Applications of low rank and sparse matrix decompositions in hyperspectral video processing

Jen-Mei Chang
California State University, Long Beach

Torin Gerhart
Western Digital Corp.

1.1   Introduction

This chapter will provide a background to those unfamiliar with hyperspectral images and detail the application of low rank and sparse decompositions in hyperspectral video analysis. Traditional RGB images capture light in the red, green, and blue portions of the visible light spectrum. This is due to the way images are displayed on computer monitors and televisions using red, green, and blue pixels. Each layer represents the amount of radiated energy being emitted at a particular wave length. Images using more than 3 layers are referred to as multispectral or hyperspectral images. These images can involve light that is outside the visible spectrum, such as infra-red (IR) and UV (ultra-violet) light. Hyperspectral images have a higher spectral resolution compared to multispectral images while being limited to a narrow spectral bandwidth. By imaging the light that is absorbed and reflected in high detail within a certain region of the electromagnetic spectrum, it is possible to identify particular materials present in the image. One major application of hyperspectral imaging is the detection of invisible gaseous chemical agents and other anomalous harmful particles. This problem comes up in many practical applications such as defense, security, and environmental safety [BLC11].

Herein, we will focus on the detection and identification of gaseous chemical plumes in hyperspectral video data. Hyperspectral video adds a time component to the processing of hyperspectral images. Processes such as gaseous diffusion can be captured, even for gases not visible to the human eye. These video sequences are typically large in size due to the fact that the images themselves are of high resolutions. Dimension reduction is often needed to make further processing easier. An algorithm to decompose a hyperspectral video sequence into a low-rank and sparse representation will be derived and applied to the detection of chemical plumes. The particular data set used in this work was provided by the Applied
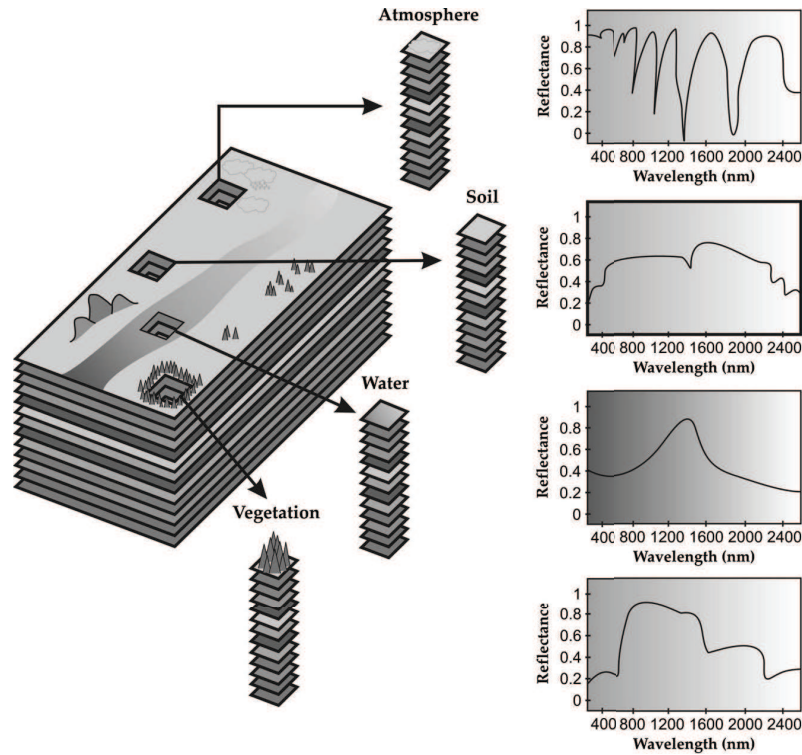
1.2   Hyperspectral Imaging

FIGURE 1.1   Example of a hyperspectral image. The image contains many layers that capture different wavelengths of reflected light. Each pixel contains information about the materials present in the scene.

Traditional RGB images consist of three layers, each corresponding to intensity values of the red, green, and blue components of the image. Hyperspectral images extend this idea by capturing hundreds of different frequencies of light. Hyperspectral imaging aims to capture the absorbed and reflected radiation of a given scene such that a region of the electromagnetic spectrum may be viewed with high spectral resolution. This spectral resolution is what differentiates hyperspectral imaging from other types of spectral imaging, such as multispectral imaging. In order to produce an accurate representation of a region of the electromagnetic spectrum, a fine discretization is required. Hyperspectral images have a frequency gap between spectral bands that is small relative to the measured spectrum.

