Towards a MEMS-based Adaptive LIDAR

Francesco Pittaluga*
NEC Labs America
francescopittaluga@nec-labs.com

Zaid Tasneem*
University of Florida
ztasneem@ufl.edu

Justin Folden*
University of Florida
jfolden@ufl.edu

Brevin Tilmon
University of Florida
btilmon@.ufl.edu

Ayan Chakrabarti
Washington University in St. Louis
ayan@wustl.edu

Sanjeev J. Koppal
University of Florida
sjkoppal@ece.ufl.edu

Abstract

We present a proof-of-concept LIDAR design that allows adaptive real-time measurements according to dynamically specified measurement patterns. We describe our optical setup and calibration, which enables fast sparse depth measurements using a scanning MEMS (micro-electro-mechanical) mirror. We validate the efficacy of our prototype LIDAR design by testing on over 75 static and dynamic scenes spanning a range of environments. We show CNN-based depth-map completion experiments which demonstrate that our sensor can realize adaptive depth sensing for dynamic scenes.

1. Introduction

Learning-enabled depth sensing has impacted every aspect of robotics. This success has prompted vision researchers to close the loop between active sensing and inference—with methods for correcting incomplete and imperfect depth measurements [44, 48], as well as those that decide where to sense next [26, 5].

However, such work is predicated on LIDAR systems that are flexible in how they make measurements. But this capability does not exist in most existing LIDAR hardware, where sampling is done in a set of fixed angles, usually modulated by slow mechanical motors.

We present a proof-of-concept, adaptive LIDAR platform that can leverage modern vision algorithms. It permits making measurements with different sampling patterns—providing a speed advantage when fewer measurements are made—and is co-located with a color camera to fully realize the benefits of deep depth completion and guided sampling.

---
*Equal Contribution

1.1. Why Foveating LIDAR?

Unlike most artificial sensors, animal eyes *foveate*, or distribute resolution where it is needed. This is computationally efficient, since neuronal resources are concentrated on regions of interest. Similarly, we believe that an adaptive LIDAR would be useful on resource-constrained small platforms, such as micro-UAVs.

Furthermore, our design uses a MEMS mirror as the scanning optics, which is compact and low-power. MEMS scanning is faster than mechanical motors, without similar wear-and-tear, and this allows for multiple fovea or regions-of-interest in a scene. Additionally, MEMS mirrors are neither limited to coherent illumination, like phase arrays, nor constrained to specific light wavelengths, like photonics-based systems.

To demonstrate depth sensing flexibility, we first train a deep neural network for depth completion and show that it delivers high quality scene geometry. We evaluate this with different sampling patterns, including those that are concentrated in a region of interest specified by vision-based control.

Our contributions include building a novel adaptive LIDAR (Fig. 1) that enables flexible deep depth completion (Fig. 4). We also confirm that LIDAR foveation improves sensing in areas of interest (Table 6). We provide analysis of receiver optics characteristics, particularly the issue of small aperture created by the MEMS mirror (Table 2). Finally, we present a working vision-based adaptive LIDAR, showing that real-time foveation is feasible (Fig. 5).

2. Related Work

Common depth modalities: Many high-quality depth sensors exist today. In Table 1 we show qualitative comparisons with these. Our sensor is the first proof-of-concept, real-time, adaptive LIDAR.

1

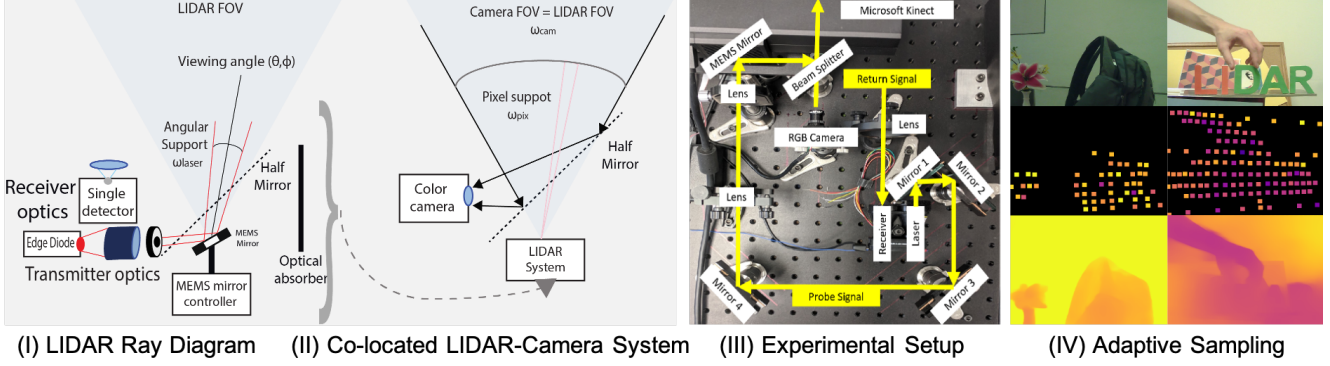| (I) LIDAR Ray Diagram | (II) Co-located LIDAR-Camera System | (III) Experimental Setup | (IV) Adaptive Sampling |

Figure 1: **Experimental setup.** We have designed a flexible MEMS mirror-modulated scanning LIDAR, as shown in (I). In (II), we co-locate this directionally controllable LIDAR with a color camera, allowing for deep depth completion of the sparse LIDAR measurements. In (III) we show a picture of the hardware setup corresponding to (I-II). The long optical path is simply an artifact of having a single circuit board for both the LIDAR receiver and transmitter. In (IV) we show adaptive sampling (middle) and deep depth completion (bottom) results captured with our Adaptive LIDAR Prototype.

