Biologically Inspired Vision Systems on configurable platforms

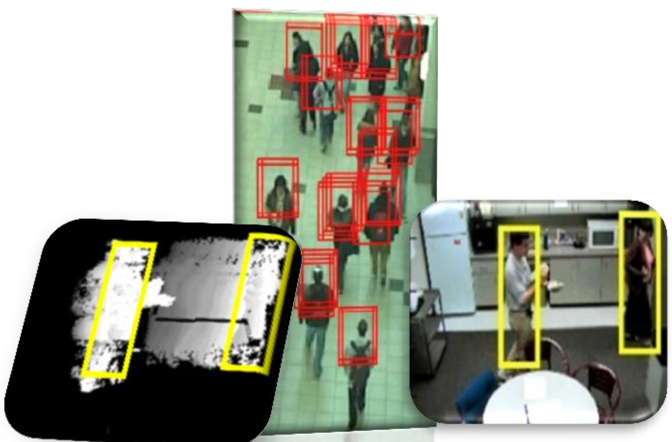
Vijaykrishnan Narayanan

This presentation includes work supported in part by DARPA, Intel Science and Technology Center on Embedded Computing, National Science Foundation and Semiconductor Research Corporation

Collaborators: Laurent Itti, USC; Kevin Irick, PSU, D. Khosla, HRL, Yang, HRL; M. Peot, Teledyne, B. Desimone and T. Poggio, MIT, J. Tstosos, York and all my students

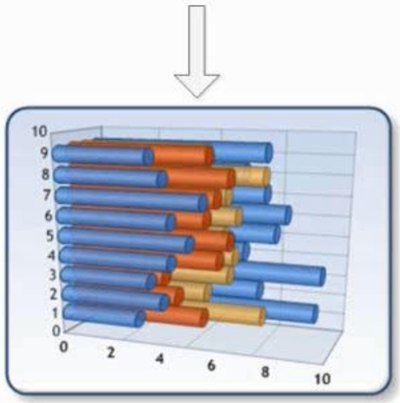
Cameras Everywhere!



Tracking, Vehicle Navigation, Augmented Reality

Form1



SmartView Digital Display

Visual Cortex Inspiration

- Can we build vision systems that detect and recognize objects as efficiently as mammals ?
  - Accuracy – Complex scenes with clutter
  - Speed – Mammals require ~150milliseconds
  - Power – less than 20 Watts?

Smart Cameras & Accelerators

Available Accelerators (abbreviated)		
Arithmetic (add, sub, mult, div,…)	Difference of Gaussians	Image Pyramid
Convolution/Correlation	Saliency (AIM and Itti) GIST	Retina Preprocessing
Color Space Conversion	Skintone Detection	SURF
Bounding Box Extraction (Connected Component)	Face Detection	Brute Force Matcher
Histogram	Support Vector Machine	Density Estimation
Image Statistics (mean and std)	HMAX Classifier	Gabor Edge Feature Extractor
2D and 3D FFT	Function Approximation (log, tanh, sigmoid)	Image Subsampling and Interpolation

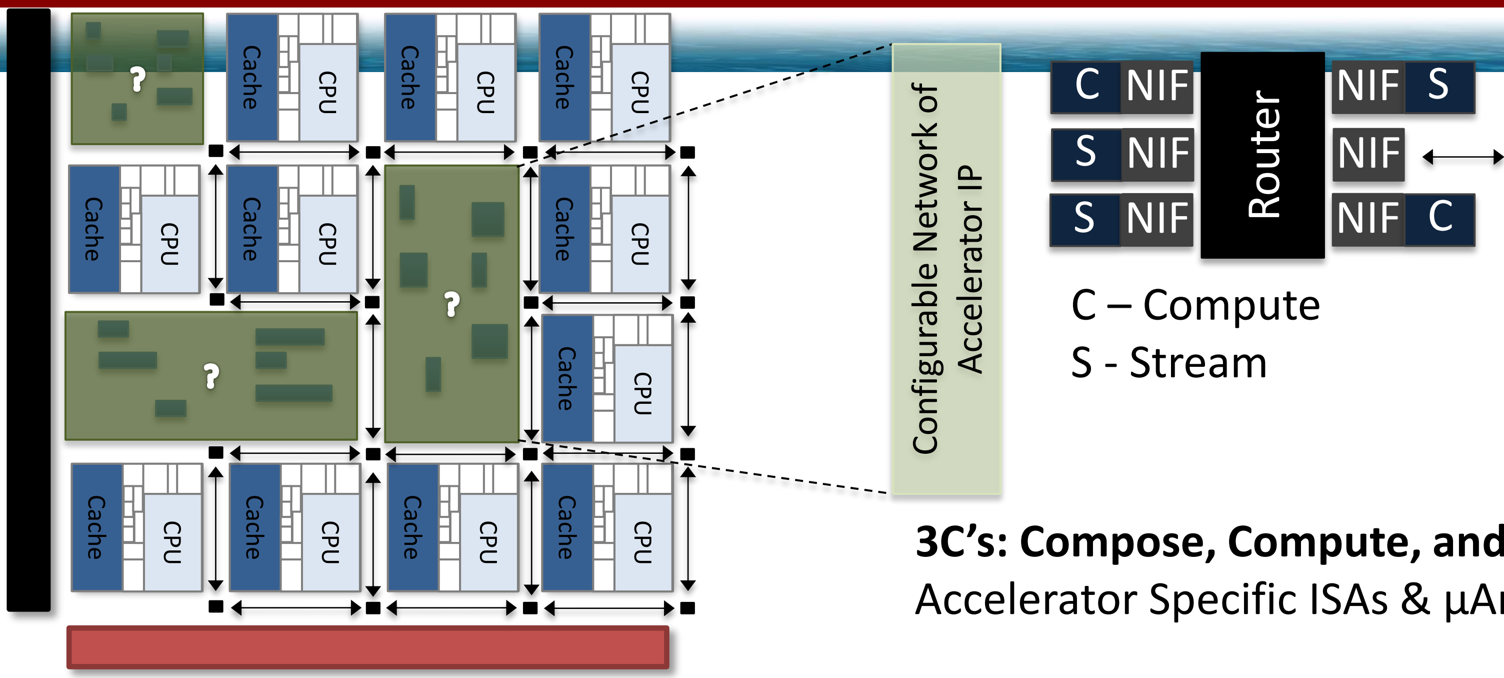
Network connectivity allows advanced video analytics tasks to be distributed across many low-cost nodes obviating the need for prohibitively expensive backend computing infrastructure

Integrated camera and computation platform allows processing to be performed directly on video stream reducing network load to backend cloud infrastructure

Library of optimized software and hardware modules support advanced video analytics on a small-footprint low-cost platform

High-level analytics are dispatched to a scalable cloud infrastructure

Smart Camera Architecture Platform



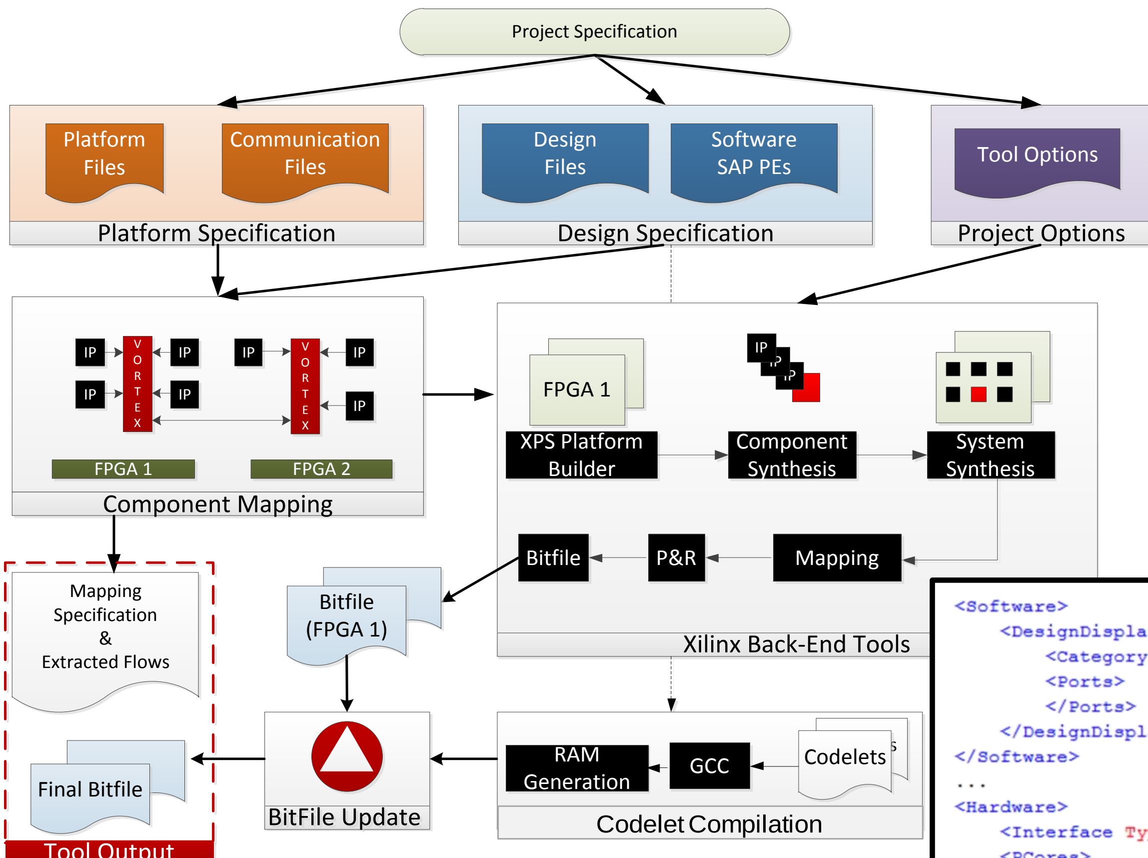
3C's: Compose, Compute, and Communicate?

