Korean Journal of Remote Sensing, Vol.26, No.6, 2010, pp.681~691

A Sequential LiDAR Waveform Decomposition Algorithm

Jinha Jung*, Melba M. Crawford*, and Sanghoon Lee †**

*Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907

**Department of Industrial Engineering, Kyungwon University, Seognam-shi, Kyunggi-do, Korea

Abstract : LiDAR waveform decomposition plays an important role in LiDAR data processing since the resulting decomposed components are assumed to represent reflection surfaces within waveform footprints and the decomposition results ultimately affect the interpretation of LiDAR waveform data. Decomposing the waveform into a mixture of Gaussians involves two related problems; 1) determining the number of Gaussian components in the waveform, and 2) estimating the parameters of each Gaussian component of the mixture. Previous studies estimated the number of components in the mixture before the parameter optimization step, and it tended to suggest a larger number of components than is required due to the inherent noise embedded in the waveform data. In order to tackle these issues, a new LiDAR waveform decomposition algorithm based on the sequential approach has been proposed in this study and applied to the ICESat waveform data. Experimental results indicated that the proposed algorithm utilized a smaller number of components to decompose waveforms, while resulting IMP value is higher than the GLA14 products.

Key Words : LiDAR, Wave decomposition, Gaussian, EM algorithm, ICESat.

1. Introduction

LiDAR (Light Detection And Ranging) provides information on 3D coordinates of the Earth's surface by actively sending out laser pulses and performing range measurements from the sensor. The range measurements are calculated by measuring the flight time of the laser pulse, and the time measurements are combined with the location (obtained from GPS (Global Positioning System)) and the attitude (obtained from IMU (Inertial Measurement Unit)) information of the system at the time of laser shot. Early LiDAR systems were limited to record small number of returns from the back scattered energy due to the hardware limitation. The volume of data acquired by the LiDAR system is enormous due to its high pulse repetition rate (PRF) of the system, and it was not feasible to record a whole spectrum of the return signal in the early systems. For this reason, early LiDAR systems extracted the location of peaks from the return signals, and the recorded peak locations were transformed into points with 3D coordinates information. However, recent advances in LiDAR hardware now enabled to record back

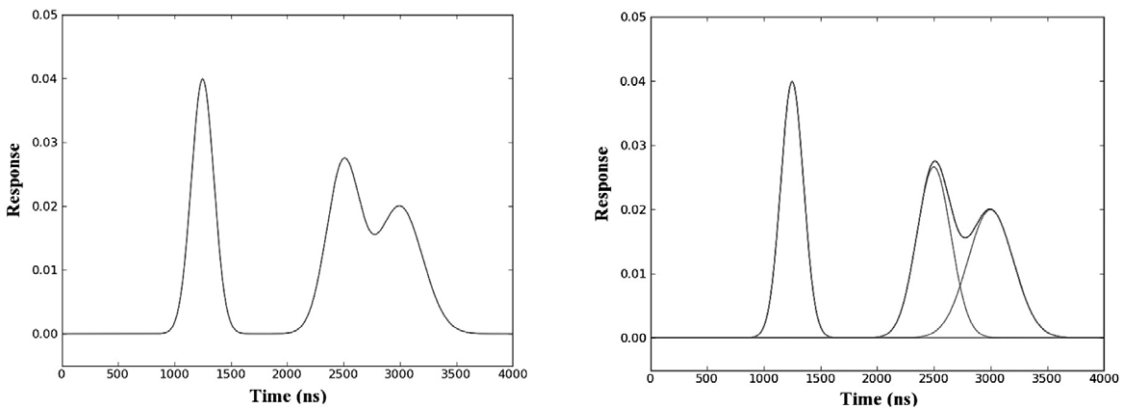
Received December 10, 2010; Revised December 22, 2010; Accepted December 26, 2010.

