An Energy-Efficient Quad-Camera Visual System for Autonomous Machines on FPGA Platform

Zishen Wan[*1], Yuyang Zhang[*2], Arijit Raychowdhury[1], Bo Yu[3], Yanjun Zhang[2], Shaoshan Liu[3]

[1] Georgia Institute of Technology [2] Beijing Institute of Technology [3] PerceptIn Inc.

([*]Equal Contributions)

Email: zishenwan@gatech.edu

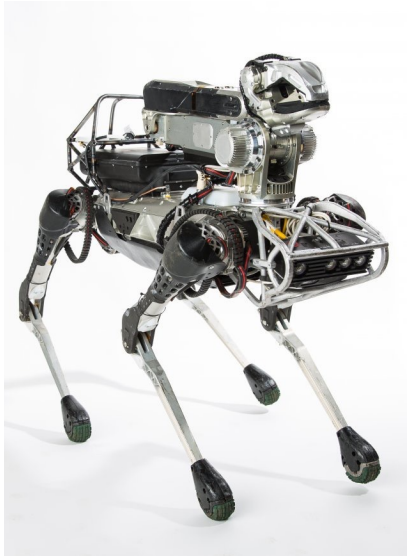
Executive Summary

- **Visual system** is the **compute bottleneck** of autonomous machine systems and a lucrative acceleration target.
- Analyze **algorithm blocks** in ORB-based vision system.
- Present an energy-efficient quad-camera visual **hardware architecture** on **FPGA** platform. Several **optimization techniques** are proposed.
- Our FPGA design achieves up to **5.6x** speedup and **34.6x** energy efficiency improvement.

Executive Summary

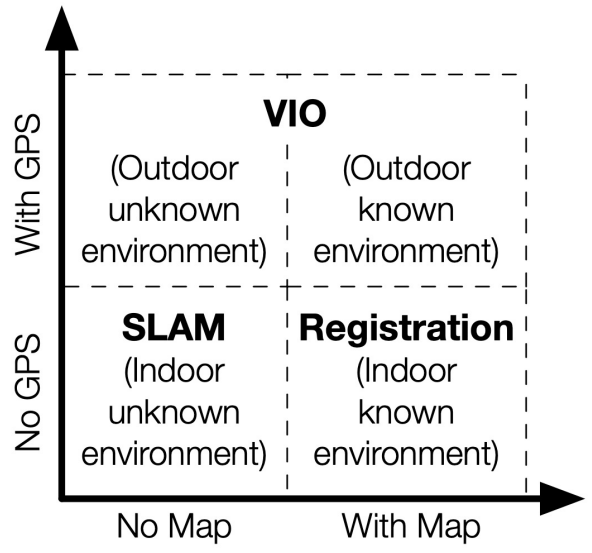
- **Visual system is the compute bottleneck of autonomous machine systems and a lucrative acceleration target.**
- Analyze algorithm blocks in ORB-based vision system.
- Present an energy-efficient quad-camera visual hardware architecture on FPGA platform. Several optimization techniques are proposed.
- Our FPGA design achieves up to 5.6x speedup and 34.6x energy efficiency improvement.

Autonomous Machines



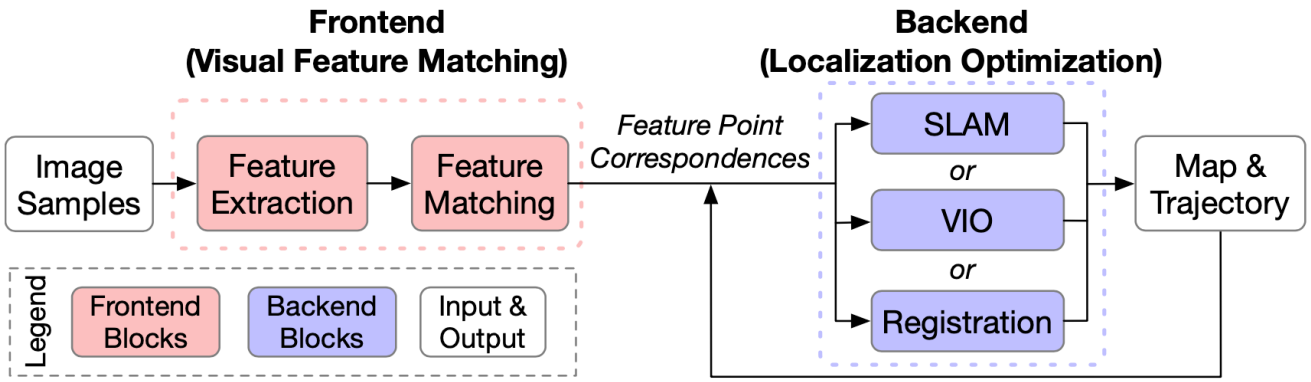
System Compute Time Profiling

- Three localization systems:
  - SLAM (Simultaneous Localization and Mapping)
  - VIO (Visual-Inertial Odometry)
  -



[Gan et al, HPCA 2020]

- Profiling Framework
  - Frontend: visual feature matching
  - Backend: Localization optimization



