Embedded smart vision system based on RBF neural processor : application to human pose classification

M.Paindavoine[1,2], E.Boattlander[2], G.Sauvage[2], G.Camenen[2], O.Brousse[1,2]

1: LEAD Universiy of Burgundy & CNRS (Dijon, France)
2: GlobalSensing Technologies (Dijon, France)

I- Introduction:

Human detection from natural scenes with artificial vision systems needs background elimination

Simple differences betwen current image and background reference image are too noise sensitive

Some robust algorithms have been introduced like Gaussian Mixture Model but calculation complexity is high

An optimization of performance vs complexity consists in bio-inspired human performances in words of detection and recognition

Simple Tasks with Human Brain vs Von Neuman Computer (like PC):

- Calculate in less than one second (398387.86 x 498.07=?)

- But regognize in less than one second this image:

Artificial vision model proposal for embedded systems:

- Arithmetic calculations used in image filtering for example:
  --> Von Neuman (or Harvard) architectures

- Object recognition from natural images:
  -->Bio-inspired Human intelligence: Artificial Intelligence on Silicon

Outline

Introduction

Artificial Intelligence on Silicon

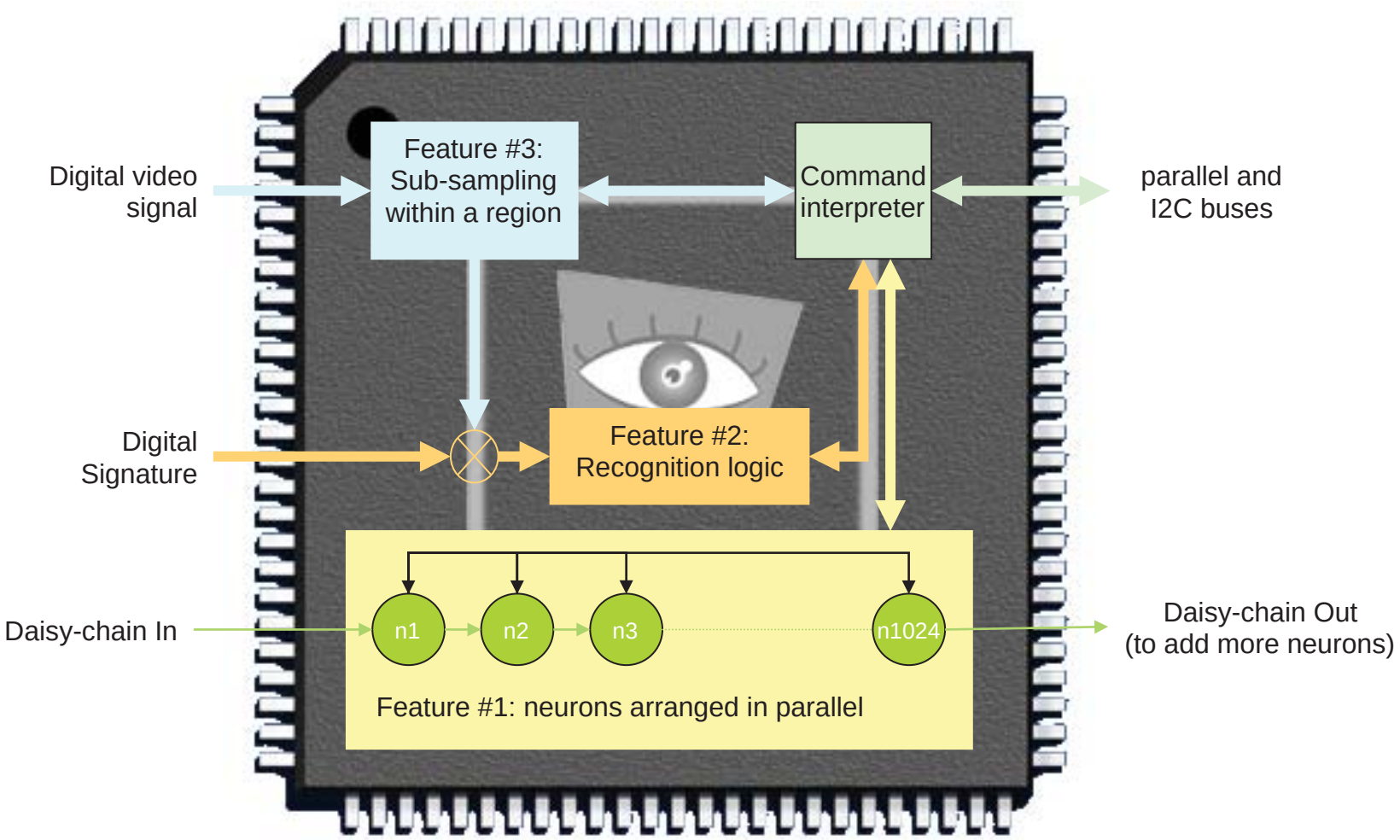
SmartNeurocam: Embedded smart vision system