† Corresponding Author: Sanghoon Lee (shl@kyungwon.ac.kr)

scatter energy which is also called a LiDAR waveform. The full waveform LiDAR system has recently attracted considerable attention of researchers since detailed information on the vertical structure of the targets can be better represented by the LiDAR waveform data than by the traditional discrete return LiDAR data. Various full waveform LiDAR systems have been developed since 1980s. The first LiDAR full waveform digitizer system was developed in the 1980s for bathymetric applications, and topographic full waveform profiling digitizer systems began to be marketed in the 1990s (Mallet and Bretar, 2009). More recently, the Laser Vegetation Imaging Sensor (LVIS) was developed as a prototype for the Vegetation Canopy LiDAR (VCL) mission (Blair *et al*., 1999). In addition to the airborne full waveform LiDAR system, the Ice, Cloud and Land Elevation Satellite (ICESat), a spaceborne large footprint full waveform LiDAR system, was launched in 2003 and operated until recently (Zwally, 2002). Even though full waveform LiDAR data provide high resolution vertical structure information, the high dimensional LiDAR waveform data inherently present challenges in processing the data. One of the most critical steps in processing the waveform data is LiDAR waveform decomposition, and it refers to the process of decomposing a return waveform into a mixture of components which are then used to characterize the original waveform data (Fig. 1). The LiDAR waveform decomposition plays an important role in LiDAR waveform processing since the resulting decomposed components are assumed to represent reflection surfaces within waveform footprints and the decomposition results ultimately affect the interpretation of LiDAR waveform data.

$$\omega(t) = \sum_{k=1}^{N} \alpha_k \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left\{-\frac{(t-\mu_k)^2}{2\sigma_k^2}\right\} + \varepsilon \quad \text{(Eq. 1)}$$

Among various kinds of mixture models, a Gaussian mixture model (Eq. 1) is the most common statistical model for the waveform decomposition process, and its parameters include the recorded time by the digitizer (t), mixing coefficients (a_k) and the mean (μ_k), standard deviation (σ_k) of each component, and background noise (ε). Decomposing a waveform into distinct components by fitting a mixture of Gaussian distributions is an unsupervised machine learning problem which involves two separate, but related problems; i) determining the



(a) (b)

Fig. 1. Example LiDAR waveform and its decomposition results; A solid blue line in (a) represents a received waveform, and solid red lines in (b) represents the decomposed components.

number of components represented in the waveform, and ii) estimating the parameters of the associated Gaussian distributions. Estimating parameters of a Gaussian mixture depends heavily on the estimated number and the initial location of the Gaussian components since the parameter estimation problem does not have a closed form solution and must be solved iteratively. Therefore, it is critical to obtain a good estimate of the number of components and good initial estimates of the parameters of the individual components of the Gaussian mixture in order to achieve rapid convergence to an optimal solution.

Researchers have decomposed the LiDAR waveform into a mixture of Gaussians using various approaches. Hofton *et al*. (2000) utilized inflection points obtained from the second derivatives of the smoothed waveforms to determine the number of components in the mixture and initial parameter estimates. They employed non-linear least squares methods such as Gauss Newton and the Levenberg-Marquardt algorithms to obtain optimal parameter estimates from the initial estimates. Persson *et al*. (2005) utilized local maxima points obtained from the first derivatives of the smoothed waveform to determine the number of components in the mixture and the initial parameter estimates, followed by optimization via the EM (Expectation-Maximization) algorithm. The EM algorithm is a well-known maximum likelihood estimation method which alternates between an expectation (E) step and a maximization (M) step. The E step computes the expectation of the likelihood function evaluated from the currently estimated parameters, and the M step updates parameters so that the updated parameters maximize the likelihood function found on the E step (Dempster *et al*., 1977). Chauve *et al*. (2007) investigated the zero crossing points of the first derivatives of the smoothed waveform in order to estimate initial number of components and derive initial parameter estimates, and then they applied a non-linear least squares method to optimize parameters of different kinds of Gaussian mixture distributions. In addition to non-linear least squares and the EM algorithms, a greedy EM algorithm has been proposed to estimate the density of a multivariate mixture of Gaussians (Vlassis and Likas, 2002). The greedy EM algorithm converges faster than the EM algorithm since parameters from the previous step are fixed and only parameters for the current step are estimated at a given time.

