

Scratching the Surface: Sensor Driven Perception and Action

Caveat: this is really about perception, the action is done by our partners!

UDRC Theme Meeting, November 2021

Presented by Andy Wallace

Thanks to Sen Wang, Joao Mota, Andreas Assmann, Marcel Sheeny, Sap Mukherjee, Zhiyang Hong, Ali Ahrabian, Abde Halimi, Yvan Petillot, Gerald Buller, Stephan Matzka and others

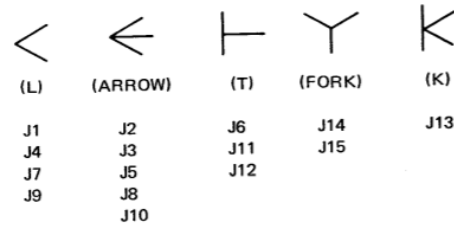
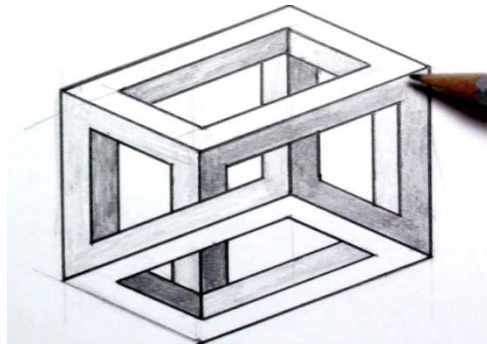
“The Treachery of Images”



1929: This isn't a pipe

26 The Psychology of Computer Vision

- ⊥ Shadow edge
- Concave edge
- + Convex edge
- Obscuring edge
- C Crack edge



2021: This is a Volkswagen



1971 (Waltz): Is this a possible object?

Scratching the Surface

In most instances of vision-guided autonomy, certainly land based autonomy, what we sense is the reflection or emission of 'light' from **surfaces**.

For autonomy, **we usually want to create a dense map** of the surrounding surfaces, and to get from A to B without making **damaging** contact with any of those surfaces.

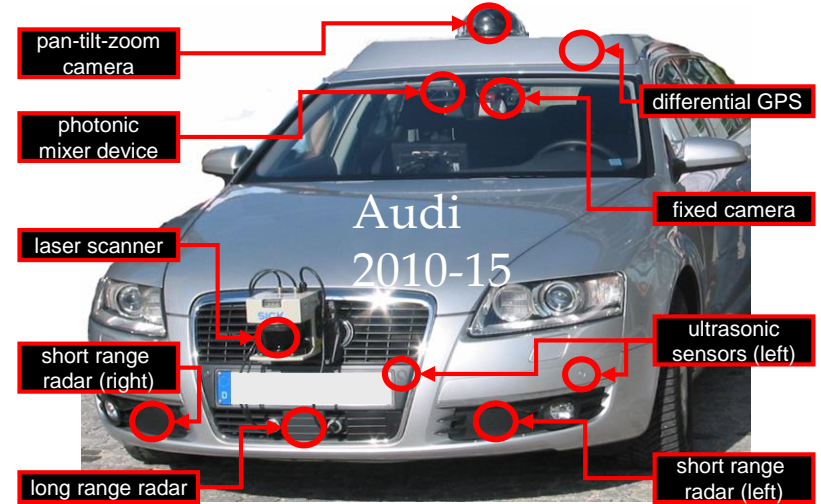
We may or may not wish **supplementary information**, to get from A to B more easily, or because we have an additional objective, understand the scene, locate a particular object or person, measure and collect samples of data etc.

Despite the alleged hostility of certain companies (e.g. Tesla) to active sensing, it seems that the **way forward in the short term is most likely through multiple, passive and active sensors**, notably Cameras (stereo?), **Radar** and **LiDAR**, to somehow get round the 'treachery of images'.

Understanding Surface Data

This talk 'scratches the surface' of some of the work we have done using (primarily) active surface sensing using radar and LiDAR.

Although generic (mostly) it has mainly been evaluated using automotive vehicles.



Chapter 1: Collecting data

One of the frustrations we have had in the last few years was an inability to find the right data, in particular coincident data from LiDAR, stereo cameras and radar 'in the wild' and in adverse conditions.

Too many people seemed to be content to process and re-process Kitty (442/367 object 2D/3D on Monday!), NuScenes, Synthia, NGSIM, ImageNet and CIFAR datasets

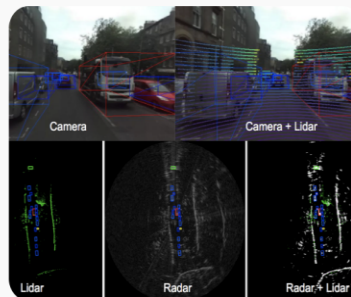
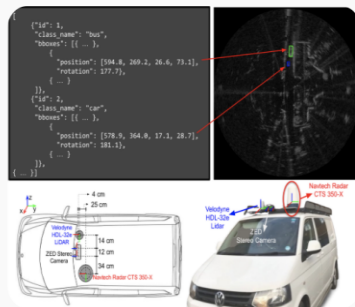
For that reason, we decided, belatedly, to collect our own data...this is now available publicly and has had several hundred downloads (since April 2021).

Welcome to Heriot-Watt RADIATE Dataset Website

Multi-Modality Radar Dataset in Adverse Weather with Object Annotation

RADIATE (RADar Dataset In Adverse weaThEr) is a high-resolution radar dataset which includes about 3 hours annotated radar images and more than 200K labelled instances on public roads. It focuses on multi-modal sensor data (radar, camera, 3D LiDAR and GPS/IMU) in adverse weather conditions, such as dense fog and heavy snowfall. It aims to facilitate research on object detection, tracking, Simultaneous Localization and Mapping (SLAM) and scene understanding using radar sensing in extreme weathers.

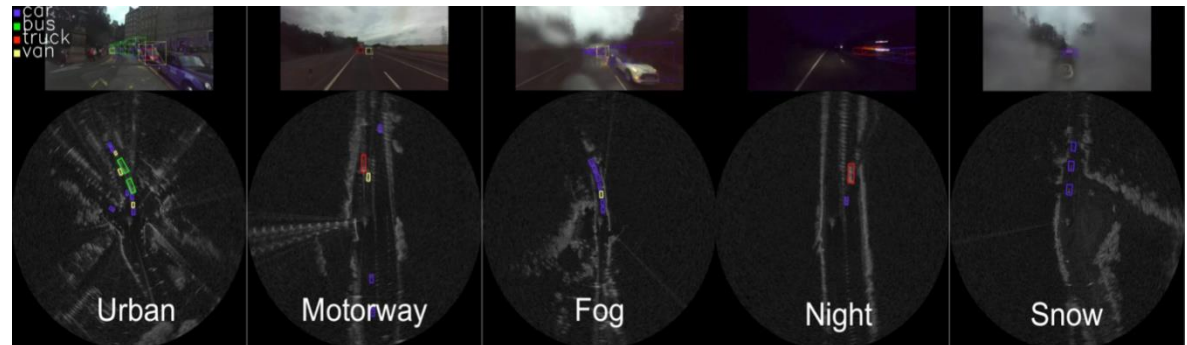
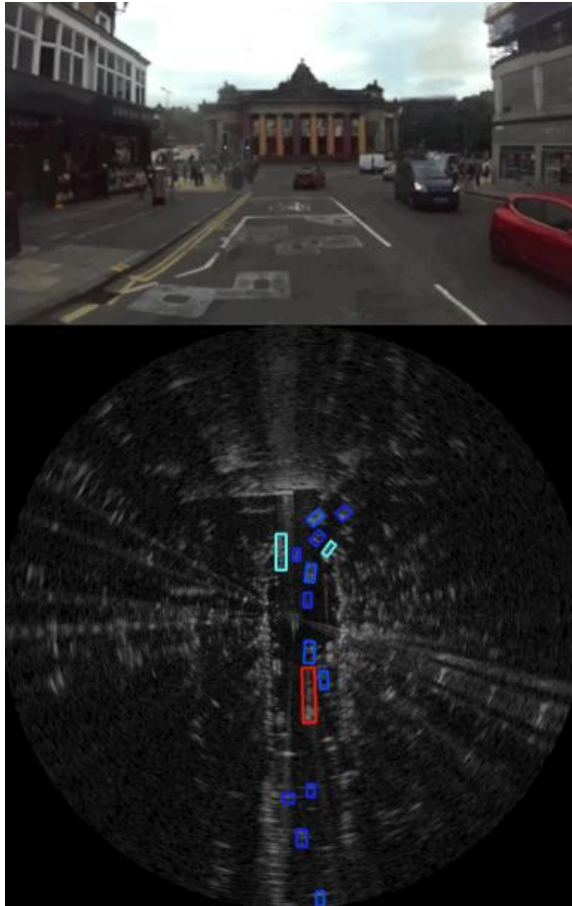
Choose to Start



The Radiate Dataset

Coincident, labelled radar, LiDAR (Velodyne) and stereo camera data is available to download together with GPS and INU data.