Hyperspectral imaging finds a lot of application in the geosciences. Measuring the absorbed and reflected radiation allows for the characterization of surface and atmospheric features in geophysical images. Figure 1.1©(2012) SPIE shows how pixels in a hyperspectral image contain spectral information that may be used to distinguish between vegetation, water, soil, and atmosphere. The most widespread application of hyperspectral imaging is the identification of particular materials present in a scene.

Hyperspectral video sequences serve as an excellent example of a problem from modern imaging science that is closely tied to the analysis of large data sets. Each hyperspectral video frame is an individual data cube, which means these video sequences require a large

amount of memory. For example, a data cube with a spatial resolution of 150×300 pixels and a spectral resolution of 150 would require 54 megabytes of physical memory (assuming it is stored as 8-byte double precision floating point values). Twenty frames alone would require over a gigabyte in the memory space. The size of these videos may prohibit loading the entire video into memory without sufficient preprocessing or dimension reduction.

Hyperspectral imaging was originally developed to be used for geology and mining applications. The imaging of core samples would detect the presence of particular mineral deposits and could aid in finding oil [Ell01]. Today, hyperspectral imaging is used in many different areas from surveillance and defense to ecology and agriculture.

1.2.1 Mixing Model

A *three-layer mixing model* is a simple method to describe the different components that comprise the spectral radiance measurement for each pixel in a long wave infrared hyperspectral image. Figure 1.2 ©(2011) JHU illustrates the different objects, or layers, that contribute to the spectral radiance measurement of the long wave infrared (LWIR) spectrometer. The three layers in this data set are the background, chemical plume, and atmosphere. Each pixel has its own radiance $L(\nu)$, and transmittance $\tau(\nu)$. The transmittance is the ratio of light leaving a surface relative to the amount of light entering the medium. Both the background and plume spectral radiances must pass through other mediums before reaching the long wave infrared spectrometer. Therefore, the spectral radiance measurement of the sensor can be represented as

$$L(\nu) = \tau_{atm}(\nu)L_p(\nu) + \tau_p(\nu)\tau_{atm}(\nu)L_b(\nu) + L_{atm}(\nu). \tag{1.1}$$

The subscripts *atm*, *p*, and *b* in Equation (1.1) refer to the atmosphere, plume, and background, respectively. This model may be modified slightly for our application. In the data set we are using, the long wave infrared spectrometers are placed within two kilometers of the release site. It is reasonable to assume that the contribution to spectral radiance from the atmosphere would be negligible, therefore the L_{atm} term from Equation (1.1) can be dropped for simplicity. In addition, it is assumed that the atmospheric transmittance does not significantly affect the spectral radiance because of the short path length, allowing most of the signal to pass through. These assumptions reduce Equation (1.1) into a simpler *two layer model* equation

$$L(\nu) = \tau_p(\nu)L_b(\nu) + L_p(\nu) \tag{1.2}$$

According to this model the spectral radiance of the scene measured by the LWIR sensors is a sum of the light emitted by the chemical plume and background mediums at the measured wavelengths in the electromagnetic spectrum [BLC11].

1.2.2 Dugway Proving Ground Data Set

The data set analyzed for this work was provided by the Applied Physics Laboratory at Johns Hopkins University as part of a Defense Threat Reduction Agency (DTRA) research grant. It consists of a series of video sequences recording the release of non-toxic chemical plumes into the atmosphere at the US Army's Dugway Proving Ground facility in Utah. Figure 1.3 ©(2011) JHU shows the three long wave infrared spectrometers (named Romeo, Victory and Tango) placed at different locations to track the release of known chemicals. The sensors capture one frame every five seconds consisting of 2 spatial dimensions and one spectral dimension. The spatial dimension of each of these data cubes is 128×320 pixels, while spectral dimension measures 129 different wavelengths in the long wave infrared (LWIR) portion of the electromagnetic spectrum. Each layer in the spectral dimension
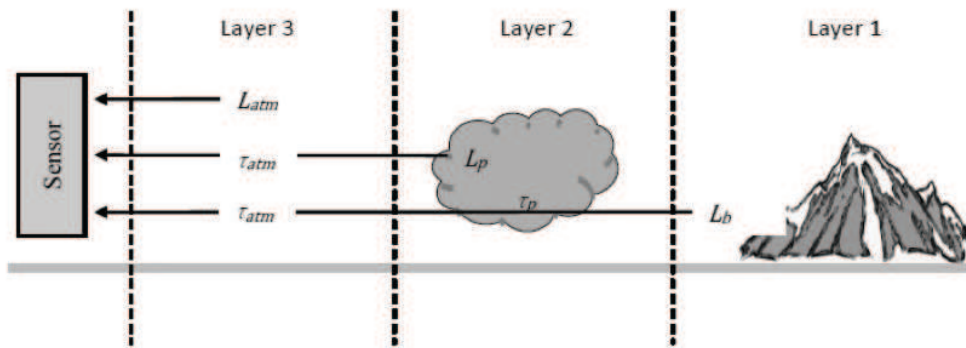
FIGURE 1.2   A three-layer mixing model depicting the composition of spectral radiance in hyperspectral images.