Although significant research has focused on developing waveform decomposition methods, several issues are still remained to be solved. In previous studies, the number of components in the mixture model was estimated before the parameter optimization step, then non-linear least squares or the EM algorithms were applied to derive final parameter estimates of components. Additional component may be added to the mixture when the estimated mixture model generated from the optimized parameters does not satisfy specified criteria. However, these approaches are problematic as the process of estimating the number of components is not only time consuming, but also not robust and tends to suggest a larger number of components than is required due to the inherent noise embedded in the waveform data, especially when the waveform is complex. There are no means of reducing the number of components before the parameter optimization process starts, which sometimes causes over-fitting and long convergence times. The main objective of this study is to develop a new sequential algorithm which decomposes a waveform into a mixture of Gaussians rapidly and accurately, with a focus on application to complex waveforms. The proposed approach is based on an effective utilization of a region growing algorithm in order to derive initial estimates of the new components, and a non-linear least squares

algorithm, a greedy EM algorithm, the EM algorithm, and a sequential EM algorithm for parameter optimization.

2. Methods

The proposed decomposition algorithm utilizes sequential approach, where the goal is to reduce computation, but provide a good approximation to the waveform. The proposed algorithm starts with a single component in the mixture, and an additional component is added to the mixture until stopping criteria are satisfied (Fig. 2). For parameter optimization, different kinds of parameter estimation techniques are utilized depending on complexity of the waveform (Fig. 2). For simple waveforms or well separated mixtures, simple non-linear least squares (Gauss-Newton or Levenberg-Marquardt algorithms) and a greedy EM algorithm are fast and may provide adequate fits. However, convergence or appropriate fits may not be achieved for more complex waveforms, so the more robust EM and sequential EM methods are applied to complex waveforms.

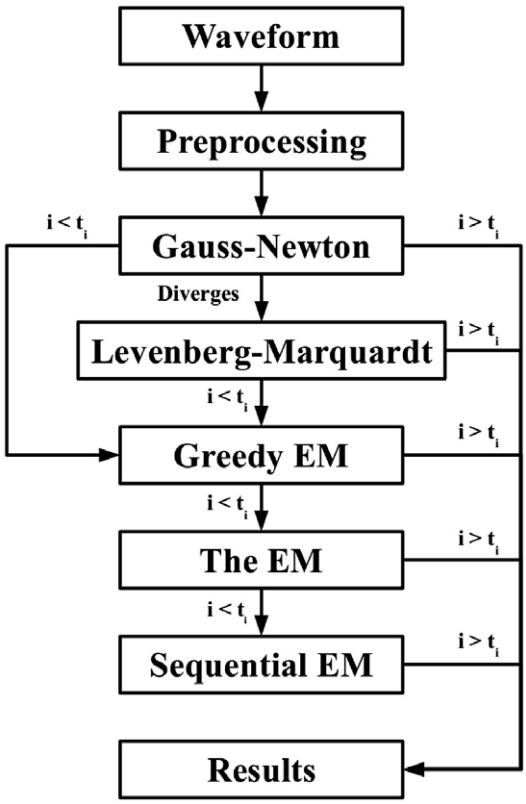


Fig. 2. Flow chart of the sequential waveform decomposition algorithm (i: improvement factor, t_i: improvement factor threshold value).