Application: Real time human pose classification

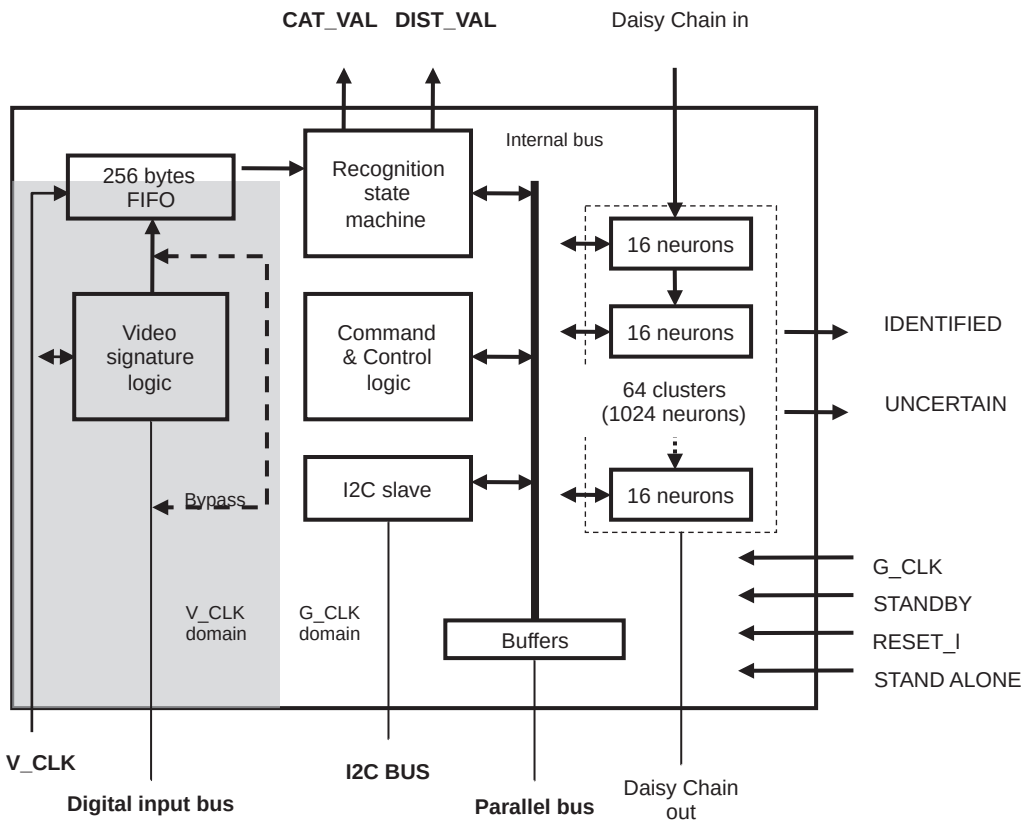
Conclusion

II- Artificial Intelligence on Silicon:

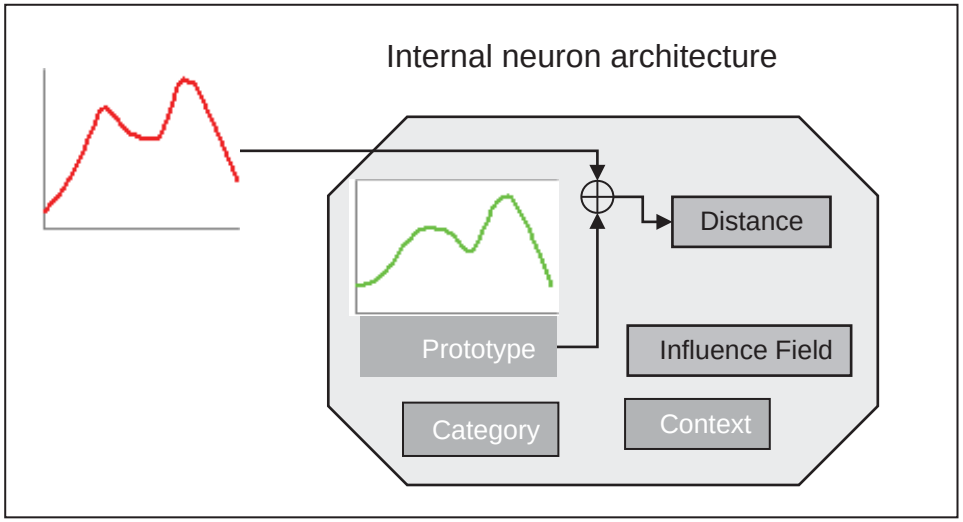
CogniMem (from General Vision): RBF & KNN models

