embedded VISION SUMMIT

# Better Image Understanding through Better Sensor Understanding

Michael Tusch, CEO Apical

2 October 2013

apical

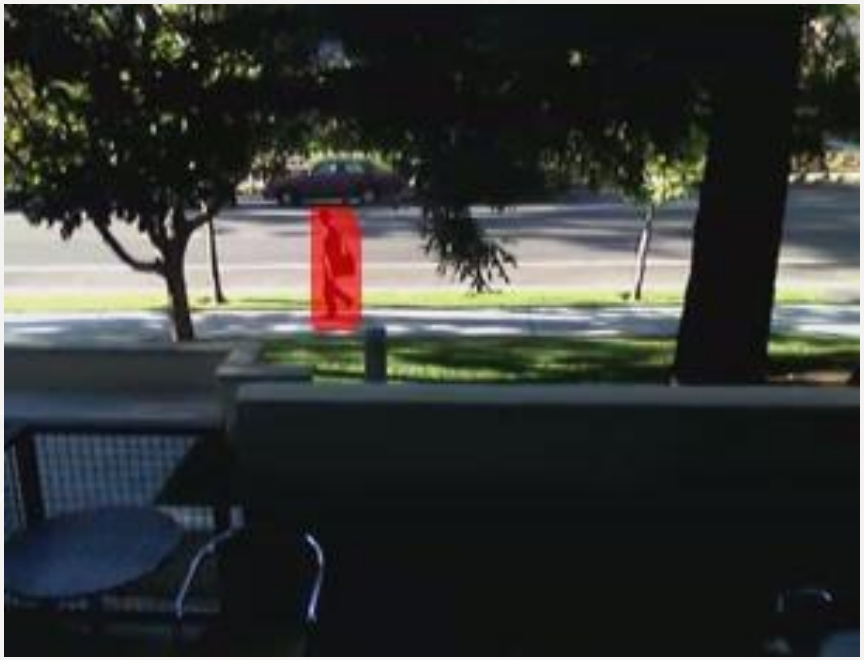
# What does Apical do?

- We design camera subsystems
  - Smartphones
  - Videoconferencing
  - Surveillance
  - Automotive
- Most of our designs are integrated into SoCs
- We're really interested in pushing the state of the art




# Analysis of this kind of data is hard

- Video is compressed
- Resolution is low
- Sensor noise levels are unknown
- Accuracy of colors is unknown



# Vision and Visualization

- ISP = Image Signal Processor
  - It's what turns a raw sensor image into an image that looks realistic to us
  - Images are formatted for display requirements
  - ISPs are not usually designed with embedded vision in mind

- Sometimes, **vision** and **visualization** are in direct conflict
  - Non-linear demosaic, color management, adaptive tone mapping, noise reduction, sharpening give great pictures but affect sensor data unpredictably
  - See my April Embedded Vision Summit presentation, "Exploiting Synergy Between Image Improvement and Image Understanding Algorithms" (http://bit.ly/11Ev5fB)

For **vision**, we want to use the sensor as a **calibrated instrument**

For **visualization**, we want to transform pixels to produce **natural imagery**

# Our theme here

By understanding how the data has been captured, how much better can we make our embedded vision system?

Let's talk about people detection/tracking…

… in difficult conditions

# How does noise affect detection reliability?

- It's harder to detect and classify objects in noisy images, of course!
  - night-time operation
  - compact sensor modules
  - fast moving objects

