Multiple views of a scene—obtained via different sensing modalities—have the potential to significantly enhance image analysis for remote sensing and other applications. This benefit is expected to be significant if the multiple views are providing independent, yet useful, information about the underlying classes in a scene. To exploit such multi-sensor information, a compressive-projection approach to the fusion of multi-sensor imagery is proposed. It is argued that random projections yield subspaces that preserve the discriminative nature of multi-sensor datasets with profound implications in a practical scenario wherein compressive measurements can directly facilitate data fusion without the need for complicated subspace-learning approaches. A case study fusing experimental hyperspectral and LiDAR data demonstrates that statistical learning in the compressive-measurement domain is not only feasible, but also provides a natural framework for sensor fusion without the need for explicit reconstruction from compressive measurements.

**Index Terms**—Random projections, Compressive sensing, data fusion

1. **INTRODUCTION**

Compressed-sensing (CS) architectures have now been investigated and developed for modern optical imaging modalities [1–10]. A variety of algorithms have been developed to recover signals from data-independent compressive measurements (often comprised of random projections implemented in hardware) [1,4,11–13]. In this paper, we argue that the capabilities of such compressive measurements (if implemented appropriately), in fact, surpass signal reconstruction, and we find that compressive measurements possess a natural ability for the fusion of multi-sensor data. To wit, we provide empirical evidence of the utility of simple, data-independent compressive measurements to facilitate effective sensor-fusion for image-analysis applications.

This work was partially supported by NASA grants NNX12AL49G and NNX14AI47G.

Modern geospatial imaging modalities include hyperspectral imagers and light detection and ranging (LiDAR) sensors. Such imagery often provides very useful information for image-analysis tasks, but is accompanied with the burden of excessive dimensionality of the resulting feature space and an inherent data glut. For instance, modern hyperspectral imagers often sample hundreds of contiguous spectral channels in the visible and infrared regions of the electromagnetic spectra, even though the intrinsic dimensionality for most applications (e.g., number of truly useful spectral channels for any particular image-analysis task) is often quite small. Likewise, next-generation LiDAR systems (full-waveform LiDAR) record the entire return pulse corresponding to each emitted laser pulse—although the degrees of freedom in such data are often small, such datasets are nevertheless of high dimension. Our application focus for this paper is hence the fusion of hyperspectral and LiDAR sensors, but, conceptually, this framework could feasibly be extended to arbitrary sensing modalities, assuming that compressive projections are meaningful, that the resulting data can be gridded onto an image, and that effective registration between the resulting images can be accomplished.

The outline of this paper is as follows. In Sec. 2, we describe the proposed compressive-measurement framework for multi-sensor image analysis and describe the Bayesian inference approach we use to learn the statistics of the data in the original, as well as compressive-measurement, domains along with alternate data-fusion approaches that we use for comparison. In Sec. 3, we summarize the experimental setup and results, making concluding remarks in Sec. 4.
representation of the information content as opposed to highly customized sensors that are tuned to a task—choosing the former versus latter allows us to trade-off simple and flexible sensing hardware with approaches that are optimized for a certain task but not necessarily generalizable.

The Johnson and Lindenstrauss lemma [14, 15] provides for preservation of the $\ell_2$ norm of $x \in \mathbb{R}^d$ (within a factor of $1 \pm \epsilon$) under a random linear transformation that maps $\mathbb{R}^d$ to $\mathbb{R}^k$, where $k = O\left(\frac{\epsilon^2 \log(1/\delta)}{\delta}\right)$, with probability at least $1 - \delta$. Traditional CS architectures (e.g., [1–4]) for optical imaging assume spatial modulation of a full frame (both spatial dimensions of an image) via a compressive-measurement projection, e.g., through a digital micro-mirror device (DMD). However, alternate architectures, particularly pertinent to hyperspectral sensing are emerging [5–10], including those where per-pixel spectral content is compressively projected via random projections [16].

2.1. Bayesian Inference in a Compressive-Measurement Domain

In this paper, we provide empirical evidence motivating a compressive data-fusion architecture for multi-view or multisensor image analysis. We contend that a lower-dimensional subspace obtained by projecting the data through a random projection matrix, is inherently suitable for image analysis without the need for reconstruction.