depicts a particular frequency starting at 7,830 nm and ending with 11,700 nm with a channel spacing of 30 nm. Low-rank and sparse decompositions of this data set will provide an estimation of the background radiation, i.e., the low-rank component, and the motion captured between frames, i.e., the sparse component.
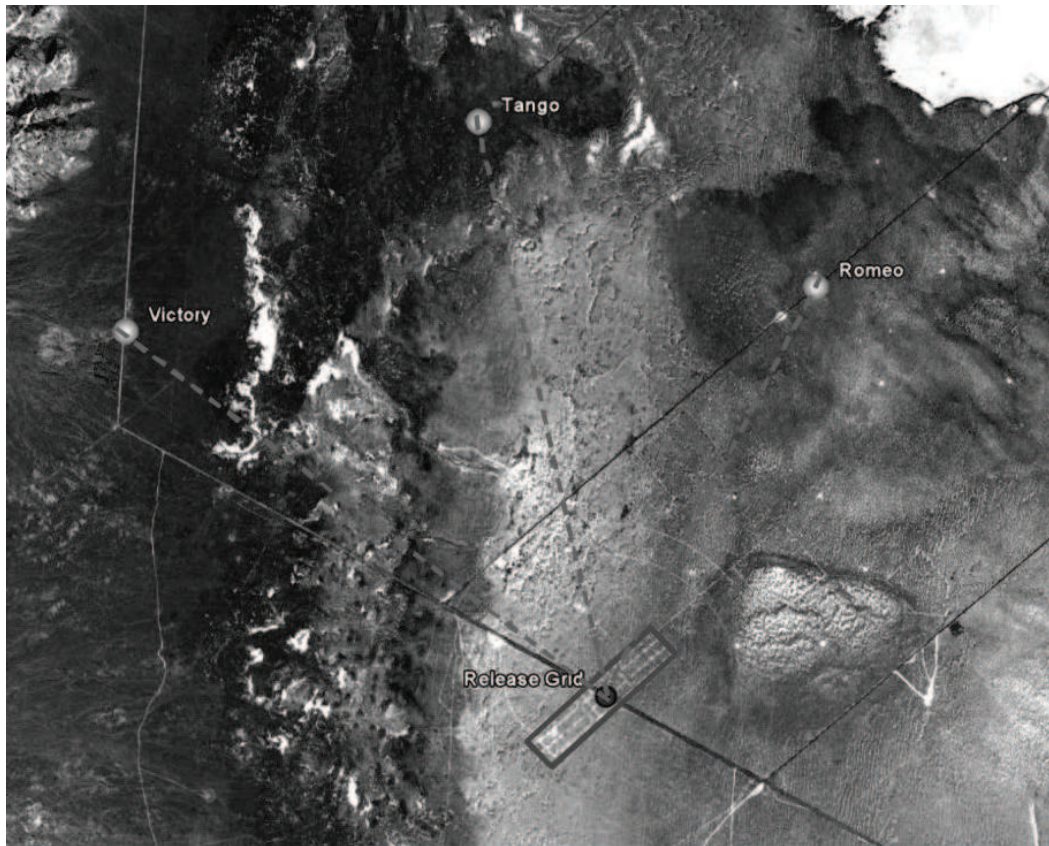

FIGURE 1.3   Placement of the three long wave infrared spectrometers to track the release of the known chemicals.

1.3   Principal Component Pursuit

This section outlines the problem of decomposing a hyperspectral video sequence into low rank and sparse matrices. A data matrix X is constructed through concatenation of video frames and decomposed into $X = L + S$ where L is a low rank matrix and S is a sparse matrix. The algorithm used to arrive at this decomposition is iterative and involves minimizing a functional consisting of the nuclear norm and the ℓ_1 norm defined on matrices. The version of the ℓ_1 norm we are using is not the typical 1-norm for matrices. It is a slightly modified version of the ℓ_1 vector norm, extended to be defined on matrices, where a matrix is seen as a long vector. This norm is used here as a sparsity inducing norm, intended to give sparse solutions. The nuclear norm is used as an approximation to the *rank* function. A minimal nuclear norm implies low-rank. By minimizing this functional, we are guaranteed solutions which are low-rank and sparse. Formally stated, the problem is

$$\underset{L,S}{\text{minimize}} \; \|L\|_\star + \lambda \|S\|_1$$
$$\text{subject to } X = L + S \tag{1.3}$$

where,

$$\|X\|_1 = \sum_{i,j} |X_{i,j}| \;\; (\ell_1 \text{ norm})$$
$$\|X\|_\star = \sum_{i=1}^{r} \sigma_i(X) \text{ (nuclear norm)} \tag{1.4}$$

