Multiclass Continuous Correspondence Learning

Brian D. Bue
(Brian.D.Bue@rice.edu)
Electrical & Computer Engineering
Rice University, Houston, TX

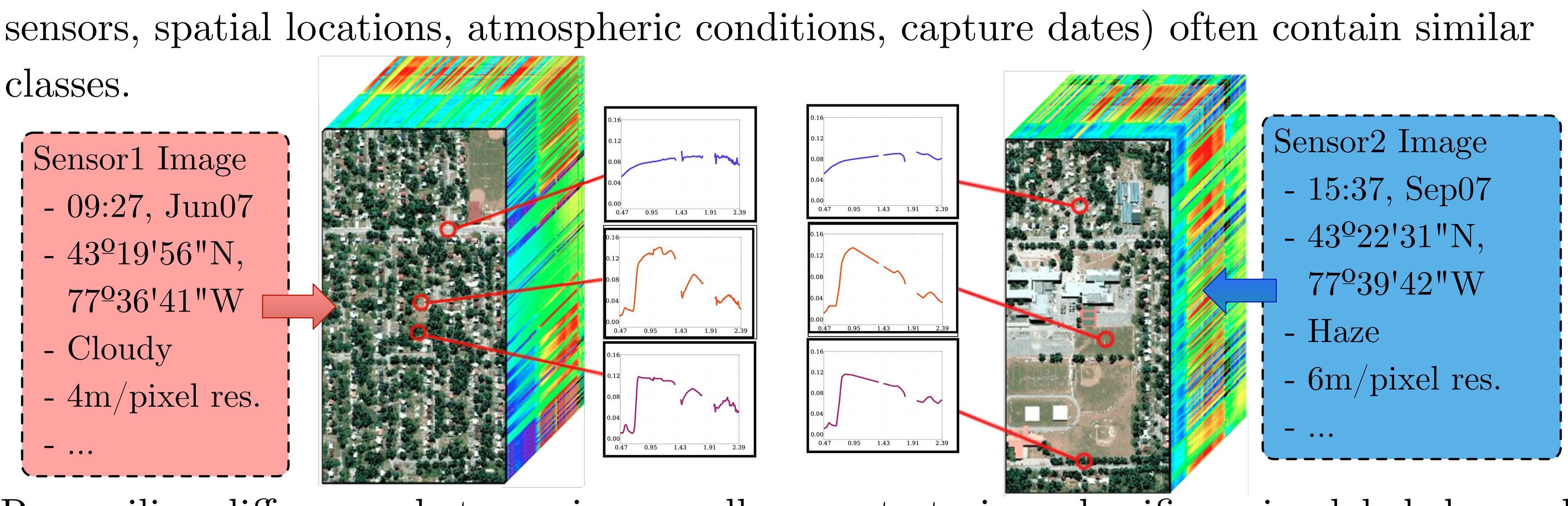
David R. Thompson
(David.R.Thompson@jpl.nasa.gov)
Jet Propulsion Laboratory
California Institute of Technology, Pasadena, CA

1. Abstract

We extend the Structural Correspondence Learning (SCL) domain adaptation algorithm of Blitzer et al. [1] to the realm of continuous signals. Given a set of labeled examples belonging to a "source" domain, we select a set of unlabeled examples in a related "target" domain that play similar roles in both domains. Using these "pivot samples," we map both domains into a common feature space, allowing us to adapt a classifier trained on source examples to classify target examples. When between-class distances are relatively preserved across domains, we can automatically select target pivots to bring the domains into correspondence.

2. Motivating Application: Domain Adaptation for Hyperspectral Imagery

Hyperspectral images include per-pixel measurements in many narrow bands of the electromagnetic spectrum. Images captured under differing conditions (e.g., different sensors, spatial locations, atmospheric conditions, capture dates) often contain similar classes.



