Heterogeneous Computing for a Smarter City

Mr. Jinshui Liu
Chief Architect for IT Hardware
Huawei IT Product Line
Smart City, Many Different Faces, All for Better Livings

- Smart Government
- Smart Healthcare
- Smart Manufacturing
- Smart Energy
- Safe City
- Smart Transport
- Smart Home
- Smart Building
- Smart Manufacturing
- Smart Energy
- Safe City
- Smart Transport
- Smart Home
- Smart Building
- Smart Manufacturing
- Smart Energy
- Safe City
- Smart Transport
- Smart Home
- Smart Building
- Smart Manufacturing
- Smart Energy
- Safe City
- Smart Transport
- Smart Home
- Smart Building
Our Smart/Safe City Mission:

- Create a Better Life
- Attract More Talents and Investments
- Promote More Business Opportunities
Super fast Computing Required to:
- Train ML/DL neural networks
- Inference ML/DL neural networks
- ...

for

• Image & Video analysis
• Facial recognition & real-time facial feature query
• NLP for voice activated services
• Billions of rows of big data real-time query
• ...

To Enable Smart City

Homogeneous Computing

Heterogeneous Computing

Heterogeneous Computing is more efficient for AI

NVIDIA DGX-1 Delivers 96X Faster Training

DGX-1 with Tesla V100
8X GPU Server
CPU-only Server

2.4 hours, 96X faster
18 hours, 40X faster
711 hours

Workload: ResNet50, 99 epochs to solution | CPU Server: Dual Xeon E5-2679 v4, 2.5GHz

Source: Nvidia 2017

Why Heterogeneous Computing for Smart City?
### Building the Base FFV Database

**Government Agencies Have Many Ways to Collect individuals’ Facial Feature Vector (FFV) Data**

<table>
<thead>
<tr>
<th>Individual-1</th>
<th>Individual-X</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFV1 for ID Card</td>
<td>FFV2 – Passport</td>
</tr>
<tr>
<td>FFV2 – Passport</td>
<td>FFV3 – DL</td>
</tr>
<tr>
<td>FFV3 – DL</td>
<td>FFV4 – Social Security</td>
</tr>
<tr>
<td>FFV4 – Social Security</td>
<td>FFV5 – Marriage Card</td>
</tr>
<tr>
<td>FFV5 – Marriage Card</td>
<td>FFV6 – Latest</td>
</tr>
<tr>
<td>FFV6 – Latest</td>
<td>FFV1 for ID Card</td>
</tr>
</tbody>
</table>

**Base FFV DB**

**Extended FFV DB**

**FFV Extraction**

<table>
<thead>
<tr>
<th>Name</th>
<th>ID#</th>
<th>Address</th>
<th>Phone #</th>
<th>Vehicle LIC #</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFV2</td>
<td>FFV3</td>
<td>FFV4</td>
<td>FFV5</td>
<td>FFV6</td>
<td></td>
</tr>
</tbody>
</table>
In China, most ID cards are valid for 5/10/20/ years. In the past, people usually need to go back to their hometowns and wait in long lines to renew their ID cards.

ID card renewal process in the past:
- Travel back to your hometown.
- Go to the authority office & wait in a long line to get your photo & fingerprint taken & application form submitted.
- Return to your living city & wait for a notice.
- Go back to your hometown, and wait in a long line again to pick up the new ID card.
- Return to your living city.

DIY ID Card Renewal: Quick, Convenient & Money Saving

1. Enter your ID number (as an index to the ID DB).
2. Take a photo of yourself.
   - Your Facial Feature Vectors generated & compared to the ID libraries in real time.
   - The New Facial Feature Vectors are also added to the system.
3. Take your fingerprints.
   - Your fingerprints are compared to the FP DB in real time.
4. Enter your current address and phone number.
   - The address and phone number are analyzed in real time.
5. Confirm your information & pay the fees.
6. Pick up or get your new id card by mail.
   - A lot of time & money are saved.
Spring Festival: Getting Home Faster

Traditional Check-in Requires Ticket-ID-Person Matching
Facial Recognition Requires Only ID-Person Matching

3-5s Check in w/ Facial Recognition,
2X Faster than manual

Source: Southern China Morning Post

ID-Person Matching is 1-to-1 FFV Matching
(get ID picture from ID number & match to captured picture)

Smart IPC
Smart Small Station
ASIC/CPU/GPU
Smart IPC & Smart Small Station:
- Facial Feature Extraction (FFV): 200/P4
- ID Card Reading
- Ticket Reading

Access

SMART SMALL STATION

Edge DC
At Railway Station
ASIC/GPU
- Facial Feature Extraction (FFV)

Center DC
At Railway Company HQ
SUSS
IDV
TMAC
Tracking

Center DC:
GPU
- Image Recognition DL Algorithm DL training

Server

AGG/CoRE

GPU
- FFV Matching against FFV-DB (500M-Lib)

Traditional Check-in Requires Ticket-ID-Person Matching

Spring Festival:
Getting Home Faster
A missing kid could be found in minutes to hours with Facial Recognition & Real-time FFV matching, if reported in time.