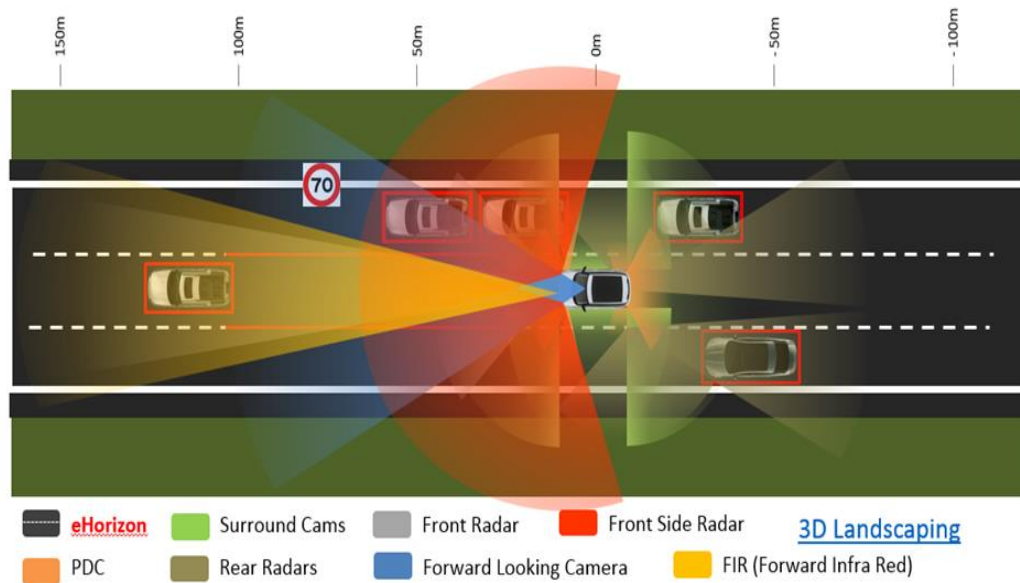
This data was collected mainly around Edinburgh and the surrounding countryside, although we also collected on- off-road data in the Cairngorms.



Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, Andrew Wallace. RADIATE: A Radar Dataset for Automotive Perception. ICRA 2021

Alternative sensing technologies: Pros and Cons

	3D Landscaping	Object Classification	Range	Operation in Adverse Weather	Operation at night
Camera	Yellow	Green	Yellow	Red	Red <small>->LWIR</small> Yellow
LiDAR	Green	Yellow	Yellow	Red	Green
Radar	Red	Red	Green	Green	Green



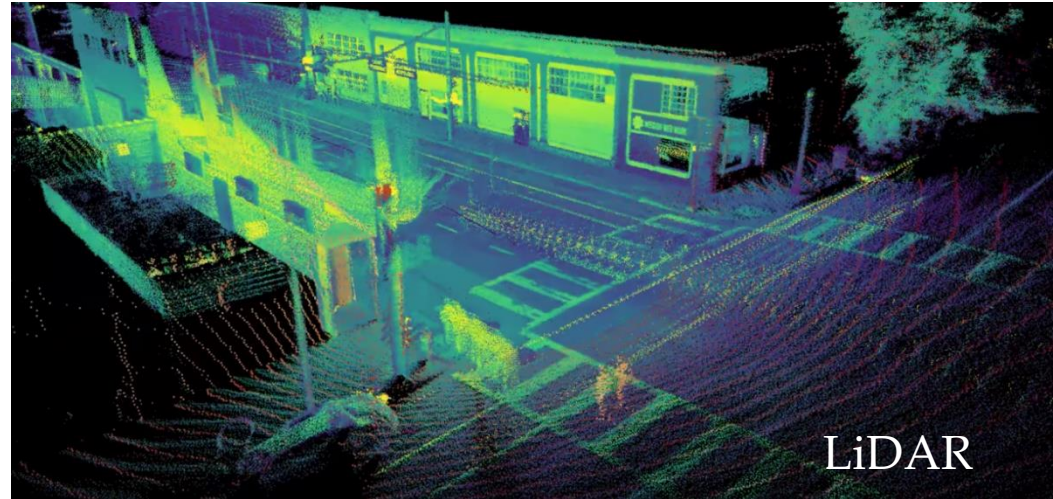
Chapter 2: LiDAR: problems and potential solutions

Good

- Direct surface measurement
- High resolution – sub-cm in (x,y,z)
- Separates the viewed scene into ‘planes’ and hence isolates ‘targets’
- Very sensitive, (at least single photon systems are)
- Can be long range – several Km

Problems

1. Mechanical scanning
2. Vehicle data tends to be regular, but sparse
3. Eye safety limits power (and range) in autonomy/assistance applications
4. Volume of data dictates slow and simple processing
5. Problems with fog/smoke/mist/snow/rain penetration



Solutions

1. Solid State Cameras
2. ‘Compressed’ sensing, larger array mosaics, video fusion
3. Sparse random projection, higher λ
4. Hardware acceleration, better algorithms
5. Higher λ , multi-return processing, fusion

Wallace, Halimi and Buller, “Full Waveform LiDAR for Adverse Weather Conditions”, IEEE Transactions on Vehicular Technology, 69(7), 7064-77, 2020.

Fast, efficient, processing: reduce the laser power and implement in parallel

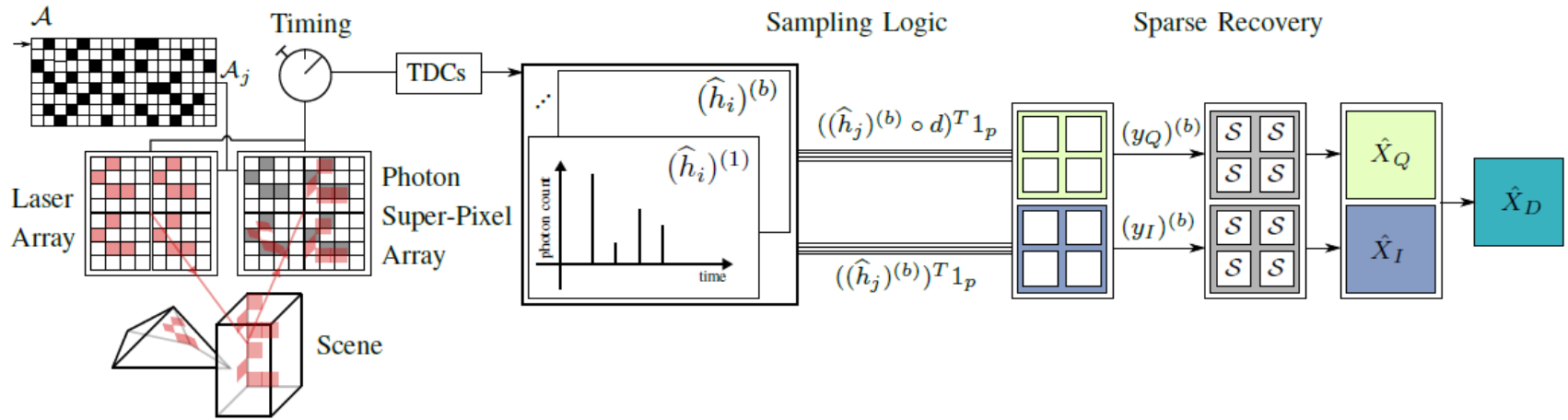
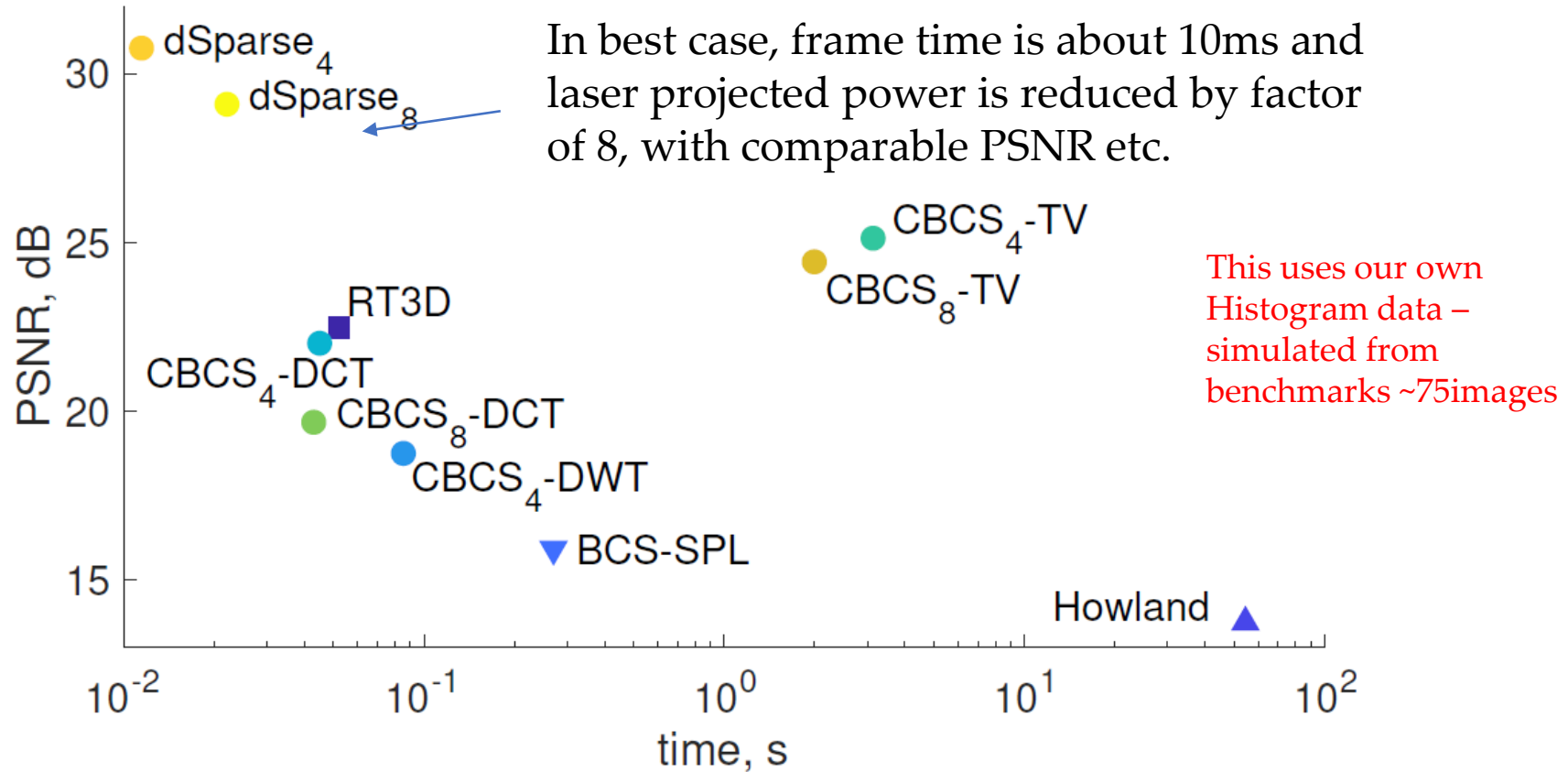


