SUB-SPOT LOCALIZATION FOR SPATIAL SUPER-RESOLVED LIDAR

Juan Castorena, Charles D. Creusere

New Mexico State University
Klipsch School of Electrical and Computer Engineering
Las Cruces, NM 88003

ABSTRACT

Third generation LIDAR full-waveform (FW) based systems collect 1D temporal profiles of laser pulses reflected by the intercepted objects to construct depth profiles along each pulse path. By emitting a series of pulses towards a scene using a predefined scanning pattern, a 3D image containing spatial-depth information can be constructed. Achieving super-resolution to resolve finer spatial details is of great interest because the spatial resolution of a LIDAR system is typically limited by the size of the spot on the target. In this study, we consider the problem of resolving range maps to resolutions smaller than the size of the spot using overlapping spots. This overlap provides the additional information needed to locate multiple objects within a spot using sparse source separation, thus achieving super-resolution.

***Index Terms*—** Full-waveform LIDAR, Super-resolution, finite rate of innovation, sparsity, source separation.

1. INTRODUCTION

LIDAR acquisition consists of emitting a laser pulse in a particular direction (corresponding ultimately to a point on a grid) and measuring the full-waveform (FW) reflected signal. This process is repeated in a predefined pattern by emitting pulses towards distinct locations which are mapped onto this 'grid' using a 2D mechanical scanning unit. Measuring FW reflections instead of single time-of-flight (TOF) provides additional information for improved range estimation and scene structure characterization (e.g, inclination, smoothness, vegetation, building roof) [1]. By further processing these contiguous FW profiles of the 3D scene, 3D images containing spatial and depth information can be constructed.

The LIDAR systems we are dealing with in this study are composed of a single pulsed laser source, a mechanical scanning unit (e.g., a rotating mirror), a single photodetector or avalanche photodiode, a high rate A/D converter, a storage media and a positioning system unit (e.g., GPS). Current LIDAR systems are equipped with A/D converters with sampling rates in the range of the GHz. With these rates, individual measured FW signals can achieve high depth/range resolution. Unfortunately, this is not the case for the spatial resolution. In general, this is constrained by the laser spot size and the precision of the mechanical scanning unit [2], [3].

In this paper, we study the problem of achieving spatial super-resolution (SR) LIDAR based on multiple lower resolution observations. This problem has been widely explored in 2D image acquisition systems [4], but to our knowledge very little work has been done with LIDAR. The most intuitive approach to achieve higher spatial resolution is to reduce the size of spot on the target (e.g., 'spot' ≤ 1 m in diameter) and increase the spatial sampling density to hundreds of millions of pulse emissions [3]. Unfortunately, collecting highly dense pulse emissions is very expensive when limited resources are available (e.g., limited storage in a remote sensor location). More importantly, the size of the spot on the target is largely controlled by the laser divergence and the distance to the reflecting objects.

Our acquisition approach to achieve SR consists, instead, of allowing for larger spot sizes (e.g., ≥ 5 m in diameter) while still using a raster scan pattern. The key difference, however, is that we scan the scene here so that the spots corresponding to successive laser pulses are overlapping. Without this overlap, sub-spot localization of multiple reflectors is infeasible because there is simply not enough information localize the individual reflectors. We say here that effective sub-spot localization is *achievable* when successful estimation of the mutual FW components is highly probable.

In essence, our problem is essentially a source separation problem [5]. The similarity comes from the fact that we attempt to separate from the FW signals those components that are common to a set of overlapping pulses. Unfortunately, this is not trivial for SR LIDAR: specifically, the number of sources is typically larger than the number of observations, implying that the resulting system of equations is underdetermined. We can, however, impose some assumptions on the FW signals and on the objects a pulse spot to promote sparsity. These, consist of assuming that the scene of interest is in the far field, that reflections are highly concentrated in time (i.e., small support) and that multiple objects are far-apart (i.e., $\geq 1/2$ pulse length). This latter assumption implies that reflections in the FW signal are disjoint.

The authors would like to thank AlanVan Nevel and Dean Cook from NAVAIR China Lake, CA for providing the LIDAR waveforms used here.