Accelerator Specific ISAs & µArchitecture

Accelerator Requirements

- Allows for Hardware Re-Use and Run-time Configurability
- Scalable Accelerator Sub-System Composed of Multiple IP
- Programming Model

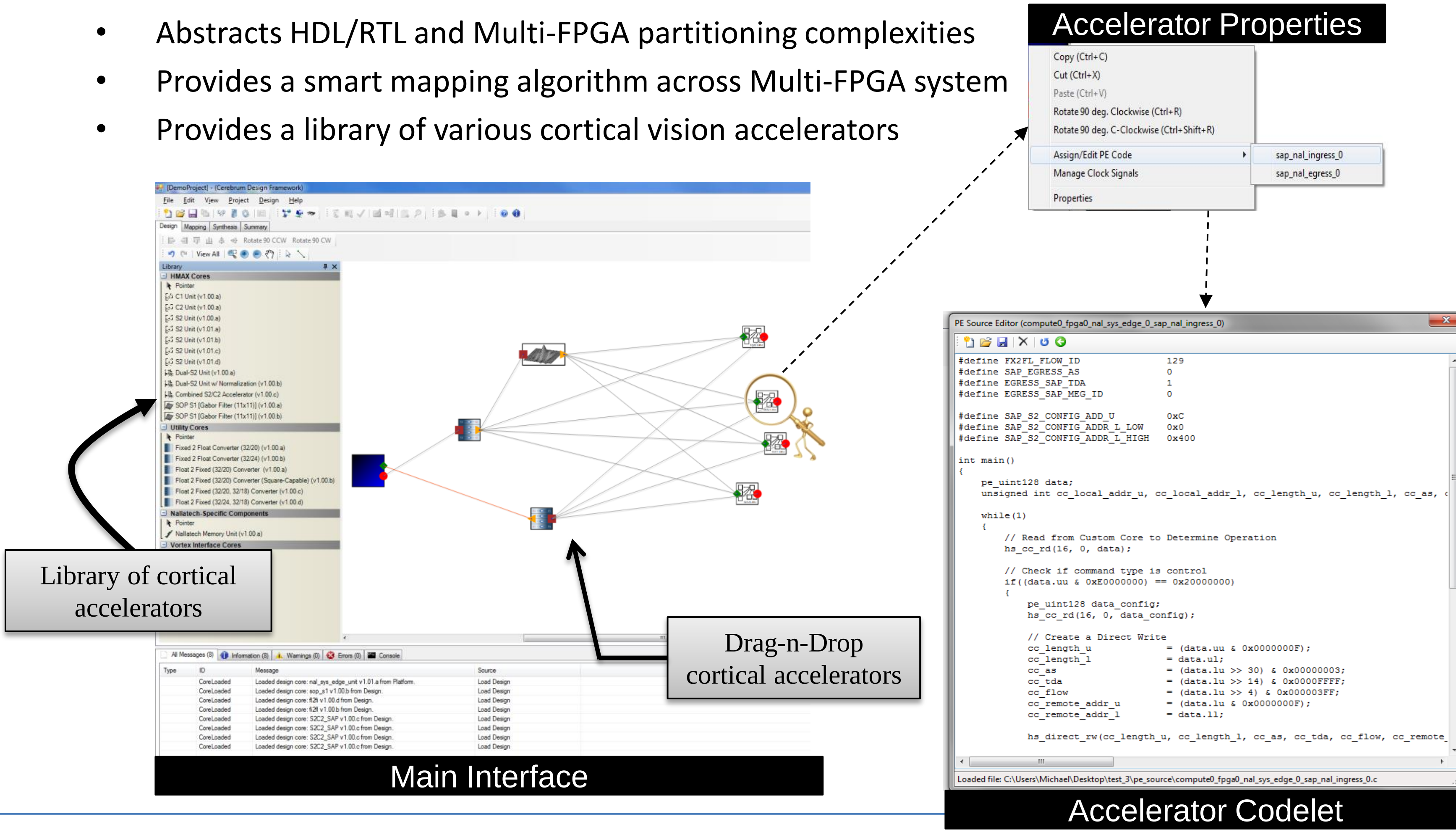
Automation Tools: From Specs to System



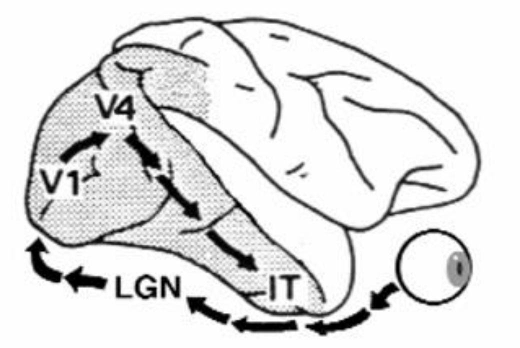

```
<Software>
    <DesignDisplay>
        <Category Name="Nallatech Interface" />
        <Ports>
        </Ports>
    </DesignDisplay>
</Software>
...
<Hardware>
    <Interface Type="SAP" PE="TRUE"> nal_rx </>
    <PCores>
        <PCore Type="nal_rx" Version="1.01.a" ... />
    </PCores>
    ...
    <Clocks>
        <"100MHz Oscillator" ... Frequency="100MHz" ... />
    </Clocks>
</Hardware>
...
```

Automation Tools: From Specs to System

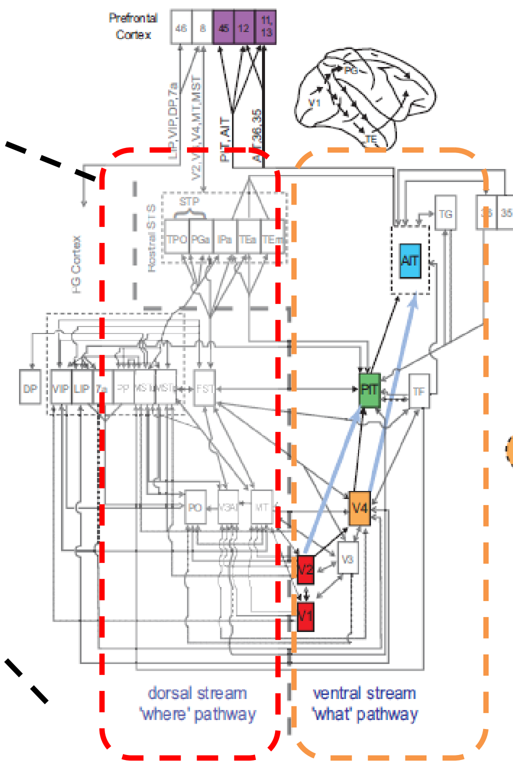
- Built for neuroscientists and researchers
- Abstracts HDL/RTL and Multi-FPGA partitioning complexities
- Provides a smart mapping algorithm across Multi-FPGA system
- Provides a library of various cortical vision accelerators