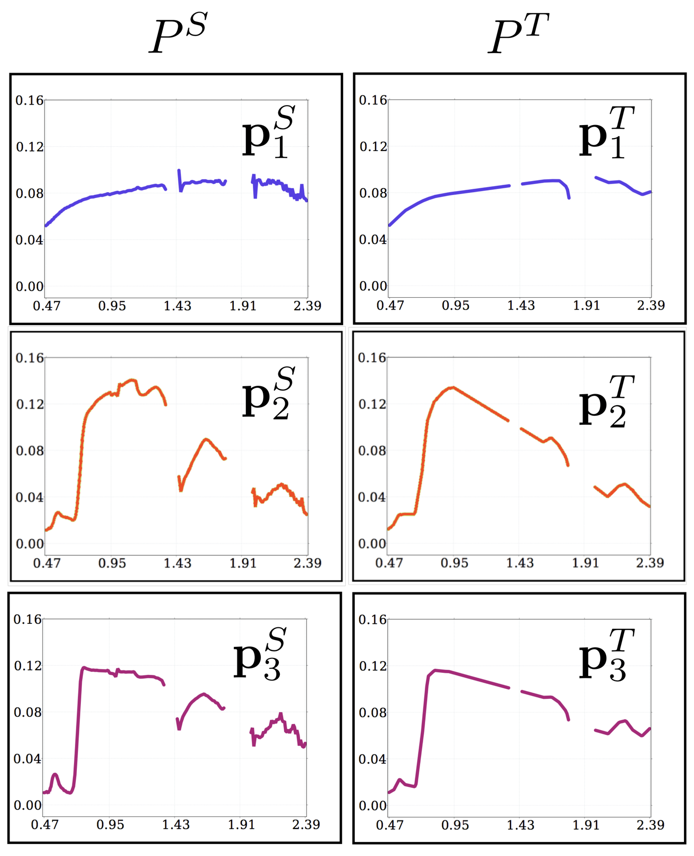
Reconciling differences between images allows us to train a classifier using labeled samples from one image (source) to label other, similar (target) images.

3. Reconciling Domains using Pivot Samples

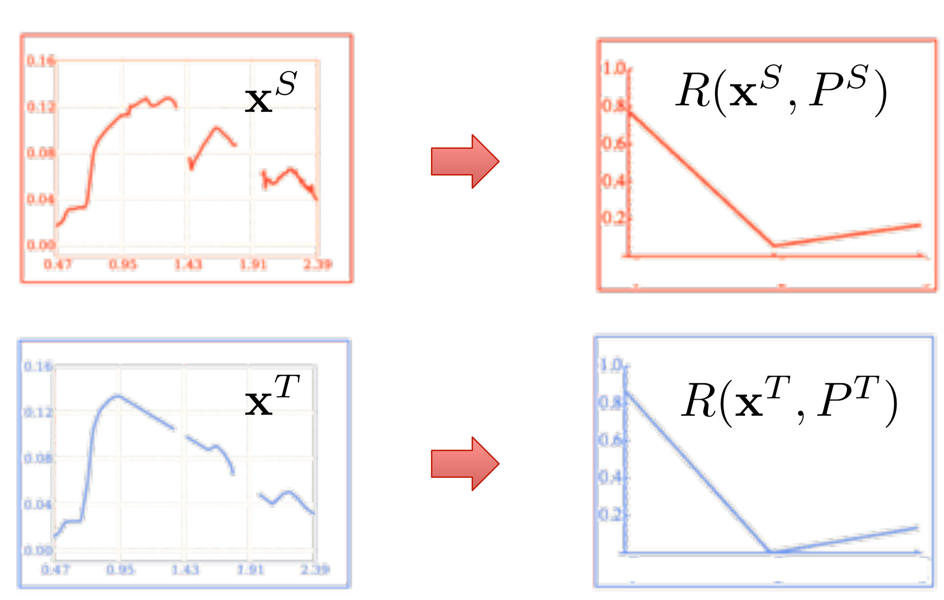
Transform sample $\mathbf{x}$ w.r.t. pivot samples $\mathbf{p}_i \in P$, distance measure $d(\cdot,\cdot)$ via:

$$R(\mathbf{x}, P) = \left(\frac{\mathrm{d}(\mathbf{x}, \mathbf{p}_1)}{\sum_{\ell=1}^{Q} \mathrm{d}(\mathbf{x}, \mathbf{p}_\ell)}, \ldots, \frac{\mathrm{d}(\mathbf{x}, \mathbf{p}_Q)}{\sum_{\ell=1}^{Q} \mathrm{d}(\mathbf{x}, \mathbf{p}_\ell)} \right) \quad (1)$$

Example: Given pivot samples...



