REVISITING THE PREPROCESSING PROCEDURES FOR ELEMENTAL CONCENTRATION ESTIMATION BASED ON CHEMCAM LIBS ON MARS ROVER

*Wei Wang, Shuangjiang Li, Hairong Qi** *Bulent Ayhan, Chiman Kwan* *Steven Vance*

University of Tennessee, EECS Dept.
Knoxville, TN, US

Signal Processing Inc.
Rockville, MD, US

Jet Propulsion Laboratory
Pasadena, CA, US

ABSTRACT

The ChemCam instrument package on the Mars rover, "Curiosity", is the first planetary instrument that employs laser-induced breakdown spectroscopy (LIBS) to determine the compositions of geological samples on another planet. However, the sampled spectra are imperfect for elemental concentration estimation because of the inevitable sampling noises, the spectra continuum and the high dimensionality, thus the preprocessing procedures (e.g., dark removal, denosing and continuum removal, etc.) are necessary to improve the quality of the spectra. This paper not only presents a comprehensive evaluation of each preprocessing techniques, but also propose to use a non-traditional denoising technique and an effective band selection approach to greatly improve the accuracy of concentration estimation. In addition, we also test various combination of each procedure to give the best preprocessing sequence. The claims are all tested on a real LIBS dataset, the experimental results demonstrate the effectiveness of the preprocessing and also validate our claims.

***Index Terms*— LIBS spectrum, Continuum removal, Denoising, Band selection, Concentration estimation.**

1. INTRODUCTION

The chemical camera (ChemCam) instrument suite is one of the remote sensing composition facilities for the Mars Science Laboratory (MSL) rover. The Laser-Induced Breakdown Spectroscopy (LIBS) was selected as one part of this suite for its capability of determining elemental compositions of rocks and soils within seven meters of the instrument [1]. The advantage of the LIBS instrument to other Spectrometers is that it is not only able to remove dust and coatings or weathering rinds from rock's surface to determine their underlying composition, but also can rapidly detect many elements, including some light elements.

LIBS is an attractive technique for in *situ* elemental analysis since it requires no sample preparation and the analysis results are also available in real time. It is a form of atomic emission spectroscopy, that works by focusing a high power laser pulse onto a target to create a plasma of excited material that emits light containing spectral lines corresponds to the atoms and ions that compose the plasma [2]. Because of the large distance between the instrument and the observed sample and also other disturbances, the LIBS spectral data will be collected in a noisy environment, with spectrally variable characteristics. In particular, the noises may be not identically distributed along the wavelength axis. In addition, the LIBS spectral emission also contains a background continuum due to Bremsstrahlung and ion/electron recombination process, which contains non-relevant spectral information [3]. Therefore, we consider a LIBS spectrum as the superposition of emission line $\mathcal{E}$, continuum $\mathcal{C}$ and noise $\mathcal{N}$. Fig.1 shows an example of one typical LIBS spectrum, which also indicates the responding wavelengths of different elements.

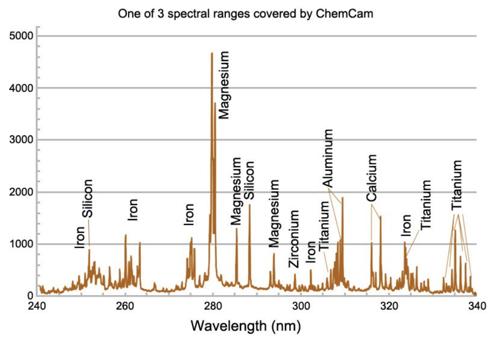


Fig. 1. An example of LIBS spectrum.

