Advances in Single-Photon Lidar for Autonomous Vehicles
Joshua Rapp, Julián Tachella, Yoann Altmann, Stephen McLaughlin and Vivek K Goyal

Abstract

The safety and success of autonomous vehicles (AVs) depend on their ability to accurately map and respond to their surroundings in real time. One of the most promising recent technologies for depth mapping is single-photon lidar (SPL), which measures the time of flight of individual photons. The long-range capabilities (kilometers), excellent depth resolution (centimeters), and use of low-power (eye-safe) laser sources renders this modality a strong candidate for use in AVs. While presenting unique opportunities, the remarkable sensitivity of single-photon detectors introduces several signal processing challenges. The discrete nature of photon counting and the particular design of the detection devices means the acquired signals cannot be treated as arising in a linear system with additive Gaussian noise. Moreover, the number of useful photon detections may be small despite a large data volume, thus requiring careful modeling and algorithmic design for real-time performance. This article discusses the main working principles of SPL and summarizes recent advances in signal processing techniques for this modality, highlighting promising applications in AV as well as a number of challenges for vehicular lidar that cannot be solved by better hardware alone.

INTRODUCTION

Humans are not particularly good drivers. When operating multi-ton vehicles at high speeds, any impairment, fatigue, or other diversion of the driver’s attention can lead to catastrophic results. Over 35,000 deaths and 2.4 million injuries resulted from automotive accidents in 2015 in the United States alone [1]. Autonomous vehicles (AVs) thus offer the promise of a revolution in transportation with the significant potential benefit of improved safety on the road. Countless other aspects of society will likewise be affected by no longer needing human drivers to control vehicles. Mobility can be improved for those unable to drive themselves due to age, inexperience, medical conditions, or lack of confidence. The economics of vehicle ownership will likely be transformed as humans transition from operators to passengers. Before AVs can reshape the transportation
landscape, numerous challenges have to be addressed. Engineers must first demonstrate the ability of AVs to perform basic driving functions safely and reliably. While driving, AVs must translate high-level goals such as route planning, safe driving, efficient energy consumption, and adherence to rules of the road into low-level decisions about the mechanics required to follow a planned trajectory. To achieve these goals, the complex control systems determining driving decisions must rely on information from sensing systems that are as good as—or preferably better than—human perception.

Autonomous system designs usually propose some form of sensor fusion, combining the strengths of various sensing modalities to overcome their individual limitations [2], [3]. In addition to classical GPS receivers, future commercial AVs are likely to include video cameras for identifying road signs and objects, short-range ultrasound sensors as currently used for parking assistance, and weather-robust radar for low-resolution position and velocity estimation. However, the centerpiece of most AV sensing systems is the lidar unit. Just as radar detects echos from radio-frequency electromagnetic waves, lidar detects reflections from optical-frequency laser illumination to generate a long-range, high-resolution point cloud corresponding to the positions and reflectivity values of millions of points in the surrounding environment. Since optical wavelengths can be easily focused into narrow beams, lidar can distinguish much smaller objects than radar, which is crucial for navigating alongside pedestrians, cyclists, and other potential hazards. Although laser ranging has been under development since the 1960s, mostly for military use, terrain mapping, or atmospheric monitoring [4], commercial lidar development has greatly accelerated since 2005, when all vehicles that completed the DARPA autonomous driving Grand Challenge employed lidar for depth mapping.

