INTELLIGENT VIDEO ANALYTICS(IVA) ON INTEL ARCHITECTURE

Aug 2018
Premchander Rao T
APJ Staff Lead –IoT and Smart Video
Core and Visual Computing Group –Intel
Optimization Notice

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804.
Legal Notices and Disclaimers (1 of 2)

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services, and processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Any forecasts of goods and services needed for Intel's operations are provided for discussion purposes only. Intel will have no liability to make any purchase in connection with forecasts published in this document.

Arduino* 101 and the Arduino infinity logo are trademarks or registered trademarks of Arduino, LLC.

Altera, Arria, the Arria logo, Intel, the Intel logo, Intel Atom, Intel Core, Intel Nervana, Intel Xeon Phi, Movidius, Saffron, and Xeon are trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2018 Intel Corporation. All rights reserved.
Legal Notices and Disclaimers (2 of 2)

This document contains information on products, services, and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications, and roadmaps. Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit www.intel.com/performance.

Cost-reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors, known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo, Pentium, Celeron, Atom, Core, Xeon, Movidius, Saffron, and others are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2018, Intel Corporation. All rights reserved.
The “Internet of things”

- The “Internet of things” (IOT) is coming to life
  - Number of “things” installed and connected is growing
  - The ultimate “thing” sensor is a camera
The “Internet of things”

• The “Internet of things” (IOT) is coming to life
  • Number of “things” installed and connected is growing
  • The ultimate “thing” sensor is a camera
• Data is exploding, small portion is used
  • And usually in retro-respective
  • Can’t be handled by Humans
• We must build smarter automated systems

* Source: IDF, April 2017
A system of systems, edge to cloud
Many different industries, many different devices..

Key vision usages (AON):
- Face detection/recognition
- Head/gaze tracking
- Gesture/pose recognition

Key vision usages:
- Intrusion detection
- Crowd monitoring
- Person/object tracking

Key vision usages:
- Obstacle detection
- Collision avoidance
- Scene analytics

Key vision usages:
- Pedestrian/vehicle detection
- Collision avoidance
- Scene analytics

Head Mounted Devices (AR/VR)
Smart Edge Cameras (DSS)
Autonomous Devices
Automotive
Amazing new capabilities

A Car
A Black Car
Detect make, model
Not the right license plate
Not the owner
Amazing new capabilities

People Detection

People Tracking

Analyze behavior/ intentions
Choosing the “right” hardware

- Consider in each device
  - Compute efficiency, parallelism (# of EU/Cores)
  - Power consumption
  - Memory hierarchy and communication
  - Programming model, APIs
**SMART CAMERAS**

- Size/Cost/Power constraints
- 10’s mS Latency
- Detection
- >.5 TOPS

**Performance / W**
(performance efficiency)

**VIDEO GATEWAYS/NVR’S**

- Aggregate video streams (..100s)
- Detect / Classify / Track / Analyze
- >2 TOPS

**Performance / $**
(channel density)

**DATACENTER / CLOUD**

- Aggregate video/data streams
- Classify / Track / Analyze
- Training
- >5 TOPS

**Performance / Sec**
(Absolute performance)
<table>
<thead>
<tr>
<th>CPU, GPU</th>
<th>FPGA</th>
<th>MOVIDIUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart Cameras</td>
<td>Video Gateways/NVR's</td>
<td>Datacenter / Cloud</td>
</tr>
<tr>
<td>Size/Cost/Power constraints 10's mS Latency Detection &gt;.5 TOPS</td>
<td>Aggregate video streams (.100s) Detect / Classify / Track / Analyze &gt;2 TOPS</td>
<td>Aggregate video/data streams Classify / Track / Analyze Training &gt;5 TOPS</td>
</tr>
</tbody>
</table>

**Performance / W** (Performance efficiency)

**Performance / $** (channel density)

**Performance / Sec** (Absolute performance)
Multi-purpose to purpose-built
AI compute from cloud to device

**HARDWARE**

**MAINSTREAM**

**INTENSIVE**

**TRAINING**

DEEP LEARNING

**INFERENCEx**

**MOST OTHER AI**

All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice.
DEEP LEARNING INFERENCE ACCELERATORS

INTEL® FPGA
Custom deep learning inference

INTEL® MOVIDIUS™ VPU
Low power computer vision & inference

INTEL® MOBILEYE EYEQ
Autonomous driving inference platform

INTEL® GNA IP¹
Ultra low power speech & audio inference

INTEGRATED GRAPHICS
Built-in deep learning inference

DATA CENTER
Small scale clusters to a few on-premise server & workstations

EDGE
User-touch end-devices typically with lower power requirements

¹GNA=Gaussian Mixture Model and Neural Network Accelerator
All products, computer systems, dates, and figures are preliminary based on current expectations, and are subject to change without notice. Images are examples of intended applications but not an exhaustive list.
Intel® Movidius™ VPU powered devices

Hikvision Intelligent Camera
Hikvision Industrial Camera
DJI Inspire 2
DJI Phantom 4 Pro
DJI Spark

