FotoNation

a Tessera company

NEXT GENERATION IMAGING FOR SMARTPHONES

Version 2.1

Who am I?

- Original Founder of FotoNation (1990’s)
  - Member Board of Governors of IEEE Consumer Electronics Society
  - Editor-in-Chief of IEEE Consumer Electronics Magazine
  - University vice-Dean & Professor
    - » E-Mail:

- pcorcoran@fotonation.com
- dr.peter.corcoran@ieee.org
- peter.corcoran@nuigalway.ie
- cesmagazine@ieee.org

CONTEXTS – WHO IS FOTONATION?

our mission is to innovate and deliver embedded imaging products, to improve picture quality, enhance the user experience and enable still and video analytics

500M smartphones shipped
60% market share in Tier 1 Smartphones
over **1.2B** imaging products worldwide
9/10 Tier 1 DSC's use our products

Licensing Business

TESSERA

NASDAQ: TSRA

Computational Imaging Business

FotoNation®

Founded in 1997 as a solution provider in connectivity business.

- In 2003 the imaging processing group has been formed
- In 2004 FotoNation’ Red Eye technology is the **first computational imaging technology** running in an embedded device, on a Nikon camera.

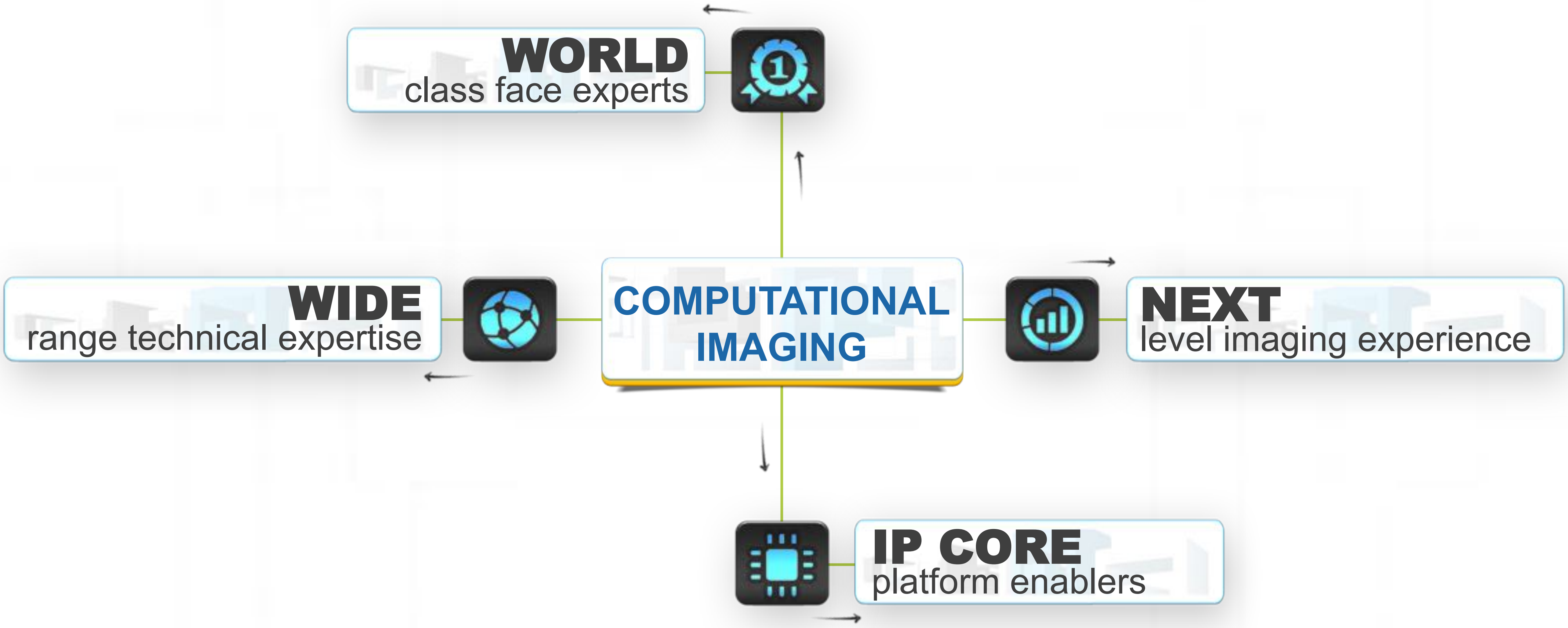
By 2007 all the main DSC & DSLR makers use FotoNation as the **imaging partner of choice**

In 2008 Tessera acquires FotoNation to build the smart technologies on top of smartphone camera modules (under DigitalOptics Corporation division)

FotoNation continued to provide technologies in the DSC/DSLR markets and became **leader** in providing technologies in the smartphone market as well.

Today, FotoNation works with the **top smartphone makers** and **embeds DSLR image quality into smartphone cameras**.

key areas of expertise

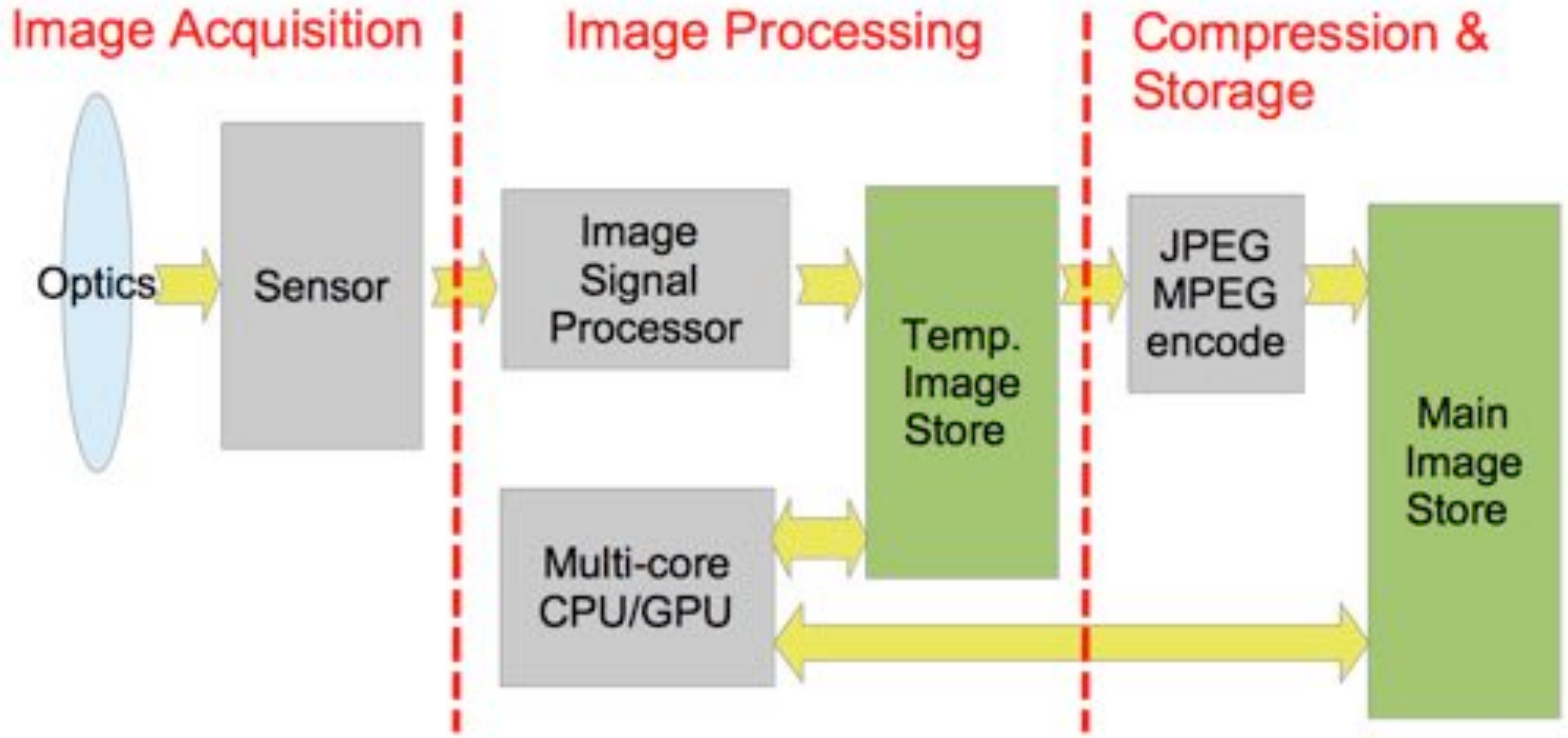


TECHNOLOGY CONTEXTS

Technology Contexts #1
Sensors & Computational Capabilities

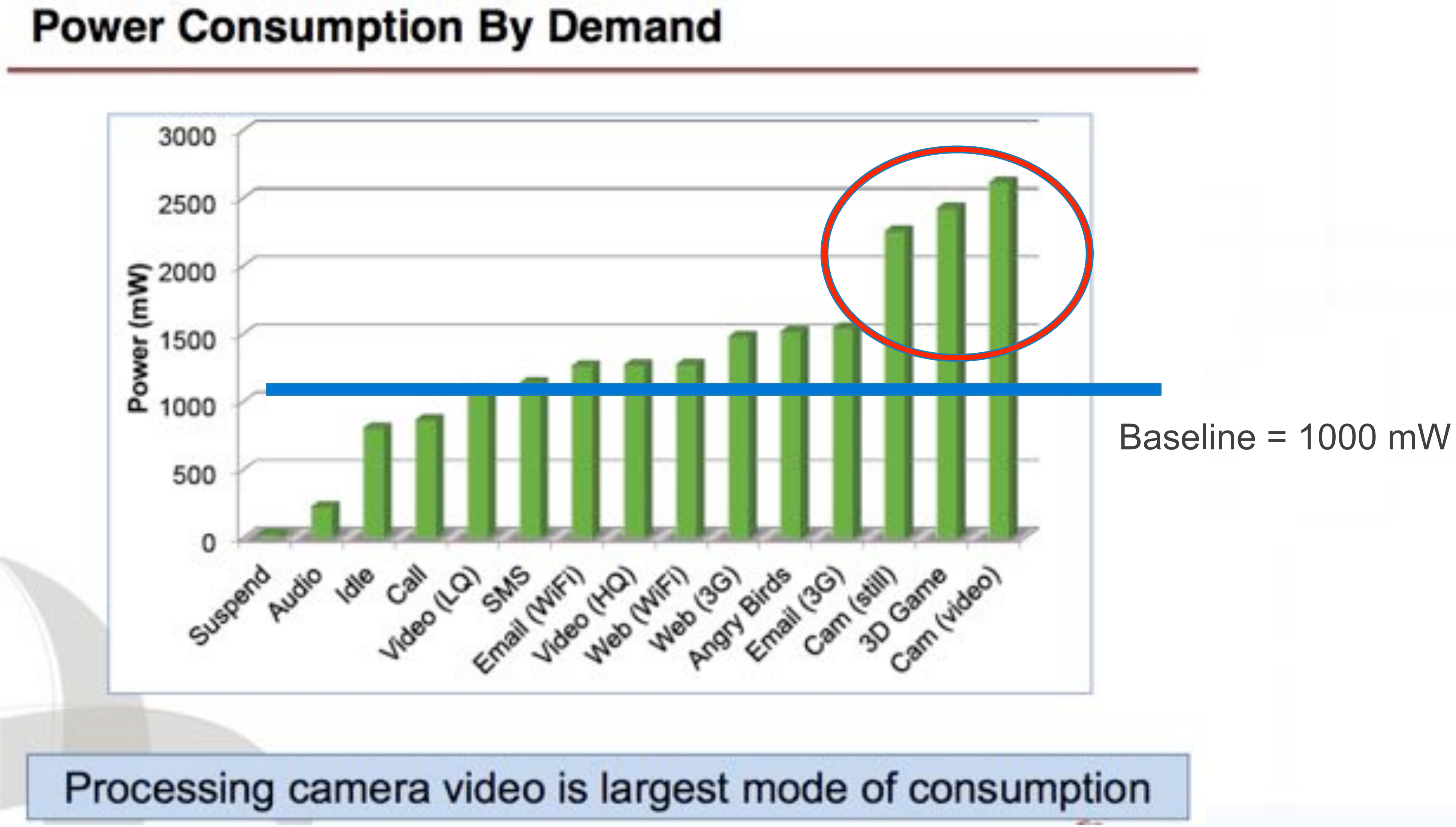
- Smartphone Technology Marches on …
  - Image Sensor Technology up to 41 Megapixels
    (e.g. Nokia Pureview)
  - Multicore CPU, Multi-GPU Architectures
    (ARM is dominant, but Intel has woken up …)
  - Multi-Camera/Multi-Optical Solutions Appearing in the Market

