

A SIGNAL DENOISING METHOD FOR FULL-WAVEFORM LIDAR DATA

Mohsen Azadbakht^{a,b,*}, Clive S. Fraser^{a,b}, Chunsun Zhang^{a,c}, Joseph Leach^b

^a Cooperative Research Centre for Spatial Information, VIC 3053, Australia

^b Department of Infrastructure Engineering, University of Melbourne, VIC 3010, Australia –

m.azadbakht@student.unimelb.edu.au, (c.fraser, leach)@unimelb.edu.au

^c School of Mathematical and Spatial Sciences, RMIT University, VIC 3000, Australia - chunsun.zhang@rmit.edu.au

KEY WORDS: Savitzky-Golay method, signal distortion, retrieval, Full-waveform LiDAR, Active Remote Sensing.

ABSTRACT:

The lack of noise reduction methods resistant to waveform distortion can hamper correct and accurate decomposition in the processing of full-waveform LiDAR data. This paper evaluates a time-domain method for smoothing and reducing the noise level in such data. The Savitzky-Golay (S-G) approach approximates and smooths data by taking advantage of fitting a polynomial of degree d , using local least-squares. As a consequence of the integration of this method with the Singular Value Decomposition (SVD) approach, and applying this filter on the singular vectors of the SVD, satisfactory denoising results can be obtained. The results of this SVD-based S-G approach have been evaluated using two different LiDAR datasets and also compared with those of other popular methods in terms of the degree of preservation of the moments of the signal and closeness to the noisy signal. The results indicate that the SVD-based S-G approach has superior performance in denoising full-waveform LiDAR data.

1. INTRODUCTION

A LiDAR system is an active remote sensing instrument that directly measures the distance between the target's surface and the sensor based on the round trip time of a laser pulse (Baltsavias 1999; Wehr and Lohr 1999; Jutzi and Stilla 2003; Wagner et al. 2006). Full-waveform Airborne Laser Scanning or LiDAR systems possess the advantages of continuous sampling to directly retrieve vertical and horizontal structural characteristics of both small and large objects, in addition to the 3D position of ground points. A variety of applications of this ranging technology, e.g. in urban planning, vegetation mapping and forestry, have already been explored with the aim of quantitative estimation of different parameters used in modelling the physical environment.

The range (distance between sensor and target), amplitude and pulse width are the most common attributes which can be derived from LiDAR waveforms by means of deconvolution/decomposition techniques (Wagner et al. 2006; Mallet and Bretar 2009; Heinzel and Koch 2011; Höfle et al. 2012). With full-waveform data, it is feasible to derive the target cross-section, which is not possible in the case of discrete return LiDAR (Höfle and Pfeifer 2007) and which is an issue of considerable interest in active remote sensing (Wagner et al. 2006; Wang et al. 2009). The accuracy of determining these parameters depends upon the shape of the waveform, and any change or shift in the waveform can cause uncertainty in the results.

On the other hand, an unavoidable phenomenon of images and signals, especially in remote sensing data, is noise that influences the resolution of outputs and restricts their capability by imposing distortion, misalignment and stretch. Preprocessing of signals in order to decrease the effect of noise is

recommended before deconvolution of waveforms and extraction of further information from the signal (Pirotti 2010). Many methods have been proposed for both noise removal and improvement of the signal-to-noise ratio (SNR) in full-waveform LiDAR data. Some methods are not appropriate in all cases and need specific assumptions. Typically, in Low-Pass filters, the high frequency segments are assumed to contain noise. However, some filters cause a shift in the signal and a change in shape of the signal (Hassanpour et al. 2012).

The FFT-based and Moving Average (MA) noise reduction methods, despite their ability to remove noise, can distort and smooth the waveform and this can affect the results of subsequent processing, such as the range values in the case of full-waveform LiDAR data. In addition, the MA filter can over smooth peaks and broaden the signals when peaks are narrower than the length of the filter, therefore reducing the amplitude of the signals. The signal accumulation approach (Kraftmakher 2006) assumes both noise with zero mean random characteristics and the signal phase to be known, although this is not the case in most situations and estimating the noise properties is not straightforward and feasible, as shown by Wu et al. (2012) and Okumura et al. (1993). The lasing and nonlasing echoes are necessary in the canonical correlation analysis proposed by Okumura et al. (1993), the nonlasing part, however, is not available for most end users (Wu et al. 2012). Some popular methods in the case of full-waveform data, such as the Wiener filter, cause oscillation and ringing effects near edges (Parrish 2007) as well as being unable to work efficiently in situations of low SNR (Hassanpour et al. 2012).