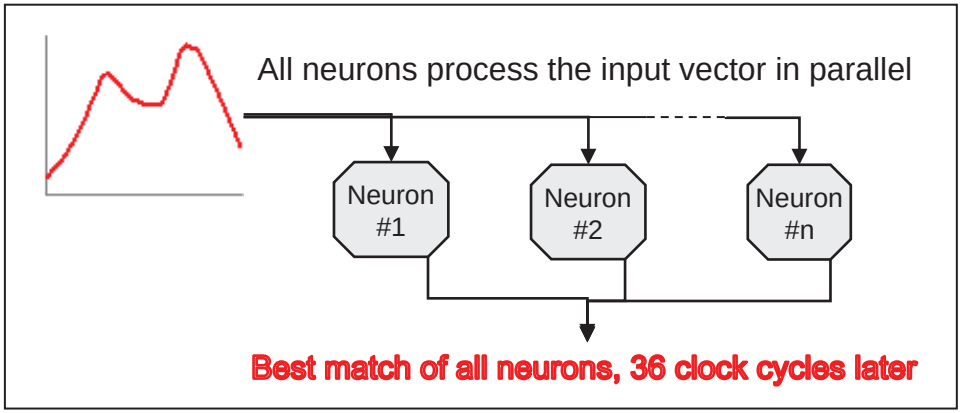


CogniMem: Detailed Block Diagram



A network of neurons in parallel

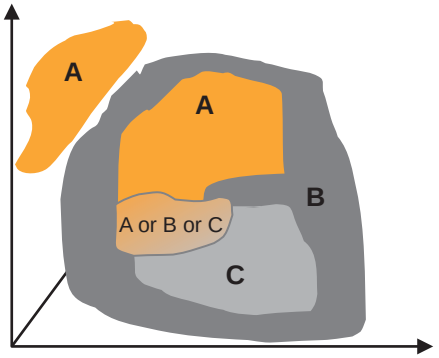




An adaptive model generator

- Learn by examples (supervised or unsupervised)
- Map decision spaces by aggregate instead of hyper planes
- Cope with non-linear, convex, disjoints and embedded categories
- Multiple space generation using different contexts
- Save and restore the contents of the neurons
- Can append more training at any time

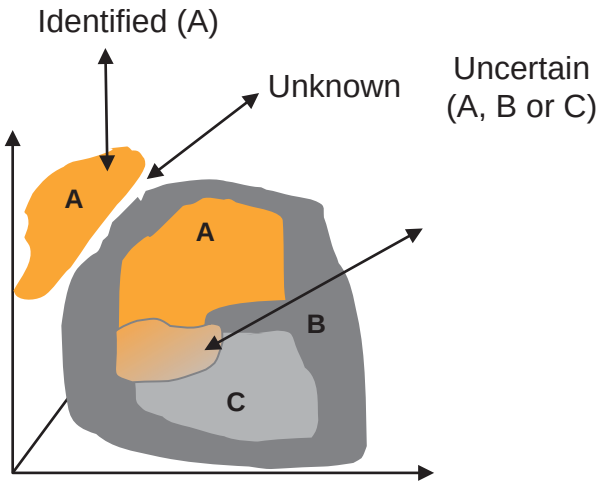
Learning = Building a **“decision spaces”** by learning examples



A high-performance classifier

- Global response readout:
  - positively identified
  - identified with uncertainty
  - unknown
- Detailed response of all the firing neurons
  - -category and confidence level (or distance)
  - -retrieved per decreasing confidence

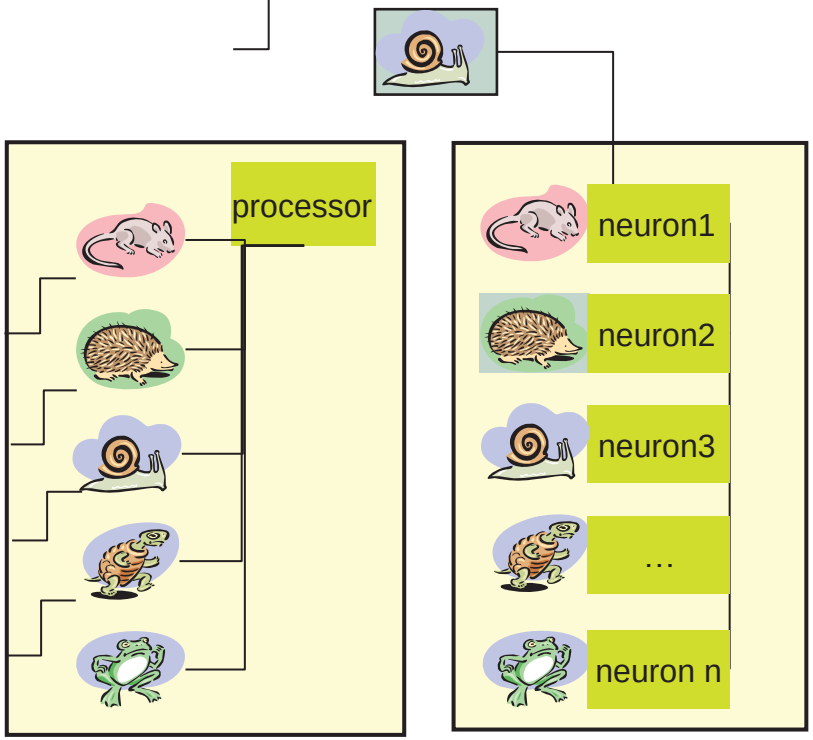
Recognition = where does the sample fall in the decision space?

