TOSHIBA

Leading Innovation

Visconti2 - A Heterogeneous Multi-Core SoC for Image-Recognition Applications

Masato Uchiyama, Hideho Arakida, Yasuki Tanabe,
Tsukasa Ike, Takanori Tamai, Moriyasu Banno

Toshiba Corporation, Kawasaki, Japan

Outline

- **Background**
- **Visconti2**
  - Overview of architecture and chip
  - CoHOG accelerator

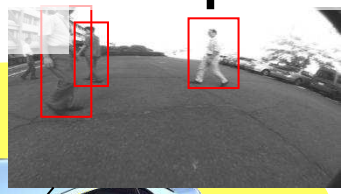
    (Co-occurrence Histograms of Oriented Gradients)
- **Real Applications**
  - Monocular Pedestrian Detection
  - Hand Gesture User Interface (UI)
- **Conclusion**

Background: Targets of Visconti2

Image recognition technology ⇨ A variety of products

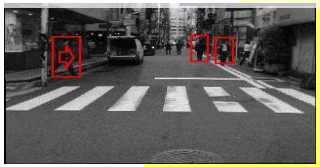
Forward collision warning

Backover prevention

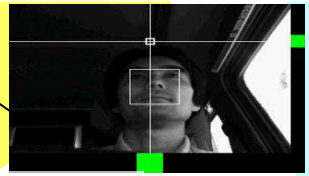


Door security

Pedestrian detection

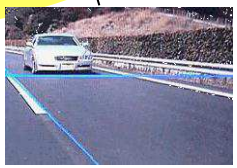


Driver monitoring



Traffic sign recognition

Lane change assistance



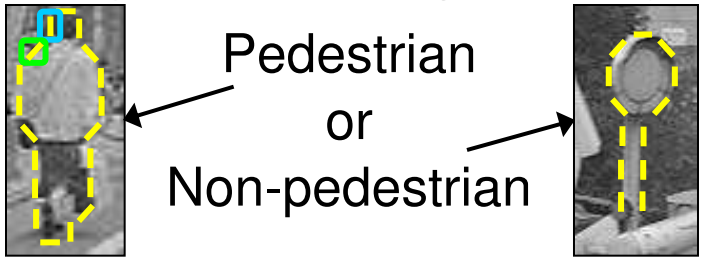
Face tracking for glassless 3D

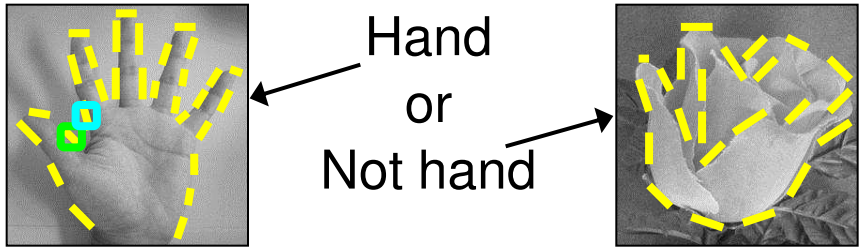
Visconti2 designed for

- **Automotive :** Advanced Driver Assistance Systems (ADAS)
- **Consumer**
- **Industry**

Background: Requirements & Approach

- **High accuracy of object recognition**





CoHOG (Co-occurrence Histograms of Oriented Gradients)
- One of the most accurate image feature descriptors
- Toshiba original (T.Watanabe et al., Proc. PSIVT 2008, pp.37-47)

- **High performance**
  - E.g. Monocular Pedestrian Detection **using CoHOG**
    ➔ 3,983ms/frame on 1GHz CPU

40x speedup required by real-time execution

- **Low power consumption**
  - Cooling without fan (< 1W in typical condition)

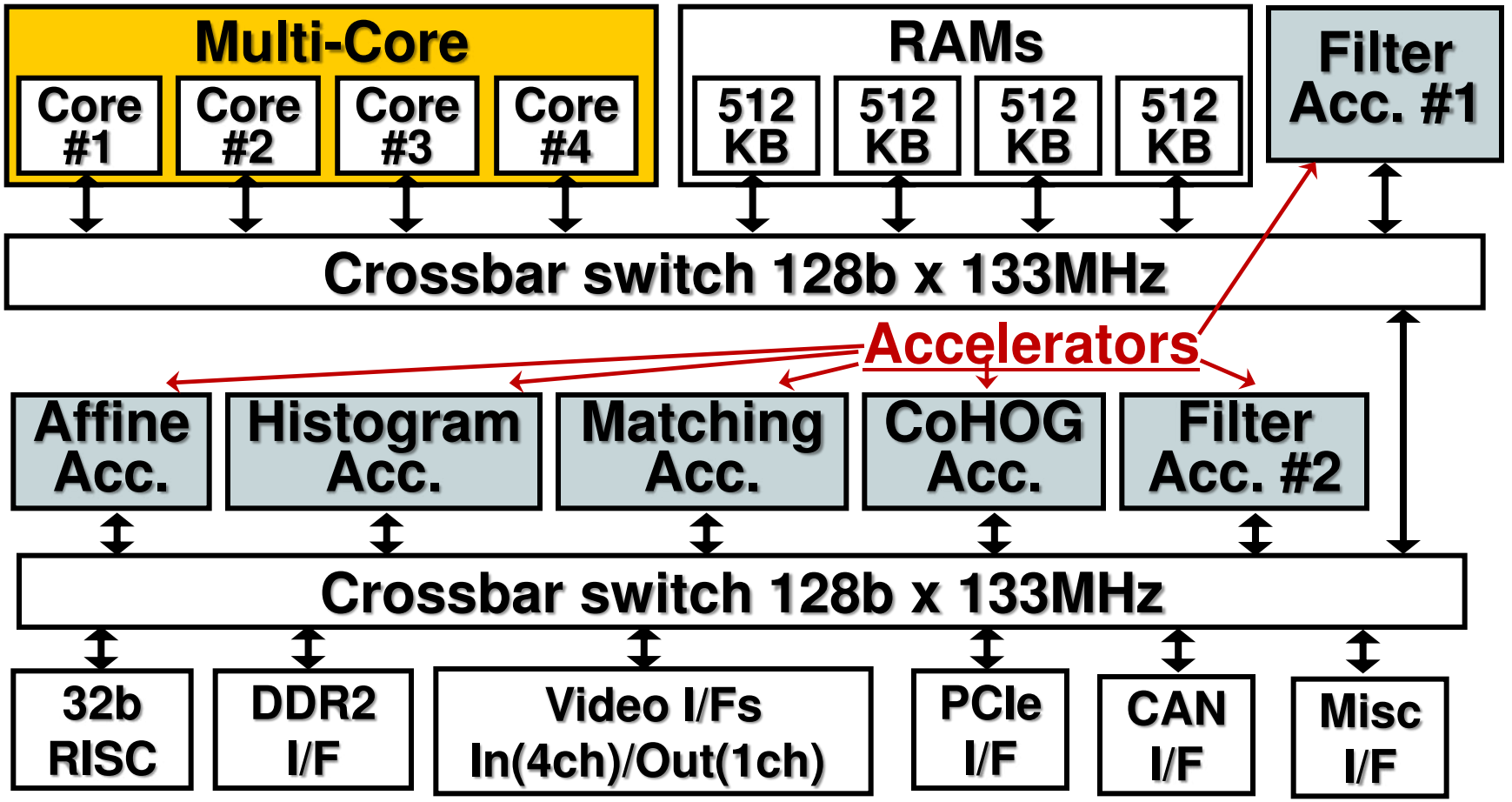
Hardware accelerators for frequently used tasks which are performance bottlenecks (CoHOG, etc.)

