IEEE Letters of the Computer Society, Volume 2 , Issue 3, 2019, pp. 28-31
DOI: 10.1109/LOCS.2019.2938446

Sample Selection with SOMP for Robust Basis Recovery In Sparse Coding Dictionary Learning

Ayan Chatterjee, *Student Member, IEEE,* Peter W. T. Yuen

Abstract—Sparse Coding Dictionary (SCD) learning is to decompose a given hyperspectral image into a linear combination of a few bases. In a natural scene, because there is an imbalance in the abundance of materials, the problem of learning a given material well is directly proportional to its abundance in the training scene. By a random selection of pixels to train a given dictionary, the probability of bases learning a given material is proportional to its distribution in the scene. We propose to use SOMP residue for sample selection with each iteration for a more robust or 'more complete' learning. Experiments show that the proposed method learns from both background and trace materials accurately with over 0.95 in Pearson correlation coefficient. Furthermore, the proposed implementation has resulted in considerable improvements in Target Detection with Adaptive Cosine Estimator (ACE).

Index Terms—dictionary learning, sparse coding, basis, reconstruction, hyperspectral.

1 Introduction

Training a dictionary from a high dimensional heteroscedastic data, like a Hyperspectral Imagery (HSI), refers to decomposing crucial spectral information of a natural scene. Sparse coding is a well-established method that decomposes the HSI to a linear collection of a few bases, called atoms, with a wide range of applications (see examples [1], [2], [3]). The group of atoms for a given scene is called a dictionary.

Given a set of signals 'Y', Sparse Coding Dictionary (SCD) aims to find the dictionary 'D' and a sparse matrix called representation 'a', s.t. $Y = Da$, D is usually with ℓ_1 norm bases i.e. $||d_i||$ = 1. One such well-known SCD algorithm proposed by Adam Charles et. al.[4] is a classical SCD solving approach and is the focus of this letter for its abilities - (i) basis that looks like real spectra, (ii) infer HSI resolution with high accuracy from multispectral images, and (iii) evidence presented reconstructing real HSI trained from another season. This dictionary assumes Laplacian prior probability distribution instead of Gaussian on the coefficients of 'a'. The main contribution of their paper was the inclusion of coefficient statistics in the cost function.

1.1 Motivation

SCD retains the higher-order statistics present in HSI. A well-learned dictionary has benefits in essential applications, one of which is target detection (TD), a binary classification problem that differentiates target material of interest with the background (refer [5] for detailed literature). However, due to a random selection of pixels with each iteration during learning, there is a variation from run to run in the current SCD in [4] which the model in [6] attempt to address with unsupervised classification like the k-means algorithm to increase the probability of learning different materials. With most abundant materials having the highest likelihood for bases of learning them, the motivation behind this letter was to increase the likelihood of bases to learn different materials present in a scene irrespective of its distribution.

1.2 Data

For experimental validation, we have used a set of real images. We have considered Selene H23 VNIR[1] and Selene H23 Dual covering 0.4 to 2.5 µm (a registered image from HySpex VNIR-1600 and SWIR-384 sensor). The Dual image has a smaller GSD of 70x70cm compared to the VNIR with a GSD of 17x34cm. Natural materials like grass, soil and tree cover over 95% of Selene scene, and artificial materials cover the remaining scene, like Orange Perspex and Green Carpet. Apart from Selene, we have also considered publicly available Paso Robles-Monterey[2] which is a high altitude AVIRIS imagery consisting of vegetation, highway, and cities, and two Virginia City images[3] which are images of a mountainous region.

2 Literature

2.1 Sparse Coding Dictionary

Given a collection of signals $Y \in \mathbb{R}^{n \times m}$ with 'm' samples and 'n' dimensions, the problem of sparse decomposition over a dictionary $D \in \mathbb{R}^{n \times p}$ with 'p' atoms is called sparse coding. The SCD algorithm in [4] selects pixels at random during training to accelerate towards convergence (see [7] for a detailed study on sample selection).