Existing approaches show varying success. The main purpose of the smoothing filters in full-waveform LiDAR is to reduce the noise level while maintaining the useful signal contents (Orfanidis 2010). This paper describes a method based on

* Corresponding author.

fitting a polynomial to signal values such that raw (noisy) samples will be substituted by values obtained from fitted smooth polynomial curves. This method, based on the Savitzky-Golay (S-G) filter (Savitzky and Golay 1964), has the potential of preserving the shape and amplitude of signals as well as reducing the noise level (Hassanpour 2007; Orfanidis 2010; Schafer 2011; Hassanpour et al. 2012). However, this is achieved at the cost of potentially retaining much more noise than the MA smoothing method (Orfanidis 2010).

In Hassanpour (2007; 2008), the S-G filter is applied in conjunction with singular vectors derived from a singular value decomposition (SVD) of a noisy Electroencephalography (EEG) signal in the time-frequency domain. The SVD-based noise reduction method is operated on a Hankel Matrix form of the noisy signal in the time domain by Hassanpour et al. (2012), followed by the S-G filter. A higher level of noise reduction can be achieved through eliminating the noise components during the SVD-based enhancement.

In this paper, we report our experience of denoising full-waveform LiDAR data. The SVD-based S-G method has been investigated and evaluated against the Moving Average approach and the S-G filter. The S-G, and SVD-based S-G filters are first described and following this the results of these methods applied to full-waveform LiDAR data are presented and compared. Experimental results are then discussed and concluding remarks are offered.

2. METHODOLOGY

2.1 Savitzky-Golay filter

As the first stage in describing the SVD-based S-G filter, the mathematical background of the S-G filter is briefly presented. The S-G filter output of length N and order d , applied to $x(n)$ takes the form of (Orfanidis 2010)

$$y(n) = \sum_{m=-M}^M b_0(m)x(n+m) \quad (1)$$

Where $x(n) = [x_{n-M}, \dots, x_{n-1}, x_n, x_{n+1}, \dots, x_{n+M}]^T$ is an N ($=2M+1$) dimensional noisy vector
 $b_0(m)$: m-th S-G filter
 $y(n)$: S-G output

The fitted polynomial (\hat{x}_m) of order d can be written as

$$\hat{x}_m = c_0 + c_1m + \dots + c_dm^d, -M \leq m \leq M \quad (2)$$

where c_i are the polynomial coefficients.

The polynomial basis vectors s_i are defined as

$$s_i(m) = m^i; -M \leq m \leq M; i = 0, 1, \dots, d \quad (3)$$

where the matrix $S = [s_0, s_1, \dots, s_d]$ with s_i as columns (with dimension of $N \times (d+1)$).

The filtered signal, \hat{x} is given as

$$\hat{x} = \sum_{i=0}^d c_i s_i = [s_0, \dots, s_d][c_0, \dots, c_d]^T = Sc \quad (4)$$

According to Hassanpour et al. (2012), the following cost function should be minimized in order to adjust the polynomial degree and window length

$$Z(d, N) = \alpha \sum_k |FWF_{Enh}(k+1) - FWF_{Enh}(k)| + (1-\alpha) \sum_k |FWF_{Enh}(k) - FWF_{Noisy}(k)| \quad (5)$$

where FWF_{Enh} corresponds to the enhanced waveform
 FWF_{Noisy} corresponds to the noisy waveform
 $\alpha \in [0, 1]$ denotes the smoothness factor

The similarity between the enhanced and noisy signal and avoidance of large and abrupt differences between consecutive samples are taken into account in this cost function.

2.2 SVD-based S-G filter

As previously mentioned, the SVD-based S-G filter represents one of the extended approaches based on the S-G smoothing that can improve the level of signal enhancement. The method commences with a singular value decomposition of a matrix H , which is here the Hankel matrix of the noisy signal:

$$H = U\Sigma V^T \quad (6)$$

where U and V are orthogonal matrices
 Σ is a diagonal matrix of singular values

By assuming an additive noise component in the noisy signal, we can write:

$$H_{Noisy} = H_{Signal} + H_{Noise} \quad (7)$$

where H_{Noisy} : received noisy Hankel matrix
 H_{Signal} : original Hankel matrix without noise
 H_{Noise} : additive noise component

Then, the matrix H_{Noisy} is divided into signal and noise subspaces by assigning a threshold to Σ and eliminating singular values of smaller than a specified magnitude by setting them to zero:

$$H_{Noisy} = U_s \Sigma_s V_s^T + U_n \Sigma_n V_n^T \quad (8)$$

where the variables with subscripts s and n represent the signal and noise subspaces.