Technology Contexts #2
Image Acquisition & Processing



Technology Contexts #3

Power Consumption (pre-4K video!)

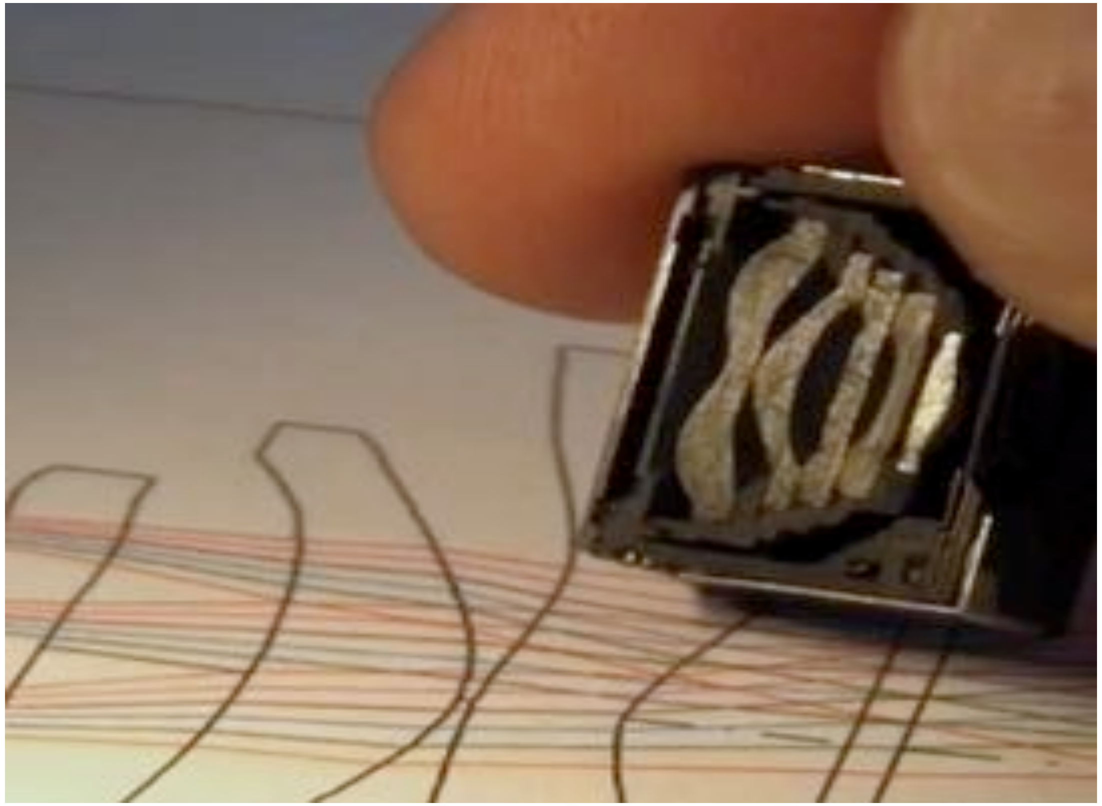


CHALLENGES

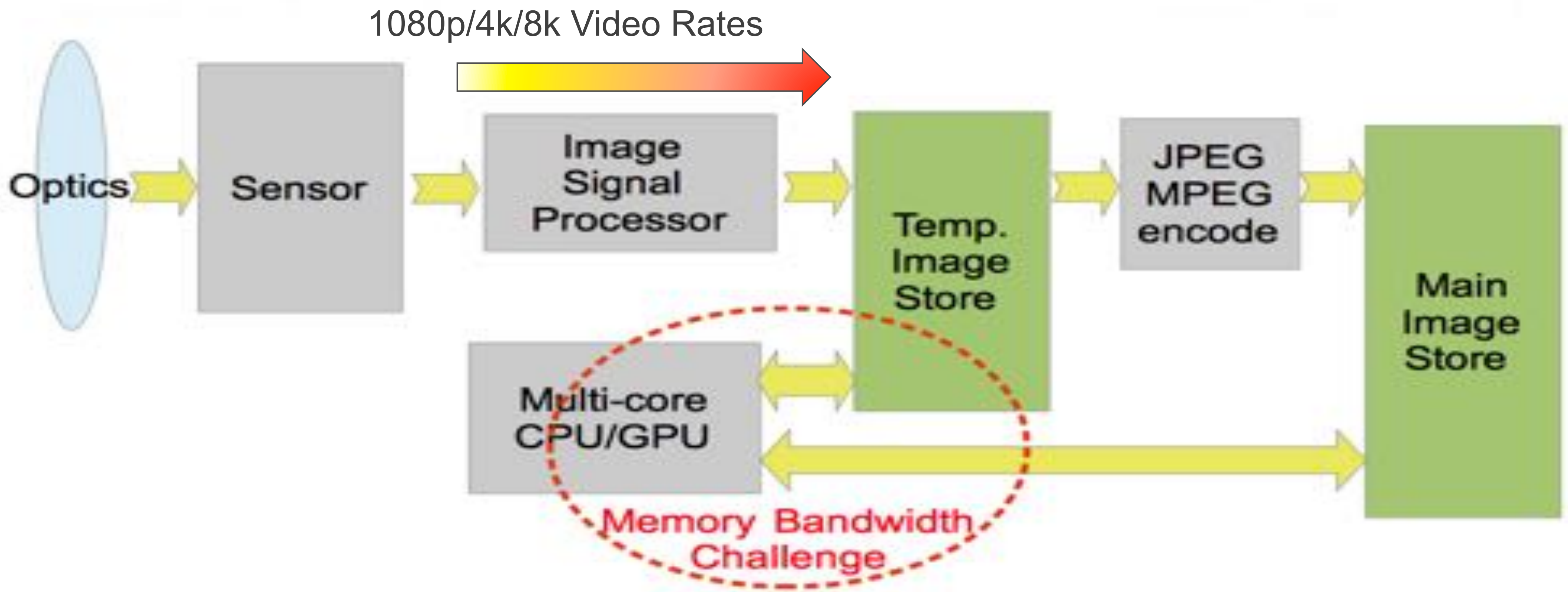
Challenge #1 - Smartphone Optical Challenges

(Lenses are manufactured from high quality plastic – Nokia 808 lens shown.)

- Lens Qualities
  - complex shapes – need plastic
  - a single moving element
    - VCM barrel or MEMS single-lens
- Optical Challenges
  - short focal length (4-5 mm)
  - small sensor size
    - Near diffraction limited
  - Use cases => WFoV optics



Challenge #2 - Memory Bandwidth Challenges



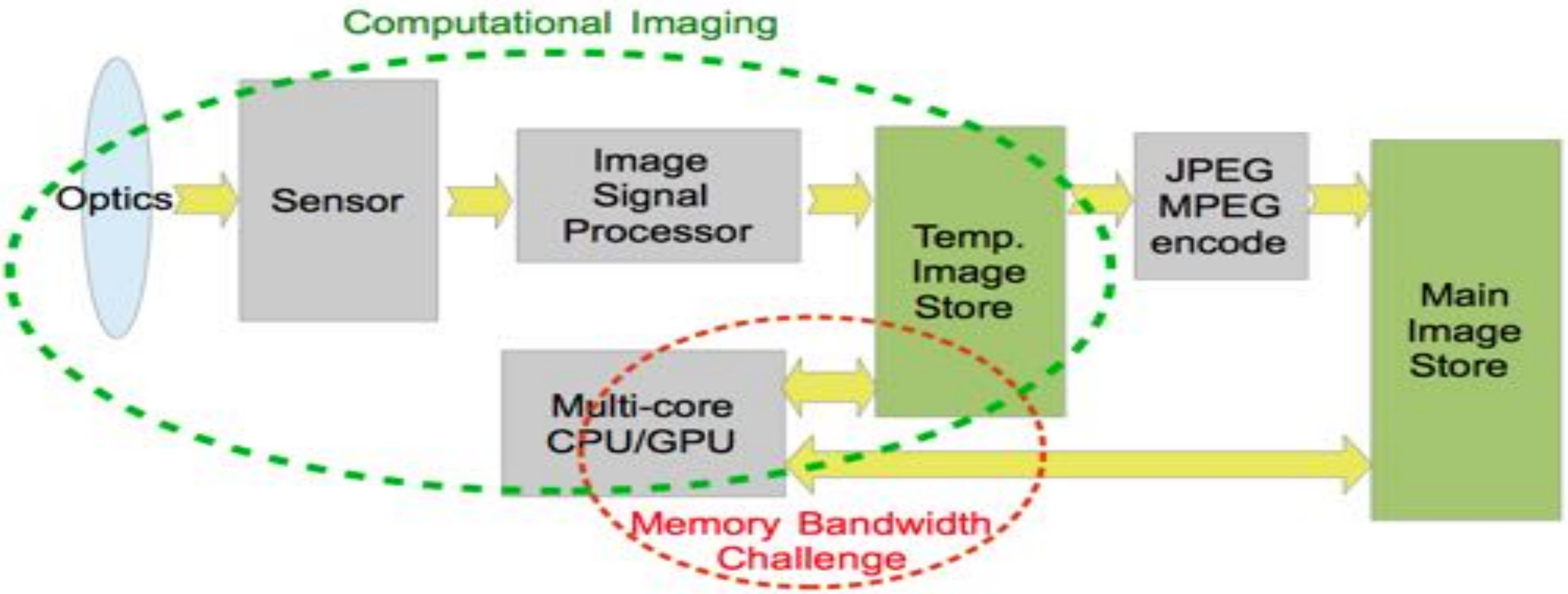
Challenge #3 – Power Consumption Challenges – GPU’s

Do you want to let a Hot-Rod designer have access to your phone’s battery … ?

