AN EVALUATION OF CLASS KNOWLEDGE TRANSFER FROM SYNTHETIC TO REAL HYPERSPECTRAL IMAGERY

Brian D. Bue, Erzsébet Merényi
Electrical and Computer Engineering
Rice University, Houston TX
{bbue, erzsebet}@rice.edu

Beáta Csathó
Department of Geology
University at Buffalo, SUNY, Buffalo NY
bcsatho@buffalo.edu

ABSTRACT

Hyperspectral imagery provides ample signal content to identify and distinguish between spectrally similar, but compositionally unique, materials, but representative training samples for all materials in a given scene are often unavailable in remote sensing applications. We propose a technique which leverages training spectra of materials from one hyperspectral image (the source), for classifying another (target) hyperspectral image, allowing class discovery in the absence of representative training samples for some classes in the target image. By locating spectra of known material species in the source image which are also known in the target image, we derive a transformation that compensates for systematic spectral differences between the images, including atmospheric or seasonal effects, calibration differences, and noise. With this transformation – applied as a similarity measure – we can adapt a classifier trained on spectra from the source image, to classify the target image. We evaluate our technique between a pair of synthetic hyperspectral images with systematic differences in the spectral signatures of corresponding materials. Then we assess the feasibility of using synthetic imagery for training a classifier to identify spectral species in similar, real images, and show knowledge transfer results between a synthetic (HYDICE-type) source image and a real low-altitude AVIRIS target image of a complex urban area.

***Index Terms*—** domain adaptation, transfer learning, synthetic, AVIRIS, DIRSIG, HYDICE

1. HYPERSPECTRAL CLASS KNOWLEDGE TRANSFER

Spectra captured by modern hyperspectral imaging platforms provide ample signal content to identify spectrally similar but distinct materials. Unfortunately, in remote sensing applications, we rarely have representative samples to train a classifier to reliably classify all known material species in a given scene. Therefore, training a classifier using samples from other, similar images is an attractive option, but poses significant challenges. A classifier must be robust to systematic differences in source and target image spectra to accurately transfer knowledge of training (or "source") classes to unlabeled test (or "target") spectra. Such differences are caused by varying atmospheric and illumination condition, differing sensor types or sensor noise, differing capture conditions (e.g. differences in spectral/spatial resolution or capture geometry), and distortions resulting from image orthorectification, or atmospheric calibration. Such differences between images are often so dramatic that spectra with equivalent material compositions will appear (visually) quite different. Additionally, the material distributions of images captured at different regions will differ, and a classifier must be able to robustly flag samples which it cannot confidently classify.

There has been considerable interest in recent years in using synthetic hyperspectral imagery for algorithm testing and development [1], [2]. One largely unexplored avenue is to use synthetic hyperspectral imagery in class knowledge transfer scenarios – by adapting a classifier trained on fully-labeled synthetic imagery for class knowledge transfer to real data, we know training data corresponds to ground truth material samples. An additional benefit in assessing synthetic-to-real class knowledge transfer is that it can potentially identify unrealistic properties of synthetic imagery, which can provide information to refine the image generation process.

In this work, we build upon our previous research in hyperspectral class knowledge transfer [3] and present an extension to our RelTrans algorithm which automatically determines a class likelihood threshold, allowing for robust class discovery in target imagery. We first evaluate our algorithm on a pair of synthetic (HYDICE-type) hyperspectral images with systematic differences in spectral signatures, and then give results for real AVIRIS image spectra with a classifier trained with samples from synthetic imagery.

2. EVALUATING CLASS KNOWLEDGE TRANSFER

The RelTrans algorithm [3] defines a mapping from training spectra from a source image $\mathbf{S}$ to spectra from an unlabeled target image $\mathbf{T}$ through a set of n_C labeled correspondences C. Each correspondence consists of a source and target spectrum of the same material species (i.e., belonging to the same source class). RelTrans calculates class likelihood values for target spectra according to their relative similarities to k_S source and target correspondence class means, and flags target spectra as "unknowns" if their likelihood values are below a threshold τ, allowing RelTrans to detect new classes in target imagery. This transformation is applicable as a similarity measure in any classification procedure if such correspondences are available. For additional details about RelTrans, see [3].