Sensing systems for AVs are evaluated on several criteria considered necessary for reliable, real-time performance, including the maximum operating range, transverse and longitudinal resolution, field of view, refresh rate, transmit power (especially with respect to eye safety), robustness to ambient light and weather conditions, processing requirements, and cost [2], [5]. Many of these factors are determined in part by hardware and manufacturing constraints, such as the quality and capability of lasers, detectors, and scanning mechanisms, which have improved through continuous investment and refinement. For instance, the cost of a lidar unit was long considered a barrier to widespread deployment in AVs, since the Velodyne systems originally used in the
Grand Challenge cost upwards of $75,000 [6], but after more than a decade of development, lidar manufacturers such as Luminar have announced units priced below $1000. On the other hand, the maximum range is one of the most crucial aspects that cannot be solved through hardware alone. The range capability is important because it affects driving performance directly: early identification of a potential hazard gives the vehicle more time to make safe driving decisions. At American highway speeds of approximately 105 km/h, for instance, a range extension of 30 m provides an additional second of reaction time. The baseline performance necessary for AVs is generally listed as the ability to detect dark objects with a reflectivity of 10% from a distance of at least 200 m. Amplitude-modulated and frequency-modulated continuous-wave lidar methods (AMCW and FMCW) have fundamental challenges restricting the maximum range of most implementations to fewer than 100 m [5]. Although the maximum range of pulsed lidar can be extended by simply increasing the laser power, vehicle power and eye safety constraints limit the feasibility of this approach.

The rapidly evolving technology of single-photon lidar (SPL) has potential to overcome the range problem. Instead of trying to sample the full-waveform transient response from a pulsed lidar illumination, SPL uses detectors that are sensitive to individual photons, enabling sensing of objects from extremely weak reflections. Such extreme sensitivity allows SPL to use eye-safe lasers and fast acquisitions while tolerating large attenuations due to very long distances or fog. For example, experiments with SPL have been used to form point clouds from standoff distances of several hundred meters up to 45 km [7], [8]. SPL is currently under commercial development for AVs by companies such as Ouster and Argo AI, which has partnered with both Ford and Volkswagen.

Beyond the specialized hardware required to detect and time-stamp individual photons, imaging with so little signal relies heavily on advanced signal processing techniques tailored to the acquisition system. Importantly, SPL systems do not conform to standard signal processing assumptions: single-photon lidar measurements are neither linear nor time-invariant, and the inherent randomness of the observations cannot be simply described by Gaussian noise models. In 2014, Kirmani et al. introduced the concept of photon-efficient imaging [9], showing that it is possible to form accurate 3D and reflectivity maps from only one photon detection for each transverse pixel location, even with half of the detected photons due to uninformative
ambient light. From a computational viewpoint, successful point-cloud reconstruction from very few photons requires two main components:

1) an acquisition system model, used to determine the probabilistic nature of any photon detection time, and

2) a scene model, used to reconstruct a point cloud by taking advantage of useful priors.

This article focuses on these two components in the context of SPL for autonomous vehicles. We begin with a description of the physical hardware, which is used to inform the acquisition system model. We then survey a number of challenges for vehicular lidar that cannot be solved by better hardware alone, including low photon counts, obscuring media, partial occlusions, etc., and present state-of-the-art signal processing solutions. Finally, we look ahead to functionalities that can be developed for SPL beyond the capabilities of the human visual system, such as anticipatory route planning enabled by seeing around corners.

**BASICS OF SINGLE-PHOTON LIDAR SYSTEMS**

Single-photon lidar comprises three main components: an illumination source, a single-photon detector, and fast timing electronics. Diode lasers are relatively inexpensive to manufacture and can achieve root mean square (RMS) pulse widths on the order of a few dozens of picoseconds. The most common choice of detector for SPL is the solid-state single-photon avalanche diode (SPAD), which consists of a reverse-biased photodiode biased above the breakdown voltage so that an individual photon incident on the SPAD can cause an avalanche of electrical charge carriers that is directly detectable as a digital signal.