Sensor	Technology	Outdoors	Textureless	Adaptive
ELP-960P2CAM	Conventional Passive Stereo	✓	✗	✗
Kinect v2	Time-of-Flight (LED)	✗	✓	✗
Intel RealSense	Structured Light Stereo (LED)	✓	✓	✗
Velodyne HDL-32E	Time-of-Flight (Laser)	✓	✓	✗
Resonance MEMS / Intel L515	Time-of-Flight (Laser)	✓	✓	✗
Robosense RS-LiDAR-M1	Solid State Time-of-Flight (Laser)	✓	✓	✗
Programmable Light curtains	Adaptive Structured Light	✓	✓	✓
Our sensor	Adaptive LIDAR	✓	✓	✓

Table 1: **Our Adaptive LIDAR vs. other common modalities.** We compare common depth modalities such as stereo [28], Kinect [35], Velodyne [15], Robosense solid state LIDAR and Resonance MEMS sensors [13, 41, 24] such as the Intel L515. Our work is closest to programmable light curtains for flexible, structured light reconstruction [3, 46]. This paper is an alternate research direction with an adaptive LIDAR.

MEMS/Galvo mirrors for vision and graphics: MEMS mirror modulation has been used for structured light [37], displays [21] and sensing [34]. In contrast to these methods, we propose to use angular control to increase sampling in regions of interest as in [43]. While MEMS mirrors have been used in scanning LIDARs, such as from NASA and ARL [13, 41, 24], *these are run at resonance with no control, while we show adaptive MEMS-based sensing.* Such MEMS control has only been shown [22] in toy examples for highly reflective fiducials in both fast 3D tracking and VR applications [31, 30], whereas we show results on real scenes. [39, 8] show a mirror modulated 3D sensor with the potential for flexibility, but without leveraging guided networks, and we discuss the advantages of our novel receiver optics compared to these types of methods. Galvo mirrors are used with active illumina-tion for light-transport [17] and seeing around corners [36]. Our closest work is the use of light curtains for flexible, structured light reconstruction [3, 46]. In contrast, ours is a MEMS-mirror driven LIDAR system with an additional capability of increasing resolution in some region of interest. In this sense, we are the first to extend adaptive control [6, 42, 8], to LIDAR imaging of real dynamic scenes.

Adaptive Scanning Lidars: Commercially available systems from AEye and Robosense are designed to improve lidar-rgb fusion for large systems such as autonomous cars. In contrast, our goal is to impact small autonomous systems and our choice of MEMS mirror modulation and our optical innovations track these goals. [47] propose a progressive pedestrian scanning method using an actively scanned LIDAR, but results are shown in simulation rather than on a hardware

platform. [42] propose directionally controlling the LI-DAR scan, but these adaptive results have been shown only for static scenes. In contrast, we show a real-time adaptive LIDAR that works for dynamic scenes.

Guided and Unguided Depth Completion: The impact of deep networks on upsampling and super-resolution has been shown on images, disparity/depth maps, active sensor data etc. [4, 11, 29, 27, 44, 38, 20] with a benchmark on the KITTI depth completion dataset [44]. Upgrading from sparse depth samples has been shown [45], and guided upsampling has been used as a proxy for sensor fusion such as the work that has recently been done for single-photon imagers [25] and flash lidar [14]. In contrast, we measure sparse low-power LIDAR depth measurements and we seek to flexibly change the sensor capture characteristics in order to leverage adaptive neural networks such as [26, 5].

3. Sensor design

Fig. 1 shows our sensor design, which consists of a small aperture, large depth-of-field color camera, optically co-located with a MEMS-modulated LIDAR sensor. If the camera has a FOV of ω_{cam} steradians and a resolution of I pixels, then the average pixel support is $\omega_{pix} = (\omega_{cam}/I)$. If the LIDAR laser's beam divergence is ω_{laser} steradians, then the potential acuity increase from LIDAR to camera is $(\omega_{laser}/\omega_{pix})$. Guided depth completion seeks to extract this potential, and a foveated LIDAR can leverage this capability.

Next, in Sect. 3.1, we discuss the MEMS-modulated transmitter optics that enable compact, low-power, fast and foveated controlled scans. The cost, however, is that MEMS mirrors act as a small aperture that reduces the received radiance, when compared to large mirrors such as galvos. In the following Sect. 3.2 we model the receiver optical design space, comparing FOV, volume and received radiance.

3.1. MEMS Mirror based Transmitter Optics

The transmitter optics consist of the pulsed light source and MEMS mirror. The LIDAR's beam is steered by the mirror, whose azimuth and elevation are given by changes in control voltages over time, $(\theta(V(t)), \phi(V(t)))$ over the MEMS mirror FOV ω_{mirror}. Controlling the scan $V(t)$ can enable attending to a region-of-interest given by an adaptive algorithm. The challenge in transmitter optics is to provide a powerful, narrow laser with low beam divergence, given by

$$\omega_{laser} \approx \frac{M^2 \ \lambda}{w_o \ \pi} \tag{1}$$

where M is a measure of laser beam quality and w_o refers to the radius at the beam waist, which we use
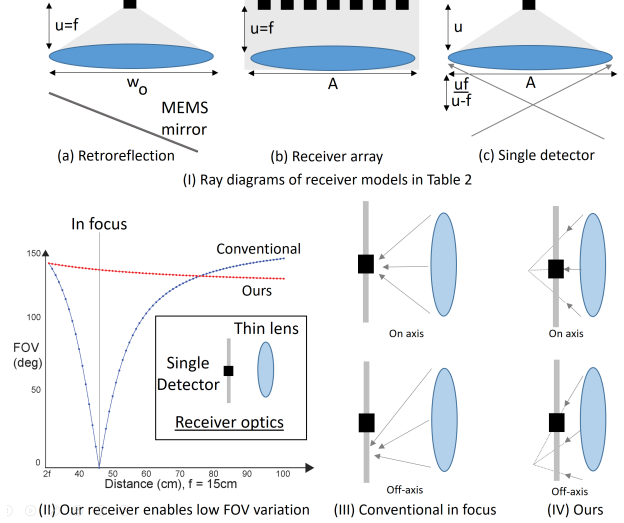

Figure 2: **Our proposed design vs. other designs.** In (I) we depict three common receiver designs, including retro-reflection (a), receiver array (b) and single detectors (c). Our design is a variant of (c), where we suggest a simple optical trick, such that the single detector is placed within the focal distance of the lens. This enables consistent FOV over range, as shown by the red curve in (II) and the designs in (III-IV). Simulations for a $f = 15mm$ unit diameter lens.

a proxy for MEMS mirror size. Previous MEMS-mirror modulated LIDAR systems cover this design space, from high-quality erbium fiber lasers with near-Gaussian profiles, used by [41] where M is almost unity, to low-cost edge-emitting diodes, such as [42] where $M \approx 300$ on the diode's major axis. Our setup follows the low-cost diode route, with an additional two-lens Keplerian telescope to reduce the beam waist to $6mm$ and an iris to match the MEMS mirror aperture instead of an optical fiber [8].

