Multimodal Data Integration Under Uncertainty

David Stracuzzi, David Robinson, Matthew Peterson,
Stephen Dauphin, Maximillian Chen, and Robert Riley

Sandia National Laboratories, Albuquerque, NM

SAND2016-1761C

Background

As platforms for sensor data collection become increasingly ubiquitous, people increasingly rely on the data for decision making in problems ranging from climate change analysis to intelligence gathering. In many cases, no one sensor provides definitive insight into a problem. In this work we study methods for combining data from different sensors and modalities in an effort to better support downstream decision making. We present preliminary results from the combination of three different sensors, optical, LiDAR, and polSAR, and look ahead toward using the uncertainty distributions to assess the relative contribution of each source to the final result.

Problem: Existing methods for sensor data integration often ignore uncertainty, leaving important information locked in the data.

Goal: Create sensor and domain-independent methods for extracting uncertainty from data.

Approach

1. Registration

1.1 Co-register samples from each sensor by geolocation.

2. Preprocessing

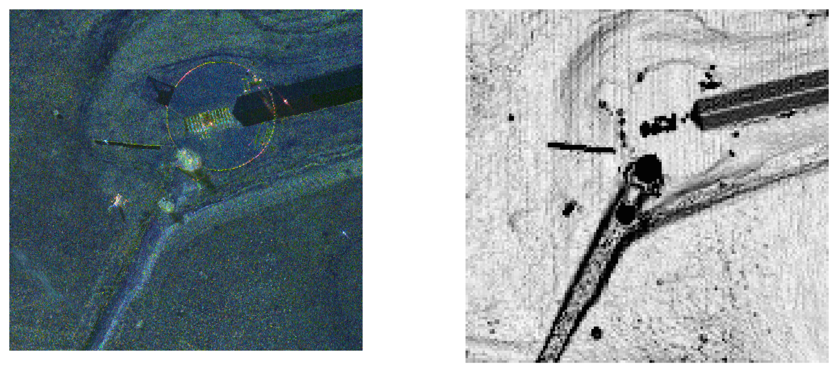
2.1 Polarimetric decomposition of SAR data

G4U: surface, dihedral, volumetric, helical, and total power (response)

H/A/alpha: entropy, anisotropy, scattering

2.2 Calculate local gradients from lidar data

2.3 Merge data from each source into a single feature vector (per pixel)



PolSAR G4U   Lidar Gradient

3. Data Analysis

3.1 Separate n pixels into K clusters via Gaussian mixture model

$$\mathcal{L}_{MIX}(\boldsymbol{\theta},\boldsymbol{\tau}|\mathbf{y}) = \prod_{i=1}^{n}\sum_{k=1}^{G}\tau_k f_k(\mathbf{y}_i|\theta_k)$$

