify the null hypothesis $\mathbf{H}_0$ of (5) while replacing $\boldsymbol{\mu}_i^{\Delta t}$ by $\mathbf{W}^{-1}\boldsymbol{\mu}_i^{\Delta t}$ in the alternate hypothesis $\mathbf{H}_1$, with $\mathbf{W}$ being the eigenvector matrix of the frame difference covariance matrix.

The motion prediction step is conducted, as presented in Section IV-A. First, the change mask $C_{t-1,t}$ is computed by thresholding the result of the change detection statistical test, so it achieves a probability of detection of 95%. Then, the estimate position of the plume $\hat{O}_t$ is obtained from $C_{t-1,t}$ and the previous position of the plume $O_{t-1}$ using (9). Note that, similar to what is done to detect the release of the plume, the statistical testing is performed on the PCA of the frame difference.

The matching step, as described in Section IV-B, involves the construction of a BPT for the current frame, which is itself governed by the definition of an initial segmentation map, a region model, and a merging criterion (see Section III-C1). For our application, the initial segmentation map was obtained using the hyperspectral watershed algorithm [61]. This technique is known to produce severe oversegmentation, but this is not a problem, since meaningful regions will be created later on in the BPT construction from the merging of initial regions. The efficiency of the hyperspectral watershed as an initial segmentation map for chemical plume tracking was already demonstrated in [17]. The region model and its associated merging criterion were, on the other hand, defined as the mean spectrum (its ability to well separate the plume from the background was shown in [17]), and the so-called spectral information divergence [49]. Following the construction of the BPT, the object detection procedure is conducted to complete the matching step. For that purpose, the spectral, spatial, and temporal features described in Section IV-B2 are retrieved for each region of the BPT representation, and the one that is the most likely to correspond to the plume instance is retrieved by evaluating the overall matching distance defined in (19) over all regions. Note that the SE involved in the computation of the temporal feature (17) is defined as a $9\times9$ square, and the threshold $\tau$ of its associated feature distance (18) is set to 80%.

To assess the performance of the proposed BPT-based tracking method, we compare the obtained results against three state-of-the-art methods.

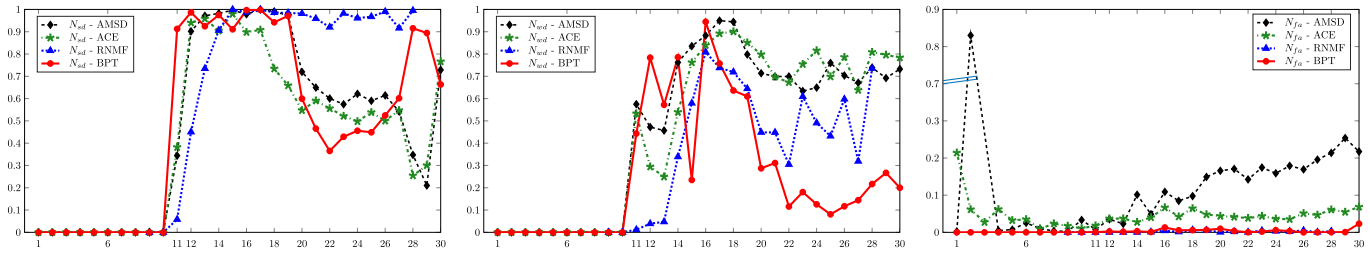
*a) AMSD detector:* It considers target pixels as anomalous with respect to a structured background model (i.e., the spectral variability of the background is described as a linear subspace), and thus performs a two-hypotheses test on the raw hyperspectral frame, therefore belonging to the class of statistical methods (see Table I). The AMSD statistic is derived from a GLRT approach and can be geometrically interpreted as the ratio between the distance of the test pixel spectrum to the subspace spanned by the sole background against the distance to the subspace engendered by the background plus the reference target [26]. The AMSD was notably investigated in [20]–[22] for the detection of chemical gas plume in hyperspectral video sequences. The implementation and the reference target emissivity spectrum provided by the JHAPL [21] were used in our comparison.

*b) ACE detector:* Unlike the AMSD, the ACE detector employs an unstructured background model (i.e., the background is simply described using its first- and second-order statistics) to perform a target detection test, hence also classifying the ACE detector as a statistical method. In such case, the GLRT approach allows for an even simpler geometrical interpretation, since the resulting ACE statistic can be viewed as the squared cosine angle between the test pixel and the reference target spectra in the whitened space. The ACE detector has also been widely used for hyperspectral chemical plume detection [24], [62], as it is known to produce globally fewer false alarms than other statistical target detection approaches.

*c) RNMF method:* This clustering-based approach [36] aims at decomposing a hyperspectral frame as low-rank matrix (accounting for the data) and a sparse matrix (representing the noise corrupting the data), the former being further factorized as the product between a cluster centroid coordinates matrix and a cluster indicator matrix, both being nonnegative. The overall decomposition is formulated as an optimization problem, solved by the alternating direction method of multipliers [63]. The RNMF results presented in Section VI are reproduced from [36], where the clustering method was applied both on aa12_Victory and aa13_Victory sequences. To account for the temporal evolution of the gas plume along the sequence, 20 frames (2 frames prior to the release of the plume and the following 18 frames featuring the diffusion for aa12_Victory, and 1 frame prior to the release and the following 19 frames with the plume for aa13_Victory) were first stacked together prior to the clustering. This method additionally requires the number of final clusters as an input, which was set to 4 (being the sky, the foreground, the mountain, and the plume) in [36].

### D. Assessing the Tracking Quality

Assessing the performances of a tracking algorithm is a well-known challenge in computer vision. Several studies have addressed the problem when ground-truth data are available. In [64], a metric is introduced to compare the trajectory of the tracked object with a reference trajectory accounting for ground truth. In [65] and [66], frame-based surveillance metrics relying on the number of true and false positives are developed. These metrics are notably used to evaluate the consistency of the tracker across the whole sequence (where a true positive is claimed when the object is present in a given frame and correctly detected by the tracker). Object-based performance metrics, such as spatial overlap between ground-truth object and tracked object and Euclidean distance between




(a) (b) (c)

Fig. 8. Percentage of (a) strong detection, (b) weak detection, and (c) false alarms for the aa12_Victory data set. Black, green, and blue dashed lines correspond to the state-of-the-art AMSD, ACE, and RNMF, respectively, while the plain red line corresponds to the proposed BPT-based method. For the false alarm plots, the $y$-axis has been broken for an easier visualization.