Consider two sensors acquiring imagery data, where the $i^{th}$ pixel generates two views: $x_1^i \in \mathbb{R}^{d_1}$, and $x_2^i \in \mathbb{R}^{d_2}$, from sensors 1 and 2 respectively. Consider orthogonal random projection matrices $P_1 : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{k_1}$ and $P_2 : \mathbb{R}^{d_2} \rightarrow \mathbb{R}^{k_2}$, that embed data from sensors 1 and 2 onto their respective compressive-measurement spaces. We propose that, for such scenarios, a compressive data-fusion framework can be described by simply stacking compressive-projection measurements from sensors 1 and 2, and performing a final data-independent random projection, $P_3 : \mathbb{R}^{k_1 + k_2} \rightarrow \mathbb{R}^{k_3}$, where $k_1 \ll d_1$, $k_2 \ll d_2$, and $k_3 \ll k_1 + k_2$. $P_1$ and $P_2$ are compressive random measurement matrices, while $P_3$ is a data-agnostic orthonormal random projection that can be used at the classifier-level. $P_3$ is helpful in further reducing the dimensionality of the underlying analysis task. This is particularly relevant when using a Bayesian classification approach and the resulting dimensionality of the ensemble of compressive measurements is large (e.g., when the dimensionality of the projected subspaces is not small compared to the number of training samples available, or when the number of sensors in the ensemble is large).

We assert that, as a direct consequence of the JL-Lemma, subspaces produced through such compressive data fusion will retain the underlying discriminative information from each sensor. Specifically, since the projections $P_1$, $P_2$, and $P_3$ preserve locality, the underlying Bayesian error in the dataset will be approximately retained under the embedding.

To provide some intuitive insights, consider the case where, at pixel $i$, sensor 1 provides a measurement that is highly discriminative (e.g., some distance metric to the mean values of all but one class is high), while sensor 2 provides a measurement that is not strongly discriminative (e.g., wherein the distance to all classes is about the same). Compressive projections of these measurements will retain this separation structure, and using the compressively fused subspace will yield a feature vector that is discriminative. Hence, a compressively fused subspace will present the ability to offer the best of all sensors in the ensemble.

Consider the space $\mathbb{R}^{k_3}$ formed by fusing compressive measurements, such that $\{x_i\}_{i=1}^N$ are $N$ training pixels for any one class in the image, $\{c_i\}_{i=1}^N$ the corresponding mode (mixture/component) labels, and $x_i = P_3(P_1x_1^i | P_2x_2^i)$ where $|$ denotes a vertical stacking operation between two vectors (i.e., $x | y = [x^T, y^T]^T$). One can infer class-conditional statistics of the data by assuming that it was generated by a Gaussian mixture model (GMM) described as:

$$c_i | \pi \sim \text{Multinomial}(\pi), \quad x_i | \{c_i = k\} \sim \text{Gaussian}(\theta_k), \quad k = 1, 2, \ldots, K,$$

where $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$, and $\theta_k$ stands for $(\mu_k, \Sigma_k)$ which are the mean vector and covariance matrix of each Gaussian distribution, respectively. Then the probability density function (pdf) can be written as the sum of $K$ Gaussian components (mixtures):

$$p(x | \omega_i) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k),$$

where $\sum_{k=1}^{K} \pi_k = 1$, and $\omega_i$ indicates class $i$.

In the absence of prior knowledge of the number of mixtures, an infinite GMM (IGMM) [17] assumes it to be $\infty$, giving us the following model:

$$\pi | \alpha \sim \text{Stick}(\alpha), \quad \theta_k \sim \mathcal{H},$$

where $\pi | \alpha \sim \text{Stick}(\alpha)$ stands for:

$$\beta_k \sim \text{Beta}(1, \alpha), \quad \pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l), \quad k \rightarrow \infty,$$

and $\theta_k \sim \mathcal{H}$ is a shorthand for:

$$\mu_k \sim \mathcal{N}(\mu, \Sigma_k/r), \quad \Sigma_k \sim \text{InverseWishart}(S^{-1}, \nu).$$

Here, $\mu$ is the mean, $r$ is the relative precision for $\mu_k$, $S^{-1}$ is the scale matrix, and $\nu$ is the number of degrees of freedom for $\Sigma_k$. Also, the Gaussian and inverse-Wishart distributions are chosen for mathematical convenience [17], because they
are conjugate priors for the Gaussian distribution, which is instrumental in deriving a closed-form solution for the posterior distribution for \( C = \{ c_i \}_{i=1}^{N} \).