- But **why** is it so hard?
- Humans don't have so much trouble…

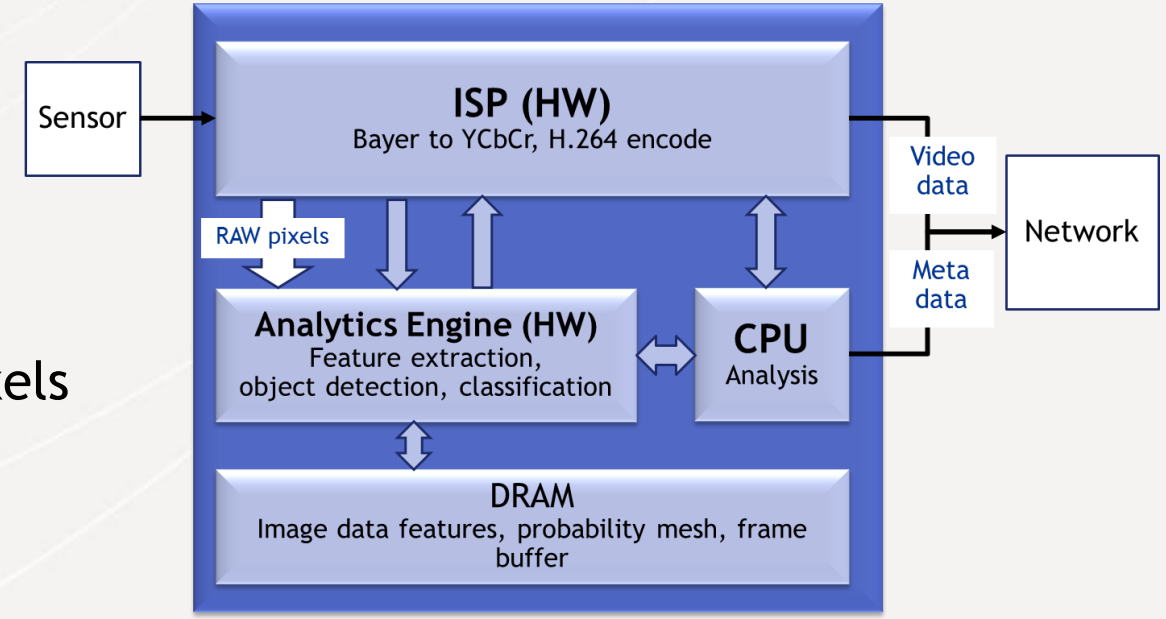
# Our test system

We want to remove all sources of compromise:

- Let's process all pixels at full frame rate at 1080P60 using multiple HOG/SVM classifiers and scales—in real time
- Let's use RAW sensor data to give maximum information

# Our test system

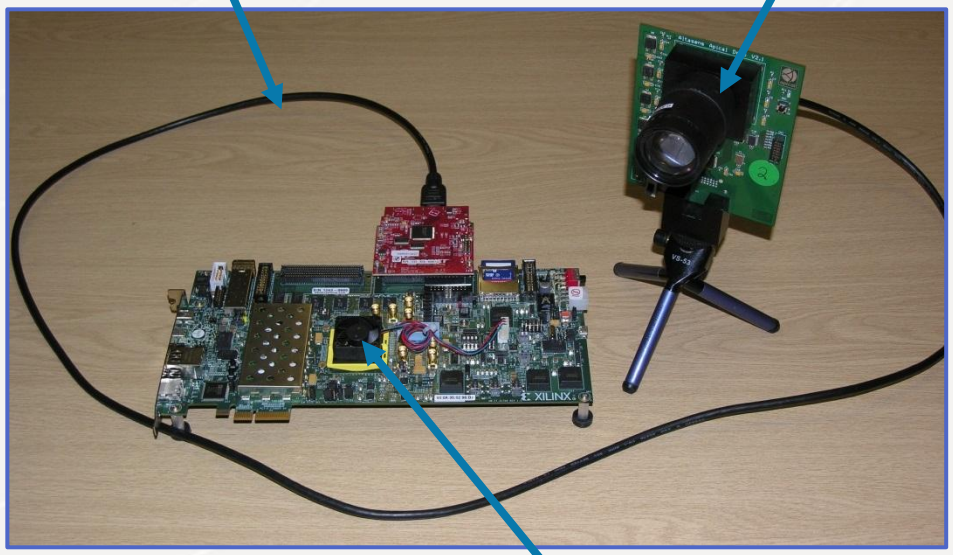
- Real-time HD
- Up to 16 classifiers running simultaneously
- Scales from 45 to 550 pixels
- Accurate tracking



# Prototype system based on Xilinx Zynq 7045 SoC

RAW sensor data over HDMI

1080P60 sensor



ISP + vision subsystem on FPGA fabric + firmware on ARM Cortex

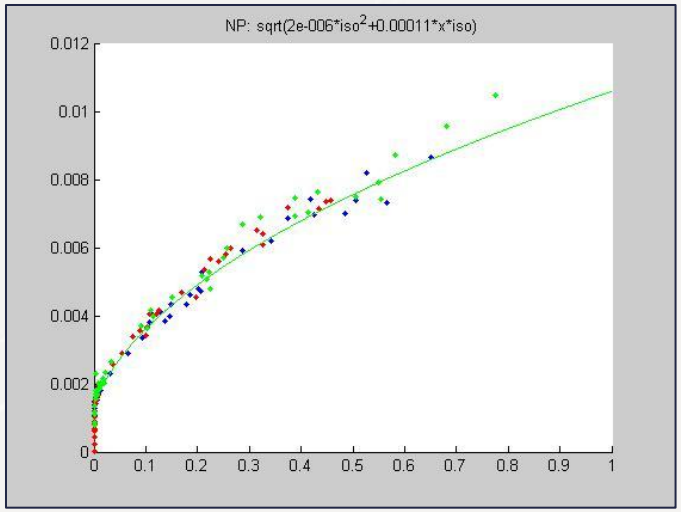
# Our experiment

- Let's analyse detection accuracy with increasing ISO (noise)
- Let's compare a "blind" method with one which is noise-aware
  - We'll use the best possible "blind" method (no compression, no non-linear adjustments etc.)
  - We'll solely add information about noise characteristics