Since a new component is sequentially added to the mixture and the parameter optimization is performed by iterative approach, initial parameters of the newly added component need to be estimated before the parameter optimization step. A region growing algorithm is utilized to derive the initial parameter estimates for the component that is introduced into the mixture. The region growing algorithm uses location of the maximum response value of the waveform as a seed point to initiate the region growing process. The region starts growing from the seed point to the left and the right simultaneously, and terminates when the response value is smaller than a specified threshold value. This region growing process prevents the cluster from including the portion of the waveform comprised of only noise. Once the region stops growing, initial parameters for the new component are estimated from the grown region and input to one of the curve fitting algorithms: non-linear least squares, greedy EM, or sequential EM. The optimized parameters are then used to reconstruct the corresponding Gaussian component, and the reconstructed component is subtracted from the waveform. If an additional component is required in the mixture, the remaining waveform is used to drive initial parameter estimates of the new component and the above process is repeated until the stopping criteria are satisfied.

Since the proposed algorithm is sequential, stopping criteria are required to determine when to stop adding an additional component to the mixture.

An improvement factor (*IMP*) which is based on the sum of squared error (*SSE*) value is used as one of stopping criteria in this study. The *IMP* represents the goodness of the fit. The *SSE* is computed as the sum of squared differences between the original and the estimated waveforms (Eq. 2), where i represents index term of the waveform, n represents the total number of terms in the waveforms, and k represents number of components in the current mixture. The initial SSE (SSE_0) assumes a value of zero for the estimated waveform for all the values of time (Eq. 3). Then, the *IMP* with k Gaussian components is defined as the ratio of difference between incremental *IMP* when compared to IMP_{k-1} and incremental *IMP* when compared to IMP_0 (Eq. 4). The *IMP* is used to determine whether the waveform approximation is adequate to justify terminating the decomposition process.

$$SSE_k = \sqrt{\sum_{k=1}^{N} (f_i - \hat{f}_{k,i})^2} \qquad \text{(Eq. 2)}$$

$$SSE_0 = \sqrt{\sum_{k=1}^{N} (f_i - 0)^2} \qquad \text{(Eq. 3)}$$

$$IMP_k = \frac{IMP_k - IMP_{k-1}}{IMP_k - IMP_0} \qquad \text{(Eq. 4)}$$

1) Preprocessing

Waveforms recorded by LiDAR systems inherently include noises due to hardware limitation of the sensor and interactions between the laser pulse and atmosphere. In order to improve decomposition results and reduce computational requirements of the decomposition process, preprocessing is performed over waveforms, and it discriminates the "signal" from the "noise." Statistics of background noise such as mean (μ_n) and standard deviation (σ_n) are estimated from the "non-signal" part of the waveform and employed to threshold the waveform and extract the "signal" part, which is then used in the decomposition stage. First, the μ_n value is subtracted from the waveform in order to remove the background noise from the waveform. Second, the first and last location of the waveform whose response value is greater than $3\sigma_n$ are identified, and the part of waveform between the first and the last locations is extracted as the "signal." The extracted signal is then fed into the sequential waveform decomposition algorithm. The pre-processing step interactively determines which part of the waveform represents the signal, and reduces computational requirements by lowering the dimension of the input data to the decomposition algorithm.

2) Sequential waveform decomposition

Two non-linear least squares fitting algorithms - the Gauss-Newton and the Levenberg-Marquardt - are used to fit the first component. Initial parameters for the component are estimated by the region growing algorithm. The Gauss-Newton based algorithm is employed first since it converges faster than the Levenberg-Marquardt algorithm. However, if the Gauss-Newton algorithm diverges, the optimization proceeds using the Levenberg-Marquardt algorithm. The Levenberg-Marquardt algorithm is applied later since the Levenberg-Marquardt algorithm is known to be more robust, but it requires more computation than the Gauss-Newton algorithm. After fitting the first component, the improvement factor is computed and compared to the specified threshold for IMP (t_i). If it is exceeded the specified t_i, the waveform decomposition process terminates, and the optimized parameters are reported as final results. Otherwise, a new component is added to the current mixture using the region growing algorithm, and a greedy EM algorithm is utilized to optimize parameters for the added component. The greedy EM algorithm provides rapid convergence because it fixes the estimated mean and standard

deviation of the previously estimate components and only updates the amplitude of previously added components and parameters of the current component. This process is repeated until the resulting IMP_k is greater than t_i or a number of components in the mixture is smaller than the specified maximum number of components (N_{max}). Parameters estimated by the greedy EM algorithm are reported as final results if no more components are required in the mixture. Otherwise, the process proceeds to the next stage which utilizes the full EM algorithm.

