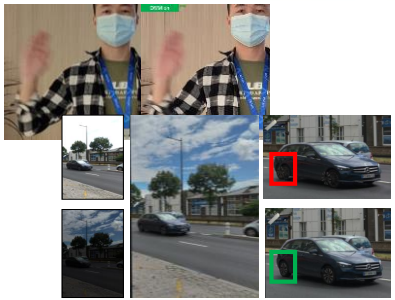
Develop Next-Gen Camera Apps Using Snapdragon Computer Vision Technologies

Judd Heape
VP, Product Management
Camera, Computer Vision and Video

Qualcomm

EVA: Engine for Visual Analytics



- Frame alignment
- Multi-frame noise reduction
- Motion-compensated temporal filtering

Motion estimation

Depth processing

- Video bokeh
- Background replacement
- Cinematic video
- XR: depth and 3D reconstruction

- Scaling, rotation and warping
- Lens distortion correction
- Rectification
- 3D zoom

Geometric correction

Engine for Visual Analytics

Indirect Time of Flight sensor (iToF)

- Accurate depth
- Face authentication
- Auto Focus

Feature extraction

Face detection

Face landmark detection

- SLAM and tracking
- VR obstacle tracking
- Autonomous parking

- Face beautification
- Face-centric 3A
- Cinematic video
- AI segmentation camera