Here, we present the RelThresh procedure (Algorithm 1), which selects an optimal threshold for class discovery with respect to the correspondences. RelThresh chooses the largest τ-value that maximizes RelTrans prediction accuracy on the (labeled) target correspondence spectra p_{best}, while ensuring no correspondence spectra are flagged. RelTrans selects this threshold by searching the range of minimum and maximum RelTrans class membership likelihood values on the set of target spectra (i.e., $\tau_{\text{range}} = \left[\min(\mathbf{L}^T), \max(\mathbf{L}^T)\right]$, where $\mathbf{L}^T$ is the RelTrans target likelihood matrix described in [3]). This allows RelTrans to find a threshold for target spectra according to similarities in the target image relative to source classes, while not discarding crucial correspondence spectra.

We provide a case study in adapting the minimum Euclidean distance classifier with RelTrans, and compare results before and after the RelTrans transform (in subsequent sections, we denote results evaluated before the transform as "MinDist" and results evaluated after the transform as "RelTrans"). We first assess the quality of the mapping between source and target images by classifying spectra from classes common to both images. If the knowledge transfer algorithm performs well, we should see similar performance classifying spectra across images ("inter-image" classification) as within

Algorithm 1 RelThresh

Input: $n_C \times k_S$ source and target correspondence likelihood matrices $\{\mathbf{L}^{cS}, \mathbf{L}^{cT}\}$, length n_C label vector $\mathbf{y}^C$, $y_i^C \in [1, k_S]$. τ search range $\tau_{\text{range}} = [\min(\mathbf{L}^T), \max(\mathbf{L}^T)]$. Total τ steps n_{step}

Output: RelTrans threshold τ_{best}.

Set best prediction $p_{\text{best}} = -\infty$, current threshold $\tau_{\text{cur}} = \max(\tau_{\text{range}})$, step size $\tau_{\text{step}} = \frac{\max(\tau_{\text{range}}) - \min(\tau_{\text{range}})}{n_{\text{step}}}$

while $\tau_{\text{cur}} > \min(\tau_{\text{range}})$ **do**

  Initialize $p \leftarrow 0$

  for $i = 1$ **to** n_C **do**

    if $\max\limits_j \mathbf{L}^{cS}(i,j) > \tau_{\text{cur}}$ and $\max\limits_j \mathbf{L}^{cT}(i,j) > \tau_{\text{cur}}$ **then**

      if $\operatorname{argmax}\limits_j \mathbf{L}^{cT}(i,j) = y_i^C$ **then** $p \leftarrow p + 1$

  if $p > p_{\text{best}}$ **then** $p_{\text{best}} \leftarrow p$, $\tau_{\text{best}} \leftarrow \tau_{\text{cur}}$

  $\tau_{\text{cur}} = \tau_{\text{cur}} - \tau_{\text{step}}$

each image ("intra-image" classification). We use the term introduced by Daumé in [4], "**domain adaptation**" for this scenario. We assess the algorithm's ability to discover unknown classes in the second scenario, "**class discovery**," by classifying target spectra without representative source classes, which the algorithm should flag as unknowns. We sample 2000 pixels from source and target images via random stratified subsampling of chosen image classes. We manually select a maximum of 50 correspondence pixels for each source class that match target pixels well in terms of absorption features and expert knowledge. Results are averaged over five randomized 50/50 training/testing splits. We select the RelTrans similarity threshold using RelThresh, and for direct comparison, we flag the same number of farthest pixels as calculated by the MinDist classifier. We report accuracy scores on unflagged pixels only – i.e., if all samples from source classes are correctly classified, and all unknown samples are flagged, the reported classification accuracy is 100%. We consider incorrectly flagged pixels (i.e., flagged pixels belonging to source classes) misclassifications. We also provide the classification accuracies for the selected classes within each image ("intra-image" accuracy) to illustrate the MinDist classification performance when we sample training and testing samples from the same image.