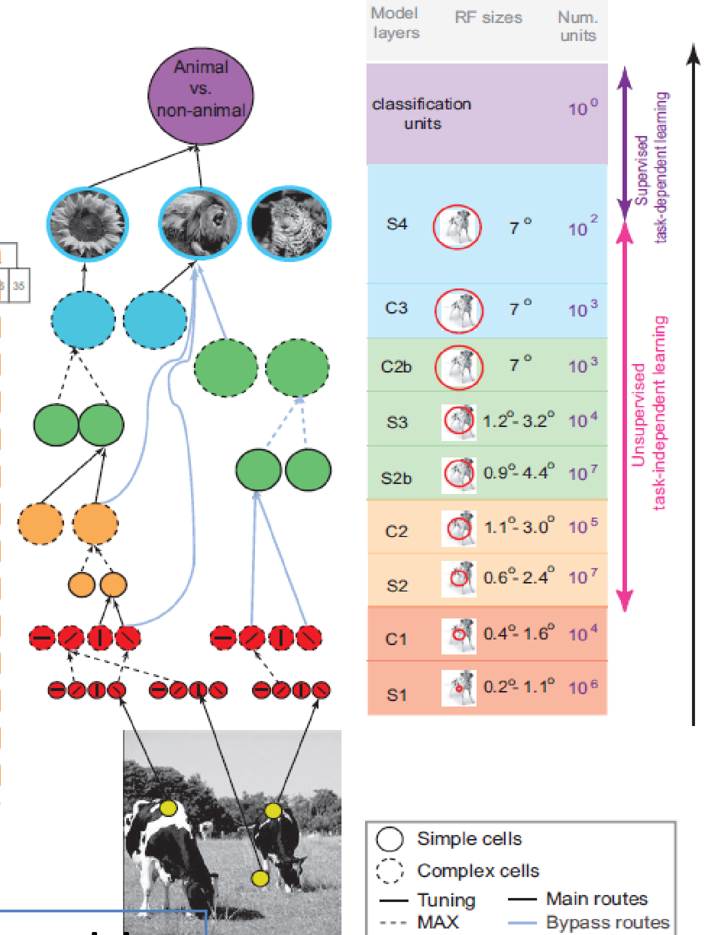


Models for Mammalian Vision



Cortex Region	HMAX Stage
V1, V2	S1, C1
V4	S2
V4 / PIT	C2





HMAX - Recognition model

Model Ventral Pathway from V_1 through V_2 and V_4 to IT

Hierarchical Simple (S) and Complex (C) cells alternate feature extraction and max pooling.

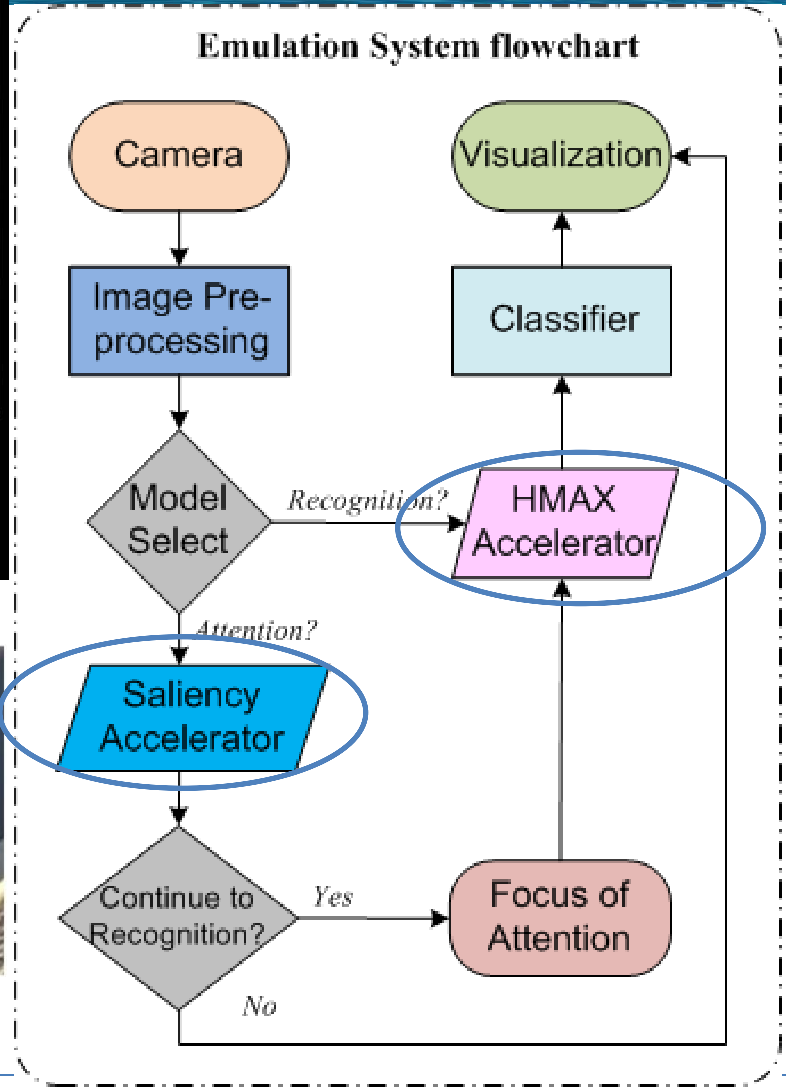
Saliency – Attention model

Model dorsal pathway which provides attention cues to recognition.

Feature extraction in I, C, O, F, M channels to obtain Saliency map.

[1] Poggio et. al. Nature Neuroscience 1999
[2] Itti et. al, SPIE 2001

System Overview



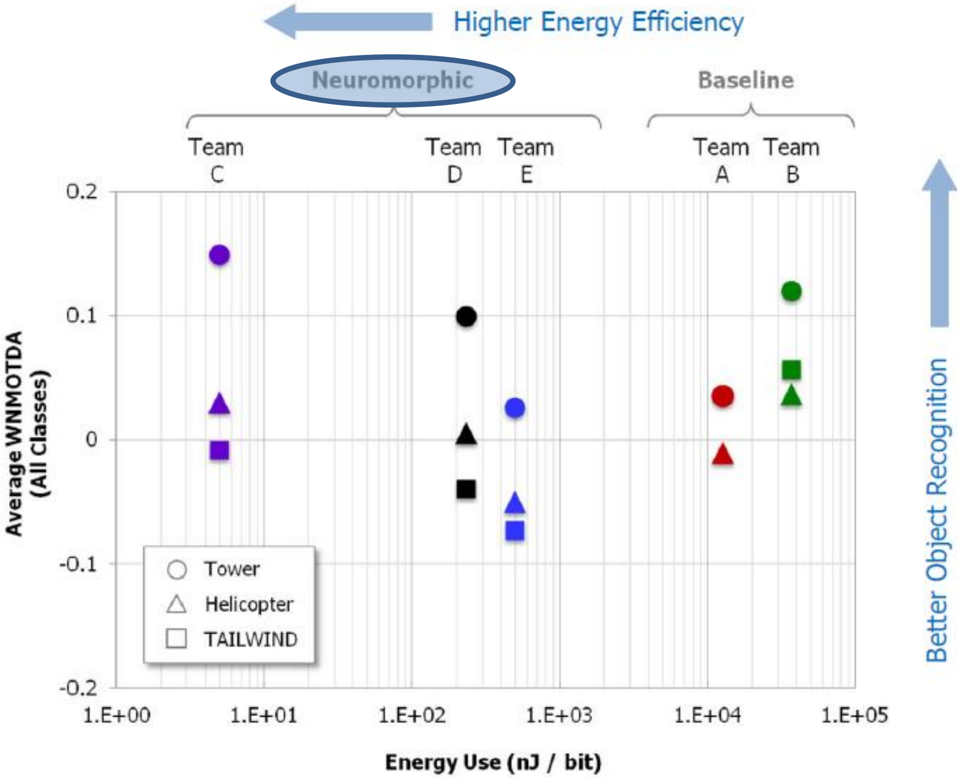
System Implementation Results

Implemented neuromorphic systems to detect and classify ten object classes

Tower
(2 MP at 30 fps)

Helicopter
(2 MP at 30 fps)

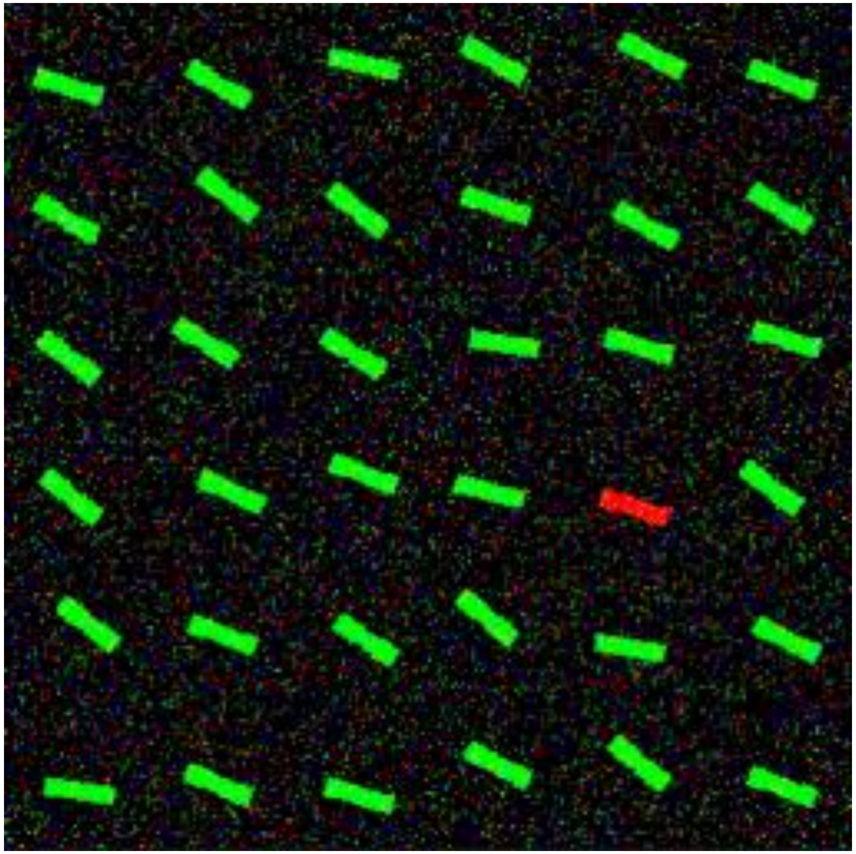
TAILWIND
(10 MP at 5 fps)