SPL is performed using a technique called time-correlated single-photon counting (TCSPC), originally developed for fluorescence lifetime imaging microscopy. The basic idea of TCSPC is that of a stopwatch: the laser starts a timer with each illumination pulse, and the timer is stopped with the detection of a photon. The time difference between the stop and start signals gives the photon’s time-of-flight, which is easily converted to a measurement of the round-trip distance through multiplication by the speed of light. Due to timing uncertainty and the presence of nuisance detections of ambient light, illumination is repeated to build up a histogram of photon detection times, from which more reliable depth estimates can be determined.
Transverse resolution is achieved using either the illumination or detection or both, each approach presenting different advantages and drawbacks. SPL systems have conventionally employed raster-scanned illumination, as illustrated in Fig. 1. A laser aimed at one spot in a scene repeatedly pulses for a certain dwell time before being redirected to the next spot. While laboratory experiments typically use a simple pair of XY galvo mirrors, and the original Velodyne systems physically rotate the lasers and detectors for a 360° field-of-view, rugged beam-steering with fewer mechanical components is an area of ongoing industrial development, with current approaches employing spatial light modulators such as optical phased arrays, MEMS mirrors, or liquid crystal metasurfaces [6]. Scanned illumination enables the use of a single-pixel or bucket detector, which is often in a confocal configuration (focused and co-axially aligned with the laser) to limit the number of photons undergoing multiple bounces or originating from ambient sources from being detected. Unfortunately, raster scanning is an inherently slow, serial process. A more desirable approach is to broadly illuminate a swath of the scene and achieve spatial resolution with an array of single-photon sensitive elements [10]. While detector arrays promise
faster, parallelized acquisition, existing arrays are still limited in their resolution, and the laser power broadly diffused over a larger area further reduces the signal strength received at each pixel location. One compromise is the use of a line illumination and line array detector, which reduces the spatial scanning to a single dimension and limits the diffusion of the laser power [11].

ACQUISITION MODELING

Despite the variations in acquisition hardware, in general each pixel will be associated with a stream of photon detection times. The time of arrival of a photon at the detector and whether that photon will be registered are inherently random processes, so the first step in making use of SPL data is to understand the probability distribution of the detection times. Factors affecting the detection time distribution can be divided into two groups based on optical or electrical interactions.

Optical Interactions

Optical factors encompass all influences on the light reaching the sensor, including the laser illumination, atmospheric attenuation, surface reflections, and ambient light. We consider the optical model for a single point in a raster-scanned lidar implementation, which is mostly analogous to light reaching a single pixel in a detector array. The laser illuminates a short-duration pulse with shape $h(t)$. Any surface within the beam width will reflect some light back, delayed by a time dependent on the distance $d_m$ to surface $m$, and attenuated by the reflectivity $r_m$, which includes the effects of the surface albedo, view angle, and optical illumination intensity. In addition, ambient light at the same wavelength as the laser—or more generally, within the spectral range of the detector—will reach the detector, although the timing of that incident light does not contain any depth information. The timing of photons incident on a detector is well-understood to be described by a Poisson process. Because the illumination is periodic with period denoted $t_r$, the process intensity in one period is approximately given as

$$\lambda(t) = \lambda_b + \sum_{m=1}^{M} r_m h(t - 2d_m/c), \quad \text{for } t \in [0, t_r),$$

(1)

where $\lambda_b$ is the intensity of the ambient light at the illumination wavelength, assumed to be constant for the duration of the acquisition, $M$ is the number of surfaces present in the field of
view, and \( c \) is the speed of light (around \( 3 \times 10^8 \) m/s in air). The first term represents uninformative background noise, while the sum combines all informative signal contributions.

**Acquisition Electronics**

The acquisition electronics determine which photon detection events are registered and what time information is stored for each event. For instance, not all photons incident on the SPAD actually cause the generation of a photoelectron, whereas thermal noise may cause spurious detections known as dark counts. Furthermore, the timing jitter of the circuitry in response to an incident photon will effectively increase the pulse width of \( h(t) \). These effects tend to be minor in practice and do not affect the processing approach. We thus typically consider \( \lambda(t) \) to describe the detection process. The photon flux, or the expected number of photons incident in one cycle, is given by \( \Lambda = \int_0^r \lambda(t) \, dt \), so the number of photon detections after \( n_r \) illumination cycles is a Poisson random variable \( N \sim \mathcal{P}(n_r \Lambda) \). As described later in the section “Detector Dead Times,” this model is only sufficiently accurate in the low-flux regime (i.e., \( \Lambda \ll 1 \)).