3.2. Receiver Optics Design Tradeoffs

From the previous section, we can denote the transmitter optics design space as a combination of laser quality M and MEMS mirror size w_o, which we write as $\Pi_t = \{M, w_o\}$. Now, we add receiver optics to the design space, which we denote as $\Pi_r = \{n, A, u, f\}$, with n^2 photodetectors in the receiver, A as the aperture, u as the distance between the photodetector array and the receiver optics, and f as the focal length of the receiver optics. Therefore the full design space consists of both receiver/transmitter optics, $\mathbf{\Pi} = \{\Pi_r, \Pi_t\}$.

We define the characterization of any instance within the design space $\mathbf{\Pi}$ as consisting of field-of-view Ω steradians, received radiance s and volume

Design	Volume	FOV	Received Radiance
Retroreflection	$\frac{\pi\ u\ w_o^2}{12}$	= MEMS FOV ω_{mirror}	$\frac{atan(\frac{w_o}{2Z})}{\omega_{\mathrm{laser}}Ztan(\frac{\omega_{\mathrm{laser}}}{2})}$
Receiver array	$u\ A^2$	$\min(2\ atan(\frac{A}{2u}),\ \omega_{mirror})$	$\frac{1}{2\ Z\ tan(\frac{\omega_{\mathrm{laser}}}{2})}$
Single detector . Conventional $(u \geq f)$ **Ours** $(u < f)$	$\frac{\pi\ u\ A^2}{12}$	$\min(2atan(\frac{A(Z-f)\lVert\frac{Zu-fu-fZ}{Z-f}\rVert}{2ufZ}),\ \omega_{mirror})$	$\frac{1}{4Zatan(\frac{A(Z-f)\lVert\frac{Zu-fu-fZ}{Z-f}\rVert}{2ufZ})tan(\frac{\omega_{\mathrm{laser}}}{2})}$

Table 2: **Receiver models.** Please see supplementary for derivations.

V denoted as $\Xi = \{\Omega, s, V\}$. The range Z is determined by the received radiance and the detector sensitivity. Computing these parameters depends on the design choices made, and we provide simulations comparing three designs: retro-reflective receivers [16] (Fig. 2I(a)), receiver arrays [8] (Fig. 2I(b)) and single-pixel detectors [42] (Fig. 2I(c)).

3.3. Simulation Setup and Conclusions

Full derivations for the three receiver designs, shown in Table 2, are in the supplementary material. The table refers to receiver sensor volume, field-of-view, and received radiance (normalized for a white Lambertian plane). The *volume* is the convex hull of the opaque baffles that must contain the receiving transducer electronics and is either a cone or cuboid. The *FOV* is the range of angles that the receiver is sensitive to, upper-bounded by the MEMS FOV ω_{mirror} [1]. *Received radiance* is the area-solid angle product used in optics [32] for a canonical LIDAR transducer, which can be loosely understood as loss of LIDAR laser dot intensity due to beam divergence and receiver aperture size when imaging a fronto-parallel, white Lambertian plane.

In our noiseless simulations, we assumed a geometric model of light. To illustrate the trade-offs, we vary the laser quality between $M = 1$ to $M = 100$, representing an ideal Gaussian beam vs. a cheap laser diode. For the same reason, we vary the MEMS mirror size w_o from $0.1mm$ (10 times larger than the TI DMD [18]) to $5mm$ (a large size for a swiveling MEMS mirror). The range of dimensions over which we explore the receiver design space are of the order of a small camera, with apertures $0cm \leq A \leq 10cm$, focal lengths $0mm \leq f \leq 50mm$ and image plane-lens distances $0mm \leq u \leq 50mm$. In Fig. 2(II)-(IV) we describe our proposed, simple modification to the conventional single-pixel receiver, where photodetector is placed on the optical axis, at a distance v larger than the focal length f. In contrast to existing work on defocusing received radiances for FOV adjustment and amplitude compensation (e.g. [32]),

we do not require special optics (e.g. split lens) and we have large off-axis FOV since the MEMS is not the aperture for the receiver. The **conclusion** from these experiments is that our sensor provides a new option for receiver design tradeoffs, explained next.

3.4. Design tradeoffs

For a full analysis of sensor design tradeoffs, please see the supplementary material. Here we compare our sensor vs. other designs. In the next section, we show how our proposed optical modification can improve single detectors even further.

Retro-reflective receivers: We show, in the supplementary, that retro-reflective designs are smaller than ours. The small retroreflective design also has the optimal FOV of the MEMS, due to co-location. Our design does have a received radiance advantage, since retroreflection requires the MEMS mirror to be the aperture for both receiver/transmitter. Fig. 3(Ia) shows how this advantage eventually trumps other factors such as laser quality ($M = 1$) or large mirrors. In the extreme case of low-cost diodes, Fig. 3(Ib), our sensor has higher received radiance at close ranges too.

Receivers arrays: Large receiver arrays take up space, but they also have a higher received radiance, due to having a bigger effective aperture, when compared with our MEMS mirror. This is demonstrated in Fig. 3(II) (right) for the particular case of $M = 100, w_o = 5mm$, favoring our design. Despite this, large arrays have higher received radiance at all depths. However, our design can always be made smaller, in every instance of these simulations.