...transform source ($\mathbf{x}^S$), target ($\mathbf{x}^T$) samples w.r.t. pivots...



...then train/classify in the pivot-based feature space (the "R-Space").

4. Multiclass Continuous Correspondence Learning (MCCL)

Goal: classify target samples in feature space defined by pivot samples with similar class relationships in both domains

Input: source data (X^S, Y^S), target data X^T, source pivots P^S
Output: predicted target labels Y^T
1. Select Q source pivots $\mathbf{p}_i^S \in P^S$ from (X^S, Y^S)
2. Select target pivots $\mathbf{p}_i^T = \mathbf{x}_\ell^T \in X^T$ according to

$$\ell = \underset{j}{\operatorname{argmin}} \, \|R(\mathbf{p}_i^S, P^S) - R(\mathbf{x}_j^T, P^S)\|, \; j \in \{1, \ldots, M\} \quad (2)$$

3. Train predictor $p : R(\mathbf{x}, P) \to Y$ using $R^S = (R(\mathbf{x}_i^S, P^S))_{i=1}^N$
Return: predictions $Y^T = (p(R(\mathbf{x}_i^T, P^T)))_{i=1}^M$, $\mathbf{x}_i^T \in X^T$

5. Measuring Pivot Set Quality using H-Divergence [2]

Goal: measure the quality of the pivot sets w.r.t. how well they reconcile the source and target domains in the R-Space (eq. 1).

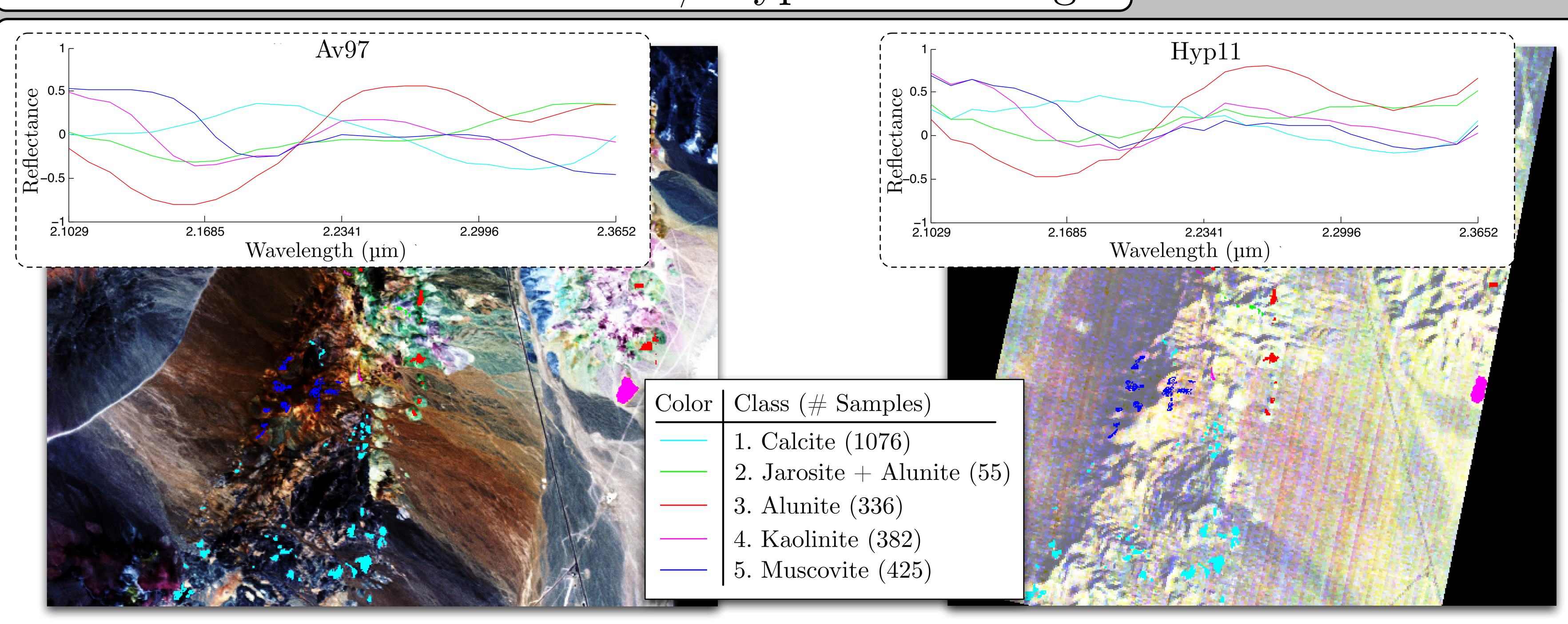
Input: pivot sets P^S, P^T, each of length $Q = \sum_{k=1}^K Q_k$
Output: divergence score H
1. **for** $k = 1$ **to** K **do**
2. Define label vector $y = ((-1)_{i=1}^{Q_k}, (1)_{i=1}^{Q_k})$ for class k pivots.
3. Train binary predictor $h : R(\mathbf{p}, P) \to \{-1, 1\}$.
4. Calculate divergence between class k source and target pivots