More fundamental to acquisition process is the effect of temporal quantization. Each photon detected by the SPAD is assigned a time stamp, where the resolution \( \Delta t \) is dictated by the TCSPC electronics. Rather than processing each of the \( N \) photon detection times \( t_1, \ldots, t_N \) separately, a histogram \([y_1, \ldots, y_{N_k}]\) of \( N_k \) time bins is typically constructed using the detection times at each pixel \((i, j)\). The observation model for each time bin is then a Poisson random variable, where the parameter integrates the Poisson process intensity over that bin:

\[
y_k \sim \mathcal{P} \left( \int_{(k-1)\Delta t}^{k\Delta t} n_r \lambda(t) \, dt \right), \quad \text{for } k = 1, \ldots, N_k.
\]

When considered together, forming a histogram for each of the \( N_i \times N_j \) transverse pixel locations yields a 3D data cube.

**Estimation Basics**

Reflectivity, depth and background information can be estimated from the photon detections according to model (2). The standard estimation procedure assumes a single surface within each
pixel \((M = 1)\) and known background rate \(b = \lambda_0 \Delta t\). Under these assumptions, the maximum likelihood (ML) reflectivity estimate \(\hat{r}\) is easily given as

\[
\hat{r} = \frac{N - bN_k}{\int_0^{\tau_r} h(t) \, dt}.
\] (3)

Finally, the ML depth estimator \(\hat{d}\) is given by cross-correlating the detection time histogram with the logarithm of \(h(t)\), also known as the log-matched filter:

\[
\hat{d} = \frac{c}{2} \arg \max_{\tau \in [0, \tau_r]} \sum_{k=1}^{N_k} y_k \log [\hat{r}h(k\Delta t - \tau) + b].
\] (4)

**Challenging Estimation Conditions**

Unfortunately, the widely varying environments encountered by an AV and the need for fast acquisition often result in conditions that can make reflectivity and depth estimation more challenging. In many cases, conventional pixel-wise estimation (3) and (4) fails under these conditions, motivating the algorithms described later in this article. A number of these conditions are depicted in Fig. 2 and described as follows.

*Few Photons:* The number of detected photons \(N\) may be small or even zero for several reasons: the number of illuminations \(n_r\) is kept low for real-time acquisition, the surface reflects very little light because it is weakly-reflective or far away, etc. Demonstrations of photon-efficient SPL used exactly one photon per pixel (ppp) [9] or an average near 1.0 ppp [12], [13], resulting in many pixels with no detections.

*Strong Ambient Light:* Estimation is particularly challenging if the ratio between the number of photons due to the laser and ambient illuminations, referred to as the signal-to-background ratio (SBR), is low. Even though optical methods (e.g., confocal configurations, bandpass filters) are used to limit the amount of ambient light that reaches the detector, strong daylight, especially when combined with a weak surface reflection, can result in far more detection events associated with background photons than from signal photons.

*Absence of Surfaces:* The most basic 3D reconstruction methods assume a single surface at each pixel location. If a pixel has no object in its line of sight \((M = 0)\), then the histogram contains only background detection events.

*Multiple Surfaces:* On the other hand, there may be reflections from multiple surfaces present at one pixel \((M > 1)\). This may occur because the light passes through a semi-transparent material
Fig. 2: Examples of recorded histograms: (a) ideal case, (b) few photons, (c) strong background illumination, (d) absence of a target, (e) multiple surfaces per imaged pixel, (f) broadening of the impulse response, (g) highly attenuating media, (h) coarse quantization, and (i) dead-time effects. The observed photon counts are shown in blue, whereas the *a priori* unknown Poisson intensity from (1) is shown in red.

such as glass. Alternatively, the pixel size or field of view increases with distance (e.g., due to the laser divergence in a scanned setting), so the spot is more likely to cover multiple surfaces. This same principle is often used in foliage-penetrating airborne lidar used for terrain mapping.

*Pulse Broadening:* Surfaces are generally assumed to be opaque and approximately normal to the illumination beam so that the reflected temporal response closely resembles the shape of the illumination pulse. However, sub-surface scattering or oblique-angled surfaces, especially at long distances, will return broadened pulse profiles [14].

*Attenuating Media:* Particles in the beam path, such as fog, smoke, rain, or snow, affect the acquired light by scattering photons in different directions after both the illumination (forward path) and reflection (return path). To some extent, the result is similar to that of a signal weakened by additional attenuation and increased background due to scattered photons [15], although the near-range effects of scattering also reshape the temporal distribution of background, with more detections at earlier times [16].