Conventional Single detector: Our approach is close to the conventional single pixel receiver, which can allow for detection over a non-degenerate FOV if it is defocused, as shown for a scanning LIDAR by [42]. When the laser dot is out of focus, some part of it activates the single photodetector. If the laser dot is in focus, the activation area available is smaller, but more concentrated. Next we describe and analyze our modification to the conventional single detector.

---

[1]For simplicity, trignometric functions are written to act on steradian quantities, but in actuality act on the apex angle of the equivalent cone.
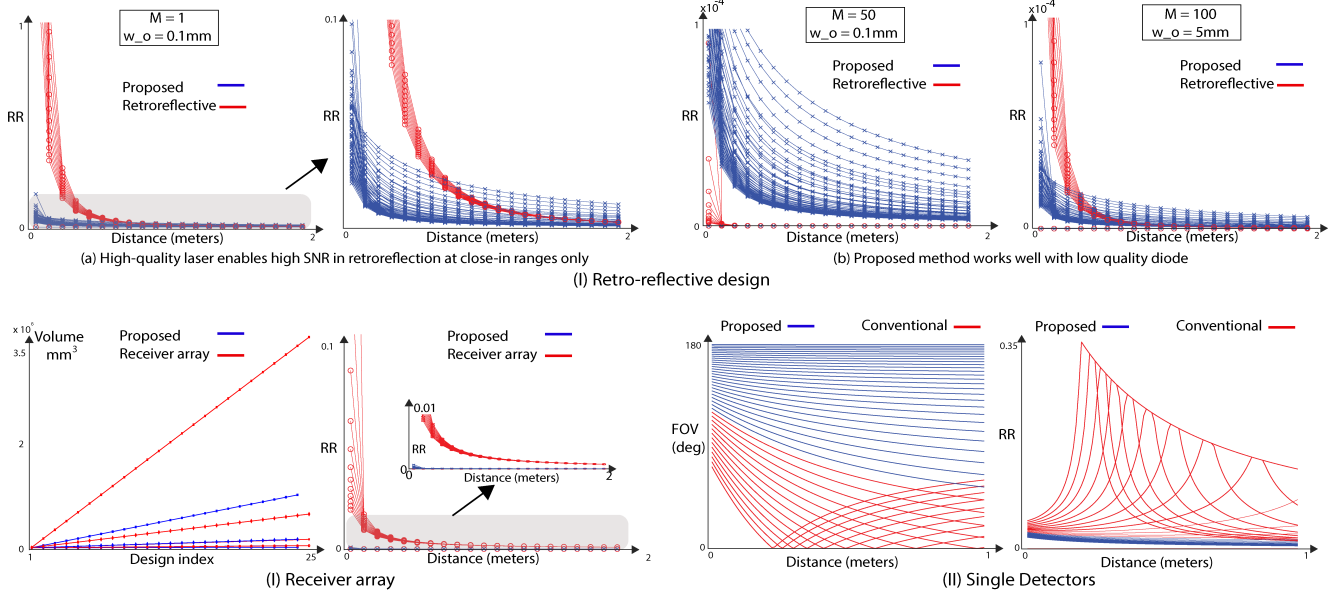
Figure 3: **Noiseless simulations comparing proposed method with other designs.** In (I) we compare the received radiance (RR) of proposed method with retroreflection for different laser qualities and mirror sizes. A high-quality laser (I)(a) enables higher RR for close-in scenes for retroreflective designs, but at large ranges, our method has higher RR. In (II)(a) we show that our proposed design has lower volume than a receiver array, across a wide range of focal lengths, but a receiver array has a higher RR (II)(b), even when compared to the best case for our sensor from (I). In (III) we compare our design with conventional single detectors, for a lens with $f = 15mm$. Although our sensor shows consistent FOV ((III) left), it is always defocused, and faces an RR cost ((III) right).

3.5. Optical modification to single detectors

Our approach is based on a simple observation; placing the image plane between the lens and the focus, i.e. $v < f$, will guarantee that the laser dot will never be in focus. For imaging photographs, this is not desirable, but for detecting the LIDAR system's received pulse, amplitude can be traded down, up to a point, as long as the peak pulse can be detected. Further, this optical setup ensures that the angular extent of the dot is nearly constant over a large set of ranges. This is further explained in the supplementary and supported by simulations (red curve) in Fig. 3(III) (left) and explained in the ray diagrams of Fig. 3(III) (right).

For the conventional approach, when $u = f$, the FOV degenerates to a small value, where received radiance is also the highest. Our design does not suffer this depth-dependent FOV variation and is consistent across the range. As shown in the right, however, this results in a low received radiance since the system is always defocused. Simulations support this, in Fig. 3(III), for $u > f$, shown in red, for settings of $f = 15mm, A = 100mm$, over a range of sensor sizes and ranges. In practice, we find consistent FOV to be more valuable than received radiance.

Data	MRE (%)	RMSE (m)	log$_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
Real	10.16	.1659	.0410	89.80	95.88	98.63

Table 3: **LIDAR Evaluation.** The table reports the mean relative error (MRE), root mean squared error (RMSE), average (log$_{10}$) error, and threshold accuracy (δ_i) of the calibrated depth measurements, relative to the "ground-truth" Kinect V2 depths, over all 75 scenes of our real dataset. The Kinect V2 has an accuracy of 0.5% of the measured range [2].