# Noise analysis

- Sensor signal-to-noise depends on luminance
  - Dark regions are more noisy than bright
- Use to adjust detection probability of objects in shadows and highlights
- Reduce false positives due to noise

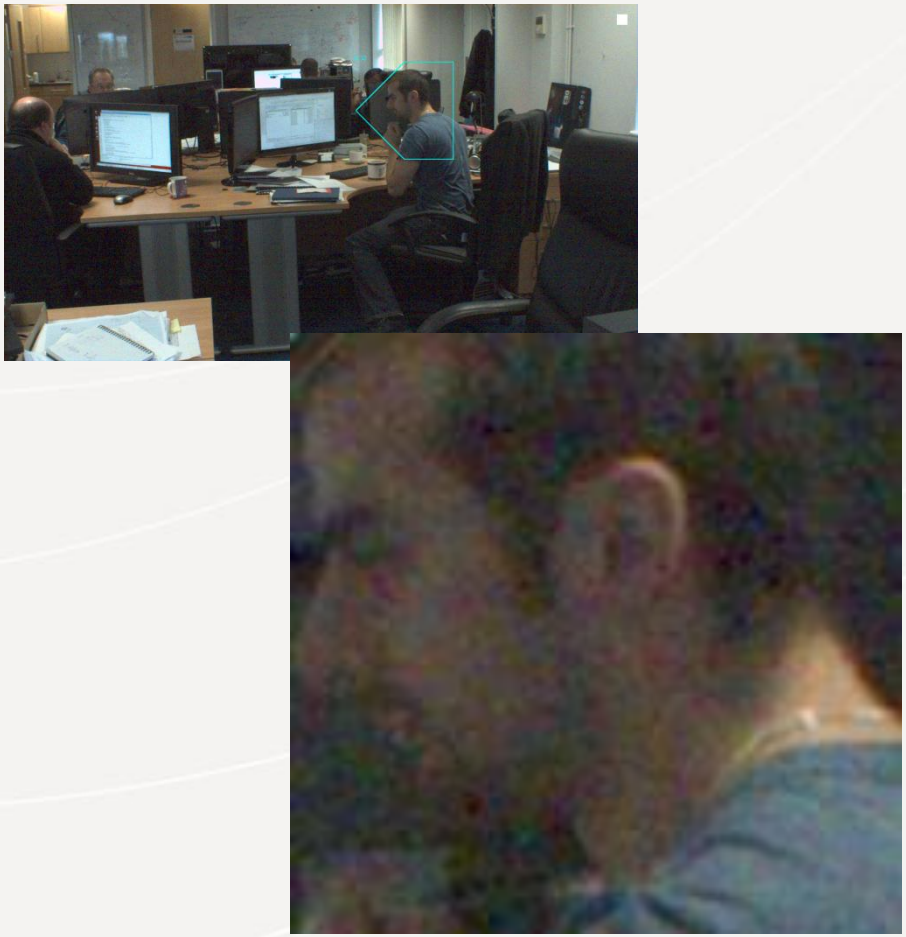
Sensor noise profile



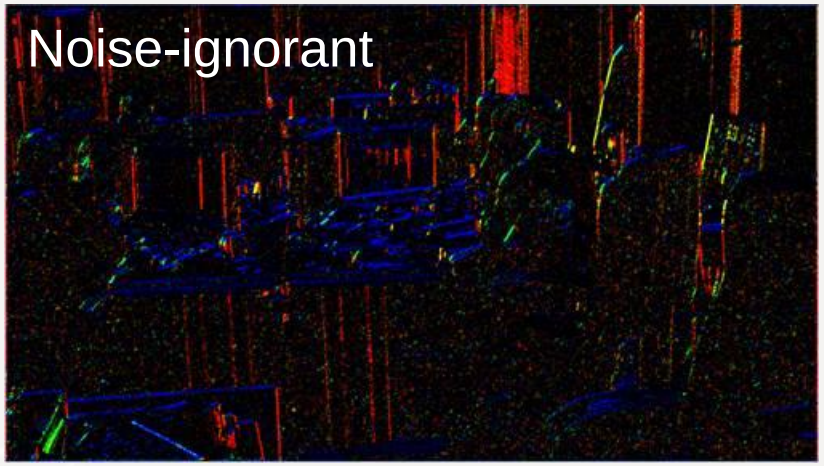

Analog noise
line, column, pattern
*Gaussian*

Quantization noise
Photon
*Poisson*

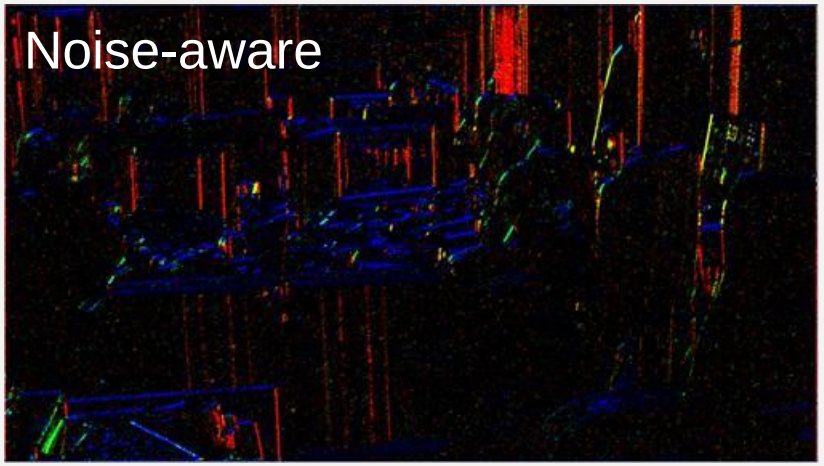
# HOG edge statistics: incorporation of noise



Noise-ignorant



Noise-aware



# ISO100

# ISO1600 (noise ignorant)

# ISO1600 (noise aware)

# Sample results: detection

| | ISO 100 | | ISO 1600 | |
|---|---|---|---|---|
| | Noise aware | Noise ignorant | Noise aware | Noise ignorant |
| Detected faces | 1255 | 1241 | 1161 | 1089 |
| Weak/missed faces | 15 | 29 | 56 | 128 |
| Upper body detected | 1248 | 1233 | 1193 | 1070 |
| Upper body weak/missed | 14 | 29 | 33 | 67 |
| False positives | 0 | 0 | 4 | 21 |
| Missed people | 0 | 0 | 2 | 16 |
| Track errors | 0 | 1 | 1 | 1 |

# Sample results: after grouping

| | ISO 100 | | ISO 1600 | |
|---|---|---|---|---|
| | Noise aware | Noise ignorant | Noise aware | Noise ignorant |
| People detection rate | 100% | 99.8% | 99.7% | 97.0% |
| Face detection rate | 97.6% | 95.6% | 91.0% | 83.0% |
| Upper body detection rate | 97.8% | 95.4% | 94.6% | 91.4% |
| False positive rate | 0% | 0% | 0.3% | 1.7% |

# Conclusion

- Incorporating sensor noise characteristics increased our detection accuracy
- Noise is one of several characteristics we need to build in to go from “video camera” to “calibrated instrument”