Outline

- **Background**
- **Visconti2**
  - Overview of architecture and chip
  - CoHOG accelerator

    (Co-occurrence Histograms of Oriented Gradients)
- **Real Applications**
  - Monocular Pedestrian Detection
  - Hand Gesture User Interface (UI)
- **Conclusion**

Chip Architecture



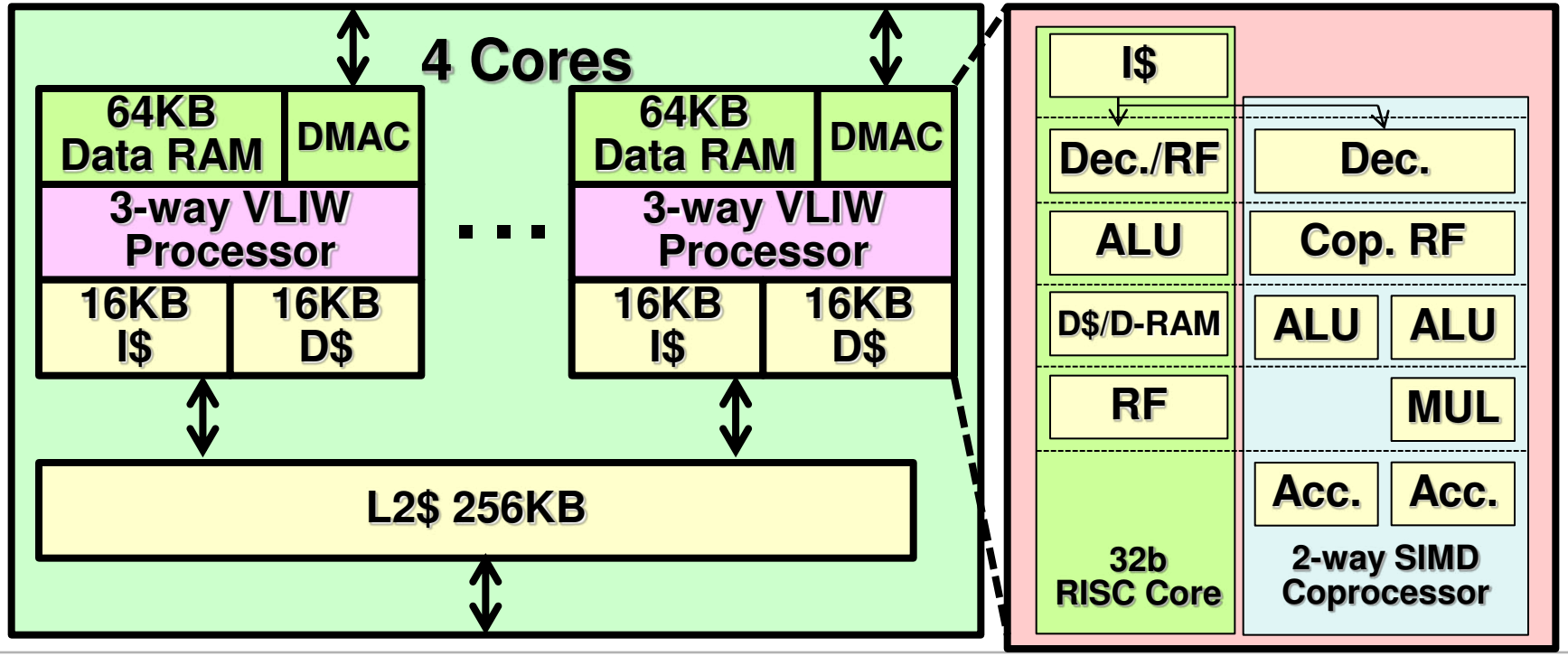
Memory Bandwidth

DDR2: Peak 2GB/sec

On-chip RAMs: 2GB/sec x 4ch.

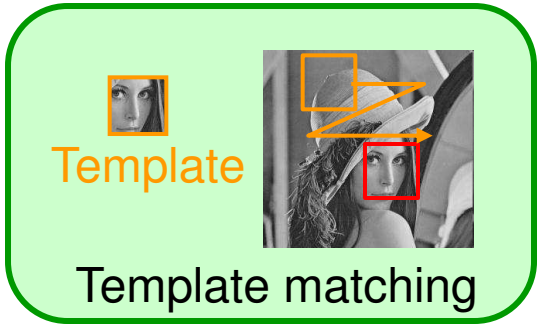
Multi-core Subsystem

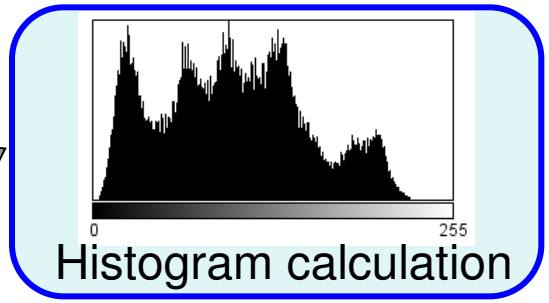
- **Four homogeneous VLIW cores with 256KB L2$**
  - **3-way VLIW core**
    - **RISC core + 2-way SIMD coprocessor (ISSCC '08[S.Nomura])**
    - **Additional 64KB data RAM and DMA controller**
  - **Exploit multi-grain parallelism**
    - **Application, task and thread level parallelism: by four cores**
    - **Data level parallelism: by SIMD coprocessor**