Qualcomm

EVA architecture

Engine for Visual Analytics

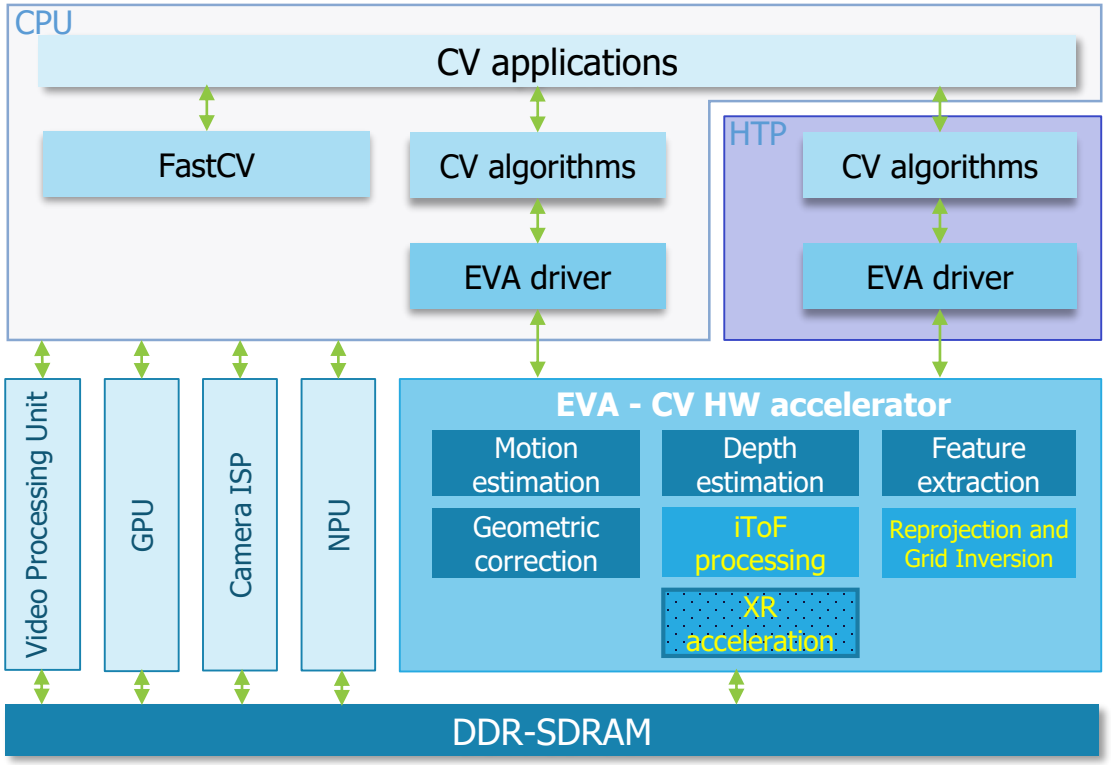
Computer Vision HW acceleration

Power savings

Performance enhancement

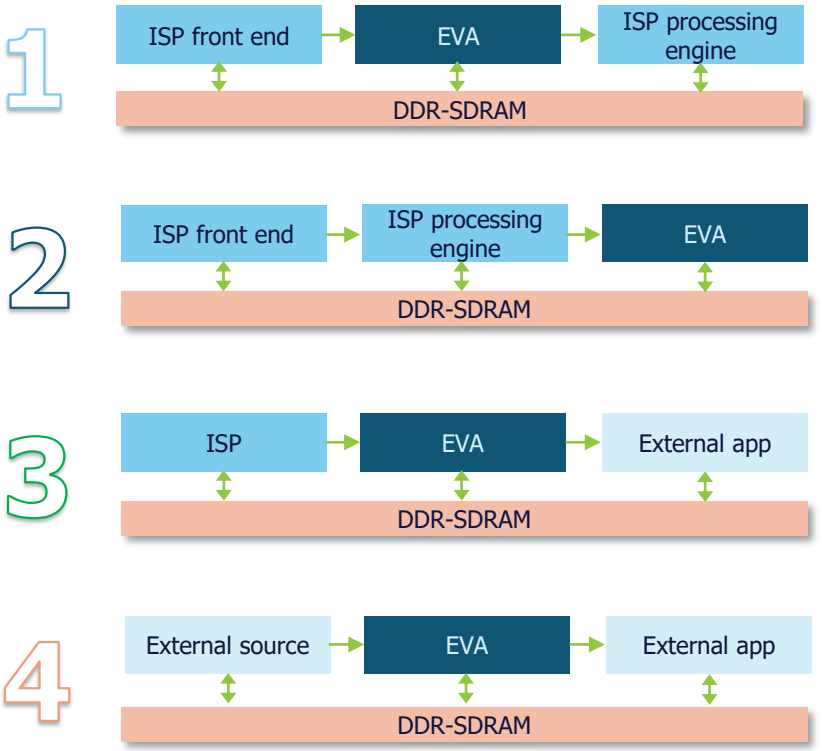
Access EVA APIs from CPU or Hexagon Tensor Processor (HTP)

Offload NPU / DSP / GPU / CPU for CV workloads

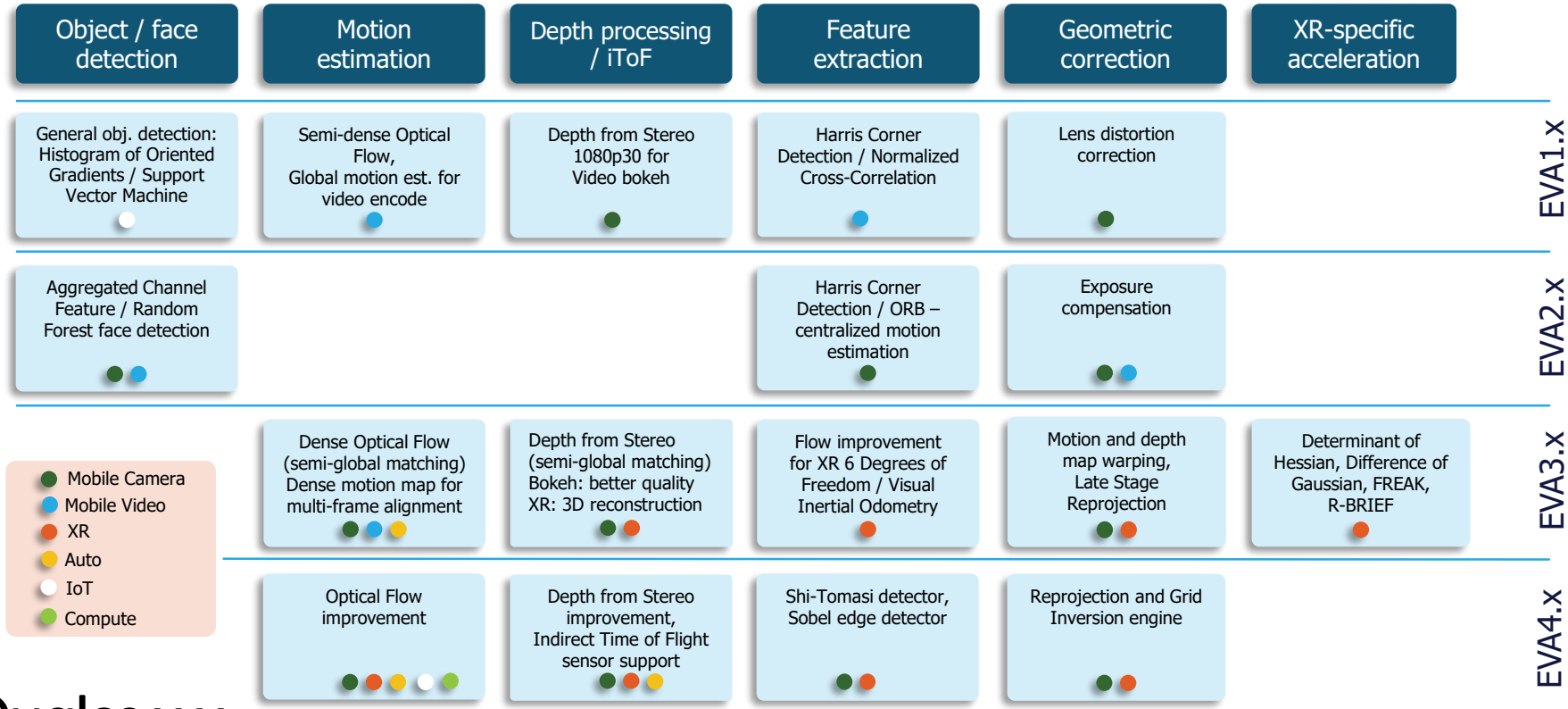


EVA access

- EVA blocks function from memory-to-memory for maximum flexibility
- APIs (for both CPU and HTP) include both synchronous APIs and asynchronous APIs
- There are direct interrupts between HTP and EVA cores for low-latency communication
- EVA includes an embedded processor primarily for task scheduling and managing hardware pipes
- EVA hardware pipes are shared between certain functions