Here, $\sigma_i(X)$'s are the singular values of X and r is the rank of X. The objective function of (1.3) has a separable structure associated with it, which allows the problem to be split into separate subproblems. The augmented Lagrangian [CLMW09] of (1.3) gives

$$\mathcal{L}_A(L, S, Y) = \|L\|_\star + \lambda \|S\|_1 + \langle Y, X - L - S \rangle + \frac{\mu}{2} \|X - L - S\|_F^2. \tag{1.5}$$

Solutions are derived from (1.5) by using the alternating direction method of multipliers (ADMM). One subproblem holds S fixed and minimizes $\mathcal{L}_A(L, S, Y)$ with respect to L, the next subproblem holds the new L fixed and minimizes $\mathcal{L}_A(L, S, Y)$ with respect to S. The subproblem involving the minimization of the nuclear norm is related to the minimization of the *rank* function. The rank function is non-convex, and finding minimizers is NP-hard. Replacing the non-convex *rank* function with a convex approximation makes the problem much easier to solve. The nuclear norm is the "best" convex approximation to the *rank* function [CT10]. The solution to the subproblem involving minimization of the nuclear norm is given by the singular value thresholding operator, $\mathcal{D}_\tau$. A more detailed derivation may be found in [CR09]. The other subproblem involving the $\|S\|_1$ term is essentially minimizing the ℓ_1 vector norm and is therefore solved by a thresholding operator, $\mathcal{S}_\tau$. The iterations obtained from (1.5) are,

$$L_{k+1} = \underset{L}{\text{argmin}} \; \mathcal{L}_A(L_k, S_k, Y_k) = \mathcal{D}_\tau(X - S_k + \mu^{-1} Y_k)$$
$$S_{k+1} = \underset{S}{\text{argmin}} \; \mathcal{L}_A(L_{k+1}, S_k, Y_k) = \mathcal{S}_{\frac{\lambda}{\mu}}(X - L_{k+1} + \frac{1}{\mu} Y_k) \tag{1.6}$$
$$Y_{k+1} = Y_k + \mu(X - L_{k+1} - S_{k+1})$$

where,

$$\mathcal{D}_\tau(X) = U \mathcal{S}_\tau(\Sigma) V^T = U \text{diag}(\max(\sigma_i - \tau, 0)) V^T$$
$$\mathcal{S}_\tau(X) = \max(0, x_{ij} - \tau).$$

When applied to video analysis, the low rank matrix the final decomposition represents the background, and the sparse matrix represents motion between frames. The algorithm requires a number of other matrices with the same size as X to be held in memory concurrently. The memory requirement of this problem is typically where challenges arise in practice. For example, each frame of the Dugway Proving Ground data set is a $128\times320\times129$ data cube. Concatenating along the spectral dimension produces a vector of length 5,283,840. A data matrix with 100 frames will be of size $5{,}283{,}840 \times 100$. There are many preprocessing techniques that can be utilized to make the task computationally feasible. For example, one can select a subset of the spectral channels based on noise or performing dimension reduction on each frame of the video sequence.

1.4   Results

The Low-Rank Sparse algorithm outlined in (1.6) was applied to a false color RGB video, created by projecting each frame onto the first three principle components. The movie generated for these results was done in RGB to demonstrate different aspects of the decomposition. Movies with more layers were processed and the results are similar to what is presented here. In early frames, the low rank approximation is able to capture the background very well. After the plume is released, the sparse component captures the movement of the plume through each channel of the video sequence. Applying this method to the original (non-reduced) video sequence results in the background matrix approximating stationary signals and the sparse component showing moving signals and noise. Each frame contained 128×320 pixels, with 3 layers. By concatenating the frames into long vectors in $\mathbb{R}^{128\times320\times3}$, a movie of 40 frames was made into the $122{,}880 \times 40$ matrix M. The original matrix had rank 40, while the low rank approximation L had rank 12.