SOLUTIONS

The Solution #1 – Computational Imaging

Early solutions used ‘software only’ approaches – these don’t really address the “Bandwidth Challenge”
But they did help explore what was possible on ISP and Applications Processor hardware.

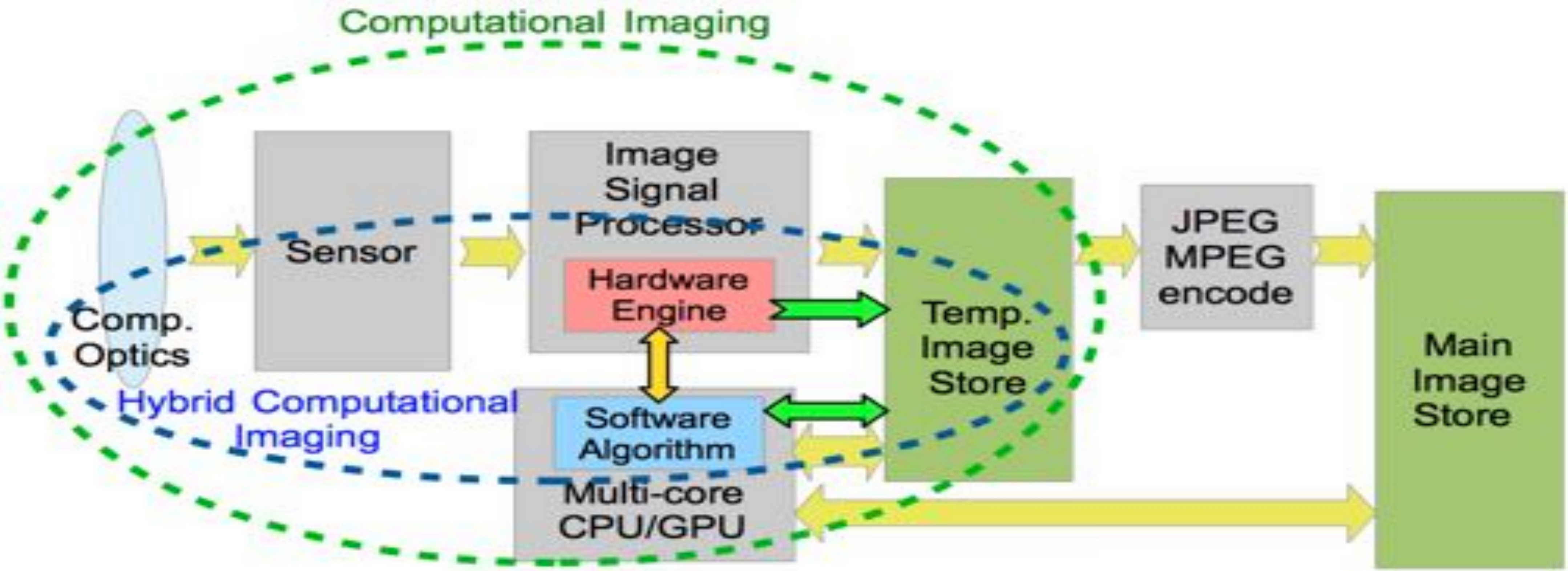


College of Engineering, Science & Informatics

The Solution #2 – Hybrid Comput'l Imaging

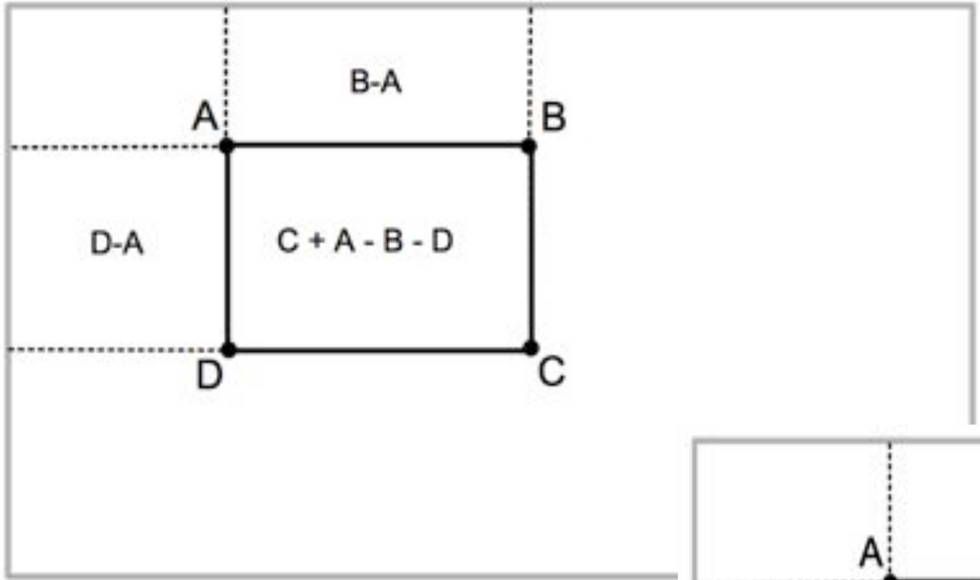
Algorithm Partitioning into hardware & software
e.g. Integral Image generated in hardware

A natural evolution was to extract the parallel elements of algorithms into hardware.



EXAMPLE #1 – AHFD
FACE DETECTION & TRACKING

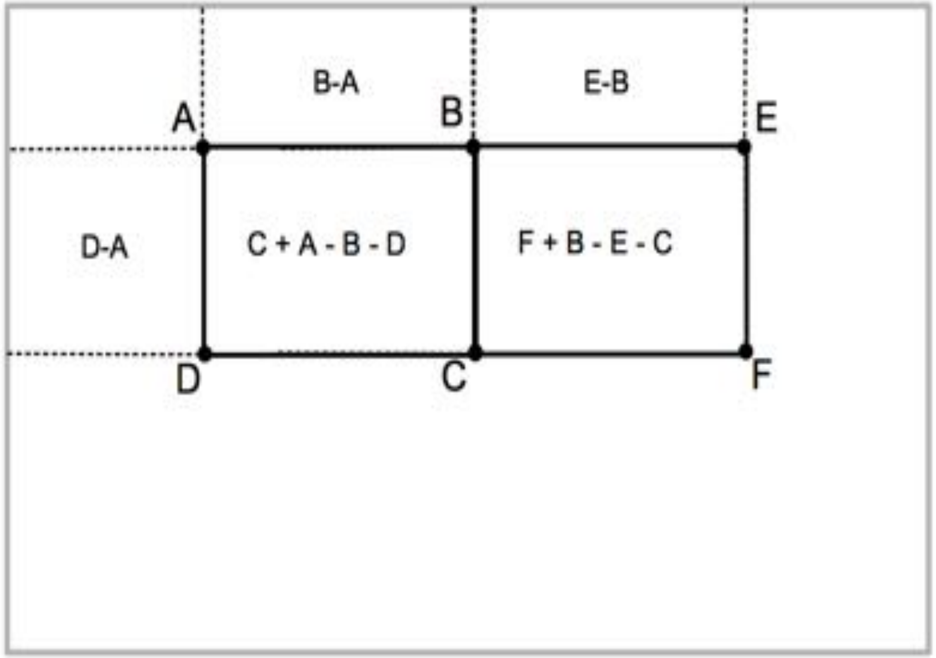
Face Detection Begins with the Integral Image



Each point of Integral Image is the sum of all previous pixel luminance values;

So determining the 'Integral Luminance' within a rectangular ares becomes a simple set of arithmetic operations:

$$
\begin{aligned}
ABCD &= C - (B-A) - (D-A) - A \\
&= C-B-D+2A-A \\
&= C+A-B-D
\end{aligned}
$$



And
BEFC = F+B-E-C

ABCD > BEFC ?

or

ABCD < BEFC ?

Face Detection uses simple rectangular Classifiers (Haar)

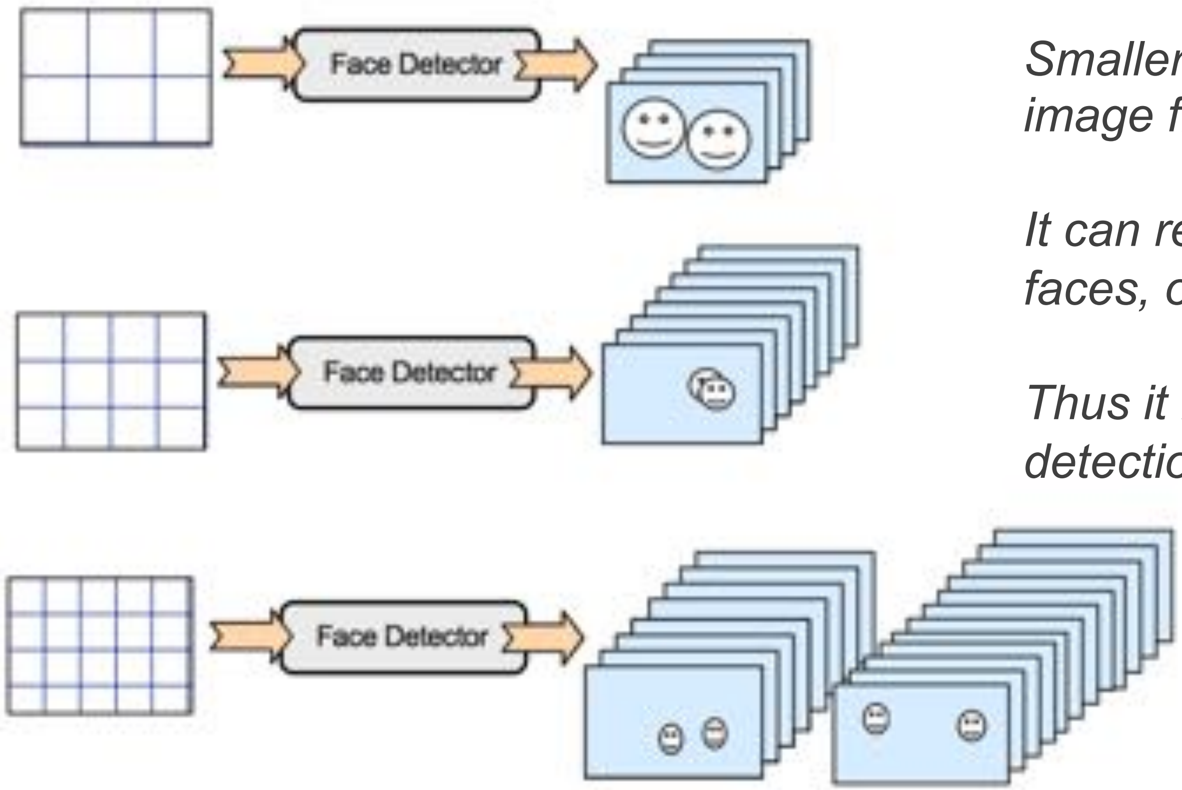
Larger classifiers detect the main structure of a face;