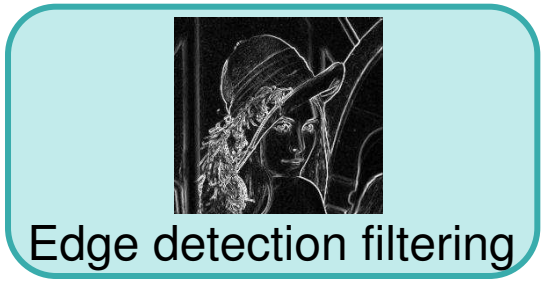


Hardware Accelerators

- Six accelerators implemented
  - **CoHOG accelerator**
  - **Matching accelerator**
  - **Histogram accelerator**
  - **Affine accelerator**
  - **Two Filter accelerators**







Realizing
"High performance with
low power consumption"

➔ **We adopted "Highly parallelized" approach rather than "High clock frequency" approach.**

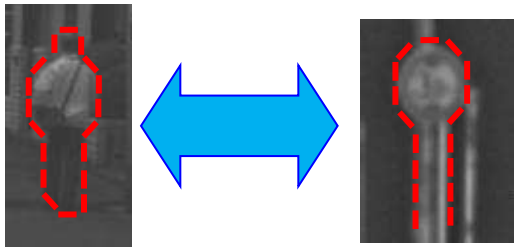
CoHOG based Recognition

- **Extension to widely-used HOG (Histogram of Oriented Gradients)**

1. Make gradient orientation image

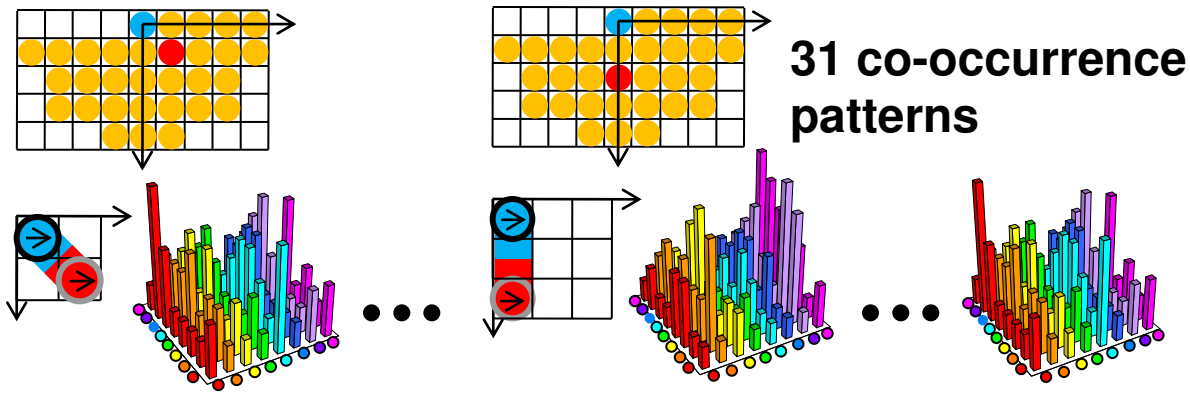
Region of Interest (ROI)