Motion was captured on each RGB layer in the sparse component S. The red component of the resulting decomposition may be seen in Figure 1.4. Notice that the frames of the low rank decomposition are virtually identical and the noise may be seen in the sparse component in the first 20 frames, as illustrated in Figure 1.4(a). On the other hand, Figure 1.4(b) shows the release of the plume in frames 21 through 40. Its motion is captured throughout the sparse component with some anomalies appearing in the low rank component towards the end of the sequence in areas where the plume has traveled. Similar results in the each of the green and blue components of the resulting decomposition may be seen in Figures 1.5 and 1.6, respectively. The final resulting RGB movies may be seen in Figure 1.7. The lighting fluctuations and noise can be seen in the sparse component of 1.7(a). As the gas plume is released, the diffusion process can be observed in the sparse component, as evidenced in 1.7(b). There, a noticeable presence of the plume can also be observed in the low rank component.

1.5   Concluding Remarks

The results presented here showed that it is possible to isolate the motion of chemical plumes from hyperspectral video data. This was achieved by using a reformulation of the principal components analysis to incorporate sparsity. The estimation of the background was done using a low-rank approximation. There are other means of background estimation such as a non-negative matrix factorization technique, which takes into account of other physical aspects of the problem as well. A better understanding of the physical process of chemical diffusion in the atmosphere is needed, such as when detection becomes impossible due to low concentrations. A practical real-time detection and tracking system for toxic chemical releases would require special mathematical and computational considerations. In order to
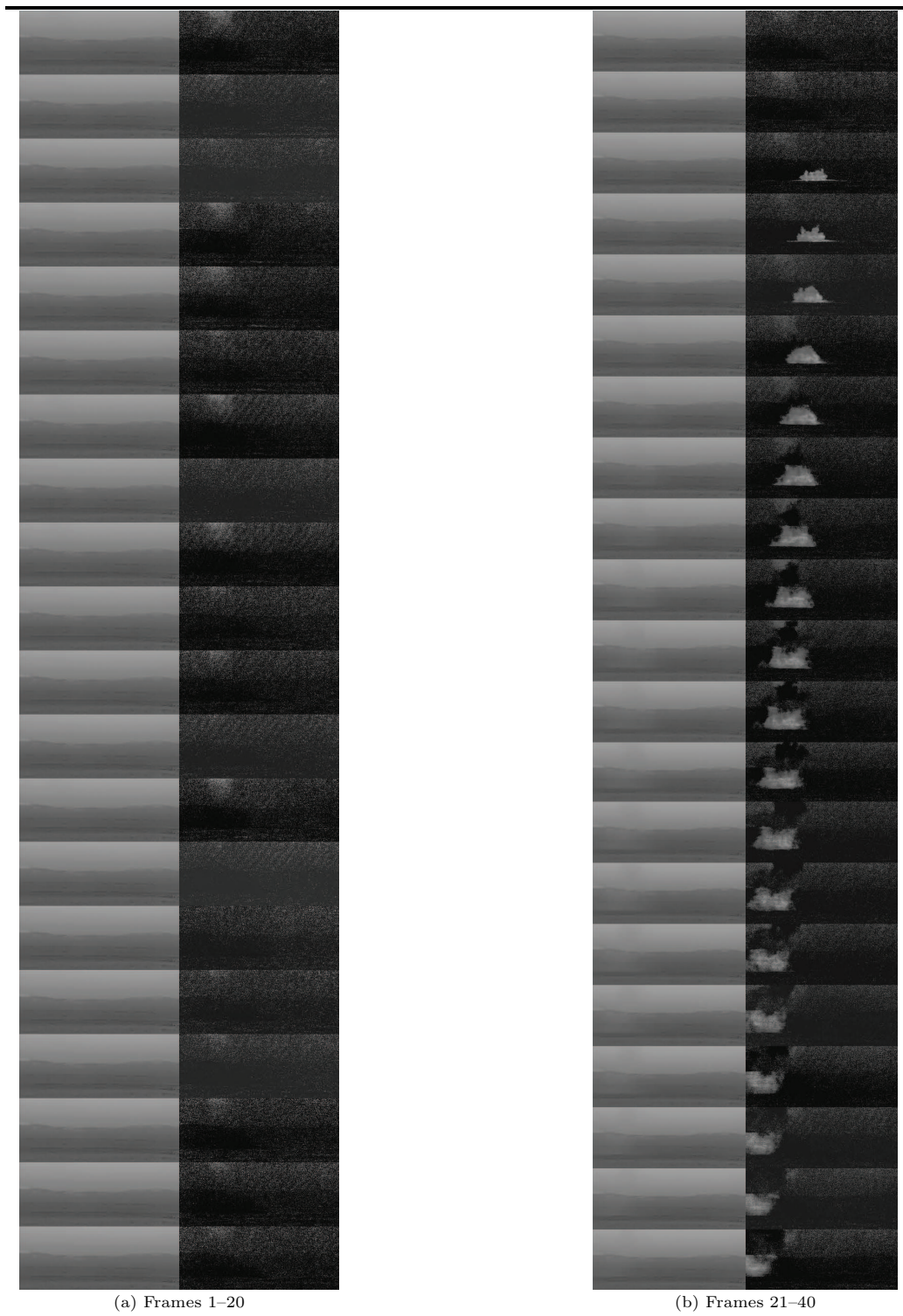
(a) Frames 1–20   (b) Frames 21–40

