Exploiting and Refining Probabilistic Depth Estimates with RGB and Triangulating Light Curtain Fusion

Anonymous CVPR 2021 submission

Paper ID ****
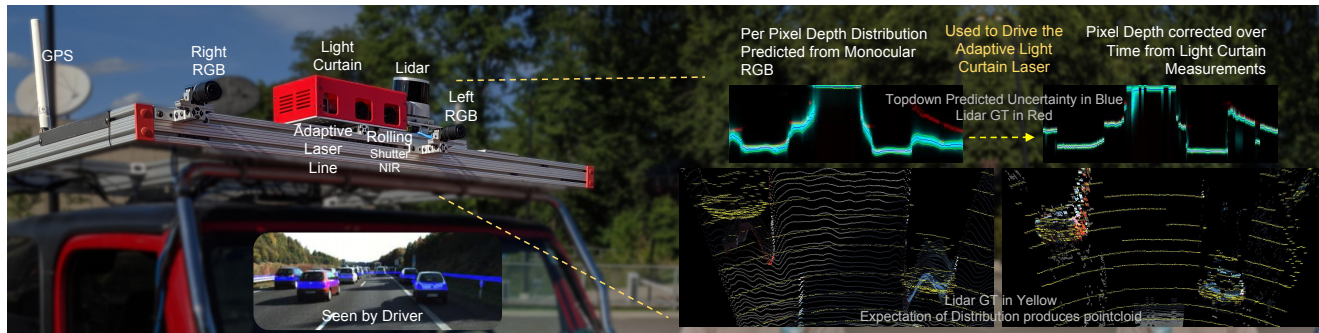
Figure 1: We present a novel method that predicts per pixel depth uncertainity from Monocular/Stereo RGB cameras, uses that information to adaptively place light curtains at regions of low uncertainty, and refines depth estimates over time

Abstract

Programmable Triangulating Light Curtains are a low-cost, high speed and high resolution approach to depth sensing, but require the user to specify a 2.5D ruled surface over where sensing should occur. Objects that intersect this surface show up with a high intensity, while those that don't do not. We have devised a novel algorithm, that exploits prior depth distributions from either Monocular or Stereo RGB cameras to sample depth in regions of the world most uncertain using the Light Curtain. We then incorporate this data to get a refined depth estimate over time. We evaluate this setup with real world real-time experiments such as in the context of Advanced driver-assistance systems (ADAS)

1. Introduction

Programmable Triangulating Light Curtains were introduced last year, and have demonstrated viability in being a low cost (1k$ vs lidar's 25k$), high spatial angular resolution ($0.02°$ vs lidar's $0.4°$), and high frame-rate (60fps vs lidar's 20fps) sensor. They have demonstrated having strong returns in the presence of artifacts such as smoke, and also in seeing small targets such as pedestrians further away. They do however, require a user to specify at what depth per pixel one want's sensed, through providing a physically possible 2.5D ruled surface that intersects the world to provide returns.

Conversely, depth estimation through RGB cameras have been a heavily researched area, but do not provide the realiability in estimates as compared to a Lidar, which can be cost prohibitive. This has made both RGB and Lidar's en-masse adoption in safety critical systems such as Advanced driver-assistance systems (ADAS) rare.

Our work, looks into marrying both modalities to counter each sensor's limitations, through the use of Geometric, Probabilistic and Physical constraints. We begin by formulating an iterative Bayesian inference approach to depth sensing using the Light Curtain alone, leveraging on Depth Probability Volume (DPV) and corresponding 2D Uncertainty Field (UF) representations. We then build upon existing planning framework based on Dynamic Programming using the Light Curtain's Galvanometer velocity and acceleration constraints. Lastly, we expand upon the state of the art in Generating and Exploiting Probabilistic Depth Estimates using Deep Learning based approaches, and use that as a prior to drive Light Curtain placement.