In this paper, we present preliminary results showing that SR reconstruction is possible. Section 2 explains the LIDAR pulse return and scanning model we have adopted while section 3 describes the acquisition approach used to achieve LIDAR SR. Section 4 presents preliminary experimental results illustrating sub-spot localization using synthetically enlarged spots created from real LIDAR, presenting an optimization approach for achieving spatial SR in the reconstructed range map. Finally, our conclusions are discussed in section 5.

2. LIDAR MODEL

2.1. Return pulse model

In our depth/range SR acquisition approach, we assume that the scene is in the far-field. In other words, the range from the sensor to the reflected objects is much larger than the diameter of the projected pulse spot. This assumption forces multiple pulse reflections (if they are present) to be highly concentrated in time. In general, such reflections depend on several factors which together determine how they affect the return pulse waveform. Examples of these factors include the emitted pulse characteristics, the angle of incidence between the pulse and the object, the roughness and reflectivity of the intercepted objects, the sensor-to-intercepted object range, the influence of previously encountered objects (including occlusion) [6]. Fortunately, we can model FW reflections using parametric signals. Specifically, in [7] we approximated FW reflections as the sum of a bandlimited and a non-uniform linear spline. In other words,

$$x_{n_{FW}}(t) = x_{n_{bl}}(t) + x_n(t). \tag{1}$$

Here, $x_{n_{FW}}(t)$ denotes the FW reflection, t and n denote, respectively, the time and the index of the spatial location towards which the pulse was emitted. The bandlimited component $x_{n_{bl}}(t)$ describes the general signal level while the non-uniform linear spline $x_n(t)$ models the multiple reflections from the objects intercepted by the laser pulse. Since $x_n(t)$ contains all of the information describing the pulse and its interaction with the scene, we base our analysis only on this signal component and discard $x_{n_{bl}}(t)$. This linear nonuniform spline component can be approximated as

$$\hat{x}_n(t) = \sum_{k=1}^{K} a_k \varphi(t - t_k) \tag{2}$$

where $\varphi^{(R+1)}(t - t_k) = \delta(t - t_k)$ with order $R = 1$ represents the linear spline functional and K is the number of knots used for the spline approximation. This problem is equivalent to that of estimating K dirac locations t_k along with weights a_k, and it can be solved very efficiently using the finite rate of innovations (FRI) method as described in [7]. In addition, the FRI approach allows us to sample based on the number of K parameters thus achieving high sampling efficiency. The recovered and sampled FW signal and its components are denoted by $x_{n_{FW}}, x_{n_{bl}}, x_n \in \mathbb{R}^{1\times P}$, respectively. Here P, denotes the number of samples implied by the FRI approach.

Since the sampled modes in the FW signal can be approximated by linear splines, we can use the matrix Ψ given by

$$\Psi = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & 0 \\ \text{P-1} & \text{P-2} & \text{P-3} & \ddots & 0 \\ \text{P} & \text{P-1} & \text{P-2} & \cdots & 1 \end{bmatrix} \tag{3}$$

(i.e., the double integration matrix operator). Using (3) and denoting the sparse signal innovations by γ_n allows us to express the linear spline components corresponding to the pulse reflections as $x_n = \gamma_n \Psi^{\mathrm{T}}$. As before, n denotes an index of the corresponding element in the grid at the given resolution.

2.2. FW complexity

LIDAR compressive sampling approaches require that one develop a model of the scene that supports sparse realizations in a realistic manner [8], [9], [10]. This is very complicated problem, however, because such models depend on many factors: e.g., on the scene's three dimensional structure (e.g., depth/range, inclined terrain, planar terrain, multiple objects, reflectivity of the materials) and on the emitted pulse characteristics (angle of incidence, the fps, pulse energy variations across the fps, divergence) [6], [1]. All of these factors contribute to the characterization of the scene response and, as such, a general model is difficult to create. Here, we exploit instead FW reflection characteristics. In [11] the authors effectively characterized FW signals as a function of scene structure (e.g., smooth, detailed and random) and spot size. This characterization allowed us to relate the FW return pulse complexity to the complexity of the scene.