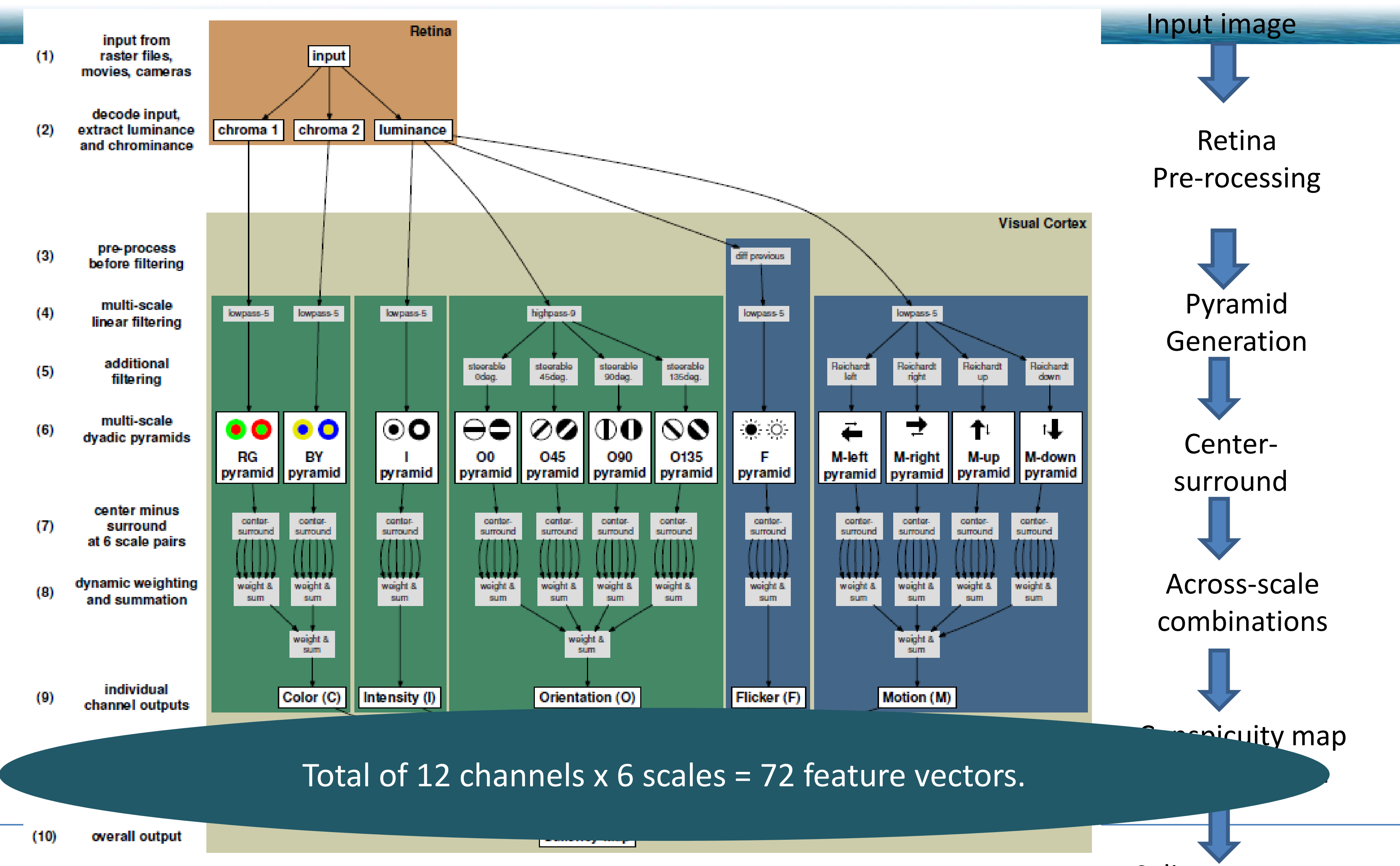
Neuromorphic approaches detect and classify well,
and use 4 orders of magnitude less energy

Still 2-3 orders less energy-efficient than brain !!

Attention



Bottom-up Saliency Model [1]

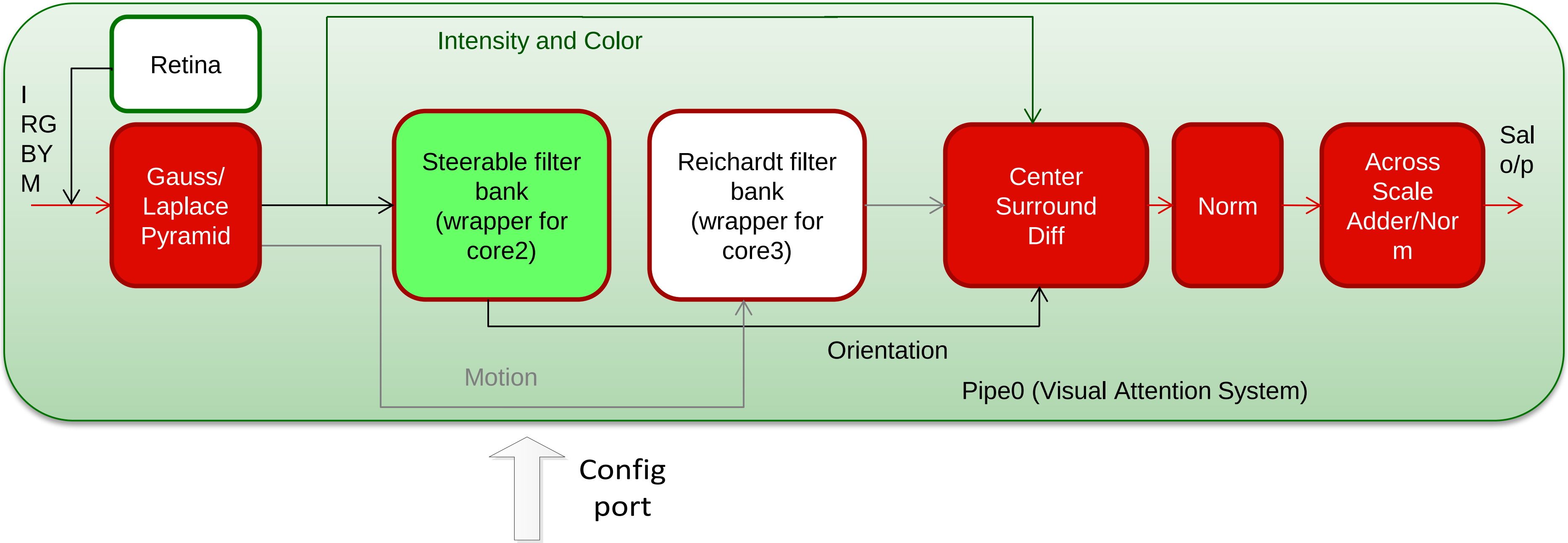


Total of 12 channels x 6 scales = 72 feature vectors.

[1] Peters, Itti, ACM Tran. Appl. Perct, 2008

Attention Pipeline

- New Instruction fetched from Instruction Queue.
  - Configures pipeline registers, data flow, selection logic.
  - Repeat all instructions, per frame.