TABLE II

DETAILED COMPUTATION TIME IN SECONDS (AVERAGE ± STANDARD DEVIATION) OF THE DIFFERENT STEPS INVOLVED IN THE PROPOSED BPT-BASED TRACKING METHOD, FOR aa12_VICTORY SEQUENCE

| | Motion prediction step | Matching step | | Total time |
|---|---|---|---|---|
| | | BPT construction | Object detection | |
| Prior to the release (9 frames) | 2.64 ± 0.02 | ✗ | ✗ | 2.64 ± 0.02 |
| After the release (20 frames) | 2.68 ± 0.15 | 13.31 ± 0.96 | 8.07 ± 0.72 | 24.07 ± 1.71 |

their respective centroids, are considered in [67]. In particular, we propose to use this notion of overlapping between ground truth and corresponding object in order to derive three metrics reflecting the performance and accuracy of the tracking. As can be seen in Fig. 7, the ground-truth map for each frame is composed of three different regions:

1) regions where the plume is strongly concentrated, denoted $\mathrm{GT}_{\mathrm{strong}}$ in the following;
2) regions where the plume is weakly concentrated, denoted $\mathrm{GT}_{\mathrm{weak}}$;
3) all remaining regions of the image, not containing any gas and denoted $\mathrm{GT}_{\emptyset}$.

We define the percentage of strong detections $N_{\mathrm{sd}}$ and of weak detections $N_{\mathrm{wd}}$ as the percentage of strongly and weakly concentrated ground-truth plume areas included in $O_t$, respectively

$$N_{\mathrm{sd}} = \frac{|O_t \cap \mathrm{GT}_{\mathrm{strong}}|}{|\mathrm{GT}_{\mathrm{strong}}|} \tag{26}$$

and

$$N_{\mathrm{wd}} = \frac{|O_t \cap \mathrm{GT}_{\mathrm{weak}}|}{|\mathrm{GT}_{\mathrm{weak}}|}. \tag{27}$$

Similarly, the percentage of false alarms $N_{\mathrm{fa}}$ is defined as the percentage of $\mathrm{GT}_{\emptyset}$ area that is wrongly comprised in $O_t$

$$N_{\mathrm{fa}} = \frac{|O_t \cap \mathrm{GT}_{\emptyset}|}{|\mathrm{GT}_{\emptyset}|}. \tag{28}$$

High values of $N_{\mathrm{sd}}$ and $N_{\mathrm{wd}}$ (theoretically, 1) along with a low value of $N_{\mathrm{fa}}$ (theoretically, 0) indicate a good detection of the plume for a given frame. The temporal performance of the tracking can be assessed by evaluating the consistency of $N_{\mathrm{sd}}$ and $N_{\mathrm{wd}}$ to remain high and of $N_{\mathrm{fa}}$ to stay low across the whole sequence.

## VI. RESULTS AND DISCUSSION

This section presents the obtained results for the proposed BPT-based approach, along with a critical comparison of its performance with respect to the state-of-the-art AMSD, ACE, and RNMF methods.

### A. aa12_Victory Sequence

*1) Results:* Quantitative results for the aa12_Victory sequence are presented in Fig. 8 and Table II. The former is composed of three plots representing the evolution of the percentages of strong detections, weak detections, and false alarms across the sequence, where the $x$-axis corresponds to the frame number and the $y$-axis is the percentage. Each plot features the $N_{\mathrm{sd}}$, $N_{\mathrm{wd}}$, and $N_{\mathrm{fa}}$ values for the proposed method and the three state-of-the-art approaches (for the RNMF method, only 20 frames among the 30 that constitute the sequence are available, hence truncated plots). Table II exhibits the detailed computation time of the proposed method. Finally, Fig. 9 displays some visual results, and is composed of ten rows and six columns. Each row corresponds to a particular frame of the sequence. Fig. 9(a)–(d) presents the binary mask of the detected plume for the state-of-the-art AMSD, ACE, RNMF, and the proposed BPT-based methods, respectively. Fig. 9(e) and (f) shows the RGB representation of the hyperspectral frame and the corresponding ground-truth map, with the same color code as described in Fig. 7. While it is impossible to show all frames by lack of room, only frames #11, #12, #14, #16, #18, #20, #22, #24, #26, and #28 are represented. We remind that the release of the plume occurs in frame #11 in both sequences. Besides, we recall that the false color RGB frames result from a strong preprocessing stage, including the visual selection of the three PCs showing the greatest contrast between the plume