EVA hardware blocks



EVA key features and KPIs

CV features	Next-generation Snapdragon EVA4.0	Snapdragon SM8450 / SM8550 EVA3.0
Optical Flow	Semi-global matching-based dense Optical Flow with quality improvement 2 modes: dense and semi-dense 1152x644 @ 60fps for dense, 1080p60 for semi-dense	Semi-global matching-based dense Optical Flow 2 modes: dense and semi-dense 1152x644 @ 60fps for dense, 1080p60 for semi-dense
Depth from Stereo	Semi-global matching-based Depth from Stereo, ± 3 pix search range 720p60, concurrent with Optical Flow	Semi-global matching-based Depth from Stereo, ± 3 pix search range 720p60, concurrent with Optical Flow
Active depth sensing	iToF Depth range 0.2-7 m, VGA @ 30fps Power saving: 1 W vs. SW iToF	-
Detectors, descriptors and matching	Pyramid generation Harris Corner, Shi-Thomasi detectors ORB, R-BRIEF and FREAK descriptors Normalized Cross-Correlation w/inline warping	Pyramid generation Harris Corner detectors ORB feature descriptor Normalized Cross-Correlation w/inline warping
Scaling, rotation and warping	Sparse-grid transform (35x27 or 67x51) Dense-grid transform (1:8) Perspective transform (3x3) 1920x1080 @ 240fps	Sparse-grid transform (35x27 or 67x51) Dense-grid transform (1:8) Perspective transform (3x3) 1920x1080 @ 240fps
Reprojection and Grid Inversion	3D depth map reprojection 3D motion map reprojection Motion vector resampling	-

Qualcomm

EVA feature APIs

EVA version	EVA features	EVA API
EVA3.0 SM8450 / SM8550	Image warping	evaWarp_Sync / evaWarp_Async
	Depth from Stereo	evaDfs_Sync / evaDfs_Async
	Normalized Cross Correlation	evaNccFrame_sync / evaNccFrame_Async
	Optical Flow	evaOF_Sync / evaOF_Async
	Feature extraction (Harris Corner Detection)	evaFeaturePoint_Sync / evaFeaturePoint_Async
	Feature Descriptor calculation & matching	evaDcm_Sync / evaDcm_Async
	Downscaler	evaScaledown_Sync / evaScaledown_Async
	Pyramid image generation	evaPyramidImage_Sync / evaPyramidImage_Async
EVA4.0 Next-gen Snapdragon	Motion vector resample	evaMVResample_Sync / evaGridResample_AsFync
	Depth reprojection	evaDepthReprojection_Sync / evaDepthReprojection_Async
	3D projection of motion field	evaMotionField3DProjection_Sync / evaMotionField3DProjection_Async

Optical Flow

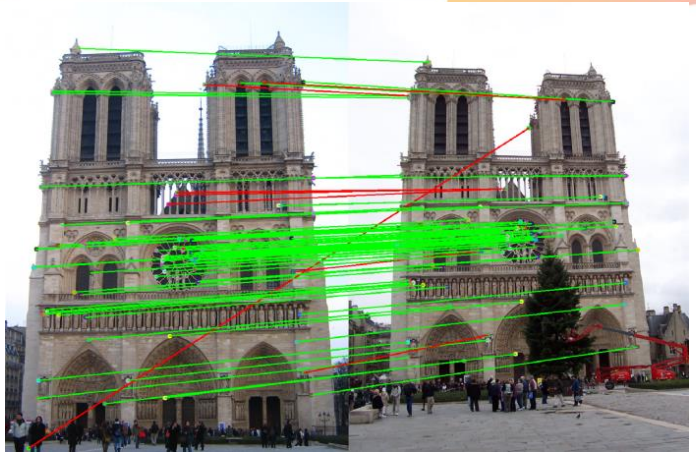
Sparse Motion

- Feature point detection and matching
- Global motion estimation

Dense Motion

- Semi-dense Optical Flow
- Dense Optical Flow
- Replace low-confidence local motion with global motion (new in EVA4)

	Semi-dense	Dense
Motion density	Every 2x2 block	Every pixel
Motion accuracy	1/8 pixel	1/16 pixel
Motion range (X,Y)	±128, ±64	±64, ±32
Max resolution	1920x1440	1920x1440
Confidence map	8-bit	8-bit
Frames per second	1920x1080 @ 60fps	1152x648 @ 60fps