In this work we consider only reflections from multiple objects whose ranges are separated by at least $1/2$ the pulse length. Such an assumption implies that the support of each of the reflections is disjoint in the FW signal. Since we assume the scene is in the far-field, reflections are highly concentrated in time. This implies that the response within the FW waveform of each object within the laser spot has very short temporal support. This property is assumed to be preserved regardless of the incidence angle between the pulse and the reflecting object.

Ideally, the FW waveform response generated by each specific object within the laser spot is a dirac function. Due to the fact that reflecting surfaces are generally not perpendicular to the angle of incidence of the laser pulse, variations are typically present. Fortunately, these are expected to be highly concentrated in time and of small support. We also assume here that scenes are generally smooth. This means that FW signals will contain, in general, information about a single reflection which implies that data is highly compressible. Although scenes are assumed to be generally smooth, scene discontinuities may appear at object boundaries and thus some FW signals may contain data from multiple reflections.

2.3. Scanning

Scanning refers to in here as the process of emitting a series of pulses, each in a slightly different direction so as to cover the scene of interest. Some examples of scanning patterns typically used are the raster and the palmer scans. To define the locations towards which a pulse is pointed and emitted prior to acquisition, a grid which covers the entire scene of interest is defined. This grid is composed of three dimensional elements with two denoting spatial coordinates and the third denoting the time index of the return pulse. The resolution of each element in the spatial dimension is determined by the spot size whereas the resolution of the third dimension is determined by the temporal reconstruction possibilities implied by the FRI recovery. Unfortunately, the size and geometry of the pulse spot (typically a circle) varies based on the range, the laser beam-divergence and the topography of the scene [2]. For now, we assume that the spot size is fixed regardless of the scanning geometry and the topography encountered.

The whole set of FW reflections obtained by scanning is denoted by $X = [x_1, x_2, ..., x_N]^T \in \mathbb{R}^{N\times P}$, indexed in the acquisition order of the raster scan. Here, N denotes the total number of pulse emissions in the scan. Each emitted pulse x_n projects a spot covering the grid element indexed by n and is directed towards the center of the corresponding grid element. Thus, the set of sampled FW reflections is expressed as

$$X = \Gamma\Psi^{\mathrm{T}} \tag{4}$$

where $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_N]^T$ is the set of signal innovations.

3. SPATIAL SUPER-RESOLUTION

3.1. Acquisition and Model

To begin our discussion about SR, we assume we have a set of FW reflections denoted by $Y = [y_1, y_2, ..., y_M]^T \in \mathbb{R}^{M\times P}$, indexed in the same order as X in (4). Each y_m corresponds to a 'large' spot pulse covering several smaller 'sub-spots'. These sub-spots are assumed to be generated by pulse reflections embedded in x_n. Thus, the problem of SR can be posed as the problem of recovering X based on the observations

$$Y = \Phi X = \Phi\Gamma\Psi^{\mathrm{T}}. \tag{5}$$

where $\Phi \in \mathbb{R}^{M\times N}$ with $M < N$ is referred to as the observation matrix. Note that (5) implies a mapping from a grid whose resolution is given by the size of the smaller sub-spots corresponding to X to a grid with larger and fewer elements corresponding to Y. We say SR is feasible if we can recover X given Y from the underdetermined system in (5).

Because the spatial information inherent in the reflected pulses is lost when a single photodetector is used, we resort here to emitting pulses with overlapping spots. Doing this allows us to localize multiple objects within the larger spot, thus achieving SR. The localization is performed by estimating the reflected components mutual to the overlapping pulses. The characteristics of this overlap are modeled in the observation matrix Φ. Within this model, we can incorporate the characteristics of the pulse; for example the amount of overlap, the energy distribution of the pulse as a function of space (e.g, typically modeled by a Gaussian distribution) and changes in the spot size and geometry due to off nadir incidence and inclined objects. In some situations, these characteristics are unknown and thus there is some uncertainty about Φ. Fortunately, we can resort to the techniques developed in blind source separation [12], [13] to estimate Φ. In this paper, however, our objective is simply to show that SR is possible and thus we limit ourselves to a very simple model of Φ.