The S-G filter will be applied on the U and V matrices of the signal subspace and the enhanced Hankel matrix H_e , which represents the best estimate of the signal subspace, can be calculated from the enhanced U and V matrices:

$$H_e = U_e \Sigma_s V_e^T \quad (9)$$

where U_e and V_e represent the enhanced left hand and right hand orthogonal matrices after applying the S-G filter Σ_s = matrix of remained singular values.

The same cost function applied in the S-G filtering is employed for the recovery of optimal parameters values.

The key to the success of the SVD-based S-G filter is the separation of the singular values of matrix Σ into signal and noise subspaces since this can affect the resulting amplitude and the range of the filtered waveform. Different values of the threshold on singular values result in dissimilar shapes of the signal and misleading results can be obtained from inappropriate parameter values. The singular values shown in Figure. 1, obtained from the full-waveform signal in Figure. 2, are normalized in the range of [0, 1] in order to easily choose the threshold value. A proper threshold value can be set on a point where the slope of the curve becomes more constant. As can be seen in Figure 2, higher values of the cost function result from setting lower thresholds on singular values, while better signal smoothing, with lower oscillation and ringing effects, can be recovered as a consequence of an appropriate threshold value. However, increasing the threshold value to more than the optimal value leads to an unveiling of more superfluous detail and to similar results as those obtained via the S-G method.

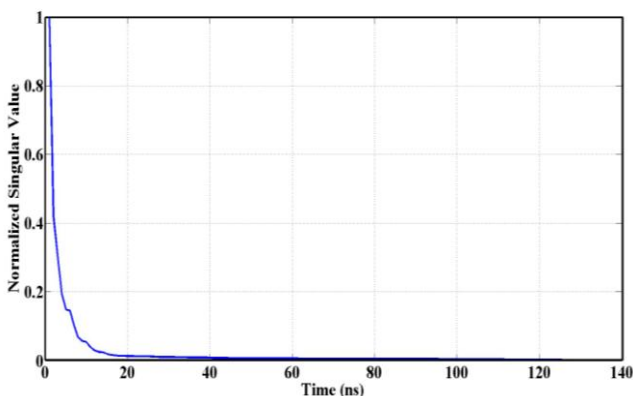


Figure 1. Normalized singular values of the matrix Σ of a given noisy signal

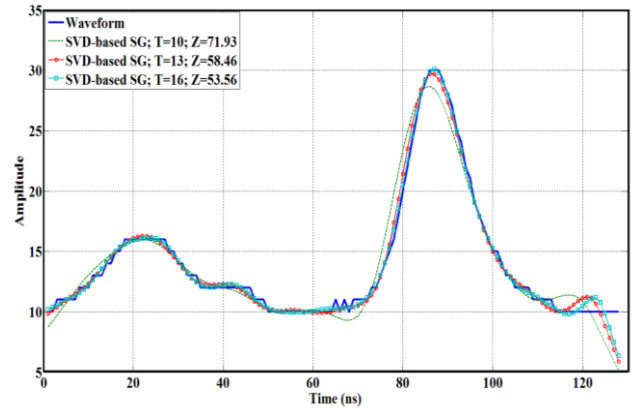


Figure 2. Influences of setting different thresholds (T) in singular values for three cost function values (Z)

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

The S-G based filters have been implemented and applied to LiDAR datasets acquired over different terrain. Comparisons of their performance with the MA approach were conducted and results will be reported in this section, following a brief description of the dataset.

3.1 Datasets

Two full-waveform LiDAR data sets were employed in the experimental evaluation of the SVD-based S-G filtering approach. The first data set, recorded with an Optech ALTM 3100 incorporating a laser of 1065 nm wavelength, covered an urban area near Verona, Italy. The second, acquired with a Riegl LMS-Q560 sensor with a 1550 nm laser, was over a forested area near Chowilla, South Australia. Shown in Figure 3 are two representative samples of the received waveforms within the two data sets. Figure.3 (a) is a waveform over flat, open terrain, whereas that in Figure.3 (b) was collected over an area yielding more complex, multi-layer returns from both the vegetation canopy and understory, and the ground.