The EM algorithm is initiated using estimated parameters from the greedy EM algorithm without adding a new component to the current mixture. The greedy EM algorithm does not allow updating parameters of previously added components other than the amplitudes of the components in the optimization process, while the EM algorithm introduces more flexibility in the optimization process by allowing updating all parameters (amplitude, mean, and standard deviation) of every component in the mixture. For this reason, the EM algorithm is more robust, but requires more computation than the greedy EM algorithm. After applying the EM algorithm, the resulting IMP_k determines whether a new component should to be added to the mixture. If IMP_k is smaller than t_i, an additional Gaussian component is added to the mixture using the sequential EM algorithm. The sequential EM algorithm derives initial parameters by applying the region growing algorithm to the remaining waveform, and applies the EM algorithm with the additional components. This process is repeated until the IMP_k is greater than t_i, or the maximum number of components allowed in the mixture (N_{max}) is reached.

3. Experimental Results and Discussion

In order to investigate the performance of the proposed algorithm, the proposed algorithm was applied to the Ice, Cloud and Land Elevation Satellite (ICESat) waveform data, and the corresponding decomposition results were compared with the GLA14 products whose decomposition results were generated by the decomposition algorithm proposed by Hofton *et al*. (2000).

1) Experimental data

The ICESat is a spaceborne full waveform LiDAR system. It was launched January 13, 2003, and its primary goal was to measure elevation changes in the arctic and the Antarctic in order to understand how changes in ice-sheet mass balance impact global sea level changes. In addition to this objective, the ICESat mission also enables precise measurement of land topography and provides information on vegetation structure by recording and processing the Geoscience Laser Altimeter System (GLAS) waveforms (Zwally *et al*., 2002). The GLAS instrument on ICESat provides various data products, of which GLA01, GLA06 and GLA14 are used in this study. The GLA01 Level 1 product provides the transmitted and received waveform from the instrument. The GLA06 Level 1 product mainly intended for providing surface elevation, but also includes the saturation index of waveform which implies the number of saturated gates. The GLA14 Level 2 product is intended to represent the potential complexities of returns from land. The GLA14 product reports the number of components in a mixture and the estimated parameters of each Gaussian component obtained from the decomposition algorithm proposed by Hofton *et al*. (2000). The GLA14 product fits a mixture of a maximum of 6 Gaussian components, and provides

the associated estimated parameters of the estimated components. From these estimated parameters, the Gaussian approximation of the waveform can be reconstructed in the direction of the laser pointing vector. In this study, the GLA01 products were used as main input to the proposed algorithm, while saturation index values from the GLA06 products were used to remove saturated waveforms from the experiment. In order to compare the waveform decomposition performance of the proposed algorithm with the previously proposed algorithms, results from the GLA14 product were compared with the decomposition results of the proposed algorithm.

22 GLA01 products were pulled out from the National Snow and Ice Data Center (NSIDC) web site. Each GLA01 product contains 64,400 waveforms, and 1,417,200 waveforms were included in the 22 GLA01 products initially (Table 1). However, waveforms from GLA01 products included not only waveforms targeted over land, but also the waveforms targeted over the ice, and the ocean. Since the goal of this study is to develop a new LiDAR waveform decomposition algorithm with a focus on complex waveforms, waveforms that were targeted over the ice and the ocean were removed from the experiment. In addition to filtering out waveforms that are not targeted over the land, noisy waveforms were also removed from the analysis. Decomposition results from noisy waveforms do not produce meaningful information, and the noisy waveforms only cause long convergence time. Two criteria were used to extract noisy waveforms from the GLA01

Table 1. List of GLA01 products used in the study and number of waveforms in the product (N_w), and number of waveforms selected ($N_{w,selected}$) for later analysis