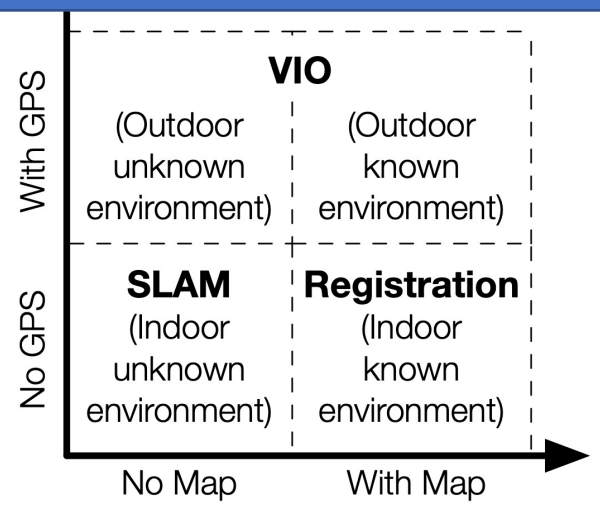
- Profiling Results

	SLAM	VIO	Registration
Frontend	54.8%	86.7%	84.6%
Backend	45.2%	13.3%	15.4%

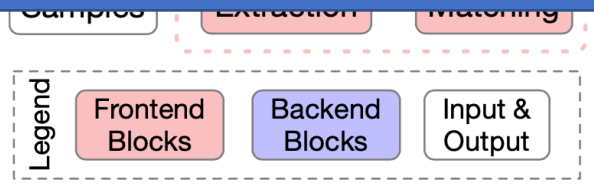
System Compute Time Profiling

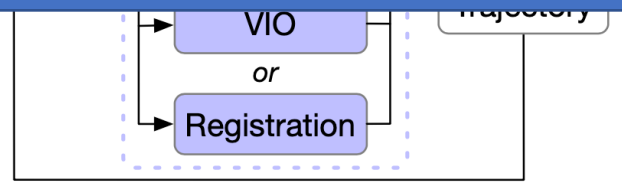
- Three localization systems:
  - SLAM (Simultaneous
- Profiling Framework
  - Frontend: visual feature matching

Visual Frontend is a lucrative acceleration target



[Gan et al, HPCA 2020]





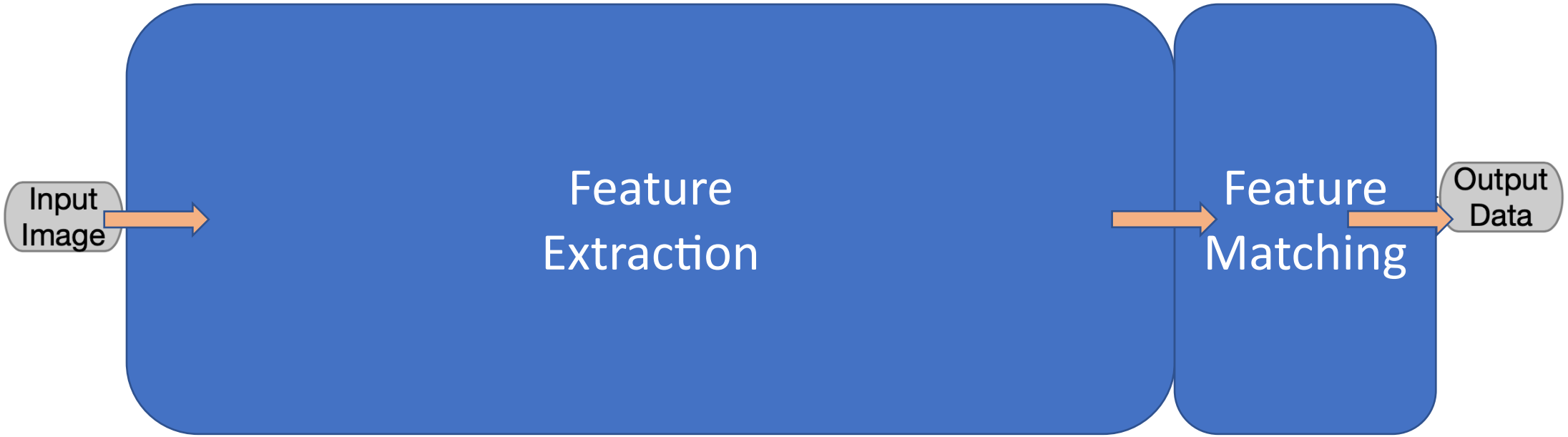
- Profiling Results

	SLAM	VIO	Registration
Frontend	54.8%	86.7%	84.6%
Backend	45.2%	13.3%	15.4%

Executive Summary

- Visual system is the compute bottleneck of autonomous machine systems and a lucrative acceleration target.
- **Analyze algorithm blocks in ORB-based vision system.**
- Present an energy-efficient quad-camera visual hardware architecture on FPGA platform. Several optimization techniques are proposed.
- Our FPGA design achieves up to 5.6x speedup and 34.6x energy efficiency improvement.

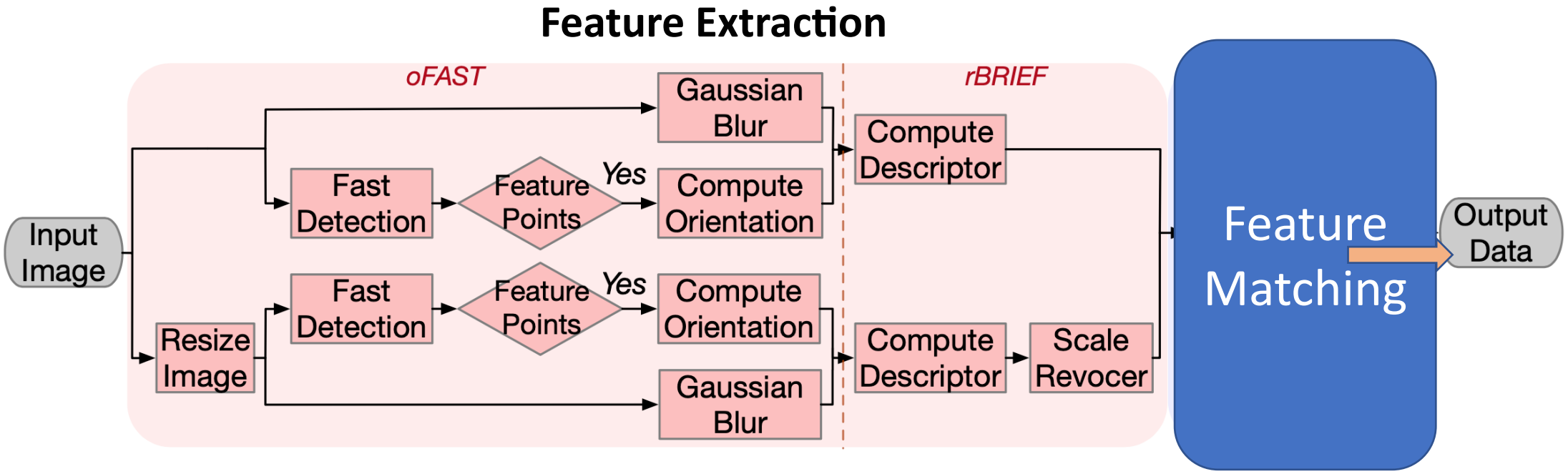
Visual Frontend: Overview



Vision Frontend

- Feature Extraction: extract feature descriptors from input images
- Feature Matching: obtain disparity and depth information