4. Towards Adaptive LIDAR

MEMS mirrors [31, 30] are fast enough to enable adaptive sensing for dynamic scenes. The question then becomes *how* to find good scan patterns, represented as voltage-dependent mirror angles over time, $(\theta(V(t)), \phi(V(t)))$. These include open loop [6, 12, 42, 8] real-time estimation of regions of interest (ROIs) as well as end-to-end learning to decide where to sense next [26, 5]. **Our contribution** here is to demonstrate LIDAR foveation for dynamic scenes with an open-loop algorithm based on motion detection [12].
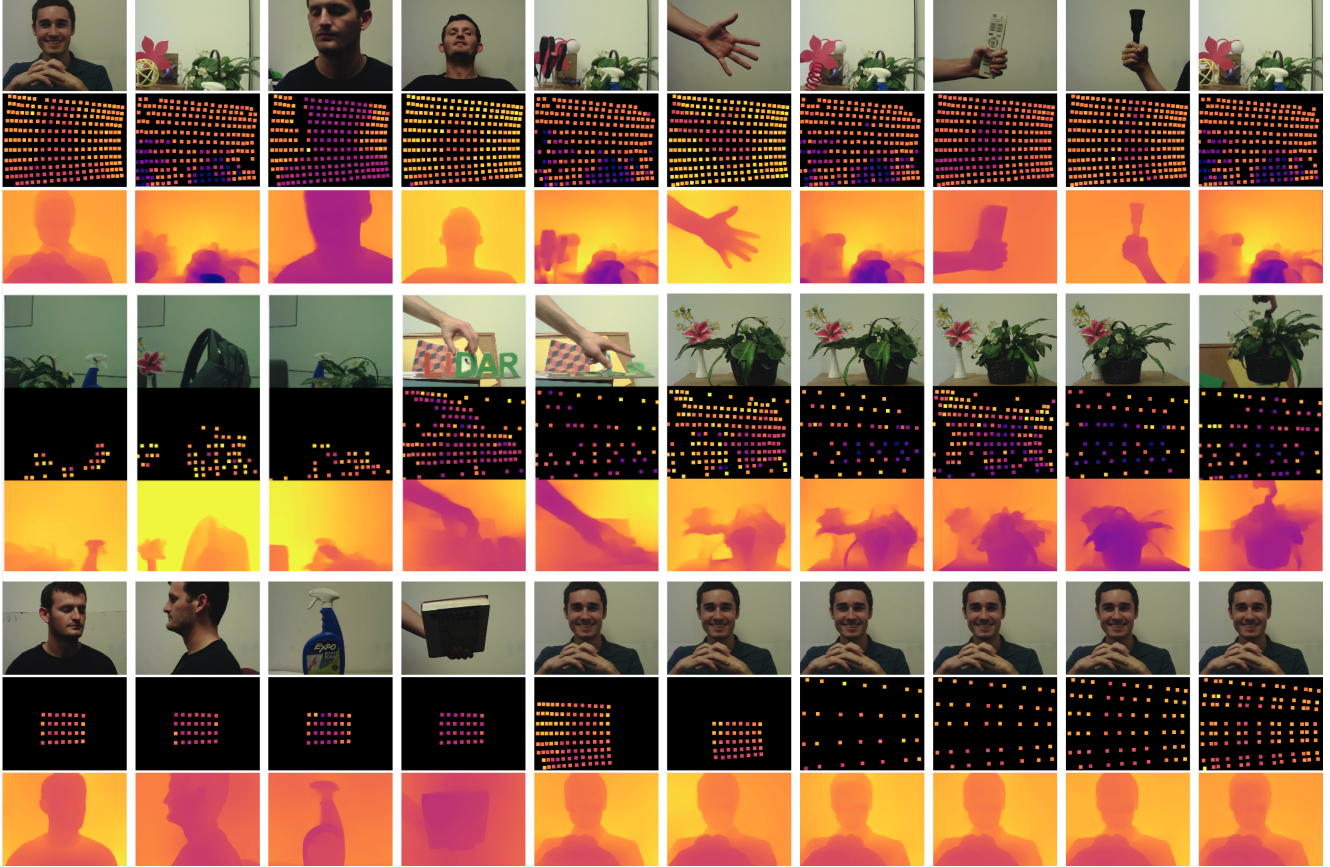
Figure 4: **Adaptive Lidar Sampling.** This figure qualitatively demonstrates the flexibility of our adaptive LIDAR by showing a range of scan patterns. In row 1, a fixed, equi-angular full FOV scan pattern was used. In row 2, the density of the scan pattern was automatically adapted according to the RGB image's entropy. In row 3, columns 1-6, constant sampling density was applied on a rectangular ROI with maximal scene entropy. In row 3, columns 7-10, the FOV of the scan pattern was kept fixed and a sweep of the sampling density was performed. Note, with no depth samples, our depth completion model defaults to monocular depth estimation from the colocated camera, since we randomly sparsified the input depth maps during training to encourage robustness to a range of sampling densities (including zero samples).

Experimental setup: Our LIDAR engine is a single beam Lightware SF30/C with an average power of $0.6mW$. This device can produce 1600 depth measurements per second at 100m. Data is captured as a stream of measurements, and each are time-stamped by the MEMS direction, given by the voltage $V(t)$. We modulate the single beam with a $3.6mm$ Mirrorcle MEMS mirror. Our current prototype has a range is $3m$ and a field-of-view of $\approx 25°$. The laser dot, in steradians, is $6 \times 10^{-4}\Omega$ and this angular support is consistent over change in MEMS mirror angle.

Calibration and validation: Since our sensor response is linear, we apply a 1D calibration to convert the LIDAR voltages into distances. We evaluate the quality of our sensor measurements and our calibra-

tion by computing the mean relative error (MRE), root mean squared error (RMSE), average ($\log_{10}$) error, and threshold accuracy (δ_i) of the calibrated depth measurements from our LIDAR. We do this relative to the "ground-truth" Kinect V2 depths, over all 75 scenes of our real dataset, and these are reported in Table 3. The Kinect V2 has an accuracy of 0.5% of the measured range [2]. Finally, we also captured 10 fronto-planar scenes (at ranges .5m-3m) and computed the RMSE of the depth measurements along the plane using the SVD method: the resulting average RSME was 0.06918m.

