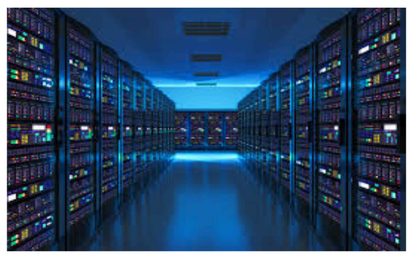
DragonFly+: FPGA-Based Quad-Camera Visual SLAM System for Autonomous Vehicles

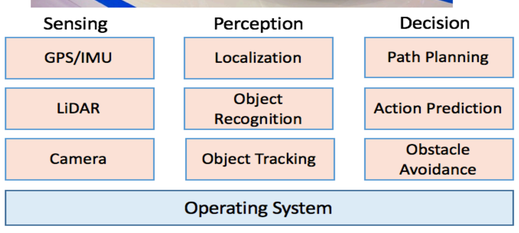
PERCEPTION INSIGHT INTELLIGENCE

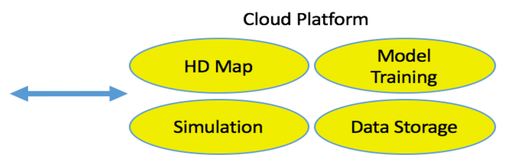
Weikang Fang*†, Yanjun Zhang*, Bo Yu†, Shaoshan Liu†

*Beijing Institute of Technology, Beijing, China †PerceptIn, Shenzhen, China

Overview of autonomous driving technologies





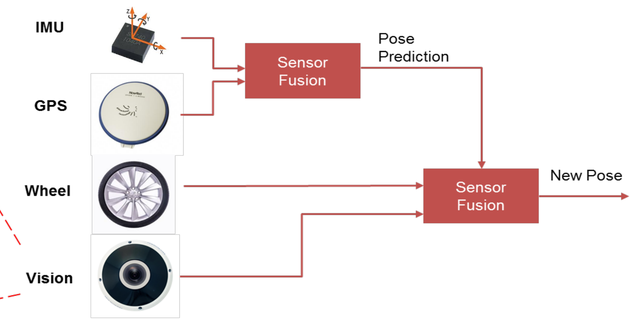


cargo carrying vehicle, passenger vehicle and four-way synchronized images with stereo 360-degree views

System Design Specifications

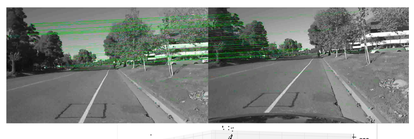
- Modular: independent hardware module for computer-vision-based localization and map generation
- SLAM-Ready: Hardware synchronization of four cameras and IMU
- Low Power: Power consumption < 10 W
- High-Performance: Four-way 720P YUV images with > 30 FPS

Computer Vision for Localization and Mapping

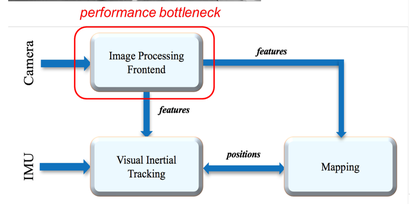


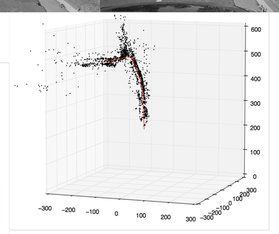
- DragonFly module: four-way synchronized images with stereo 360-degree views
- Embedded with IMU and GPS, interface with wheel
- Online real-time localization and offline high precision map generation

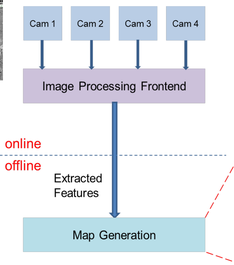
Online Real-Time Localization and Offline High Precision Visual Map Generation