In the inference model, given observations, our goal is to infer the latent variables \( C = \{ c_i \}_{i=1}^{N} \) and \( \Theta = \{ \theta_k \}_{k=1}^{K} \) from their posterior distribution, which, however, cannot be computed analytically. A Markov Chain Monte Carlo (MCMC) method can be used to estimate the posterior. We do so by Gibbs sampling, a MCMC method which iteratively samples each variable conditioned on the current values of all the other variables. Upon convergence, these samples will well approximate the posterior distribution \([17]\). Specifically, \( \theta_k (k = 1, \ldots, K) \) is sampled from:

\[
P(\theta_k | C, X, \Theta_{-k}, \pi, \alpha) \propto \prod_{c_i = k} P(x_i | c_i, \theta_k) P_H(\theta_k)
\]

(8)

where \( \Theta_{-k} = \{ \theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots, \theta_K \} \), and \( P_H(\theta_k) \) is the probability of \( \theta_k \) under \( H \). Similarly, \( c_i (i = 1, \ldots, N) \) is sampled from:

\[
P(c_i = k | C_{-i}, x_i, \Theta, \pi, \alpha) \propto P(x_i | c_i, \Theta) P(c_i = k | C_{-i}, \alpha)
\]

(9)

where the prior

\[
P(c_i = k | C_{-i}, \alpha)= \begin{cases} \frac{n_{k-1}}{N-1+\alpha}, & c_i \in C_{-i}, \\ \frac{1}{N-1+\alpha}, & c_i \notin C_{-i}. \end{cases}
\]

(10)

such that \( C_{-i} = \{ c_1, \ldots, c_{k-1}, c_{k+1}, \ldots, c_N \} \) refers to all indices except \( i \), and \( n_{k-1} \) refers to the number of observations assigned to the \( k \)th cluster, excluding the \( i \)th observation \( x_i \).

During the training process, we infer the model parameters given the training dataset \( X \) per class, using Gibbs sampling. We can use this model to construct class-conditional likelihoods, based on the inferred parameters. During testing/classification, given a new sample \( x^* \), we can compute the likelihoods of \( x^* \) conditioned on each class, which can then be used in decision making as

\[
\omega_{\text{predicted}} = \arg \max_i p(x^* | \omega_i).
\]

(11)

3. EXPERIMENTAL SETUP AND RESULTS

The “University of Houston” (UH) dataset\(^1\) used in this work was acquired by the National Center for Airborne Laser Mapping (NCALM) over the University of Houston campus and the neighboring urban area. This dataset contains a hyperspectral image and a LiDAR point cloud for this area.

Hyperspectral imagery is a volumetric imagery dataset that captures spectral reflectance profiles per pixel in the scene. On the other hand, traditional discrete-return LiDAR systems are active sensors that generate a point cloud wherein each point represents peaks of reflected laser returns, with the goal of localizing and characterizing shapes of objects. Cutting-edge LiDAR systems are capable of recording the entire shape (waveform) of the return pulse—it is reasonable to assume that such full-waveform data provides more information for analysis tasks as compared to point-cloud data.

In the absence of collocated hyperspectral imagery and full-waveform LiDAR data, we simulate full-waveform LiDAR data from discrete-return point clouds. We simply grid LiDAR point clouds onto a reference grid formed by the hyperspectral image, resulting in a multi-sensor dataset that represents co-registered hyperspectral signatures and a proxy-waveform per pixel in the image. The waveform is approximated by binning elevations in a cell to yield a histogram that approximates average intensity as a function of elevation per pixel, akin to a true waveform. Both images in this dataset have a spatial resolution of 2.5 m. The hyperspectral image contains 144 spectral bands over the 364- to 1046-nm wavelength range. The corresponding co-registered proxy-waveform LiDAR feature associated with each pixel has 80 bands whose values correspond to elevation in meters above sea level. 200 samples per class are chosen for testing, while the number of training pixels per class is varied from 30 through 120. There are 15 classes in this scene. The proxy-waveform LiDAR data is used as a case study to simulate a multi-sensor imaging framework that can, in principle, be acquired via compressive random projections, and as a proxy for real-waveform LiDAR data—in a scenario with real waveform LiDAR collected simultaneously with hyperspectral imagery, a similar multi-sensor dataset can be obtained by appropriate georeferencing of the LiDAR waveforms. Fig. 1 shows the full-waveform features within the range from 5 to 40 m since the waveforms are all zero out of this range.

To quantify the discriminative capacity of compressively fused subspaces, we employ Bayesian inference statistical learning and classification of multi-sensor imagery. We note an additional benefit of compressive data fusion—it implicitly facilitates dimensionality reduction, making it possible to apply Bayesian approaches to image analysis in the measurement domain directly. The final dimensionality after pro-

\(^1\)http://hyperspectral.ee.uh.edu/?page_id=459
Finally, abilities through a logarithmic opinion pool (LOGP) [18, 19].

measurement domain of each source, and fuse posterior prob-

our classification strategy individually in the compressive-

analysis.