With no assumptions made on X, recovery from the underdetermined system in (5) is very difficult. However, we can exploit the structure in X to construct optimization programs that recover X exactly. In general, note that X contains the set of FW reflections and that Φ considers the columns $X^{(i)}$ of X jointly for $i \in (1, P)$. Thus, based on the assumptions described in section 2.2—namely that the scene is in the far-field, that reflections are highly concentrated and that each of these reflections is disjoint with each other—we can enforce some strong constraints on the structure of X. The first two assumptions imply that X is, in general, very sparse. The reflections embedded in the FW return pulses of X will tend to occur in blocks and to be of very small support. The third assumption implies that reflections from multiple objects in one FW return pulse are not likely to have the same support as similar reflections in the remaining FW pulses in X. Alternately, one could say that a reflection's support will be highly likely to be disjoint with the supports of other reflections.

If one further modifies (5) using the inverse of the second order operator Ψ, then one can instead analyze

$$Y_\Psi = Y(\Psi^{\mathrm{T}})^{-1} = \Phi\Gamma. \tag{6}$$

Under this form, the problem becomes recovering Γ which is more sparse than X in general. As such, it is highly probable that the joint sparsity of the columns $\Gamma^{(i)}$ is higher, and this, in turn, implies that recovery can be improved.

4. A SIMPLE EXAMPLE

The data used here was obtained from NAVAIR China Lake, CA and collected using the VISSTA ELT LADAR system. The pulse emission rate of the system is 20 Khz where each pulse is of 1.5 ns duration (calculated using 1/2 the maximum amplitude). Reflections are sampled at a rate, of 2 GHz and quantized using an 8-bit A/D converter. The dataset used here was obtained by imaging a pickup truck through a chain like fence both positioned approximately perpendicular to the pulse transmission path. Figure 1 illustrates the range map obtained by processing the collected FW signals using the VISSTA system.

Fig. 1. 3D image with spatial resolution and ranges.

To show the feasibility of SR recovery, we simulate two overlapping pulses of large spot size using the collected data. The spots of these two simulated pulses cover the same area as three pulses with smaller spot sizes. The overlap we consider corresponds exactly to the size of one of the pulses of small spot size. These FW's are chosen from the locations in Figure 1 at the boundary between the chain link fence and the truck. Based on this, we use $X \in \mathbb{R}^{3\times P}$ shown in Figure 2c and obtain $Y \in \mathbb{R}^{2\times P}$ shown in Figures 2a and 2b using (5) and the simple observation matrix model

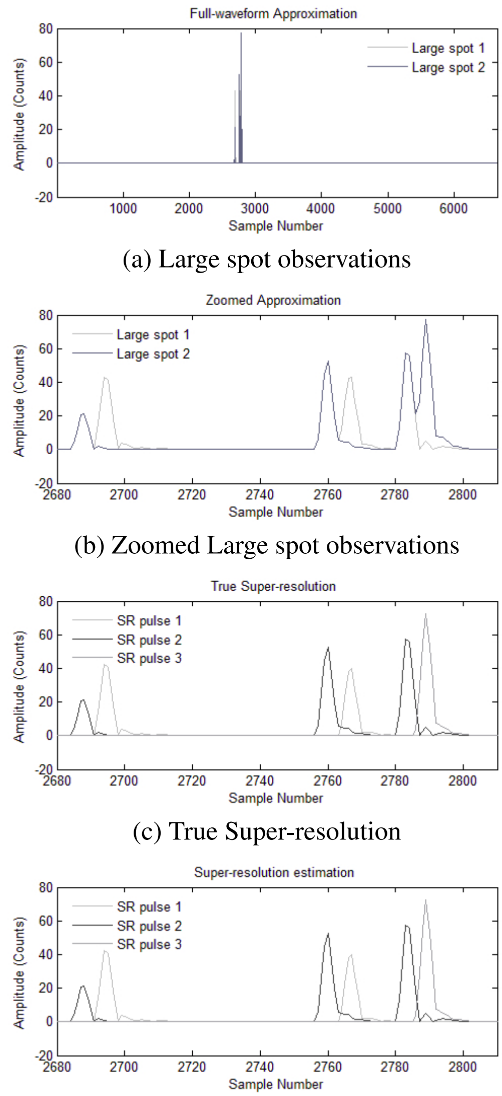
$$\Phi = \frac{1}{3}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{7}$$