*Challenges due to Nonideal Acquisition Electronics*

Although the effects of the electronics have thus far been treated as minor, further challenges requiring more careful modeling arise if the actual electronics significantly deviate from the ideal.
Coarse Temporal Quantization: The ability to accurately resolve transient information depends on the width of the histogram bins. For raster-scanning systems, the bin resolution that can be achieved currently is on the order of picoseconds, which is typically much less than the duration of pulse $h(t)$, so quantization effects on the depth estimation can be neglected. However, the timing resolution of detector arrays is usually coarser for each element than for a single-pixel device due to hardware and readout constraints. The single-photon-sensitive elements and timing electronics can easily be constructed as separate elements for a single pixel, whereas in 2D arrays, the timing electronics must be integrated on-chip for each pixel, resulting in a trade-off between the fill factor of the photo-sensitive detector and timing components. This becomes particularly problematic if the bin size $\Delta_t$—the least-significant bit of the temporal quantization—is larger than the duration of pulse $h(t)$. In that case, the depth resolution that can be achieved is quantization-limited and can make object detection and recognition more difficult (see Fig. 2(h)). One proposed solution is to use longer-duration illumination pulses, which distributes photon detection events from the same surface over multiple time bins [10]. This acts as a sort of non-subtractive dithering, which through averaging of enough detections can provide sub-bin resolution. A more photon-efficient strategy is to take advantage of the short pulse durations and implement subtractive dithering via time shifts in the system synchronization. By inserting sub-bin delays into the synchronization signal between the laser pulse and the detection timing, efficient estimators can recover depths with sub-bin resolution [17].

Additional Array Considerations: Another current limitation of array manufacturing constraints is spatial non-uniformity. While raster-scanning imaging with a single-pixel detector has essentially identical system properties for each laser location, array elements have neither the same light sensitivity nor identical noise characteristics across the device. In particular, arrays often present “hot” pixels with overwhelming numbers of dark counts, or “dead” pixels with inadequate light sensitivity. The inputs from these pixels must then be omitted, or at least accounted for, in the reconstruction process.

Detector Dead Times: Unfortunately, the circuitry required for SPADs to be single-photon sensitive both precludes the ability to resolve numbers of photons and also requires a reset period or dead time following each detection. Photons incident on a SPAD during a dead time are missed, and the dependence of a detection on the most recent detection time means the
sequence of detection times can no longer be described by a Poisson process. Missed detections occur most frequently in the high-flux regime when $\Lambda \ll 1$ is no longer valid (e.g., when imaging bright objects such as retro-reflective street signs that reflect many photons), and cause distortions of the detection time histogram that may result in erroneous depth and reflectivity estimates, thereby making accurate localization or object recognition more difficult. The simplest way to avoid dead time distortions is to attenuate the incident light to reduce the probability of missed photons, but attenuation is impractical for automotive lidar, since it is already a challenge to recognize dark and distant objects in real time without additionally weakening the input signal. Instead, dead time can be mitigated by recognizing that the sequence of detection times forms a Markov chain [18]. A good approximation to the dead-time distorted detection time distribution is then given by the stationary distribution of the Markov chain, which can be pre-computed and tabulated for different light intensity values. Additional methods exist for correcting histograms when range gating can also be used [19]. Thus, the distributions for both bright and dark objects can be accommodated and correctly matched, resulting in accurate point clouds.

**Scene Reconstruction**

So far, we have discussed the acquisition process and associated challenges for individual SPL histograms. Here, we consider the problem of processing the complete data cube of $N_i \times N_j$ pixels with $N_k$ histogram bins, indicated in tensor form as $y \in \mathbb{Z}_{>0}^{N_i \times N_j \times N_k}$. We denote the set of depth positions, reflectivity values and background levels as $d$, $r$, and $b$. The 3D reconstruction task consists of recovering a 3D point cloud $(d, r)$ and background levels $b$ from the photon detection events $y$, as illustrated in Figure 3. This task is an ill-posed inverse problem due to the random noise affecting the measurements, the depth uncertainty related to the breadth of the impulse response, and potentially missing data issues, such as the presence hot or dead pixels. The classical maximum likelihood estimate using cross-correlation (4) does not provide robust estimates when very few photons are collected by the detector. To improve these estimates, we can exploit a priori knowledge on the scene’s structure, restricting the reconstructed scene to a set of plausible point clouds. This knowledge is incorporated using regularization functions by
Scene reconstruction
Photon detections
3D point cloud