Fig. 2: Proposed compressive block LiDAR sampling. A laser array is partitioned into B blocks, each of which illuminates the scene independently. Information about the received photons is collected into histograms, from which $y_Q^{(b)}$ and $y_I^{(b)}$ in (5) are formed within each block b . A depth image \hat{X}_D is constructed by processing these measurements in parallel, which entails solving several instances of (6), represented by \mathcal{S} and possibly with some post-processing, and forming \hat{X}_Q and \hat{X}_I .

$$\begin{aligned}
 & \underset{x_Q}{\text{minimize}} \quad \frac{1}{2} \|Ax_Q - y_Q\|_2^2 + \alpha_Q \|\Theta x_Q\|_1 \\
 & \underset{x_I}{\text{minimize}} \quad \frac{1}{2} \|Ax_I - y_I\|_2^2 + \alpha_I \|\Theta x_I\|_1,
 \end{aligned} \tag{6}$$

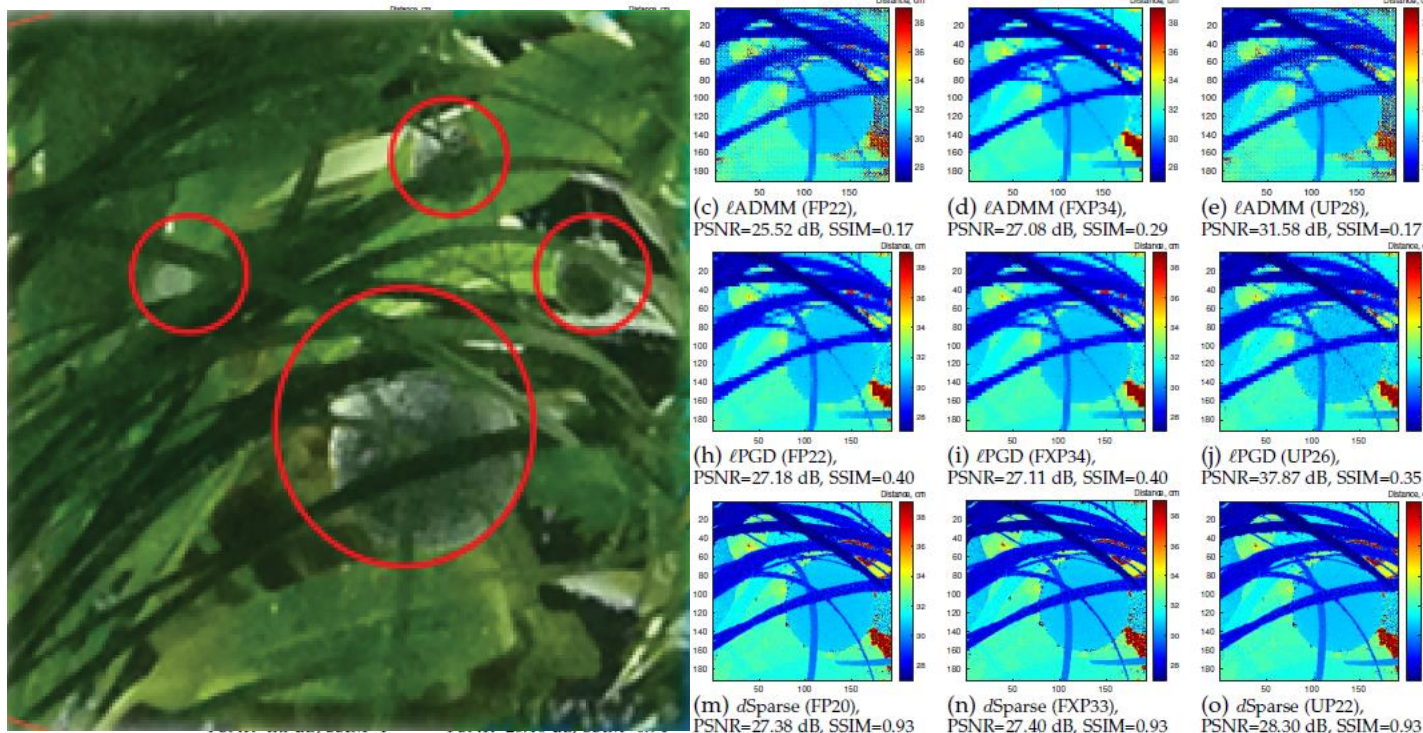
Examples of Θ include the Wavelet and DCT transforms, or a difference matrix (which expresses the fact that x_Q and x_I have sparse gradients ([22], [23])). Such assumptions enable us to estimate x_Q and x_I concurrently using CS methods, as

Results: fast parallel GPU-based LiDAR processing



A Aßmann, B Stewart, J. Mota and AM Wallace, “Compressive super-pixel LiDAR for high frame rate 3D Depth Imaging” IEEE Global Conference on Signal and Image Processing, 2019
(Extended work and submission/revision to IEEE-TCI November 2021)

Next, reduce the arithmetic precision (FPGA)



Data from a previous DSTL project on underwater mine detection (192x192 pixels)

Fig. 8: A visual comparison of the final depth map reconstruction using real underwater scene photon count data for a real depth scene from [17]. The image size is 192×192 .

“For dSparse ... with customized floating point, the resource saving is 85% in LUT, 80% in both DSP and BRAM. and 87% power saving is achieved with over 67% data ratio and 75% processing latency reductions.” p.s. **best parallel block time ~0.01ms.**

Wu, Aßmann, Stewart and Wallace, “Energy Efficient Approximate 3D Image Reconstruction”, IEEE Transactions on Emerging Topics in Computing, 2021,

Why does it matter?: use approximation to reduce size, weight and power and make laser systems safe and covert ...



VL53L0X integrates a leading-edge SPAD array

NanEye: 1mmx1mm, 320x320, 50fps



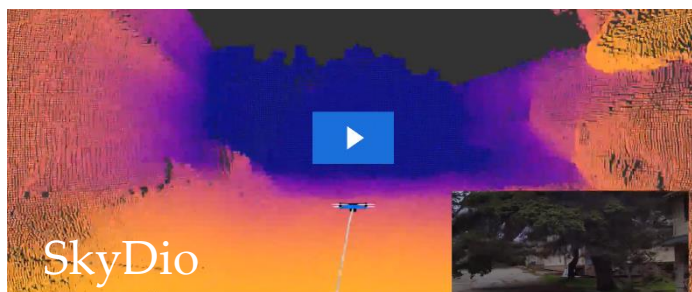
POWER BUDGET

A medium size drone

- Consumes 100-200W
- 20 to 40 minute flight time

1 Watt of compute reduces flight time by ~10s
4g of payload reduces flight time by ~10s

Final compute constraint is mostly driven by size/weight of the board

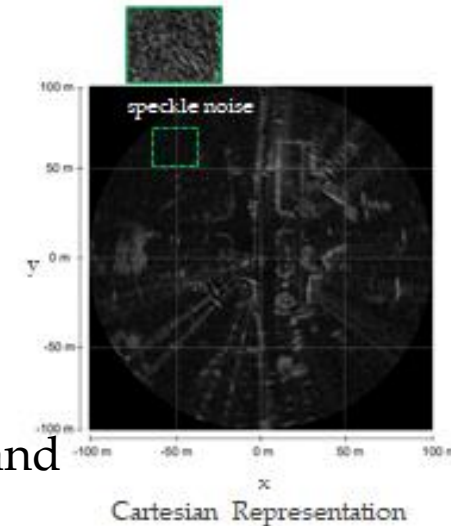


2021 Skydio
“Hot Chip” talk

.... for mapping, object recognition, scene segmentation and motion planning

360° frequency modulated continuous wave (FMCW) radar

Chapter 3: Radar Image Analysis: Mapping and Navigation



While our radar senses surface reflection, it scans in azimuth only, and the data is a range-azimuth power density map.