Index	File Name of GLA01 Product	N_w	$N_{w,selected}$
1	GLA01_028_2111_001_1343_4_01_0001.DAT	64,400	1,512
2	GLA01_028_2111_002_0280_2_01_0001.DAT	64,400	8,028
3	GLA01_028_2111_003_0280_2_01_0001.DAT	64,400	7,336
4	GLA01_028_2113_001_1343_4_01_0001.DAT	64,440	1,860
5	GLA01_028_2113_002_0280_2_01_0001.DAT	64,440	10,694
6	GLA01_028_2115_001_1343_4_01_0001.DAT	64,400	4,251
7	GLA01_028_2115_002_0167_4_01_0001.DAT	64,400	12,395
8	GLA01_028_2115_002_0354_2_01_0001.DAT	64,440	8,326
9	GLA01_028_2115_002_1343_4_01_0001.DAT	64,400	4,388
10	GLA01_028_2115_003_0280_2_01_0001.DAT	64,440	4,701
11	GLA01_028_2117_001_1343_4_03_0001.DAT	64,400	1,768
12	GLA01_028_2117_002_0048_4_02_0001.DAT	64,400	5,159
13	GLA01_028_2117_002_0280_2_02_0001.DAT	64,440	11,156
14	GLA01_028_2117_002_0354_2_02_0001.DAT	64,400	13,421
15	GLA01_028_2119_001_1343_4_02_0001.DAT	64,440	7,790
16	GLA01_028_2119_002_0048_4_02_0001.DAT	64,440	8,469
17	GLA01_028_2119_002_0167_4_06_0001.DAT	64,400	6,336
18	GLA01_028_2119_002_0405_4_02_0001.DAT	64,400	6,433
19	GLA01_028_2121_002_0048_4_03_0001.DAT	64,400	8,177
20	GLA01_028_2121_002_0220_2_02_0001.DAT	64,440	6,760
21	GLA01_028_2121_002_0280_2_02_0001.DAT	64,440	12,782
22	GLA01_028_2121_002_0405_4_02_0001.DAT	64,400	7,927
Sum		1,417,200	159,669

products. One criterion was the saturation index from GLA06 products, and any waveform whose saturation index is greater than 0 was removed from the experiment. Second criterion was based on the maximum response of the waveform. One characteristics of noisy waveform is its low maximum response value. When waveforms are reflected over targets without interference, the maximum response value of the waveform is usually much higher than the noisy part of the waveform. ICESat waveforms are digitized in 8-bit resolution, and the response values range from 0 to 255. Any waveform whose maximum response is smaller than 50 was considered to be a noisy waveform and removed from the experiment. Based on the above selection scheme, 159,669 waveforms among 1,417,200 waveforms were selected and used in the experiment (Table 1).

2) Decomposition results

The proposed LiDAR waveform decomposition algorithm was applied to 159,669 waveforms selected in the data preparation stage. User specified threshold values of t_i = 0.95 and N_{max} = 6 were used in this study. In addition to comparing decomposition results of overall waveforms, it is also important to compare the decomposition results depending on whether waveforms are simple or complex since both the proposed algorithm and the previously developed decomposition algorithm are expected to perform well for simple waveforms. Waveforms were classified into two classes based on the decomposition results of the GLA14 products. Waveforms that resulted in a single component in the decomposition results from the GLA14 products were classified as a simple waveform class, and the number of simple waveforms is listed in the third column ($N_{w,smp}$) of Table 2. Similarly, waveforms that resulted in more than one component in the decomposition results from the GLA14 products were classified a complex waveform class, and the number of complex waveforms is listed in the fourth column ($N_{w,cmp}$) of Table 2. Initially, it was expected that all waveforms targeted over the land would show up in the GLA14 products, but it turned out that several waveforms that were selected in the data preparation stage were missing from the GLA14 products. This was because not all waveforms targeted over the land actually hit the land, especially when waveforms are targeted close to the boundary between the land and the ocean or the ice. The number of missing waveforms selected in the data preparation stage but missing in the GLA14 products is listed in fifth column (N_{na}) of Table 2, and the missing waveforms were not included in the