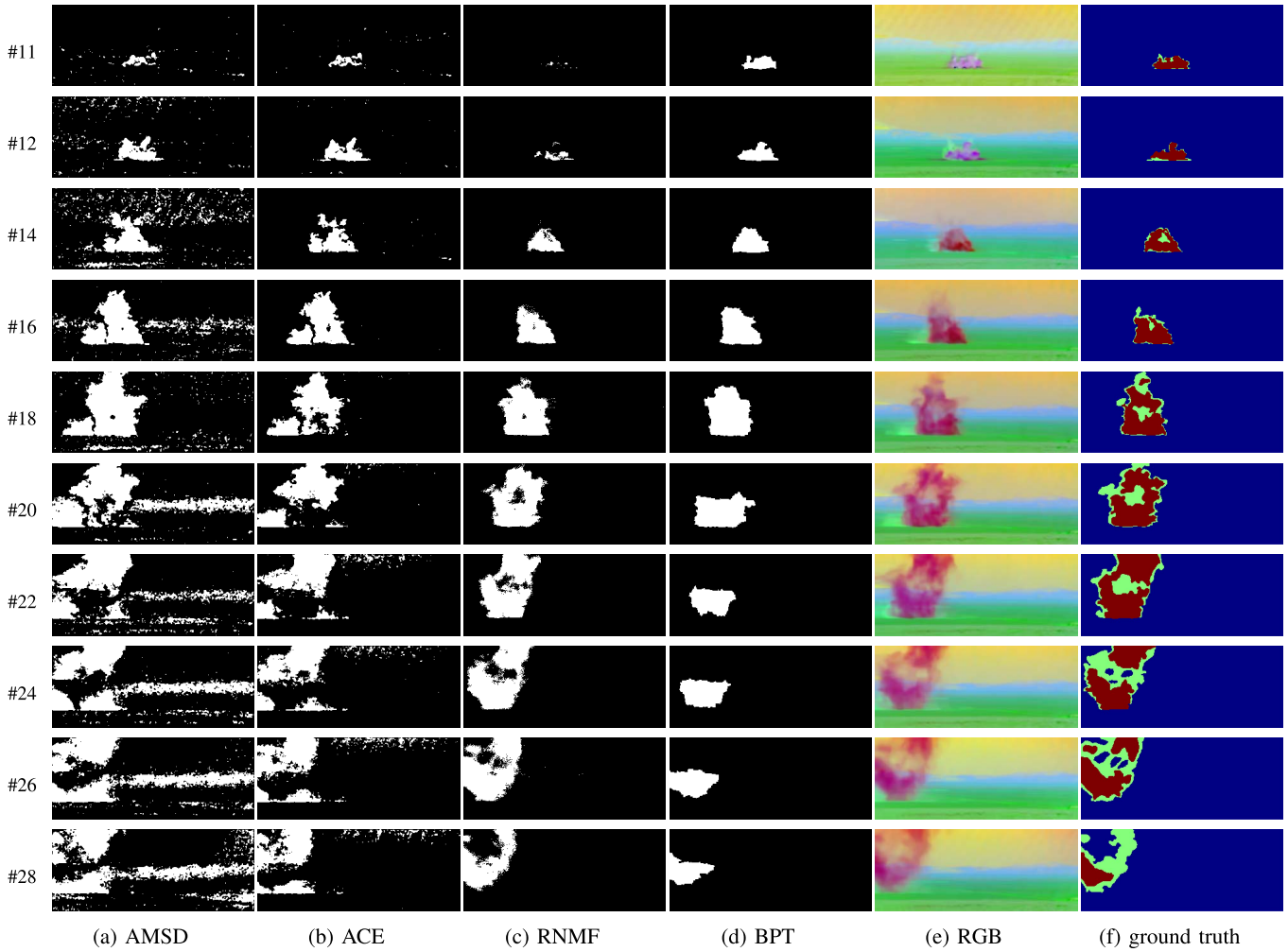



Fig. 9. Visual results of the tracking for ten frames of aa12_Victory sequence. (a)–(d) Binary mask of the detected plume for the AMSD, ACE, RNMF, and the proposed BPT-based methods, respectively. (e) and (f) False color RGB representation of the hyperspectral frame and its ground truth, respectively.

and the background, as well as local contrast enhancement and adjustment to ensure the color of the various elements composing the scene (the plume, the foreground, the mountain, and the sky) to remain relatively similar across the sequence. Thus, the false color RGB frame does not reflect the difficulty of separating the plume from the background.

*2) Discussion:* We now analyze in details the three plots displayed in Fig. 8.

*a) About the strong detection plot:* The evolution of the percentage of strong detection $N_{\text{sd}}$ is shown in Fig. 8(a), and several observations arise from its analysis.

The evolution of the $N_{\text{sd}}$ curves for the AMSD, the ACE, and proposed BPT-based methods features a similar trend: remaining over 90% of strong detections until frame #19, they both suffer a drop from frame #20 on (and even earlier for the ACE), before rising again at the end of the sequence. At this point of the sequence, the plume equally overlays the foreground and the sky, as shown in Figs. 7(a) and 9, and all three methods suffer from this split. For both statistical approaches, the bottom half of the plume is poorly detected, and the small cloud of dust (triggered by the explosive release and appearing as bright green on the left-hand side of the red plume) is erroneously identified as the gas plume. For the proposed BPT-based method, on the other hand, this flaw comes from the *one object, one node* paradigm that was adopted for the hierarchical object detection process. As a matter of fact, both halves of the plume (the one overlaying the foreground and the other covering the sky) are supported by nodes in different branches in the BPT decomposition of the frames. Only the bottom half subsequently remains correctly tracked by the proposed method, as it can be observed in Fig. 9. This notably explains why the percentage of strong detection $N_{\text{sd}}$-BPT is divided by 2 between frames #19 and #22. The percentage of strongly concentrated plume $N_{\text{sd}}$ is increasing again for the proposed BPT-based method from frame #25 on, because the top half of the plume is gradually disappearing from the frame of view, leaving only residuals that are classified as weakly concentrated in the ground-truth map.

The behavior of $N_{\text{sd}}$-RNMF is slight different as it remains consistently over 90% between frame #15 and frame #28 (the last available one for this method), but it gradually increases for the first four frames of the sequence, as if the clustering method was unable to efficiently differentiate between the plume signature and the background for these initial frames of

the sequence. A plausible (yet rather surprising) explanation for this observation is that the plume is "too much" concentrated in those first frames. As the RNMF clustering method is operating on the 20 frames stacked together, it may consider that those highly concentrated plume pixels do not belong to the same cluster as the more diffused plume pixels, and misclassifies them as foreground.

*b) About the weak detection plot:* The evolution of the percentage of weak detection $N_{\text{wd}}$ is displayed in Fig. 8(b). The ground-truth data labeled as weakly concentrated corresponds to areas where the plume is diffused and thin, and is likely to be mistaken with the background. In addition, there are globally less pixels labeled as weakly concentrated, hence the percentage value $N_{\text{ws}}$ being more impacted, even for errors of a few pixels. For these reasons, the $N_{\text{wd}}$ values are lower than the $N_{\text{sd}}$ ones for all four methods and the majority of the frames.

There are, however, some common patterns with the strong detection plot [Fig. 8(a)]. For the proposed BPT-based method, the $N_{\text{wd}}$ value remains relatively high for the first half of the frames and significantly drops for the second half. This is in accordance with the explanation that the plume is splitting over the sky and the foreground, and only the bottom half is tracked. The top part, overlaying the sky, quickly diffuses and is labeled as weakly concentrated in the ground truth, therefore missed by the tracker. As for the strong detections, the weak detection values $N_{\text{wd}}$ for the state-of-the-art RNMF clustering remain low for the first four frames featuring the release, confirming that the method does not assign those pixels (even if moderately concentrated) to the plume cluster.

*c) About the false alarm plot:* The percentage of false alarms $N_{\text{fa}}$ is presented in Fig. 8(c). The $y$-axis has been cut for a better visualization. The false alarms plots shall be analyzed in two time intervals: prior to and after the release of the plume.

For the first case, observing the $N_{\text{fa}}$-BPT plot, one can see by that the percentage of false alarms remains equal to 0% prior to the release occurring at the 11th frame. This implies that the proposed method does not generate any false alarm before the appearance of the plume, meaning that the tracking is triggered at the right time, which is a strongly desirable property for an object tracking algorithm operated in passive surveillance scenarios. In comparison, the AMSD and ACE consistently produce false alarms before the release of the plume. This could be problematic in a scenario where further processing relying on a precise detection of the plume release is needed.

Once the plume has appeared in the sequence, the RNMF and proposed BPT-based method only generate a tiny amount of false alarms (no more than 2% of all background pixels). While not necessarily implying that all plume pixels are correctly detected, this observation signifies that background pixels are, however, not confused with the plume. For the proposed method, in particular, it indicates that the BPT is able to properly separate the plume from the background during the construction of the BPT, suggesting that the mean spectrum, even if relatively standard, is an appropriate choice for the region model. For the two statistical methods, however, the amount of false alarms relatively significant: while being constantly around 5% for the ACE detector, it globally increases up to over 20% for the last frames of the sequence for the AMSD. Looking, in particular, at Fig. 9, one can see that the AMSD detects as plume a portion of the mountain in the background, while the ACE detector scores several false alarms in the sky region.