3. EXPERIMENTS USING SYNTHETIC IMAGERY

First, we evaluate class knowledge transfer between two synthetically generated images. The synthetic images we study are generated with the RIT Digital Imaging and Remote Sensing Image Generation (DIRSIG) [5] model, and are both from the urban RIT "Megascene" [1]. Our target image for experiments in this section, which we denote DIRSIG^1 (previously described in [3] and [6]) is a 400x400 pixel, 4m/pixel resolution image, with 210 bands over 0.4-2.5 microns, modeled after the HYDICE [7] instrument. The second image, DIRSIG^2, also generated with the HYDICE sensor model, is a "cleaner" version of DIRSIG^1 with minimal atmospheric effects and less noisy illumination conditions (i.e., fewer shadow pixels). We convert image radiances to reflectances using the empirical line (ELM) method using the software package ENVI [8]. After atmospheric calibration, we remove noisy bands in the extreme short and long wavelengths and the water vapor saturation bands – leaving 160 of the original 210 bands for analysis – and perform illumination normalization by dividing each spectrum by its Euclidean norm. All subsequent plots of spectra are given in terms of wavelength (μm) vs. (Euclidean normalized) reflectances.

Domain Adaptation: We test the capability of the RelTrans transform to generalize to target spectra with systematic differences from source spectra by drawing training samples from two different versions of DIRSIG^2. The first, $\mathrm{DIRSIG}^2_{\mathrm{G}}$, we preprocess as we previously described. The second, $\mathrm{DIRSIG}^2_{\mathrm{B}}$, is the result of a poor ELM atmospheric calibration due to incorrectly pairing a "dark" and featureless radiance spectrum to a field reflectance spectrum corresponding to a different material with prominent absorption features. Figure 1 gives the mean signatures for correspondence spectra in images DIRSIG^1, $\mathrm{DIRSIG}^2_{\mathrm{G}}$ and the $\mathrm{DIRSIG}^2_{\mathrm{B}}$ image.

We summarize domain adaptation results in Table 1 (top). When we sample source spectra from image $\mathrm{DIRSIG}^2_{\mathrm{G}}$, we see equivalent MinDist and RelTrans accuracies, which is not surprising, given that the $\mathrm{DIRSIG}^2_{\mathrm{G}}$ and DIRSIG^1 spectra are nearly identical. When we sample source spectra from $\mathrm{DIRSIG}^2_{\mathrm{B}}$, we see major improvements when classifying RelTrans-transformed spectra over MinDist. MinDist tends to misclassify samples from the "Siding, Brick, Mix Brown, Fair" class (Figure 1, top, right) as "Wood, Stained, Red, Old, Weathered" (Figure 1, bottom, right). Visually, these classes are similar in image $\mathrm{DIRSIG}^2_{\mathrm{B}}$, but less so in DIRSIG^1. These classes are difficult to reconcile without leveraging subtle intra-image/intra-class relationships which RelTrans is able to exploit. Also, note (Table 1) the RelThresh procedure selects τ values that do not incorrectly discard any samples in both $\mathrm{DIRSIG}^2_{\mathrm{G}}$ and $\mathrm{DIRSIG}^2_{\mathrm{B}}$ cases.