3.2 Comparison of different methods

The performance of the SVD-based S-G method applied to the recorded noisy waveform data has been compared with the MA smoothing approach and the S-G method. In the SVD-based S-G filter, small oscillations in the waveform can be moderated and diminished after utilizing the SVD, and noise will be removed using the S-G low-pass filter. Here, the smoothness factor is set to 0.7 to yield results with satisfactory degrees of smoothness, and the maximum polynomial degree (d) and window length (N) are set at 5 and 20, respectively. High values of the polynomial degree can result in retention of more of the higher frequencies, increasing cut-off frequency and hence allowing more noise to be preserved (Orfanidis 2010).

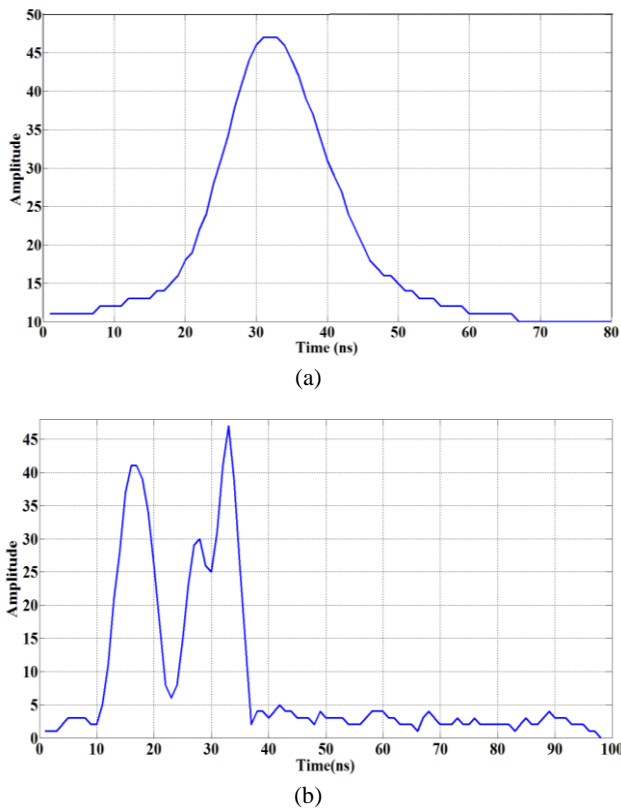


Figure 3. Samples of received waveforms over different terrain, from the Optech ALTM 3100 (a) and Riegl LMS-Q560 (b)

The performance of the S-G low-pass filter is improved by exploiting the SVD of the signal and separating the noise and signal subspaces. As can be seen from Figure 4, although the value of the cost function for the S-G filter is clearly smaller than for the SVD-based S-G filter, the resulting signal from the former still remains noisy, with the filtered signal being only slightly smoother than the original waveform. It is noticeable that although these two methods can preserve the main moments of the waveform (the amplitude and the pulse width), the SVD-based S-G method is more robust in alleviating noise, and small abrupt fluctuations exhibited in the result of the S-G method can be smoothed more efficiently.

However, it is noteworthy that the smoothed waveform, resulted from the SVD-based S-G filter, is affected by fluctuation near to the end of the signal. This problem is related to the form of the Hankel matrix and can be reduced largely by substituting this matrix with a Toeplitz form of the noisy signal. The enhanced version of the SVD-based S-G filter, after replacing the Hankel matrix by a more generalized Toeplitz form, is depicted in Figure 5.

Figure 6 illustrates the difference between the MA and SVD-based S-G filtering results for a complex waveform with three distinct peaks and small distances between peaks. The SVD-based S-G filters demonstrate the ability to largely preserve the pulse width and signal amplitude. On the other hand, the amplitude of the waveform cannot be recovered properly using the MA approach, particularly in instances of sharper peaks (mostly from the ground returns). Thus, estimates of the signal amplitude can be significantly underestimated by the MA approach. In the case of the second peak, the amplitude is markedly influenced by the higher amplitude of the

neighbouring peak, resulting in a significantly distorted form of the signal, which will cause incorrect interpretation of the data.