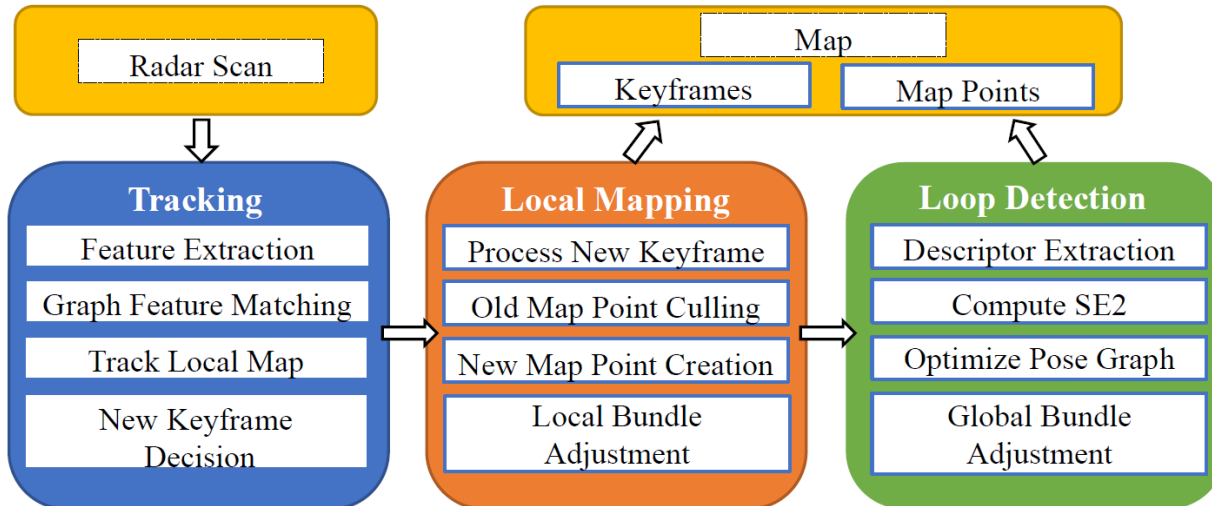
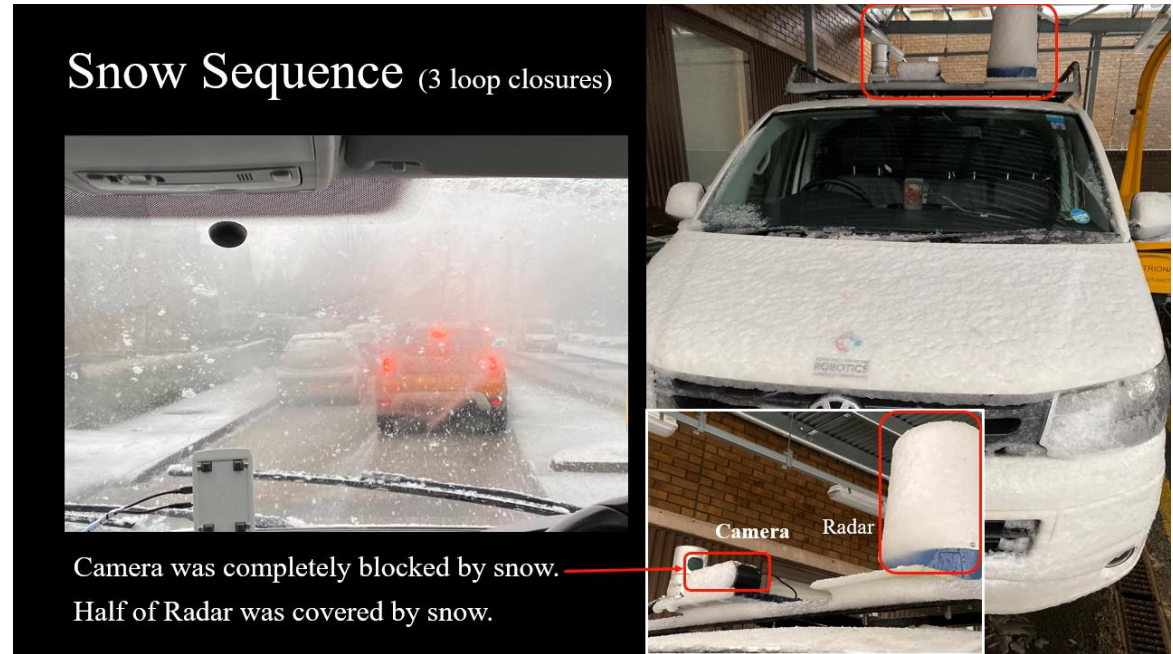
Research at the Universities of Birmingham and Edinburgh (and elsewhere) has been directed towards the acquisition of high resolution 3D surface maps** (looking a TDOA and SAR), but we have been constrained to a 2D plan view.

So, it might correspond to a surface, occupancy grid planning model.

** covered at July 2020 Theme meeting

Radar Location and Mapping

This procedure has been tested at night, in dense fog and heavy snowfall, in a GPS denied environment.



Hong, Petillot, Wallace, Wang:
“Radar SLAM: A Robust SLAM System for All Weather Conditions”
Recommended for publication in International Journal of Robotics Research, November 2021
Preprint -arXiv:2104.05347v1

Object Classification at 300GHz

In this example, we used a small set of six objects, viewed in isolation for training, in cluttered scenes for evaluation.

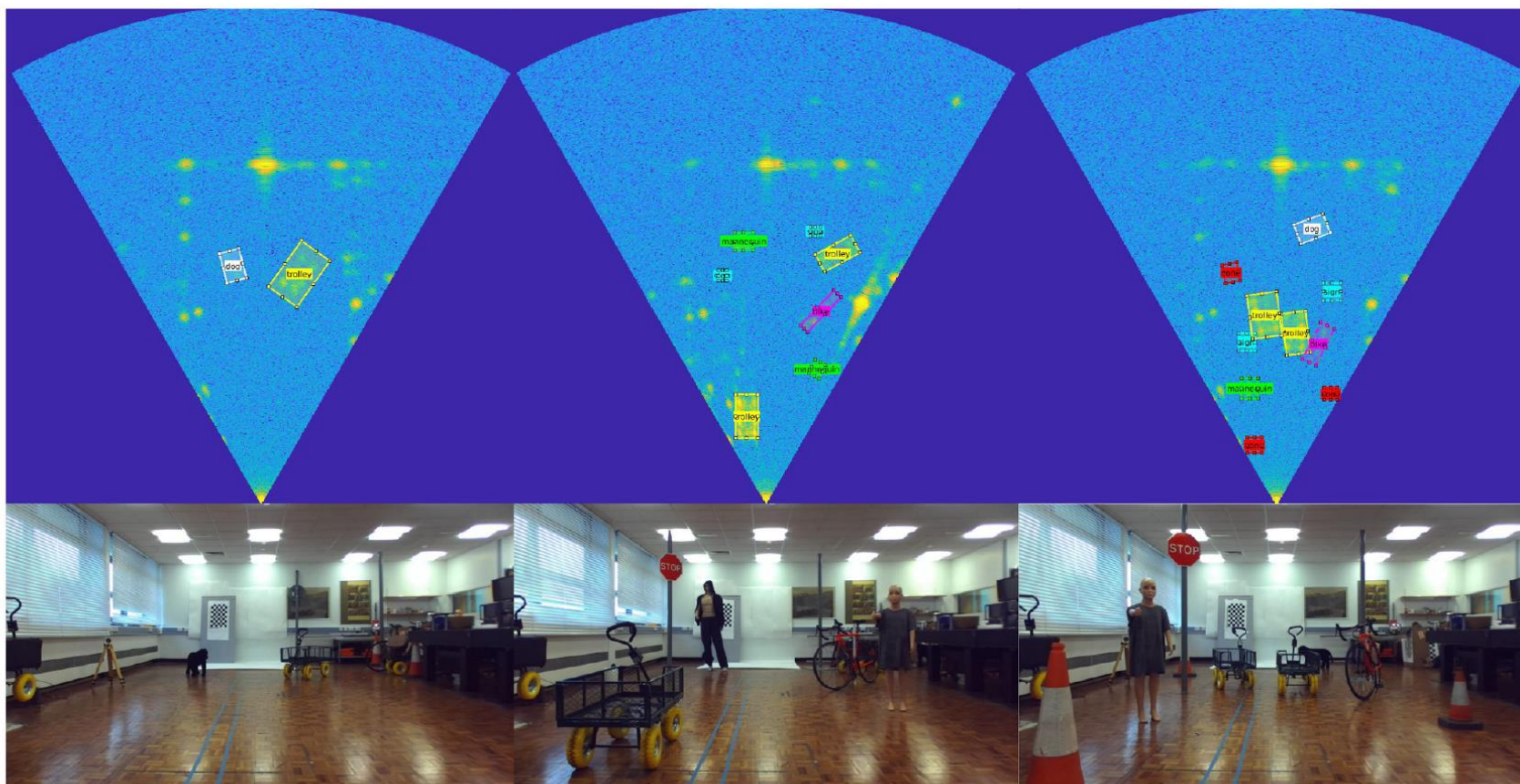
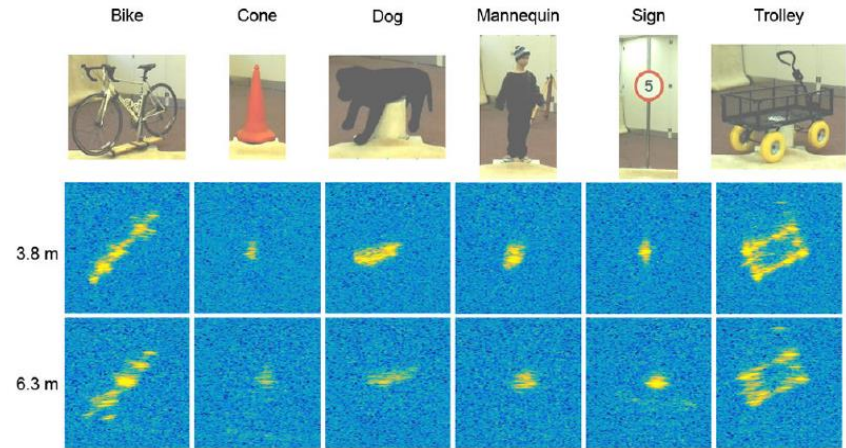
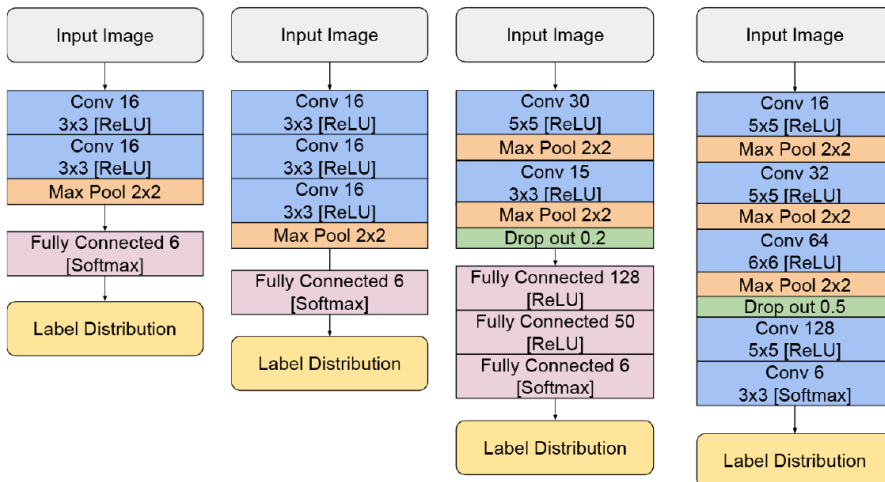