*d) About the computational time:* Table II presents the computational time of the various steps of the proposed BPT-based tracking method. The motion prediction step (with includes the PCA transformation, the GLRT computation for the change detection process, and the estimate of the new position) is performed for every newly acquired frame in the sequence, while the matching (composed of the BPT construction and the object detection procedure) is achieved only when some motion has been identified in the sequence. We recall that the used FIRST sensor acquires a hyperspectral frame every 5 s.

The most demanding step is the construction of the BPT, which takes approximately 13 s, despite the use of the hyperspectral watershed as an initial segmentation map to lighten the computational burden. Nevertheless, it allows to reduce the number of initial regions from $128 \times 320 = 40\,960$ to 2750 regions in average, while still producing a severe oversegmentation. This effect could be mitigated by employing another strategy (such as superpixels [68] or mean shift clustering [69]) to derive the initial segmentation, at the risk of undersegmenting the gas plume if this latter is of limited size. The object detection process takes approximately 8 s, most of which is due to the computation of the spectral feature distance (15), which requires the inversion of a pooled covariance matrix for all regions in the BPT structure. Finally, the motion prediction step takes less than 3 s per frame, which is already compatible with the real-time processing constraint inherent to a passive surveillance context. The overall computational time adds up to about 24 s per frame for the motion prediction and the matching steps, when run on a laptop computer with an Intel Core i7-4700HQ at 2.40-GHz processor and MATLAB R2015a.

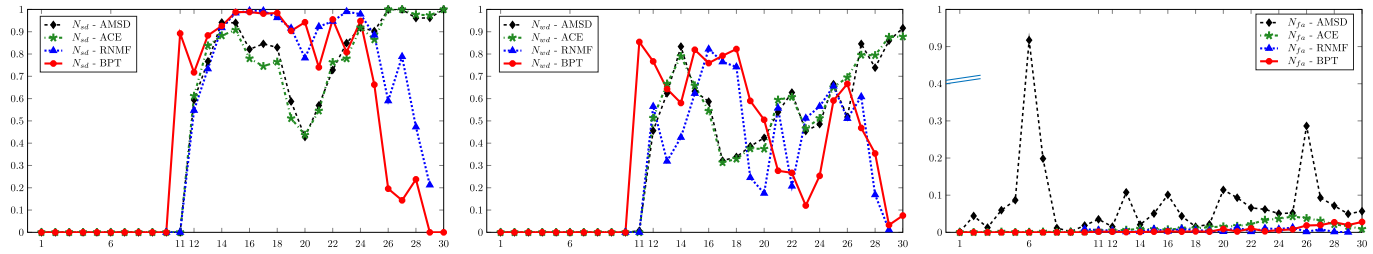
Although obtaining real-time performance is beyond the scope of this paper, the computational time could be further reduced by implementing the BPT algorithm in a more efficient language (such as C/C++). Speedup strategies, such as using the output of the motion prediction step as a mask to guide the local coarseness of the initial segmentation map, could, in addition, be employed.

### *B. aa13_Victory Sequence*

*1) Results:* Quantitative results for the aa13_Victory sequence are presented in Fig. 10 for the $N_{\text{sd}}$, $N_{\text{wd}}$, and $N_{\text{fa}}$ plots, and in Table III for the computational time. Visual results are provided in Fig. 11. Their display is strictly identical to the one of the aa12_Victory sequence.

*2) Discussion:* We now analyze in details the three plots displayed in Fig. 10.

*a) About the strong detection plot:* Fig. 10(a) displays the percentage of strong detections $N_{\text{sd}}$. The first observation that




(a) (b) (c)

Fig. 10. Percentage of (a) strong detection, (b) weak detection, and (c) false alarms for the aa13_Victory data set. Black, green, and blue dashed lines correspond to the state-of-the-art AMSD, ACE, and RNMF, respectively, while the plain red line corresponds to the proposed BPT-based method. For the false alarm plots, the $y$-axis has been broken for an easier visualization.

TABLE III

DETAILED COMPUTATION TIME IN SECONDS (AVERAGE ± STANDARD DEVIATION) OF THE DIFFERENT STEPS INVOLVED IN THE PROPOSED BPT-BASED TRACKING METHOD, FOR aa13_VICTORY SEQUENCE

| | Motion prediction step | Matching step | | Total time |
|---|---|---|---|---|
| | | BPT construction | Object detection | |
| Prior to the release (9 frames) | 2.62 ± 0.01 | ✗ | ✗ | 2.62 ± 0.01 |
| After the release (20 frames) | 2.64 ± 0.03 | 11.57 ± 0.47 | 6.79 ± 0.34 | 21.00 ± 0.78 |

can be made is that, unlike the proposed BPT-based approach, all AMSD, ACE, and RNFM methods miss the release of the plume at the 11th frame. This is confirmed by the first row of Fig. 11, which shows that all three state-of-the-art methods present some isolated false alarms, while the BPT is able to accurately isolate the appearing plume from the background. It confirms the robustness of the implemented change detection approach to perform the tracking initialization, and the subsequent object detection procedure, as the proposed method is able to lock on the plume directly as it appears in the sequence.

Contrarily to the aa12_Victory sequence, both the $N_{\text{sd}}$-BPT and $N_{\text{sd}}$-RNMF plots have a close behavior, remaining over 70% of strong detections until frame #25 before significantly dropping for the last frames of the sequence. The visual analysis of Fig. 11 suggests that the track has been completely lost for the BPT-based method. The main reason is that the gas has become so diffused as this point of the sequence that the change detection test no longer detects change between two consecutive frames, resulting in an empty change mask $C_{t-1,t}$. Consequently, the output $\hat{O}_t$ of (9) with an empty $C_{t-1,t}$ is equal to the previous object position $O_{t-1}$. In this situation, the motion prediction step becomes trapped in one part of the image (being the background mountain in this case) and the track is lost. A possible solution to overcome this issue would be to regularly reset the motion predictor by estimating the change between the current frame and a frame prior to the release instead of using two consecutive frames and a more limited motion.

The AMSD and ACE approaches produce similar results, with a drop at the middle of the sequence before increasing again, overall outperforming both the RNMF and BPT-based methods.

*b) About the weak detection plot:* The evolution of the percentage of weak detection $N_{\text{wd}}$ is shown in Fig. 10(b).

