nodar
Re-envision safety.

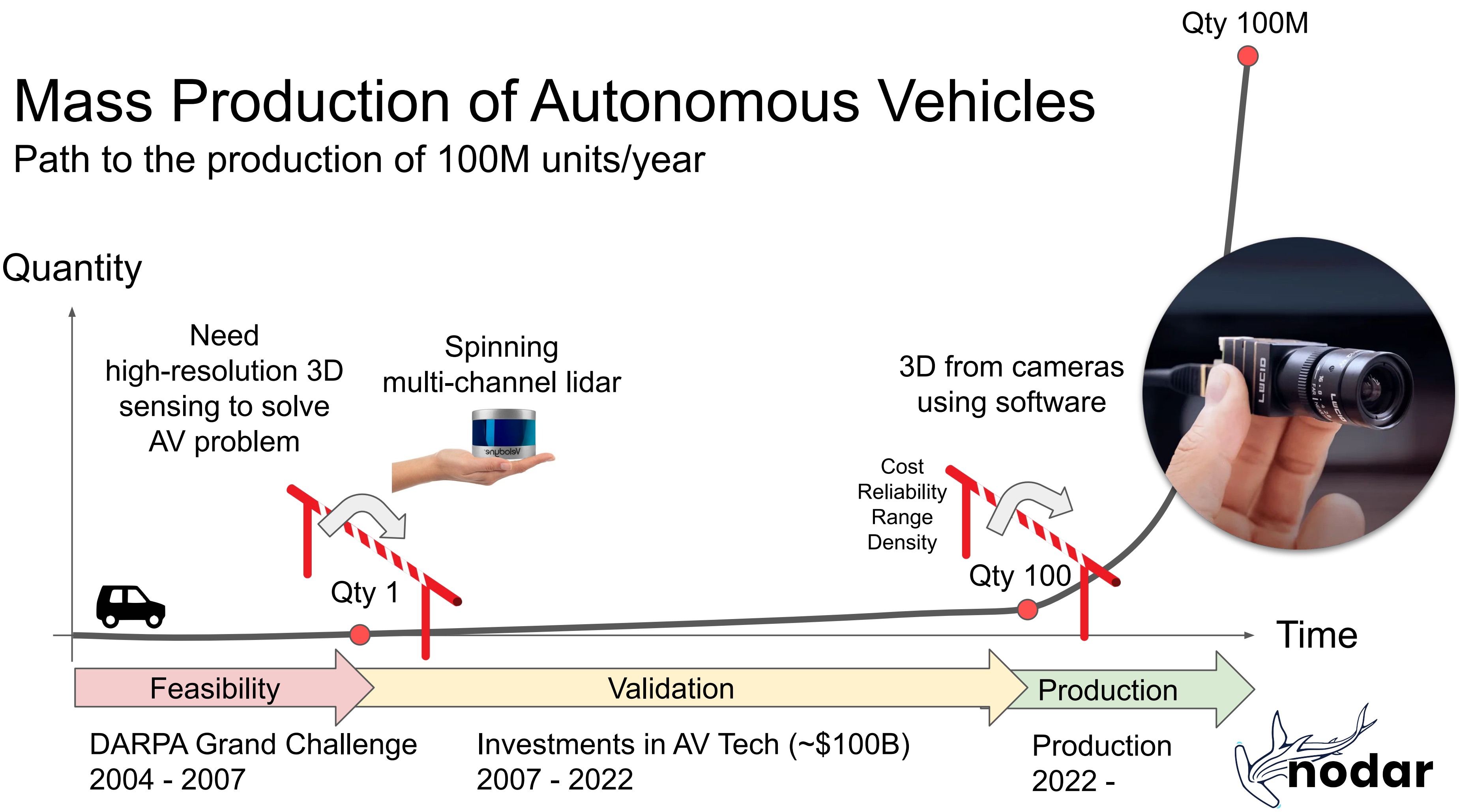
NODAR 3D Vision System

Enabling Mass Production of Autonomous Vehicles

Hot Chips 34
August 21-23, 2022

Mass Production of Autonomous Vehicles

Path to the production of 100M units/year



Wide-Baseline Stereo Vision Camera

Exquisitely dense and accurate point clouds to 1000+ meters