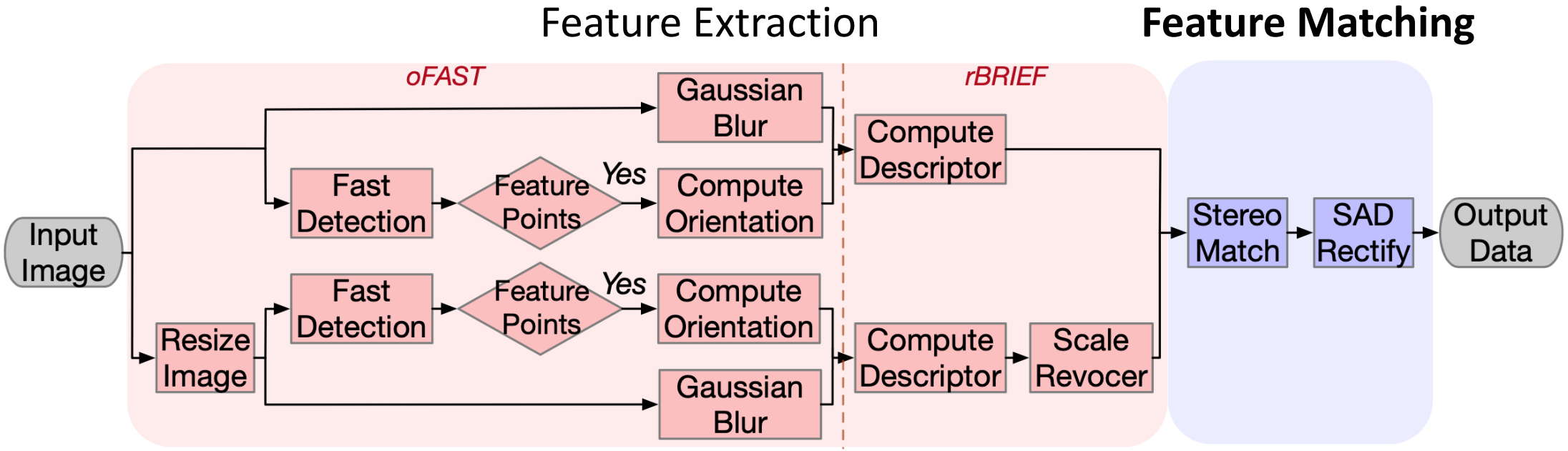
Visual Frontend: Feature Extraction



Feature Extraction (ORB)

- oFAST (Feature from Accelerated Segment Test): feature detection
- BRIEF (Binary Robust Independent Elementary Features): feature description

Visual Frontend: Feature Matching



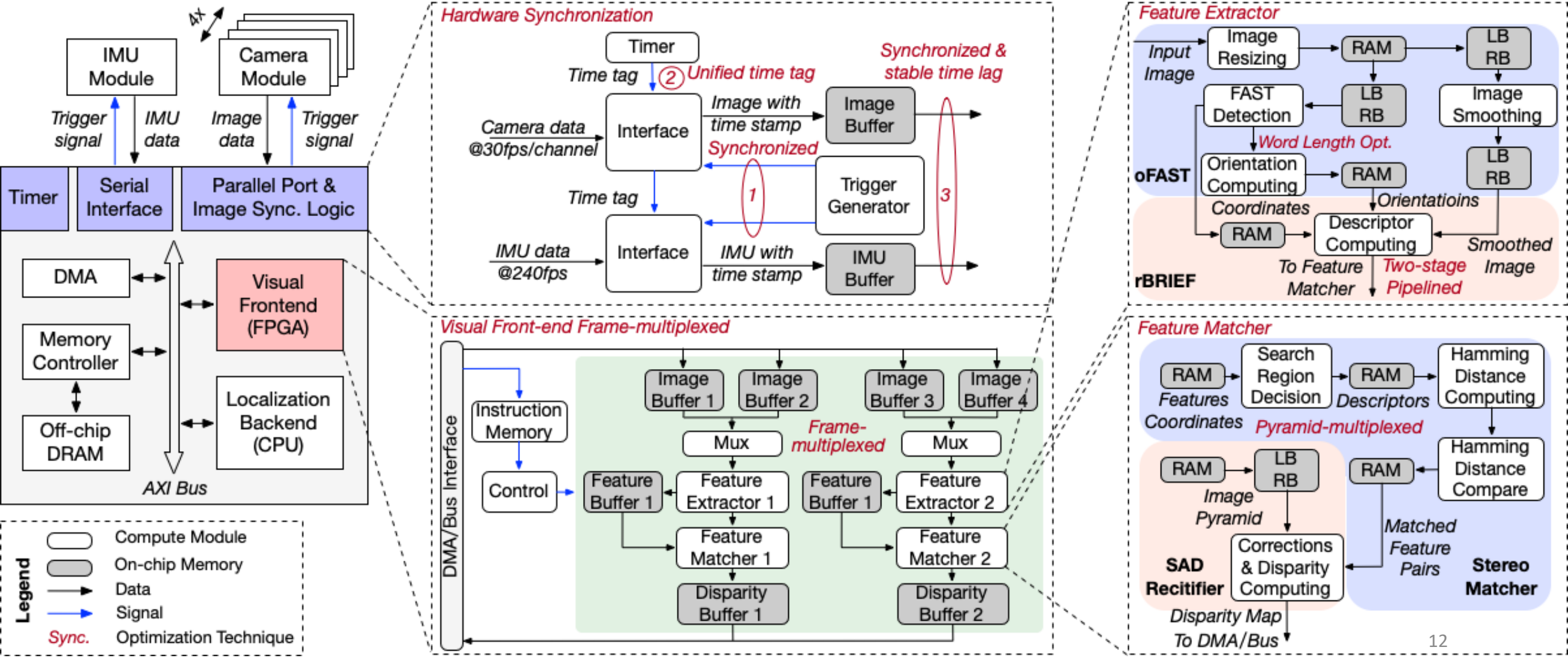
Feature Matching

- Stereo Matching: matches feature points in a stereo image pair
- Rectification: SAD (Sum of Absolute Differences) rectification -> depth information