For all compared methods, the obtained $N_{\text{wd}}$ plot is more chaotic than the $N_{\text{sd}}$ ones, as the curves oscillate a lot and do not show much consistency from one frame to the other. One can, however, remark a peak centered around the 25th and 26th frames for the proposed BPT-based method. As a matter of fact, the method has just lost the track of the object and is settling in a particular region of the image instead. In frame #26, this region coincides with the wake of the plume, explaining the spike in the curve.

*c) About the false alarm plot:* Fig. 10(c) exhibits the percentage of false alarms $N_{\text{fa}}$. The $y$-axis has been cut for a better visualization. Similar to the aa12_Victory data set, the percentage of false alarms should be analyzed in two parts, being prior to and after the release of the plume in the sequence.

For the first ten frames, the observations are almost similar to those made for the aa12_Victory sequence. The AMSD generates false alarms for all frames prior to the release, contrarily to the ACE, which scores this time almost no false alarms at all. The proposed BPT-based method remains at 0% all along, validating the fact that the tracking algorithm is not triggered prior to the release.

For all remaining 20 frames, the RNMF and the proposed method do not generate a lot of false alarms, confirming that both methods are able not to confuse the background pixels with the plume. The loss of track for the BPT-based method is illustrated by the increase in percentage of false alarms from frame #25 on, remaining nevertheless moderate (no more than 5%). While the AMSD still confuses part of the background mountain with the plume, the ACE detector this time produces less false alarms. However, and similar to the aa12_Victory data set, both statistical methods tend mistake the bright green cloud of dust, located directly below the chemical plume.

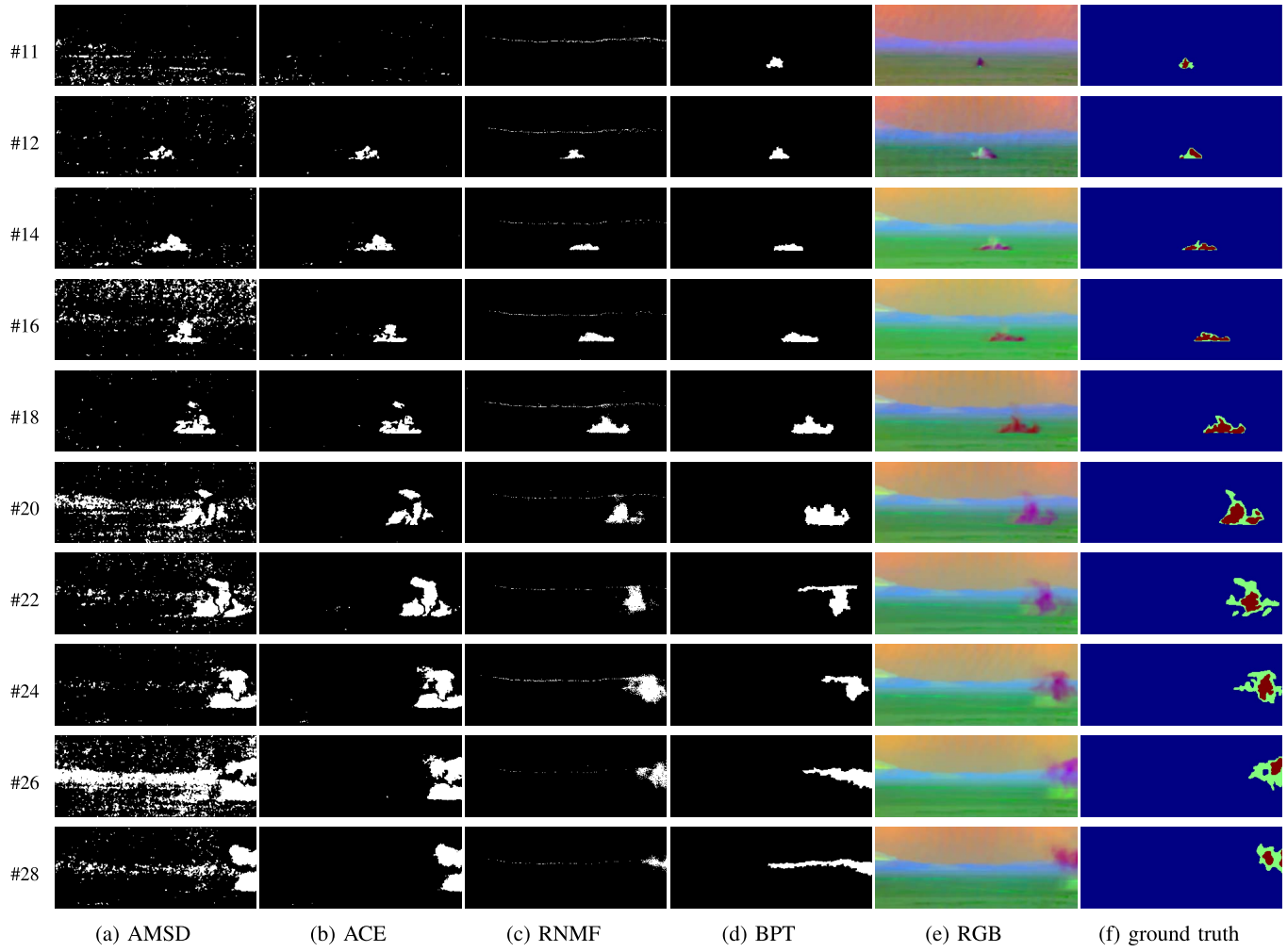



Fig. 11. Visual results of the tracking for ten frames of aa13_Victory sequence. (a)–(d) Binary mask of the detected plume for the AMSD, ACE, RNMF, and proposed BPT-based methods, respectively. (e) and (f) RGB representation of the hyperspectral frame and its ground truth, respectively.

*d) About the computational time:* The computational time for the aa13_Victory sequence is presented in Table III. The running time is of similar order than for aa12_Victory sequence, with the construction of the BPT being the slowest step, followed by the object detection process and the motion prediction step.

### C. Critical Analysis and Comparison of Obtained Results

We finally conduct a critical comparison and list the advantages and drawbacks of the proposed method and all three compared state-of-the-art approaches.