Similar on HOG

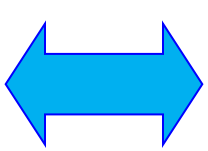


Different on CoHOG

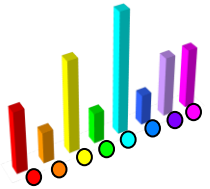
2. Calculate co-occurrence histogram



31 co-occurrence patterns



HOG

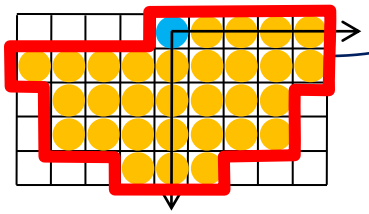


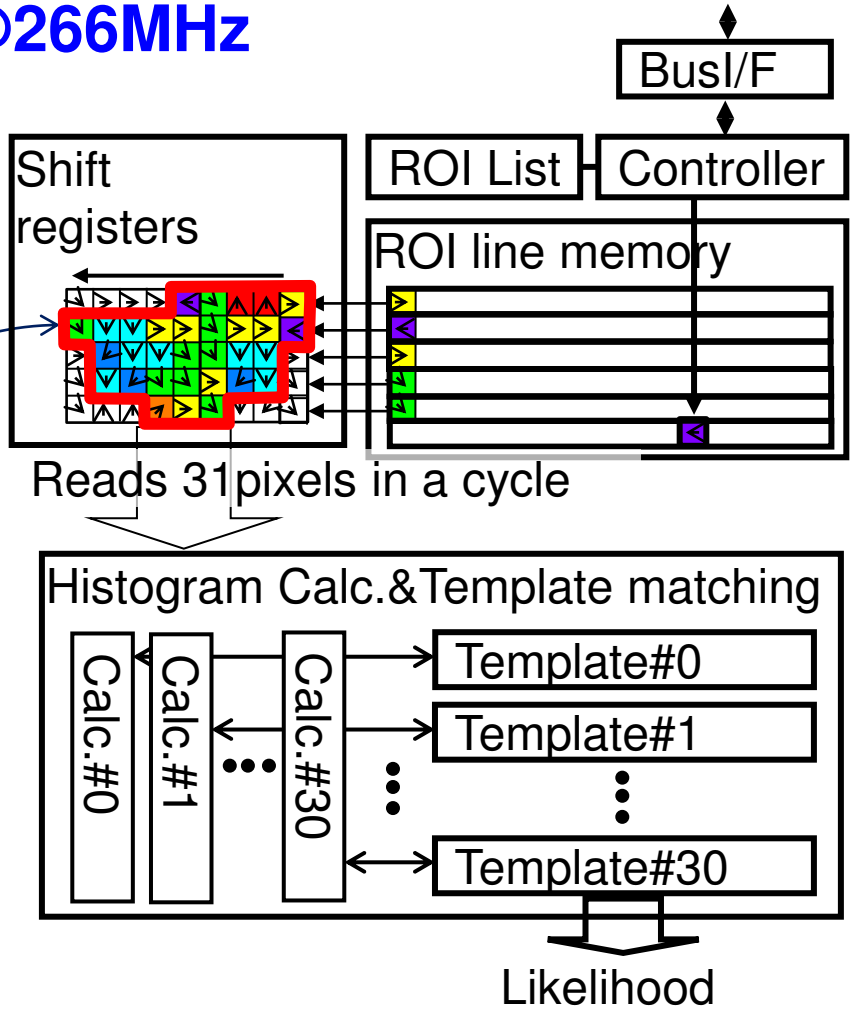
8 gradient orientations

Higher accuracy

CoHOG Accelerator

- **Throughput: 1 pixel / clock @266MHz**
  - **31 co-occurrence pairs are calculated in a clock cycle.**
    - **31 x 3 arithmetic operations**
    - **31 x 2 data references**
    - **Pixel range check**





Over 400,000 ROIs/sec

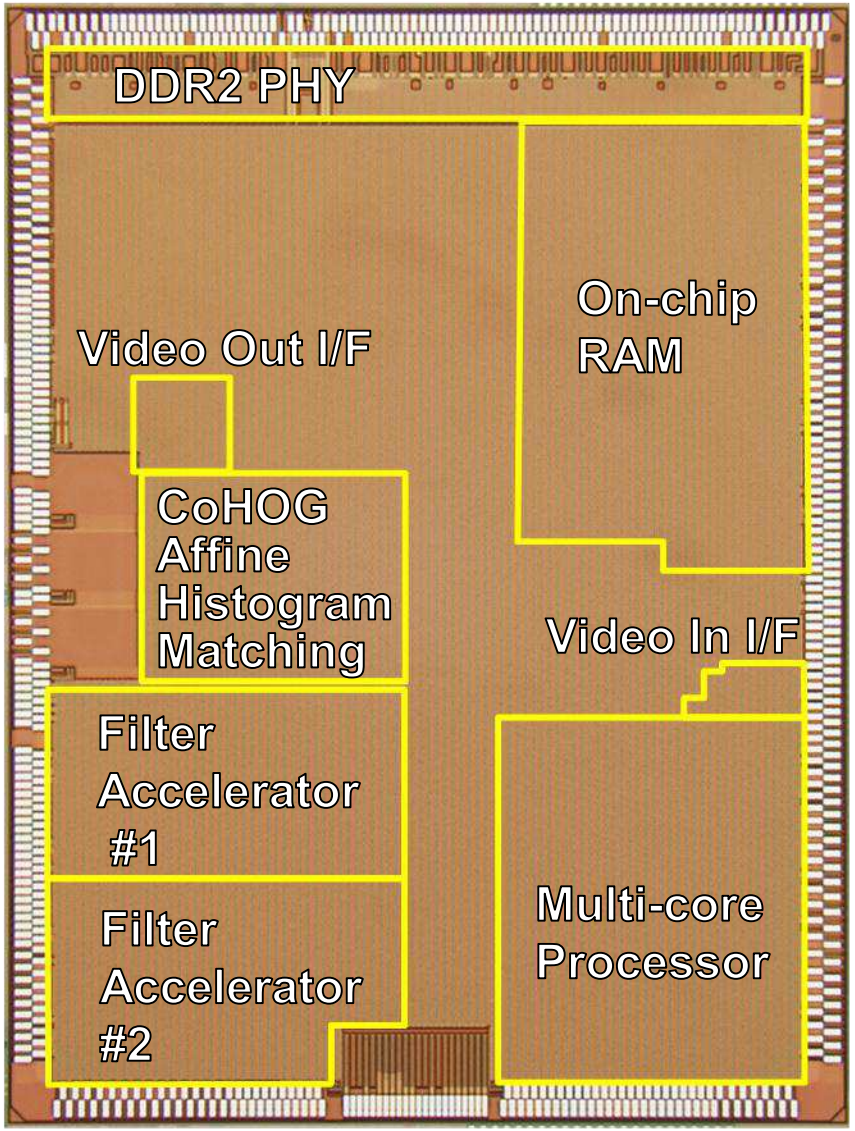
(18 x 36 pixels/ROI)

400,000 ROIs/sec is enough for our target applications.

Features and Chip Micrograph

Process	**40nm**
Chip Size	**44.54mm^2**
Supply Voltages	
Core	**1.1V**
DDR2/PCIe PHY	**1.8V**
I/O	**3.3V**
Performance	
Total peak performance	**464GOPS**
Power efficiency	**620GOPS/W**

(Y.Tanabe et al., Proc. ISSCC 2012, pp.222-223)



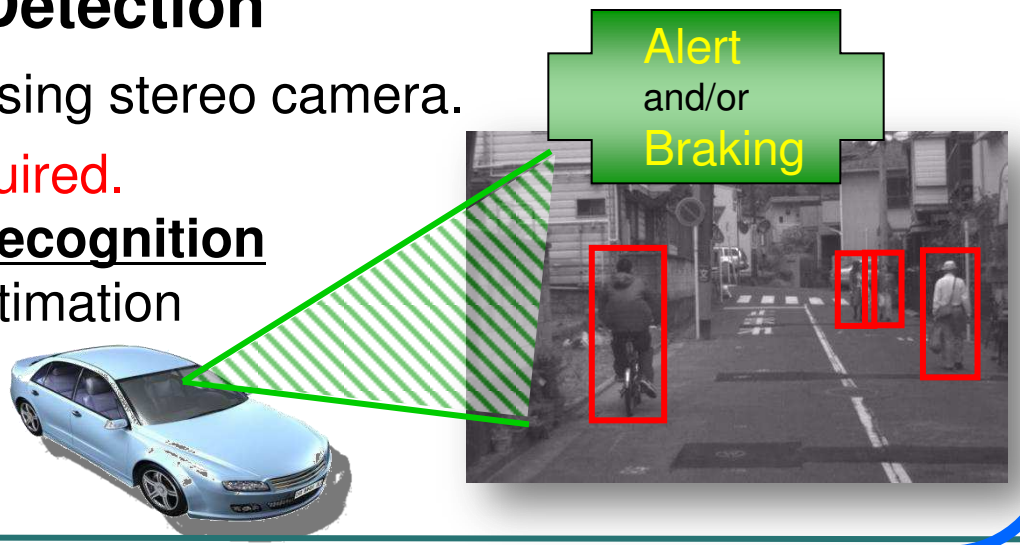
Outline

- **Background**
- **Visconti2**
  - Overview of architecture and chip
  - CoHOG accelerator