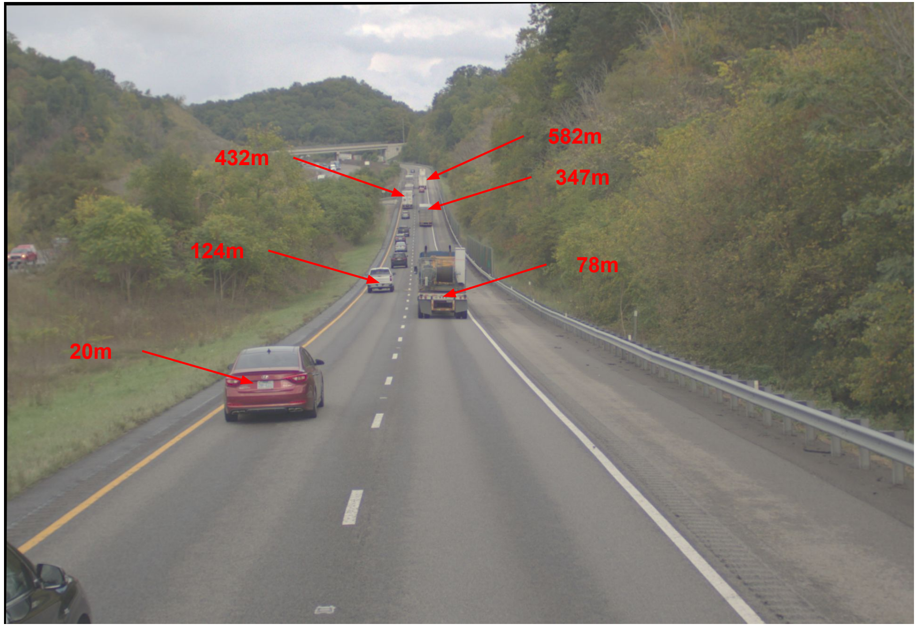
nodar

Long Range Data Collection

- Captures bridge crossing and reconstructs accurately as ego truck passes under bridge that casts a strong shadow on the road

- Captures repetitive patterns of the road railing barriers on right hand side

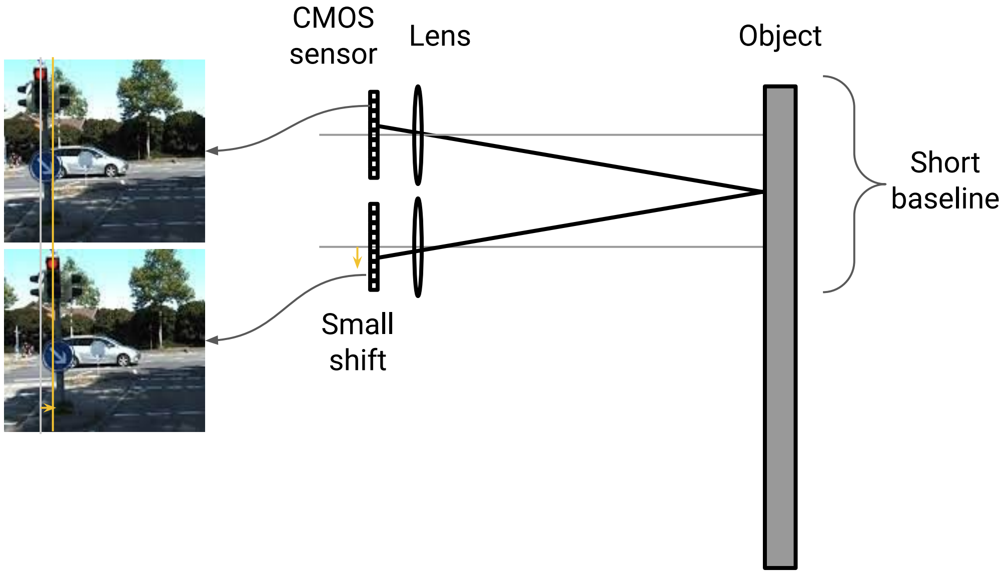
- Captures vehicles from near (12.2 m) to far (1.5 km) range