*a) Statistical AMSD and ACE detectors:* Thanks to their simplicity of use and their computational efficiency (less than 1s per frame for both of them), AMSD and ACE are popular target detectors for the detection and tracking of chemical plumes. Their main disadvantage is that they both require the knowledge of the target spectrum to be operated. While this latter is, in practice, estimated with spectral libraries, it nevertheless implies that such libraries are well constituted, with the sought samples matching the used sensor characteristics in terms of number of bands and spectral frequencies. Moreover, sample measurements are often conducted in-lab, with next-to-perfect conditions that do not necessarily conform the real data collecting situation. Their use of the temporal information contained in the sequence is also limited, as they commonly take only advantage of pixels that were labeled as "background" with high confidence in $\mathcal{I}_{t-1}$ to update some parameters (the background endmembers in the AMSD case and the whitening matrix for the ACE detector) for frame $\mathcal{I}_t$. There is, however, no real temporal consistency constraint for the pixels labeled as "plume," which leads both detectors to suffer from many false alarms (even if this effect is less severe for the ACE detector). This could, for instance, be detrimental when a precise detection of the plume release is required for further (and possibly uncorrelated) operations.

*b) Clustering-based RNMF method:* The results outputted by the RNMF clustering method are visually the closest to the ground-truth data. Its formulation, as presented in [36], however, suffers from two major drawbacks that would limit its potential use for real chemical plume tracking scenarios. First, it is necessary to stack several consecutive frames together (including some empty background frames), in order to enhance the capability of the algorithm to separate the plume from the background. This point is not only problematic to perform frame-by-frame tracking, since several consecutive

frames must be grouped together prior to the clustering step, but it also implies that the release point of the plume in the sequence is known beforehand. Second, the clustering operation requires to fix *a priori* the number of clusters to employ, which can be difficult to set optimally for complex background scenes.

*c) BPT-based object tracking approach:* The obtained results for aa12_Victory sequence put in evidence a limit of the *one object, one node* paradigm that was adopted for the object detection process. This latter indeed assumes that the whole tracked object is captured as one region in the hierarchical structure, and this may not always be true, as it was demonstrated on the aa12_Victory sequence by the plume splitting in two different branches of the BPT. The extension to a *one object, multiple nodes* paradigm, required to accommodate such case, is, however, not straightforward as there is also a risk of confusion with a *multiple objects, one node each* situation. Nevertheless, handling both possibilities appears as the next improvement of the proposed methodology, and it is part of our future research avenues. Note that the challenge for that purpose does not concern the adaptation of the object detection procedure itself, but rather the motion prediction step, and more particularly the derivation of the change mask that must be able to differentiate to which part of the changing map is linked each moving object, in order to output several new estimate positions.

The experiments conducted on the aa13_Victory sequence showed another limitation regarding the propagation of errors that can lead to a loss of track. As a matter of fact, both the motion prediction step and the matching step greatly depend on each other results, as they are sequentially conducted, and a poor output of any of these steps can potentially compromise the whole tracking process. A possible solution to prevent this effect could be to regularly check that the shape of the estimate position obtained by (9) is similar to the output of a change detection process between the current frame $\mathcal{I}_t$ and an empty background frame, and reset the tracker if the test is not conclusive. Conducting this test in an objective and generic fashion is, however, a challenging task, since it requires to define up to which point two shapes are similar or not, and this could vary from one application to the other. Improving this point is also part of our future research objectives.

In spite of the two aforementioned issues, the proposed BPT-based tracking method features several advantages with respect to the presented state-of-the-art approaches. First, it really makes the most out of the temporal dimension of the sequence through the motion prediction step, while the matching stage itself relies on the spectral and spatial information contained in each hyperspectral frame. The final output has a real tracking connotation, with the tracked object appearing as connected, unlike the three state-of-the-art methods for which the pixels labeled as "plume" are not handled as a semantic object as a whole. Another asset of the proposed method is that it does not require any prior information on the spectral composition of the object to track, which is contrarily incorporated and updated on-the-fly. Last but not least, the tracking can be performed on a frame-by-frame basis, even though real-time performances are not reached yet (and are beyond the scope of this paper), allowing for passive surveillance applications.

## VII. Conclusion

In this paper, we presented a new methodology to perform hyperspectral object tracking based on the hierarchical analysis of hyperspectral video sequences, and applied it to the tracking of chemical gas plumes. Like classical object tracking algorithms for color video sequences, the proposed method was decomposed in a motion estimation step and a matching step, performed sequentially. The motion estimation involved the derivation of a so-called binary change mask between consecutive frames with a GLRT approach, based on the assumption that every pixel spectrum can be written as a linear combination between a moving object and a fixed background. The change mask was then combined with the known position of the object in the previous frame to derive an estimate of its new position in the current frame. The matching step was handled as an object detection process. The use of a BPT to that purpose allowed to drastically reduce the object search space by representing a hyperspectral frame as a limited set of hierarchically organized candidate regions. The matching involved the definition of a set of reference features for the sought object and the evaluation of every candidate region features against those reference ones.

The proposed method was applied to the tracking of chemical gas plumes in two different LWIR hyperspectral video sequences. This challenging application is, up to now, mostly addressed in the literature with the use of anomaly detectors that do not make the most out of the temporal information in the sequence, in addition to requiring the knowledge of the target spectrum, or clustering-based method that stack several consecutive frames, hence not being operable in any real-time scenario. We manually delineated some ground-truth data for all frames of the two sequences, and corresponding performance metrics were introduced in order to compare our proposed method with three state-of-the-art methods, namely, the statistical AMSD and ACE detectors and the clustering-based RNMF approach. A critical evaluation of the obtained results was then conducted, allowing to identify the advantages and drawbacks of all compared methods. In particular, despites the promising results of the proposed methodology, we established several points to investigate in the future, such as the relaxation of the *one object, one node* paradigm and the management of errors propagation leading to a loss of track. Further experiments will also be conducted on other hyperspectral sequences of various physical natures in order to evaluate the robustness of the proposed methodology with respect to the underlying application.

## Acknowledgment

The authors would like to thank the John Hopkins Applied Physics Laboratory for providing the data sets, the reference target spectra, and the AMSD implementation. They would also like to thank D. Pauwels for her help regarding the ground-truth delineation. G. Tochon was with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology when the materials presented in this paper were conducted.

## REFERENCES


[1] J. R. Schott, *Remote Sensing: The Image Chain Approach*. Oxford, U.K.: Oxford Univ. Press, 2007.

[2] P. S. Thenkabail, J. G. Lyon, and A. Huete, *Hyperspectral Remote Sensing of Vegetation*. Boca Raton, FL, USA: CRC Press, 2012.

[3] F. D. van der Meer *et al.*, "Multi- and hyperspectral geologic remote sensing: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 14, no. 1, pp. 112–128, 2012.

[4] M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," *Water SA*, vol. 33, no. 2, pp. 145–151, 2007.

[5] G. Elmasry, M. Kamruzzaman, D.-W. Sun, and P. Allen, "Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review," *Crit. Rev. Food Sci. Nutrition*, vol. 52, no. 11, pp. 999–1023, 2012.