- State-of-the-art HOG/SVM-based detection and classification algorithms can work with high reliability even in noisy conditions, provided noise is accurately modelled

# Why are we doing all this?

# Understanding behavior

# Understanding behavior

# Resources

- **Check out our live system in the demo room:** www.apical.co.uk/assertive-vision
- **How Does Camera Performance Affect Analytics?** http://www.embedded-vision.com/platinum-members/apical/embedded-vision-training/documents/pages/camera-performance-analytics
- **April 2013 Embedded Vision Summit Technical Presentation: "Exploiting Synergy Between Image Improvement and Image Understanding Algorithms," Michael Tusch, Apical** http://www.embedded-vision.com/platinum-members/apical/embedded-vision-training/videos/pages/april-2013-embedded-vision-summit
- **September 2012 Embedded Vision Summit Presentation: "Harnessing Hardware Accelerators to Move from Algorithms to Embedded Vision," Michael Tusch, Apical** http://www.embedded-vision.com/platinum-members/apical/embedded-vision-training/videos/pages/september-2012-embedded-vision-summit-presentation
- **September 2012 Embedded Vision Summit Presentation: "Image Sensor Options and Trends for Embedded Vision," Eric Gregori, BDTI** http://www.embedded-vision.com/platinum-members/bdti/embedded-vision-training/videos/pages/september-2012-embedded-vision-summit-pr
- **Apical New Product Presentation at the March 2012 Embedded Vision Alliance Summit** http://www.embedded-vision.com/industry-analysis/video-interviews-demos/2012/06/15/apical-new-product-presentation-march-2012-embed