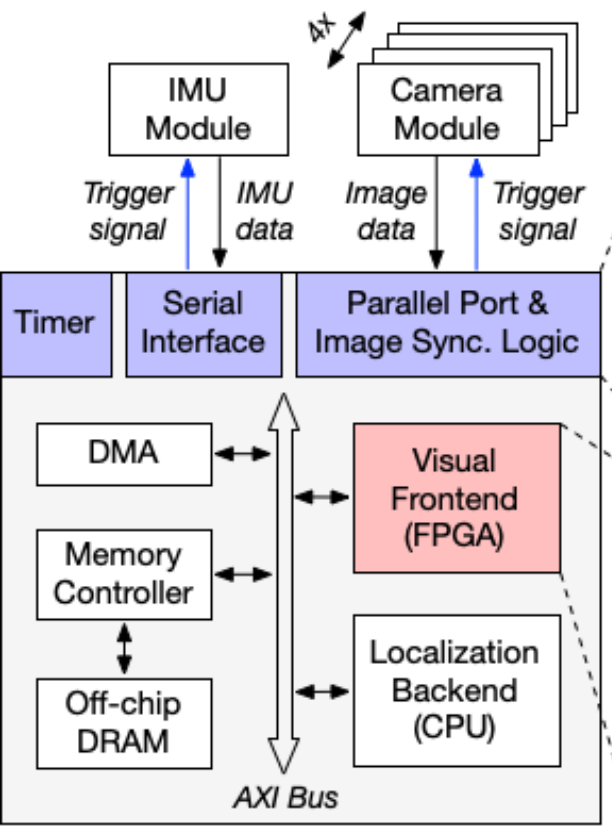
Executive Summary

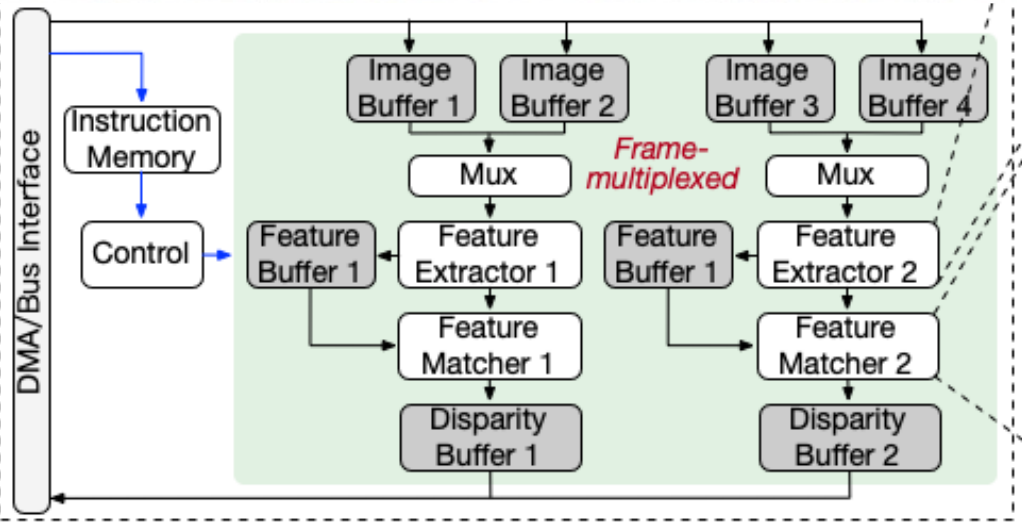
- Visual system is the compute bottleneck of autonomous machine systems and a lucrative acceleration target.
- Analyze algorithm blocks in ORB-based vision system.
- **Present an energy-efficient quad-camera visual hardware architecture on FPGA platform. Several optimization techniques are proposed.**
- Our FPGA design achieves up to 5.6x speedup and 34.6x energy efficiency improvement.