• *A. Chatterjee, and P. Yuen are with Centre for Electronic Warfare, Information and Cyber, Cranfield Defence and Security, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA, UK.*

Manuscript submitted May 29, 2019. Revised August 10, 2019. Accepted August 26, 2019. Correspondence: ayan@outlook.com

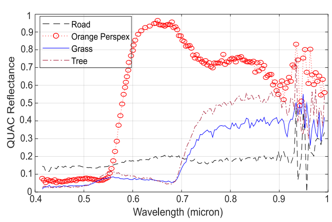
[1]This scene was acquired by HySpex VNIR-1600 sensor at Porton Down range (Long 51°8′19.7″N Lat 1°39′16.9″W to 51°7′41.7″N 1°40′8.5″W) on 12 August 2014 BST 12:00:04. QUAC was applied using ENVI software for atmospheric compensation.
[2]AVIRIS dataset (flight name: f150615t01p00r11)
[3]https://www.spectir.com

Algorithm 1 Dictionary Learning Algorithm, refer [4]

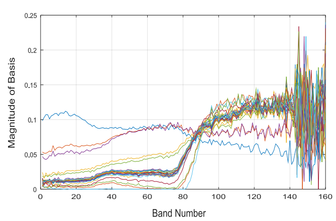
Import HSI Image as 'Y', number of dictionary atoms 'n', user-defined step size 's' and step-size decay 'd', and maximum number of iterations as 'M'
Initialize dictionary as 'D' as 'n' random numbers
for `i = 1 to M` **do** ▷ Iterate till convergence
  choose pixels 'x' (x ⊂ Y) ▷ Sample selection
  $a_x = \min||x - Da_x||_p$ ▷ Infer coefficients
  $\Delta D_x = (x - D \cdot a_x) \times a_x$
  $D \leftarrow D + s \times \Delta D_x$ ▷ Dictionary Update
  s = s × d ▷ Step size update

User-defined estimates like step size and decay, and least squares function to infer coefficients remain unchanged from [4]. An inner loop is required if more than one sample 'x' is selected per iteration to estimate 'a'.

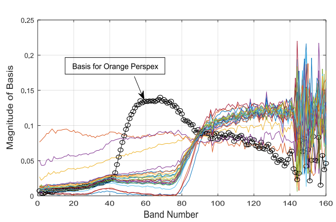
To investigate the inconsistencies of the above mentioned SCD from iteration to iteration, we did a run of 50,000 (or 50k) iterations on Selene H23 VNIR scene. We learnt that the background materials are learnt well but trace materials once learnt were forgotten over the next few iterations due to their lower probability of existence[4] in random sample selection (evidenced in figure 1).



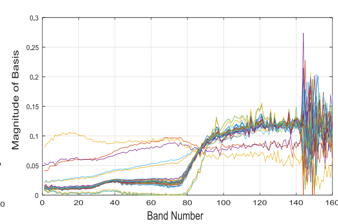
(a) Ground Truth (GT) pixels



(b) Dictionary at 15k$^{\text{th}}$ Iteration



(c) Dictionary at 23k$^{\text{th}}$ Iteration



(d) Dictionary at 35k$^{\text{th}}$ Iteration

Fig. 1: SCD progression with iteration on Selene H23 VNIR

At this stage, there is certainly is a need for an error evaluation criteria at each step during training that would help in sample selection to (i) select *weakly* learned samples (ii) avoid already learned samples. Algorithms that propose to learn trace 'target' materials along with the background, e.g., in [8] require *a priori* information of the target materials for classification before sample selection. A priori information of materials is usually not known in real scenes. Also, unsupervised classification suggested in [6] of separating various materials followed by a random selection of a single pixel from each cluster at each iteration is one way to increase the likelihood of selecting different materials. We will use k-means to cluster and show that although this approach produced improved accuracy than random selection, the proposed implementation has produced an even lower error.

[4]Orange Perspex covers ≈1% of Selene scene.