nodar

Stereo Vision Principle

Wider baseline gives longer range

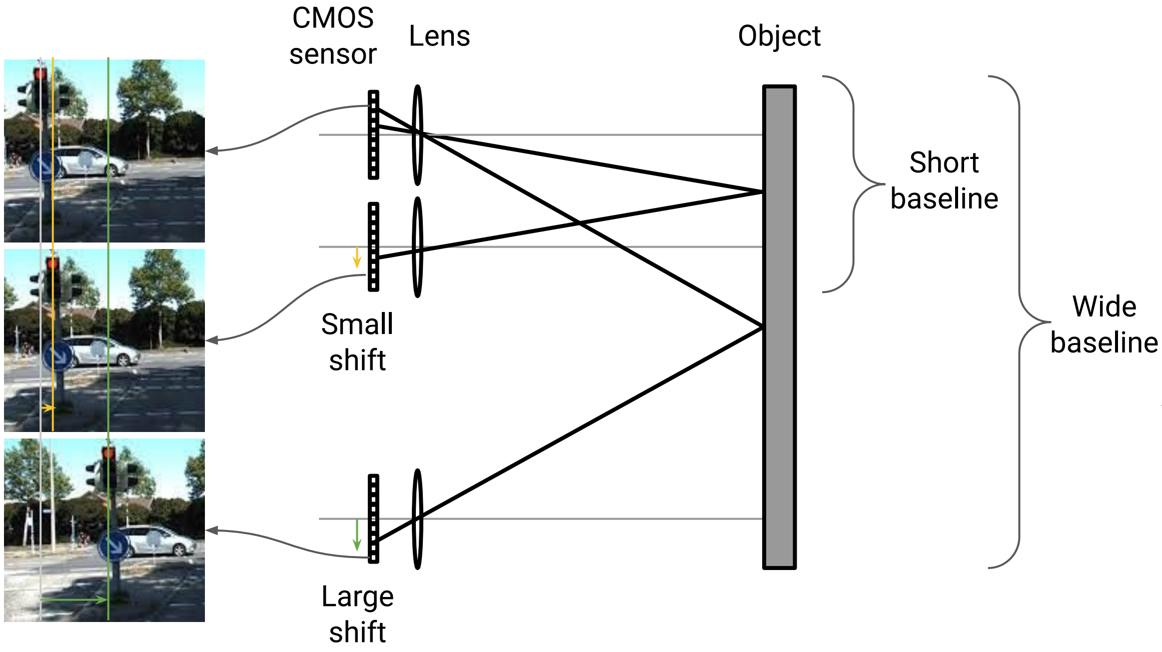


Short baseline stereo vision has trouble discerning the shift of the image at long ranges

nodar

Stereo Vision Principle

Wider baseline gives longer range



Short baseline stereo vision has trouble discerning the shift of the image at long ranges

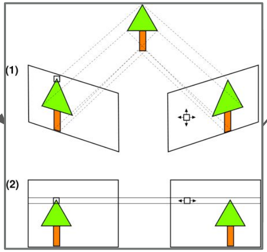
Wide baseline gives sensor access to longer ranges ***but*** need to solve calibration problem for stereo cameras mounted on vehicles where maintaining 0.01° optical alignment is virtually impossible with shock and vibration

nodar

Stereo Vision Principle

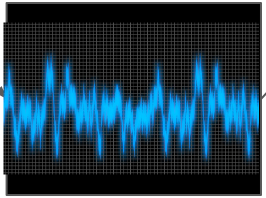
NODAR solves decade-old online calibration problem

You can't ship an engineer with a product.



So researcher have been working on online calibration using natural scenes for the last 30+ years.

But published algorithms did not work on natural scenes



or compute fast enough to correct the camera parameters within the timescale of the road and engine vibrations