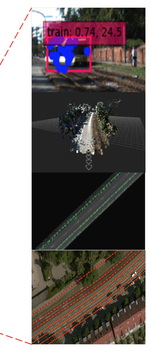


performance bottleneck









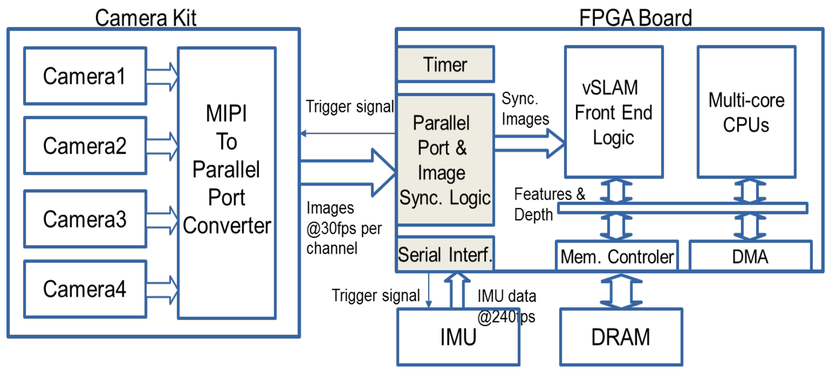
Layer 4: semantic Information

Layer 3: spatial features

Layer 2: ground features

Layer 1: digital map with lane-level annotation

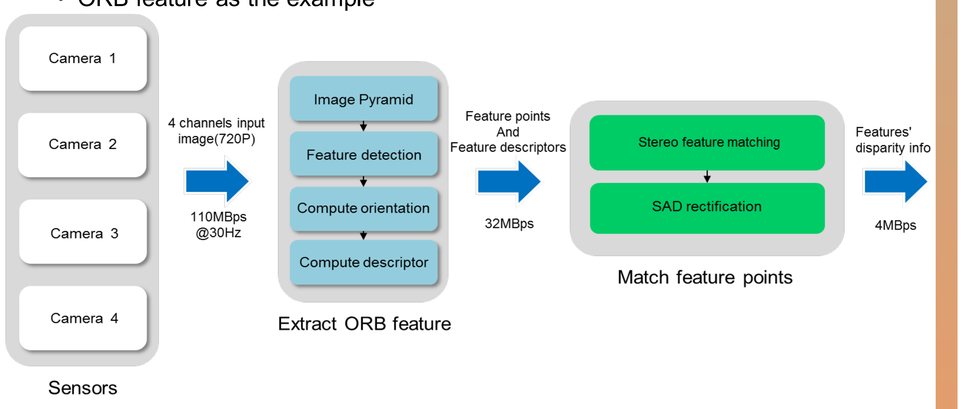
DragonFly+ FPGA Based Visual SLAM System



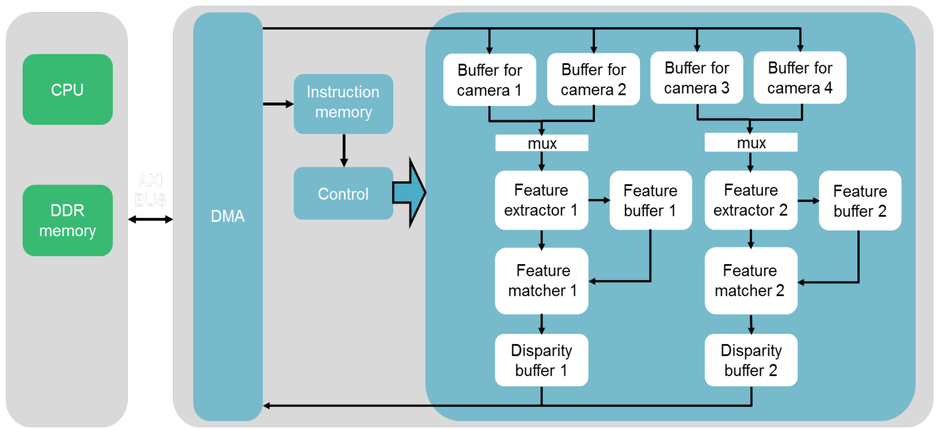
- Camera Interface
  - Feature: **Hardware based images synchronization** and **Direct IO**
  - Four cameras are triggered by FPGA at the same time
  - Captured images are tagged time stamp by camera interface
  - Captured images are sent to vSLAM front end logic directly
- IMU Interface
  - IMU is triggered by FPGA
  - IMU data is tagged time stamp by IMU interface

- vSLAM Front End Logic:
  - Input: camera images
  - Generic function: calculate features and obtain 3D information
  - Implementation: FPGA logic
- Multi-core CPUs:
  - Input: features and 3D information
  - Generic function: multi-view image optimization