Sparse motion

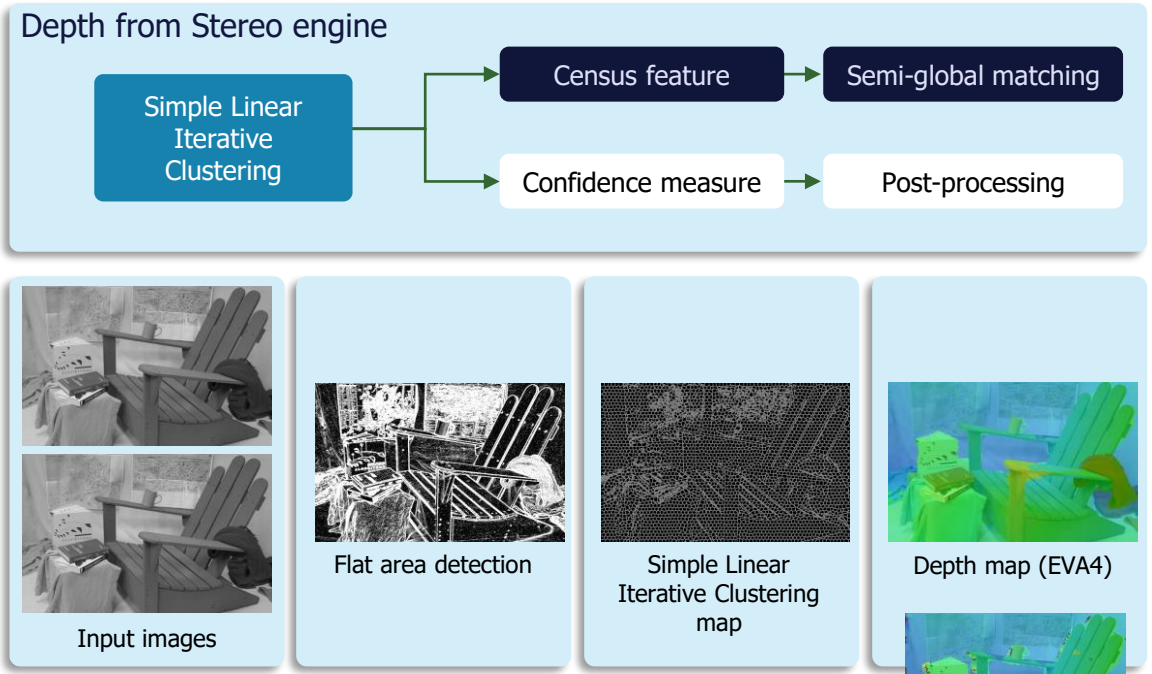
Dense motion

Depth from Stereo estimation

Depth from Stereo

- Sum of Absolute Difference matching between left and right images
- Super-pixel segmentation on Simple Linear Iterative Clustering
- Feature extraction and matching
- Confidence map and post-processing

	KPI
Depth density	Every pixel
Disparity accuracy	1/16 pixel
Disparity level	[0, 63]
Max resolution	720p 60fps
Confidence map	8-bit



Geometric Correction Engine

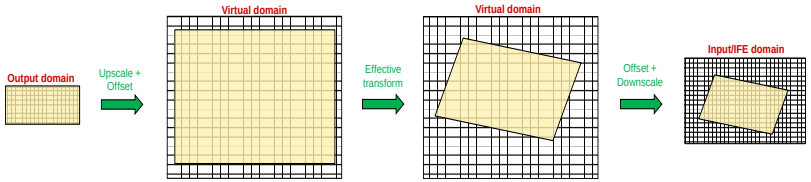
Low-power, high-quality warping

- Maps output pixels to input pixels

Effective transformation

- Sparse-grid transformation (35x27 or 67x51)
- Dense-grid transformation (8 pixel grid)
- Perspective transformation (3x3 transform)

Effective transform



Use Cases

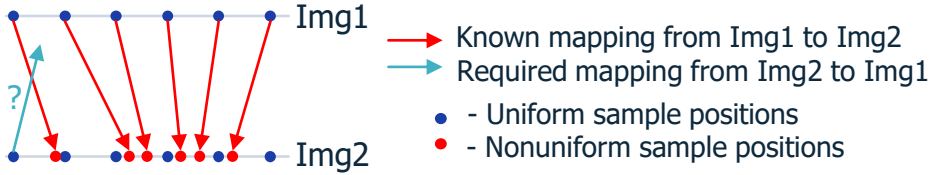
- Lens distortion correction
- Motion vector grid composition
- Rectification

Lens distortion correction

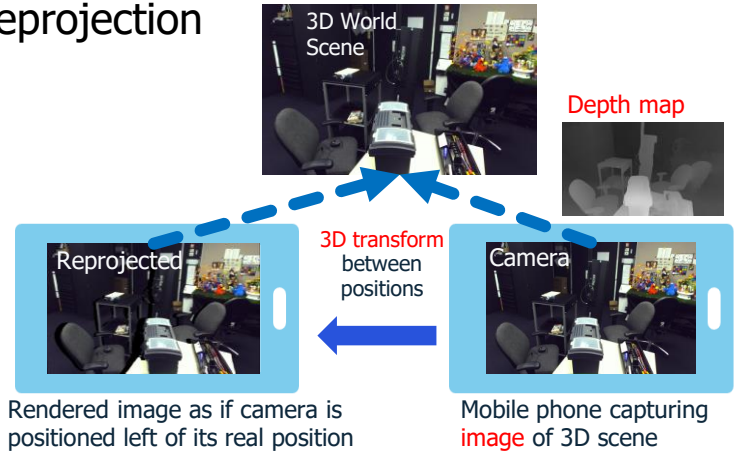
Rectification

Reprojection and Grid Inversion

Grid Inversion



Reprojection



Use cases

- Temporal deblur
- Super slow motion
- Video concurrency transform composition
- Frame interpolation for latency reduction