or produce the alignment accuracy needed to see 1000+ meters

until NODAR's Hammerhead Vision System

Stereo Vision Capabilities

Image from Ford Open Dataset

Previous Generation

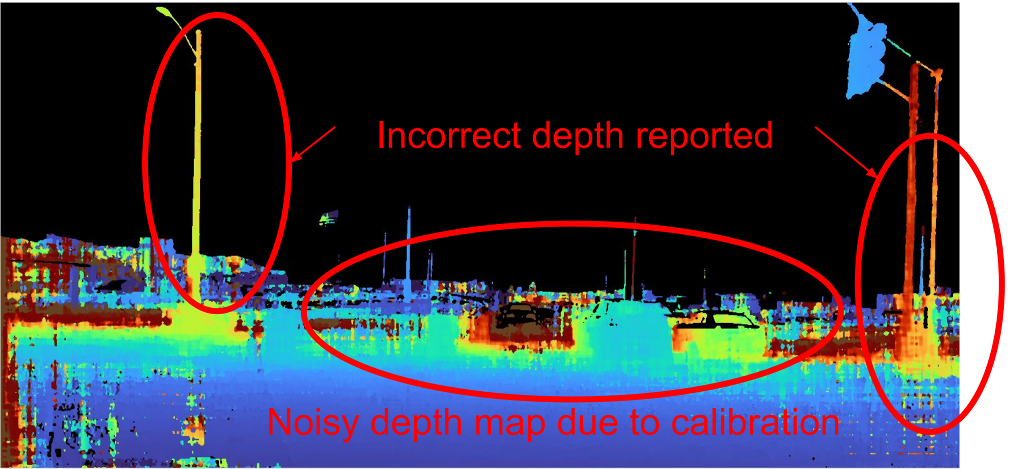
Short Baseline and Static Calibration

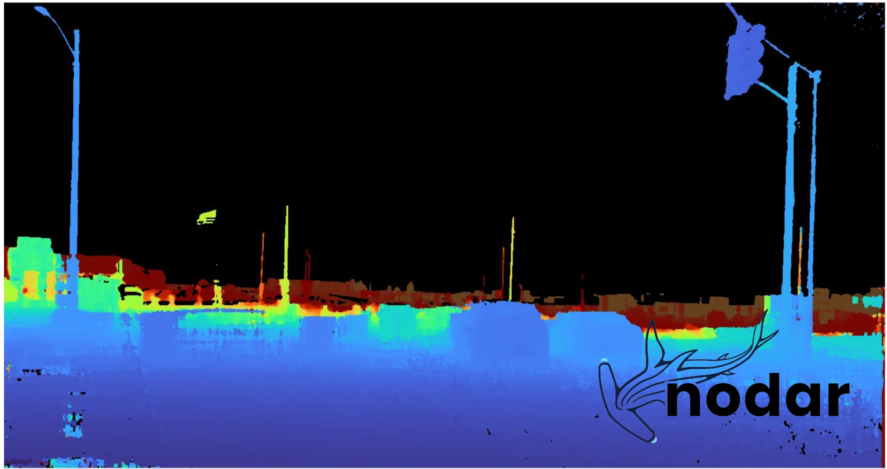
- Poor long-range 3D reconstruction
- Poor minimum range
- Poor vibration/shock tolerance

Next Generation

Wide Baseline and Online Calibration

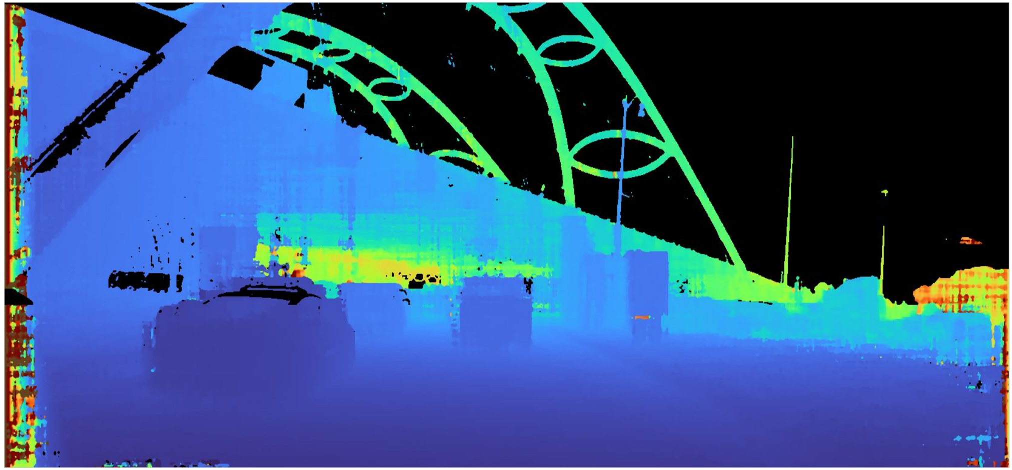
- Lidar-like+ 3D point cloud reconstruction
- Excellent minimum range
- Excellent vibration/shock tolerance