    (Co-occurrence Histograms of Oriented Gradients)
- **Real Applications**
  - Monocular Pedestrian Detection
  - Hand Gesture User Interface (UI)
- **Conclusion**

Real Applications

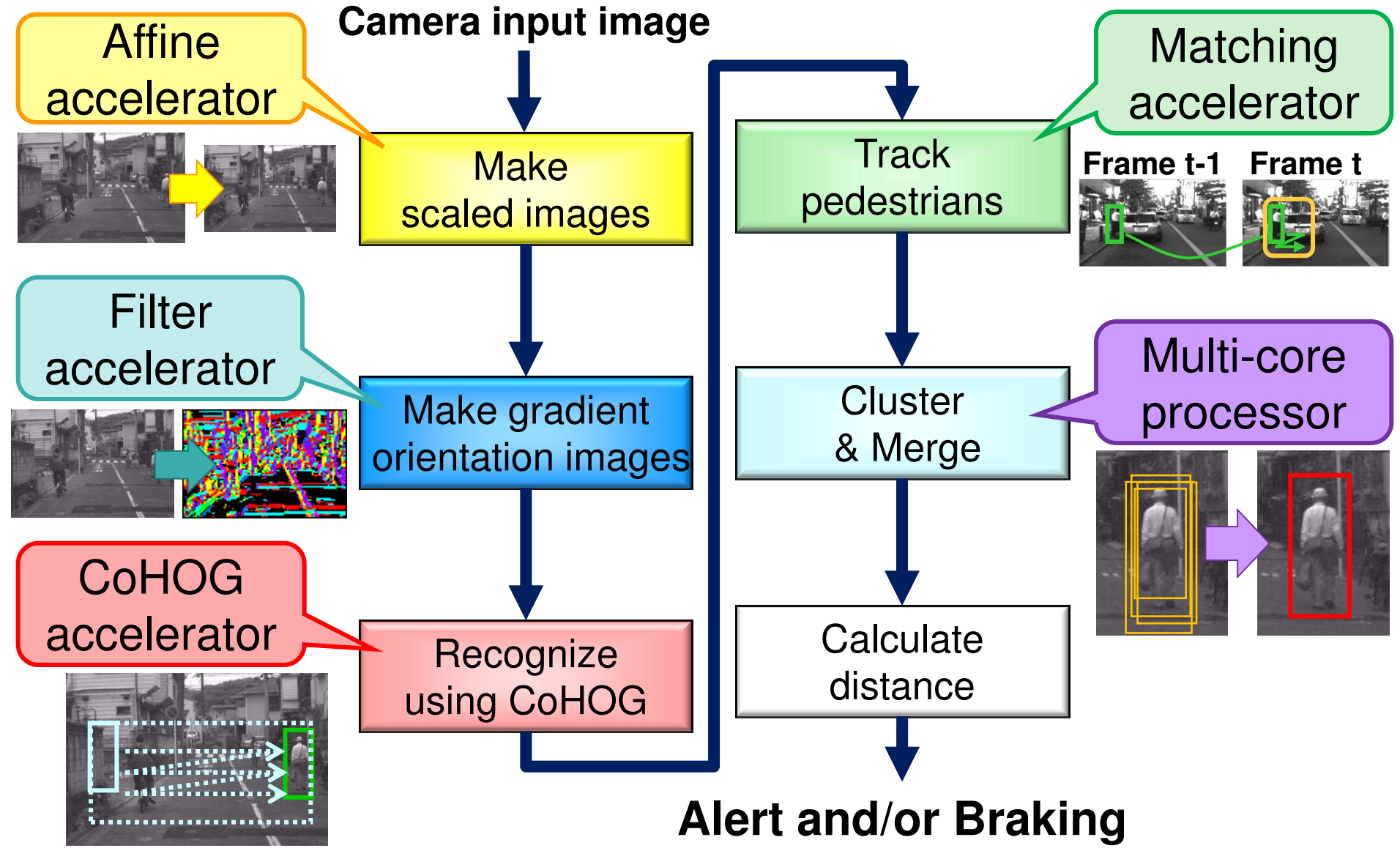
- **Monocular Pedestrian Detection**
  - System cost is lower than using stereo camera.
  - Huge computations are required.
    (**Sliding window CoHOG recognition** is used instead of depth estimation based on stereo matching with stereo camera.)



- **Hand Gesture UI**
  - Hand recognition is applied to many ROIs (**sliding window CoHOG recognition**).
  - High frame rate is required.

Command examples

Pedestrian Detection : Processing Flow

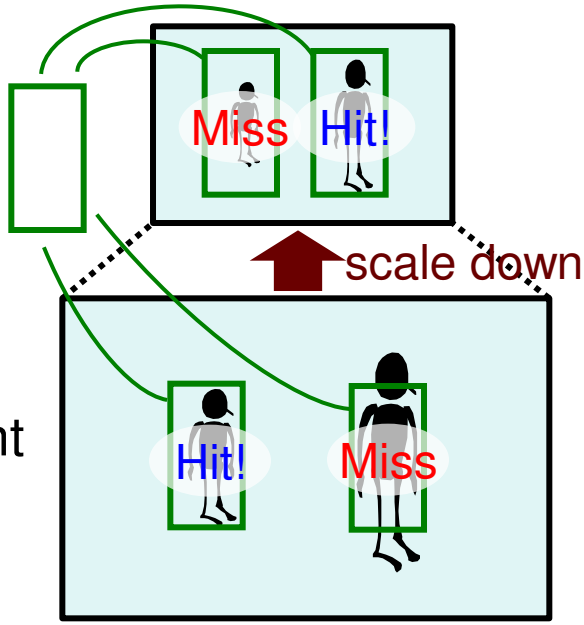


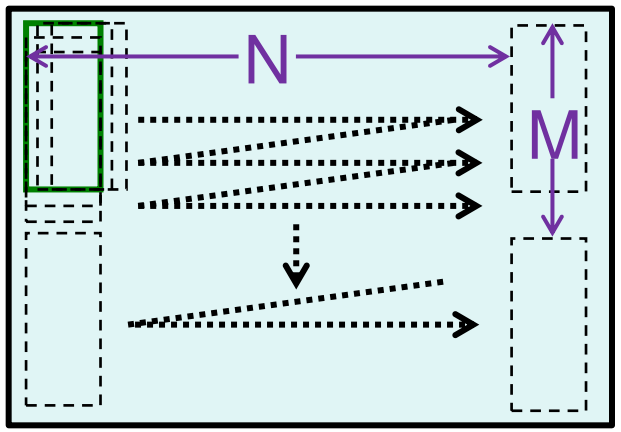
Pedestrian Detection : CoHOG Recognition

