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 $\rightarrow$ 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**
  - Pedestrian or Non-pedestrian
  - Hand or Not hand

  **CoHOG** (Co-occurrence Histograms of Oriented Gradients)
  - One of the most accurate image feature descriptors

- **High performance**
  - E.g. Monocular Pedestrian Detection using CoHOG
    \[ \text{3,983ms/frame on 1GHz CPU} \]
    \[ \text{40x speedup required} \]

- **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

- Multi-Core
  - Core #1
  - Core #2
  - Core #3
  - Core #4

- Crossbar switch 128b x 133MHz

- RAMs
  - 512 KB
  - 512 KB
  - 512 KB
  - 512 KB

- Filter Acc. #1

- Accelerators
  - Affine Acc.
  - Histogram Acc.
  - Matching Acc.
  - CoHOG Acc.
  - Filter Acc. #2

- Crossbar switch 128b x 133MHz

- Interfaces
  - 32b RISC
  - DDR2 I/F
  - Video I/Fs
  - PCIe I/F
  - CAN I/F
  - Misc I/F

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)

2. Calculate co-occurrence histogram

31 co-occurrence patterns

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)$^{36}$

400,000 ROIs/sec is enough for our target applications.
### Features and Chip Micrograph

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>40nm</td>
</tr>
<tr>
<td>Chip Size</td>
<td>44.54mm²</td>
</tr>
<tr>
<td>Supply Voltages</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>1.1V</td>
</tr>
<tr>
<td>DDR2/PCIe PHY</td>
<td>1.8V</td>
</tr>
<tr>
<td>I/O</td>
<td>3.3V</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Total peak performance</td>
<td>464GOPS</td>
</tr>
<tr>
<td>Power efficiency</td>
<td>620GOPS/W</td>
</tr>
</tbody>
</table>

(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

- move
- select
- cancel
Pedestrian Detection : Processing Flow

Camera input image
- Make scaled images

Affine accelerator
- Make gradient orientation images

Filter accelerator
- Recognize using CoHOG

CoHOG accelerator
- Track pedestrians

Matching accelerator
- Cluster & Merge

Matched accelerator
- Calculate distance

Multi-core processor
- Alert and/or Braking
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**
  
  \[ 650 \text{ ROIs / image @ VGA} \]

- **Performance requirement of CoHOG recognition**

  \[
  \begin{align*}
  &500 \text{ (sliding window ROIs on average)} \\
  &\times 20 \text{ (scaled images)} \\
  &\times 10 \text{ (frame / sec)} \\
  =& 100,000 \text{ ROIs/sec} \\
  < \text{CoHOG accelerator : 400,000 ROIs/sec}
  \end{align*}
  \]
Pedestrian Detection: Execution Time

- Execution time per frame

10fps = 100 msec/frame

Visconti2

1GHz CPU

Make scaled images

Make gradient images

Recognize using CoHOG

Track & Merge

3983 msec

50x faster

Real-time Execution

Visconti2 execution breakdown

Make scaled images

Make gradient images

Recognize using CoHOG

Track & Merge

5x

63x

70x

10x

79.3 msec

50x faster
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**
- move
- select
- cancel
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
  
  **1GHz CPU Visconti2**
  
  Pre-processing | Detection part
  
  **15fps = 66.7msec/frame**
  
  Pre-processing | Detection part
  
  **54.3 msec**
  
  **10x faster**
  
  Post-processing
  
  Pre-processing | Detection part
  
  **540.9 msec**
  
  Post-processing
  
- Execution time per frame in tracking mode
  
  **30fps = 33.3msec/frame**
  
  1GHz CPU Visconti2
  
  Pre-processing | Tracking part
  
  **176.5 msec**
  
  Pre-processing | Tracking part
  
  **29.0 msec**
  
  **6x faster**
  
  Pre-processing | Tracking part
  
  Real-time Execution
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, 25°C

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/