2.2 SD-SOMP Algorithm

Unmixing is the identification of *end-member* pure-pixels in a given scene assuming a convex cone (see [9] for an overview). Unmixing with Self-Dictionary assumes that there are pure-pixels in the scene and attempts to select a handful of sample pixels, one at a time, that are the end-members for the scene, called Self-Dictionary (SD). The closest explanation of SD is the pure-pixel index algorithm. The objective of SD problem is to minimise the ℓ_0-norm (or rows) of signal samples selected, is written as:

$$\min||X||_{\text{row},0} \text{ in } |Y - YX| \leq \epsilon, \text{ refer [10]} \tag{1}$$

The recently published greedy algorithm, SD-SOMP in [10], constructs SD by nominating the pixel with the maximum Simultaneous Orthogonal Matching Pursuit (SOMP) residue to be an end-member at each iteration. The algorithm continues till the pixel selected in '*ith*' iteration satisfies $||y_i - Da||_2 \leq$ acceptable error (the dictionary 'D' are all the pixels selected previously and representation 'a' is constrained to sum-to-one). The algorithm goes on to automate the stopping tolerance term with error estimates from HySime (see [11]).

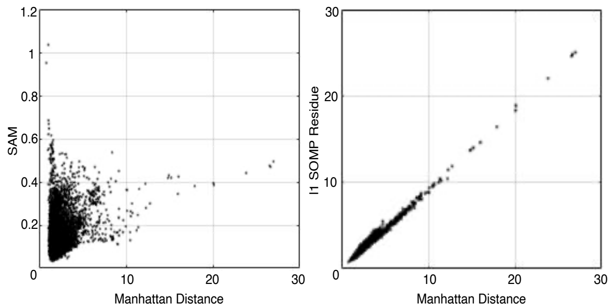
3 Proposed Implementation

If the pixel with maximum SOMP residue in SD-SOMP algorithm is within acceptable error tolerance, then the rest of the scene has converged, suggesting that SOMP residue has a relationship with the fitting error. The relationship for a given pixel 'y' w.r.t. a dictionary 'D' can be derived as:

$$\begin{aligned}
&\text{p-norm fitting error: } ||y - Da||_p \\
&\implies ||Iy - DIa||_p, \text{ by Identity matrix property} \\
&\implies ||Iy - D(D^{-1}D)a||_p \text{ or } ||Iy - (DD^{-1})(Da)||_p \\
&\implies ||Iy - (DD^{-1})(y - n)||_p, \text{ from y = Da + n} \\
&\implies ||(I - DD^{-1})y + (DD^{-1})n)||_p \qquad \text{(2a)} \\
&\approx \underbrace{||(I - DD^{-1})y||_p}_{\text{SOMP residue}}, \forall n << y \qquad \text{(2b)}
\end{aligned}$$

The above equation (2) demonstrates how, when noise is negligible, the SOMP residue is equal to the p-norm error. This approach is computationally much more efficient estimating the approximate fitting error at each iteration without the need to estimate sparse representation. In practice, pseudoinverse (usually Moore-Penrose pseudoinverse) is used to invert the non-square dictionary matrix, noise term is inclusive of material variability and other external factors deviating from an ideal signal 'y', and an HSI like Paso has mean signal-to-noise estimate of 39dB.

Adjacency and scattering exist in the atmosphere, and, the full pixels we think we see with our naked eye are not full pixels. Furthermore, due to the existence of sub-pixel materials, using SOMP residue is more effective than using other existing data curation methods using spectral angle to address the problem, like [12] using ORASIS algorithm[13]. The comparison is shown in figure 2.



(a) MD with SAM (b) MD with SOMP

Fig. 2: Scatter plot of Manhattan distance (MD) against (a) minimum SAM value from learned dictionary atoms, and (b) ℓ_1 SOMP residue. Pearson correlation coefficient of (a) is 0.2096, and (b) is 0.9630.

Higher SOMP residue is an indication for higher reconstruction error, and the *least learnt* pixels among the highest values. Replacing sample selection in each iteration in algorithm 1 with the pixels with maximum SOMP residue to train a model, guarantees that the dictionary has learned from all pixels of the scene when it converges. Figure 3 shows the convergence of maximum error residual.