Use cases

- Depth and RGB sensor alignment
- Gaze correction
- XR: Late Stage Reprojection
- 3D Spatial Alignment & Translation: eliminate parallax jump
- 3D zoom
- 3D camera motion special effect

Qualcomm

Feature detectors and descriptors

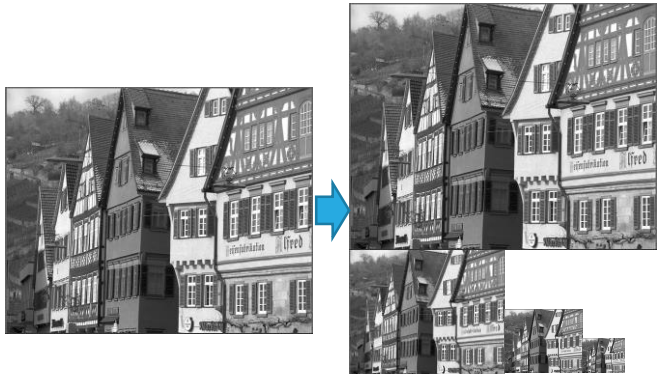
Pyramid generation
- Up to 6 levels

Detectors
- Harris Corner detector
- Shi-Tomasi detector
- Sobel edge detector

Descriptors
- Oriented FAST detector and rotated BRIEF descriptor (R-BRIEF)

Matching
- Normalized Cross-Correlation and binary for R-BRIEF

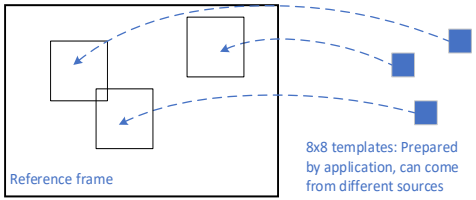
	KPI
Modes	8x8 cell or zone
Max resolution	UHD 3840x2160
Frame rate	1920x1080 @ 60fps
Descriptor size	256 for R-BRIEF



Pyramid generation

Harris Corner vs. Shi-Tomasi

Sobel edge map



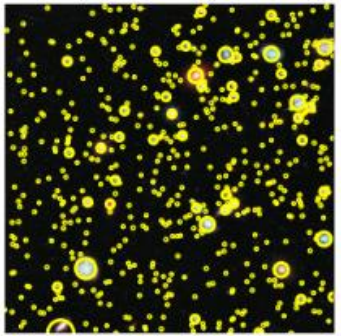
Normalized Cross-Correlation

XR Accelerator: detectors & descriptors

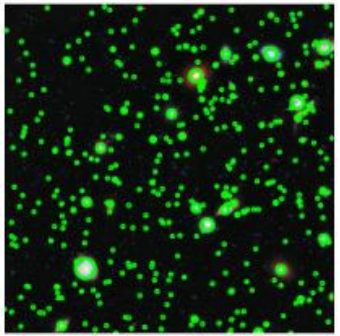
Blob detectors and edge map
- Determinant of Hessian
- Difference of Gaussian
- Laplacian of Gaussian

Descriptors
- Rotated BRIEF descriptor
- FREAK descriptor

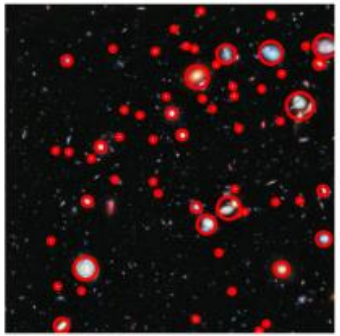
Matching
- Binary matching



Laplacian of Gaussian



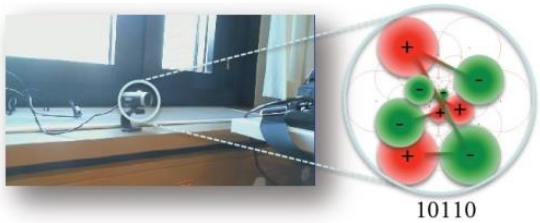
Difference of Gaussian



Determinant of Hessian

	KPI
Modes	8x8 cell or zone
Max resolution	UHD 3840x2160
Frame rate	1920x1080 @ 60fps
Descriptor size	256 for R-BRIEF, 512 for FREAK

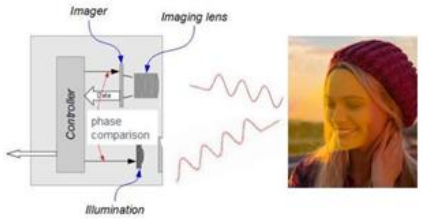
FAST and Rotated BRIEF (ORB)



FREAK descriptor

Indirect Time of Flight (iToF)