Smaller classifiers match with local features – nose, mouth, eyebrows;

If enough classifiers match at a sufficient level, statistically you can confirm the presence of a face ;

Computationally expensive because you have to scan many windows

Scan Step Size	Full HD (1080p)	HD (720p)	480p	240p
4	125875	55125	23920	8525
8	45188	13659	5928	2079
12	13904	6090	2622	936



Smaller scanning window size means more scanning per image frame;

It can require many image frames to completely scan for faces, or it takes lot of computational hardware;

*Thus it is better to **track faces** and run **two processes** – detection & tracking – at the same time ...*

Tracked (Locked) faces are likely to appear in the next image frame:
(i) in approximately the same location (if the face was moving you can estimate the offset)
(ii) with approximately the same size (so no need to scan all face sizes – just proximate ones!)
And you can 'skip' regions with 'known' faces in your detection process

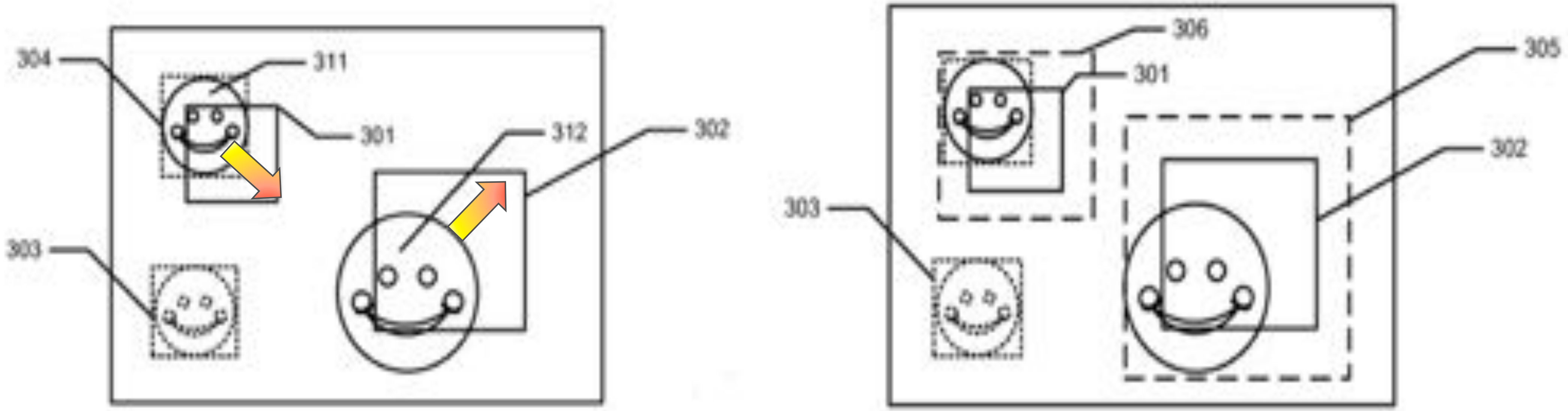
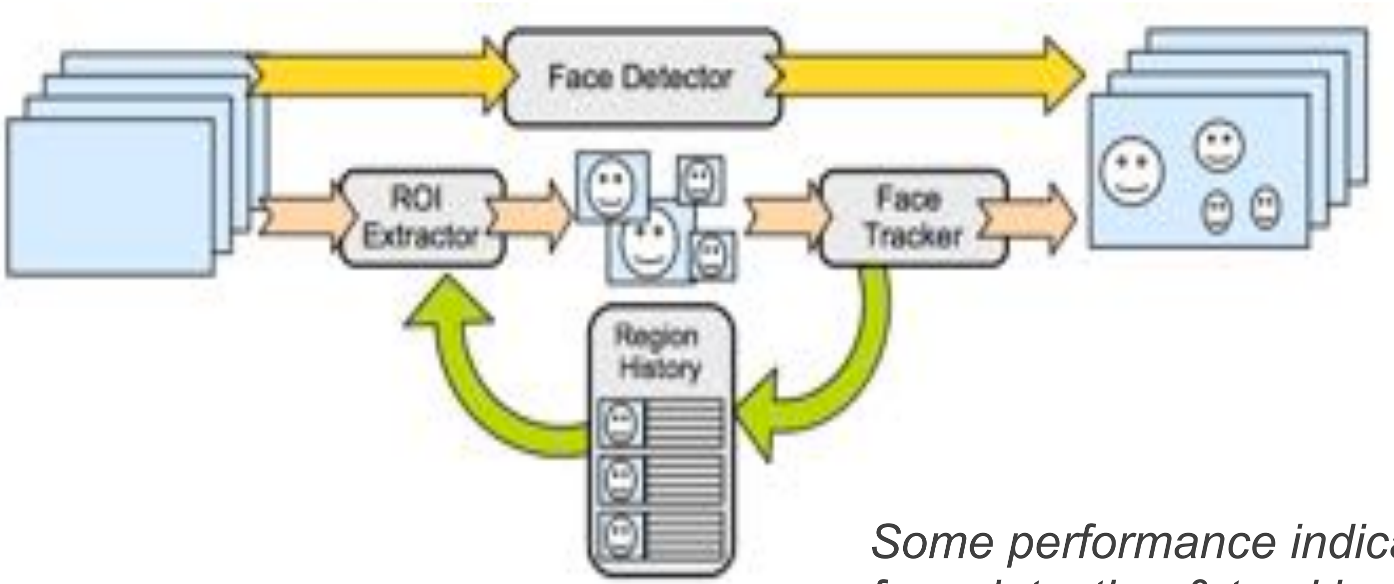


Figure 13: Predicted regions – where we expect to find a face in the next image frame.
(Original image from author – taken from Corcoran, P., Steinberg, E., Petrescu, S., Drimbarean, A., Nanu, F., Pososin, A., & Biglol, P. (2008). U.S. Patent No. 7,315,631. Washington, DC: U.S. Patent and Trademark Office.)

So a Face Tracker follows "Locked" faces while a Face Detector continually looks for 'new' faces.

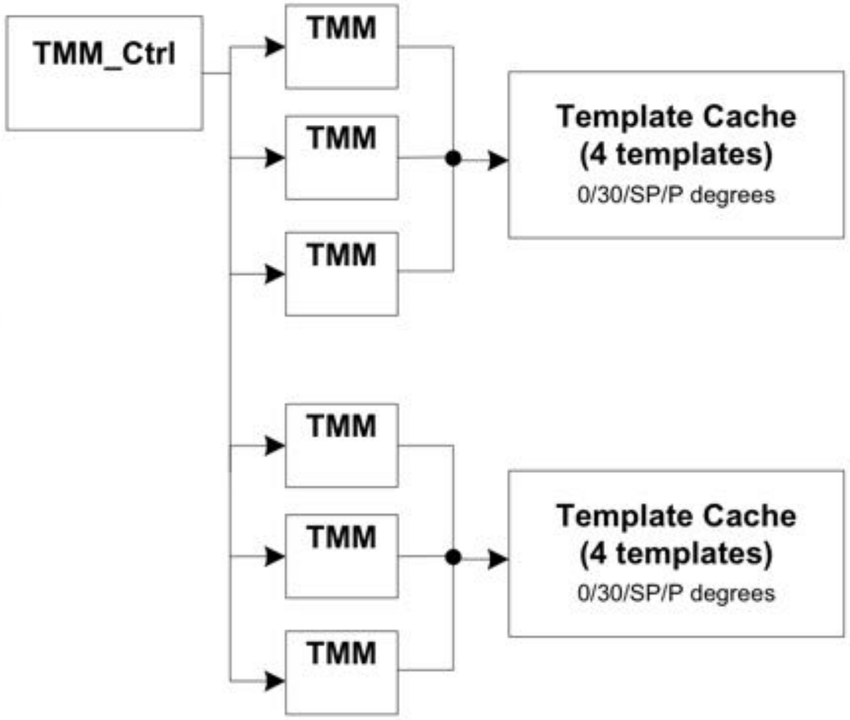


Some performance indicators for software and hardware accelerated face detection & tracking:

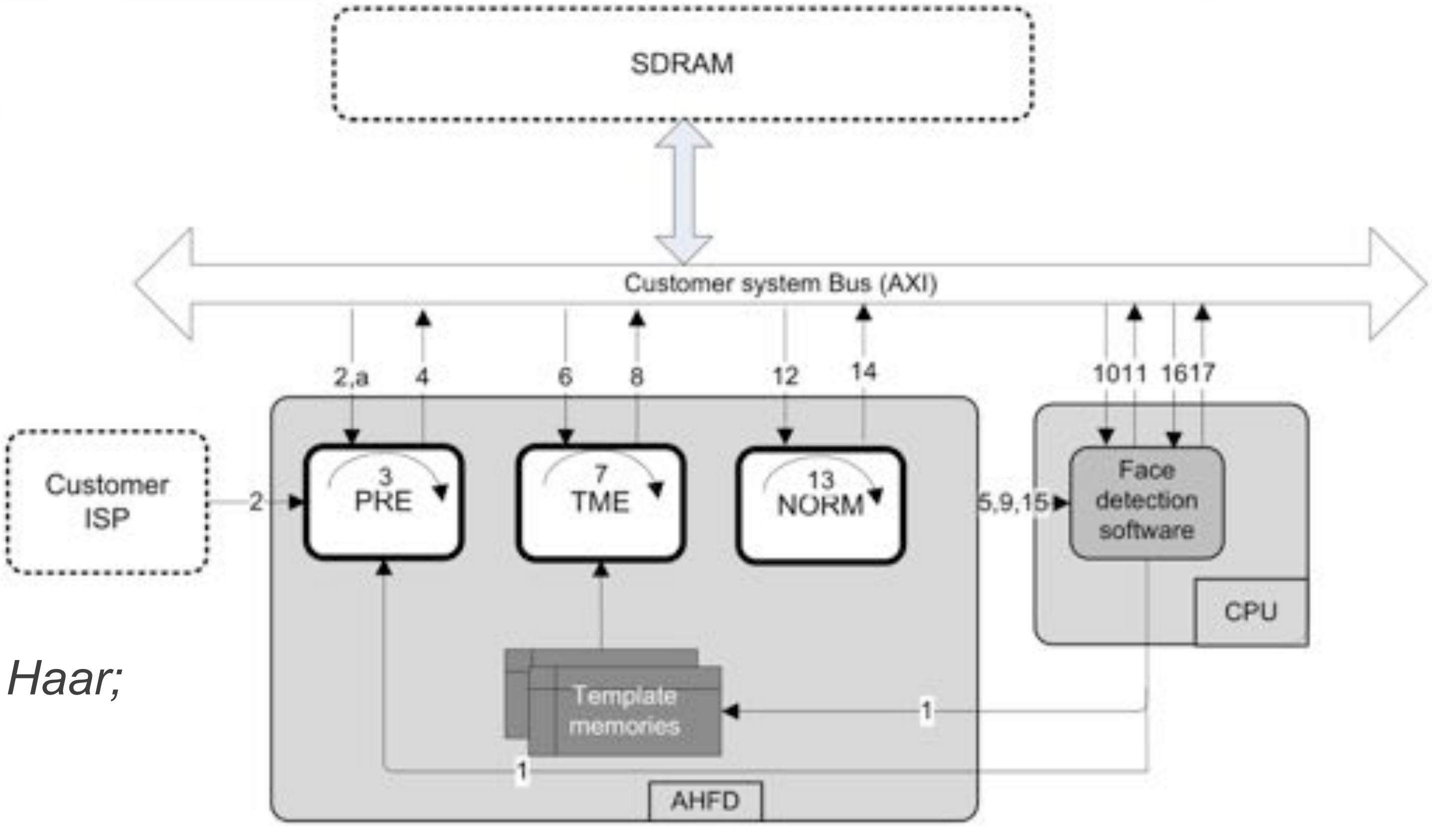
Naturally this can be implemented in Hardware