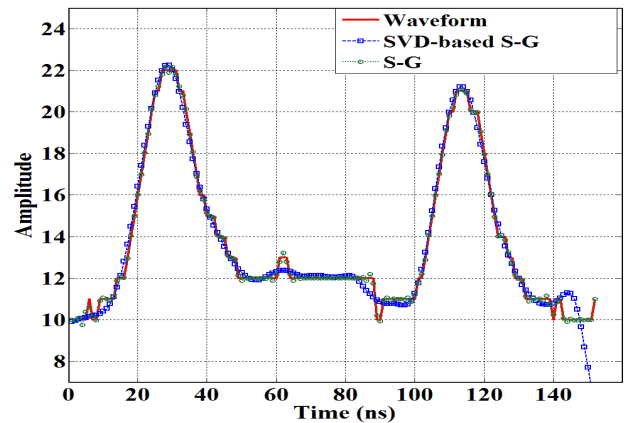


Figure 4. Comparison of the SVD-based S-G and S-G filters

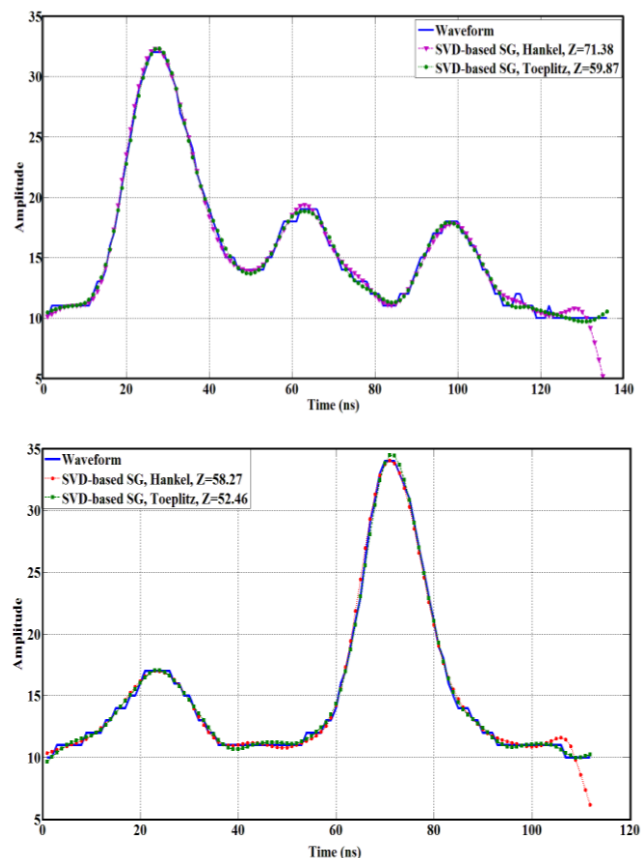


Figure 5. Comparison of the SVD-based S-G filters for Hankel and Toeplitz form of two noisy waveforms, along with their corresponding cost function values (Z)

A simple quantitative comparison based on the value of the utilized cost function confirms that S-G filtering outperforms MA filtering in terms of producing a smaller discrepancy between the filtered and original waveform. This implies that the SVD-based S-G filter preserves both the moments and, principally, the amplitude of the waveform.

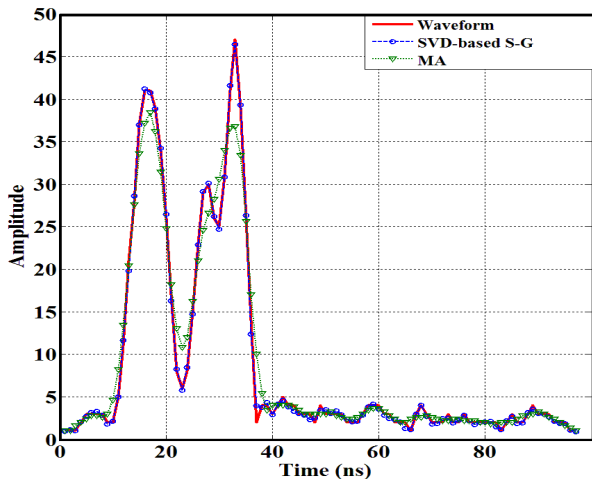


Figure 6. Smoothing a waveform via the SVD-based S-G and MA methods

Another approach to moderate oscillations yielded by the SVD-based SG filter on the Hankel form of the noisy signal is to weight the singular values. For this purpose, a normalized weight between 0 and 1 can be considered with evenly distributed values based on the total number of signal bins. This approach is yet to be implemented.