Table 2. Description of waveform decomposition results

Index	$N_{w,selected}$	$N_{w,smp}$	$N_{w,cmp}$	N_{na}	N_f
1	1,512	458	881	87	86
2	8,028	4,375	3,249	103	301
3	7,336	3,537	2,938	184	677
4	1,860	825	381	324	330
5	10,694	3,488	5,086	149	1,971
6	4,251	2,211	875	633	532
7	12,395	3,789	3,450	3,112	2,044
8	8,326	3,060	3,563	1,398	305
9	4,388	2,153	1,123	893	219
10	4,701	1,700	2,190	91	720
11	1,768	349	1,272	60	87
12	5,159	3,013	1,054	632	460
13	11,156	5,930	5,081	81	64
14	13,421	6,467	5,926	834	194
15	7,790	3,214	3,169	567	840
16	8,469	4,063	2,096	1,845	465
17	6,336	2,558	1,774	1,703	301
18	6,433	2,570	1,923	1,228	712
19	8,177	3,356	2,215	1,902	704
20	6,760	2,371	3,474	387	528
21	12,782	6,288	6,294	47	153
22	7,927	3,095	3,458	691	683
Sum	159,669	68,870	61,472	16,951	12,376

comparison. Even though noisy waveforms were filtered out from the data preparation stage, some of noisy waveforms were still remained, and some waveforms were not successfully decomposed for this reason. In order to minimize effects the noisy waveforms in the comparison, only waveforms that were successfully decomposed by both the proposed algorithm and the GLA14 products were included in the comparison. It was considered as a decomposition failure when 1) the estimated parameters contain any value (*NaN*), or 2) the resulting *IMP* value is negative. The number of waveforms which are not successfully decomposed either in the GLA14 products or by the proposed algorithm is listed in the sixth column (N_f) of Table 2.

3) Decomposition results comparison

Decomposition results of the proposed algorithm and the GLA14 products were compared in term of the average number of components used in the decomposition process and the resulting *IMP* value. In order to compare the decomposition performance of algorithms depending on complexity of waveforms, two different comparisons were conducted in this study. One comparison focused on the decomposition results when all waveforms ($N_{w,smp}$ + $N_{w,cmp}$) were included in the comparison; the other comparison focused on the decomposition results when only complex waveforms ($N_{w,cmp}$) were included in the comparison.

Fig. 3 shows an average number of components used in the decomposition process, and it revealed that the proposed algorithm utilized a smaller number of components for both comparisons, while the proposed algorithm utilized much smaller number of components to decompose complex waveforms. Fig. 4 shows an average IMP value of the proposed algorithm and the GLA14 products, and it indicated that resulting mean IMP value of the proposed

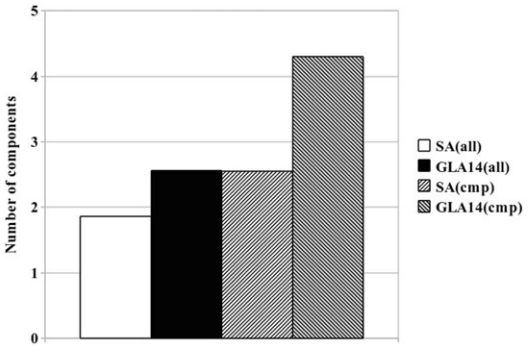


Fig. 3. Average number of components used in the decomposition process (SA(all): Results from the proposed sequential algorithm when all waveforms were included, GLA14(all): Results from GLA14 products when all waveforms were included, SA(cmp): Results from the proposed algorithm when only complex waveforms were included, GLA14(cmp): Results from GLA14 products when only complex waveforms were included in the comparison).

