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

Quantity

- Need high-resolution 3D sensing to solve AV problem
- Spinning multi-channel lidar
- 3D from cameras using software

Time

- DARPA Grand Challenge 2004 - 2007
- Investments in AV Tech (~$100B) 2007 - 2022
- Production 2022 -
Wide-Baseline Stereo Vision Camera
Exquisitely dense and accurate point clouds to 1000+ meters
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
Stereo Vision Principle
Wider baseline gives longer range

Short baseline stereo vision has trouble discerning the shift of the image at long ranges
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.
Stereo Vision Principle

NODAR solves decade-old online calibration problem

You can’t ship an engineer with a product.

So researchers 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

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

Image from Ford Open Dataset

Incorrect depth reported
Noisy depth map due to calibration
Stereo Vision Capabilities - Bridge example

Depthmap with NODAR auto calibration software

Dangerous situation: says that bridge is farther away than it really is (and that there is no space to drive under it)
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

Left image

Online Calibration

Stereo Correspondence

Depth map

Right image
Processing block diagram

Left image

Online Calibration

Stereo Correspondence

Depth map

Right image
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)

NODAR Cost Function Approach
Robust under large range of scenes, computed efficiently, and no assumption of flat road surface

Industry Standard

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

---

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)

Input:
- 22 camera parameters

Rectification of image left and right image

Correspondence search along epipolar lines for all image patches

Pass/Fail correspondence values for all patches

Sum number of passing patches

Output

Cost Function

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

Online Calibration

Stereo Correspondence

Right image

Depth map
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
Power vs. Applications for Long-Range Stereo Cameras
Decreasing Power Consumption Unlocks More Markets

<table>
<thead>
<tr>
<th>Application</th>
<th>Maximum Power Consumption (Watts)</th>
<th>Resolution</th>
<th>FPS</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drones/UAVs</td>
<td>&lt;5 W*</td>
<td>VGA</td>
<td>10-20 FPS, 400 meters</td>
<td></td>
</tr>
<tr>
<td>Last-Mile Delivery</td>
<td>&lt;20 W*</td>
<td>2 MP</td>
<td>5-30 FPS, 50 meters</td>
<td></td>
</tr>
<tr>
<td>Consumer Vehicles</td>
<td>&lt;50 W*</td>
<td>5-8 MP</td>
<td>5-30 FPS, 200+ meters</td>
<td></td>
</tr>
<tr>
<td>Robo-taxis/Shuttles</td>
<td>&lt;100 W*</td>
<td>5-8 MP</td>
<td>5-30 FPS, 200+ meters</td>
<td></td>
</tr>
<tr>
<td>Commercial Vehicles</td>
<td>&lt;300 W*</td>
<td>5-8 MP</td>
<td>5-30 FPS, 400+ meters</td>
<td></td>
</tr>
</tbody>
</table>

* Compute power available on these platforms is roughly proportional to the vehicle mass (because kinetic energy is \( \frac{1}{2} \text{mv}^2 \))
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
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