- **A number of scaled images are generated by Affine accelerator.**
  - A template is used to match with the scaled images:
    - To detect pedestrians in different distances
    - To detect pedestrians with different body height
- **Sliding window CoHOG recognition**

  ➔ 650 ROIs / image @ VGA
- **Performance requirement of CoHOG recognition**

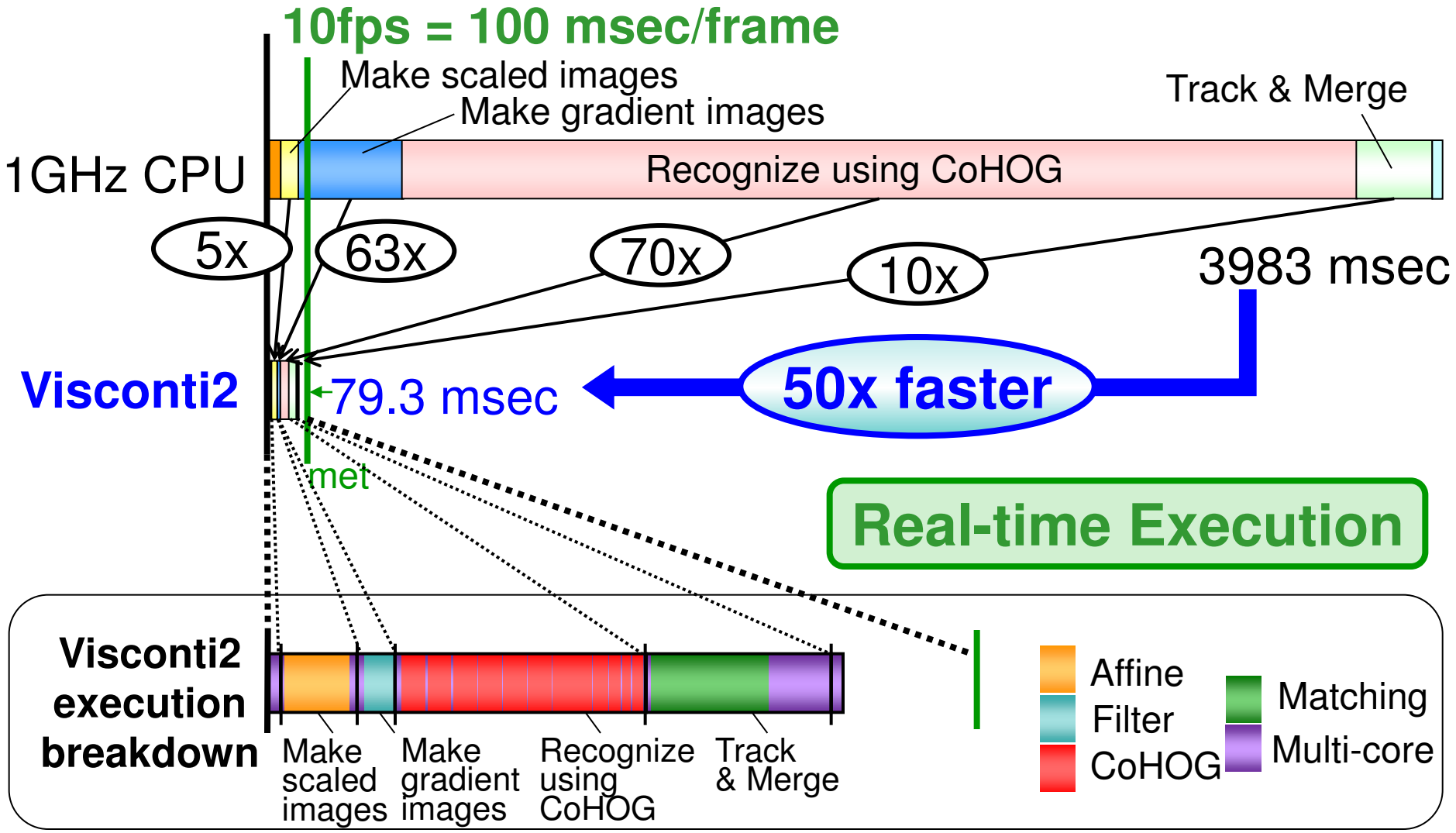
  500 (sliding window ROIs on average)
  x 20 (scaled images)
  x 10 (frame / sec)
  = 100,000 ROIs/sec
  < CoHOG accelerator : 400,000 ROIs/sec





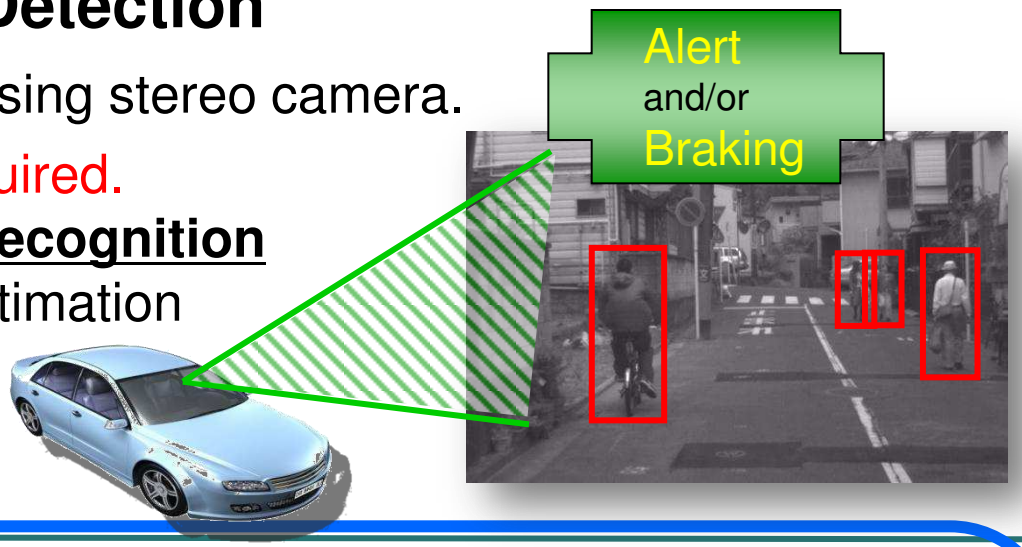
Pedestrian Detection : Execution Time

- **Execution time per frame**



Real Applications

- **Monocular Pedestrian Detection**
  - System cost is lower than using stereo camera.
  - Huge computations are required.
    (**Sliding window CoHOG recognition** is used instead of depth estimation based on stereo matching with stereo camera.)