FIGURE 1.4   (a) Frames 1–20; (b) Frames 21–40 of the red component of the decomposed video sequence. The left panel displays the low rank approximation and the right panel shows the sparse component within each strip.
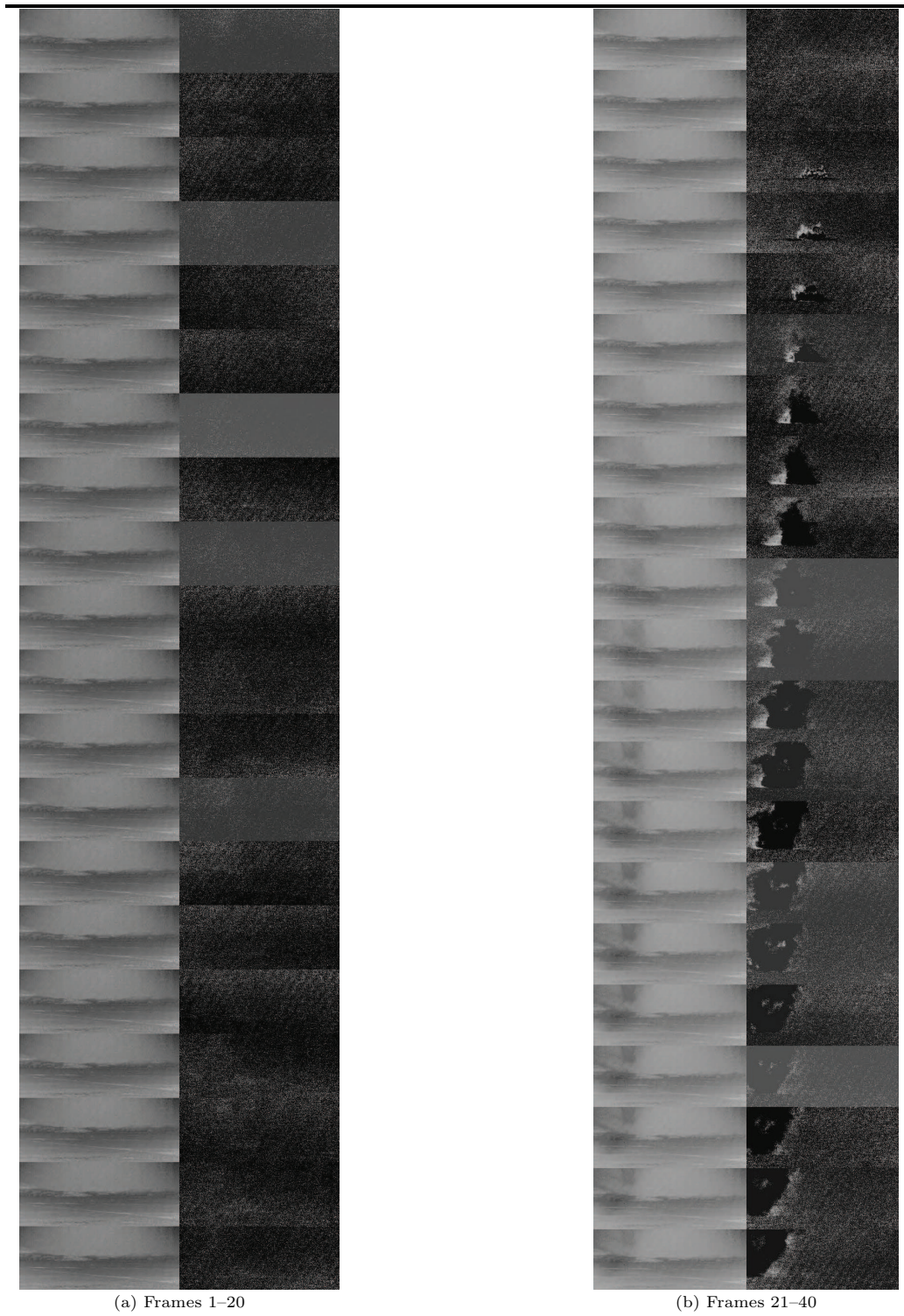
(a) Frames 1–20   (b) Frames 21–40