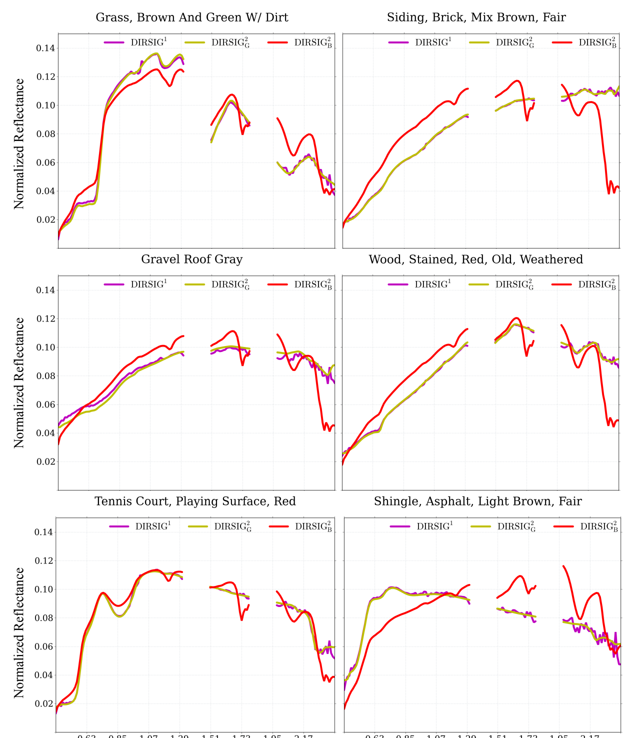

Fig. 1: Mean spectra for manually-selected correspondences between DIRSIG^1 (magenta), $\mathrm{DIRSIG}^2_{\mathrm{G}}$ (yellow) and $\mathrm{DIRSIG}^2_{\mathrm{B}}$ (red) images. Spectra from images DIRSIG^1 and $\mathrm{DIRSIG}^2_{\mathrm{G}}$ are similar after ELM atmospheric calibration, while those from the poorly calibrated image $\mathrm{DIRSIG}^2_{\mathrm{B}}$ are considerably distorted.

Class Discovery: Next, we evaluate algorithm performance when some target classes are not present in the source training spectra. We sample from five additional classes from the DIRSIG^1 image (Figure 2), and classify them using the same source classes shown in Figure 1. Table 1 (bottom) gives results for this scenario. Because accuracy here is limited by the number of source class samples, scores of 56% are the highest attainable without thresholding. With thresholding, when we sample source spectra from $\mathrm{DIRSIG}^2_\mathrm{G}$, we observe a significant improvement in classification accuracy by both MinDist and RelTrans, as distances between spectra are preserved across the two images. When source samples are taken from $\mathrm{DIRSIG}^2_\mathrm{B}$, MinDist achieves about 50% of the $\mathrm{DIRSIG}^2_\mathrm{G}$ scores due to considerable differences in *inter-image* spectral distances. RelTrans is unaffected by this change, since the systematic distortions in $\mathrm{DIRSIG}^2_\mathrm{B}$ do not alter *intra-image* distances, which RelTrans exploits for *inter-image* classification.

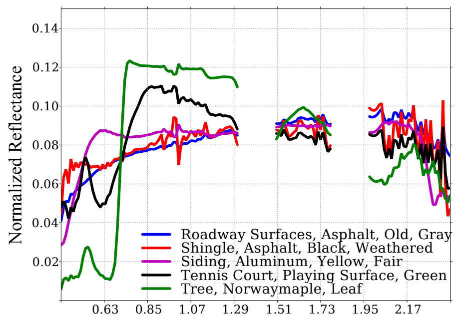

Fig. 2: Mean spectra of DIRSIG^1 target samples for evaluating class discovery. These spectra belong to different material classes as source classes (Figure 1) and should be flagged as unknowns.

Domain Adaptation	$\mathrm{DIRSIG}^2_\mathrm{G}(99)$ vs. $\mathrm{DIRSIG}^1(99)$	$\mathrm{DIRSIG}^2_\mathrm{B}(84)$ vs. $\mathrm{DIRSIG}^1(99)$
MinDist	99 / 99 [100]	73 / 73 [100]
RelTrans	100 / 100 [100]	100 / 100 [100]
Class Discovery	$\mathrm{DIRSIG}^2_\mathrm{G}(96)$ vs. $\mathrm{DIRSIG}^1(98)$	$\mathrm{DIRSIG}^2_\mathrm{B}(83)$ vs. $\mathrm{DIRSIG}^1(98)$
MinDist	56 / 93 [92]	24 / 50 [17]
RelTrans	56 / 99 [100]	56 / 99 [100]