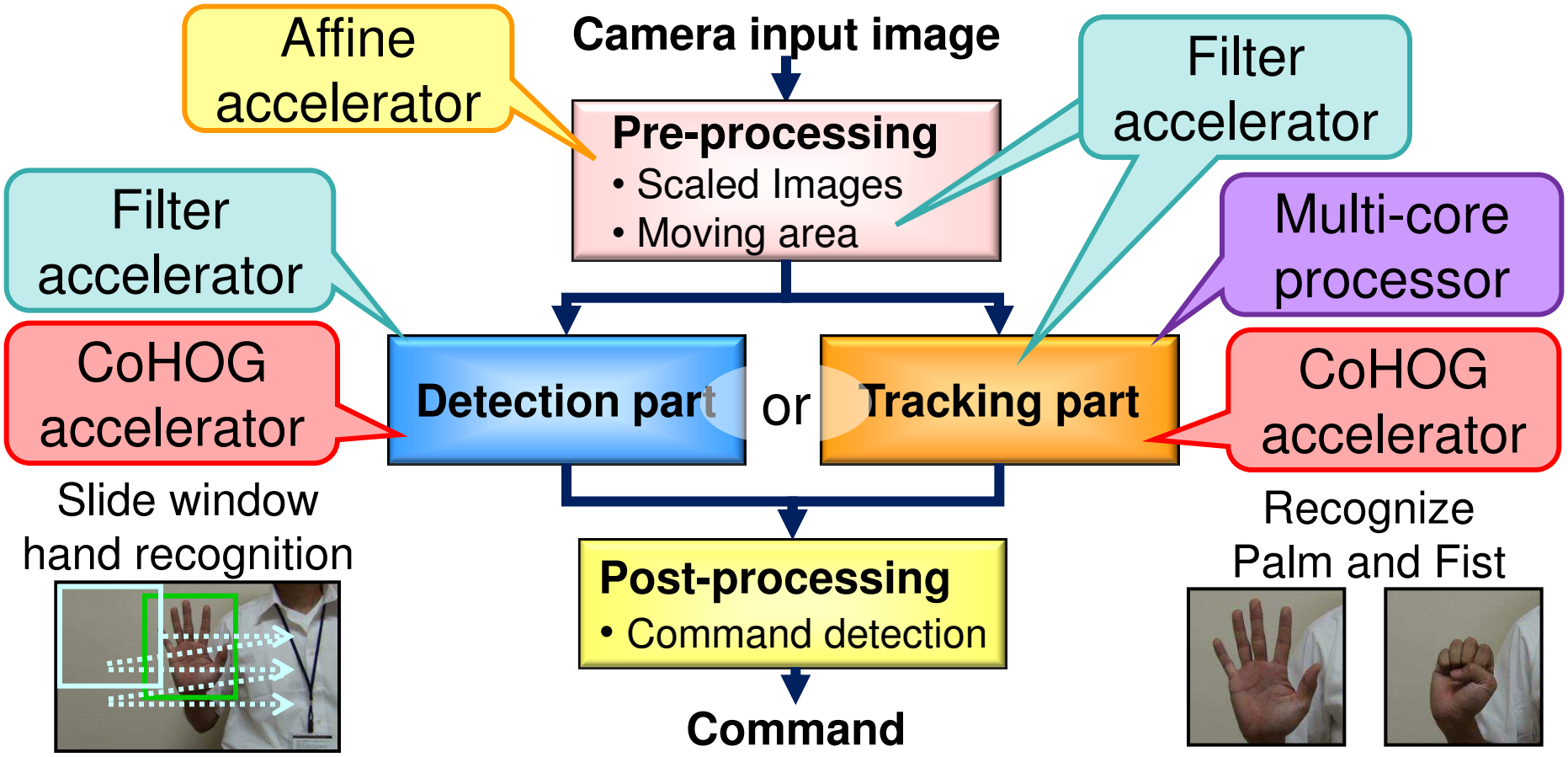


- **Hand Gesture UI**
  - Hand recognition is applied to many ROIs (**sliding window CoHOG recognition**).
  - High frame rate is required.