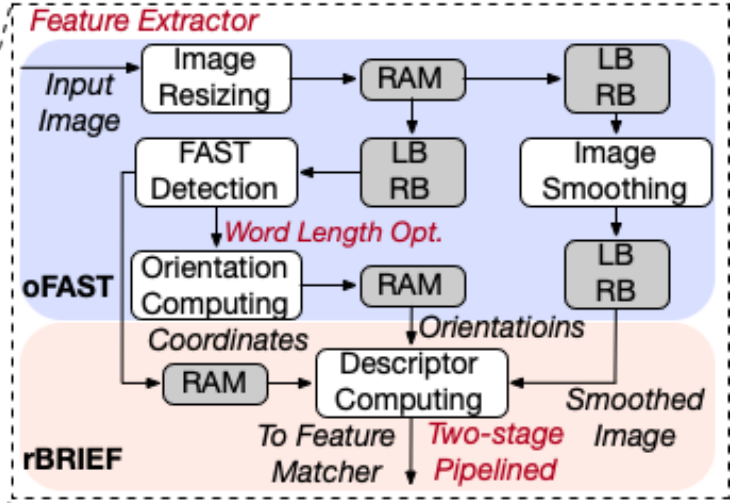
Hardware Architecture: Overview

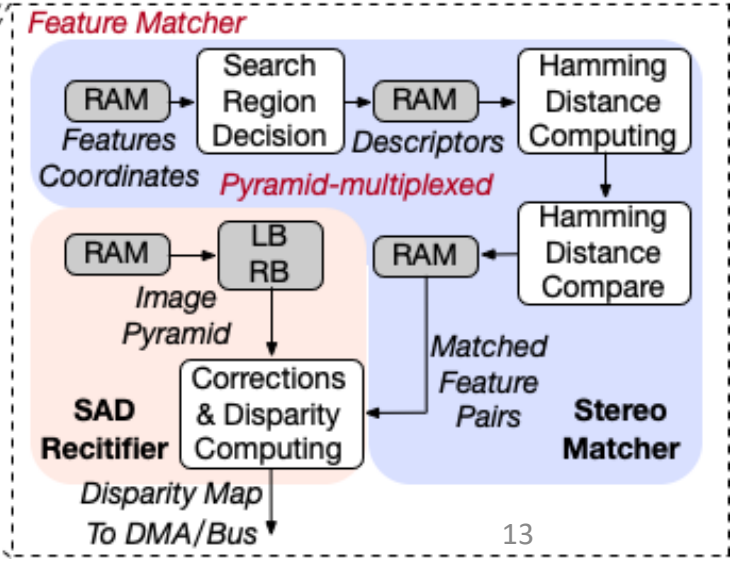


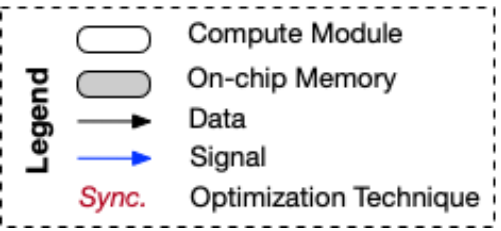
Hardware Architecture: Overview



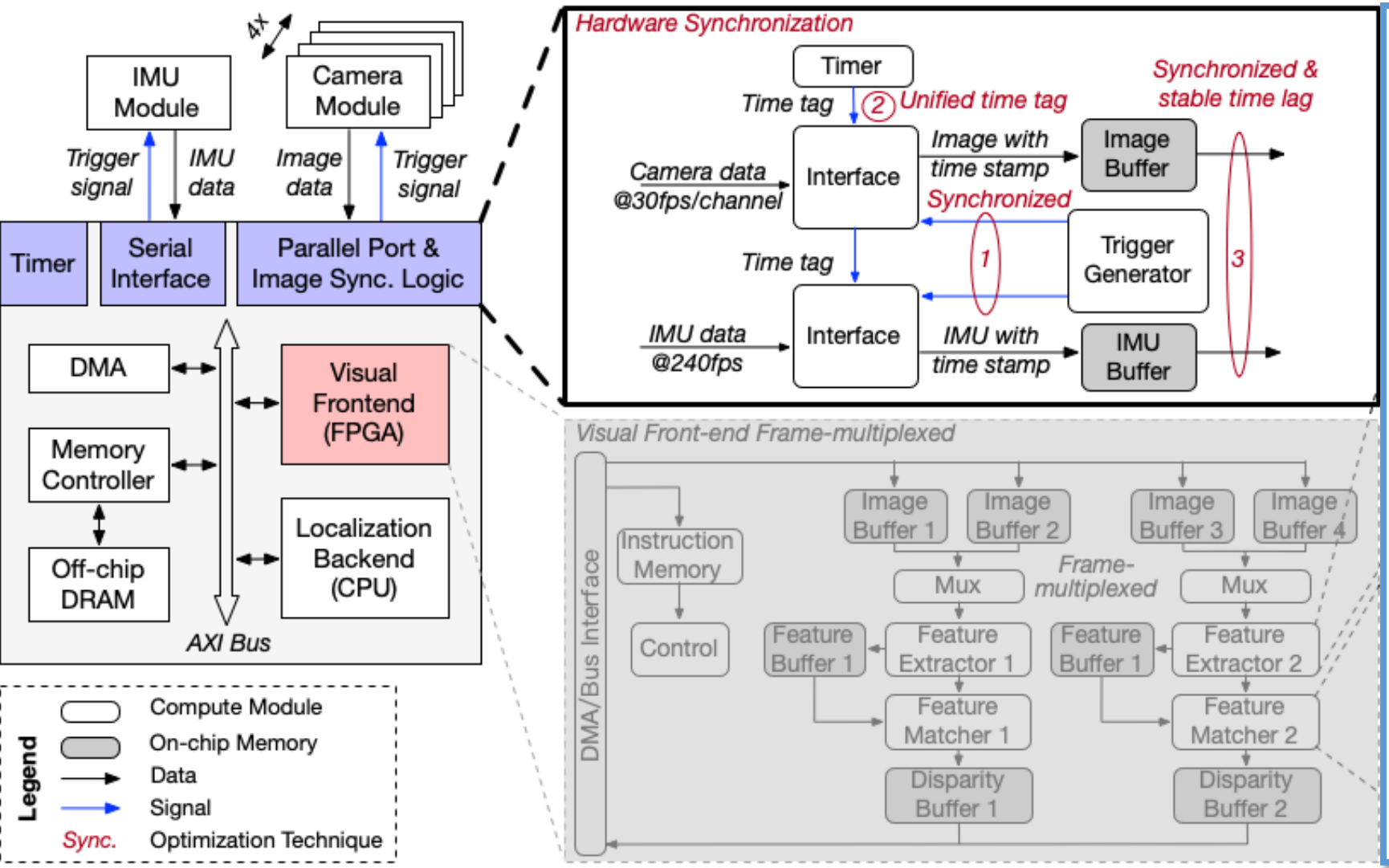






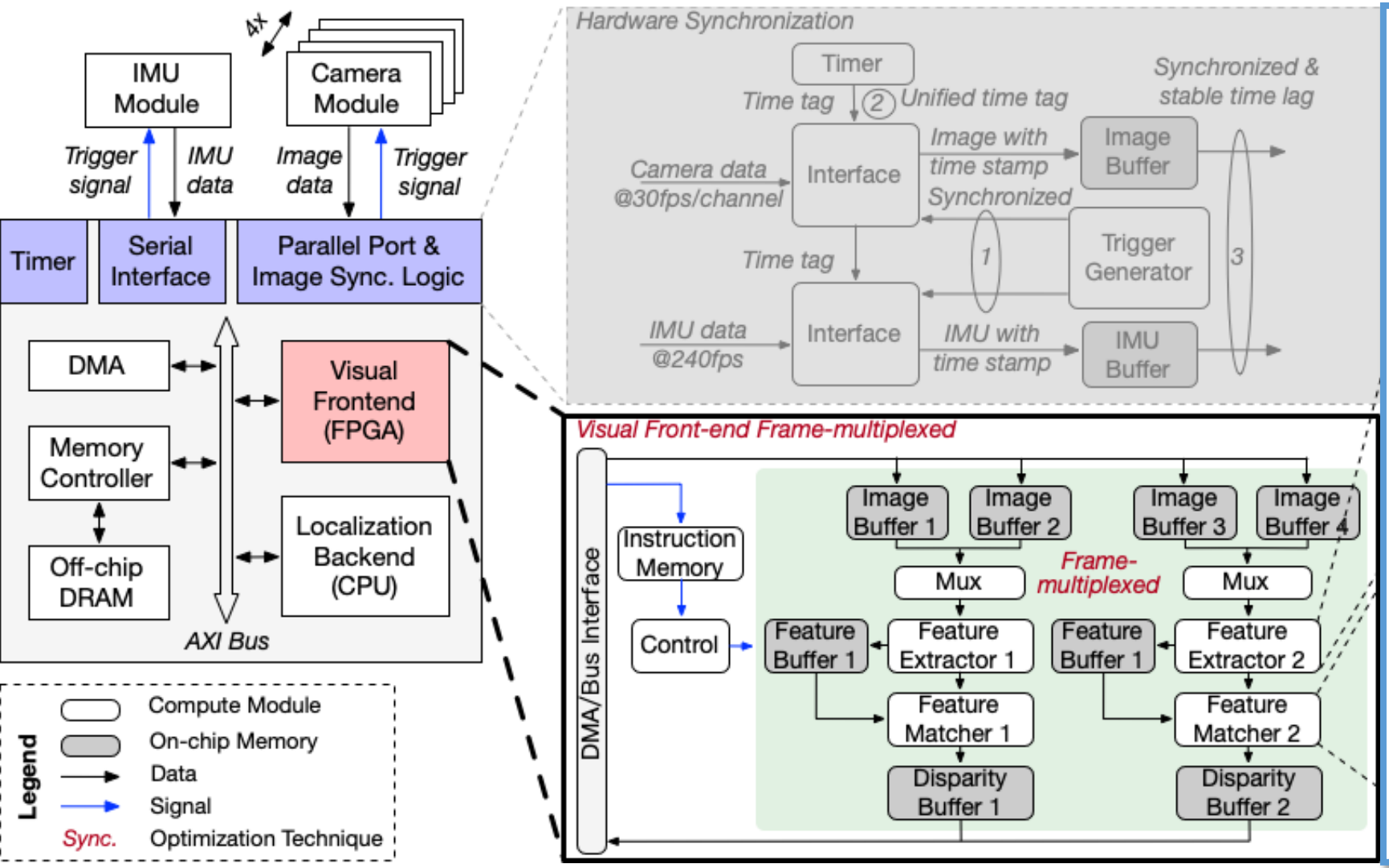


Hardware Synchronization Interface

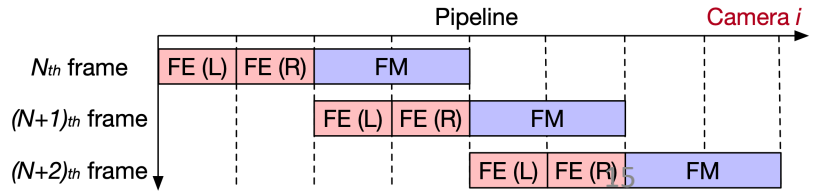


- Software synchronization leads to variable delay among 4 input images
- A direct IO architecture
- Hardware synchronization step:
  - Trigger generator for camera and IMU
  - Unified time tag
  - Synchronized & stable time lag

Frame-Multiplexed Visual Frontend



- Feature extraction: 7.28ms
  Feature matching: 14.59ms
- Two camera channels share one feature extractor
- Two identical hardware modules process two stereo cameras in parallel
- Pipeline



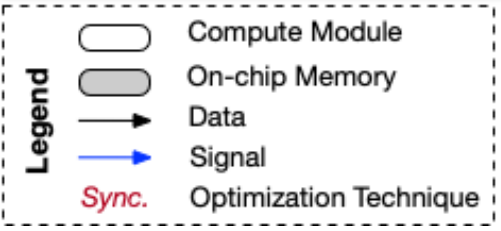
Hardware Architecture: Feature Extractor

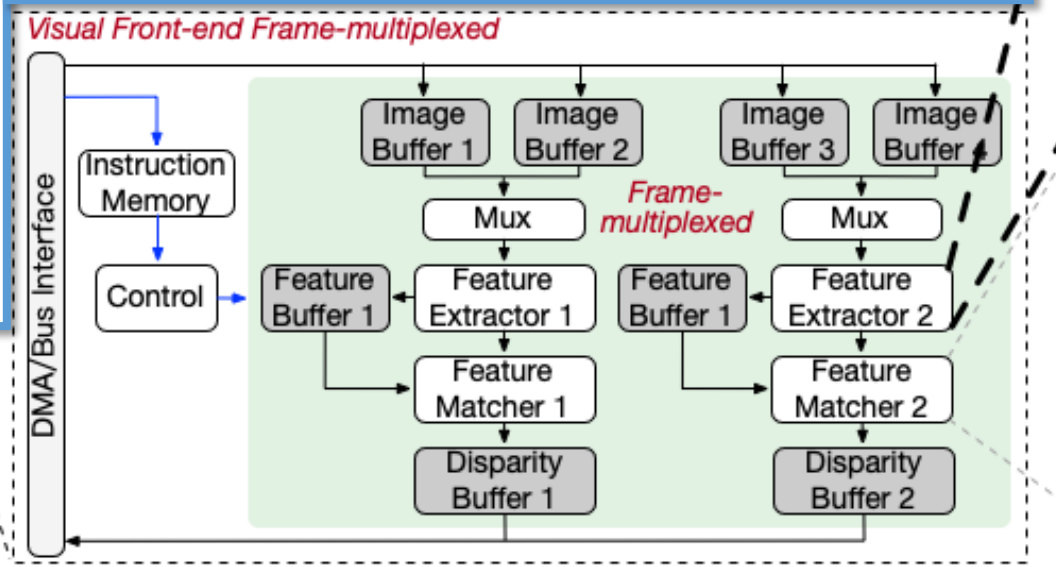
Feature Extractor Module:

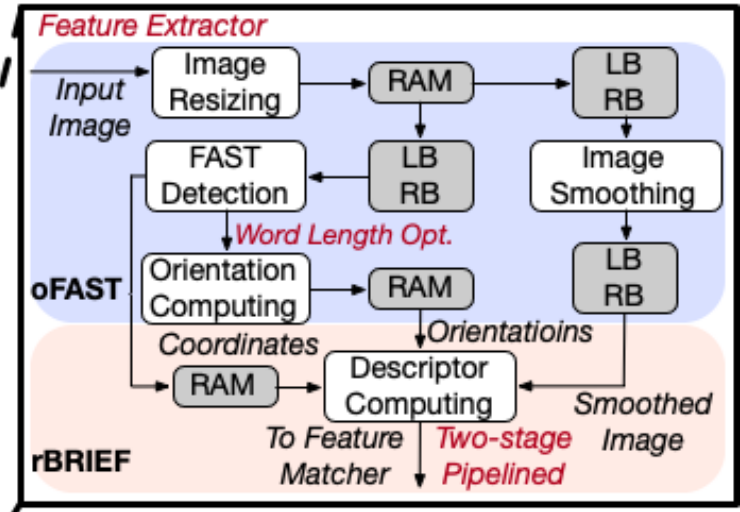
- Image Resizing
- FAST Detection
- Orientation Computing
- Image Smoothing
- Descriptor Computing

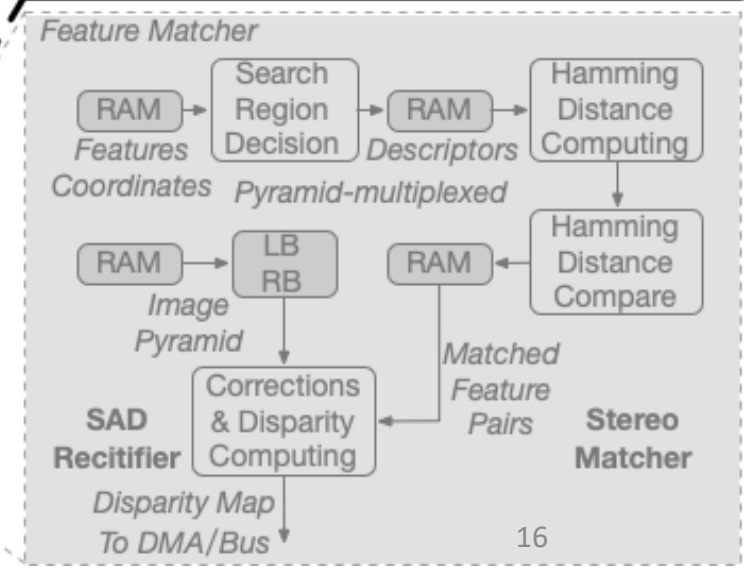
Design Traits

- Orientation computing: word length optimization
- Descriptor computing: synchronized two-stages shifting line buffers









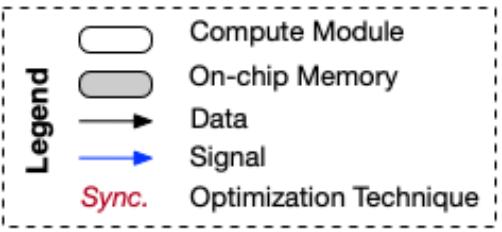
Hardware Architecture: Feature Matcher

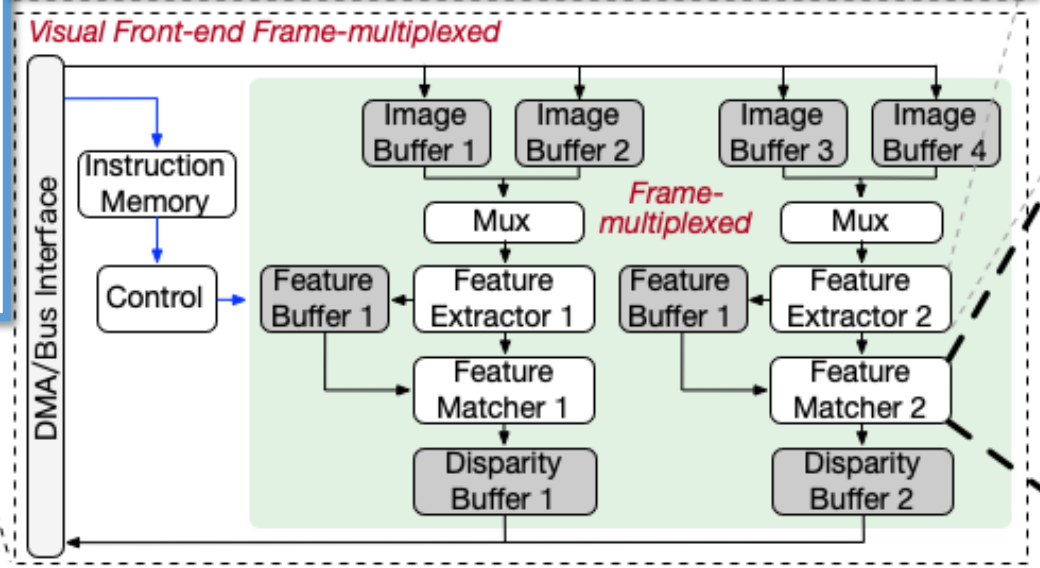
Feature Matcher Module:

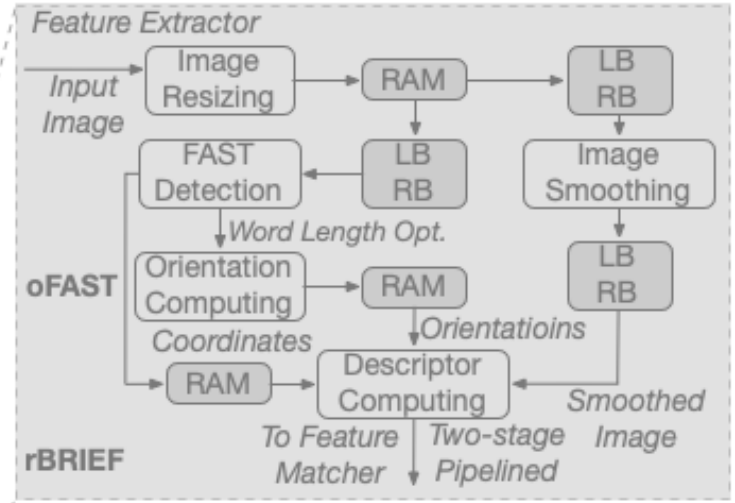
- Search Region Decision
- Distance Computing and Compare
- Correction and Disparity Computing

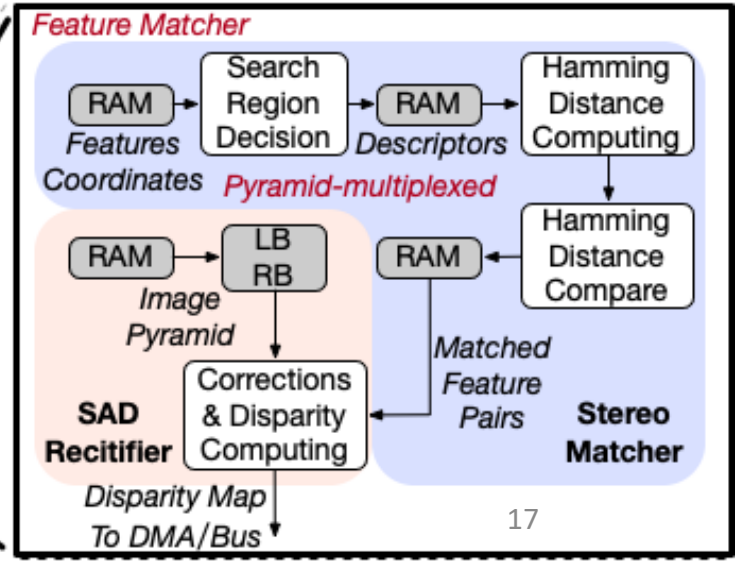
Design Traits

- Image pyramid-multiplexed scheme









Executive Summary

- Visual system is the compute bottleneck of autonomous machine systems and a lucrative acceleration target.
- Analyze algorithm blocks in ORB-based vision system.
- Present an energy-efficient quad-camera visual hardware architecture on FPGA platform. Several optimization techniques are proposed.
- **Our FPGA design achieves up to 5.6x speedup and 34.6x energy efficiency improvement.**

Evaluation Results

Hardware Platform: Xilinx Zynq Ultrascale+ XCZU9EG MPSoC
Operating Frequency: Feature extraction: 203MHz
Feature matching: 230MHz
Total Resources: 274K LUTs, 548K FF, 912 BRAMs, 2520 DSPs
Resource Consumption:

Resource	Modular Used (640×480)			Total Used	
	FE	FM	Ctrl.	640×480	720×1280
LUT	96850	40034	1759	138643 (51%)	177196 (65%)
Flip-Flop	54100	12694	479	67273 (12%)	82730 (15%)
BRAM	271	0	0	271 (30%)	785 (86%)
DSP	32	0	0	32 (1%)	109 (4%)

- Evaluate on two different image resolutions: 640x480 and 720x1280
- FE consumes over 2/3 of frontend resource -> FE-multiplexing

Evaluation Results

Accuracy Analysis

	# Feature Points	# Matched Pairs	# Effective Depth Value
Software	961.2	211.5	107
FPGA	961.1	211.7	107.3
Error	-0.1	+0.2	+0.3

Performance and Power Evaluation

		Perform. (*fps*)	Perform. (%)	Power (*W*)	Power (%)
	Image Resolution: 640x480				
Accelerator Comparison	**FPGA (Ours)**	**69**	-	**1.63**	-
	FPGA (Liu, DAC'19)	56	81.16%	1.94	1.19×
	FPGA (Fang, FPT'17)	67	97.1%	4.56	2.80×
	Image Resolution: 1280x720				
CPU/GPU Comparison	**FPGA (Ours)**	**50.7**	-	**2.31**	-
	Nvidia TX1	9	17.75%	7	3.03×
	Intel i7 Core	15	29.59%	80	34.63×

Conclusion

- We identify that **vision frontend** is the **unified compute bottleneck** of various localization systems.
- We propose an ORB-based **real-time and energy-efficient visual system** for autonomous machines on **FPGA platform**.
- We present a **hardware synchronization** scheme to support multi-image channels and IMU for reliable localization.
- Several optimization techniques, including **frame-multiplexing, parallelisms, and pipelines** are exploited to reduce accelerator latency and energy.
- Compared with Nvidia TX1 and Intel i7, our design achieves **5.6× and 3.4× speedup in frame rate**, and **3× and 34.6× improvement in energy efficiency**.

Thank You

Email: zishenwan@gatech.edu