Traditional LIBS data preprocessing involves subtracting light background, removing noise and removing electron continuum. In this paper, we revisit the preprocessing procedures of the LIBS data and expect that the spectra after preprocessing can provide better performance for multivariate model based (e.g., PLS) elemental abundance estimation. Since a non-laser "dark" spectrum will be taken in close temporal proximity to the LIBS spectra, the information of the light background is always given with the raw spectral data. Therefore, we mainly focus on denoising (Dn) and continuum removal (CR). Usually, the continuum removal is done after denoising because the impulsive noises might affects the extraction of continuum. However, we find the impulsive noises seldom affect the CR step because each of the element response

*Thanks to XYZ agency for funding.

will give a large value that noises can not be reached easily, and thus guess the denoising should be taken after CR since the procedure of CR also might magnify undesired noises.

As for CR in this work, we adopted a standard continuum removal algorithm introduced in [4]. The spectral continuum occurs when the interactions of a large number of atoms, ions spread out all the discrete emission lines of an object, so they can no longer be distinguished. Since most remotely sensed spectra are composed of mixtures instead of pure materials, by removing the continuum, we are permitted to compare individual features of each element from a common baseline.

As for LIBS denoising, we presented a comparative analysis between two distinct approaches: the traditional Wavelet Decomposition Transformation (WDT) [5] and the Block-Matching and 3D Filtering (BM3D) [6]. WDT has been widely used in areas of signal processing. The WDT based approaches do not require any particular assumptions about the nature of the signal, permits discontinuities and spatial variation, and also exploits the spatially adaptive multiresolution of the wavelet transform. While BM3D represents the state-of-the-art in the field of image/video denoising. It is an approach of non-local adaptive nonparametric filtering based on enhanced sparse representation in transformed domain. Though the LIBS spectra are in 1-D space, we can tailor each spectrum into a virtual image to be denoised.

In addition to these routine preprocessing procedures, we also find that there are only a few wavelengths give responses for each element, though all the responses will span the whole range 240-905 nm. Therefore, some of the wavelengths may not give response to any of the elements in consideration. These useless wavelengths not only give very limited contribution in elemental concentration estimation, but also increase the dimension of the spectrum, resulting in possible Hough effect and higher computational cost. To solve this problem, we add one more procedure: band selection, after the aforementioned procedures. With lower dimension, the spectral analysis can be done much faster with even higher estimation accuracy.

2. CONTINUUM REMOVAL

The continuum of a spectrum is defined as a continuous, convex hull draped over the source spectrum at its *high points*. To removed the continuum existed in a spectra, we use the algorithm in [4]. We first locate all the high points on the spectrum, and try to combine the line segments that connect every two adjacent high points without cross the spectrum. By doing so, this approach assures that the resulting continuum is always convex and does not cross through the original spectra. The linear data interpolation method is used when connecting two points. In additional, because the visible and near infrared (VNIR) portion of the spectrum has by far the strongest relative contribution of the continuum. We can evaluate the CR only in the VNIR range or on the whole spectrum range,

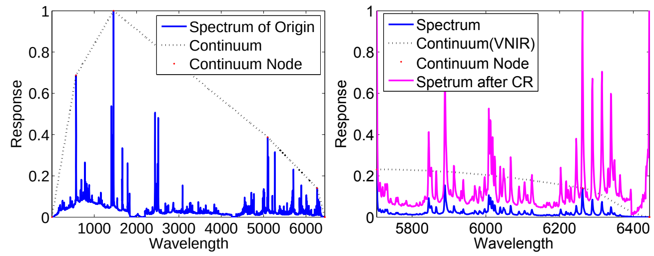


Fig. 2. Illustration of Continuum Removal. Left, red points are the identified high points on all spectrum ranges and the dash line is the extracted continuum; right, the magenta spectrum is the one after continuum removal on VNIR spectrum range.

respectively. The Fig. 2 shows an example of CR applied on the whole range and VNIR range. The left sub-figure shows the *high points* and the continuum extracted from the whole spectrum range using linear interpolation. After the continuum is extracted, the source spectrum will be divided by its continuum spectrum, resulting in a continuum removed spectrum containing normalized reflectance values ranging from 0.0 to 1.0. The right sub-figure in Fig.2 shows the example spectrum as well as the resulting continuum removed spectra when CR only applied on VNIR.