Table 1: Domain adaptation (top table) and class discovery (bottom table) results for source images $\mathrm{DIRSIG}^2_\mathrm{G}$ and $\mathrm{DIRSIG}^2_\mathrm{B}$ vs. target image DIRSIG^1 before / after thresholding. Values in () give the "intra-image" MinDist classification accuracy within the corresponding image (i.e., training/testing samples taken from the same image). Values in [] give the percentage of correctly flagged unknown samples. No target samples belong to unknown classes in the domain adaptation scenario (i.e., no samples should be thresholded), while 439 of the 1000 target samples are unknowns in the class discovery scenario.

4. SYNTHETIC-TO-REAL KNOWLEDGE TRANSFER

Here, we aim to transfer material classes from the synthetic $\mathrm{DIRSIG}^2_\mathrm{G}$ image to a real low-altitude AVIRIS image. This represents the case when we wish to transfer class knowledge between two geographically distinct, but similar hyperspectral images. This image was captured on Nov 5, 1998, over Ocean City, MD [9], with spatial resolution of 4m/pixel. We manually select correspondence spectra, and sample target spectra from, the segmentation map described in [10] and [11]. As verified from field data, segments in this map represent distinct material types, and most of them are clearly (albeit non-uniquely) associated with objects – for instance: tennis courts, building rooftops, roads and parking lots.

Domain Adaptation: We select pixels from six segments (Figure 3) which correspond well in terms of spectral characteristics and expert knowledge of materials present in the $\mathrm{DIRSIG}^2_\mathrm{G}$ image. We match segment **C** (from the segmentation map provided in [11]) to "Tennis Court, Playing Surface, Green," **G** to "Roadway Surfaces, Asphalt, Old, Gray," **U** to "Roof Shingle, Asphalt, Brown and Red Blend, Fair," **L** to "Grass, Brown and Green w/ Dirt," **T** to "Roof, Gravel, Gray," and **f** to "Wood, Stained, Red, Old, Weathered." Note that we select these matches based on expert knowledge and spectral characteristics of materials, *not the objects to which the materials belong*. For instance, we know from field knowledge that segment **G** is made of rooftop materials with spectral features indicative of asphalt, and we pair it with the "Roadway Surfaces, Asphalt Old Gray" synthetic image class, as they share the same material composition.

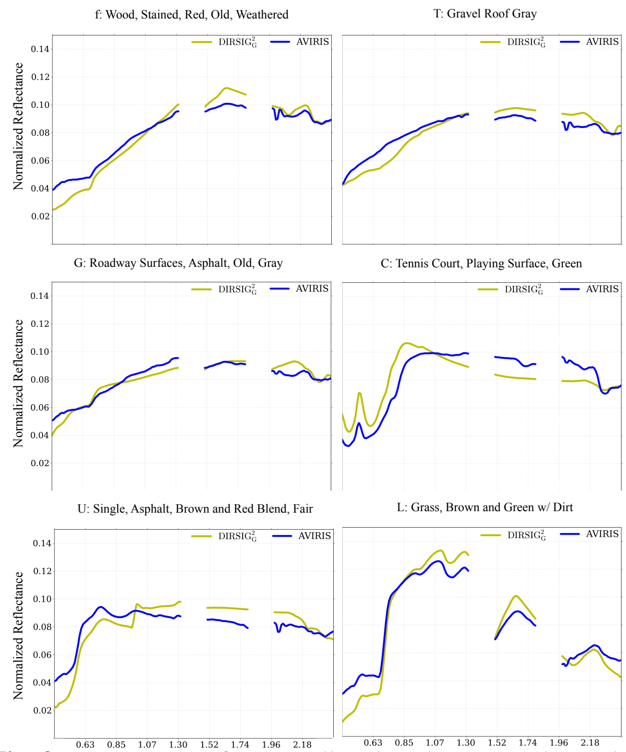

Fig. 3: Mean spectra for manually-selected correspondences between $\mathrm{DIRSIG}^2_\mathrm{G}$ (yellow), and AVIRIS (blue) images. Correspondence spectra are selected based on expert field-knowledge of material classes and similarities in absorption features.