Configuration	**Object Detectors** ***(in parallel in one image frame)***	**Face Size in Pixels** ***(full HD resolution @60fps)***	**Face Size in Pixels** ***(4K resolution @30fps).***	**Detection Ratio**	**False Positive Ratio**
Software only (QVGA)	0.2 to 1	63x63	126x126@30fps	>=90%	<= 0.5%
Software only (VGA)	0.2 to 1	32x32	64x64@30fps	>=90%	<= 0.5%
Hardware accelerated (fullHD)	6	18x18@30fps, 33x33@60fps 50x50@120fps	36x36@30fps 66x66@60fps	>=90%	<= 0.5%

*At the core is a **scalable** template matching engine;*



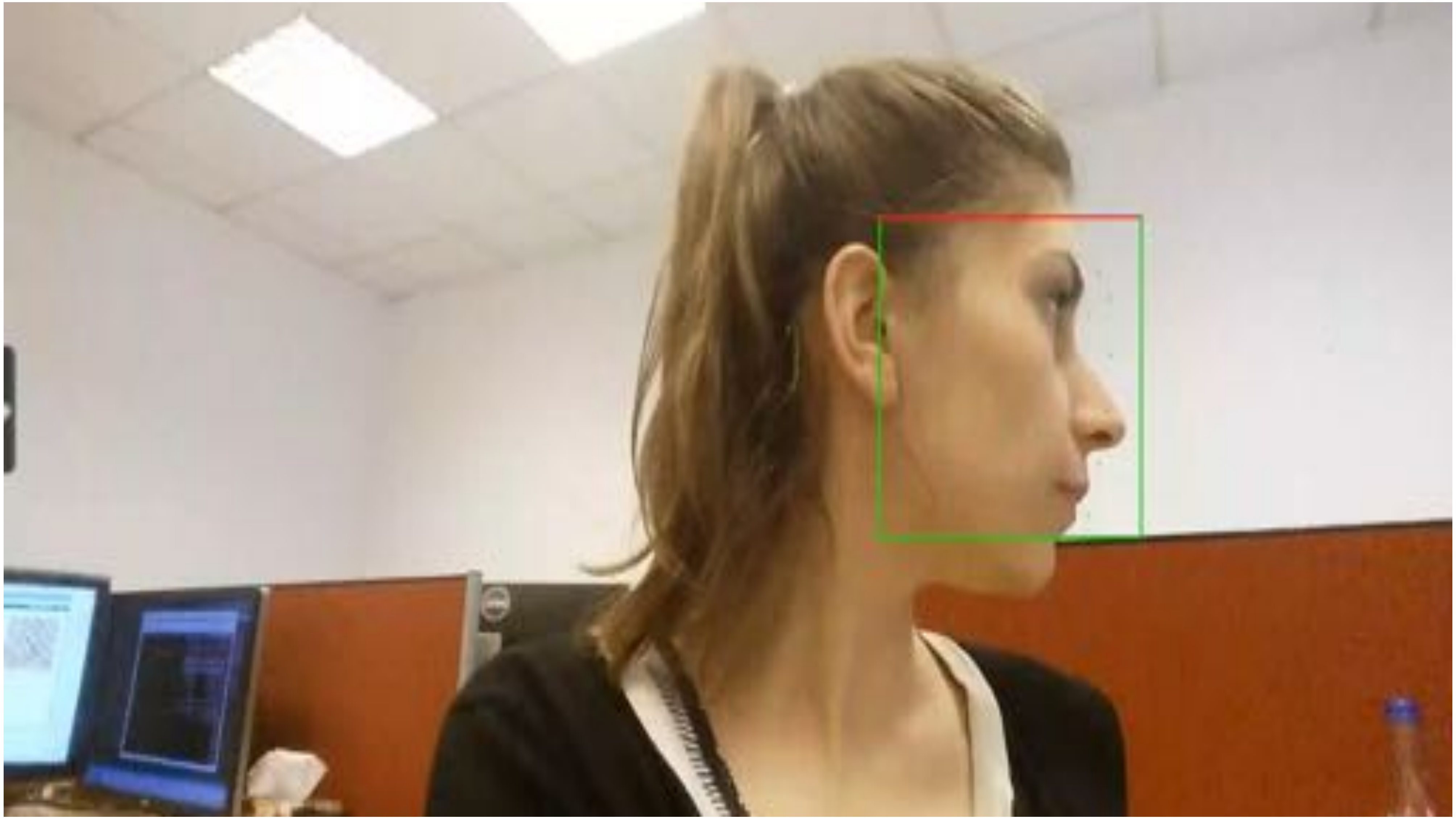
This is embedded in a smart pre-scaling & cacheing infra-structure that processes image data as it is clocked through the Image Signal Processor ;



The template engine can do much more than Haar;
It can do various HoG, LBP, Census, etc ...

AHFD Hardware - Multi-Face Detection

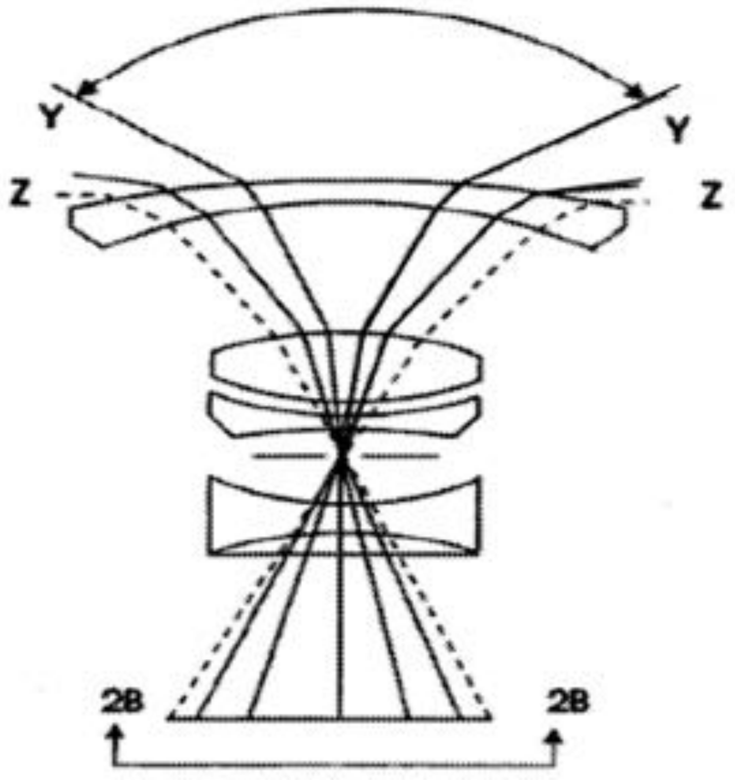
Multi-Pose & Position Face Detection/Tracking

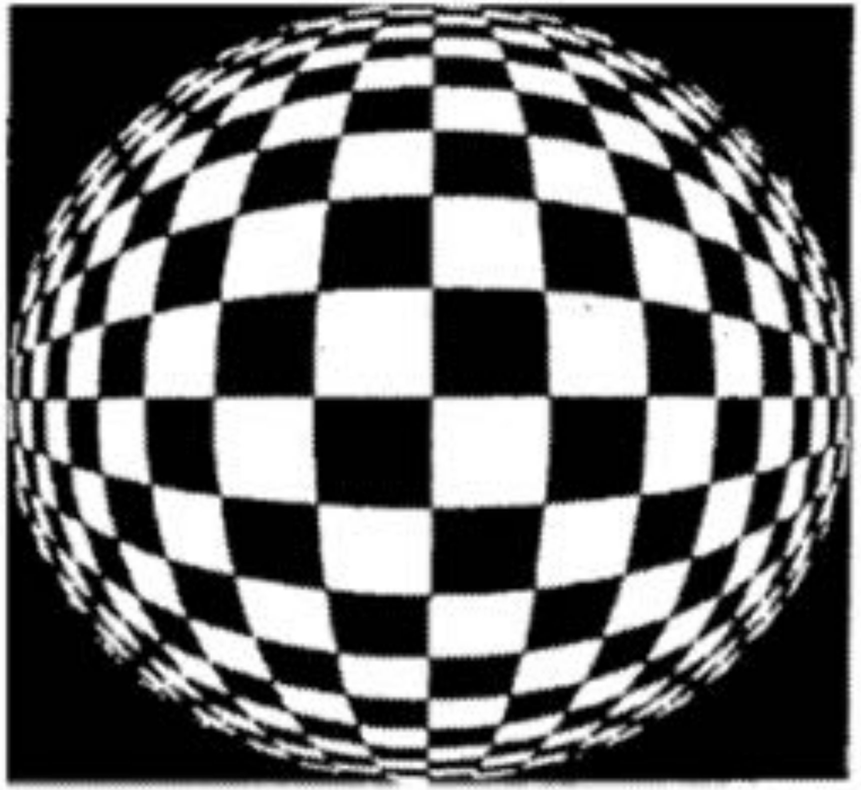


Full-Body Detection in Crowded Scenes

EXAMPLE #2 – SGDE
SMART GEOMETRIC DISTORTION ENGINE

The Problem to be Solved? (1)