As was stated earlier in (1), we use here only the part of our model pertaining to the spline approximation given by X. After Y is obtained, we assume no knowledge about X other than sparsity, and attempt to recover X based on the following optimization program

$$\min \|X^{(i)}\|_1 \text{ subject to } \|Y^{(i)} - \Phi X^{(i)}\|_2 \leq \epsilon. \tag{8}$$

Since our goal is to find X, we independently recover the columns of X denoted by $X^{(i)}$ for $i \in (1, P)$. To speed up the recovery process and based on the fact that typical FW reflections are highly concentrated in time and in the far field, we avoid recovering $X^{(i)}$ when the columns $Y^{(i)} \in \mathbb{R}^{M\times 1}$ have zero energy. In addition, the case when both observations in $Y^{(i)}$ are very similar is also trivial. The program in (8) is repeatedly computed for all $i \in (1, P)$.

The result of applying (8) on the observations Y is shown in Figure 2d. Note that very precise recovery is achieved as can be seen by comparing the obtained result with the true FW signals shown in 2c. In general, Figure 2d indicates that our SR acquisition approach and recovery algorithm is capable of separating, and thus localizing, the part of the signal pertaining to the reflection occurring in the spot overlap. This in turn implies that reflections corresponding to the non-overlapping regions are also localized. Therefore, sub-spot localization was achieved here, demonstrating the possibilities of our approach for achieving LIDAR SR.



(a) Large spot observations

(b) Zoomed Large spot observations

(c) True Super-resolution

(d) Super-resolution recovery

Fig. 2. Example showing sub-spot localization.

5. CONCLUSION

In this research, we present a LIDAR acquisition approach which consists of the collection of FW reflections of large and overlapping laser pulse spots to achieve spatial super-resolution (SR). The overlap is used to localize multiple objects within the laser spot, thus achieving sub-spot localization. For localization, source separation for sparse sources was implemented successfully. This sparsity condition applies under the assumptions that the scene is in the far-field and that multiple-objects are not to close to each other. A simple example with real data was included, demonstrating the potential of our acquisition approach to achieve SR. The results show that given the assumptions stated here, precise recovery of the SR map is possible.

6. REFERENCES

[1] C. Mallet, F. Bretar, M. Roux, U. Soergel, and C. Heipke, "Relevance assessent of full-waveform lidar data for urban area classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 71–84, October 2011.

[2] Y. Sheng, "Quantifying the size of a lidar footprint: A set of generalized equations," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 419–422, July 2008.

[3] B. Schwarz, "Lidar mapping the world in 3d," *Nature Photonics*, vol. 4, pp. 429–431, July 2010.

[4] S.C. Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.

[5] J.F. Cardoso, "Source separation using higher order moments," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1989, pp. 2109–2112.

[6] C. Mallet, F. Lafarge, M. Roux, U. Soergel, F. Bretar, and C. Heipke, "A marked point process for modeling lidar waveforms," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3204–3221, December 2010.

[7] J. Castorena and C.D. Creusere, "Compressive sampling of lidar: Full-waveforms as signals of finite rate of innovation," in *The 20th European Signal Processing Conference (EUSIPCO-2012)*, August 2012, pp. 984–988.

[8] A. Kirmani, A. Colaco, F.N.C. Wong, and V. K. Goyal, "Exploiting sparsity in time of flight range acquisition using a single time-resolved sensor," *Optics Express*, vol. 19, no. 22, October 2011.

[9] G.A. Howland, P.B. Dixon, and J.C. Howell, "Photon-counting compressive sensing laser radar for 3d imaging," *Applied Optics*, vol. 50, no. 31, pp. 5917–5920, November 2011.

[10] P. T. Boufounos, "Depth sensing using active coherent illumination," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 5417–5420.

[11] J. Castorena, C.D. Creusere, and D. Voelz, "Using finite moment rate of innovation for lidar waveform complexity estimation," in *IEEE Conference Record of the Fourty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2010, pp. 608–612.

[12] A. Belouchrani, K. Abed-Meraim, and J.F. Cardoso, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[13] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001.