2. Prior Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation

ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. [3] [2]

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. [1]

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3. Sensor Setup
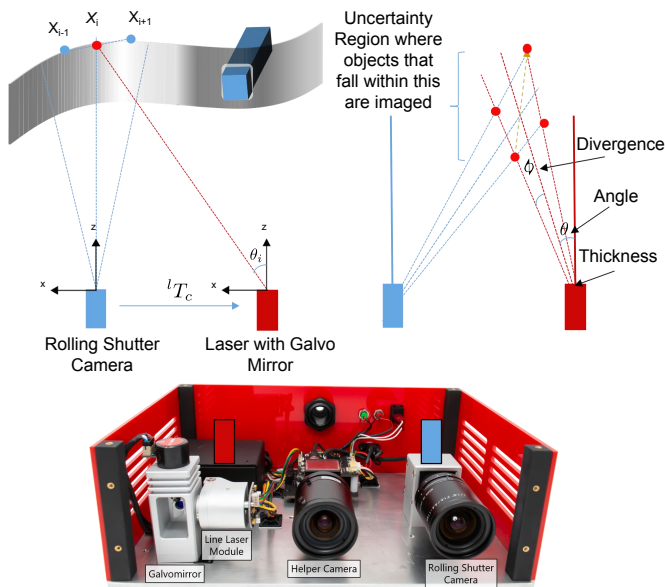
3.1. Light Curtain Basics

Figure 2: The Programmable Light Curtain device, consisting of a steerable laser, an IR camera and a microcontroller, capable of generating a slice in space to image in 3D

The Light Curtain consists of a rolling shutter NIR camera flipped vertically, a Line Laser module and a Galvomirror. The rolling shutter NIR camera runs in sync with the laser, where each row of the camer can be thought of as a

plane with some divergence going out vertically into space. The laser/projector fires a similar vertical sheet of light as a plane into space, which intersects with the camera's vertical plane to produce an intersecting volume. Any objects that lie within this space will then be imaged by the camera. We we will know the 3D position of said objects. Hence, a top-down XZ profile can be provided in the NIR camera's frame, to generate a ruled 3D curtain surface at 50fps
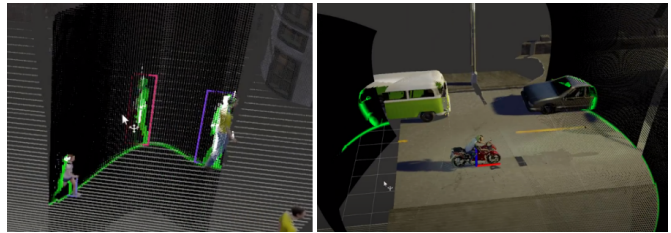
3.2. Simulator

Figure 3: Light Curtain Simulator in CARLA environment

We also designed and open sourced a light curtain simulator to test and evaluate our algorithms, allowing for the Light Curtain paramters such as NIR instrinsics, Laser/NIR extrinsincs, Galvomirror speed etc. to be controlled.
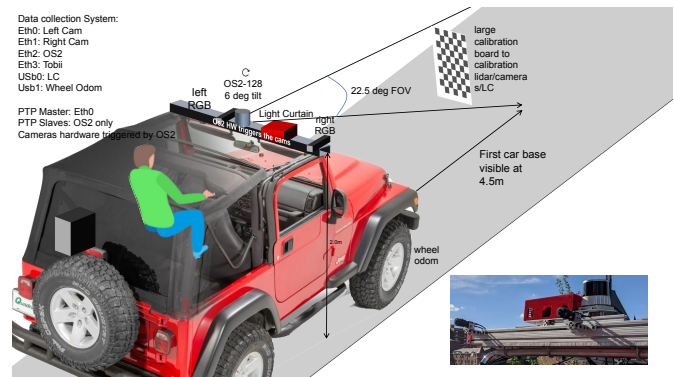
3.3. Sensor Array

Figure 4: Sensor Array consisting of a Stereo camera pair, the Light Curtain device and a 128 Beam Lidar

Our sensor array consists of a Stereo Camera Pair with a baseline of 0.7m, the Light Curtain device located close to the main/left camera to minimize depth/volume transformation parallax artifacts, and an Ouster OS2-128 Lidar for algorithm accuracy validation and to assist in training depth estimation networks. All sensors are calibrated to the common reference frame of the left RGB camera.

