Tools for Creating Next-Gen Computer Vision Apps on Snapdragon

Judd Heape
VP Product Management for Camera, Computer Vision and Video Technology
Computer Vision in Snapdragon
Three function levels to provide comprehensive CV solutions

1. CV hardware
   Acceleration blocks to support and enable hardware, software and system designs in Snapdragon platforms

2. CV Algorithms to
demonstrate complete workflows that provide state-of-the-art solutions to certain perception problems

3. CV end-to-end
   Applications in mobile, XR, Automotive and IOT market segments to enable unique and enhanced user experiences

© 2022 Qualcomm
# Engine for Visual Analytics (EVA): Computer Vision Hardware Blocks

<table>
<thead>
<tr>
<th>Object/Face Detection</th>
<th>Optical Flow</th>
<th>Depth Estimation</th>
<th>Feature Extraction</th>
<th>Geometry Correction</th>
<th>XR &amp; 3DR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object Detection</strong></td>
<td><strong>Semi-dense OF</strong></td>
<td><strong>DFS 1080p@30</strong></td>
<td><strong>Feature – video encode</strong></td>
<td><strong>Lens Distortion</strong></td>
<td><strong>XRA - DoH, DoG</strong></td>
</tr>
<tr>
<td>HOG/SVM</td>
<td>GMO for video encode</td>
<td>Video Bokeh</td>
<td>with 30% BR reduction</td>
<td>Correction</td>
<td>FREAK, R-BRIEF</td>
</tr>
<tr>
<td><strong>ACF/RDF Face</strong></td>
<td><strong>Dense OF (SGM based)</strong></td>
<td><strong>DFS (SGM based)</strong></td>
<td><strong>Flow improvement</strong></td>
<td><strong>Exposure Compensation</strong></td>
<td>(In EVAa 3.5)</td>
</tr>
<tr>
<td><strong>Detection</strong></td>
<td><strong>Dense motion map for multi-frame processing, sensor alignment</strong></td>
<td>• Bokeh better quality</td>
<td>for XR 6DoF/VIO</td>
<td><strong>Motion and depth</strong></td>
<td><strong>DoH, DoG</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Visual special effect</td>
<td>map warping</td>
<td><strong>map warping</strong></td>
<td><strong>FREAK, R-BRIEF</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• XR 3D reconstruction</td>
<td>LSR (in EVAa 3.5)</td>
<td><strong>LSR</strong></td>
<td>(In EVAa/3.5)</td>
</tr>
</tbody>
</table>

- Mobile Camera / Video
- XR
- Auto
- IoT

© 2022 Qualcomm
Optical Flow

Sparse Motion
- Feature Point Detection, Local and Global Motion
- Various Detector and Descriptors (Harris, DoH, DoG, FREAK)

Dense Motion
- Semi-dense Optical Flow (sDOF)
- Dense Optical Flow (DOF)
- Hybrid Deep Learning based Motion + OF Segmentation Enhanced

<table>
<thead>
<tr>
<th></th>
<th>Semi-Dense OF</th>
<th>Dense OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion Density</td>
<td>Every 2x2 block</td>
<td>Every pixel</td>
</tr>
<tr>
<td>Motion Accuracy</td>
<td>1/8 pixel</td>
<td>1/16 pixel</td>
</tr>
<tr>
<td>Motion Range (X,Y)</td>
<td>±128, ±64</td>
<td>±64, ±32</td>
</tr>
<tr>
<td>Max Resolution</td>
<td>1920x1080</td>
<td>1152x648</td>
</tr>
<tr>
<td>Confidence Map</td>
<td>8-bit</td>
<td>8-bit</td>
</tr>
<tr>
<td>Frames per Second</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

© 2022 Qualcomm
Depth from Stereo (DFS)

- Super-pixel Segmentation on SLIC
- Feature Extraction and Matching
- Confidence Map and Post Processing

**DFS Engine**

- SLIC (Simple Linear Iterative Clustering)
- Census Feature
- SGM
- Confidence Measure
- Post Processing

<table>
<thead>
<tr>
<th>Depth Density</th>
<th>Every pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disparity Accuracy</td>
<td>1/16 pixel</td>
</tr>
<tr>
<td>Disparity Level</td>
<td>[0,63]</td>
</tr>
<tr>
<td>Max Resolution</td>
<td>720P@60FPS</td>
</tr>
</tbody>
</table>
Geometric Correction Engine (GCE)

Low-power High-quality Warping
- ICA 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
- Dense Grid Transformation
- Perspective Transformation
- Sparse Grid Transformation

© 2022 Qualcomm
Normalized Cross Correlation

NCC Supports Two Modes
- Patch to Frame Mode
- Frame to Frame Mode

Frame Matching Using Harris Corners and NCC

Patch-to-frame mode
- 8x8 templates: Prepared by application, can come from different sources