DJI Mavic Pro
Uniview IP Camera
Dahua Industrial Camera
Moto 360° Camera
Google Clips
More Analytics on the Edge → Faster Respond Time → More Controllability

Less Band-Width, Less Storage Required Upstream
Deployment issues with DL Inference

- DL Training and inference are essentially very different domains
  - Frameworks contain too much for training
  - Hardware requirements are very different (100W is pretty normal)
  - Performance goals are different (batch size, latency vs. throughput)
  - Environment is different (development vs. deployment)

- Hard to find solution that does DL inference properly
Open Visual Inference and Neural Network Optimization (OpenVINO™) Toolkit and Components

Inside the OpenVINO™ Toolkit

Intel® Deep Learning Deployment Toolkit

Model Optimizer
Convert and Optimize

Inference Engine
Optimized Inference

Component tools

Traditional Computer Vision—All SDK Versions
Optimized Computer Vision Libraries

- OpenCV*
- OpenVX*

Increase Processor Graphics Performance—Linux* Only

- Intel® Media SDK (Open Source Version)
- OpenCL™ Intel Integrated Graphics Drivers and Runtimes

Linux for FPGA Only

- FPGA Runtime Environment (RTE) (from Intel® FPGA SDK for OpenCL™)
- Bit Streams

Trained Models

GPU = CPU with Intel Integrated Graphics Processing Unit
VPU = Intel® Movidius™ Vision Processing Unit
Performance Improvement Using the OpenVINO™ Toolkit
Comparison of Frames per Second (FPS)

Faster Results on Intel Hardware

1Accuracy changes can occur w/FP16
The benchmark results reported in this deck may need to be revised as additional testing is conducted. The results depend on the specific platform configurations and workloads utilized in the testing, and may not be applicable to any particular user's components, computer system or workloads. The results are not necessarily representative of other benchmarks and other benchmark results may show greater or lesser impact from mitigations. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. Configuration: Intel® Core™ i7-6700K CPU @ 2.90 GHz fixed, GPU GT2 @ 1.00 GHz fixed Internal ONLY testing, performed 4/10/2018 Test v312 30 – Ubuntu* 16.04, OpenVINO™ 2018 RC4. Tests were based on various parameters, such as model used (these are public), batch size, and other factors. Different models can be accelerated with different Intel hardware solutions, yet use the same Intel software tools. Benchmark Source: Intel Corporation.
## Video Analytics in OpenVINO™ Toolkit

<table>
<thead>
<tr>
<th>Pre-trained Model</th>
<th>Supported Samples</th>
<th>CPU</th>
<th>Integrated Graphics</th>
<th>FPGA</th>
<th>VPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>face-detection-adas-0001</td>
<td>Interactive face detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>age-gender-recognition-retail-0013</td>
<td>Interactive face detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>head-pose-estimation-adas-0001</td>
<td>Interactive face detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>emotions-recognition-retail-0003</td>
<td>Interactive face detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vehicle-license-plate-detection-barrier-0007</td>
<td>Security barrier camera</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vehicle-attributes-recognition-barrier-0039</td>
<td>Security barrier camera</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>license-plate-recognition-barrier-0001</td>
<td>Security barrier camera</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>road-segmentation-adas-0001</td>
<td>Image segmentation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>semantic-segmentation-adas-0001</td>
<td>Image segmentation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>person-detection-retail-0001</td>
<td>Object detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>person-attributes-recognition-crossroad-0031</td>
<td>Crossroad</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pedestrian-detection-adas-0002</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pedestrian-and-vehicle-detector-adas-0001</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>person-detection-retail-00013</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>face-detection-retail-0004</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>face-person-detection-retail-0002</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>person-vehicle-bike-detection-crossroad-0078</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vehicle-detection-adas-0002</td>
<td>Any SSD-based sample</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Customer Success (just a few of many)

**Hikvision**
"Hikvision is collaborating with Intel on End-to-End AI/DL solutions from front end Movidius camera to backend servers. We are excited by the prospect of moving to Myriad X. We are also working with Intel on Intel's newly released toolkit CV SDK to achieve higher performance and shorten the development cycle. Hikvision is looking forward to building a strong, long term relationship with Intel to establish tech leadership in AI/DL based solutions." - *Dr. Pu Shiliang, Chief Scientist, Hikvision*

**Dahua**
"We are using Intel® Core i5 CPU along with Intel Arria®10 FPGA on our newly designed AI NVR product. By leveraging Intel's Computer Vision SDK toolkit, we are able to seamlessly implement our own customized deep learning solutions onto Intel's various platforms, perfect for applications with low latency requirements. Accelerating our system with Intel FPGAs has enhanced the deep learning capabilities for our real-time vision based solutions enabling a higher performance/watt/dollar. We are also using Intel Movidius Myriad product family for analytics at endpoints including surveillance and machine vision cameras. As technology and algorithms continue to evolve, we look forward to working with Intel to establish leadership in AI based solutions."