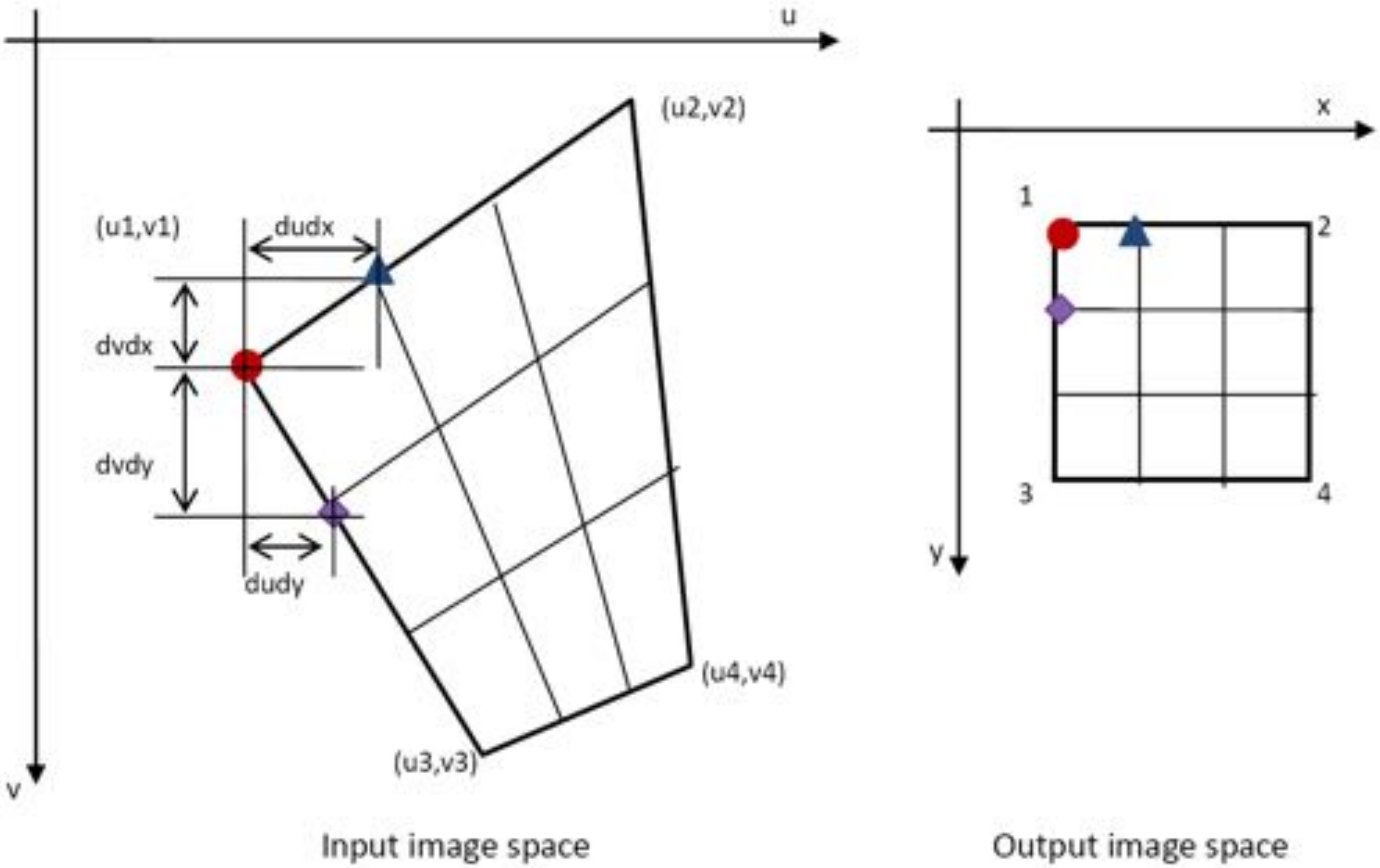


Or given this …

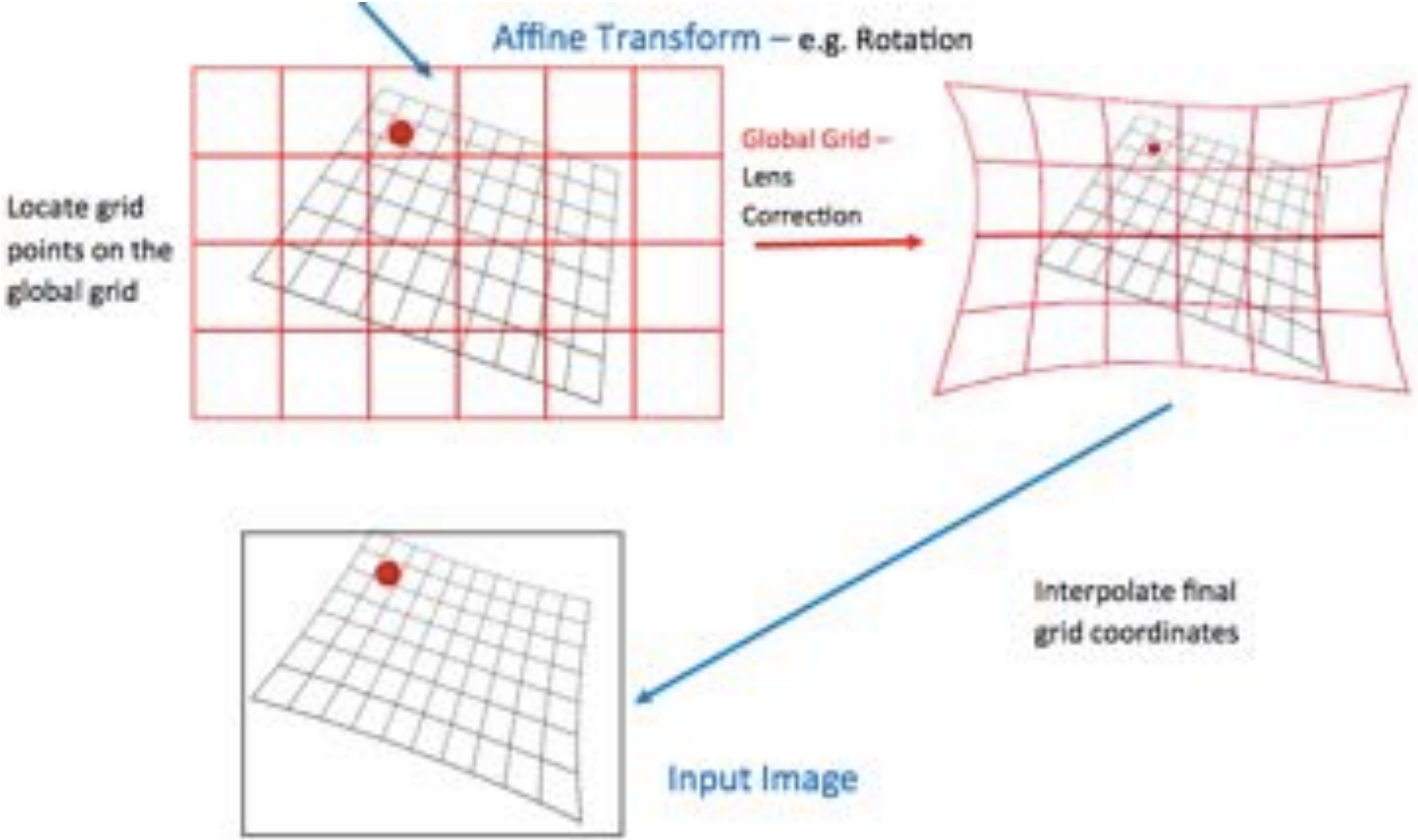
How can I get this?

MAPPING THE IMAGE FRAME

From sensor to output …

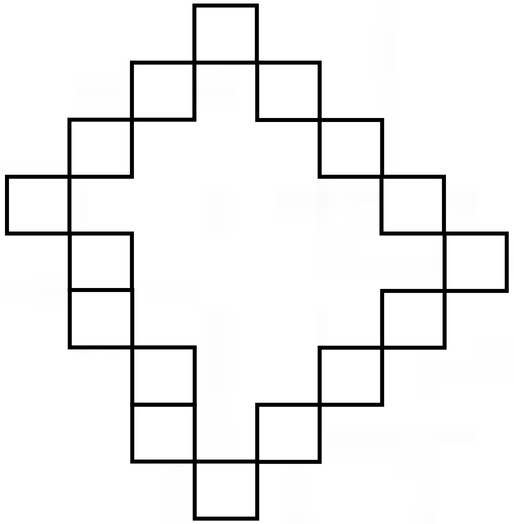


Is not easy …



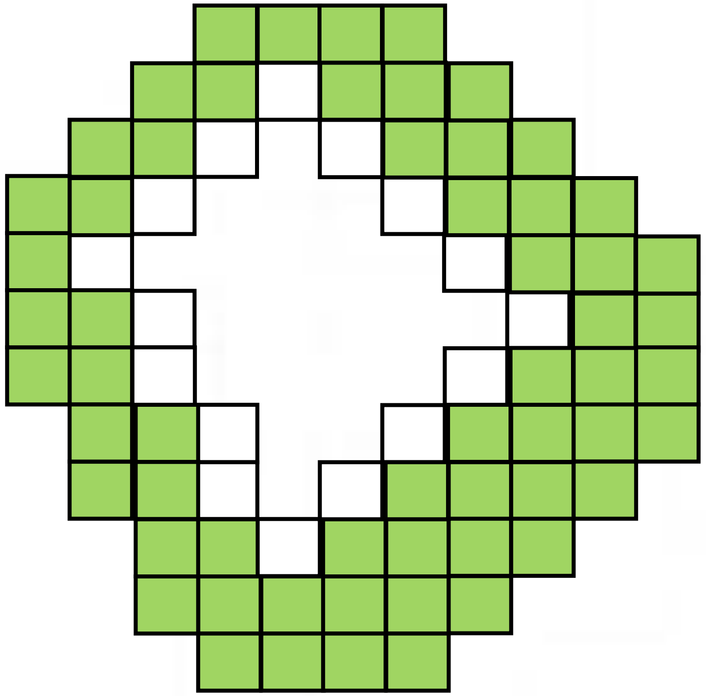
Affine Transform – e.g. Rotation
Locate grid points on the global grid
Global Grid – Lens Correction
Interpolate final grid coordinates
Input Image