Stereo Vision Capabilities - Bridge example

Left frame

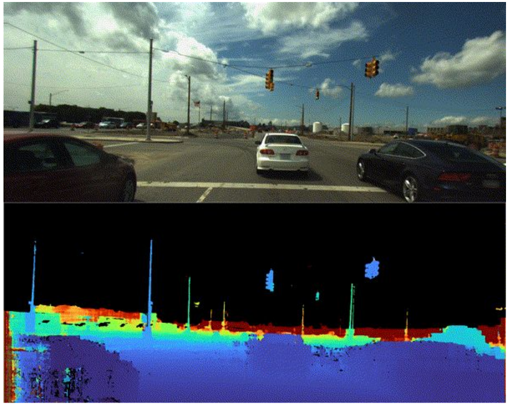


Depthmap with NODAR auto calibration software

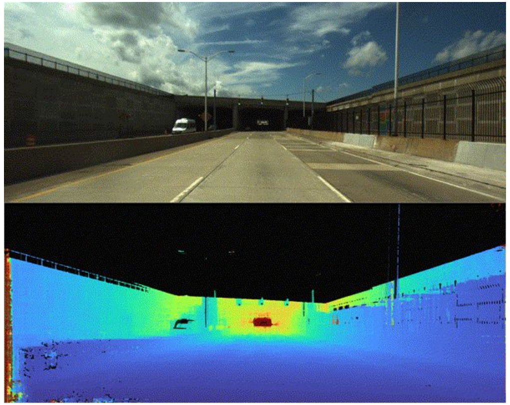
Depth map from Ford rectification

nodar

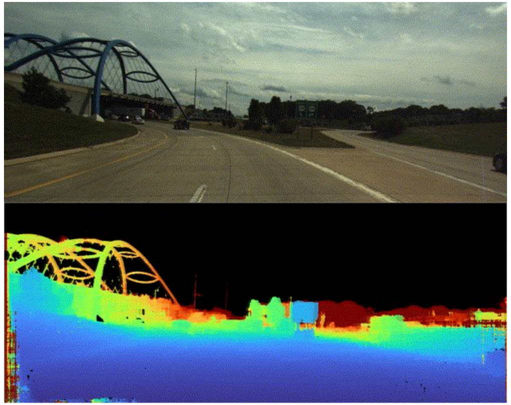
Robust Stereo Vision for Vehicles



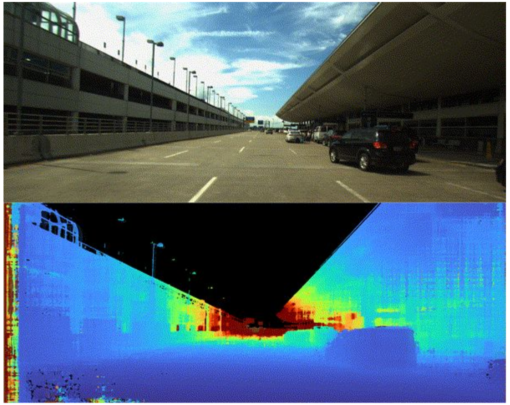
Case 1: Construction site



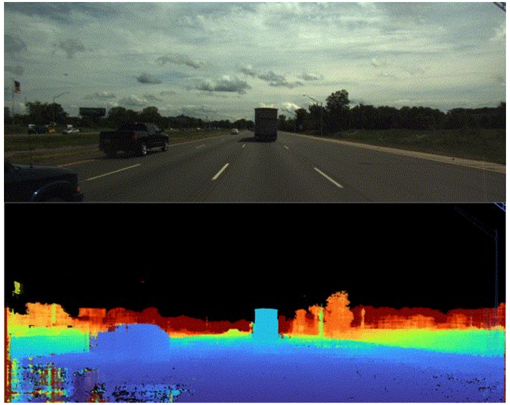
Case 2: Tunnel



Case 3: Girder bridge



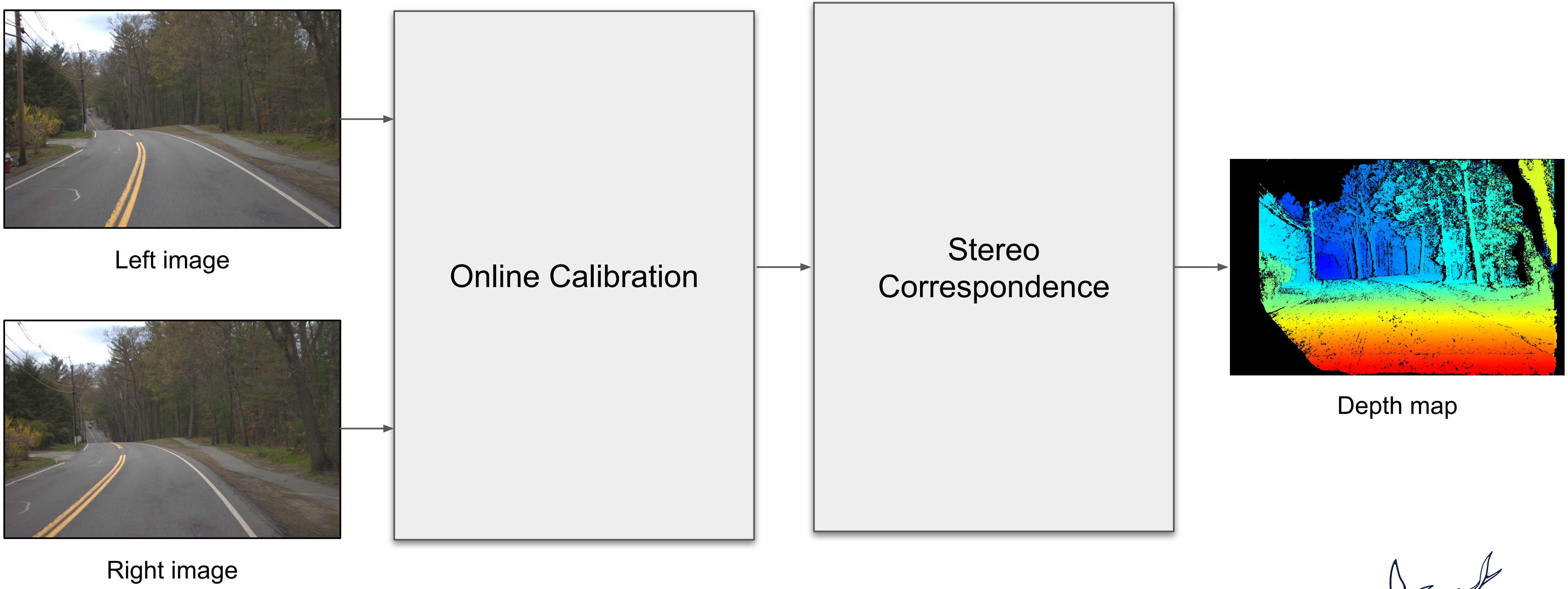
Case 4: Airport



Case 5: Overcast sky

Ford AV open dataset

Processing block diagram



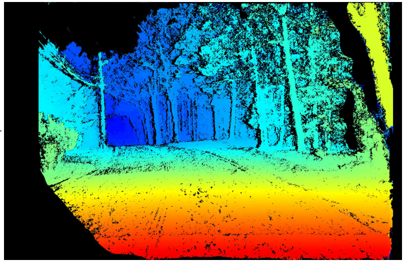
Processing block diagram

Left image

Right image

Online Calibration

Stereo Correspondence



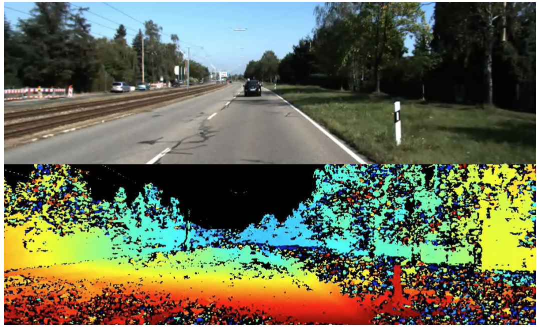
Depth map

Autocalibration Technology

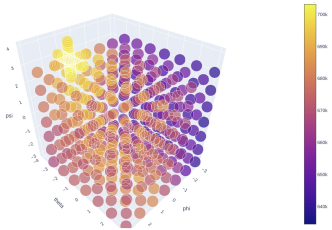