Fig. 9 Multiple object dataset. Above: 300 GHz radar image. Below: reference RGB image

Experiments in the Lab with 300GHz data: several architectures, with and without transfer learning from Mstar, data augmentation. On isolated objects, classification rates are 80-90%



status used

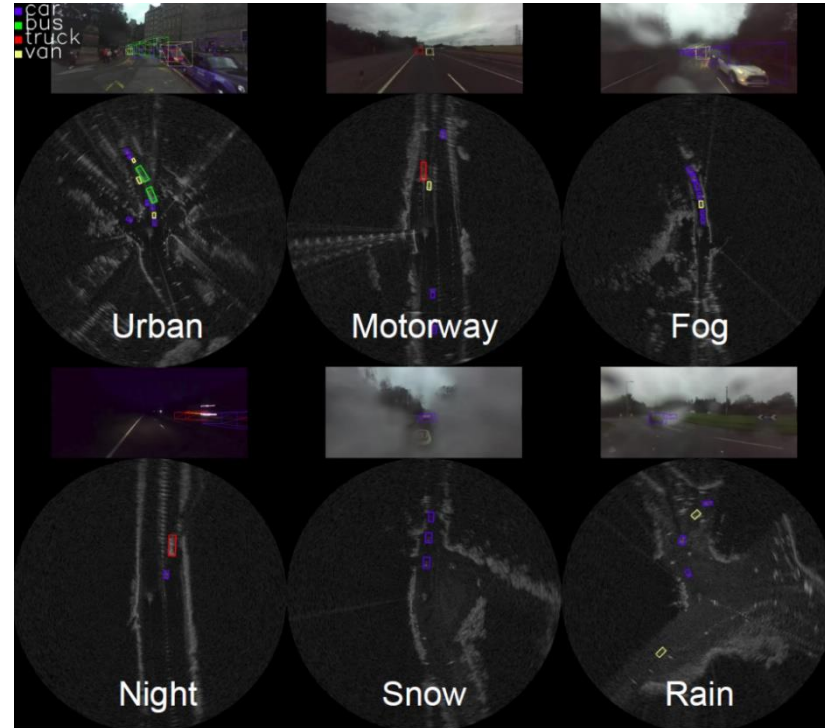
AP	Overall	#Objects < 4				4 ≤ #Objects < 7				#Objects ≥ 7				Short	Mid	Long
		Overall	Short	Mid	Long	Overall	Short	Mid	Long	Overall	Short	Mid	Long			
bike	65.78	81.7	50.0	88.89	75.0	53.82	0.0	59.92	66.67	66.36	N/A	67.2	N/A	25.0	69.31	71.43
bike (ours)	71.94	90.98	50.0	94.44	100.0	64.55	100.0	59.99	100.0	53.43	N/A	48.92	N/A	75.0	67.17	100.0
cone	42.29	51.25	50.0	61.11	50.0	66.44	65.08	83.33	28.57	34.35	62.5	23.19	60.44	52.71	24.22	
cone (ours)	49.96	66.67	50.0	66.67	100.0	69.23	66.67	83.33	35.07	62.5	13.33	0.0	62.07	37.04	45.31	
dog	26.72	45.07	55.56	47.32	30.95	25.22	33.33	36.67	16.67	13.68	50.0	N/A	7.5	48.0	38.33	9.44
dog (ours)	67.26	86.43	88.89	99.05	11.11	64.74	83.33	60.42	50.0	30.91	40.0	N/A	0.0	68.0	85.02	6.67
mannequin	42.05	71.28	53.33	87.02	55.56	28.3	42.42	34.18	7.5	44.47	14.29	58.82	19.25	34.72	54.16	14.94
mannequin (ours)	41.15	82.61	83.33	92.86	33.33	28.56	45.45	26.67	14.55	36.28	14.29	52.94	10.92	45.83	49.94	10.62
sign	49.36	41.3	0.0	44.71	40.0	59.77	N/A	62.5	57.89	40.35	N/A	41.94	38.46	0.0	50.57	49.28
sign (ours)	77.26	87.01	100.0	90.09	66.67	74.49	N/A	80.19	70.27	76.92	N/A	77.78	76.19	100.0	82.52	72.13
trolley	84.61	90.23	84.72	99.92	76.47	87.35	100.0	91.66	68.81	78.51	95.99	85.78	26.67	96.29	91.69	61.94
trolley (ours)	90.08	97.08	97.62	99.44	94.12	94.16	100.0	96.42	81.36	83.82	100.0	87.79	44.37	99.01	92.91	75.11
ine mAP	51.86	63.47	48.94	71.49	54.66	53.48	48.17	61.38	41.02	46.29	55.69	59.25	23.01	44.07	59.46	38.54
mAP (ours)	66.28	85.13	78.31	90.43	67.54	65.96	79.09	65.06	66.58	52.74	54.2	56.15	26.3	74.99	69.1	51.64

Object detection in the wild

The wild data is challenging - networks were trained on cars, vans, trucks, buses, motorbikes and bicycles.

Motion was NOT used. Tracking and/or Doppler analysis may give better results BUT scenes are cluttered and cars and pedestrians stop!

Cameras and LiDAR were not effective.



		Overall	Static	Motorway	Urban	Night	Rain	Fog	Snow
Radar	Faster RCNN ResNet-50 Trained on Good and Bad Weather	53.57	88.19	44.47	42.58	73.02	48.08	70.69	22.45
	Faster RCNN ResNet-50 Trained on Good Weather	52.77	88.08	49.03	35.05	64.03	43.50	62.02	27.63
	Faster RCNN ResNet-101 Trained on Good and Bad Weather	54.43	87.86	47.54	42.09	74.22	51.79	63.04	26.70
	Faster RCNN ResNet-101 Trained on Good Weather	52.90	87.98	46.44	36.26	64.40	42.51	56.99	17.77
Lidar	Faster RCNN ResNet-50 Trained on Good and Bad Weather	19.91	40.59	11.18	20.17	19.38	15.53	19.48	1.23
	Faster RCNN ResNet-50 Trained on Good Weather	17.49	40.87	11.03	17.35	14.13	9.31	15.02	1.82
	Faster RCNN ResNet-101 Trained on Good and Bad Weather	20.66	40.25	11.74	20.17	19.18	16.85	36.41	1.82
	Faster RCNN ResNet-101 Trained on Good Weather	17.81	41.80	10.85	20.84	13.39	16.17	13.07	1.30

Table 7: AP results on each scenario using rectangular bounding boxes.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{AP} = \int_0^1 p(r) dr$$