Finally, by considering a non-square Hankel matrix and reducing the number of its columns (and consequently its rank), it is possible to obtain more consistent results. Indeed, a smaller number of columns of the Hankel matrix can help to lessen the oscillation and more effectively remove noise by producing a smaller number of singular values. The influence of the number of columns is insignificant on the obtained results according to their cost function values for the modified thresholds on singular values (Table 1). As seen in Table 1, the threshold value is raised by increasing the number of columns of the Hankel matrix, allowing more singular values to contribute to denoising the signal and, accordingly, to increasing the cost function value. Although differences between these values are not considerable, it is recommended that the smallest number of columns be used such that the number is not smaller than the highest filter window length. Figure 7 illustrates the result of the related signal with 20 selected columns of the Hankel matrix. There is no evidence of oscillation and/or fake peaks at the tail-end of the signal, compared to the results from full-rank Hankel matrix in Figures 4-6.

No. columns	Threshold on singular values	Cost function value
20	4	47.74
50	9	47.81
80	12	48.10

Table 1. Different number of columns in the Hankel matrix with their related cost function values

From the comparison between the SVD-based SG filter with a square and non-square Hankel matrix, it emerges that the result is a more robust preservation of the signal shape, with the cost function value being slightly smaller, as per Figure 8. It is noticeable that while the MA result leads to a smaller discrepancy value with the noisy waveform, it also markedly underestimates the amplitude of the first pulse in addition to causing a near distortion of the second pulse.

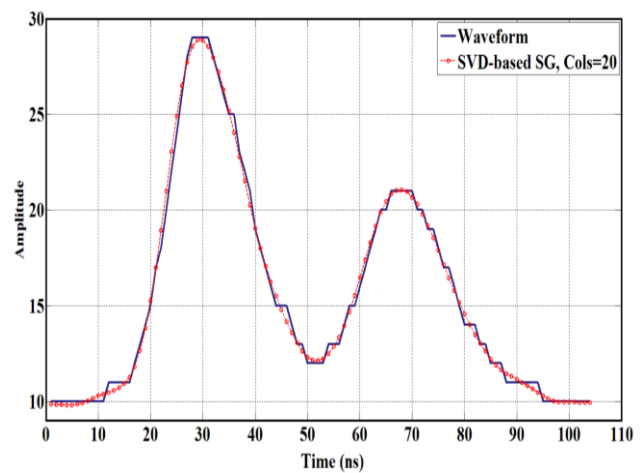


Figure 7. SVD-based S-G smoothing result with its specified number of columns of the Hankel matrix

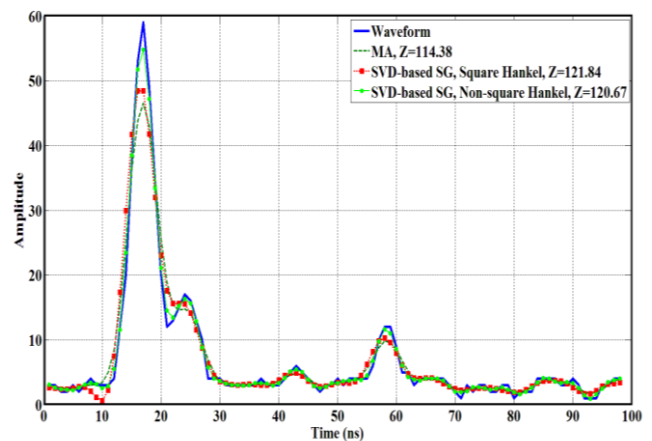


Figure 8. Comparison of the SVD-based S-G smoothing results together and with the MA

4. CONCLUDING REMARKS

The S-G low-pass filter has been applied across several fields of science, e.g. spectrometry, and to many types of signals. Here, it has been applied to noise removal in full-waveform LiDAR data. Moreover, it has been shown that the SVD-based S-G method exhibits an enhanced performance compared to both the S-G and MA approaches to noise reduction and LiDAR waveform smoothing, with better retention of the pattern of the signal and avoidance of induced distortion of the waveforms. The latter characteristic is essential in full-waveform LiDAR data processing since any deformation and distortion in the shape of the signal can lead to errors in range estimation and further processing will be adversely affected.