NODAR's patented calibration tech enables automotive applications with significant shock and vibration

Keypoint Matching Approach

Fails when descriptors are similar (windows in urban environments and active stereo illumination)

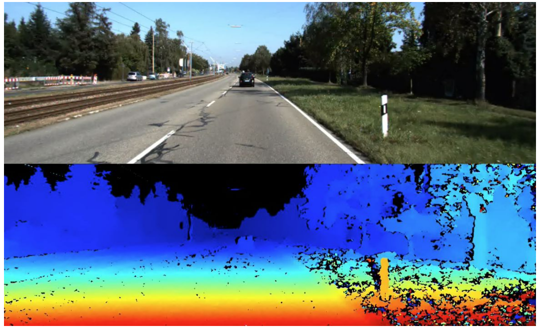


Industry Standard



NODAR Cost Function Approach

Robust under large range of scenes, computed efficiently, and no assumption of flat road surface

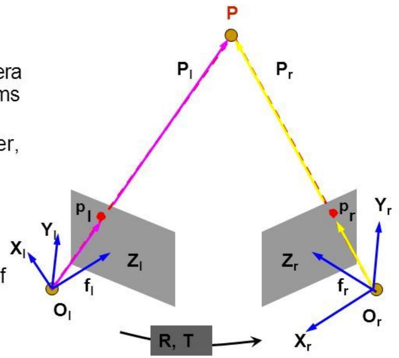


NODAR

Calibration is an optimization problem

Rectification requires 6 extrinsic and 18+ intrinsic camera parameters. NODAR efficiently, quickly, and accurately searches camera parameters to support off-road environments with high levels of shock and vibration, which is the key innovation for supporting long-baseline stereo vision in vehicles.