3. SPECTRAL DENOISING

3.1. Wavelet Decomposition (WDT)

Wavelet analysis is becoming one of the most useful tools in signal processing areas. In related literature, [5, 7] provided several examples of signal denoising works applying wavelet transform.

To remove noises, the noisy spectrum is transformed into the wavelet domain and expanded at different wavelet scales. Then, the decomposed wavelet coefficients will be filtered by a threshold, and only those with values greater than the threshold will be kept. Finally, the denoised spectrum is obtained by transforming the modified wavelet coefficients back into the time domain. The choice of the threshold value is a key step. On one hand, a large threshold results in a lost of useful information. On the other hand, a small one does not remove the noise to a satisfactory extent. Donoho [8] proposed a method for defining a kind of soft threshold T for signal denoising based on wavelet decomposition.

$$T = (M/0.6745) \times \sqrt{2\ln(N)} \tag{1}$$

where M is the medium value computed iteratively at each scale and N is the length of the signal at each scale. Then the soft threshold value is defined as:

$$W_{jk} = \begin{cases} W_{jk} & \text{for} \quad |W_{jk}| \geq 1.25T \\ \frac{1.25T(|W_{jk}|-0.75T)}{sgn(W_{jk})\cdot 0.5T} & \text{for} \quad 0.75T \leq |W_{jk}| < 1.25T \\ 0 & \text{for} \quad |W_{jk}| < 0.75T \end{cases} \tag{2}$$

Finally, all the wavelet coefficients are filtered by this strategy before transformed back into the time domain.

3.2. Block-Matching and 3D Filtering (BM3D)

Dabov et al. proposed a patch-based strategy that exploits image self-similarities and gives state-of-the-art results for image denoising [6]. Similar to approach of nonlocal means filtering [9], they reconstruct patches by finding similar ones in the image (block matching), stacking them together into a 3D block, and denoise the block using hard or soft thresholding with a 3D orthogonal dictionary (3D filtering). In conjunction with a combination of weighted averages of overlapping patches, Kaiser windows, and Wiener filtering to further improve results, the BM3D has proven to be very efficient and gives better results than regular non-local means.

The BM3D approach is realized by four successive steps: grouping 2D fragments in image into 3D arrays, 3D transformation of the group array, shrinkage of the transformed spectrum and inverse 3D transformation. In this paper, we use the BM3D to eliminate noise in the LIBS spectra. Though, the spectra are all 1-D signals, we can tailor them into a 2D matrices as virtual images, then the denoising based on BM3D can further attenuate noises by considering the repeated fragments on each spectrum that without obvious elemental responses.