4.1. Depth completion

We now describe depth completion for the foveated measurements of our LIDAR. This builds on existing

Data	Method	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	Mono	8.55	.3800	.0361	90.56	98.08	98.56
	Ours	**5.89**	**.2488**	**.0245**	**97.69**	**99.68**	**99.92**
Real	Mono	28.26	.3711	.1090	50.14	87.38	96.00
	Ours	**12.29**	**.1668**	**.0395**	**85.86**	**95.89**	**99.18**

Table 4: **Base Comparison to Monocular Depth Estimation.** As a baseline, we compare to state-of-the-art monocular depth estimation [1] (Mono) to our depth completion method (Ours) on a sub-sampled version of the NYUv2 Depth [33] (NYU) dataset and on our real dataset (Real). Both the monocular depth estimation and depth completion methods were trained only on NYUv2 data. To account for this, monocular depth estimates were scaled by the ground-truth median, as in [1]. Such scaling was not performed for depth completion predictions since the sparse LIDAR samples provide a reference absolute depth.

work [44, 48] where the sparse depth measurements are captured by our flexible LIDAR sensor and the "guide" image is captured by a RGB camera that is co-located with the sensor. We train a DenseNet-inspired [19] encoder-decoder network to perform RGB-guided depth completion of sparse measurements.

Architecture. We adopt [1]'s encoder-decoder network architecture, except that our network has 4 input channels, as it expects a sparse depth map concatenated with an RGB image. The encoder component of our network is the same as DenseNet 169 minus the classification layer. The decoder component consists of a three convolutional blocks followed by a final 3×3 convolutional layer. Each bilinear upsampling block consists of two 3×3 convolutional layers (with a leaky ReLU), and 2×2 max-pooling.

Optimization. We adopt [1]'s loss as a weighted sum of three terms:

$$L(y, \hat{y}) = \lambda L_{depth}(y, \hat{y}) + L_{grad}(y, \hat{y}) + L_{SSIM}(y, \hat{y}) \quad (2)$$

where y and $\hat{y}$ denote the ground-truth and estimated depth maps respectively, and λ denotes a weighting parameter, which we set to 0.1. The remaining terms are defined as in [1] which has the full expressions.

Datasets and Implementation: We perform our evaluations using two datasets: a real dataset captured with our LIDAR system and a simulated Flexible LIDAR dataset generated by sub-sampling the NYUv2 Depth dataset [33]. The real dataset consists of pairs of RGB images and sparse depth measurements of 75 different scenes captured with our LIDAR system. For each of the 75 scenes, we also capture a dense "ground-

Data	FPS	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	30	5.89	.2488	.0245	97.69	99.68	99.92
	24	5.88	.2430	.0244	97.97	99.70	99.92
	18	5.59	.2261	.0233	98.50	99.77	99.94
	12	5.65	.2255	.0236	98.52	99.77	99.94
	6	5.15	.1879	.0217	99.32	99.91	99.98
Real	30	12.29	.1668	.0395	85.86	95.89	99.18
	24	12.09	.1644	.0446	86.34	96.04	99.26
	18	11.57	.1578	.0430	87.27	96.61	99.30
	12	11.59	.1558	.0435	88.26	97.01	99.33
	6	11.19	.1537	.0422	88.10	97.19	99.26

Table 5: **Evaluation of Depth Completion.** This table conveys three key features of our system: (1) It highlights, the trade-off between frame rate and depth uncertainty, which impacts real-time applications; (2) it provides a quantitative evaluation of the robustness of our depth completion algorithm to varying sampling densities; and (3) provides an illustrative example of our system flexibility, which can be leveraged for a range of applications. For frame rates of 30, 24, 18, 12 and 6, the samples per frame were 28, 40, 60, 104 and 231 respectively.

Data	Method	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	Full FOV	5.52	.2392	.0231	98.24	99.86	99.98
	Foveated	**4.81**	**.1845**	**.0202**	**99.50**	**99.97**	**100**
Real	Full FOV	15.72	0.1925	.0566	80.30	99.79	99.36
	Foveated	**13.36**	**.1589**	**.0497**	**83.24**	**97.80**	**99.46**

Table 6: **Depth Completion on Foveated Lidar Data.** "Foveated" means that the scan pattern was automatically adapted to densely sample a region of interest in the scene. "Full FOV" means that a scene independent equi-angular scanning pattern was utilized. In all cases, the "Foveated" and "Full FOV" scan patterns contain the same number of samples (hence, the equivalent frame rates). Results are evaluated at 30 FPS. Both Full FOV and Foveated errors are computed only in identical regions of interest, showing foveation increases accuracy.

truth" depth map using a Kinect V2 depth sensor that is stereo calibrated with our LIDAR system. All real dataset images are used exclusively for testing. The simulated dataset is split into non-overlapping train, test, and validation scenes.

We train the model described in section 4.1 on a simulated Flexible LIDAR dataset generated by sub-sampling the NYUv2 Depth dataset. During training,
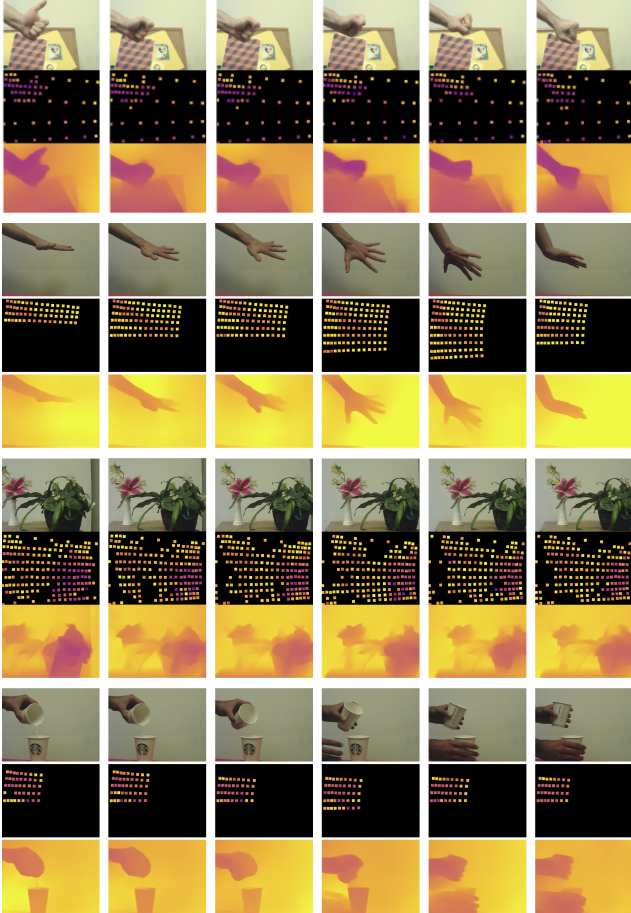
Figure 5: **Motion-based adaptive sensing.** As the object moves, we use background subtraction to detect the region of interest and the MEMS-modulated LIDAR puts the samples where the object is located. Please see supplementary video.