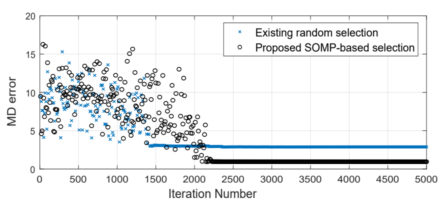
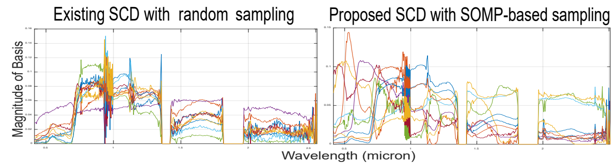


Fig. 3: Decrease in maximum error with iteration.

4 Results

Experiments conducted have been converged for ℓ_2 norm and with ℓ_2 SOMP residue to keep the comparison similar with the methods - SCD[4], the popular K-SVD[14], SD-SOMP[10], CoNMF[15], MVSA[16], and VCA[17] algorithm. We previously mentioned that one of the reasons for using SCD in [4] is that the learned atoms look like a real pixel. Existing SCD algorithm with random selection is referred as SCD-Rand, random selection after k-means clustering with 200 clusters where one pixel is selected from each cluster as SCD-kmeans, and the proposed SOMP selection as SCD-SOMP minimised for mean residue. Figure 4 show that using the proposed method, we learned both background and trace materials. Correlation is estimated with the Pearson Correlation coefficient.

In another run, the first 1k lines is used for training with 50 atoms each for both Selene and Paso. Reconstructed Virginia City image was trained on 1807-1211 scene. For the 10k iterations, the time taken by the existing approach with random selection on Selene H23 Dual is 115.58 min, and the proposed with 132.7 min (SOMP estimated on NVIDIA



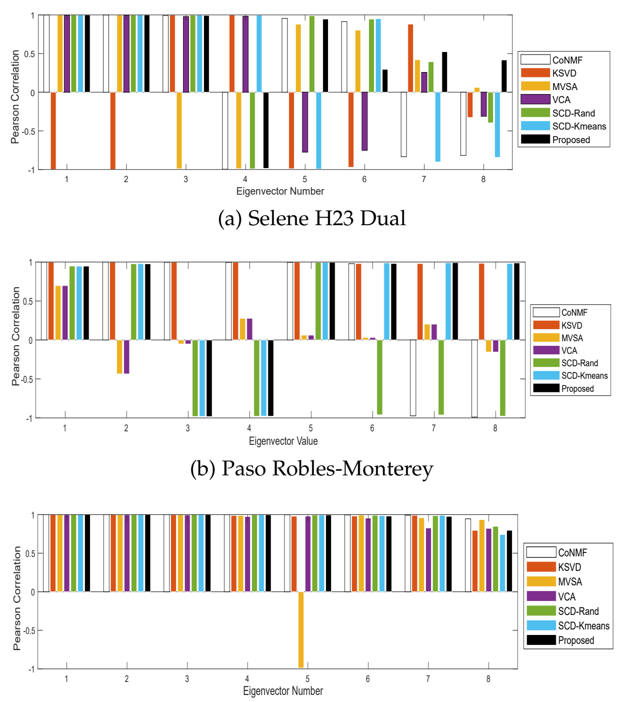
(a) Learned Dictionaries

(b) Grass (proposed), Correlation = 0.9556

(c) Orange Perspex (proposed), Correlation = 0.966

Fig. 4: Learned atoms on Selene H23 Dual

GTX 1060 GPU). Eigendecomposition with the correlation matrix conclude, figure 5, that proposed method has more positively correlated eigenvectors with GT[5].



(a) Selene H23 Dual

(b) Paso Robles-Monterey

(c) Virginia City 1807-1220

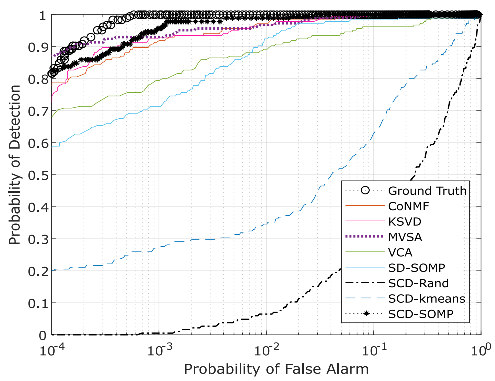
Fig. 5: Correlation of eigenvectors with GT.