$$H_k = \frac{1}{2Q_k} \left(\sum_{i=1}^{Q_k} \mathcal{I}(h(\mathbf{p}_i^S, P^S) = y_i) + \sum_{i=Q_k+1}^{2Q_k} \mathcal{I}(h(\mathbf{p}_i^T, P^T) = y_i) \right)$$

Return: $H = \frac{1}{K} \sum_{i=1}^K H_k$ (3)

Interpretation: small H $\Longrightarrow$ domains well reconciled.

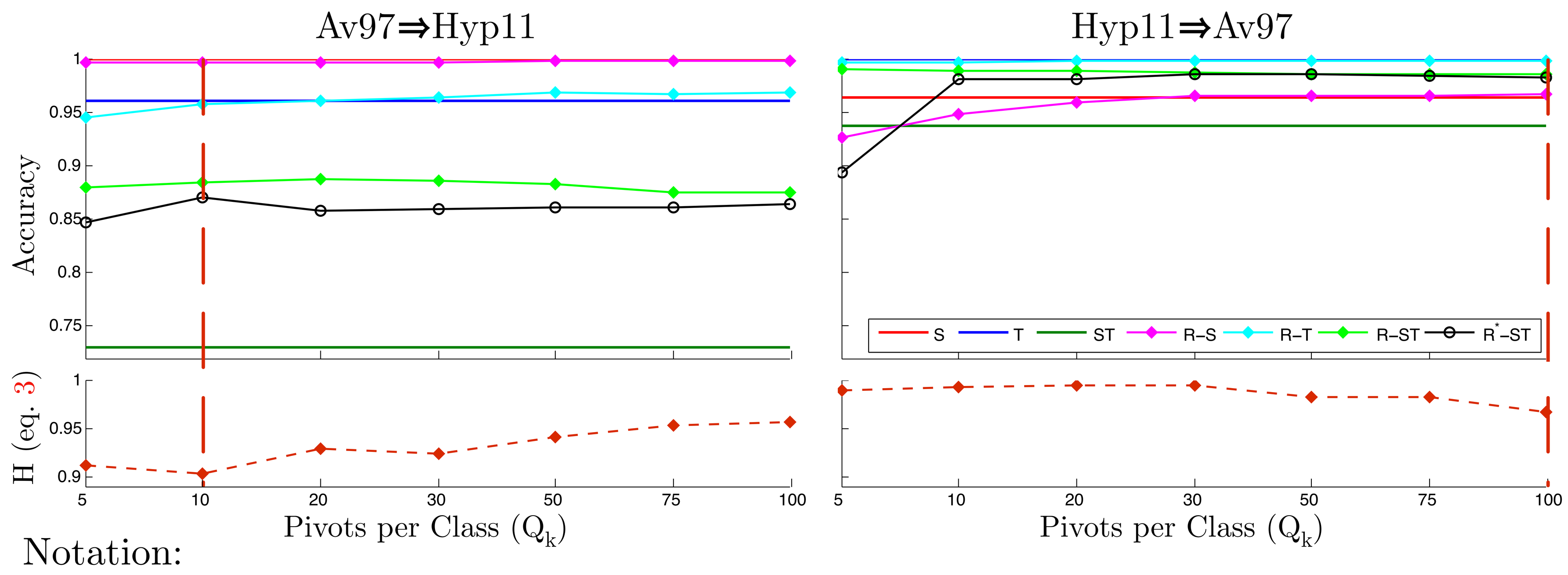
6. Data and Results: AVIRIS97 / Hyperion11 Images