Fig. 3: Illustration of a single-photon lidar dataset containing two surfaces (the man behind camouflage netting from [20]). The graph on the left shows the histogram of a given pixel. The limited number of collected photons and the high background level makes the reconstruction task extremely challenging. In this case, processing the pixels independently yields poor results, while it can be improved by considering *a priori* knowledge about the scene’s structure.

\[
(\hat{d}, \hat{r}, \hat{b}) = \arg \min_{d,r,b} \log p(y|d, r, b) + \rho_1(d) + \rho_2(r) + \rho_3(b).
\]

where \(p(y|d, r, b)\) is given by the observation model (e.g., (2)), and \(\rho_1, \rho_2, \text{ and } \rho_3\) are interpreted as the regularization terms for the depth, reflectivity, and background respectively. In the context of Bayesian statistics, these terms are interpreted as log-prior distributions. The objective in (5) assumes that depth, reflectivity and background parameters are *a priori* independent, however some models also capture dependencies between them by using a non-separable \(\rho(d, r, b)\).

The 3D reconstruction task is challenging for several reasons. First, the cost function in (5) is generally non-convex and even multi-modal. Second, meaningful and well-defined priors that capture correlations of 2D manifolds (surfaces) are difficult to construct. Third, making an all-encompassing algorithm that can handle varying scenarios such as those shown in Fig. 2 is a difficult task. Finally, the algorithms have to be able to process very large volumes of data while providing fast estimates for any posterior decision making.

The SPL literature contains a wide variety of 3D reconstructions algorithms, differing both in the assumptions about the signal model (1) and the regularization assigned to the unknown
parameters (5). We distinguish two main families of algorithms. The first group of methods assumes exactly one object per pixel \((M = 1)\), reducing the 3D reconstruction problem to the estimation of depth, reflectivity and background images. Most methods in this group use total variation regularization for both depth and reflectivity images [9], [12], [21]. The second group of algorithms, namely multi-depth methods, relax this assumption and attempt to infer a more general 3D point cloud, relying on priors defined in 3D space. In the following, we will discuss different approaches to reconstruction, from robustness to background illumination to real-time scene reconstruction of multiple surfaces per pixel. Fig. 4 shows a comparison between classical cross-correlation estimation and advanced 3D reconstruction algorithms in several challenging scenarios.

Robust Imaging

SPL systems in autonomous vehicles must be able to handle conditions with both weak signals and strong ambient light simultaneously. A fundamental goal of scene reconstruction algorithms is thus to provide accurate depth estimates for low SBR. Initial photon-efficient methods emphasized censoring background detections by comparing detection times at neighboring pixels before applying background-free estimation techniques [9], [12]. More recent algorithms also take advantage of two additional observations. At a pixel level, photons due to ambient illumination are uniformly distributed across the histogram, whereas photons due to a target are clustered following the shape of the impulse response of the device. At an image level, background levels resemble a passive image of the scene, allowing reconstruction algorithms to benefit from additional spatial correlation. The single-depth method introduced in [21] uses the first idea of discarding non-clustered photon detections. Following an iterative multiscale approach, pixels with very low SBR are integrated with neighbouring pixels, increasing the probability of finding a cluster of signal photons (see Fig. 4(a)). Alternatively, an SPL data cube can be fused with a high-resolution 2D photograph via a Convolutional Neural Network to improve the detail of depth and reflectivity reconstructions [22].