Tiles are not regular …



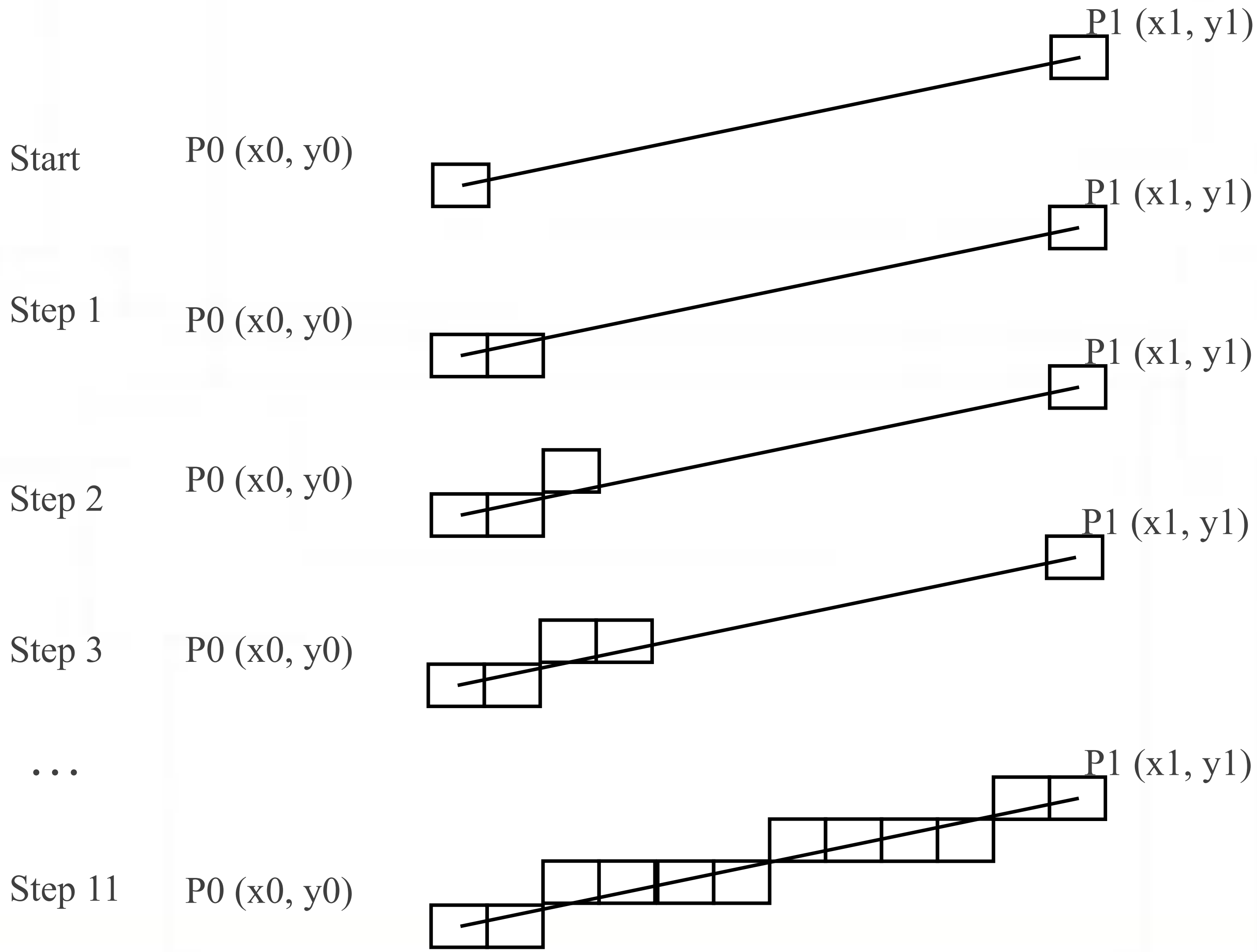
Coordinates found by the Edge Tracer

Each of these pixels must be the center of a 4x4 area



Green area is added by the Tile Border Extender

Lines are not straight …



SGDU is a FotoNation **IP core** designed to support flexible high quality resampling without the burden of GPU

- Enables great flexibility, allowing **static corrections** (e.g. sensor tilt, lens barrel distortion), **dynamic corrections** (where grid changes from one frame to the other, such in the case of VIS) and any **concurrent combinations of both**

Enabling technology for a number of applications

- Lens distortion correction (geometrical and colour aberrations) for Wide Field of View **(>80 degrees**)
- **Dynamic field of view** (WFoV support with capability of multiple windows electronic pan, zoom and tilt with individual distortion correction per window)
- **Video Stabilization** correction with RS correction

Supports combinations of both static and dynamic correction grids

And many others

- 3D convergence correction
- Split screen display /picture in picture support
- Multiple face based zoom windows with individual perspective correction for video telecoms applications.
- Simple video effects (moving windows with alpha blending) and region based warping (e.g. face effects)

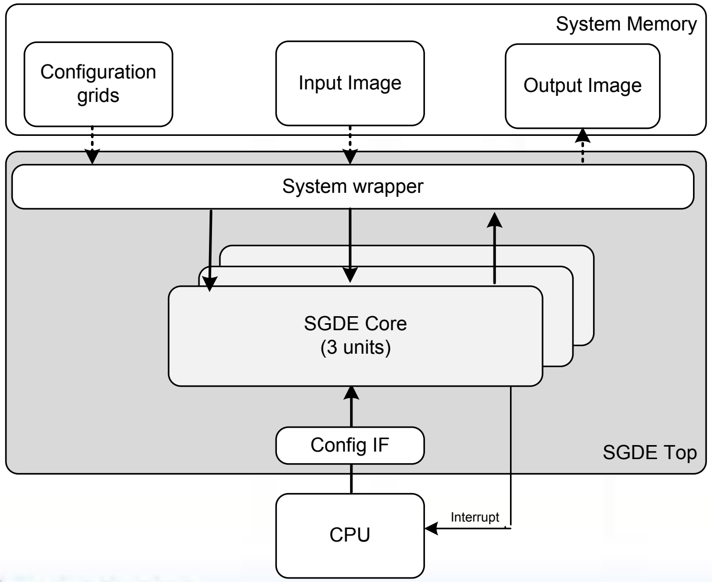
SGDE Performance and HW configurations

Performance

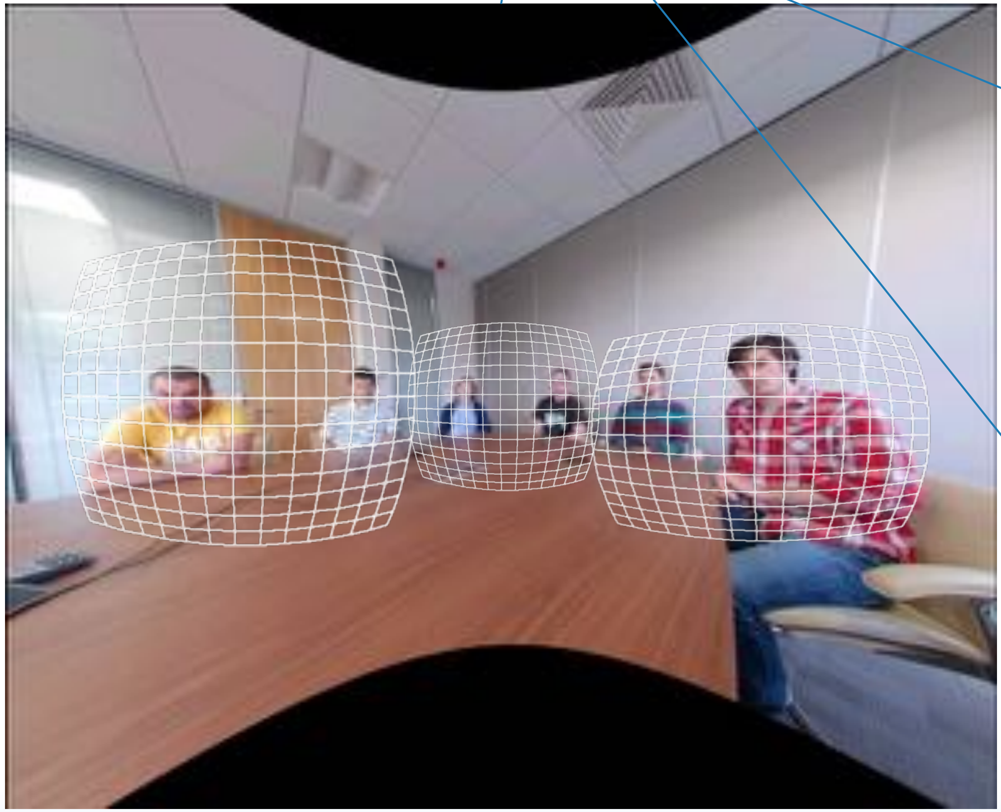
- Single pass through the input image – Low memory bandwidth requirement;
- Multiple output streams support with up to **256 programmable windows** with individual correction grids (each with its own zoom, pan and specific geometric correction support)
- Support for 3 color/pixel transformation, with 5 bits sub-pixel computation precision
- Support for 8k@60 fps with bi-cubic quality resampling

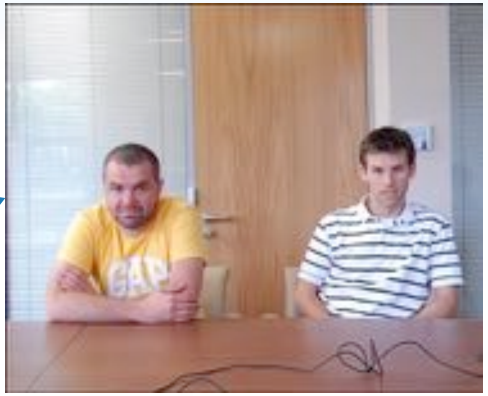
Example – teleconference use case:

3xSGDU cores for **4k@60fps** support)



SGDU





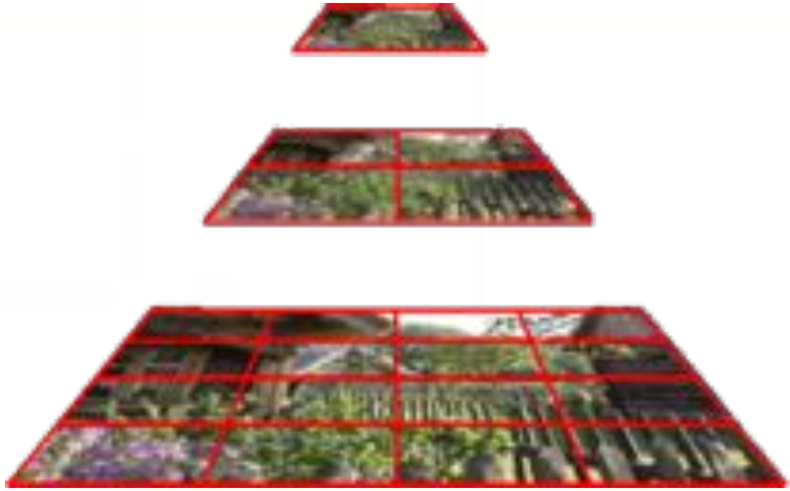
SGDE – Flexible for a number of use-cases (wplay, multiple streams with individual correction, etc..)

SGDE – Flexible for a number of use-cases (wplay, multiple streams with individual correction, etc..)

EXAMPLE #3 – HRE
HIERARCHICAL REGISTRATION ENGINE

HRE is the IP core implementation of FotoNation' **motion analysis technologies**

Starting from the coarsest level, motion information is further analyzed, its accuracy increases which each level until the last level is reached



Hierarchical approach: motion from previous levels used as starting point and accuracy is increased with subsequent level

HRE is an **enabling** technology which require good understanding of the local and global motion within the scene.