Marcel Sheeny, Andrew Wallace and Sen Wang, “300 GHz Radar Object Recognition based on Deep Neural Networks and Transfer Learning”, IET Proceedings on Radar, Sonar and Navigation, 14(10), 1483-1493, 2020. (and ICRA paper)

Behaviour prediction using video and radar data

An LSTM encoder-decoder structure predicted the future trajectories and manoeuvres of vehicles in the radar data over a 3-5s (30-50 frames) time window

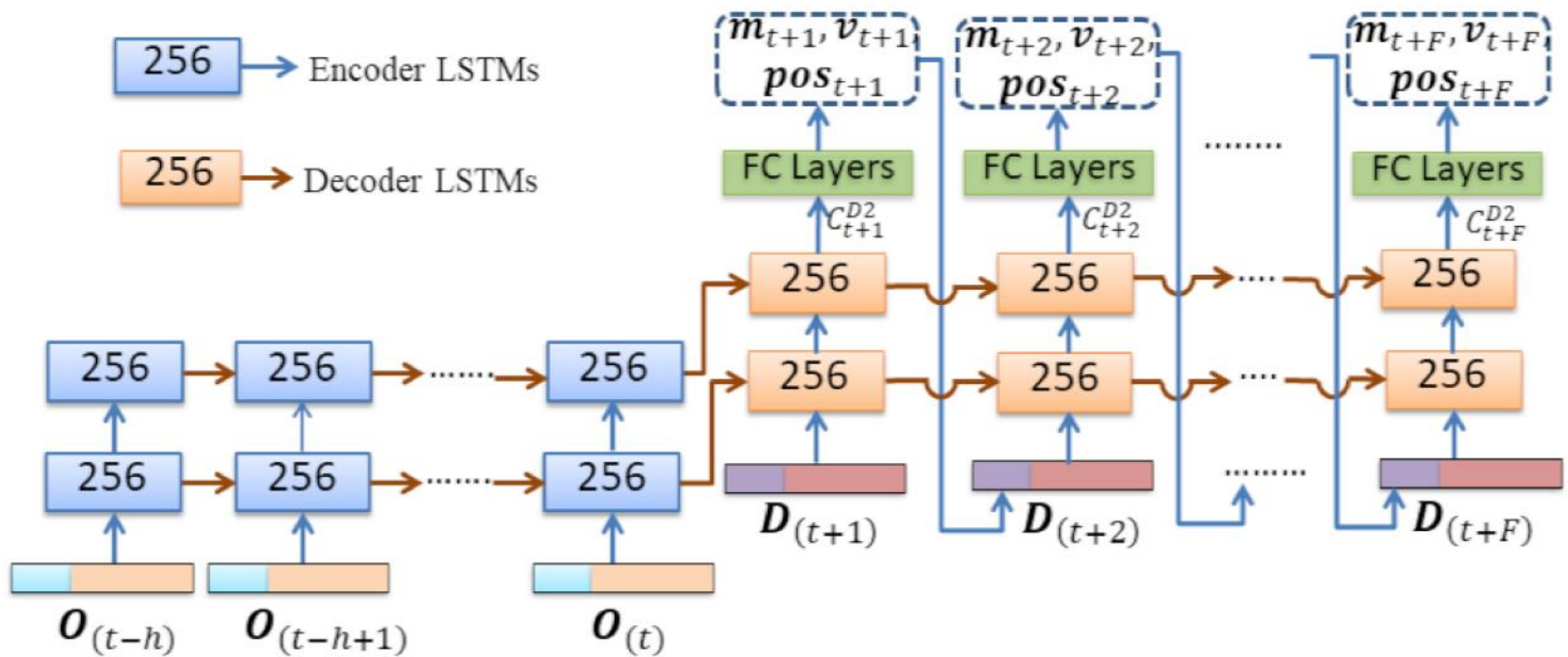
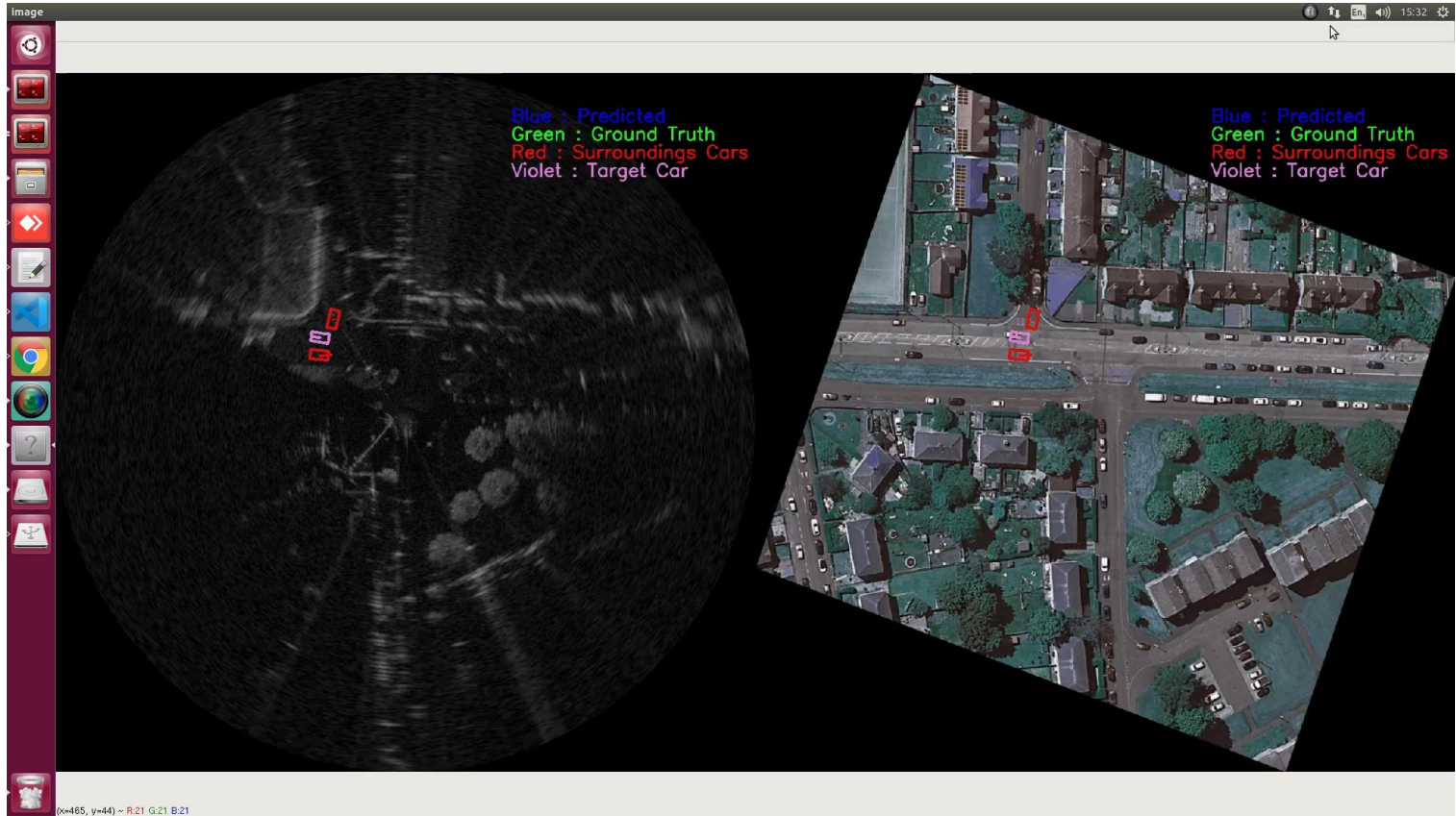


FIGURE 4. The proposed LSTM [40] based encoder-decoder structure, using the LSTM cell with 256 neuron count. FC Layers are the fully-connected layers shown in Fig. 5. $O_{(t-h)}$ to $O_{(t)}$ are the encoder-inputs shown in Fig. 2(a) consisting of the TV's features (sky-blue) and nearest ns SV's features (orange). $D_{(t+1)}$ to $D_{(t+F)}$ are the decoder-inputs shown in Fig. 2(b) consisting of the predicted entities of the TV (purple) and the predicted entities of nearest ns SVs (red).

Video showing behaviour prediction (Sighthill)



Quantitative Results: trajectory and manoeuvre prediction

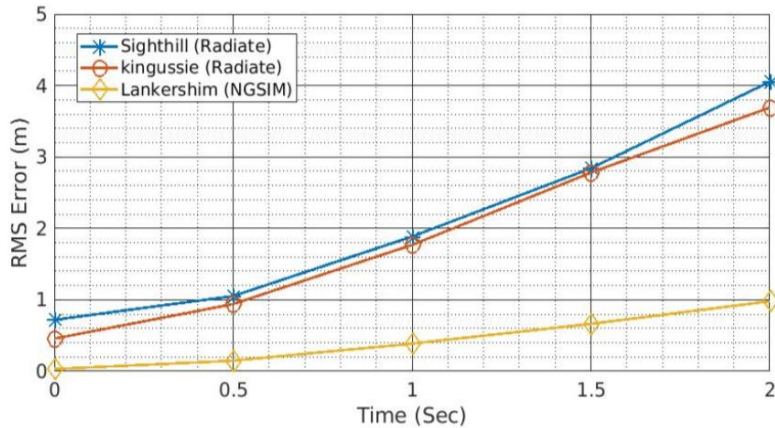


Fig. 7: The RMSE comparison of Lankershim (NGSIM), Kingussie (Radiate) and Sighthill (Radiate) dataset. This figure shows their average mean squared errors for the prediction time horizon from 1s to 2s.