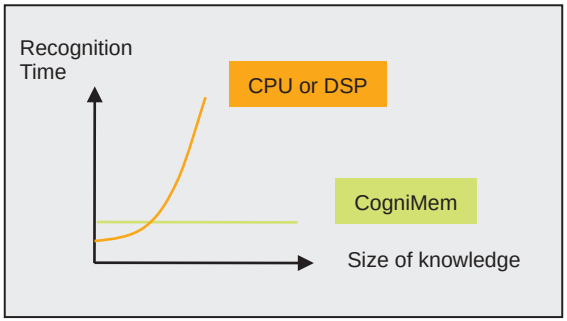


High speed performance

-Constant recognition time independent from the number of neurons in use

- 9.47 usec to broadcast a pattern of 256 bytes to all neurons (@ 27Mhz)
- 1.3 µsec to read to the best match





Comparison with other architectures

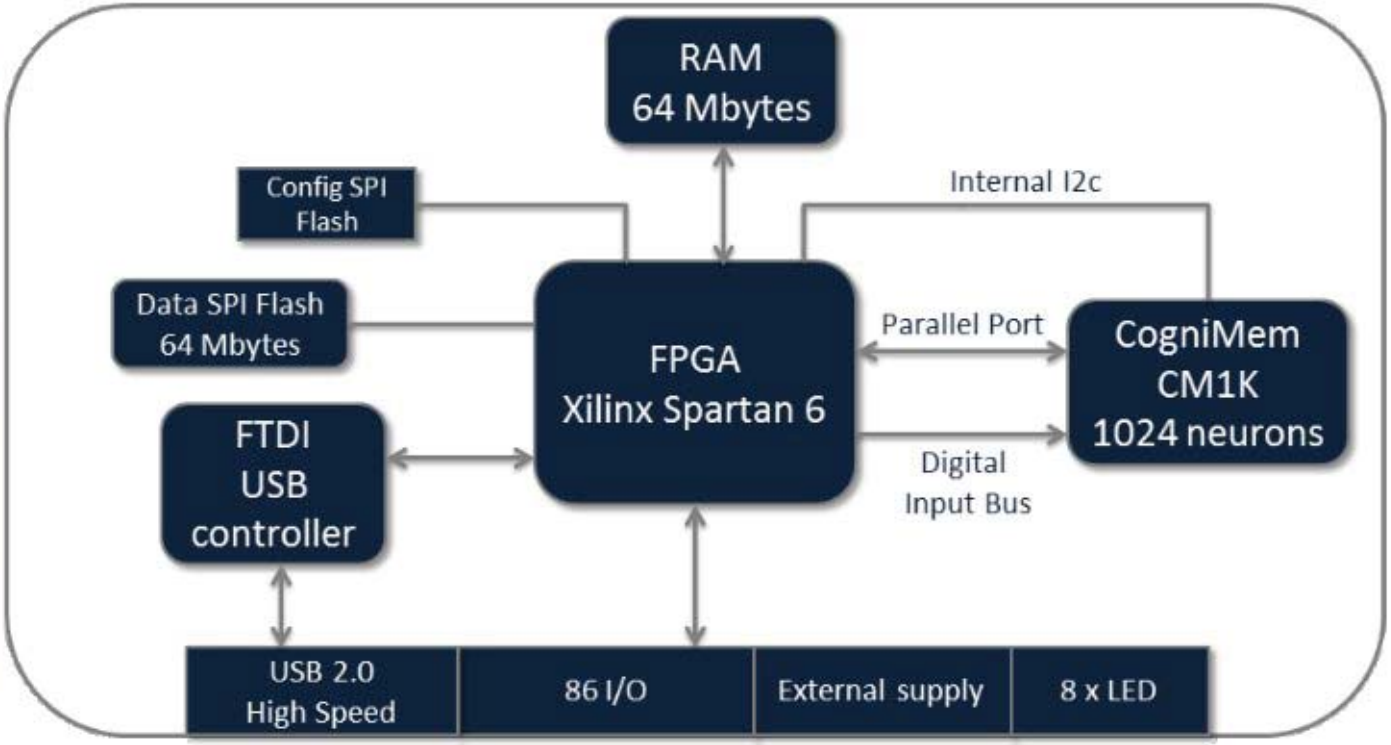
Benchmark: 256-bytes vector at the input of the NN

PC pentium	DSP Shark	DSP Tiger	Neural Processor
2 GHz	100 MHz	300 MHz	27 MHZ
T= 1714 µs	T=7925 µs	T=949 µs	T=10 µs

- 170 times faster than a computer running at 2.4 Ghz
- Capable of **100.000 recognition of a 256-bytes vector /sec**
- 120 times less power than a Pentium (**0.5w vs 60w**)