Command examples

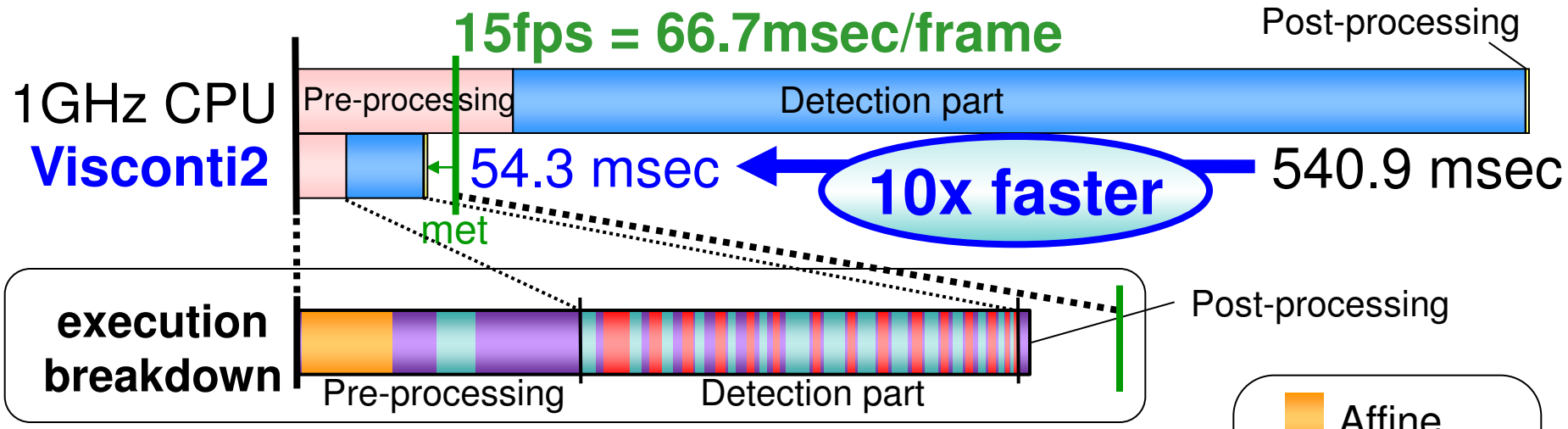
Hand Gesture UI : Processing Flow

- **Switching between two processing modes**
  - **Detection mode** : sliding window hand recognition @ 15fps
  - **Tracking mode** : trajectory recognition @ 30fps

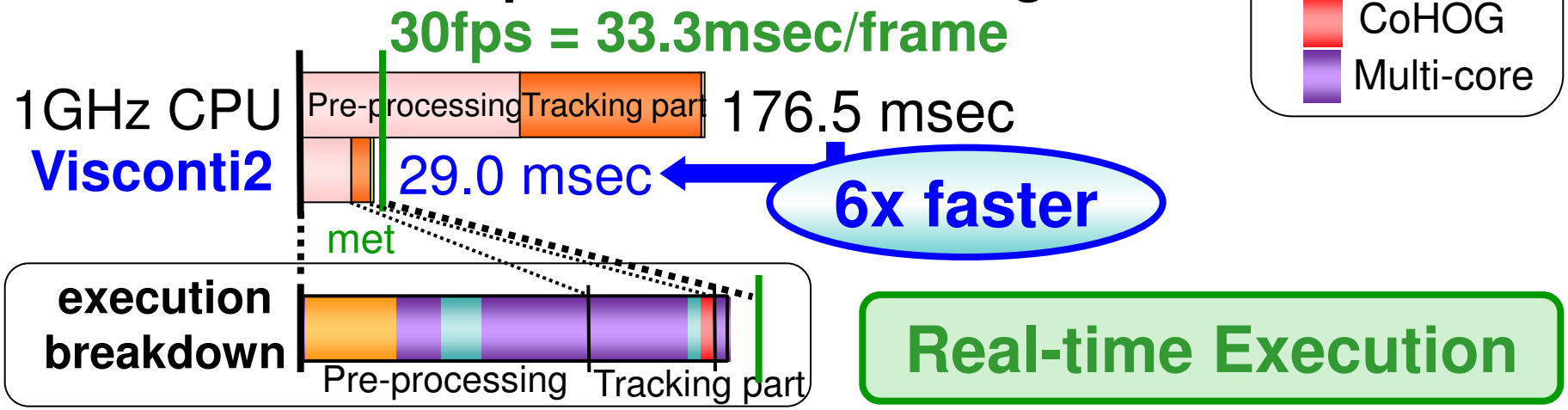


Hand Gesture UI : Execution Time

- **Execution time per frame in detection mode**



- **Execution time per frame in tracking mode**

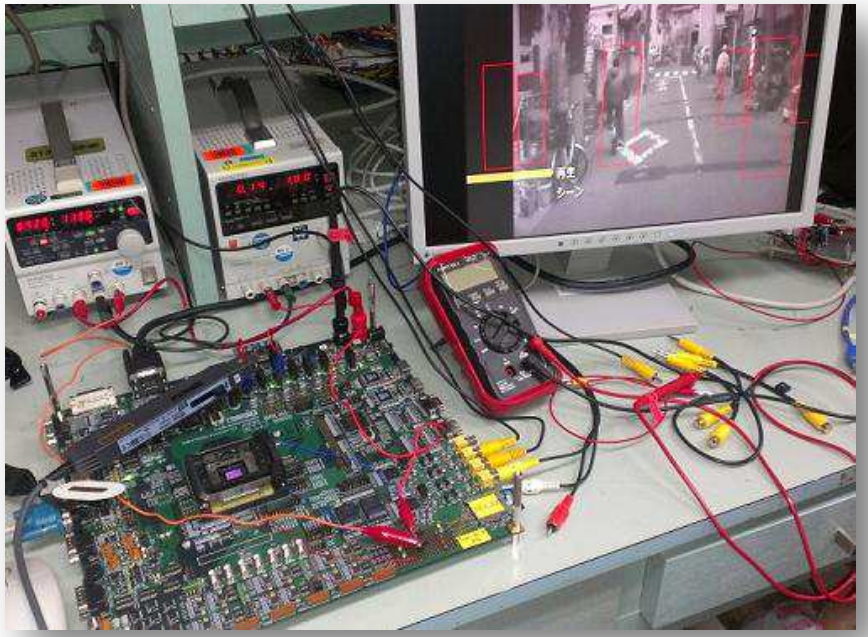


Evaluation of Power Consumption

- **Monocular Pedestrian Detection**
  - Chip total : 870mW
    - Core (1.1V) : 356mW
    - PHY(1.8V) : 460mW
    - I/O (3.3V) : 54mW
- **Hand Gesture UI**
  - Chip total : 891mW
    - Core (1.1V) : 363mW
    - PHY(1.8V) : 472mW
    - I/O (3.3V) : 56mW

< 1W : Cooling without fan

Typical condition:
Process center sample, 25ºC



Evaluation board and
power measurement environment

Conclusion

- **Visconti2 is a heterogeneous multi-core SoC dedicated for image recognition.**

  Visconti2 achieves:

  – Accurate recognition
    - CoHOG based image recognition is implemented.

  – High performance with low power consumption
    - We implemented six highly parallelized hardware accelerators.
    - Under 1W power consumption is achieved. (typical condition)

- **Two real applications on Visconti2 using HW accelerators are demonstrated.**

  – Monocular Pedestrian Detection

  – Hand Gesture User Interface

- **Visconti2 status: ES ready**

  http://www.semicon.toshiba.co.jp/eng/product/assp/selection/automotive/infotain/visconti/

TOSHIBA

Leading Innovation