These results are comparable to SoA on video data

Mukherjee, S., Wallace, A.M. and Wang, S., "Predicting Vehicle Behaviour using Automotive Radar and Recurrent Neural Networks", IEEE Journal of Intelligent Transportation Systems, 2, 254-268, 2021

TABLE III: Confusion matrix for maneuver classification at Lankershim junction (NGSIM dataset).

		Actual		
		Straight	Left	Right
Predicted	Straight	2346050	39300	35750
	Left	58050	962150	22950
	Right	30750	20700	350650

TABLE IV: Confusion matrix for maneuver classification at Kingussie junction (Radiate dataset).

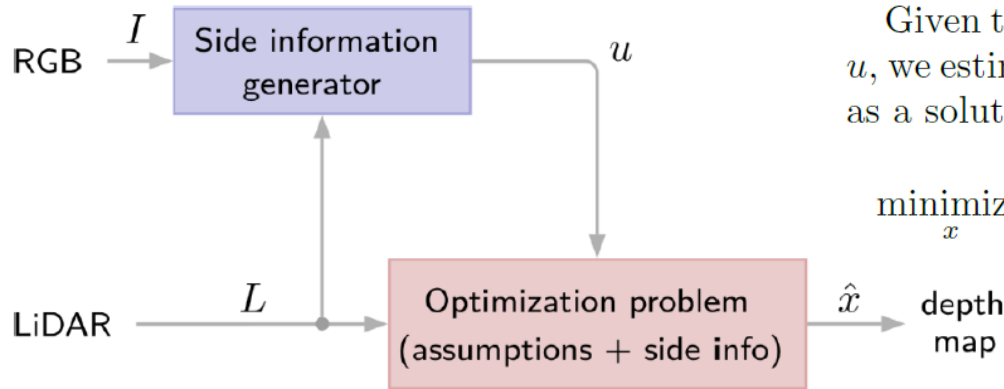
		Actual		
		Straight	Right	Left
Predicted	Straight	4755	6	–
	Right	4	3335	–
	Left	–	–	–

TABLE V: Confusion matrix for maneuver classification at Sighthill junction (Radiate dataset).

		Actual		
		Straight	Left	Right
Predicted	Straight	2800	14	20
	Left	22	1400	8
	Right	46	10	1320

Chapter 4: “Fusion” and full waveforms

Creating dense depth images from sparse data using intensity images



Given the sparse LiDAR map L and the side information u , we estimate the dense map of the scene, denoted $\hat{x} \in \mathbb{R}^n$, as a solution of

$$\text{minimize}_x \frac{1}{2} \|Sx - b\|_2^2 + \beta \|WHx\|_1 + \gamma \|Dx - u\|_1 \quad (1)$$

Fig. 1. Diagram of our method. Given an RGB image I and LiDAR data L , it generates a vector u that is used as side information in the optimization problem. The output is a dense depth map \hat{x} .

No physics? No semantics?

- Waltz Line labelling (1971)
- Beattie - Edge detection for semantically based early visual processing (1985)
- Zhang Physical modelling and combination of range and intensity edge data (1993)

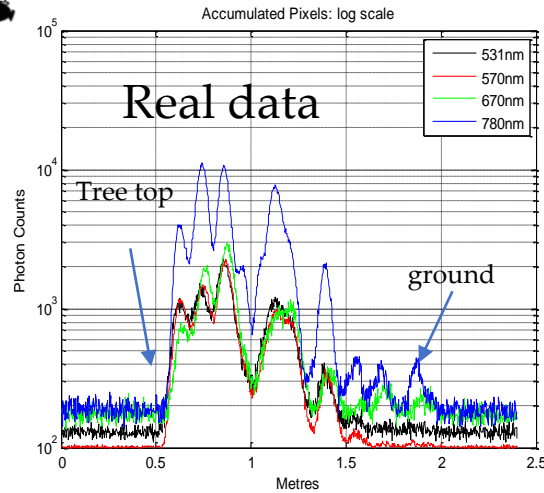


Full waveform processing



Outgoing

Returning



We can analyse the obscuring medium to measure bark and leaf area and NDVI

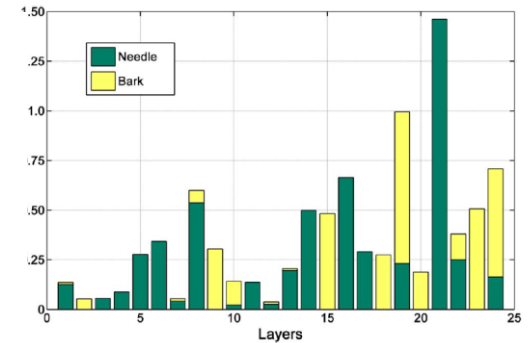
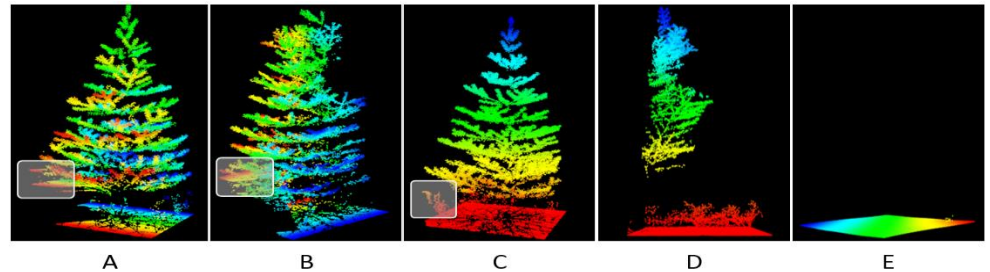


Fig. 15. Measured area profiles for leaf and bark as a function of canopy depth. These are measured as m²/m² at the irregularly spaced layer positions shown in Fig. 6. This assumes a random distribution of areas in each layer so that subsequent layers have occluded material.

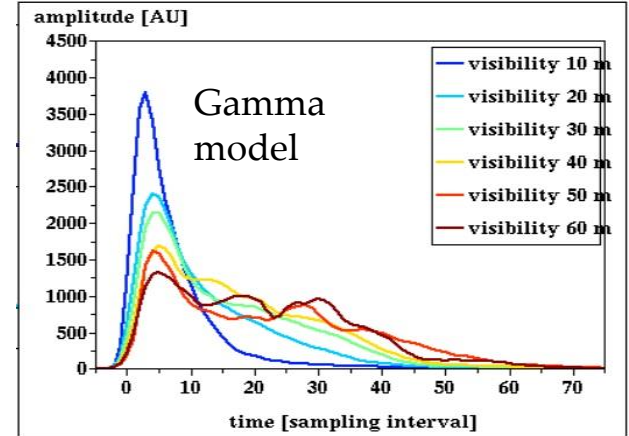
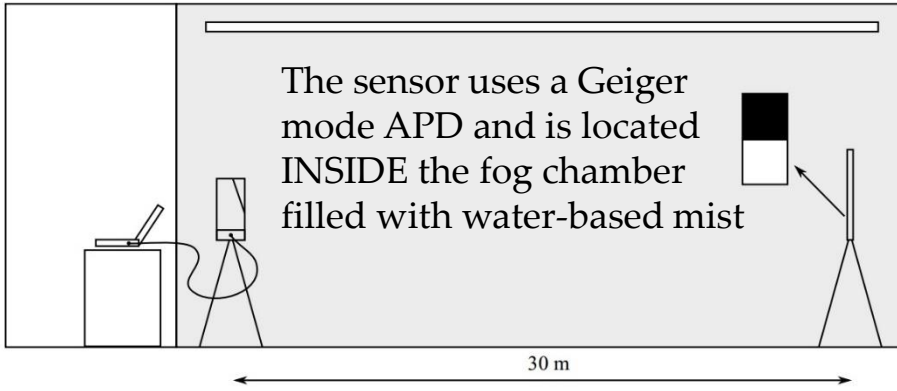
We can detect vehicles or rogue vegetation below the tree canopy.



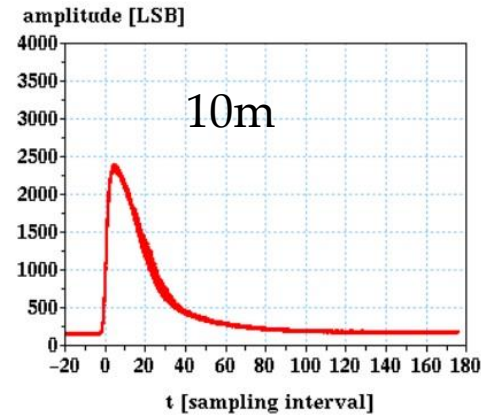
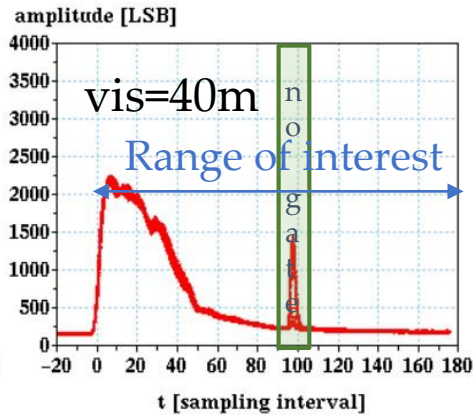
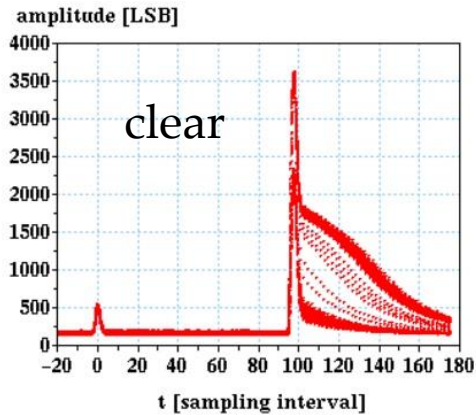
Multispectral!