Imaging Complex Scenes

In a general setting, an \textit{a priori} unknown but usually reduced number of surfaces can be found at each imaged pixel. Target detection algorithms focus on cases where at most one surface is
Fig. 4: Examples of 3D reconstructions in challenging imaging conditions: The top row shows ML estimates by cross-correlation, whereas the second row shows restorations using advanced methods that also incorporate spatial priors. See [7] and [23] for acquisition details. (a) Robust single-depth algorithms (e.g., [21]) can form accurate point clouds for targets like the mannequin head at 325 m, even when short acquisition times and strong ambient illumination result in few signal photons and an extremely low SBR. (b) Multi-depth algorithms such as ManiPoP [14] can handle complex scenes, such as in the 3-km acquisition depicted here, which has some pixels with zero or multiple surfaces. The estimated degree of impulse response broadening could be used in computing surface normals. (c) A real-time multi-depth algorithm [24] is designed to process 3D data at video rates.

present \((M \leq 1)\), which encompass a wide range of practical scenes. In contrast, multi-depth algorithms relax assumptions on the number of surfaces, estimating the number of objects \(M\) at each pixel.

In the target detection setting, simply thresholding the reflectivity estimates obtained by a single-depth algorithm is generally not robust to background illumination. Hence, specific target detection algorithms estimate an additional binary image indicating the per-pixel presence or absence of a target. In [23], spatial correlations are promoted using an Ising model and inference is performed using a reversible-jump Markov chain Monte Carlo algorithm (RJ-MCMC). While
being fully unsupervised (the hyperparameters are estimated within the Markov chain), this method requires execution times on the order of hours. Recent alternatives achieve faster results by computing a per-pixel posterior probability of target presence and processing the pixels independently in parallel.

The general multi-depth assumption includes the previously discussed algorithms as special cases, at the expense of solving a harder problem. We identify two main strategies. The first approach aims at estimating a 3D volume of reflectivity values, where only a few non-zero values correspond to the 3D points, rewriting (2) for a each pixel as

\[
y \sim P(Hr + 1_{N_k}b),
\]

where \( H \) is the convolution matrix of \( h(t) \), \( 1_{N_k} \) is a unitary vector of \( N_k \) entries, and \( r = [r_1, \ldots, r_{N_k}]^T \) is the reflectivity vector, which has small \( \ell_0 \) pseudo-norm \( ||r||_0 = M \). In this direction, different convex relaxations combine the \( \ell_1 \) and \( \ell_2 \) norms to promote sparsity within histograms and take advantage of correlations across pixels (e.g., [20]). These formulations have the advantage of having a unique solution, which can be obtained using standard proximal-gradient or alternating direction method of multipliers (ADMM) optimization techniques. The downsides are the expensive memory requirements (at least one dense 3D cube has to be stored in memory) and the lack of exactly sparse solutions, generally relying on a post-processing step to further sparsify the output. A second strategy directly estimates a 3D point cloud, where the dimension of the parameter space (i.e., the number of 3D points) is \textit{a priori} unknown. The first step in this direction was the algorithm in [25], which infers the point positions using an RJ-MCMC algorithm to handle the varying dimensions but does not consider spatial correlations between the point positions. The more recent ManiPoP model [26] includes spatial correlations, using a spatial point process prior to model the manifold structure of 3D surfaces. Inference is again performed with an RJ-MCMC algorithm, but the overall computational cost is significantly reduced by the use of carefully tailored proposals. In [14], ManiPoP was extended to an even more general observation model, including the effects of attenuating media and impulse response broadening (see Fig. 4(b)). While being robust to a wide range of conditions, these algorithms still require execution times on the order of seconds or minutes, hindering any real-time decision making in autonomous vehicles. This shortfall is addressed by algorithms tackling real-time performance.
Real-time Imaging