Pipe_id=0, config_data= 32'h00000A /32'h00000B /32'h00000C
32'h00000A-> bypass steerable bypass reichardt and use Gaussian pyramid
32'h00000B-> use steerable bypass reichardt along with Laplacian pyramid
32'h00000C-> bypass steerable use reichardt and use Gaussian pyramid

Attention: Object Detection

Performance Comparison in Frames per sec (FPS) for 640x480

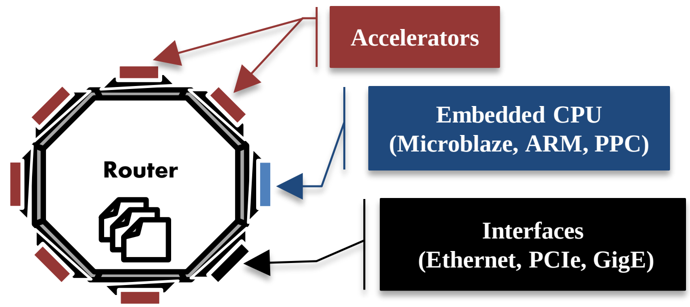
	CPU[1]	GPU [2]	Proposed Accelerator
	Intel Xeon dual-core CPU (2.8 GHz)	Nvidia GeForce 8800 (GTX) (1.35 GHz)	1 x Virtex6 SX475T FPGAs (100 MHz)
CIO	19.48	94.25	169.55
CIOFM	14.99	NA	100.06

- **Spatial Saliency (CIO) - Speedups of 8.7X over CPU, 1.8X over GPU and 1.89X over FPGA implementations.**
- **Full Saliency (CIOFM) Speedups of 6.6X over CPU impl.**
- Power Efficiency (FPS/Watt): 11.2X over GPU

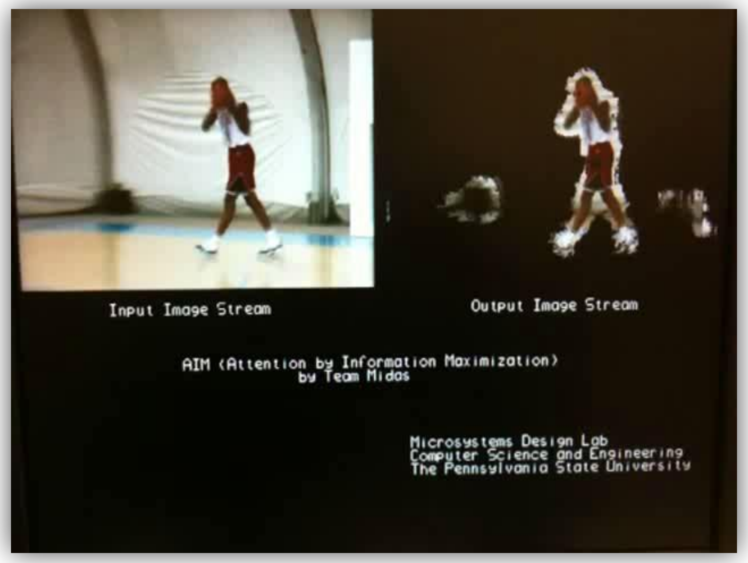
Video Analytics Accelerator Platform

Configurable On-chip Communication Infrastructure + Embedded Video Analytics Accelerators

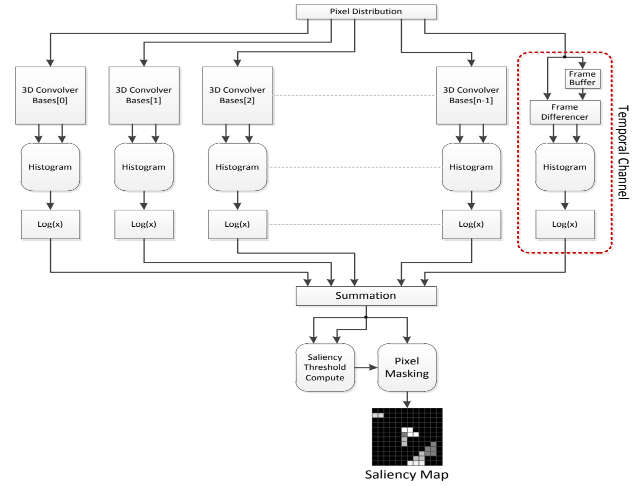
Expansion Card Platform



On-chip Infrastructure



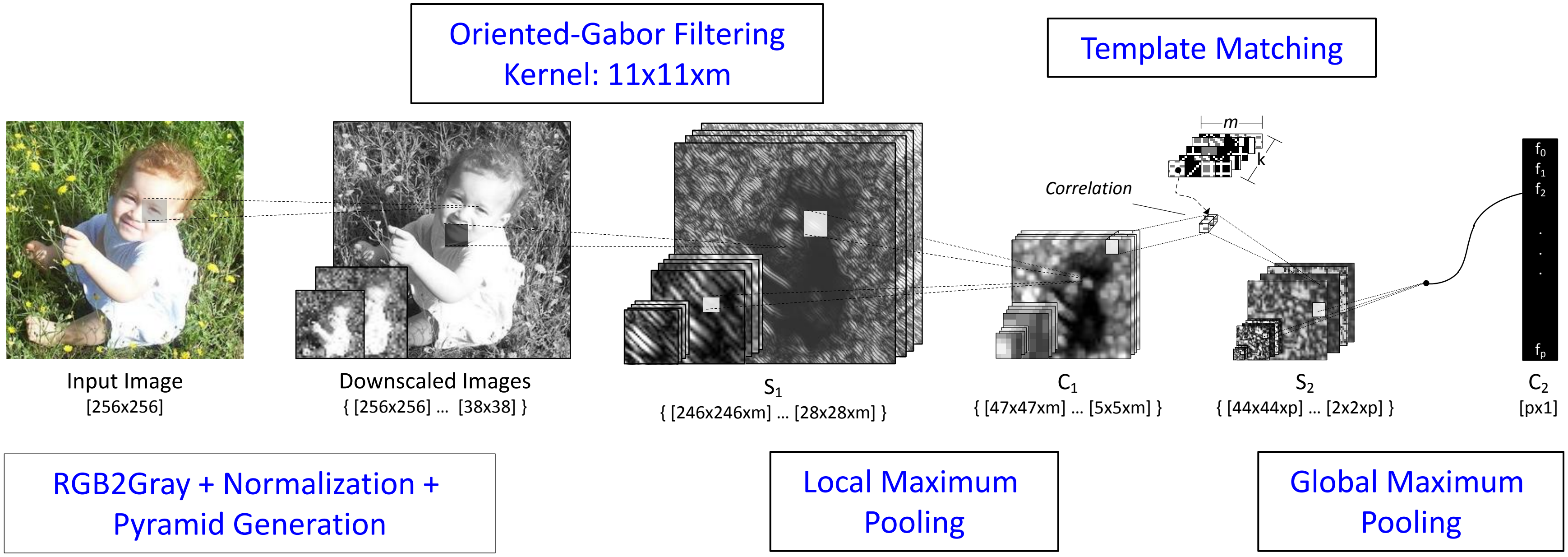
Accelerated Application
(Image Saliency AIM)



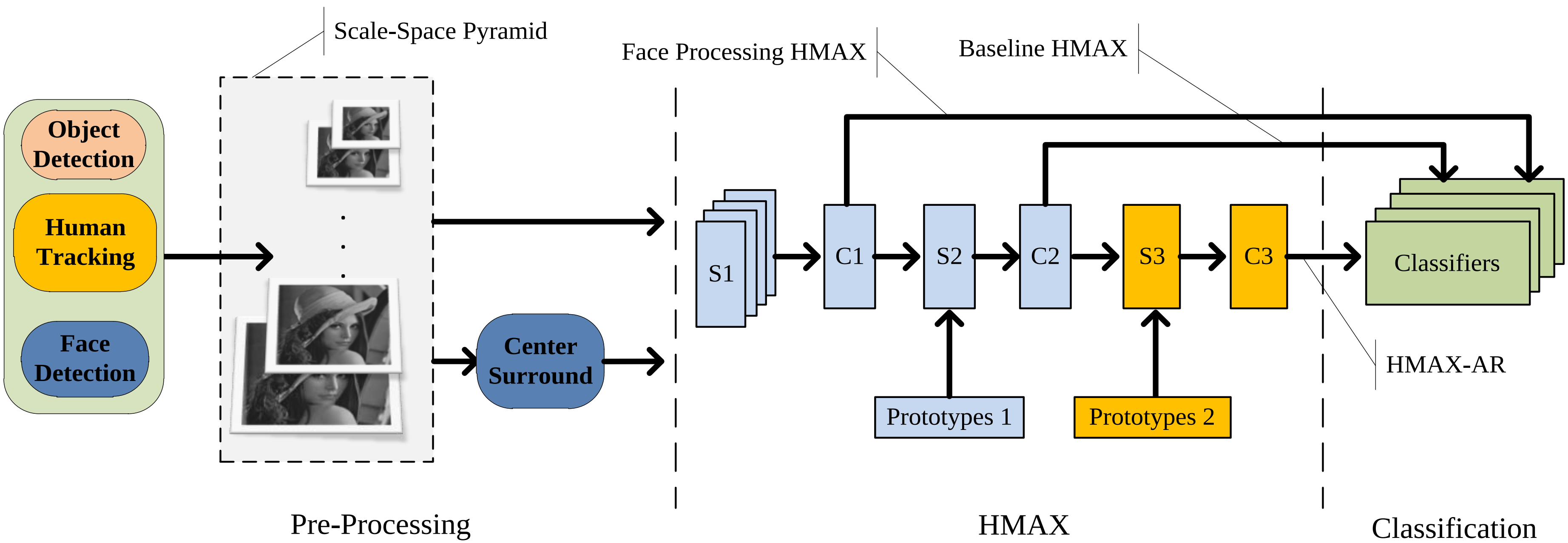
Custom Hardware Accelerator

Classification Model: HMAX

- Riesenhuber & Poggio, 1999
- A cortical model for object classification, that models the ventral path in the visual cortex.