- Algorithm overview
  - ORB feature as the example



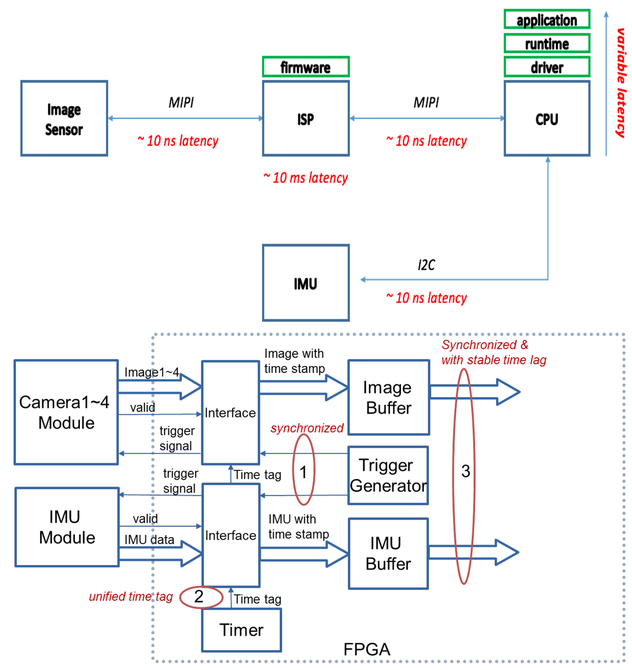
- vSLAM front end hardware architecture
  - Two identical hardware to process two stereo cameras in parallel
  - Two camera channels share one feature extraction hardware to reduce power and hardware



Hardware Synchronization

Problem: software synchronization leads to variable delay among the four images, making it impossible to achieve reliable visual SLAM results

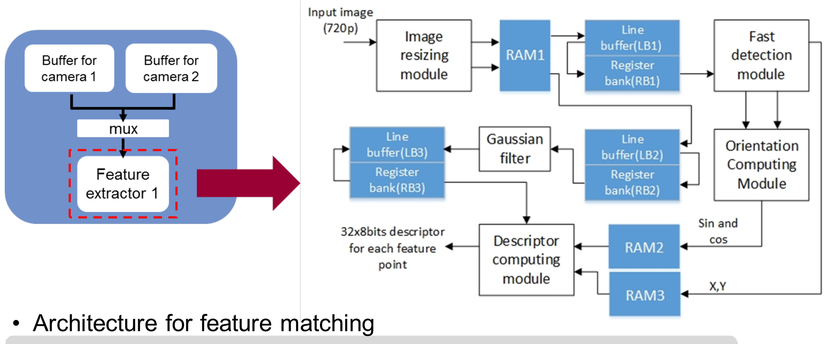
Requires Hardware Synchronization!



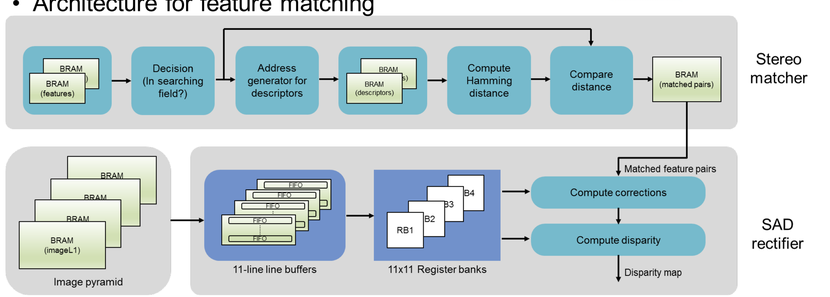
- Hardware(Trigger Generator) generates synchronized trigger signals for camera and IMU
- Input image and IMU are tagged by a unified time tag
- Input image and IMU are synchronized at the interface
- Images and IMUs for vSLAM are with a certain time lag, which is guaranteed by this hardware based triggering and synchronization method

Image Processing Frontend Acceleration

- Architecture for feature extractor



- Architecture for feature matching



Evaluation Results

- **Target platform : Xilinx ZYNQ Ultrascale+ XCZU17EG MPSOC**

Table I : Performance of feature extractor and matcher

	Latency	Frequency	Power
Feature extractor	7.9ms	203MHz	0.52W
Feature matcher	16.1ms	230MHz	0.07W

Table II : Resource consumption of feature extractor and feature matcher

	LUT	FF	BRAM	DSP
Feature extractor	70483	62597	205	32
Feature matcher	42134	11372	68	8
Control logic and buffers	8445	1894	147	10

Table III and IV : DragonFly+ 4-channel vSLAM front end system, compared with software solution

	LUT	FF	BRAM	DSP
Total	423403	846806	796	1590
Used	233679	152832	693	90
Utilization	55%	18%	87%	6%

	Power (W)	Performance (FPS)
DragonFly+	2.31	42
Nvidia TX1	7	9
Intel Core i7	80	15

Conclusions

- PerceptIn's DragonFly car utilizes computer-vision-based sensor fusion to achieve affordable and reliable real-time localization
- Processing four-way 720p synchronized images imposes tremendous stress on DragonFly car's computing system, hence we designed and implemented DragonFly+
- DragonFly+ achieves and exceeds the design goals: modular, SLAM-ready, low-power, and high-performance
  - 3x more power efficient and delivers 5x of computing power compared to Nvidia TX1
  - 34x more power efficient and delivers 3x of computing power compared Intel Core i7