4. Formulation

We begin by looking at a Light Curtain only problem of adaptively discovering the depth of a scene. We wouldn't know the path along the rays/pixels on which an object may lie/intersect, hence we choose to go with a sampling based approach. We treat each ray as initially having either a uniform distribution, or a gaussian with a large variance in the center, and we attempt to formulate our problem with a Recursive Bayesian update approach.

4.1. Representation

Our state space is represented as a tensor of some fixed resolution [320x240] with each pixel in image I encoding the depth value $\mathbf{d}(u,v)$ as a bernoulli distribution $P(\mathbf{d}(u,v))$. D_c represents the list of depths quantizing the space of each pixel defined within (d_{min}, d_{max}) of some fixed size N [64], resulting in a Depth Probability Volume (DPV) tensor of size [64, 240, 320]

$$D_c = \{d_0...d_n\} \quad d_i = d_{min} + (d_{max} - d_{min}) * t \quad (1)$$
$$P(\mathbf{d}(u,v)) = I(u,v)$$
$$\sum_d (P(\mathbf{d}(u,v))) = 1 \quad E[P(\mathbf{d}(u,v))] = \mathbf{d}(u,v) \quad (2)$$

While an ideal sensor could choose plan a path to sample in the full 3D volume, our Light Curtain device only has control over a collapsed XZ space. Hence, we select a subset of rays that correspond to a plane that we wish to sample, and generate an Uncertainty Field (UF) as follows:

$$P(\mathbf{d}(u)) = \frac{\sum_{u,v} P(\mathbf{d}(u,v)).\mathbf{1}}{\sum_{u,v} \mathbf{1}}$$
$$where \; h_{min} > E[P(\mathbf{d}(u,v))] > h_{max} \quad (3)$$
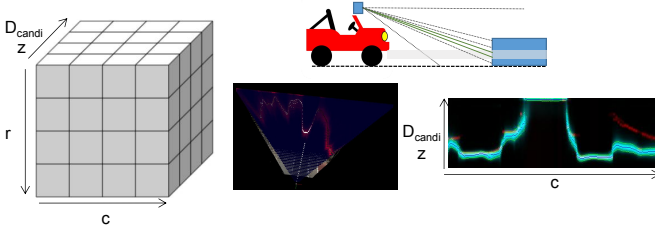

Figure 5: Our state space consisting of a Depth Probability Volume (DPV) its corrsp Bird's Eye Uncertainty Field (UF)

4.2. Curtain Planning

With the Uncertainty Field (UF) extracted from the state space, we can use this to figure out where to place light curtains. We build upon prior work solving Light Curtain placement as a Constraint Optimization and Dynamic Programming problem. A single light curtain placement is defined by a set of T control points $\{\mathbf{X}_t\}_{t=1}^T$. We wish to maximize the objective $J(\mathbf{X}_1, \ldots, \mathbf{X}_T) = \sum_{t=1}^T UF(\mathbf{X}_t)$ where $UF(\mathbf{X})$ is the uncertainty field probabilies at the anchor location of $\mathbf{X}$.

The control points $\{\mathbf{X}_t\}_{t=1}^T$, where each $\mathbf{X}_t$ lies on the the camera ray $\mathbf{R}_t$, must be chosen to satisfy the physical constraints of the light curtain device: $|\theta(\mathbf{X}_{t+1}) - \theta(\mathbf{X}_t)| \leq \Delta\theta_{max}$ with θ_{max} being the maximum angular velocity of the Galvomirror. The problem is also discretized such that $\mathbf{X}_t \in D_c$ and also lies along $\mathbf{R}_t$

$$\arg \max_{\{\mathbf{X}_t\}_{t=1}^T} \sum_{t=1}^T UF(\mathbf{X}_t) \qquad \text{where } \mathbf{X}_t \in D_c$$
subject to $|\theta(\mathbf{X}_{t+1}) - \theta(\mathbf{X}_t)| \leq \Delta\theta_{max}, \; \forall 1 \leq t < T \quad (4)$
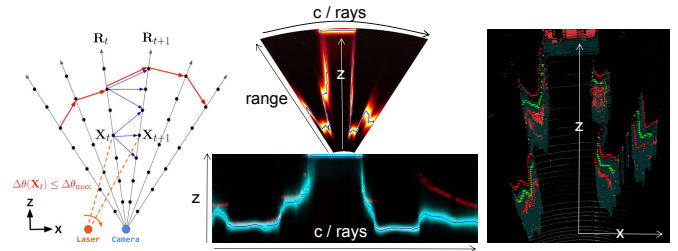