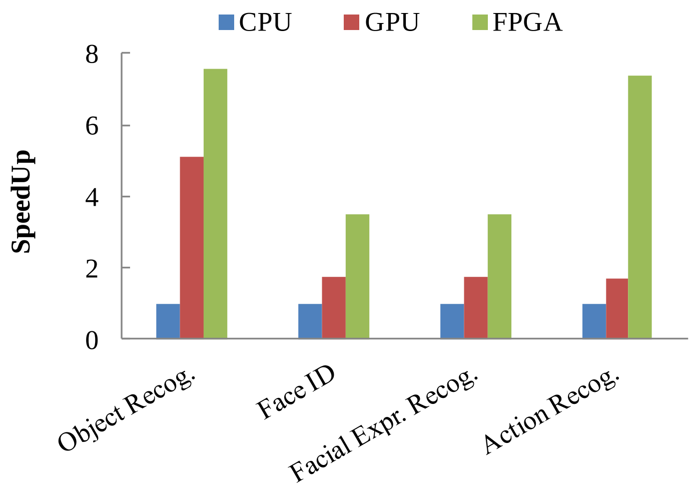


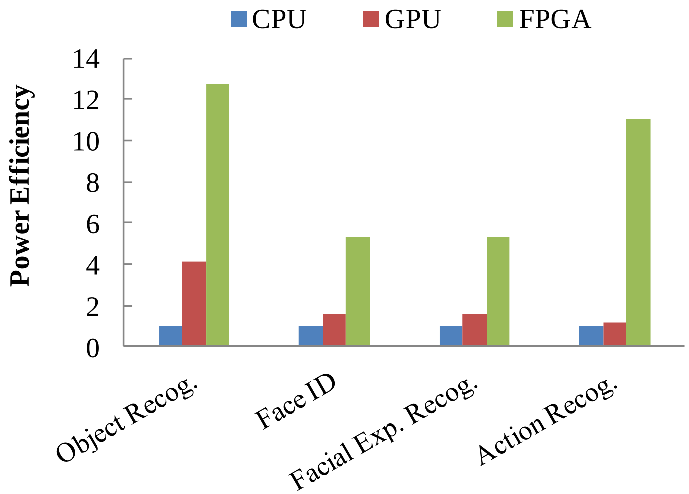
Enhanced HMAX Recognition System



Experimental Setup

HMAX Accelerator





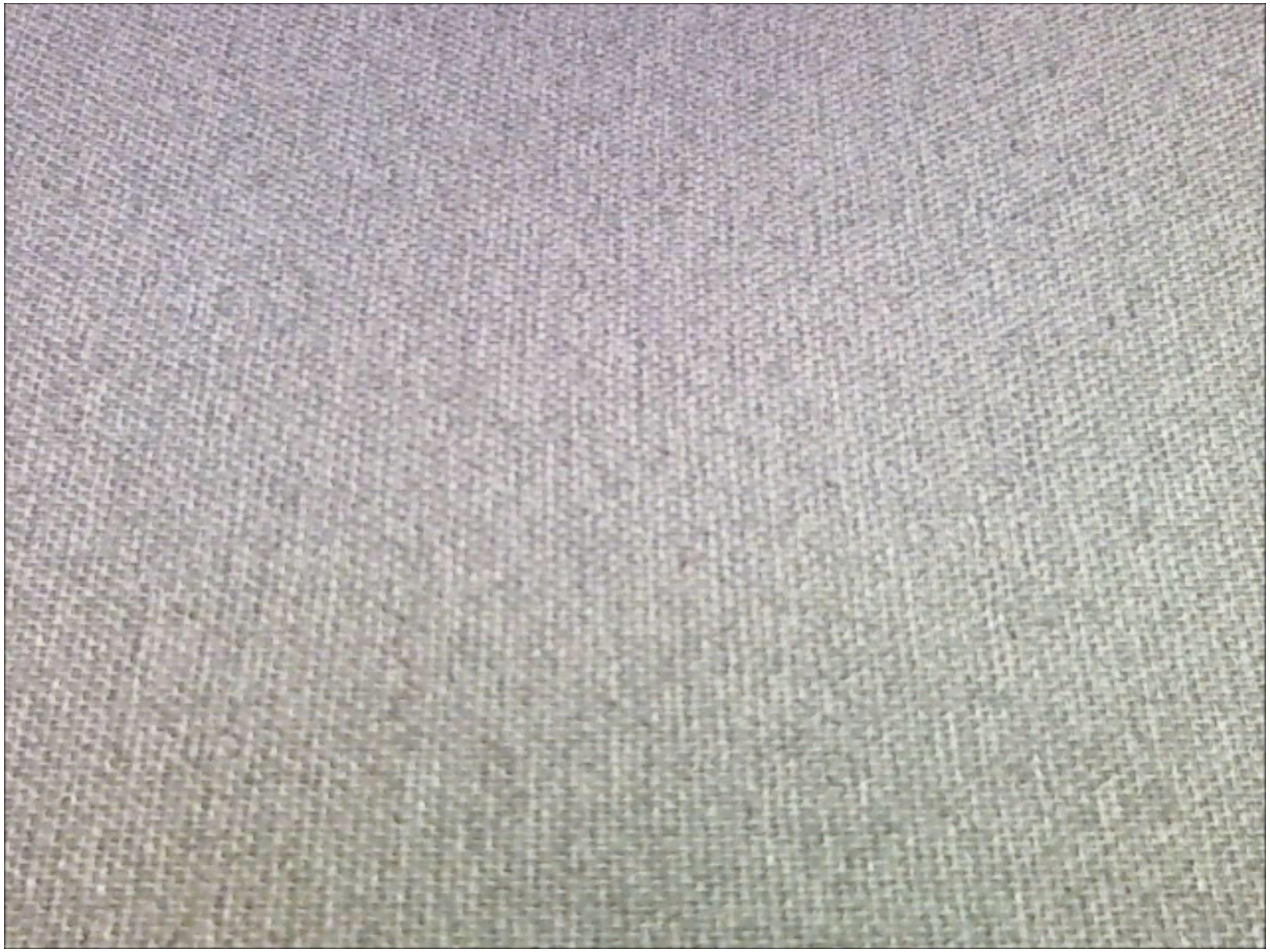
Speedup: up to 7.6X (4.3) compared to CPU (GPU)

Power-Efficiency: up to 12.8X (9.7X) compared to CPU (GPU)

CPU: 12-Core Xeon CPU @ 2.4 GHz, **GPU**: Nvidia Tesla M2090 board (T20A GPU @ 1.3 GHz)
FPGA: 4 Virtex-5 SX240 FPGAs