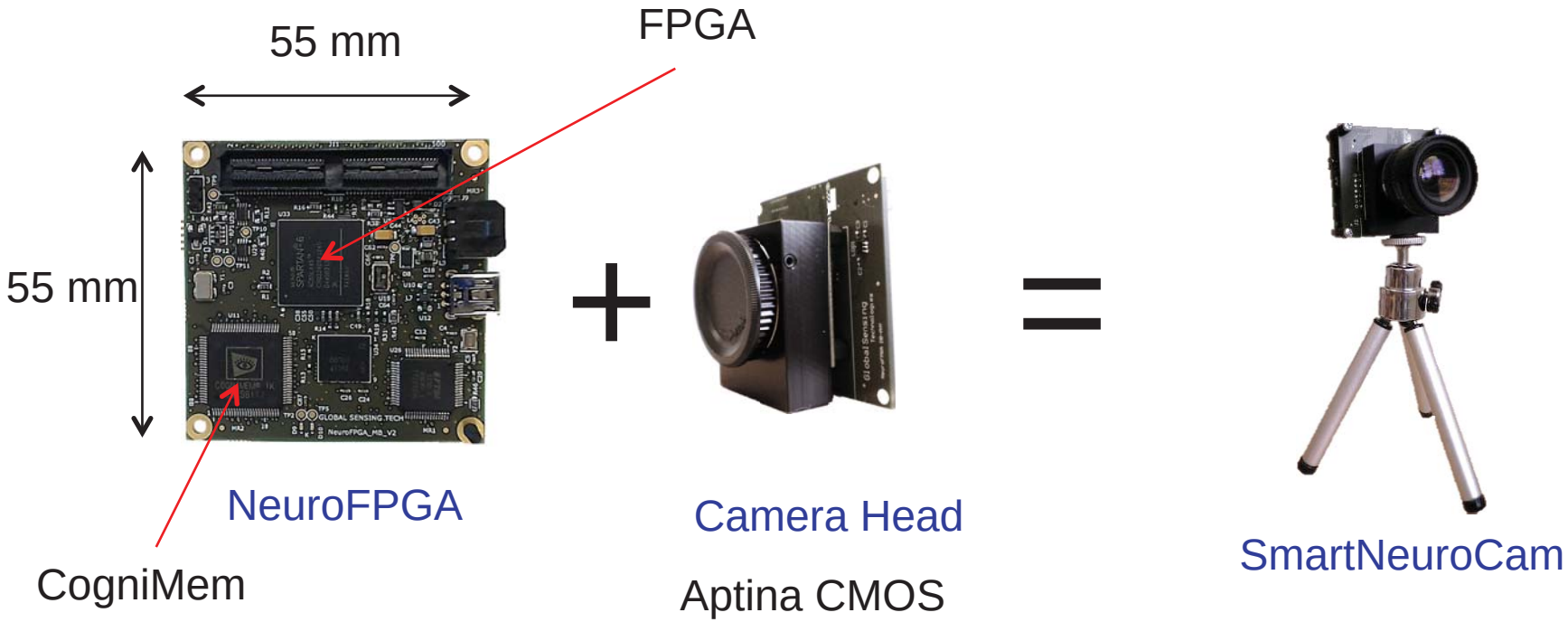
III-Embedded smart vision system (1/2):

SmartNeuroCam

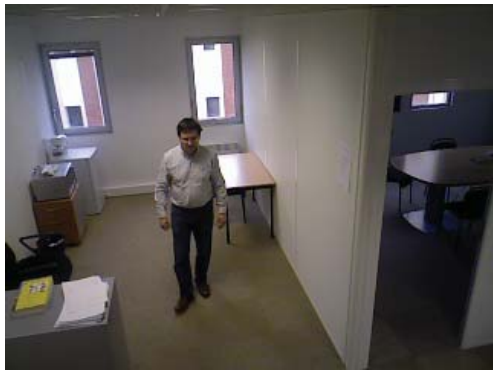


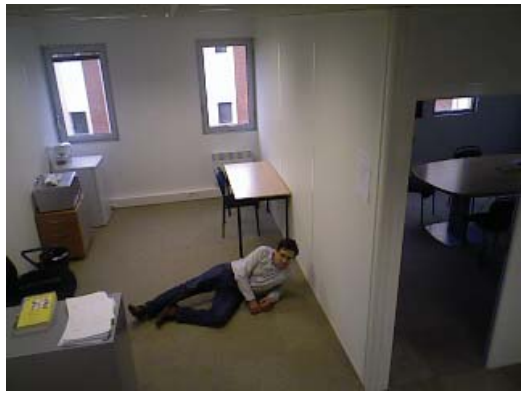
NeuroFPGA board

Embedded smart vision system (2/2):



IV- Application: Real time human pose classification





1) Learning:

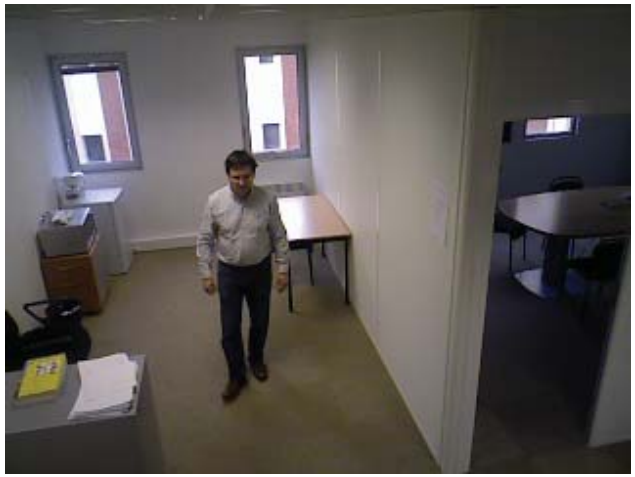
- Sequence acquisition with different poses
- Background elimination → pre-processed image
- Region Of Interest (ROI) off line learning

2) Real time recognition:

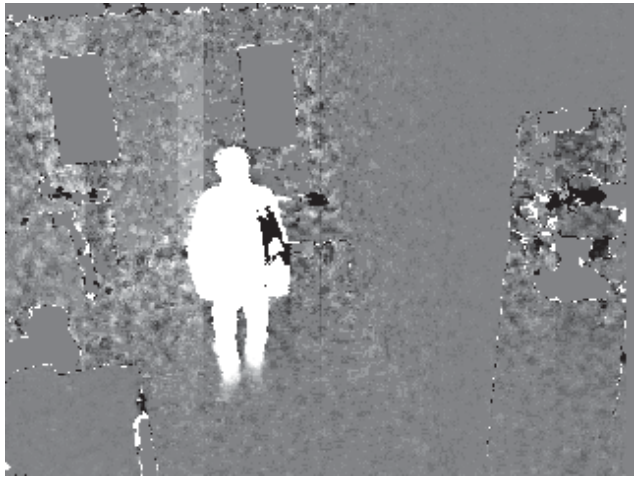
-Current image acquisition
-Background elimination using FPGA
-ROIs scanning (from pre-processed image)
-ROI classification using CogniMem neural processor

Background Elimination

Based on image differencies

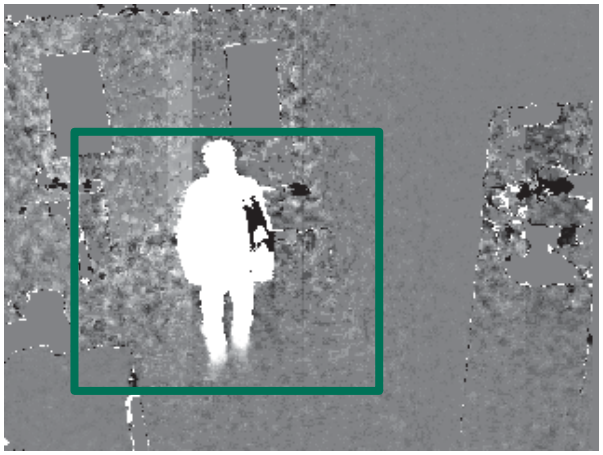


Orignal Image

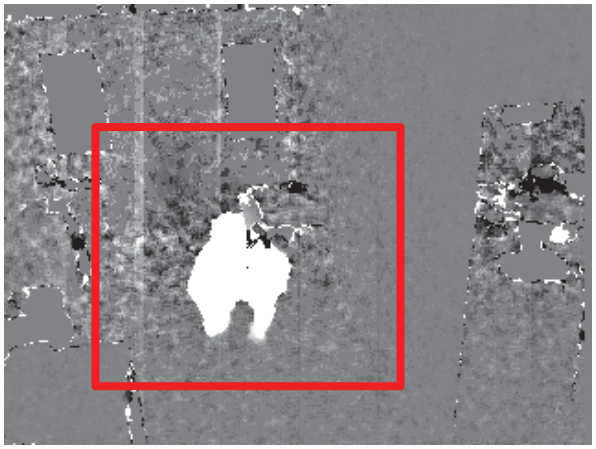


Pre-Processed Image

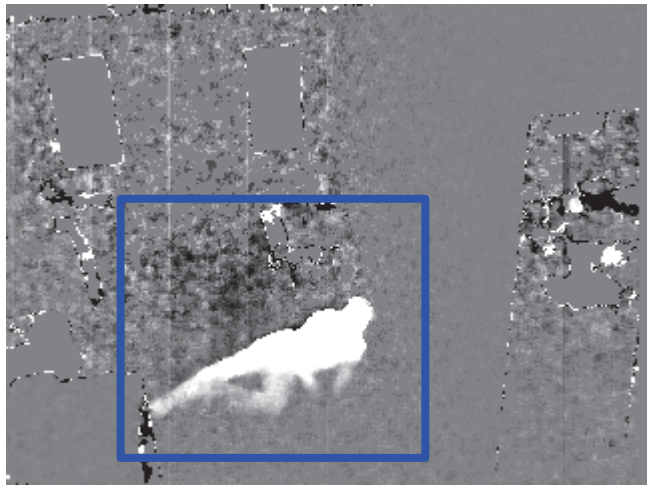
ROI off line learning



Standing Pose

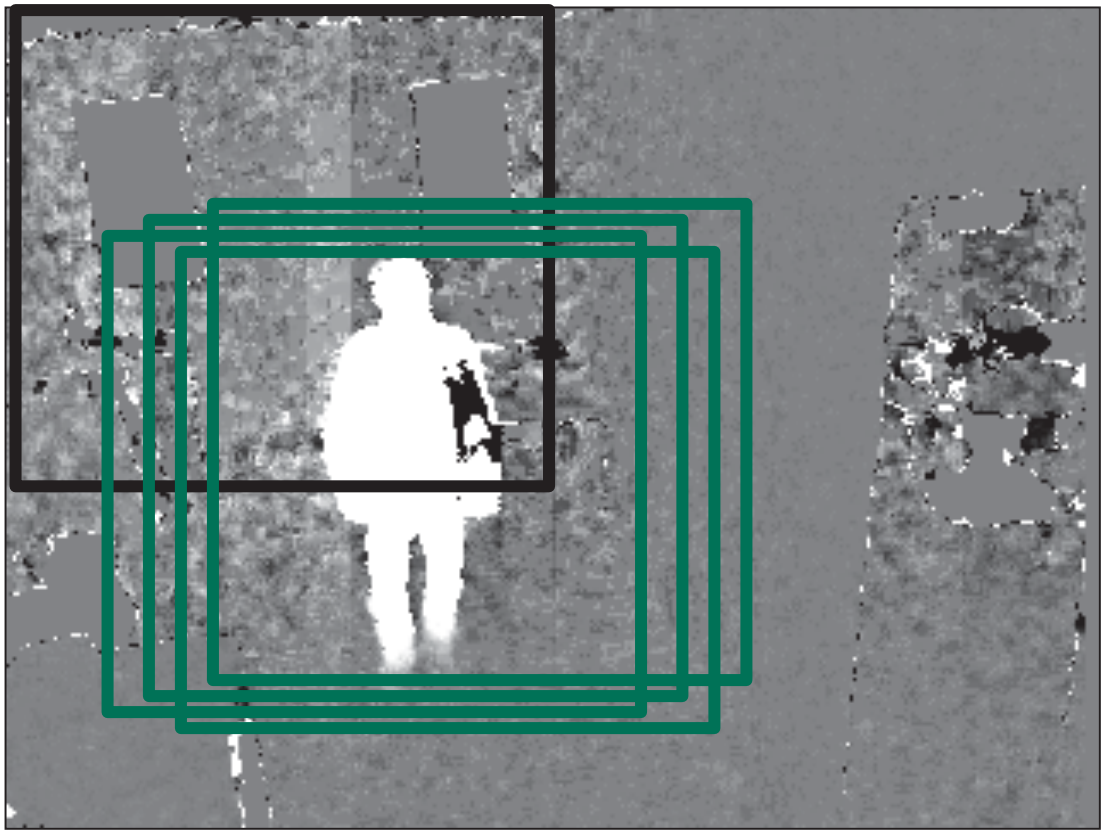


Crounching Pose