Figure 6: Left: Light Curtain constraint graph subject to max angular velocity of Galvomirror. Right: Placing an optimal curtain along the highest probability region per column of rays

In the figure above, we have placed a single curtain along the highest probability region per column of rays, but ideally, we should place more curtains spanning the uncertaintiy of those distributions. To do this, we generate corresponding entropy fields $H(\mathbf{X})_i$ to be fed to the planner from $UF(\mathbf{X})$ based on two approaches: $m0$ attempts to normalize each ray's distribution $P(\mathbf{d}(u))$ and warp the distribution such that a selected span from the mean is maximized. $m1$ attempts to sample a point on the ray given $P(\mathbf{d}(u))$.

As seen in Fig. 7, strategy $m0$ is able to generate fields that adaptively place additional curtains around a consistent span around the mean, but is unable to do so in cases of multimodal distributions. $m1$ on the other hand is able to place a curtain around the 2nd modality, albeit with a lower probability. The inconsistency from ray to ray in $m1$ however, may be impossible to image due to it exceeding the acceleration bounds, hence a spline fit is used with control points every 5 to 10 rays, resulting in an imagable but non-flat curtain placement. We see the effects of both in later experiments.
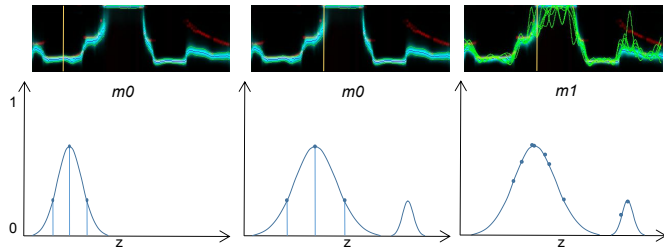
Figure 7: We look at a depth distribution of one of the rays in UF (yellow line), and figure out where additional curtains (blue points) can be placed such as to maximize information gained. Observe that $m1$ is able to handle multimodal distributions

4.3. Observation Model

We now have returns from curtains $\mathbf{C}$ with each $\mathbf{C}_i$ containing $[x, y, z, i]$, the 3D position and intensity value in the same spatial resolution as I, planned based on a particular policy, conditioned on our prior distribution $P(\mathbf{d}_{t-1})$ at time t. We need to convert $\mathbf{C}_i$ into a likelihood distribution $P(\mathbf{c}_t|\mathbf{d}_t)$, such that this Hidden Markov Model representation and sum of log likelihoods holds true:

$$P\left(\mathbf{d}_0, ..., \mathbf{d}_k, \mathbf{c}_1, ..., \mathbf{c}_k\right) = P\left(\mathbf{d}_0\right) \prod_{t=1}^{k} P\left(\mathbf{c}_t|\mathbf{d}_t\right) P\left(\mathbf{d}_t|\mathbf{d}_{t-1}\right)$$

$$\log\left(P\left(\mathbf{c}_t|\mathbf{d}_t\right)\right) = \sum_{i=0}^{n}\left(\log\left(P\left(\mathbf{c}_{it}|\mathbf{d}_t\right)\right)\right)$$

$$(5)$$

References

[1] Ancha, Raaj, Hu, Narasimhan, and Held. Active perception using light curtains for autonomous driving. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2020. 2

[2] Whittaker Joe Bartels, Jian Wang and Srinivasa G. Agile depth sensing using triangulating light curtains. In *2019 IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[3] Jian Wang, Joseph Bartels, William Whittaker, Aswin C Sankaranarayanan, and Srinivasa G Narasimhan. Programmable triangulation light curtains. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 19–34, 2018. 2