FIGURE 1.5   (a) Frames 1–20; (b) Frames 21–40 of the green component of the decomposed video sequence. The left panel displays the low rank approximation and the right panel shows the sparse component within each strip.
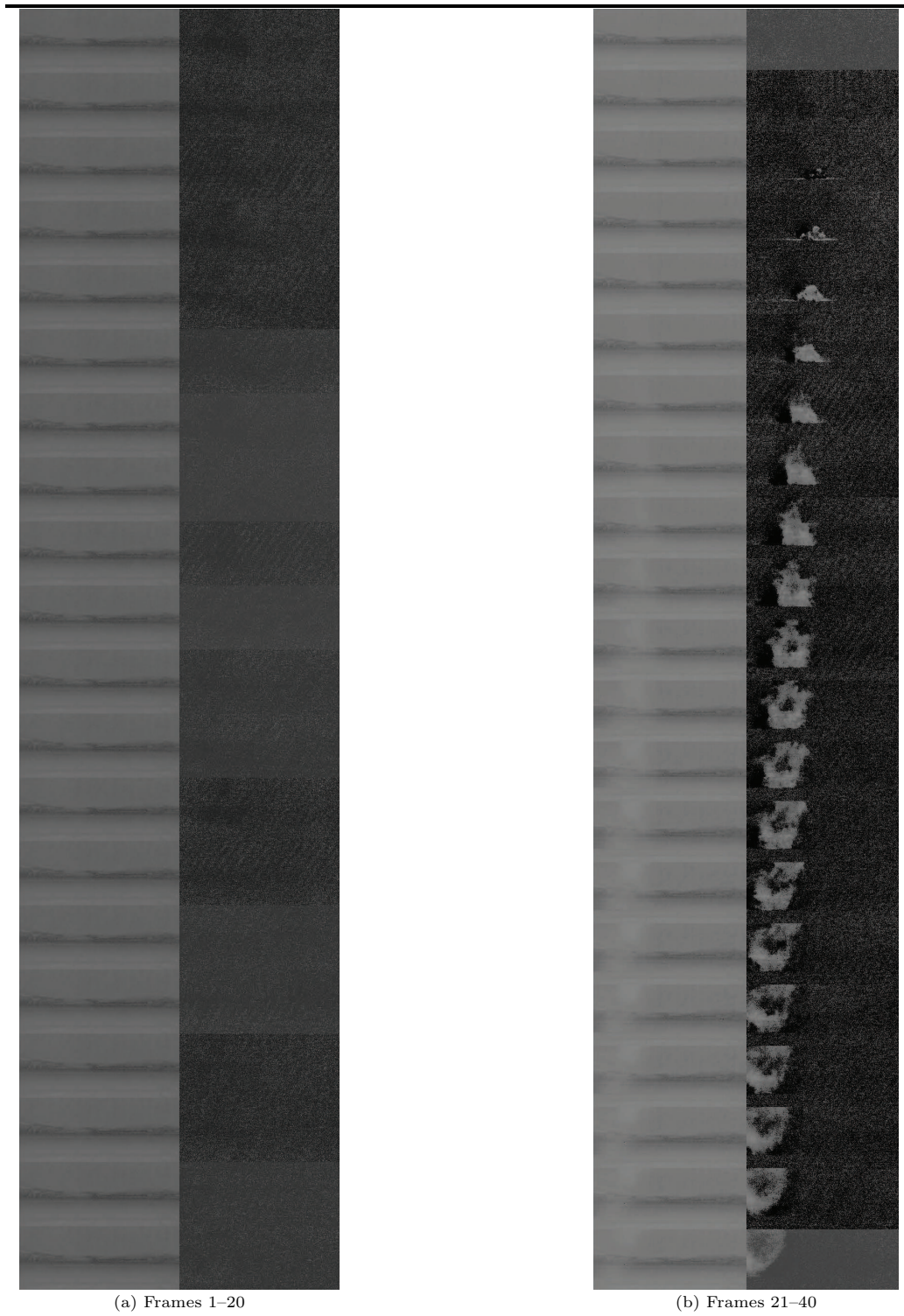
(a) Frames 1–20