**GE Healthcare**
"Exceeded target performance by 6x...flexible, high-performance solution for a new era of smarter medical imaging. Our partnership with Intel allows us to bring the power of AI to clinical diagnostic scanning and other healthcare workflows in a cost-effective manner."

**GeoVision**
"GeoVision is going to release GV-VMS, GV-Smart Factory solution in Q3 which use Intel CV SDK to power up the performance of the systems."
- George Tai, CEO, GeoVision
Putting It All Together

- A major challenge is to get all the tools and libraries to work together in the best possible way to minimize development time and optimize system power/performance.
- A good way to abstract the workload is to use an end-to-end pipeline.
• Traditional Computer Vision:
  • OpenCV used in most computer vision projects
  • OpenCL™ as accelerator direct coding language
GT2  
Intel® HD Graphics  
24 EUs, 1 MFX  

GT3  
Intel® Iris™ Graphics  
48 EUs, 2 MFX  

GT4  
Intel® Iris™ Pro Graphics  
72 EUs, 2 MFX
VDBox: A BitStream decoder
Encoding acceleration

VEBox: HW acceleration for video enhancement/frame processing operations

EUs: “Execution Units”
3D rendering, Encode acceleration
API to access Intel hardware accelerated encoding, decoding & processing

- H.265 (HEVC)
- H.264 (AVC)
- MPEG-2 and more

- Resize, Scale, Deinterlace
- Color Conversion, Composition
- Denoise, Sharpen and more
  - Frame Rate Conversion
  - Color space conversions
  - Composition/alpha blending
  - Scaling

Benefits
- Cross OS and cross-platform API
- Support new processors without code changes
- Better performance & quality
- Performance tuning tools (specific platforms)
  - Media-SDK Analyzer
  - Media-SDK Tracer
- Open Sourced.
Decode ➔ Pre-Processing ➔ Inference ➔ Post-Processing ➔ Encode

Load model ➔ 4 frames
Decode → Pre-Processing → Inference → Post-Processing → Encode

GPU loading weights

GPU running DL model 4 times..
OpenVINO Customer Success stories

**GE Medical**: 14x performance improvement on Xeon for medical imaging neural network, beating target performance by 6x

**Philipps Medical**: 38x performance improvement in a lung segmentation app, 132x performance improvement in a bone-age-prediction model

**Hikvision**: 2x performance improvement in Xeon performance for inference server in security and surveillance. Reducing and even eliminating need for GPU

**Agent VI**: 10x performance improvement in their application on Core i3 CPU.
# OpenVINO™ toolkit Technical Specifications

## Intel® Platforms

<table>
<thead>
<tr>
<th>CPU</th>
<th>6th-8th generation Intel® Xeon® and Core™ processors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intel® Pentium® processor N4200/5, N3350/5, N3450/5 with Intel® HD Graphics</td>
</tr>
</tbody>
</table>

## Target Solution Platforms

### 6th-8th generation Intel® Core™ processor with Intel® Iris™ Pro graphics and Intel® HD Graphics

- Ubuntu* 16.04.3 LTS (64 bit)
- Microsoft Windows® 10 (64 bit)
- CentOS® 7.4 (64 bit)

### 6th-8th generation Intel® Xeon® processor with Intel® Iris™ Pro Graphics and Intel® HD Graphics (excluding e5 product family, which does not have graphics)

- Ubuntu 16.04.3 LTS (64 bit)
- Windows 10 (64 bit)
- CentOS 7.4 (64 bit)

### Intel® Arria® FPGA 10 GX development kit

- Ubuntu* 16.04.3 LTS (64 bit)
- CentOS 7.4 (64 bit)

### Intel® Programmable Acceleration Card with Intel® Arria® 10 GX FPGA operating systems

- Ubuntu* 16.04.3 LTS (64 bit)
- CentOS 7.4 (64 bit)

## VPU

- Intel® Movidius™ Neural Compute Stick

## Development Platforms

- 6th-8th generation Intel® Core™ and Intel® Xeon® processors

## Additional Software Requirements

<table>
<thead>
<tr>
<th>Linux* build environment required components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OpenCV 3.4</strong> or higher</td>
</tr>
<tr>
<td><strong>CMake</strong> 2.8 or higher</td>
</tr>
<tr>
<td><strong>GNU Compiler Collection (GCC) 3.4</strong> or higher</td>
</tr>
<tr>
<td><strong>Python</strong> 3.4 or higher</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Microsoft Windows* build environment required components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intel® HD Graphics Driver</strong> (latest version)™</td>
</tr>
<tr>
<td><strong>Intel® C++ Compiler 2017 Update 4</strong></td>
</tr>
<tr>
<td><strong>OpenCV 3.4</strong> or higher</td>
</tr>
<tr>
<td><strong>CMake</strong> 2.8 or higher</td>
</tr>
<tr>
<td><strong>Microsoft Visual Studio® 2015</strong></td>
</tr>
</tbody>
</table>

## External Dependencies/Additional Software

View Product Site, detailed System Requirements

---

1 Graphics drivers are required only if you use Intel® Processor Graphics (GPU).