Wallace et al. "Design and of Evaluation of Multi-spectral LiDAR for the Recovery of Arboreal Parameters" IEEE Transactions on Geoscience and Remote Sensing, 52(8), 4942-4954, 2014

Detecting surfaces through Obscuring Media



Mean waveforms, fog only



Detecting surfaces through Obscuring Media

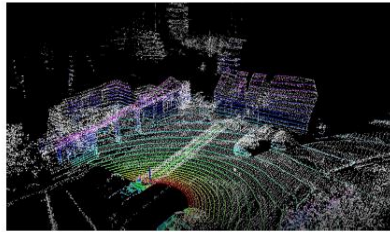


Fig. 7: Accumulated LiDAR 3D points using vehicle motion. The wall of the building in front and parked cars are clearly visible

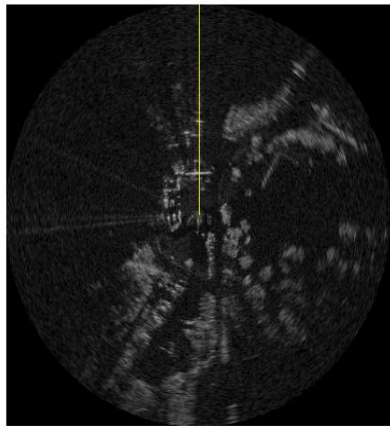


Fig. 8: The radar image; again the wall and cars are visible. The yellow line denotes the line of sight for the real radar power waveform in Fig. 9

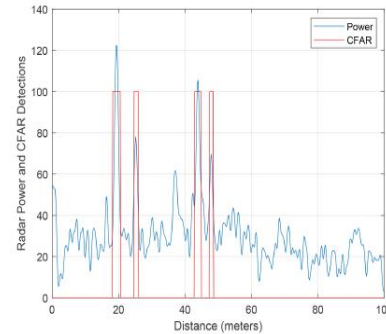


Fig. 9: Radar power spectrum for Navtech image captured at the location shown in Fig. 6. A result from CFAR detection (guard cells=10, training cells=300 false alarm rate =0.1) is also shown.

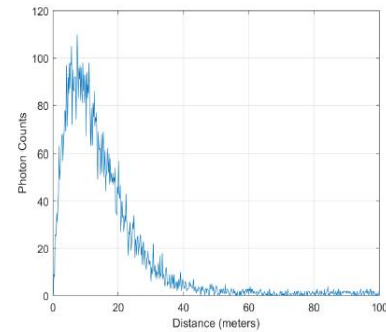


Fig. 10: LiDAR magnitude Spectrum corresponding to Navtech image. This uses real 3D data but generates waveform synthetically. The LiDAR return is at 19.44 meters.

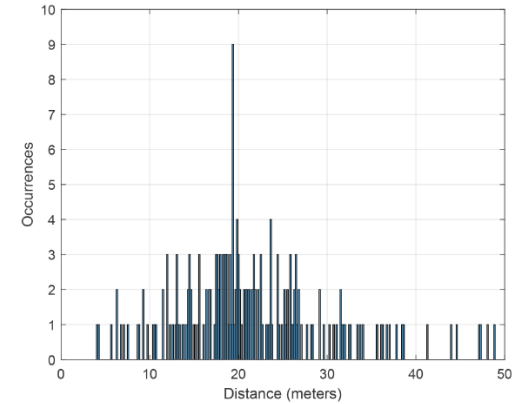


Fig. 12: Histogram of signal detections without radar waveform

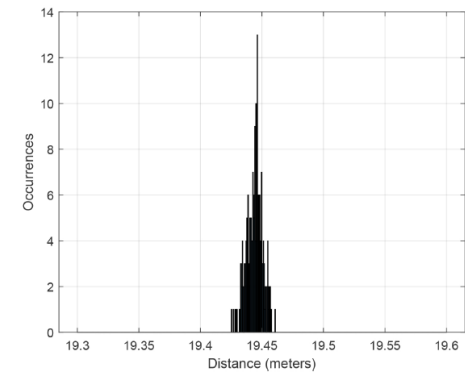


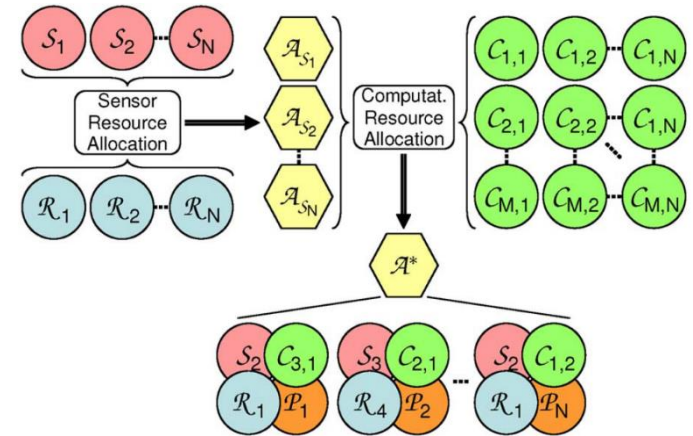
Fig. 11: Histogram of signal detections with radar waveform

Wallace, Mukherjee, Toh, Ahrabian “Combining automotive radar and LiDAR for surface detection in adverse conditions”, IET Proceedings on Radar, Sonar & Navigation. 2021; 1-11

Attentive Resource Allocation

Sometimes, you have to be attentive
(as are humans when they drive cars)

We have used an attentive multi-objective strategy based on utility theory to select computational resource to selected sensors and regions to perform processes (recognise, estimate TTC etc.)

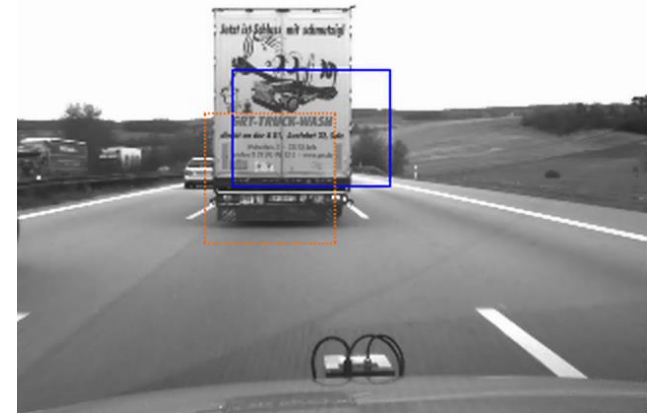


The allocation strategy



Fig. 15. TTC values for the example road traffic scene. The minimum TTC inside each region is used as t_{TCC} and drawn in HSV color space.

Example: Priority given to time to collision



Example: Attentive regional processing

Summary: Scratching the Surface

- We have been looking at the use of surface data for autonomous vehicle and driver assistance, primarily using sensor suites, to take advantage of complementary strengths and weaknesses of Radar, LiDAR and Video data
- Highlights
 1. Collected and distributed a labelled (200,000 actors) radar, LiDAR and stereo dataset in all weathers
 2. Simultaneous location and mapping using our 'wild' 79GHz radar data
 3. Studies of object recognition using 300GHz (Lab) and 79GHz radar data
 4. Driver behaviour prediction using benchmark birds-eye video and our wild radar data
 5. Dense depth map creation using concurrent sparse LiDAR and video data
 6. Full waveform LiDAR analysis for seeing through obscuring media
 - Fusion with 79GHz radar image data
 7. Efficient, resource-saving processing using random laser projection, parallel block processing and reduced precision (GPU and FPGA)
 8. An attentive, multi-sensor processing architecture using utility theory and multi-objective optimisation
- Finally, the devil is in the detail. I appreciate this talk has been light on detail but all this work has been published or is under review.

Are we making slow progress?

THEN (1993-97, first photon counting imaging LiDAR)

Wallace, Massa, Buller and Walker, "Laser ranging using time correlated single photon counting", UK Patent GB 2 306 825 A, filed 18.10.95, first paper published 1997.

AND NOW (2021)

[How Multi-Beam Flash Lidar Works | Ouster](#)

“I’m excited to announce that Ouster has been granted foundational patents for our multi-beam flash [lidar technology](#)” (850nm)

“Ouster is the first company to commercialize a high performance SPAD (single photon avalanche detector) and VCSEL (vertical cavity surface emitting laser) approach.”

“The second chip in our flash lidar is our custom designed CMOS detector ASIC that incorporates an advanced single photon avalanche diode (SPAD) array.”