Digitally enhanced Video Image Stabilization solution takes advantage HRE (with gyro/acc input) and SGDE to enable all degrees of freedom hand jitter correction (including rotation) as well as rolling shutter compensation

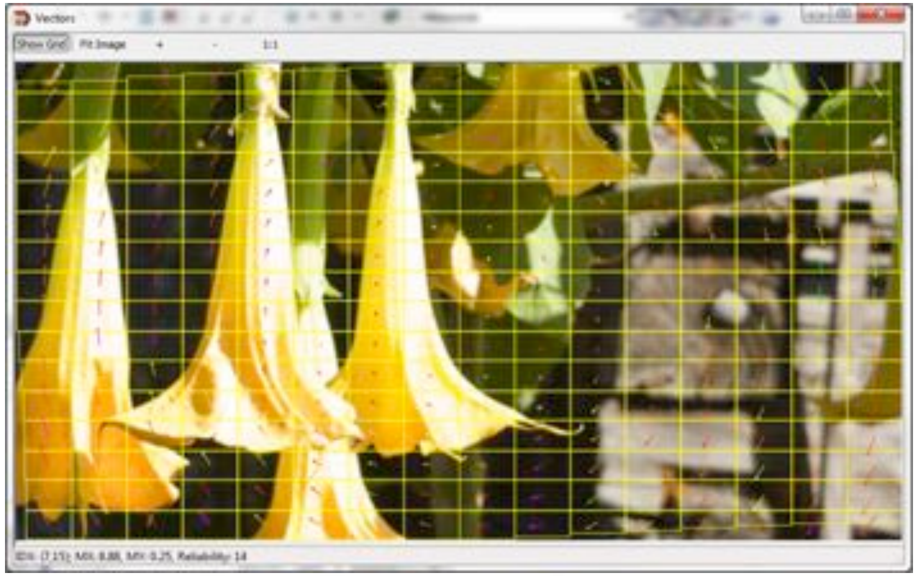
Other features using HRE

- Panorama Capture with enhancements for moving objects
- 3A's and Continuous Autofocus
- Any multi-frame which requires motion analysis (**HDR**, **Night Portrait**, etc)

HRE Performance and HW configuration

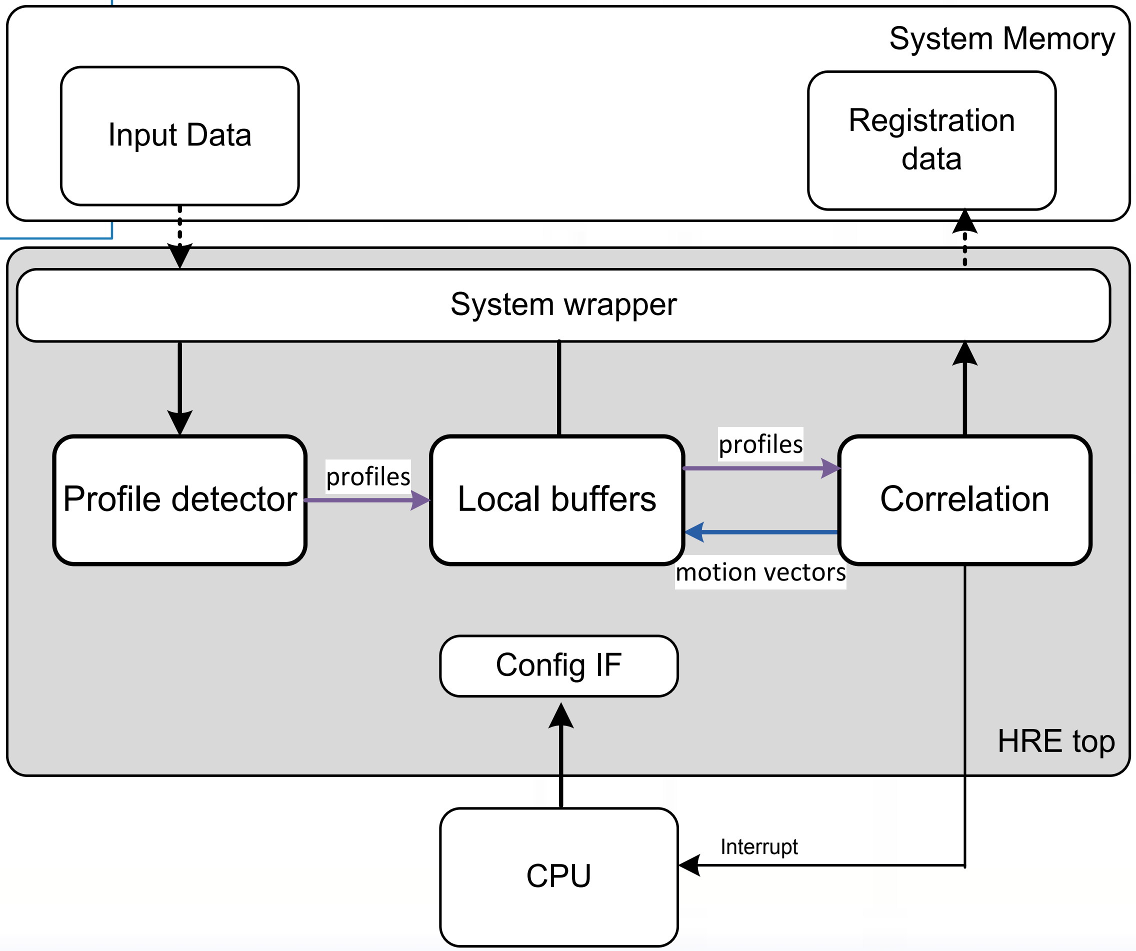
Performance

- Can run **8k@60fps**
- Can also support bigger frame resolution (**5Mp, 8Mp, … up to 20Mp**) at a lower frame rate
- Provides good accuracy with any input size up to 4k pixels per line
  - Sub-pixel motion estimation allows motion estimation at lower resolution than the input data
- Works with distorted input images up to 10%
- The resulting displacement map is up to 16x16



Graphical representation of the motion information obtained with HRE

HRE schematic



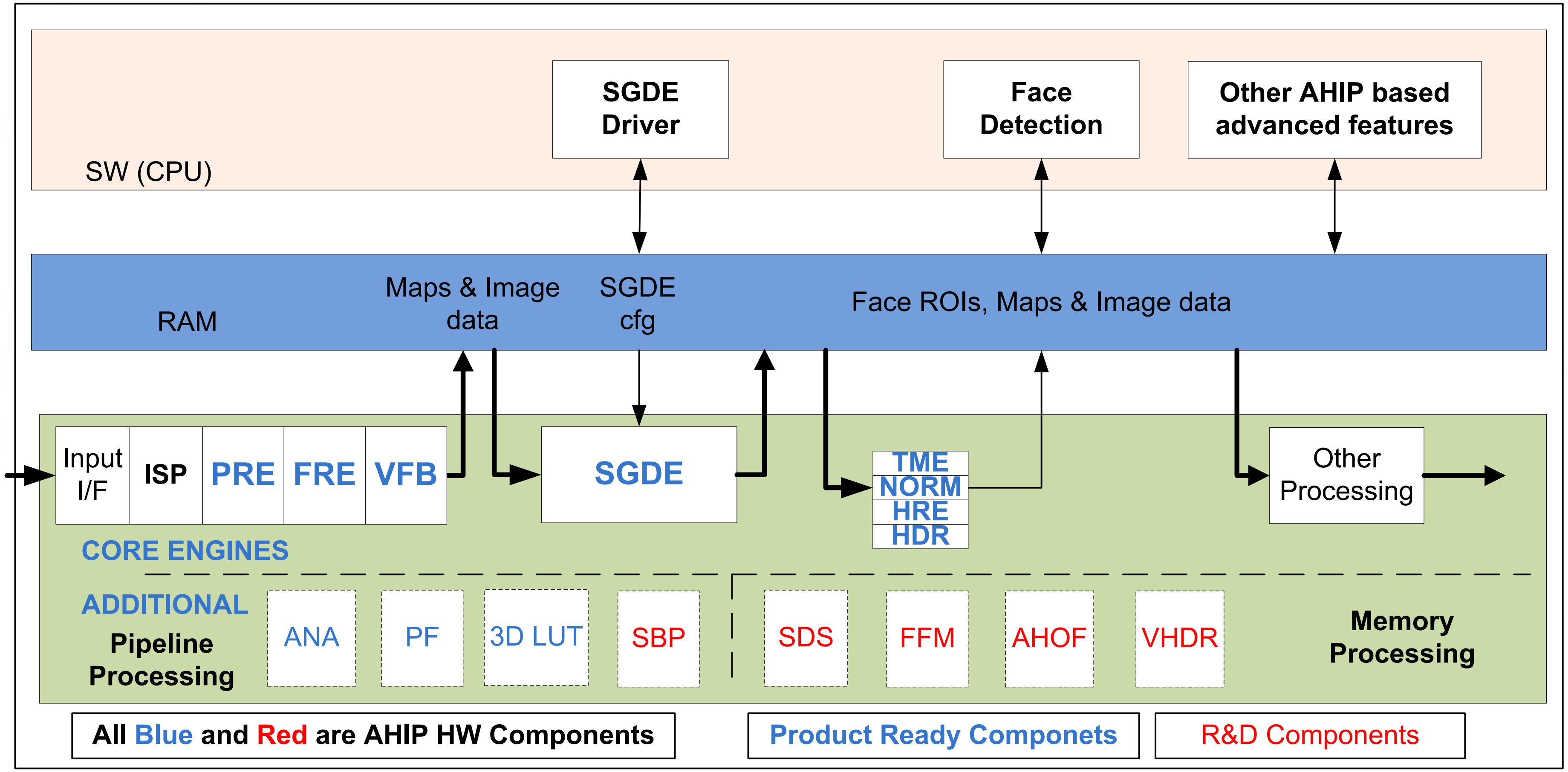
An experiment using the Qualcomm **MSM8974AC Snapdragon 801 with Adreno 330 GPU** with bicubic resampling implemented using shaders (in an effort to match SGDE image quality) showed that the **GPU was unable to maintain a useful frame rate** even while operating at full capactity and consuming a significant amount of power.

SETUP	INPUT	OUTPUT	SPEED	POWER
SW QCM	3 Mp	full HD	43.5 ms (23 FPS)	750 mW
SDGE (1 cores)	3 Mp	full HD	7.5 ms (60 FPS)@300Mhz	6mw
SGDE (1 cores)	Up to 20Mp	4K	30 ms (30 FPS)@600Mhz	10mW
SGDE (3 cores)	Up to 20Mp	4k	15ms (60fps)@300Mhz	18mW

Digitally enhanced Video Image Stabilization running on top of HRE and SGDE

FUTURE VISION (AVAILABLE TODAY!)

Tested, dedicated IP cores combined in an Advanced Hardware for Image Processing (AHIP)



Legend: PRE – pre-analysis engine; FRE – Frame Registration Engine; SGDE – Smart Geometrical Distortion Engine; HDR – High Dynamic Range; TME – Template Matching Engine; ANA – Adaptive Nonlinear Amplification; PF – Purple Fringing; VFB – Video Face Beautification; SBP – Selective Background Processing; AHOF – Advanced Hardware Object Filtering; FFM – Face Feature Model; SDS – Stereoscopic Depth Sensing; VHDR – Video HDR; HRE – Hierarchical Registration Engine; NORM – normalization engine

FotoNation®

a Tessera company

Thank You