Frame-to-frame mode
- Templates: All in the same frame
Face Detection

Deep Learning based Face Detection (FD)
- Min Face Size: 32x32
- Detection Accuracy: 95%
- 1080p@60FPS
- Multiple cameras supported

Under Non-Ideal Conditions
- Strong Backlight
- Full Profile
- Occlusions – Face Masks, Hats, Glasses, Sunglasses

© 2022 Qualcomm
EVA Architecture and Access

- The EVA APIs are exposed both from the CPU and Hexagon Processor sides
- It includes both synchronous APIs and asynchronous APIs
- There are direct interrupts between the Hexagon Processor and EVA cores for low latency communication
- EVA includes embedded CPU primarily for task scheduling and hardware pipes
- EVA hardware pipes are shared between certain functions
<table>
<thead>
<tr>
<th>Feature</th>
<th>EVA API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Warping</td>
<td>evaWarp_Sync / evaWarp_Async</td>
</tr>
<tr>
<td>Depth from Stereo (DFS)</td>
<td>evaDfs_Sync / evaDfs_Async</td>
</tr>
<tr>
<td>Normalized Cross Correlation (NCC)</td>
<td>evaNccFrame_Sync / evaNccFrame_Async</td>
</tr>
<tr>
<td>Optical Flow (OF)</td>
<td>evaOF_Sync / evaOF_Async</td>
</tr>
<tr>
<td>Feature Extraction (HCD)</td>
<td>evaFeaturePoint_Sync / evaFeaturePoint_Async</td>
</tr>
<tr>
<td>Feature Descriptor Calc &amp; Matching</td>
<td>evaDcm_Sync / evaDcm_Async</td>
</tr>
<tr>
<td>Downscaler</td>
<td>evaScaledown_Sync / evaScaledown_Async</td>
</tr>
<tr>
<td>Pyramid Image</td>
<td>evaPyramidImage_Sync / evaPyramidImage_Async</td>
</tr>
</tbody>
</table>
EVA SDK Simulator

- Query and populate init configuration parameters.
- Query and populate frame configuration parameters.
- Create EVA session.
- Initialize feature handle.
- Load input data.
- Allocate memory for output buffers.
- Start EVA session.
- Execute feature sync mode.
- Compare output with ground truth.
- Deallocate buffers and terminate EVA session.

[1;32m[EVA_SDK INFO] [0m Output matches ground truth
[1;32m[EVA_SDK INFO] [0m Sample code finished execution successfully.

C:\Workspace\Release\Snap\EVA-Mobile-3.0.6\build\code\Release>.

Command Prompt

: IS NOT recognized as an internal or external command, operable program or batch file.

C:\workspace\snap\eva-mobile-3.0.6\build\code\release>eva_sdk.exe --help
Reset XML Parameter tree in Memory:
Usage
  .\eva_sdk.exe [options]
Options
  --feat eva-feature-string
  --mode eva-mode-number
  --data path-to-data-folder
  --help
Features
  DCM : Descriptor Calculation and Matching
  Modes - 1 (eva_dcm_cal) / 2 (eva_dcm_cal_match)
  DFS : Depth From Stereo
  DS : Scaledown
  FPX : Feature Point Extraction
  Modes - 1 (eva_fpx_peak_8x8) / 2 (eva_fpx_zone)
  NCC : Normalized Cross Correlation
  Modes - 1 (eva_ncc_patch_mode) / 2 (eva_ncc_frame_mode)
  WNCC : Warped Normalized Cross Correlation
  Modes - 1 (eva_wncc_mono_mode) / 2 (eva_wncc_stereo_mode)
  OF : Optical Flow
  PVRFPX : Pyramid Image and Feature Point Extraction
  PYRIMG : Pyramid Image
  WARP : Image Warping
  Modes - 1 (PERSPECTIVE) / 2 (GRID) / 3 (LMC)
CV Use Case 1
Depth Map from Stereo Cameras (DFS)

Applications

• Accurate Camera/Video Bokeh effect

• Background replacement in video recording or Zoom call

• AR/VR (3D Reconstruction, Video Passthrough, Occlusion)
Applications

• Accurate Camera/Video Bokeh effect
Key Benefits of EVA

- Register multiple frames with local motion compensated
- Remove ghosting artifacts in combined video frames
CV Use Case 4
Dense Motion Map (DMM) for Video MFHDR

Key Benefits of EVA

• Estimating and compensating for motion is key to achieve high quality HDR video

• Remove ghosting artifacts in combined video frames

• Running global motion and local motion estimation simultaneously requires large amount of computation power
CV Use Case 5
Face Detection (FD) and Face Landmark Detection (FLD)

Applications

• Gender/Expression/Emotion/Gaze detection
• Avatar animation
• Geometric personalization

Qualcomm Deep Learning-based 3D face landmark detection reaches high accuracy in locating **115 or 300** facial landmarks
Start Developing on Snapdragon

Capture at higher FPS

Extend battery life

Tap into hardware-accelerated CV features with an SDK not previously available
For access to the SDK contact:

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