[6] D. Lorente, N. Aleixos, J. Gômez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment," *Food Bioprocess Technol.*, vol. 5, no. 4, pp. 1121–1142, 2012.

[7] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, p. 010901, 2014.

[8] M. A. Calin, S. V. Parasca, D. Savastru, and D. Manea, "Hyperspectral imaging in the medical field: Present and future," *Appl. Spectrosc. Rev.*, vol. 49, no. 6, pp. 435–447, 2014.

[9] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[10] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.

[11] E. Trucco and K. Plakas, "Video tracking: A concise survey," *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 520–529, Apr. 2006.

[12] A. Banerjee, P. Burlina, and J. Broadwater, "Hyperspectral video for illumination-invariant tracking," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, 2009, pp. 1–4.

[13] H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 44–51.

[14] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2000, pp. 142–149.

[15] S. J. Young, "Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging," Aerosp. Corp., Aerosp. Rep. ATR-2002(8407)-1, 2002.

[16] V. Farley, A. Vallières, A. Villemaire, M. Chamberland, P. Lagueux, and J. Giroux, "Chemical agent detection and identification with a hyperspectral imaging infrared sensor," *Proc. SPIE*, vol. 6739, p. 673918, Oct. 2007.

[17] G. Tochon, J. Chanussot, J. Gilles, M. Dalla Mura, J.-M. Chang, and A. Bertozzi, "Gas plume detection and tracking in hyperspectral video sequences using binary partition trees," in *Proc. 7th Int. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2014.

[18] D. G. Manolakis, S. E. Golowich, and R. S. DiPietro, "Long-wave infrared hyperspectral remote sensing of chemical clouds: A focus on signal processing approaches," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 120–141, Jul. 2014.

[19] T. S. Spisz, P. K. Murphy, C. C. Carter, A. K. Carr, A. Vallières, and M. Chamberland, "Field test results of standoff chemical detection using the first," *Proc. SPIE*, vol. 6554, p. 655408, Apr. 2007.

[20] J. B. Broadwater, T. S. Spisz, and A. K. Carr, "Detection of gas plumes in cluttered environments using long-wave infrared hyperspectral sensors," *Proc. SPIE*, vol. 6954, p. 69540R, Apr. 2008.

[21] J. B. Broadwater, D. Limsui, and A. K. Carr, "A primer for chemical plume detection using LWIR sensors," Dept. Nat. Secur. Technol., John Hopkins Appl. Phys. Lab., Tech. Rep. NSTD-11-0661, 2011.

[22] E. M. O'Donnell, D. W. Messinger, C. Salvaggio, and J. R. Schott, "Identification and detection of gaseous effluents from hyperspectral imagery using invariant algorithms," *Proc. SPIE*, vol. 5425, pp. 573–582, Aug. 2004.

[23] A. Vallières *et al.*, "Algorithms for chemical detection, identification and quantification for thermal hyperspectral imagers," *Proc. SPIE*, vol. 5995, p. 59950G, Nov. 2005.

[24] D. G. Manolakis and F. M. D'Amico, "A taxonomy of algorithms for chemical vapor detection with hyperspectral imaging spectroscopy," *Proc. SPIE*, vol. 5795, pp. 125–133, May 2005.

[25] E. R. Larrieux, "Performance evaluation of chemical plume detection and quantification algorithms," M.S. thesis, Dept. Elect. Comput. Eng., Northeastern Univ., Boston, MA, USA, 2009.

[26] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[27] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[28] D. G. Manolakis, G. A. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," *Proc. SPIE*, vol. 4049, pp. 2–17, Aug. 2000.

[29] S. W. Sharpe, T. J. Johnson, R. L. Sams, P. M. Chu, G. C. Rhoderick, and P. A. Johnson, "Gas-phase databases for quantitative infrared spectroscopy," *Appl. Spectrosc.*, vol. 58, no. 12, pp. 1452–1461, 2004.

[30] (Apr. 2014). *Pacific Northwest National Laboratory IR Database*. [Online]. Available: http://nwir.pnl.gov

[31] E. Hirsch and E. Agassi, "Detection of gaseous plumes in IR hyperspectral images using hierarchical clustering," *Appl. Opt.*, vol. 46, no. 25, pp. 6368–6374, 2007.

[32] T. Gerhart *et al.*, "Detection and tracking of gas plumes in LWIR hyperspectral video sequence data," *Proc. SPIE*, vol. 8743, p. 87430J, May 2013.

[33] E. Merkurjev, J. Sunu, and A. L. Bertozzi, "Graph MBO method for multiclass segmentation of hyperspectral stand-off detection video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 689–693.

[34] H. Hu, J. Sunu, and A. L. Bertozzi, "Multi-class graph Mumford–Shah model for plume detection using the MBO scheme," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2015, pp. 209–222.

[35] E. Merkurjev, T. Kostić, and A. L. Bertozzi, "An MBO scheme on graphs for classification and image processing," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 1903–1930, 2013.

[36] J. Qin *et al.*, "Detecting plumes in LWIR using robust nonnegative matrix factorization with graph-based initialization," *Proc. SPIE*, vol. 9472, p. 94720V, May 2015.

[37] J. Delon, "Midway image equalization," *J. Math. Imag. Vis.*, vol. 21, no. 2, pp. 119–134, 2004.

[38] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[39] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 262–272, Feb. 2012.

[40] G. Martin and A. Plaza, "Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 380–395, Apr. 2012.

[41] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 561–576, Apr. 2000.

[42] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1430–1443, Apr. 2013.

[43] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image segmentation using binary partition trees," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 1273–1276.

[44] M. A. Veganzones, G. Tochon, M. Dalla-Mura, A. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3574–3589, Aug. 2014.

[45] A. Alonso-González, S. Valero, J. Chanussot, C. López-Martínez, and P. Salembier, "Processing multidimensional SAR and hyperspectral images with binary partition tree," *Proc. IEEE*, vol. 101, no. 3, pp. 723–747, Mar. 2013.

[46] L. Drumetz *et al.*, "Binary partition tree-based local spectral unmixing," in *Proc. 7th Int. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2014.

[47] S. Valero, P. Salembier, and J. Chanussot, "Object recognition in hyperspectral images using Binary Partition Tree representation," *Pattern Recognit. Lett.*, vol. 56, pp. 45–51, Apr. 2015.