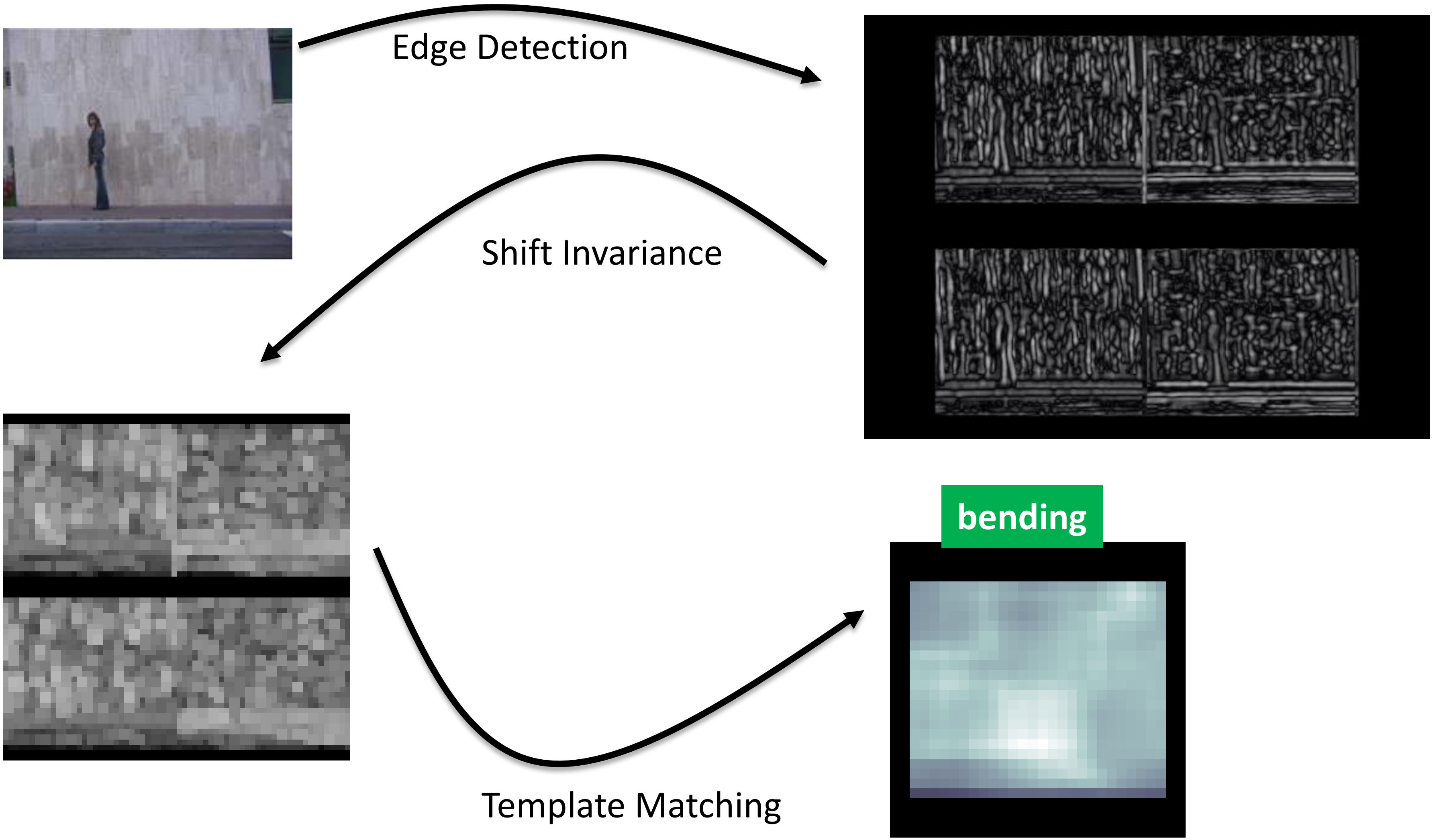
Compared to a single threaded 2.4GHz CPU, the accelerator delivers 73X speedup and 25.8X more power efficiency