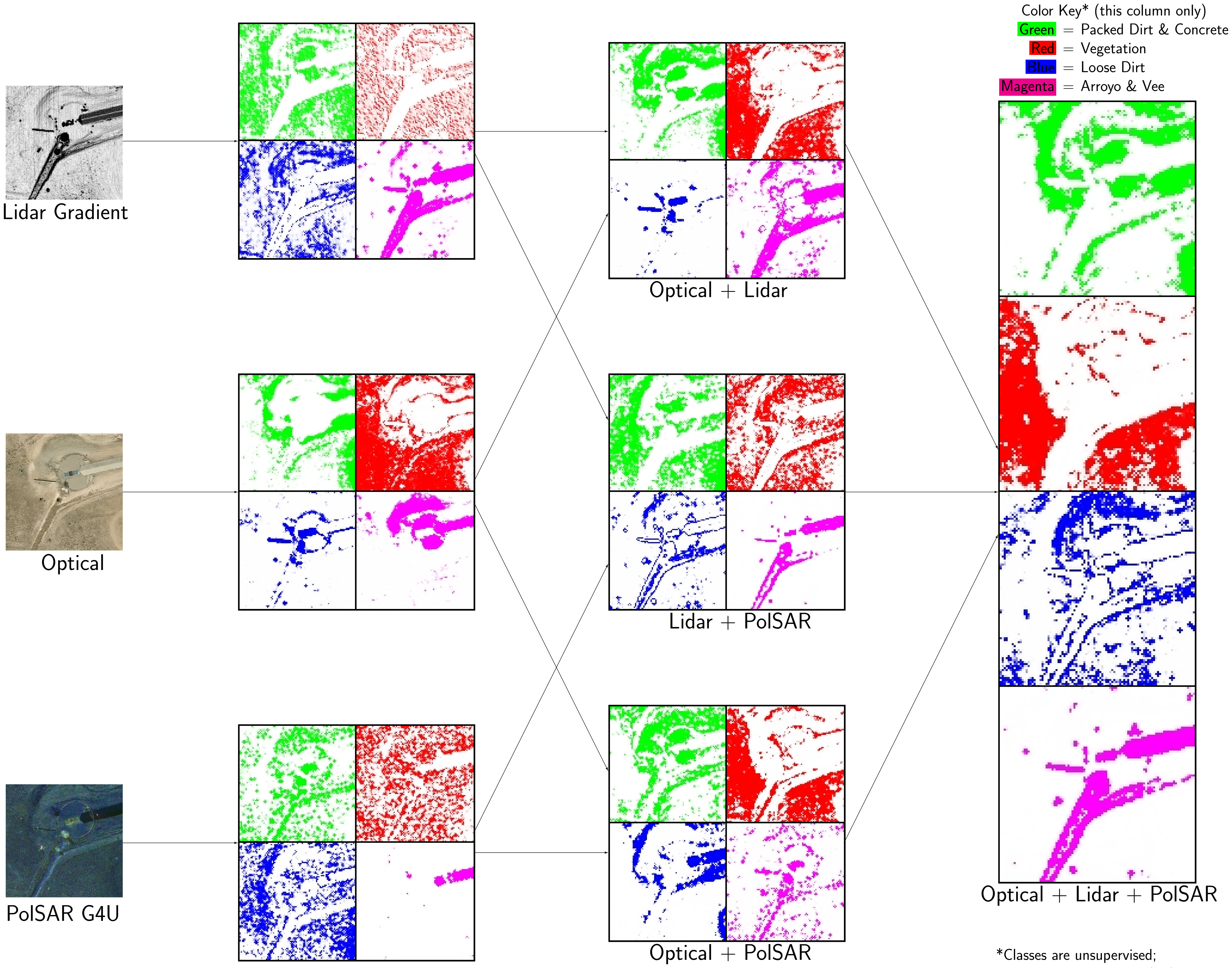
3.2 Perform hierarchical agglomeration to obtain classifications for labels l

$$\max \mathcal{L}_{CL}(\boldsymbol{\theta},\mathbf{l}|\mathbf{y}) = \prod_{i=1}^{n} f_{l_i}(\mathbf{y}_i|\theta_{l_i})$$

3.3 Apply expectation maximization to estimate model parameters

3.4 Compute the Bayesian Information Criterion (BIC) for each model. Select model with maximum BIC value (trades model complexity against data fit).

Preliminary Results



*Classes are unsupervised; semantic names are author's guess

Next Steps

Alternate Approaches: Distance weighted Chinese restaurant process

- Better accounting for spatial relationships
- Simplifies extraction of uncertainty distributions

Uncertainty Analysis:

- Models are underspecified — many possible solutions
- Extract uncertainty distributions for class probabilities
- Separate from class *confusion* (probability distrs.)

Piecewise Analysis:

- Ideally, analyze sources separately, then 'convolve'
- Computationally efficient; simplifies later analyses
- *Challenge:* Handle differences in unsupervised class semantics across sources

Value of Information:

- Identifies local contribution of each sensor / source
- Used to optimize data collection and analysis
- Need to identify appropriate measures

References

Shane R. Cloude and Eric Pottier.
An entropy based classification scheme for land applications of polarimetric SAR.
IEEE T. Geoscience and Remote Sensing, 35(1):68–78, 1997.

Chris Fraley and Adrian E. Raftery.
Model-based clustering, discriminant analysis, and density estimation.
Journal of the American Statistical Association, 97:611–631, 2002.

Gulab Singh, Yoshio Yamaguchi, and Sang Eun Park.
General four-component scattering power decomposition with unitary transformation of coherency matrix.
IEEE Transactions on Geoscience and Remote Sensing, 51(5):3014–3022, May 2013.

Acknowledgements

Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy National Nuclear Security Administration under contract DE-AC04-94AL85000.