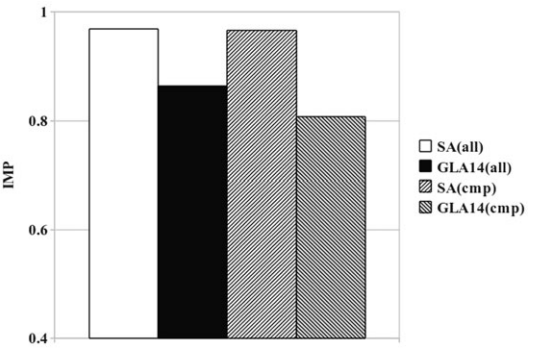


Fig. 4. Average IMP value (SA(all): Results from the proposed sequential algorithm when all waveforms were included, GLA14(all): Results from GLA14 products when all waveforms were included, SA(cmp): Results from the proposed algorithm when only complex waveforms were included, GLA14(cmp): Results from GLA14 products when only complex waveforms were included in the comparison).

algorithm was greater than the GLA14 products for both comparisons, while the difference was increased when only complex waveforms were included in the comparison. In the comparison with all waveforms, the proposed algorithm utilized 29.69 % less number of components to decompose waveforms, while the resulting average IMP is 12.11 % higher than the GLA14 products. In the comparison with only complex waveforms, the proposed algorithm utilized

40.7 % less number of components, while the resulting average IMP is 19.75 % higher than the GLA14 products. In general, the proposed algorithm utilized a smaller number of components to fit Gaussian mixture over the waveforms than the GLA14 products, while it provides better fit to the waveforms. These results indicated that the proposed algorithm outperforms the previous waveform decomposition algorithm in general and its performance improves as complexity of waveforms is increased.

4. Conclusion

LiDAR waveform decomposition plays an important role in LiDAR data processing since the resulting decomposed components are assumed to represent reflection surfaces within waveform footprints and the decomposition results ultimately affect the interpretation of LiDAR waveform data. The previous approaches estimated the number of components in the mixture before the parameter optimization step, and it tended to suggest a larger number of components than is required due to the inherent noise embedded in the waveform data. A new LiDAR waveform decomposition algorithm based on the sequential approach has been proposed in this study and applied to the ICESat waveform data. The experimental results indicated significant improvement over the previous approaches. The proposed algorithm utilized a smaller number of components to decompose waveforms, while the resulting IMP value is higher than the GLA14 products. The experimental also indicated that the improvement was increased when waveforms are complex. The proposed algorithm appears to be a promising direction for processing waveform data acquired over land by ICESat. Potential improvements would be even more significant for the future ICESat II mission, which will require processing of much more data due to longer duty cycles for the sensor. The proposed approach also has potential for adaptation to analysis of waveform data collected by current airborne scanning LiDAR systems, where the number of waveforms is enormous and the processing requirements are prohibitive.

Acknowledgment

This research was supported by a grant (07KLSGC03) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Land, transport and Maritime Affairs of Korean government and by the Kyungwon University Research Fund in 2010.

References

Blair, J. B., Rabine, D. L, and Hofton, M. A., 1999. The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64: 115-122.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Deseilligny, M. P., and Peuch, W., 2007. Processing full-waveform LiDAR data: Modelling raw signals. *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, pp. 102-107.

Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*, 39(1): 1-38.

Hofton, M. A., Blair, J. B., and Minster, J., 2000. Decomposition of Laser Altimeter Waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4): 1989-1996.

Mallet, C. and Bretar, F., 2009. Full-waveform topographic LiDAR: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1): 1-16.

Persson, A., Soderman, U., Topel, J., and Ahlberg, S., Visualization and analysis of full-waveform airborne laser scanner data. *ISPRS Workshop on Laser scanning 2005*, pp. 103-108.

Vlassis, N. and Likas, A., 2002. A greedy EM algorithm for Gaussian mixture learning, *Neural Processing Letters*, 15(1): 77-87.

Zwally, H. J., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirne, and R. Thomas., 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land, *Journal of Geodynamics*, 34: 405-445.