Object Recognition

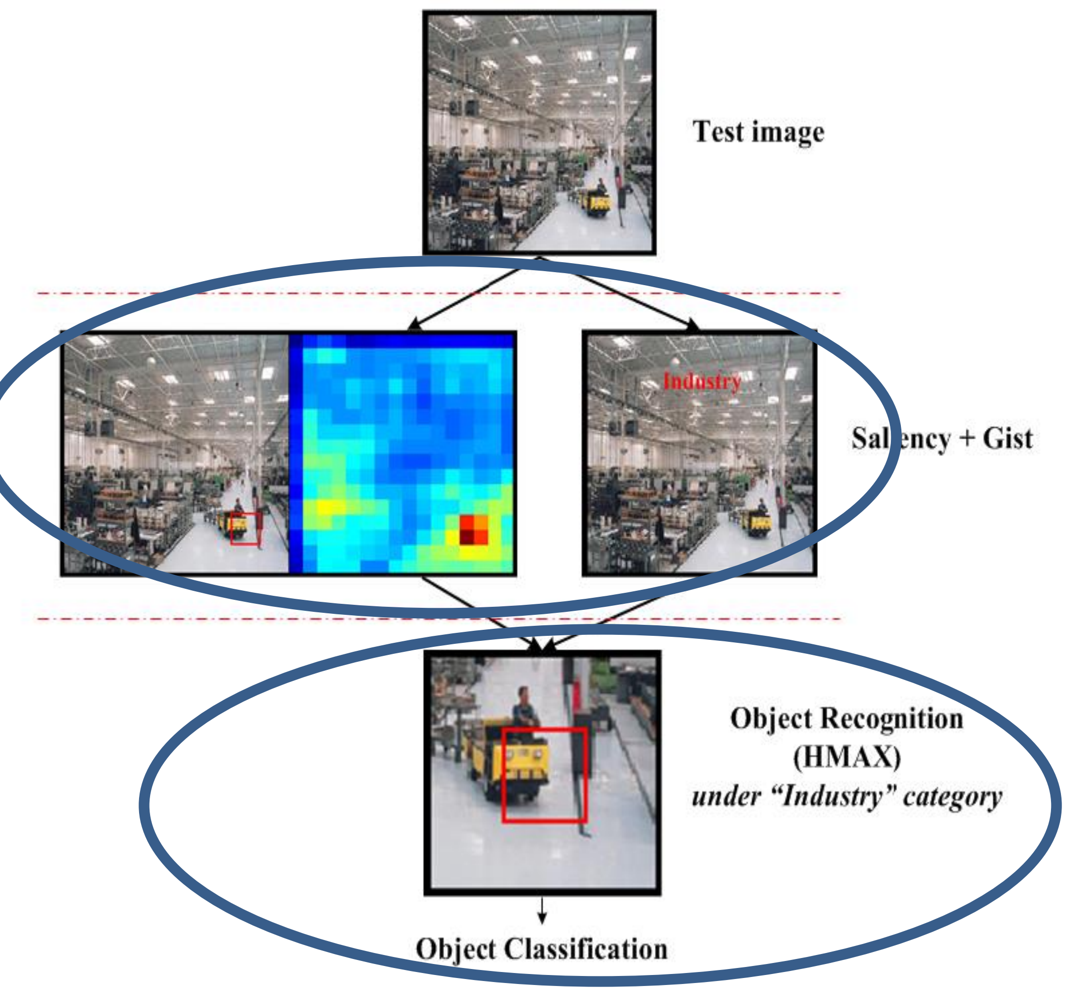


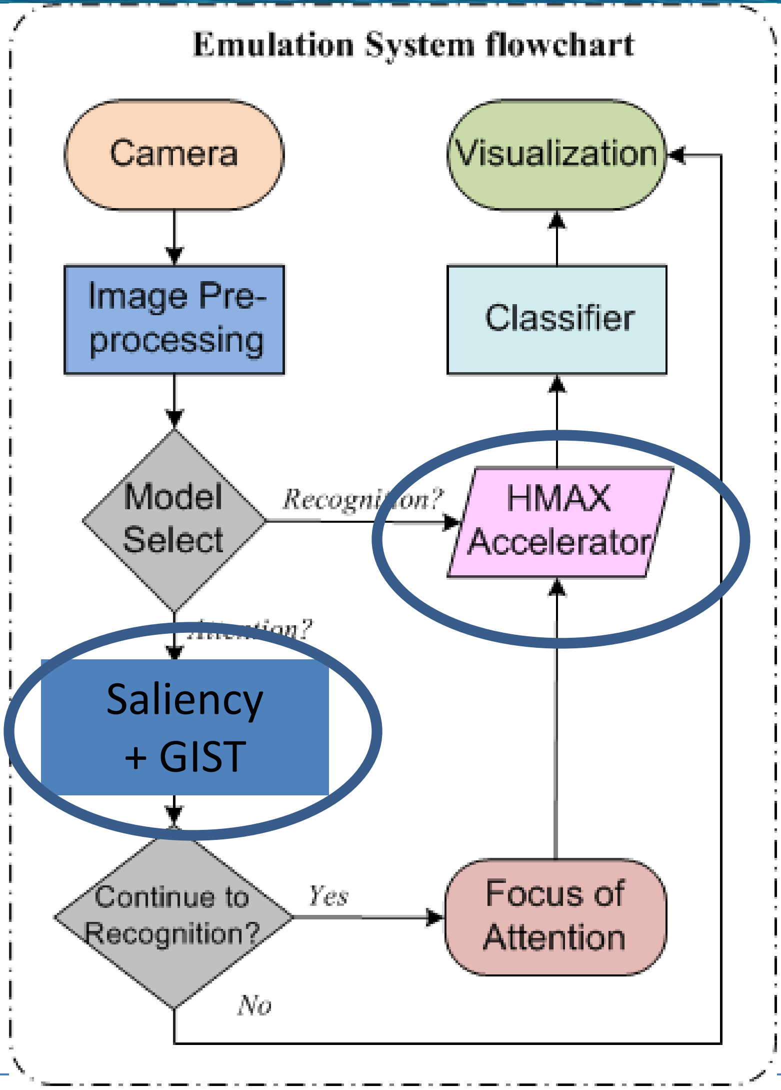

Happy, Sad, Surprise

Action Recognition



System Overview



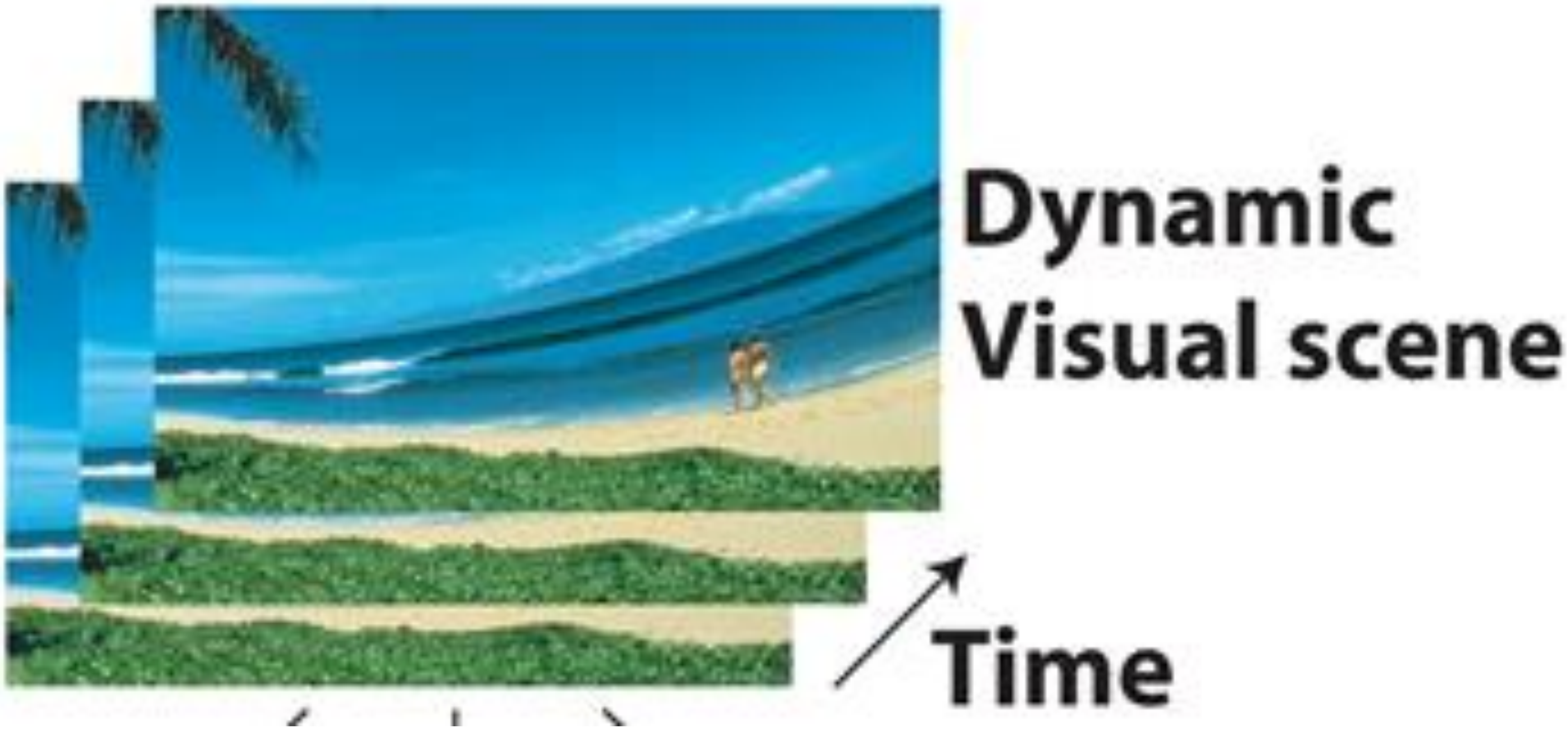


# of scene categories	# of object classes per category	HMAX-only Accuracy	Gist + HMAX New Accuracy
4	3	69.17 %	77.92 %
	2	77.50 %	84.38 %
3	3	72.22 %	82.22 %
	2	83.33 %	92.50 %

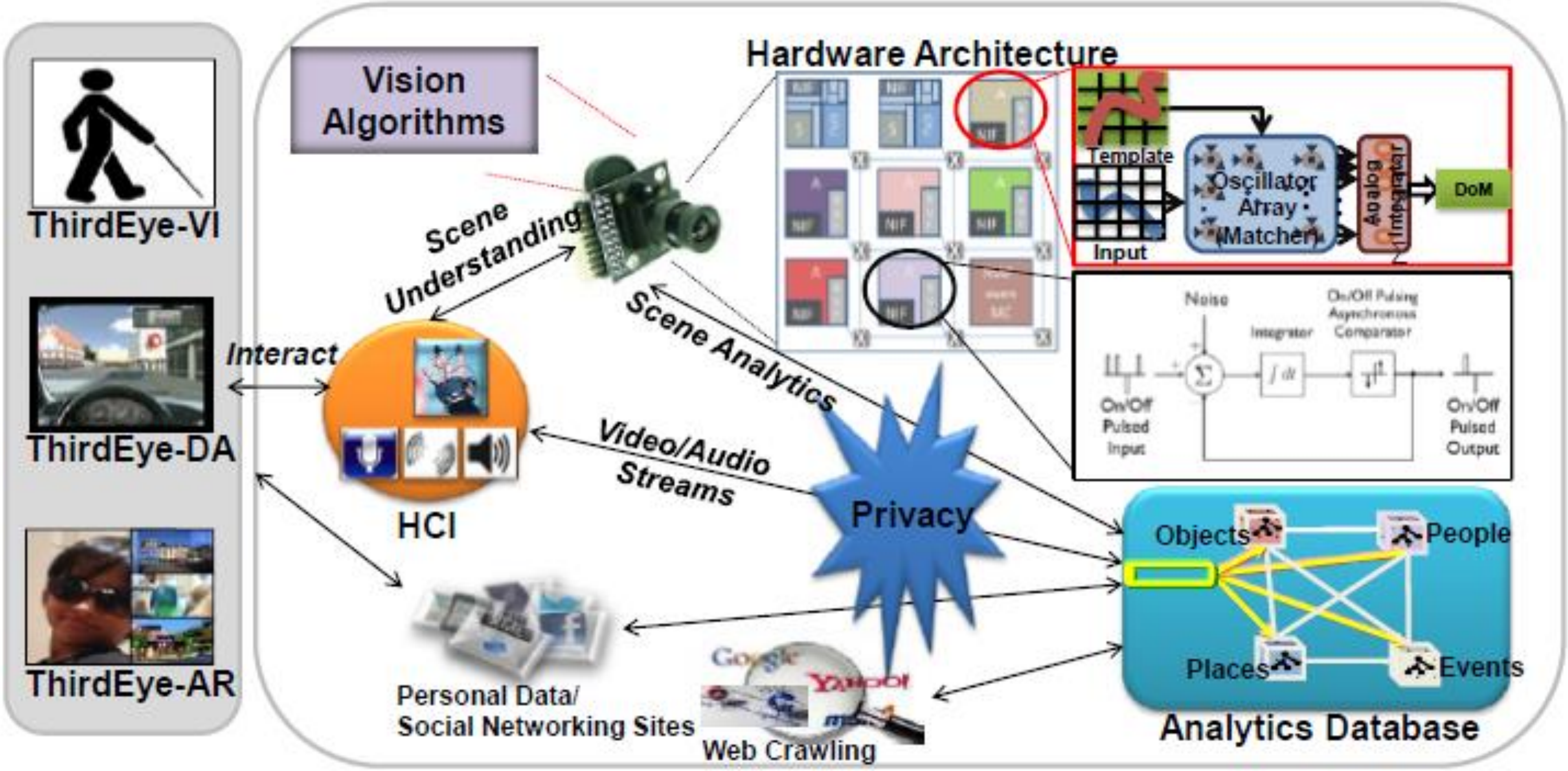
Scene Category	Object Class
highways	car
	stop sign
	motorbikes
beaches	ketch
	ferry
	crab
forests	panda
	leopards
	beaver
buildings	cup
	laptop
	chair

Algorithmic Choices

Couple walking on a beach



Social Impact


ThirdEye-VI
ThirdEye-DA
ThirdEye-AR
Vision Algorithms
Hardware Architecture
Template
Input
Oscillator Array (Matcher)
Analog Integrator
DoM
Noise
Integrator
On/Off Pulsing Asynchronous Comparator
On/Off Pulsed Input
On/Off Pulsed Output
Scene Understanding
Interact
HCI
Scene Analytics
Video/Audio Streams
Privacy
Objects
People
Places
Events
Analytics Database
Personal Data/
Social Networking Sites
Web Crawling

Thank you

QUESTIONS ?