[5]The dimensions of the scenes are:

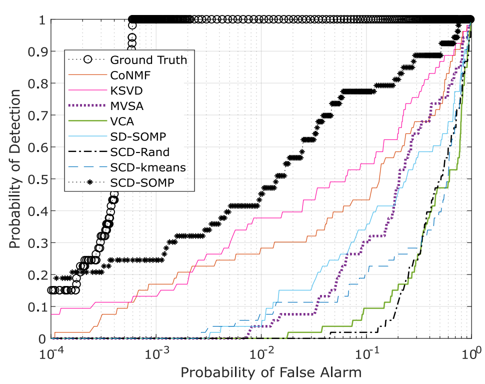
- Selene H23 VNIR: 3752 lines, 1600 samples, and 160 bands
- Selene H23 Dual: 1876 lines, 380 samples, and 448 bands
- Paso Robles-Monterey: 5115 lines, 741 samples, and 224 bands
- Virginia City 1807-1211: 6349 lines, and 1807-1220: 6758 lines (both scenes are with 320 samples and 178 bands)

5 Target Detection

We previously mentioned that one of our purposes of using SCD is in target detection. ACE is one of the classical TD methods whose individual distance measures is the square of the Mahalanobis distance. As a consequence of learning with the proposed SOMP selection to increase the probability of absorbing both the background materials and the trace materials in the dictionary, we expect an increase in TD accuracy. Applying the proposed implementation of two target materials on Selene scene, the detection result of the same experiment from figure 5 is shown in figure 6.



(a) Orange Perspex (easy target material)



(b) Green Carpet (hard target material)

Fig. 6: ROC Curve using ACE for detection.

6 Conclusion

We suggested random sample selection during the training of SCD algorithm in [4] be replaced by samples with maximum SOMP residue for applications like TD. SOMP residue is a way to estimate the fitting error without estimating representation. Using this approach, we are able to successfully demonstrate the ability of bases in a dictionary to be more robust, learning different materials in a scene irrespective of its abundance distribution.

Acknowledgment

This work was supported in part by The Defence Science and Technology Laboratory (DSTL) under project number DSTLX-1000103251. The authors would like to thank DSTL for providing Selene dataset, and extend their gratitude to Dr Adam Charles for responding to email communications about his journal article on SCD algorithm.

References

[1] D. Gao, Z. Hu, and R. Ye, "Self-dictionary regression for hyperspectral image super-resolution," *Remote Sensing*, vol. 10, no. 10, 2018.

[2] X. Han, J. Yu, J. Luo, and W. Sun, "Reconstruction from multispectral to hyperspectral image using spectral library-based dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1325–1335, March 2019.

[3] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectralspatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4186–4201, Aug 2015.

[4] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, Sep. 2011.

[5] K. B. William F. Basener, Brian Allen, "Geometry of statistical target detection," *Journal of Applied Remote Sensing*, vol. 11, no. 1, pp. 1 – 22 – 22, 2017.

[6] M. A. Toksz, K. G. Toker, and G. G. Gngr, "An automatic dictionary construction framework for sparsity-based hyperspectral target detectors," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 8, pp. 1264–1268, Aug 2018.

[7] R. Giryes, Y. C. Eldar, A. M. Bronstein, and G. Sapiro, "Tradeoffs between convergence speed and reconstruction accuracy in inverse problems," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1676–1690, April 2018.

[8] Y. Niu and B. Wang, "Hyperspectral target detection using learned dictionary," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1531–1535, July 2015.

[9] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, April 2012.

[10] X. Fu, W. Ma, T. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1128–1141, Sep. 2015.

[11] P. Benner, V. Novakovi, A. Plaza, E. S. Quintana-Ort, and A. Remn, "Fast and reliable noise estimation for hyperspectral subspace identification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1199–1203, June 2015.

[12] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 219–228, Jan 2012.

[13] J. Bowles, D. Gillis, and P. Palmadesso, "New improvements in the orasis algorithm," in *2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484)*, vol. 3, March 2000, pp. 293–298 vol.3.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[15] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6076–6090, Oct 2016.

[16] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[17] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.