[48] S. Valero, P. Salembier, J. Chanussot, and C. M. Cuadras, "Improved binary partition tree construction for hyperspectral images: Application to object detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2011, pp. 2515–2518.
[49] S. Valero, P. Salembier, and J. Chanussot, "Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 2565–2568.
[50] F. Calderero and F. Marques, "Region merging techniques using information theory statistical measures," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1567–1586, Jun. 2010.
[51] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, "Using multiple segmentations to discover objects and their extent in image collections," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 1605–1614.
[52] T. Malisiewicz and A. A. Efros, "Improving spatial support for objects via multiple segmentations," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2007.
[53] H. Lu, J. C. Woods, and M. Ghanbari, "Binary partition tree for semantic object extraction and image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 3, pp. 378–383, Mar. 2007.
[54] O. Salerno, M. Pardàs, V. Vilaplana, and F. Marques, "Object recognition based on binary partition trees," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 2. 2004, pp. 929–932.
[55] V. Vilaplana, F. Marques, and P. Salembier, "Binary partition trees for object detection," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2201–2216, Nov. 2008.
[56] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[57] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*, 3rd ed., vol. 709. Hoboken, NJ, USA: Wiley, 2012. [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470178965.html
[58] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, vol. 197. Hoboken, NJ, USA: Wiley, 2009.
[59] A. Gillespie, S. Rokugawa, T. Matsunaga, J. S. Cothern, S. Hook, and A. B. Kahle, "A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1113–1126, Jul. 1998.
[60] V. Farley, A. Vallières, M. Chamberland, A. Villemaire, and J.-F. Legault, "Performance of the FIRST: A long-wave infrared hyperspectral imaging sensor," *Proc. SPIE*, vol. 6398, p. 63980T, Oct. 2006.
[61] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.
[62] I. Leifer *et al.*, "Comparing imaging spectroscopy and *in situ* observations of Chino dairy complex emissions," in *Proc. 9th Int. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Aug. 2016.
[63] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," in *Proc. Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, vol. 61. 2009, pp. 1–18.
[64] C. J. Needham and R. D. Boyle, "Performance evaluation metrics and statistics for positional tracker evaluation," in *Computer Vision Systems*. Berlin, Germany: Springer, 2003.
[65] F. Yin, D. Makris, and S. Velastin, "Performance evaluation of object tracking algorithms," in *Proc. 10th IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 25–32.
[66] J. Black, T. Ellis, and P. Rosin, "A novel method for video tracking performance evaluation," in *Proc. Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, 2003, pp. 125–132.
[67] F. Bashir and F. Porikli, "Performance evaluation of object detection and tracking systems," in *Proc. 9th IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, vol. 5. Jun. 2006, pp. 7–14.
[68] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[69] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

**Guillaume Tochon** (S'14–M'16) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 2012, and the Ph.D. degree in signal and image processing from Université Grenoble Alpes, Grenoble, and the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-lab), Grenoble, in 2015.

From 2015 to 2016, he was a temporary Assistant Professor with Grenoble INP and the GIPSA-lab. Since 2016, he has been an Assistant Professor with the Graduate School of Computer Science and Advanced Techniques (EPITA), Paris, France. He is conducting his research with the EPITA Research and Development Laboratory (LRDE). His research interests with LRDE include mathematical morphology, pattern recognition, and data fusion for remote sensing applications, with a focus on hyperspectral imagery.

**Jocelyn Chanussot** (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement, French National Defense Department, Arcueil, France. From 2015 to 2017, he was a Visiting Professor with the University of California at Los Angeles, Los Angeles, CA, USA. Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005 and an Associate Professor from 2005 to 2007. He has been a Visiting Scholar with Stanford University, Stanford, CA, USA, the Royal Institute of Technology, Stockholm, Sweden, and the National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland.

He is currently a Professor of Signal and Image Processing with Grenoble INP. He is conducting his research with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot has been a member of the Institut Universitaire de France, Paris, France, since 2012. He was a member of the IEEE Geoscience and Remote Sensing Society Administrative Committee from 2009 to 2010 and in charge of membership development. He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008. He was the founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE GRS-S Chapter Excellence Award. He was a co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. He was the Chair from 2009 to 2011 and a Co-Chair of the GRS Data Fusion Technical Committee from 2005 to 2008. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He was an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2005 to 2007 and *Pattern Recognition* from 2006 to 2008. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. He was a Guest Editor of the PROCEEDINGS OF THE IEEE in 2013 and a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE in 2014. Since 2007, he has been an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

**Mauro Dalla Mura** (S'08–M'11) received the Laurea (B.E.) and Laurea Magistrale (M.E.) degrees in telecommunication engineering from the University of Trento, Trento, Italy, in 2005 and 2007, respectively, and the joint Ph.D. degree in information and communication technologies (telecommunications area) from the University of Trento and in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2011.

In 2011, he was a Research Fellow with Fondazione Bruno Kessler, Trento, where he was involved in conducting research on computer vision. He is currently an Assistant Professor with the Grenoble Institute of Technology, Grenoble, France. He is conducting his research with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble. His research interests include the fields of remote sensing, image processing, and pattern recognition, particularly mathematical morphology, classification, and multivariate data analysis.

Dr. Dalla Mura was a recipient of the IEEE Geoscience and Remote Sensing Society (GRSS) Second Prize in the Student Paper Competition of the 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2011 and a co-recipient of the Best Paper Award of the *International Journal of Image and Data Fusion* from 2012 to 2013 and the Symposium Paper Award for IEEE IGARSS 2014. He has been the President of the IEEE GRSS French Chapter since 2016, where he was a Secretary from 2013 to 2016. He has been on the Editorial Board of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2016.

**Andrea L. Bertozzi** (M'14) received the B.A., M.A., and Ph.D. degrees from Princeton University, Princeton, NJ, USA, in 1987, 1988, and 1991, respectively, all in mathematics.

She was on the faculty of the University of Chicago at Illinois, Chicago, IL, USA, from 1991 to 1995, and Duke University, Durham, NC, USA, from 1995 to 2004. From 1995 to 1996, she was the Maria Goeppert-Mayer Distinguished Scholar with the Argonne National Laboratory, Lemont, IL, USA. In 2012, she was appointed as the Betsy Wood Knapp Chair for Innovation and Creativity. Since 2003, she has been with the University of California at Los Angeles, Los Angeles, CA, USA, as a Professor of Mathematics, where she is currently the Director of Applied Mathematics. Her research interests include image inpainting, image segmentation, cooperative control of robotic vehicles, swarming, fluid interfaces, and crime modeling.

Dr. Bertozzi is currently a fellow of the American Academy of Arts and Sciences, the Society for Industrial and Applied Mathematics, the American Mathematical Society, and the American Physical Society. She was a recipient of the Sloan Foundation Research Fellowship, the Presidential Career Award for Scientists and Engineers, and the SIAM Kovalevsky Prize in 2009.