depths were randomly scaled to prevent the network from overfitting to the color camera used to capture the RGB images in the NYUv2 dataset. For optimization, we used Adam [23] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e - 8$, a learning rate of 0.0001, and a batch size of 4. The learning rate was dropped to 0.00001 after $94k$ iterations. The first layer used Xavier initialization, whereas other layers were initialized with the pre-trained weights from [1] for monocular depth estimation on NYUv2. To augment the data, we first randomly resize the input such that the smallest dimension varies between $640, 832$, or 1024. We then apply a random crop to the reduce the size to 640×480. In addition, the RGB channels were randomly shuffled.

As a 'sanity check', we confirm that guided depth completion outperforms monocular depth estimation using a state-of-the-art network [1] in Table 4.

4.2. Motion-based Foveated Depth Sampling

Our flexible platform allows us to ask if foveated LIDAR sampling improves depth measurements. We evaluate our guided depth completion network on LIDAR data captured with two different sampling regimes, full field-of-view sampling and foveated sampling in regions of interest, at various frame rates.

Table 5 shows our evaluation for full field-of-view sampling. Table 6 demonstrates that foveation improves reconstruction in a region of interest, with qualitative results in Figures 4 and 5.

We also perform foveated sampling in real-time, using an open-loop motion-based system to determine the scan patterns. For a dynamic scene, a foveating LIDAR can have fewer samples in the right places, decreasing latency and improving frame-rate. In Fig. 5, we show objects moving across the scene. At each instance, the system performs background subtraction to segment a motion mask. This mask drives the LIDAR sampling, which has less points than a full dense scan would have, and therefore has higher sampling rate. In each result, ROI sampling density was identically dense, and the rest-of-the-scene density was different and sparser. The amortized frame rates for the real-time foveated sequence in rows 1, 2, 3 and 4 of Fig. 5 are 20 FPS, 13 FPS, 9 FPS and 24 FPS. Without foveation, dense sampling over the entire scene would result in a frame rate of 6FPS, which is much lower. Note that as the object changes position, the ROI changes and the LIDAR senses a different area. If temporal sampling is not the focus, then the method can instead densely sample the points onto the region of interest, increasing the angular resolution (i.e. zooming). Finally, we note that all results include depth completion of the measurements, showing high-quality results.

5. Discussion

Limitations: Our LIDAR engine has a $3m$ range, which enables initial feasibility tests and is appropriate for certain tasks such as gesture recognition. Range extension is achievable, since the Lightware LIDAR electronics engine has a 100m outdoor range and is only reduced by unnecessary optical losses. For future prototypes we wish to remove these losses with a GRIN lens, as done by [13].

Conclusions: Our LIDAR prototype enables the kind of adaptive sensing, which, thus far, has only been shown in simulation and follows a recent trend in computational photography to use data-driven approaches inside the sensor [7, 10, 9, 40].

References

[1] I. Alhashim and P. Wonka. High quality monocular depth estimation via transfer learning. *arXiv preprint arXiv:1812.11941*, 2018. 7, 8

[2] C. S. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V. Chan. A 0.13 um cmos system-on-chip for a 512 × 424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 mhz and 2 gs/s adc. *IEEE Journal of Solid-State Circuits*, 50(1):303–319, 2015. 5, 6

[3] J. R. Bartels, J. Wang, W. Whittaker, and S. G. Narasimhan. Agile depth sensing using triangulation light curtains. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7900–7908, 2019. 2

[4] R. Battrawy, R. Schuster, O. Wasenmüller, Q. Rao, and D. Stricker. Lidar-flow: Dense scene flow estimation from sparse lidar and stereo images. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7762–7769. IEEE, 2019. 3

[5] A. Bergman, D. Lindell, and G. Wetzstein. Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates. *ICCP*, 2020. 1, 3, 5

[6] G. M. Breivik, J. T. Thielemann, A. Berge, Ø. Skotheim, and T. Kirkhus. A motion based real-time foveation control loop for rapid and relevant 3d laser scanning. In *CVPR 2011 WORKSHOPS*, pages 28–35. IEEE, 2011. 2, 5

[7] A. Chakrabarti. Learning sensor multiplexing design through back-propagation. In *Advances in Neural Information Processing Systems*, pages 3081–3089, 2016. 8

[8] S. Chan, A. Halimi, F. Zhu, I. Gyongy, R. K. Henderson, R. Bowman, S. McLaughlin, G. S. Buller, and J. Leach. Long-range depth imaging using a single-photon detector array and non-local data fusion. *Scientific reports*, 9(1):8075, 2019. 2, 3, 4, 5,

[9] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific reports*, 8(1):12324, 2018. 8

[10] H. G. Chen, S. Jayasuriya, J. Yang, J. Stephen, S. Sivaramakrishnan, A. Veeraraghavan, and A. Molnar. Asp vision: Optically computing the first layer of convolutional neural networks using angle sensitive pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 903–912, 2016. 8