While offline processing of scanned stationary targets has offered compelling evidence of the potential of SPL systems, practical vehicular deployment requires real-time performance. Crucial to speeding up point cloud formation is the parallelization of both the acquisition and processing procedures. Initial real-time lidar development at the MIT Lincoln Laboratory required expensive computation to be run on a large server cluster [27]. More recently, a single-pixel SPL device illuminating short-range scenes with structured (Hadamard) patterns from a digital micromirror device (DMD) was shown to be capable of processing up to 12 frames per second for, enabled by direct inversion of the patterns [28]. As SPAD arrays have improved to capture very high rates of time-stamped photon detections in parallel, the bottleneck for real-time performance has moved to the reconstruction algorithms. New algorithms are designed to take advantage of the parallel computations in general-purpose graphics processing units (GPGPUs), which are compact and inexpensive enough for use in autonomous vehicles. Under a single-depth assumption, the algorithm in [29] avoids the construction of histograms, updating the depth and reflectivity estimates with each new photon arrival. The multi-depth algorithm in [24] uses off-the-shelf point cloud denoisers from the computer graphics community as part of a proximal-gradient optimization algorithm similar to plug-and-play strategies in image processing. The combination of fast parallel denoisers and parallel gradient updates allows acquisition and processing up to 50 frames per second (see Fig. 4(c)). In this direction, future methods will have to incorporate other prior knowledge without slowing down the reconstruction process significantly, relying on an efficient combination of algorithmic structure and computing hardware.

Anticipatory Imaging

In addition to fast and precise formation of point clouds, which is the primary goal of lidar systems, SPL has the potential to extend its functionality beyond the capacity of human vision. A growing research area is non-line-of-sight (NLOS) imaging: forming images of, or at least detecting, objects outside of the direct field of view (see Fig. 5). The principle behind such work is that, even after multiple bounces, intensity and transient measurements of light may retain some information about each interaction between the source and detector. Thus, diffuse surfaces like roads or walls could act as mirrors, allowing vision around obstacles.
There are two main challenges to NLOS imaging: 1) diffuse reflectors scatter light in all directions, scrambling directional information about the light paths, and 2) the light returning to a detector after multiple diffuse bounces is extremely weak. The single-photon sensitivity of SPL systems has made them a popular choice for tackling the second challenge. The first problem continues to be a hot topic of research, with novel computational strategies (e.g., [31], [32]) employed to make the most of the weak signals.

One recent approach uses a standard coaxial SPL configuration to scan a grid of points on a visible wall [30]. While the strong first peaks contain the usual line-of-sight lidar signal, later parts of the temporal histogram contain contributions from light that has undergone multiple bounces before reaching the detector. Clever manipulation of the image formation model reveals that the photon detection time histogram measurements can be expressed as a 3D convolution, enabling a straightforward, closed-form solution that can be implemented with a fast Fourier transform. The resulting algorithm has lower computational and memory requirements than previous approaches using back projection or linear inverse formulations, thus even enabling some real-time demonstrations. The method is especially powerful for use with hidden retro-reflective surfaces, which return much more light than diffuse surfaces do. Since retro-reflective coatings are already in widespread use on roads, including on street signs and within bicycle reflectors, NLOS imaging with SPL may one day be capable of alerting an AV to potential road hazards occluded from view.
FUTURE DIRECTIONS FOR SINGLE-PHOTON LIDAR

Single-photon lidar is already a promising tool for providing fast, high-resolution depth sensing for autonomous vehicles. On-going hardware development of the illumination and detection systems will continue to bring down costs and improve acquisition performance. Nonetheless, signal processing for future SPL systems will still need to address the main questions of how to model the acquisition process, and how to incorporate prior knowledge of 3D structures for point cloud formation. Future extensions to SPL will incorporate more information from alternative lidar acquisitions, e.g., using multiple wavelengths or polarizations, or from other sensing modalities such as standard RGB images or radar. These other sources of information could improve the robustness of the imaging system, generally at the cost of longer or more complex acquisitions and reconstructions. Deep neural networks will likely play an important role in identifying correlated information and fusing the various modalities. The ever-growing demand for better spatial (array size and timing) and spectral resolution can only be met with highly-scalable signal acquisition and processing techniques. Developing compressed sensing techniques for photon-counting lidar, following single-pixel SPL, wavelength-time coding and color-coded apertures, could potentially reduce both the acquisition and reconstruction time. These methods will continue to require a better understanding of the trade-offs between acquisition (photon budget, laser power, acquisition model), reconstruction complexity (memory requirements, parallel or serial architecture, execution time) and estimation performance. Given the success demonstrated thus far and the available opportunities for improvement, SPL will ultimately be one of the factors enabling autonomous driving and the transformation of the transportation industry and beyond.

REFERENCES