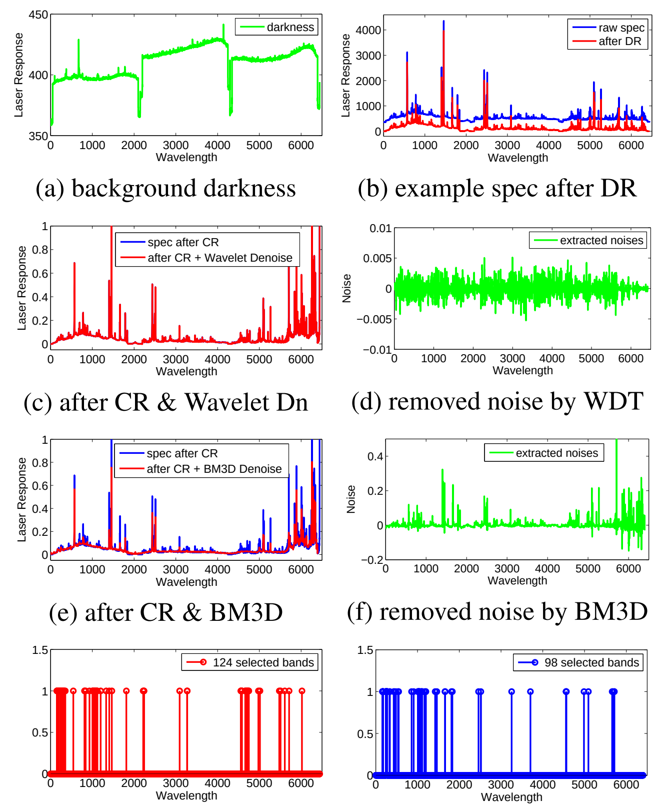
4. BAND SELECTION

As we mentioned before in introduction, many of the bands on LIBS spectrum do not give response in presence of elements under consideration. In order to extract the most effective wavelengths for improving the accuracy of concentration estimation and also reduce the spectral dimension to save computational cost, we adopt the Sequential Floating Forward Selection (SFFS) algorithm [10] to find the optimal band subset for better concentration estimation. Since we are using Partially Least Square Regression (PLS) for estimation, the criteria function in SFFS is chosen as the PLS estimation accuracy. Without any pre-knowledge about the band's effectiveness, the SFFS always constructs the "best" combination of wavelength bands to maximize the index of the criteria function using a non-exhaustive way.

Though the SFFS uses an efficient way to perform band selection, the dimension of LIBS can be as high as 6000, the SFFS algorithm still need quite a long time to finish the band selection process. Based on the original observation, we also take a preliminary band selection before SFFS by filtering according to the standard deviation (**StdFilter**) of each band in the measured dataset. Then, the dimension of the spectra that be input into the SFFS can be reduced to a half or even one third by removing those bands with small variation. Then, the SFFS can be done in a much shorter time.

5. EXPERIMENTS

The data used in our experiments is a standard LIBS spectral library developed by Los Alamos National Laboratory (LANL) (download at the website [1]). The spectral library was developed using 66 rock standards in the form of pressed powder disks, glasses, and ceramics to minimize heterogeneity on the scale of the observation (350-550 m dia.). The standards covered typical compositional ranges of igneous materials and also included sulfates, carbonates, and phyllosilicates. Each sample has 4 spectra, and each of which is averaged from 50 measures. The concentration ground truth of the compositional elements is also provided, thus we can quantitatively evaluate the quality of the preprocessed LIBS spectra for elemental concentration estimation.



(a) background darkness (b) example spec after DR

(c) after CR & Wavelet Dn (d) removed noise by WDT

(e) after CR & BM3D (f) removed noise by BM3D

(g) after CR&WDT, 124 BS (h) after CR&BM3D, 98 BS

Fig. 3. Intermediate results after each procedure.

As introduced before, our preprocessing involves dark removal (DR), continuum removal (CR), spectral denoising (Dn) and band selection (BS). We first show the intermediate results after each procedure in Fig.3. We can observe that DR does not change the spectrum much, but for CR, which applied only on VNIR wavelengths in Fig.3, it does modified the shape of the spectrum a lot and the highest response values at different wavelength ranges are normalized into similar scale. As for the two different denoising approaches under consideration, we can find that the wavelet based denoising

Table 1. Accuracy of elemental concentration estimation in each intermediate step. (Note: LA-CR denotes linear CR applied on ALL wavelengths; Li-on-All denotes linear CR on whole range; Preprocessed denotes the preprocessed raw data by standard procedures in [3].)

Data Type	subtype	PLS LV	**Avg. Err**
Raw		19	1177.53
+DR		19	1183.92
+DR+CR	Li-on-ALL	19	**1052.91**
	Li-on-VNIR	17	1087.99
+DR+Dn	Wavelet	19	1178.19
	BM3D	19	**1139.71**
+DR+W-Dn+CR	Li-on-VNIR	19	1079.93
+DR+B-Dn+CR	Li-on-ALL	27	1044.29
+DR+LA-CR+Dn	Wavelet	27	1042.34
$\cdots$ + StdFilter	std ≥ 0.02	27	1076.07
$\cdots$ + SFFS-BS		23	976.45
+DR+LA-CR+Dn	BM3D	17	**1026.99**
$\cdots$ + StdFilter	std ≥ 0.02	17	1045.88
$\cdots$ + SFFS-BS		25	884.05
+DR+LV-CR+Dn	+ SFFS-BS	27	**767.75**
PreProcessed*		15	1382.12