Fuse-PD

compressive measurements prior to reconstruction or image

thereby underscoring the motivation for performing at-sensor

Table 1. Overall classification accuracy (0.0–1.0) when using hyperspectral data, waveform LiDAR data, compressive projection after feature fusion (Fuse-SP), decision fusion after compressive projection (Fuse-PD), and compressive fusion

<table>
<thead>
<tr>
<th># Training Samples</th>
<th>CS-Hyperspectral</th>
<th>CS-LiDAR</th>
<th>Fuse-SP</th>
<th>Fuse-PD</th>
<th>Compressive Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.814 (0.018)</td>
<td>0.457 (0.017)</td>
<td>0.761 (0.022)</td>
<td>0.817 (0.018)</td>
<td>0.833 (0.012)</td>
</tr>
<tr>
<td>50</td>
<td>0.815 (0.021)</td>
<td>0.474 (0.017)</td>
<td>0.776 (0.020)</td>
<td>0.823 (0.019)</td>
<td>0.837 (0.013)</td>
</tr>
<tr>
<td>80</td>
<td>0.823 (0.016)</td>
<td>0.482 (0.016)</td>
<td>0.792 (0.015)</td>
<td>0.831 (0.016)</td>
<td>0.848 (0.011)</td>
</tr>
<tr>
<td>120</td>
<td>0.821 (0.013)</td>
<td>0.486 (0.011)</td>
<td>0.798 (0.017)</td>
<td>0.837 (0.014)</td>
<td>0.856 (0.009)</td>
</tr>
</tbody>
</table>

Finally, Compressive Fusion connotes compressive data fu-

sion of incoming compressive measurements, followed by

a single Bayesian classifier. We note that the last two ap-

proaches perform well, with the latter being marginally better

than the decision-fusion approach. In each case, we used an

IGMM-based model to infer class-conditional statistics that

were used for the resulting Bayesian classification. We report

classification accuracy as a function of the number of training

samples used per class in order to highlight the efficacy of

these compressive measurements for robust classification in

a small-sample-size scenario—facilitated by implicit dimen-

sionality reduction.

We also note another observation in these experiments—

several classes in the dataset required more than one Gauss-

ian mode in the IGMM representation, owing either to non-

Gaussian distributions, or, in the extreme scenario, to multi-

modal distributions (e.g., similar materials in sun and shade).

We noticed that compressive measurements often preserved

the number of modes required by IGMM for such classes in

the lower-dimensional subspace, verifying that these sub-

spaces are often highly suitable for direct image classification.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrated empirically that compressive

random projections actually constitute a natural framework

for multi-sensor fusion without the need for extensive signal

reconstruction. This is particularly relevant to imaging archi-

tectures wherein compressive sensing occurs at a pixel level

(e.g., spectrally), as opposed to spatial modulation via DMDs.

One such architecture is proposed in [16].

Although hardware implementations of such a compres-

sively sensed multi-sensor optical platform will pose

unique challenges (including georeferencing between multi-

ple views), we argue that the results obtained in this paper

demonstrate the immense promise that such systems hold for

robust image analysis—particularly in the context of simulta-

neous acquisition of complementary information. Further, by

staying in a compressive-measurement domain, we would not

only avoid large volumes of datasets, but also potentially yield

robust image analysis directly without the need for explicit

signal reconstruction. Of course, if signal reconstruction is a

requirement in a workflow (e.g., for visualization of spectra),

a variety of reconstruction algorithms exist that can ensure

a robust recovery of spectral content from such compressive

measurements. In ongoing work, we are exploring the ben-

efit of such a data-fusion paradigm with collocated visible,

shortwave-infrared, and thermal-infrared imagery that rep-

resent an important scenario in geospatial imaging owing to

different imaging-hardware constraints when sensing in these

three very different parts of the electromagnetic spectrum.

Finally, we will also extend this analysis to emerging pushb-

room and whiskbroom CS (e.g., [16]) architectures wherein
different pixels in a pushbroom linear detector may undergo
different projections.
5. REFERENCES


[18] H. Wu and S. Prasad, “Infinite Gaussian mixture models for robust decision fusion of hyperspectral imagery and full waveform LiDAR data,” in IEEE Global Conference on Signal and Information Processing, Austin, TX, December 2013, pp. 1025–1028.