Vision Artificielle
Embarquée Neuro-Inspirée

Michel Paindavoine

LEAD UMR CNRS Université de Bourgogne – Dijon, France
paindav@u-bourgogne.fr

GlobalSensing Technologies (GST) – Dijon, France
www.globalsensing.eu
Introduction

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

Simple differences between 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.
Biological Vision
(from Simon Thorpe)

- Feedforward processing
- Only a few milliseconds per processing step
- One spike per neuron
- Very sparse coding
- Processing without context based help

How does state of the art computer vision compare?
Cerveau Humain: quelques chiffres:

Nombre de processeurs (neurones): $86.10^{12}$ dont:

- $16.10^{12}$ dans le cortex
- $4.10^{12}$ dans le système visuel

Fréquence de fonctionnement: 1kHz

Vitesse de conduction: 1-2 m/s

Puissance consommée: 20 W
• Simple Tasks with Human Brain vs Von Neuman Computer (like PC):
  - Calculate in less than one second \( (398387.86 \times 498.07 =?) \)
  - But recognize in less than one second this image:

![Eiffel Tower](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:
    --> Neuro-inspired Human intelligence: Artificial Intelligence on Silicon
Outline

Introduction

Vision Neuro-Inspired models

Artificial Intelligence on Silicon

Embedded smart vision system

Application: Real time human pose classification

Work in Progress
Neuro-Inspired Models: The Hmax

Hmax Model:

Serre et al, IEEE PAMI 2007

Poggio et al., J Neurophysiol, 2007
Neuro-Inspired Models: The Hmax

**Hmax Model: operators**

**S1: (Simple Cell): Spatio-Temporal Directionals filters**
- Simple filters like Sobel
- Gabor filters
- Wavelet based filters

**C1: (Complex Cell)**
- Local Maximum

**S2: (Simple Cell)**
- Convolution with a patch

**C2: (Complex Cell) → V4**
- Neural Network like RBF
Neuro-Inspired Models: The Hmax

S1 layer using Gabor filters

2D Gabor filter

\[
F(x, y) = \exp \left( -\frac{(x_0^2 + y_0^2)}{2\sigma^2} \right) \times \cos \left( \frac{2\pi}{\lambda} x_0 \right)
\]

\[
\begin{pmatrix}
  x_0 \\
  y_0
\end{pmatrix} =
\begin{pmatrix}
  \cos \theta & \sin \theta \\
  -\sin \theta & \cos \theta
\end{pmatrix}
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]
Neuro-Inspired Models: The Hmax

Gabor Filters

Original Image
Neuro-Inspired Models: The Hmax

Gabor Filters

Original Image
Neuro-Inspired Models: The Hmax

Yu & Slotine 2008
Neuro-Inspired Models: The Hmax

Hmax Model performances

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Proposed</th>
<th>[Serre 07]</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplanes</td>
<td>96.0</td>
<td>96.7</td>
<td>94.0 [Fergus 03]</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>98.0</td>
<td>98.0</td>
<td>95.0 [Fergus 03]</td>
</tr>
<tr>
<td>Cars (Rear)</td>
<td>96.0</td>
<td>99.8</td>
<td>84.8 [Fergus 03]</td>
</tr>
<tr>
<td>Leaves</td>
<td>92.0</td>
<td>97.0</td>
<td>84.0 [Weber 00]</td>
</tr>
</tbody>
</table>
Neuro-Inspired Models: Convolutional Neural Network (CNN)

Based on a neural "face/non face" classifier
- Gather a training set of examples
  - 25,212 positive (faces) and 25,487 negative (non faces)
  - using virtual faces, bootstrapping
- Build and train a Convolutional Neural Architecture
  - cascade of convolutions and subsampling
  - 951 trainable weights
  - learnt by backpropagation

Robustness
- minimal face size: 30x30 pixels
- poses (±25° in-plane and ±60° out-of-plane rotation)
- illumination, weak contrast, low resolution, etc.
- partial occlusions (glasses, beard, etc.)
Neuro-Inspired Models: Convolutional Neural Network (CNN)
Neuro-Inspired Models: Convolutional Neural Network (CNN)

Image Database: CMU

90.3% good detection and 3% false detection
Neuro-Inspired Models: Spiking Neurons

Image Pre-Filtering

Spike detection and classification

(Thorpe S. et al.)