mainly removes white noise with small scale. In contrast, the BM3D based denoising removes noise with much larger scale and gives a flat spectrum at non-response wavelength ranges. This may due to the reason that BM3D also explores the self-similarity and be able to identify which patch may correspond to noise based on the statistical property. Finally, we compare the selected bands after two denoising approaches, 124 and 98 effective bands are selected out for final elemental concentration estimation, respectively. Because the BM3D removes noise to a further extent, the number of selected bands is even smaller than the WDT based denoising.

We propose two claims in this paper: First, the denoising should be taken after CR in preprocessing procedures; Second, band selection should be taken as an additional procedure for elemental concentration estimation. We thus quantitatively evaluate each of the preprocessing procedures based on the estimation accuracy, and seek for the optimal sequence by different combinations. We use a 66-fold cross validation for evaluation, where only one sample with 4 spectra will be tested each round and the averaged result is used as the estimation for the one sample in test. We also scan a range of values for choosing the optimal latent parameter for PLS regression to guarantee a fair comparison. From the experimental results as shown in Table 1, we have several observations: First, the CR procedure is more effective than denoising, this is intuitive since CR magnifies the useful information on the spectra; Second, the BM3D based denoising can perform better than traditional wavelet based approach; Third, the concentration estimation indeed achieves higher accuracy when put denoising after CR; Fourth, the additional band selection can greatly improve the accuracy of the concentration estimation, and the estimation speed can be greatly boosted after StdFilter though the StdFilter does not help in accuracy improvement; Fifth, our preprocessed spectra can provide much higher estimation accuracy compared to the preprocessed data via standard procedures in [3], and the estimation error can be reduced to almost a half (767.75 vs 1382.12).

6. CONCLUSION

We revisited the LIBS spectral preprocessing procedures for accurate elemental concentration estimation. With observation of the LIBS spectral property, we identified effective algorithms for continuum removal, spectral denoising and band selection. In addition, by comprehensive comparisons in experiments, we also determined the optimal preprocessing sequences to achieve high estimation accuracy. The experimental results confirmed the use of our strategy for improved performance of elemental concentration estimation via LIBS.

7. REFERENCES

[1] "http://pds-geosciences.wustl.edu/missions/mer/geo-mer-datasets.htm," .

[2] K. Miziolek, V. Palles, and I. Schechter, "Laser induced breakdown spectroscopy LIBS: Fundamentals and applications," in *NewYork: Cambridge Univ. Press*, 2006.

[3] R. C. Wiens, S. Maurice, and et al., "Pre-flight calibration and initial data processing for the chemcam LIBS instrument on the mars science laboratory MSL rover," *Spectrochimica Acta, B.*, vol. under review, 2013.

[4] F. A. Kruse and A. B. Lefkoff, "Knowledge-based geologic mapping with imaging spectrometers," *Remote Sensing of Environment*, vol. 81, no. 3, pp. 3–28, 1994.

[5] J. Pan, Q. Guo, and B. Jiang, "Wavelet packet transform based minimum risk denoising algorithm for astronomical spectra," in *ICFSKD*, 2007.

[6] K. Dabov, A. Foi, and V. Katkovnik, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE TIP*, vol. 16, no. 8, pp. 2080–2095, 2007.

[7] F. Lei and J. Gong, "Denoise in the pool alarm system based on wavelet transform," in *ICMT*, 2010.

[8] D. Donoho and L.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of American Stat. Assoc.*, vol. 12, no. 90, pp. 1200–1224, 1995.

[9] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *IEEE CVPR*, 2005.

[10] P. Pudil, J. Novo, and J. Kittler, "Floating search methods in feature selection," *PR Letter*, vol. 15, no. 11, 94.