Active depth sensor



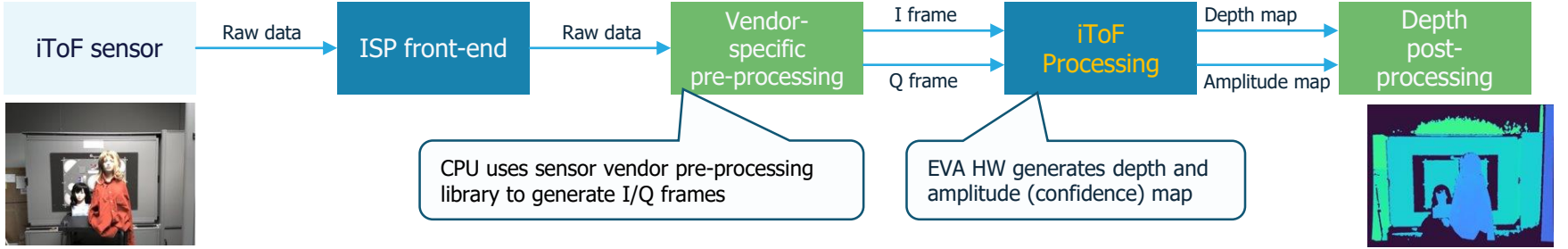
Indirect Time of Flight

- Measurement of phase shift: phase shift is proportional to distance
- Very small pixel size, standard CMOS technology, enables high resolution (QQVGA, VGA up to 720p)

Comparison of depth solutions

	Depth from Stereo	iToF
Light source	-	Pulsed or continuous light
Power consumption	Low (<100 mW)	High (0.5-1.5 W)
Latency	Low (<8 ms on 720p)	Mid (<10 ms)
Accuracy	Mid	High (<2% * depth)
Resolution	High (VGA, 720p or higher)	Mid to low (VGA, HQVGA or lower)
Depth range	Varies (4 m – 10 m)	Sensor dependent (5 m – 7 m)
Best use cases	Ample ambient light	Indoor

Indirect Time of Flight processing

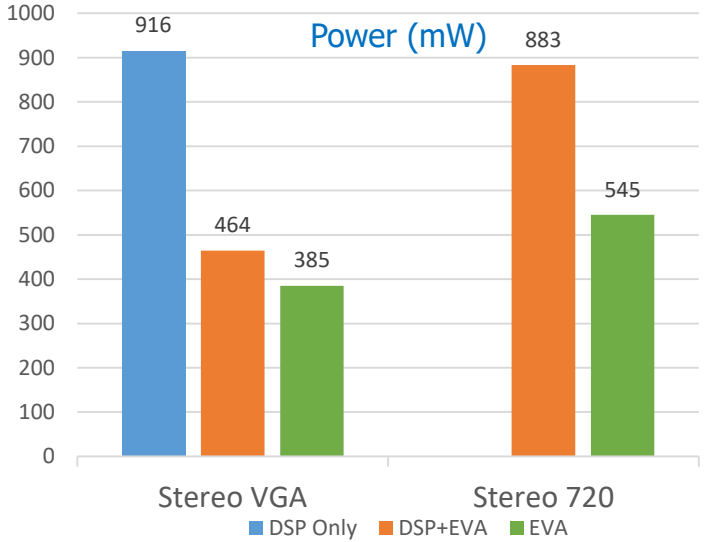


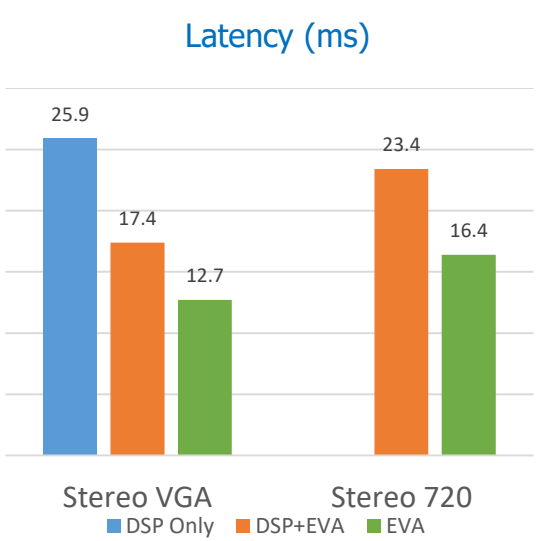
	KPI
Image resolution	HQVGA / QVGA / VGA / 720p / QuadVGA / 1080p; min: 32x24, max: 1920x1080
Frequency	Single or dual frequency supported
Mode	Flood, spot
Frame rate	720p60, 1080p30
Input format	Signed 16-bits for In-phase (I) and Quadrature (Q)
Output format	Unsigned 16-bits for depth (1-bit masking + 15-bits depth, normalized / unnormalized) Unsigned 16-bits for amplitude map (confidence)
Power consumption	~20x power saving compared to SW processing

The EVA hardware advantage: Power consumption & latency reduction

Three main advantages of running your workload on EVA:

1. Low power consumption
2. High performance (low latency)
3. NPU / DSP / GPU / CPU offloading





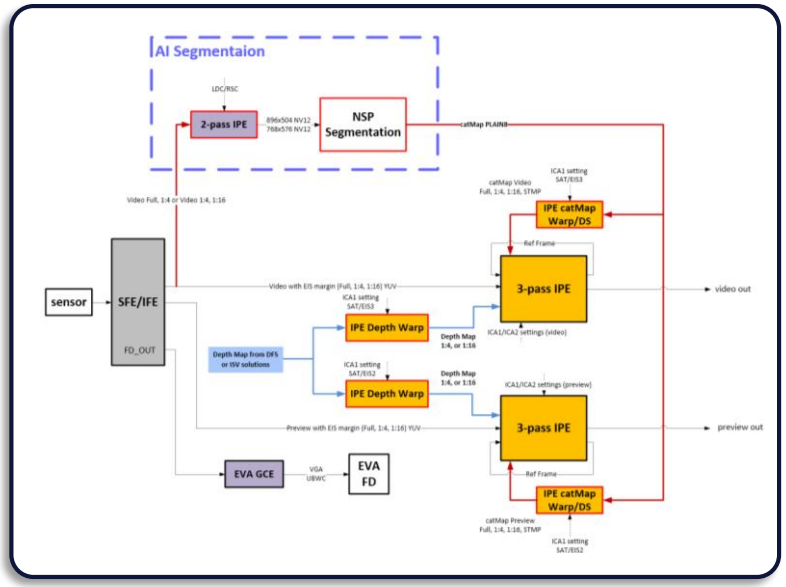
6 Degrees of Freedom processing improvement with power & latency reduction

CV Use Case 1:
Depth map from stereo cameras

Applications

- Accurate camera / video bokeh effect
- Background replacement in video conference or video recording
- AR/VR depth mapping