[11] Z. Chen, V. Badrinarayanan, G. Drozdov, and A. Rabinovich. Estimating depth from rgb and sparse sensing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 167–182, 2018. 3

[12] S. Ferrari. Track coverage in sensor networks. In *2006 American Control Conference*, pages 7–pp. IEEE, 2006. 5

[13] T. P. Flatley. Spacecube: A family of reconfigurable hybrid on-board science data processors. 2015. 2, 8

[14] T. Gruber, F. Julca-Aguilar, M. Bijelic, W. Ritter, K. Dietmayer, and F. Heide. Gated2depth: Real-time dense lidar from gated images. *arXiv preprint arXiv:1902.04997*, 2019. 3

[15] D. S. Hall. High definition lidar system, June 28 2011. US Patent 7,969,558. 2

[16] R. Halterman and M. Bruch. Velodyne hdl-64e lidar for unmanned surface vehicle obstacle detection. Technical report, SPACE AND NAVAL WARFARE SYSTEMS CENTER SAN DIEGO CA, 2010. 4

[17] T. Hawkins, P. Einarsson, and P. E. Debevec. A dual light stage. *Rendering Techniques*, 5:91–98, 2005. 2

[18] L. J. Hornbeck. Architecture and process for integrating dmd with control circuit substrates, May 28 1991. US Patent 5,018,256. 4

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 7

[20] T.-W. Hui, C. C. Loy, and X. Tang. Depth map super-resolution by deep multi-scale guidance. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2016. 3

[21] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an interactive 360 light field display. *ACM Transactions on Graphics (TOG)*, 26(3):40, 2007. 2

[22] A. Kasturi, V. Milanovic, B. H. Atwood, and J. Yang. Uav-borne lidar with mems mirror-based scanning capability. In *Proc. SPIE*, volume 9832, page 98320M, 2016. 2

[23] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 8

[24] K. T. Krastev, H. W. Van Lierop, H. M. Soemers, R. H. M. Sanders, and A. J. M. Nellissen. Mems scanning micromirror, Sept. 3 2013. US Patent 8,526,089. 2

[25] D. B. Lindell, M. O'Toole, and G. Wetzstein. Single-photon 3d imaging with deep sensor fusion. *ACM Transactions on Graphics (TOG)*, 37(4):113, 2018. 3

[26] C. Liu, J. Gu, K. Kim, S. Narasimhan, and J. Kautz. Neural rgb-to-d sensing: Depth and uncertainty from a video camera. *arXiv preprint arXiv:1901.02571*, 2019. 1, 3, 5

[27] J. Lu and D. Forsyth. Sparse depth super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2245–2253, 2015. 3

[28] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981. 2

[29] F. Mal and S. Karaman. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2018. 3

[30] V. Milanović, A. Kasturi, N. Siu, M. Radojičić, and Y. Su. "memseye" for optical 3d tracking and imaging applications. In *Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International*, pages 1895–1898. IEEE, 2011. 2, 5

[31] V. Milanović, A. Kasturi, J. Yang, and F. Hu. A fast single-pixel laser imager for vr/ar headset tracking. In *Proc. of SPIE Vol*, volume 10116, pages 101160E–1, 2017. 2, 5

[32] J. Mudge. Range-compensating lens for non-imaging active optical systems. *Applied optics*, 58(28):7921–7928, 2019. 4

[33] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 7

[34] S. K. Nayar, V. Branzoi, and T. E. Boult. Programmable imaging using a digital micromirror array. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004. 2

[35] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136. IEEE, 2011. 2

[36] M. O´ Toole, D. B. Lindell, and G. Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018. 2

[37] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998. 2

[38] G. Riegler, M. Rüther, and H. Bischof. Atgv-net: Accurate depth super-resolution. In *European Conference on Computer Vision*, pages 268–284. Springer, 2016. 3

[39] T. Sandner, C. Baulig, T. Grasshoff, M. Wildenhain, M. Schwarzenberg, H.-G. Dahlmann, and S. Schwarzer. Hybrid assembled micro scanner array with large aperture and their system integration for a 3d tof laser camera. In *MOEMS and Miniaturized Systems XIV*, volume 9375, page 937505. International Society for Optics and Photonics, 2015. 2

[40] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):114, 2018. 8

[41] B. L. Stann, J. F. Dammann, M. Del Giorno, C. DiBerardino, M. M. Giza, M. A. Powers, and N. Uzunovic. Integration and demonstration of mems–scanned ladar for robotic navigation. In *Proc. SPIE*, volume 9084, page 90840J, 2014. 2, 3,

[42] Z. Tasneem, D. Wang, H. Xie, and S. J. Koppal. Directionally controlled time-of-flight ranging for mobile sensing platforms. In *Robotics: Science and Systems*, 2018. 2, 3, 4, 5,

[43] B. Tilmon, E. Jain, S. Ferrari, and S. Koppal. Foveacam: A mems mirror-enabled foveating camera. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2020. 2

[44] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger. Sparsity invariant cnns. In *2017 International Conference on 3D Vision (3DV)*, pages 11–20. IEEE, 2017. 1, 3, 7

[45] W. Van Gansbeke, D. Neven, B. De Brabandere, and L. Van Gool. Sparse and noisy lidar completion with rgb guidance and uncertainty. *arXiv preprint arXiv:1902.05356*, 2019. 3

[46] J. Wang, J. Bartels, W. Whittaker, A. C. Sankaranarayanan, and S. G. Narasimhan. Programmable triangulation light curtains. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 19–34, 2018. 2

[47] T. Yamamoto, Y. Kawanishi, I. Ide, H. Murase, F. Shinmura, and D. Deguchi. Efficient pedestrian scanning by active scan lidar. In *Advanced Image Technology (IWAIT), 2018 International Workshop on*, pages 1–4. IEEE, 2018. 2

[48] Y. Zhang and T. Funkhouser. Deep depth completion of a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 175–185, 2018. 1, 7