L: AVIRIS 1997 (Av97) sample locations and whitened class means
R: Hyperion 2011 (Hyp11) sample locations and whitened class means

Each sample (pixel) is a 28 dimensional, L^2 normalized, whitened vector
Av97 class samples linearly separable, Hyp11 classes nonlinearly separable (esp. classes 1,2)
Source pivot samples = Q_k nearest samples to each class mean
Target pivot samples selected using Equation 2.

Source ⇒ Target Accuracy vs. Pivots per Class



Notation:
- S, T: Source, Target Accuracy
- ST: Source-to-Target Accuracy (no domain adaptation)
- R-S, R-T: "R-Space" Source, Target Accuracy
- R-ST: "R-Space" Source-to-Target Accuracy, pivots from labeled Target data
- R*-ST: "R-Space" Source-to-Target Accuracy, pivots from unlabeled Target data

Dashed red vertical line indicates # pivots selected via minimizing equation 3.
All accuracies produced with a linear multiclass SVM [3], 10 fold cross-validation

7. Observations

- We observe maximum classification accuracy with the pivot set with the minimum divergence.
- Domain adaptation accuracy (R*-ST) closely approaches target (T) accuracy with informative pivots (Hyp11⇒Av97 scenario)
- Adaptation challenging when distances (i.e., class separability) not well preserved across domains (Av97⇒Hyp11 scenario)
- R-S, R-T accuracy equivalent to S, T accuracy when pivots well-selected.
  - Can eliminate some degenerate pivot sets by filtering cases where R-S accuracy < S accuracy

8. Conclusions

- We described a multiclass extension for continuous feature spaces (MCCL) to the Structural Correspondence Learning algorithm.
- Capturing structured class relationships with pivot samples allows us to map source & target domains to a common space
- When between-class distances are preserved across domains, we can select pivots using unlabeled target data by minimizing the empirical H-divergence, yielding performance competitive to selecting pivots from labeled target data

9. Future Work

- Optimize pivot selection by selecting samples minimizing eq. 3
- Formalize theoretical limits of MCCL adaptation accuracy, particularly in multiclass scenarios

References

[1] Blitzer, J., McDonald R., & Pereira, F. Domain adaptation with structural correspondence learning. Proc: Conference on Empirical Methods in Natural Language Processing, 2006.
[2] Kifer, D., Ben-David, S., & Gehrke, J., Detecting change in data streams. Proc: Very large databases 2004.
[3] Chang, C.C & Lin C.J., LIBSVM: a library for support vector machines. ACM TIST, 2(3):27, 2011.

Acknowledgements

A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, with support from an AMMOS Multimission Operations Systems technology development program. U.S. Government Support Acknowledged.