- Intrinsic Parameters
  - Characterize the transformation from camera to pixel coordinate systems of each camera
  - Focal length, image center, aspect ratio
- Extrinsic parameters
  - Describe the relative position and orientation of the two cameras
  - Rotation matrix R and translation vector T



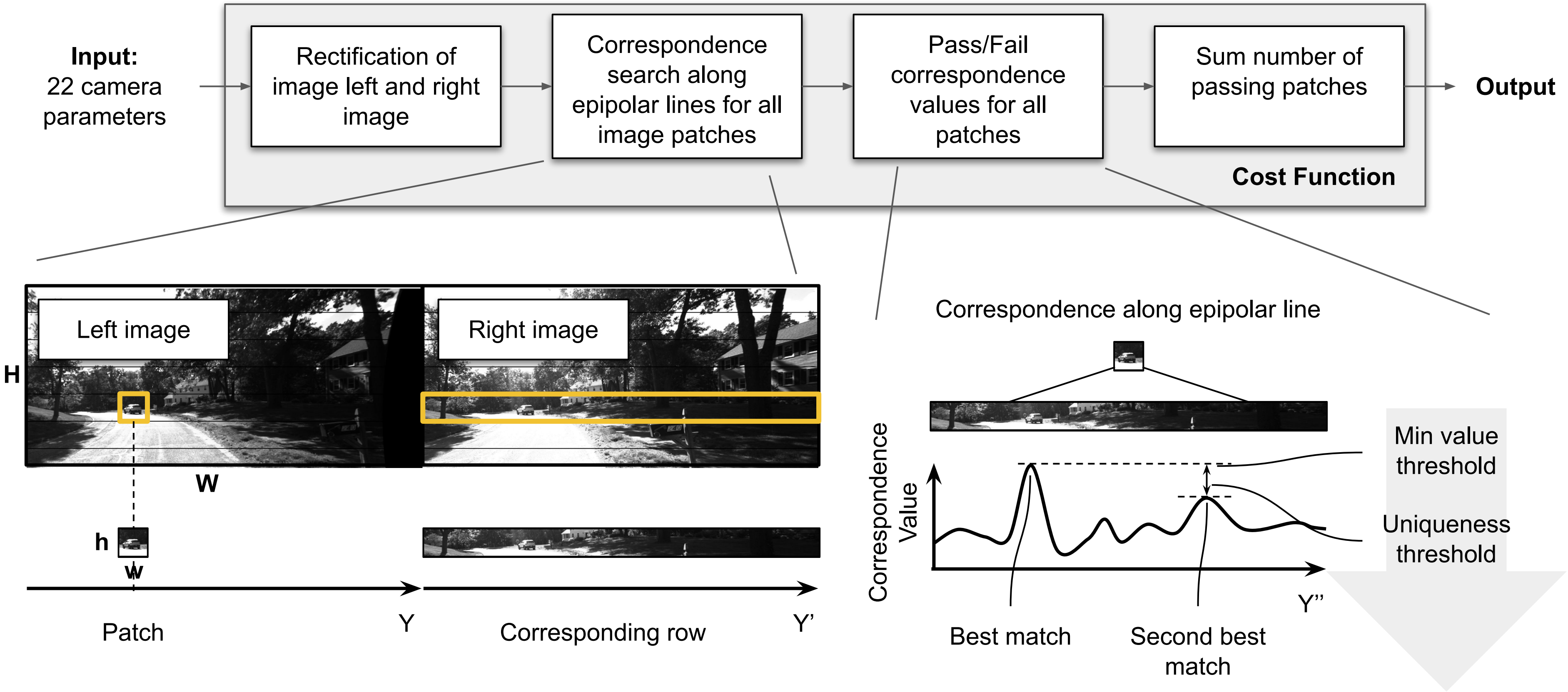
1. Roll (°)
2. Pitch (°)
3. Yaw (°)
4. Camera location x (m)
5. Camera location y (m)
6. Camera location z (m)