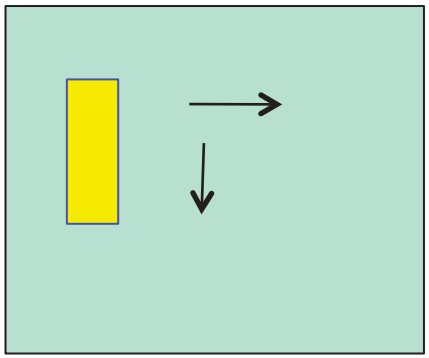
Lying Pose

Real time classification: ROI scanning



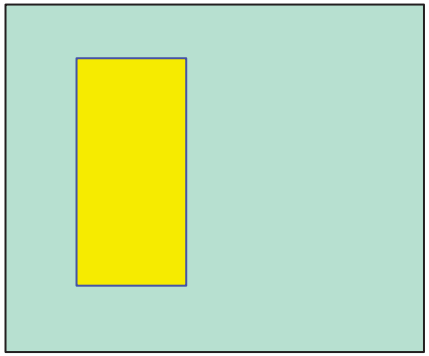
ROI classification obtained with CogniMem

Real Time Image scales analysis < 40ms



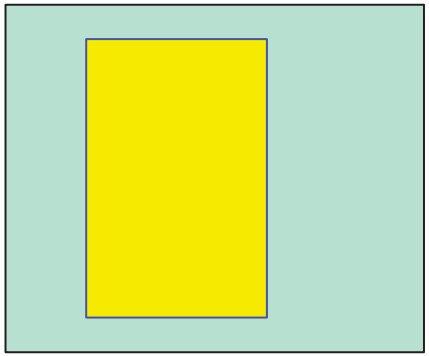
Scale 1: Small pattern sizes
ROI= 80x200 with step 8

→2500 ROI to analyze
→2500x10microsec= 25ms



Scale 2: Medium pattern sizes
ROI=125x300 with step 8

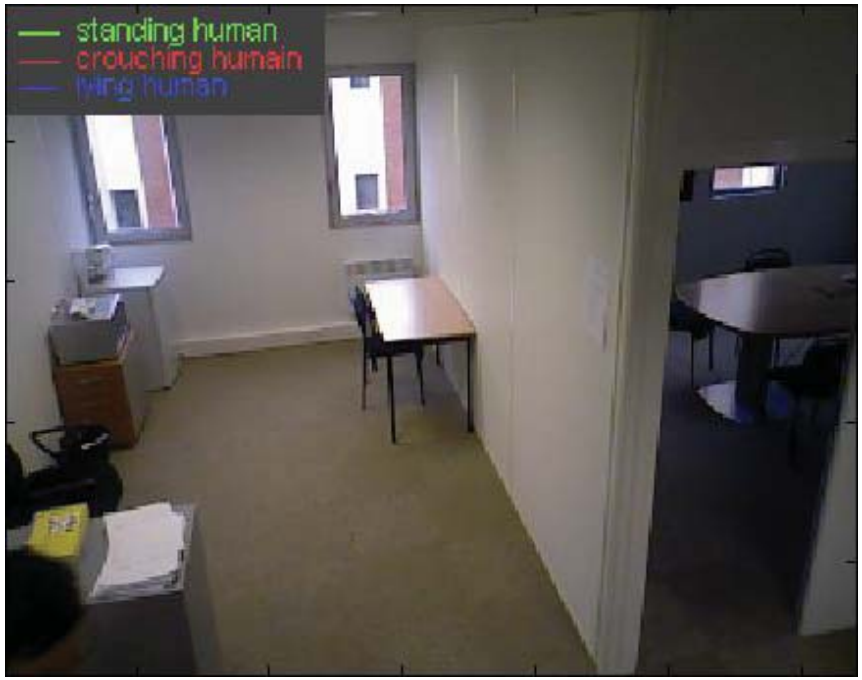
→1000 ROI to analyze
→1000x10microsec= 10ms