Video bokeh flow

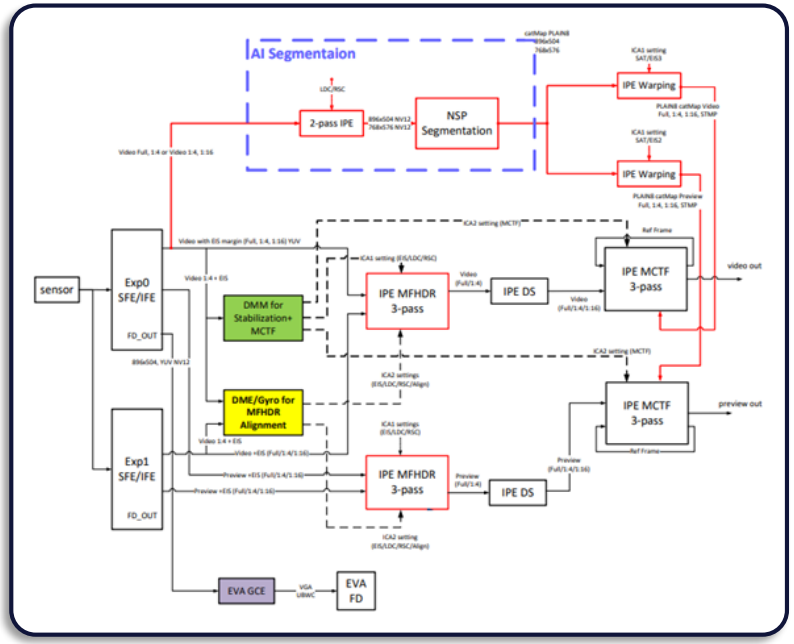


CV Use Case 2:
Dense motion map for video multi-frame HDR

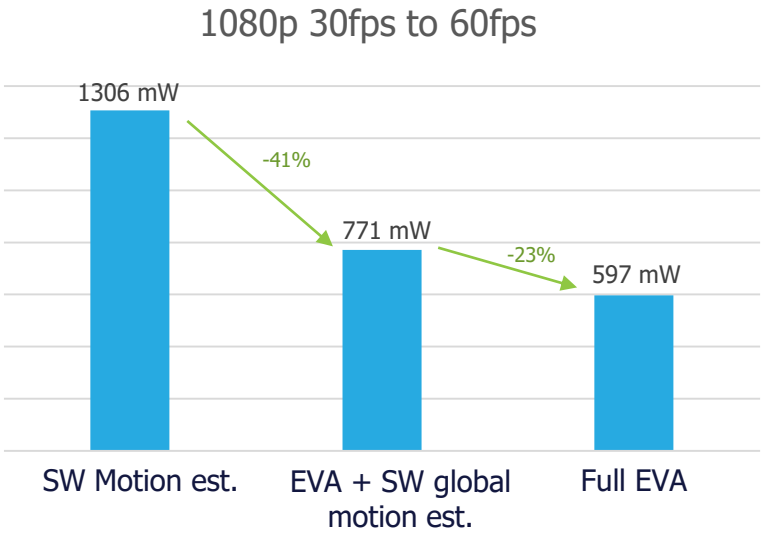
Key Benefits of EVA

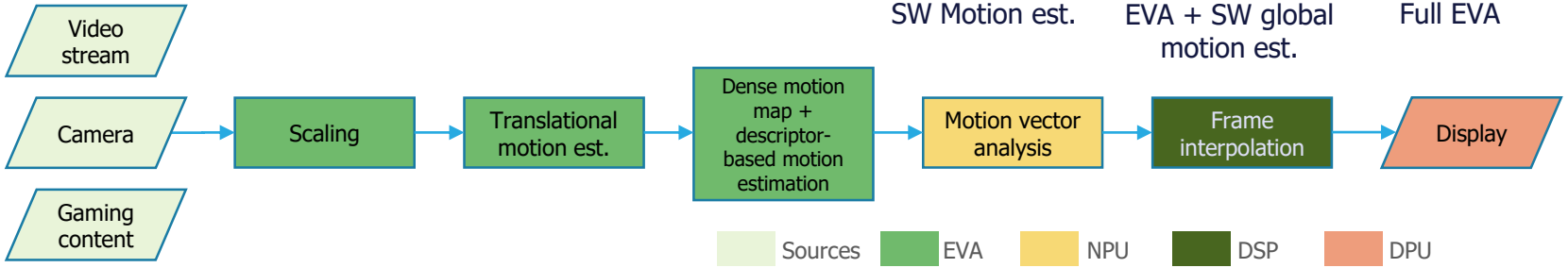
- Estimating and compensating for motion is key to achieve high-quality HDR video
- Removes ghosting artifacts in combined video frames
- Running global motion and local motion estimation simultaneously requires a large amount of computation power
- Superset of motion compensated temporal filtering (noise reduction) use case

Multi-frame HDR flow



CV use case 3:
Motion estimation in frame rate conversion

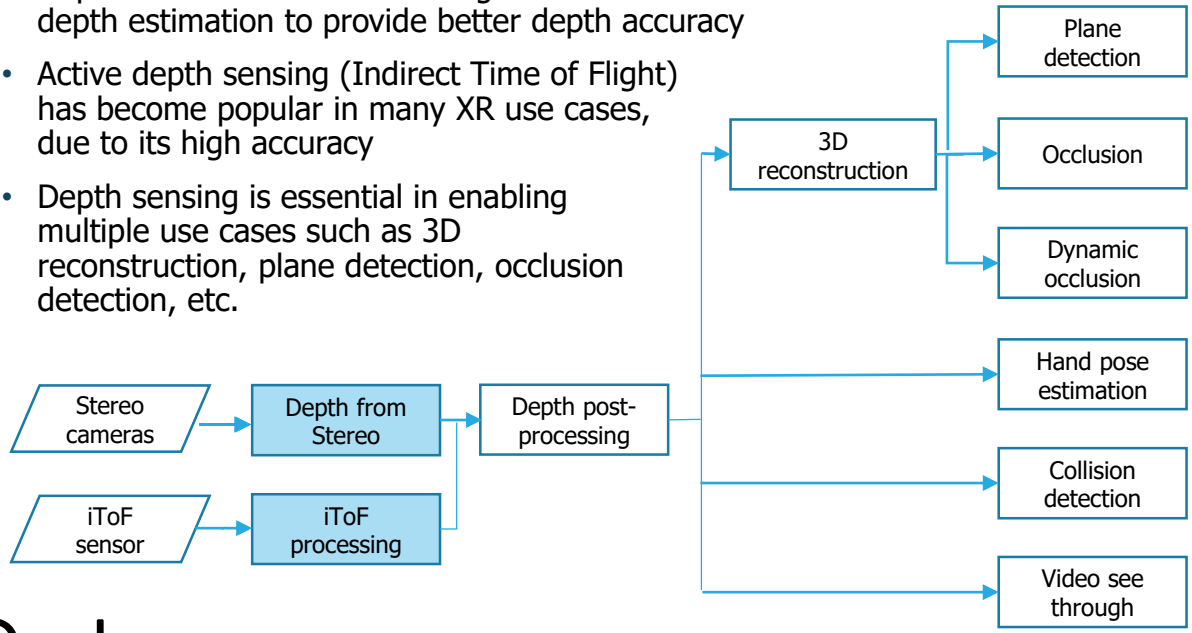




CV use case 4:
Depth in XR perception

Key benefits of EVA

- Depth estimation is a key block in the XR perception pipeline
- Depth from Stereo can work together with ML-based depth estimation to provide better depth accuracy
- Active depth sensing (Indirect Time of Flight) has become popular in many XR use cases, due to its high accuracy
- Depth sensing is essential in enabling multiple use cases such as 3D reconstruction, plane detection, occlusion detection, etc.



Qualcomm

Start developing on Snapdragon!

For access to the EVA SDK, contact:

Xin Zhong
Director, Product Management
xzhong@qti.qualcomm.com

Resources

Qualcomm® Mobile AI

Mobile AI | On-Device AI | Qualcomm

Qualcomm Technologies & Google NAS

Qualcomm Technologies and Google Cloud Announce Collaboration on Neural Architecture Search for the Connected Intelligent Edge | Qualcomm

Qualcomm Wireless Academy

Fundamentals of AI

Available for free until October 2023

Dr. Vinesh Sukumar
Senior Director, Product Management – AI/ML
vinesuku@qti.qualcomm.com

Thank you!

Qualcomm