Left Camera
1. Focal length x
2. Focal length y
3. Principal point x
4. Principal point y
5. Lens distortion, radial, k1
6. Lens distortion, radial, k2
7. Lens distortion, radial, k3
8. Lens distortion, tangential, p1
9. Lens distortion, tangential, p2

Right Camera
1. Focal length x
2. Focal length y
3. Principal point x
4. Principal point y
5. Lens distortion, radial, k1
6. Lens distortion, radial, k2
7. Lens distortion, radial, k3
8. Lens distortion, tangential, p1
9. Lens distortion, tangential, p2

24-dimensional optimization problem
~100 elements per dimension
$100^{24} = 10^{48}$ search space
Assuming 1 ns per point
$\rightarrow 3 \times 10^{31}$ years $\gg$ Age of the universe (10^{10} years)
A challenging problem!

Definition of Cost Function (Highly Parallelizable)



H*W*2hw*D Ops/cost function evaluation:

H = 1860, W = 2880, h = 5, w = 5, D = 256 → 68 Gops/cost function evaluation

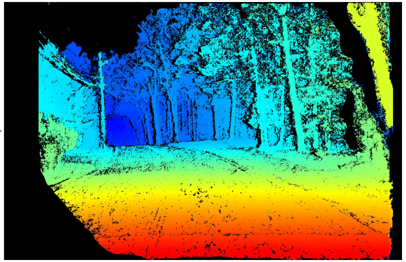
Processing block diagram


Left image


Right image

Online Calibration

Stereo Correspondence


Depth map

nodar

Stereo Correspondence

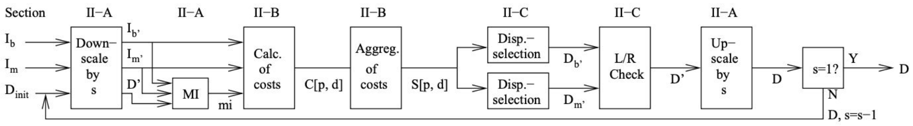
Match corresponding pixels in left and right images

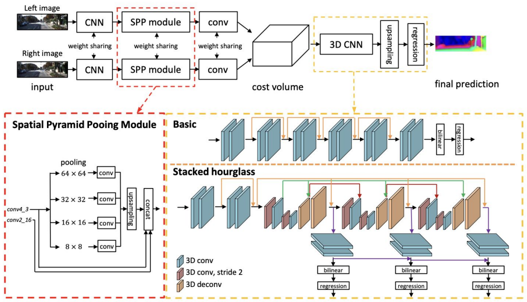
Signal Processing Algorithms

- 1D search (along epipolar lines)
- Faster
- Does not hallucinate
- Generalizable
- Example: Semi-Global Block Matching, 5MP image, **127G ops/frame**

Deep Learning Algorithms

- 2D search (convolutions)
- Slower
- Could hallucinate
- Not generalizable
- Example: PSMNet, 5 MP image, **9604G ops/frame**





Optimal solution depends on application and compute resources

nodar

Power vs. Applications for Long-Range Stereo Cameras

Decreasing Power Consumption Unlocks More Markets

Drones/ UAVs
<5 W*
VGA, 10-20 FPS, 400 meters

Last-Mile Delivery
<20 W*
2 MP, 5-30 FPS, 50 meters

Consumer Vehicles
<50 W*
5-8 MP, 5-30 FPS, 200+ meters

Robo-taxis/ Shuttles
<100 W*
5-8 MP, 5-30 FPS, 200+ meters

Commercial Vehicles
<300 W*
5-8 MP, 5-30 FPS, 400+ meters

Maximum Power Consumption (Watts)

Hammerhead Vision System Next Year

Hammerhead Vision System Today (using Nvidia HW)

nodar

* Compute power available on these platforms is roughly proportional to the vehicle mass (because kinetic energy is ½ mv²)

Limitations in existing silicon platforms and the future

- The online calibration algorithms currently run on general purpose GPUs, which consumes too much power for smaller platforms (such as drones)
- To make this a “solved” problem across all autonomous platforms would require an ASIC for
  - Rectification with ability to quickly modify the look-up tables
  - Correspondence-computation accelerator

nodar

Summary

- High-resolution 3D sensing is necessary for autonomous vehicles
- Wide-baseline stereo vision provides a commercially viable path to mass production
- Next generation stereo vision has two innovations:
    - Online calibration of independent camera modules on platforms with shock and vibration
    - More accurate stereo correspondence algorithms
- Likely to see adoption of independently-mounted stereo vision cameras in other markets such as robotics, which has similar economics and platform costs as passenger vehicles

nodar