Scale 3: Large pattern sizes
ROI= 240x400 with step 8

→250 ROI to analyze
→250x10microsec= 2.5ms

Results overlayed on the original color images



V- Conclusion and Perspectives

Validation of a real time human pose classification using an embedded smart vision system.

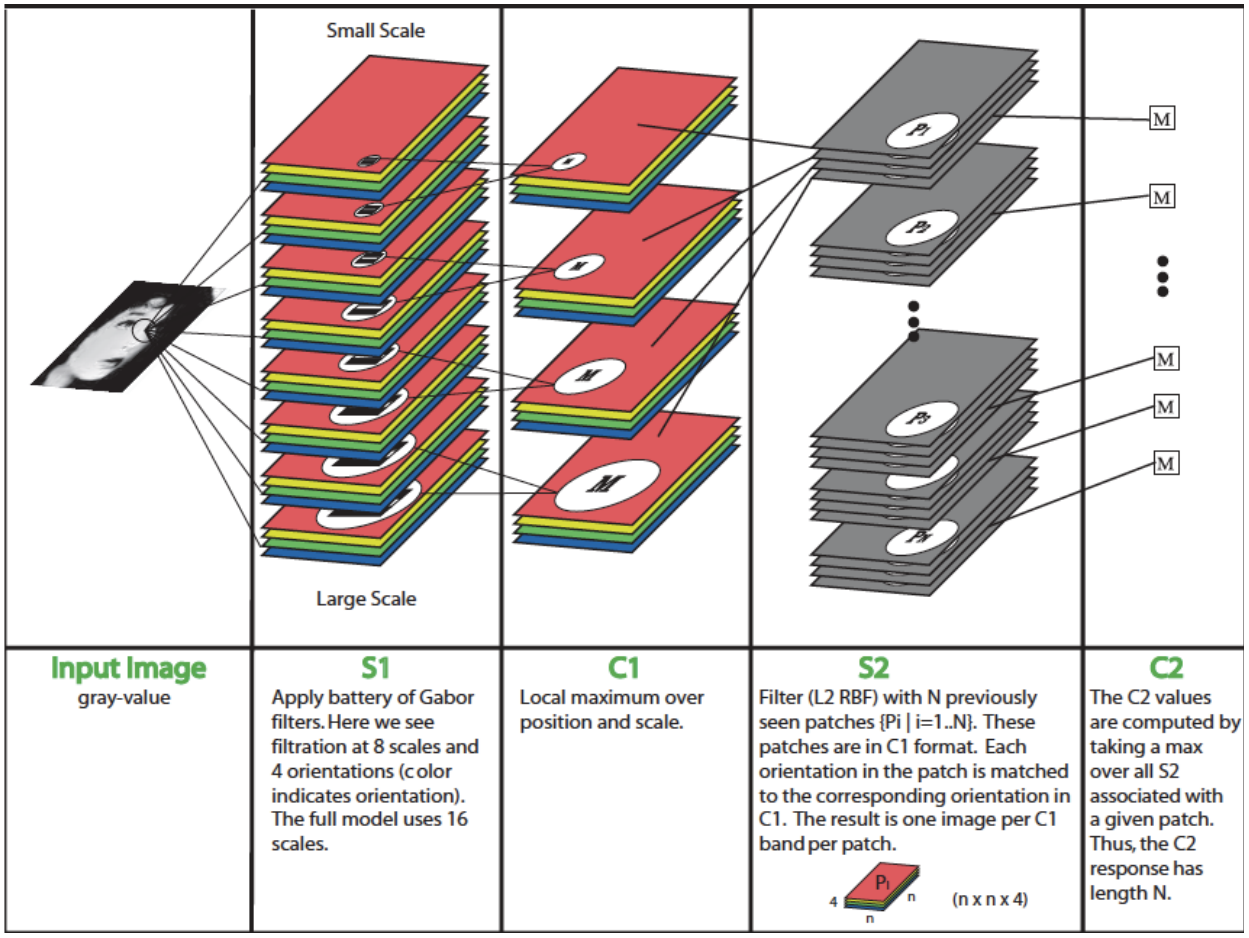
Demonstration of the interest of using Artificial Neural Network for human pose classification (95% of good classification).

Demonstration of the interest of using Neural Processor in words of low-power consumption and speed.

Under study:

- To obtain more robust features, implementation of efficient low-level image pre-processing (like Gabor filter, wavelets, …..) → **Hmax model**
- Implementation inside the same FPGA image pre-processing and classifier based on RBF

Hmax Model Implementation onto FPGA ARTIX-7



Serre et al, IEEE PAMI 2007