Elimination of smaller noise-related singular values in the SVD-based S-G filter, along with smoothing the singular vectors (U and V), results in smoother signals together with minimal change in the moments of the signal. In the current implementation, the parameters associated with the SVD-based S-G filter are optimized both by minimizing the predefined cost function in order to achieve a filtering that best preserves the moments of the recorded waveform, and by empirically adjusting the threshold applied to singular values. Substitution of the Hankel matrix by a Toeplitz form, along with use of a

lower rank matrix approximation of the Hankel form of the noisy signal, achieves better noise reduction and improvement in the results. This has the importance of reducing the computational cost by initially eliminating small singular values. The number of matrix columns then depends upon the maximum window length of the filter, while the threshold value does not play a significant role in signal distortion.

5. REFERENCES

- Baltsavias, E. P. (1999). "Airborne laser scanning: basic relations and formulas." *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(1999), pp. 199-214.
- Hassanpour, H. (2007). Improved SVD-based technique for enhancing the time-frequency representation of signals. *IEEE International Symposium on Circuits and Systems*.
- Hassanpour, H. (2008). "A time-frequency approach for noise reduction." *Digital Signal Processing* **18**(5), pp. 728-738.
- Hassanpour, H., A. Zehtabian and S. J. Sadati (2012). "Time domain signal enhancement based on an optimized singular vector denoising algorithm." *Digital Signal Processing* **22**(5), pp. 786-794.
- Heinzel, J. and B. Koch (2011). "Exploring full-waveform LiDAR parameters for tree species classification." *International Journal of Applied Earth Observation and Geoinformation* **13**(1), pp. 152-160.
- Höfle, B., M. Hollaus and J. Hagenauer (2012). "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data." *ISPRS Journal of Photogrammetry and Remote Sensing* **67**(2012), pp. 134-147.
- Höfle, B. and N. Pfeifer (2007). "Correction of laser scanning intensity data: Data and model-driven approaches." *ISPRS Journal of Photogrammetry and Remote Sensing* **62**(6), pp. 415-433.
- Jutzi, B. and U. Stilla (2003). Laser pulse analysis for reconstruction and classification of urban objects. In: Ebner H, Heipke C, Mayer H, Pakzad K (eds) *Photogrammetric Image Analysis PIA'03*. International Archives of Photogrammetry and Remote Sensing, Munich.
- Kraftmakher, Y. (2006). "Noise Reduction by Signal Accumulation." *The Physics Teacher* **44**(8): 528-530.
- Mallet, C. and F. Bretar (2009). "Full-waveform topographic lidar: State-of-the-art." *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(1), pp. 1-16.
- Okumura, H., T. Sugita, H. Matsumoto and N. Takeuchi (1993). Noise reduction method for lidar echo data based on multivariate analysis method. *Proceeding of IEEE Geoscience and Remote Sensing Symposium*.
- Orfanidis, S. J. (2010). *Introduction to Signal Processing*, Prentice Hall.
- Parrish, C. E. (2007). Exploiting full-waveform lidar data and multiresolution wavelet analysis for vertical object detection and recognition. *IEEE International Geoscience and Remote Sensing Symposium*.
- Pirotti, F. (2010). "IceSAT / GLAS Waveform Signal Processing for Ground Cover Classification : State of the Art." *Italian Journal of Remote Sensing* **42**(2), pp. 13-26.
- Savitzky, A. and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by simplified least squares procedures." *Analytical Chemistry* **36**(8), pp. 1627-1639.
- Schafer, R. W. (2011). "what is a Savitzky-Golay Filter?" *Signal Processing Magazine, IEEE* **28**(4), pp. 111-117.
- Wagner, W., A. Ullrich, V. Ducic, T. Melzer and N. Studnicka (2006). "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner." *ISPRS Journal of Photogrammetry and Remote Sensing* **60**(2), pp. 100-112.
- Wang, Y., J. Zhang, A. Roncat, C. Künzer and W. Wagner (2009). "Regularizing method for the determination of the backscatter cross section in lidar data SIGNALS." *Optical Society of America* **26**(5), pp. 1071-1079.
- Wehr, A. and U. Lohr (1999). "Airborne laser scanning—an introduction and overview." *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(1999), pp. 68-82.
- Wu, J., J. a. N. van Aardt, J. McGlinchy and G. P. Asner (2012). "A Robust Signal Preprocessing Chain for Small-Footprint Waveform LiDAR." *IEEE Transactions on Geoscience and Remote Sensing* **50**(8), pp. 3242-3255.