Tables 2 (top) and 3 provide overall and per-class accuracy scores for this scenario, respectively. Again, we observe noteworthy performance increases with the RelTrans transformation. Moreover, RelTrans misclassifications are generally more intuitive than MinDist.. MinDist misclassifies all "T: Roof, Gravel, Gray" samples as either "f: Wood, Stained, Red, Old, Weathered", "G: Roadway Surfaces, Asphalt, Old, Gray", or "C: Tennis Court, Playing Surface, Green," whereas RelTrans correctly classifies 35% of the same pixels, with misclassifications falling into classes **G** and **f**, but not (dissimilar) class **C**.

Class Discovery: We select five AVIRIS material classes without strong correspondences to the $\mathrm{DIRSIG}^2_\mathrm{G}$ classes to evaluate synthetic-to-real class discovery capabilities. Figure 4 shows the mean spectra for these classes. Table 2 (bottom) gives classification results for class discovery. Here, the maximum accuracy with 5 unknown classes is near 54% without thresholding, and we see similar performance to the synthetic scenario in Section 3 both before and after the RelTrans transform.

With automatic threshold selection, RelTrans achieves compa-

Domain Adaptation	$\mathrm{DIRSIG}_{\mathrm{G}}^{2}$(98) vs. AVIRIS(83)	Class Discovery	$\mathrm{DIRSIG}_{\mathrm{G}}^{2}$(92) vs. AVIRIS(82)
MinDist	45 / 45 [100]	MinDist	26 / 31 [77]
RelTrans	72 / 72 [100]	RelTrans	43 / 74 [100]

Table 2: Domain adaptation (left table) and class discovery (right table) results for source image $\mathrm{DIRSIG}_{\mathrm{G}}^{2}$ vs. target image AVIRIS before / after thresholding. Values in () give the "intra-image" MinDist classification accuracy for the corresponding image, and values in [] give correctly flagged unknown sample percentages. 461 of 1000 target samples are unknowns in the class discovery scenario.

Segment Label: Material Class	MinDist	RelTrans
C: Tennis Court, Playing Surface, Green	7	100
G: Roadway Surfaces, Asphalt, Old, Gray	55	26
L: Grass, Brown and Green w/ Dirt	63	93
T: Gravel Roof Gray	0	100
U: Shingle, Asphalt, Brown and Red Blend, Fair	57	100
f: Wood, Stained, Red, Old, Weathered	28	83

Table 3: Per-class accuracy for $\mathrm{DIRSIG}_{\mathrm{G}}^{2}$ source classes and AVIRIS target classes in the domain adaptation scenario.

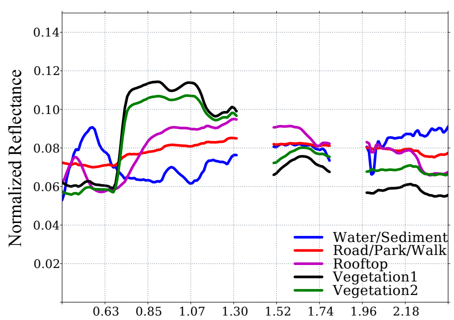

Fig. 4: AVIRIS class mean spectra used in class discovery evaluation. These spectra correspond poorly to source classes (Figure 1) and should be flagged as unknowns.

rable classification accuracy (74%) to the domain adaptation case when all material classes are known (72%), correctly thresholding 100% of all unknown samples. Table 4 gives the per-class percentages of unknown samples which each classifier correctly flags as unknowns. Because the unknown target classes are dissimilar from the source classes, we see high accuracies in both cases. However, MinDist regularly misclassifies AVIRIS material class "Vegetation2" as "Tennis Court, Playing Surface, Green" due to similar absorption features common to both classes. MinDist classifies "Road/Park/Walkway" and "Rooftop" as the "Roadway Surfaces, Asphalt, Old, Gray" $\mathrm{DIRSIG}_{\mathrm{G}}^{2}$ material class – which are misclassifications that RelTrans correctly resolves, but may be semantically acceptable, given their similar material classes.