1% (107 spikes)  2% (215 spikes)  5% (537 spikes)  10% (1075 spikes)
Artificial Intelligence on Silicon
Artificial Intelligence on Silicon

Pre-Processing

• Analog Retina
  (Delbruck, LETI, UB, TIMA, ...)

• CMOS sensor + FPGAs
  (ISL-ELSI, GST, ...)

• CMOS sensor + ASICs
  (like Features Extractor from BVS)

• CMOS sensor + GPUs

Classification

• FPGAs
  Example: RBF, CNN for
  Face Recognition
  (Paindavoine et al., IEEE TNN
  2003 and IEEE TCSVT 2008)

• ASICs:
  L-Neuro (Philips, 1990)
  ZISC (IBM, 1993)
  CogniMem (GV, 2007)
  NeuFlow (LeCun, 2012)
Artificial Intelligence on Silicon: CogniMem (General Vision)

Feature #1:
- Neurons arranged in parallel

Daisy-chain In

Feature #3:
- Sub-sampling within a region

Digital video signal

Digital Signature

Feature #2:
- Recognition logic

Daisy-chain Out (to add more neurons)

Parallel and I2C buses

Command interpreter

Artificial Intelligence on Silicon: CogniMem (General Vision)
Artificial Intelligence on Silicon: CogniMem, a Neuron Network

All neurons process the input vector in parallel. The best match of all neurons is determined 36 clock cycles later.
Artificial Intelligence on Silicon : CogniMem

A RBF classifier family

- 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?
Artificial Intelligence on Silicon : CogniMem

Comparison with other architectures

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

<table>
<thead>
<tr>
<th>PC pentium</th>
<th>DSP Shark</th>
<th>DSP Tiger</th>
<th>Neural Processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 GHz</td>
<td>100 MHz</td>
<td>300 MHz</td>
<td>27 MHz</td>
</tr>
<tr>
<td>T=1714 µs</td>
<td>T=7925 µs</td>
<td>T=949 µs</td>
<td>T=10 µs</td>
</tr>
</tbody>
</table>

- 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**)
Embedded Smart Vision System (GST)

SmartNeuroCam

NeuroFPGA board
**Embedded Smart Vision System (GST)**

- **CogniMem**
- **NeuroFPGA**
- **FPGA**
- **Aptina CMOS Image sensor:**
  - 752 x 480 pixels @ 60fps
- **Camera Head**

Equals

**SmartNeuroCam**
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

Application:
Real Time Human pose classification (1/6)
Application:
Real Time Human pose classification (2/6)

Background Elimination

Based on spatio-temporal filters

Original Image

Pre-Processed Image
Application:
Real Time Human pose classification (3/6)

ROI off line learning

Standing Pose
Crouching Pose
Lying Pose
Application:
Real Time Human pose classification (4/6)

Real time classification: ROI scanning

ROI classification obtained with CogniMem
Application: Real Time Human pose classification (5/6)

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

Real Time Image scales analysis < 40ms
Application:
Real Time Human pose classification (6/6)

Results overlayed on the original color images
Work in progress: RBF Neurons onto FPGA (1/3)

Implementation of >100 RBF Neurons in the Spartan6 FPGA inside NeuroFPGA board (cooperation with ISL-ELSI)
Work in progress: Spike Camera (2/3)

Implementation inside NeuroFPGA board of spike generation (cooperation with Simon Thorpe: CERCO and SpikeNet Technology)
Conception d’un processeur intégrant au sein d’un même chip des fonctions de traitement du signal et des fonctions neuronales: Hmax, CNN

Objectif: <1w

(collaboration GST, CEA, Morpho, SEB, CTDec, Vitec, Ardpi)
Conclusion

Démonstration des performances des systèmes embarqués neuro-inspirés:

- Systèmes Neuro-Inspirés (Réseaux de neurones Artificiels) pour la détection et le suivi de personnes (95% bonne détection).

- Avantages Processeurs neuronaux:
  **Apprentissage, faible consommation vs vitesse**

Adaptable à différents cas applicatifs:

- Industriel (reconnaissance de défauts),

- Robotique (Guidage),

- Sécurité (Détection, suivi de personnes)

- Médical (Détection chutes)