(b) Frames 21–40

FIGURE 1.6 (a) Frames 1–20; (b) Frames 21–40 of the blue component of the decomposed video sequence. The left panel displays the low rank approximation and the right panel shows the sparse component within each strip.
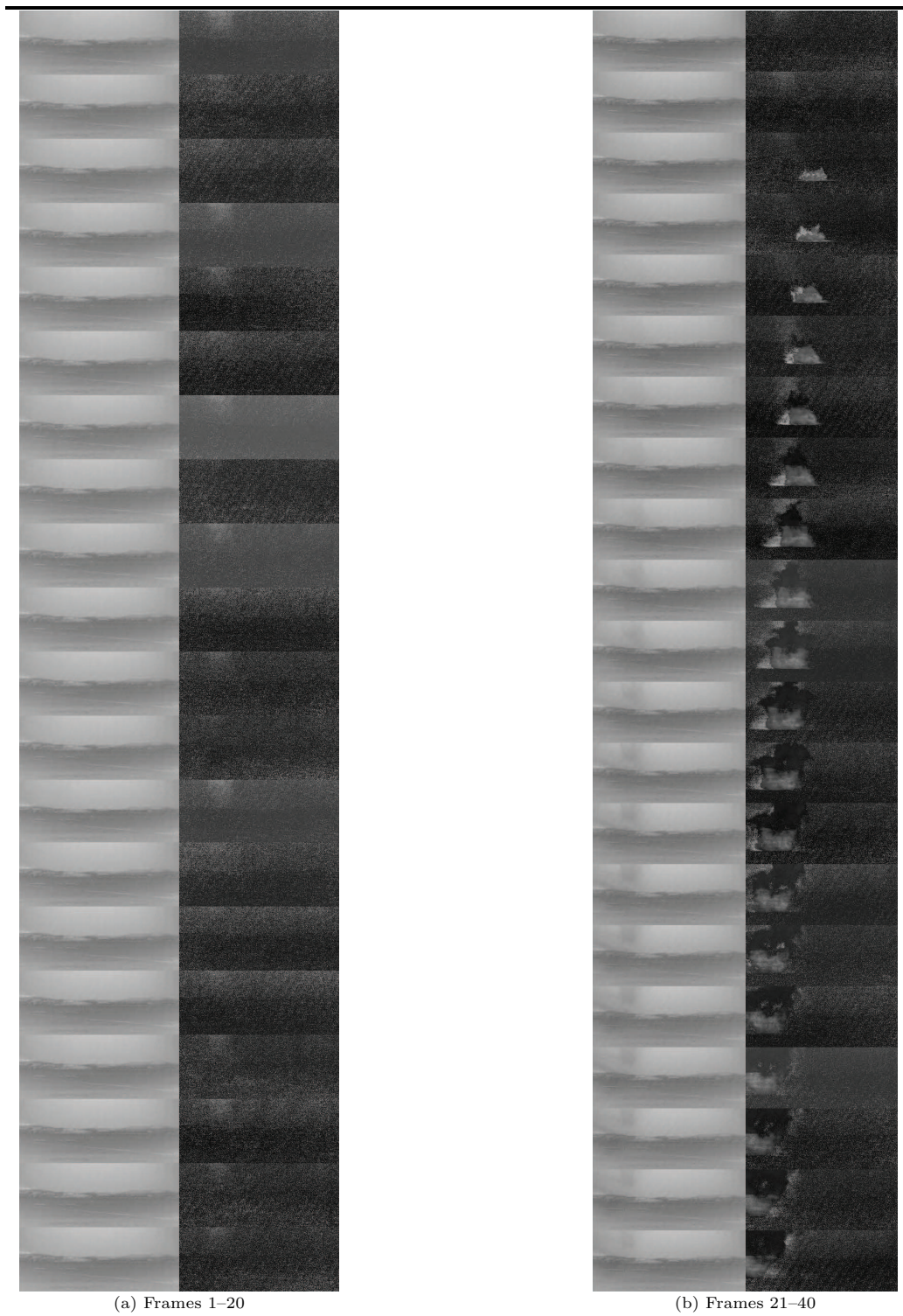
(a) Frames 1–20   (b) Frames 21–40

FIGURE 1.7   (a) Frames 1–20; (b) Frames 21–40 of the red, green, and blue components of the decomposed video sequence combined.

provide the accuracy and reliability required for such a system in the real world, many more physical and computational experiments need to be considered.

References

1. J. Broadwater, D. Limsui, and A.K. Carr. A primer for chemical plume detection using LWIR sensors. Technical report, National Security Technology Department, Johns Hopkins University, 2011.

2. E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of ACM*, 58(1):1–37, 2009.

3. E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

4. E. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

5. J. Ellis. Searching for oil seeps and oil-impacted soil with hyperspectral imagery. *Earth Observation Magazine*, 10(1):1058–1064, 2001.