5. DISCUSSION AND FUTURE WORK

It may be beneficial to employ other similarity measures other than the Euclidean distance in the RelTrans transform, as the Euclidean distance may fail to capture distinctions between absorption features. Our recently proposed Continuum-Intact/Continuum-Removed (CICR) measure [11] and an adaptive extension of CICR that we are presently working on, may better capture these distinctions, and may further improve class knowledge transfer results.

We are experimenting with approaches to improve the the RelTrans correspondence mapping. One potential technique described

	Water / Sediment	Road / Park / Walk	Rooftop	Vegetation1	Vegetation2
MinDist	100	90	78	100	44
RelTrans	100	100	100	100	100

Table 4: Percentage of correctly flagged unknowns for AVIRIS class discovery scenario.

by Balcan et al. in [12] calculates an orthonormal basis that spans the space of the correspondences. This eliminates highly correlated / redundant correspondence classes which can reduce the quality of the RelTrans transform.

Acknowledgements: We would like to thank David Pogorzala and Prof. John Kerekes at the RIT Digital Imaging and Remote Sensing (DIRS) Laboratory for their gracious assistance in generating the DIRSIG image data used in this work. We would also like to thank Dr. David Thompson and Dr. Kiri Wagstaff of the NASA Jet Propulsion Laboratory for their helpful input on topics related to metric learning and class knowledge transfer.

6. REFERENCES

[1] C Salvaggio, L Smith, and E Antoine, "Megacollect 2004: hyperspectral collection experiment of terrestrial targets and backgrounds of the RIT megascene and surrounding area (Rochester, New York)," *Proceedings of SPIE*, p. 555, 2005.

[2] M Parente, JT Clark, A Brown, and J Bishop, "Simulation of the image generation process for CRISM spectrometer data," *IEEE WHISPERS 2009*, pp. 1–4, Jun 2009.

[3] B Bue and E Merényi, "Using spatial correspondences for hyperspectral knowledge transfer: evaluation on synthetic data," *IEEE WHISPERS 2010*, 2010.

[4] H Daumé III and D Marcu, "Domain adaptation for statistical classifiers," *Journal of Artificial Intelligence Research*, vol. 26, no. 1, pp. 101–126, 2006.

[5] JR Schott, SD Brown, RV Raqueno, HN Gross, and G Robinson, "Advanced synthetic image generation models and their application to multi/hyperspectral algorithm development," *Proceedings of SPIE*, vol. 3584, pp. 211, 1999.

[6] E Merényi, K Tasdemir, and W Farrand, "Intelligent information extraction to aid science decision making in autonomous space exploration," *Proceedings of SPIE*, vol. 6960, 2009.

[7] R Basedow, D Carmer, and M Anderson, "HYDICE system: implementation and performance," *Proceedings of SPIE*, vol. 2480, pp. 258–267, 1995.

[8] Research Systems Inc, *ENVI 4.6 Users Guide*, 2008, 1196 pp.

[9] B Csathó, WB Krabill, J Lucas, and T Schenk, "A multisensor data set of an urban and coastal scene," *International Archives of Photogrammetry and Remote Sensing*, vol. XXXII (3/2), pp. 26–31, Jan 1998.

[10] E Merényi, B Csathó, and K Tasdemir, "Knowledge discovery in urban environments from fused multi-dimensional imagery," *Proc. 4th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing Data Fusion over Urban Areas*, pp. 1–13, 2007.

[11] B Bue, E Merényi, and B Csathó, "Automated labeling of materials in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, pp. 4059—4070, Jun 2010.

[12] M Balcan, A Blum, and S Vempala, "Kernels as features: On kernels, margins, and low-dimensional mappings," *Machine Learning*, pp. 79–94, Jan 2006.