Finding out a Missing Kid in Minutes

- Shopping Mall
- Street
- Railway Station
- Highway Entry
- Subway Station
- Airport
- Bus

Alert & Dispatch

Control Center DC Servers
- Report
- Submit Picture
- Facial Vector
  - N:1 FFV Matching
  - 20K-200K/s FFV Matchings: GPU
- Dispatching
- Locating
- Facial Vectors
  - 20K-200K/s
- GPU

Capture Pictures

Report
Submit Picture
Facial Vector
Dispatching
Locating
Facial Vectors
GPU

Found!
Catching a Known Suspicious Suspect in Minutes

Source: Some pictures from BBC News
Catching Red Light Violation: Reduce Traffic Jam

To Identify Red Light Violations Is a Time-consuming Many (Violation)-to-Many (DB) FFV Matching Process

Violation Detection & ID Recognition in Real Time

FFV Matching against FFV-DB at Edge DCs:
- 20M-People City & 2M for each Edge DC
- 15M FFV-Records per Edge DC, Others at Center DC
- 20K HD Cameras & 1 Violation/s/camera at Peak
  - 20K Violations/s at peak
  - 20K x 15M FFV Matchings/s = 300G FFVMs/s
  - 300G/s x 2KB = 600TB/s!! Raw Memory Bandwidth
  About 1000 V100 GPU HBM2 BW!!

Smart IPC & Smart Small Station:
- Violation Detection
- Small Image Extraction from Large(SIEL)
- Facial Feature Extraction(FFV): 200/P4

Edge DC:
- Violation Detection
- Small Image Extraction (SIEL)
- Facial Feature Extraction(FFV)
- FFV Matching against FFV-DB (15M local)

Center DC:
- FFV Matching against FFV-DB (500M-Nation)
- History Activity Tracking
- Image Recognition DL Algorithm Training

Many to Large FFV DB Matching Requests!

AGG/CoRE: Video Cloud

SUSS: Suspect Surveillance

IDV: ID Verification

TMAC: Traffic Monitoring Analysis & Control

Center DC Server

Edge DC Server

Access

Smart IPC

Smart Small Station
Huawei Heterogeneous Computing for Smarter Cities

Rich HC Product Portfolio for Smart City Video Surveillance & Intelligent Analysis: G5500/G2500/G1500
G5500: Modular Design for Quick Upgrade & Maintenance & System Reliability & Availability
G5500: High Performance & Scalable for 350 W GPU & 255 W x86 CPU & 2S+32-DIMM CPU Node
G5500: Zero-Touch Topology Change & Large NVMe SSD or HDD Storage w/o Need for External NAS
Shenzhen Smart Transport Powered By Huawei Atlas+GPU

**Safer Cities Creating a Better Life, Attracting More Talents & Investments, Promoting More Businesses**

---

**Challenges**

- Legacy devices are prone to fail and difficult to maintain. No sudden event detection function is available.
- More than 100M images per day are uploaded to be analyzed in time.
- Complex services, multiple algorithms & applications to support

**Huawei Solution**

- G5500 w/ High Performance GPU for 100M Pictures analysis & 500 HD video streams real-time structured analysis per day per 4U chassis
- Container-based deployment for traffic volume detection, red light violation & sudden event detection on the same platform, resource pooling and removal of silos of resources

---

**Customer Benefits**

- Resource pooling for multiple AI algorithms on the same hardware, reduced CAPEX & human resource investment
- Signal optimized adjustment cycle shortened from 3 months to 7 days
- Vehicle speed up by 9% for critical road segments
- Traffic jam wait time reduced by 24% in rush hours

**Project Summary**

- 3.3M vehicles, 2430 intersections & 480 vehicles per km² in Shenzhen, intelligent transportation control & management required to ensure smooth traffic
- "Huawei Atlas+GPU+FusionInsight" solution awarded due to high performance, high density, modular design, standardized & openness.

---

**Sudden Event Quick Detection**

**Redlight Violator Detection**

**Traffic Volume & Direction D&A**

**LIC Recognition & Fake LIC Detection**
Safer Cities Creating a Better Life, Attracting More Talents & Investments, Promoting More Businesses

Powerful GPU Servers for:
- 10,000s Facial Feature Vector (FFV) Extractions per Second
- 100B Rows of FFV Query + 100B Rows Structured Data Query in Seconds
- Powerful V100-based NVLink GPU Cluster for Fast DL Training

Result:
- Real-time: Facial & Vehicle LIC Recognition, Analysis & Multi-dimension Big Data Query & Analysis for data streams from 20,000 HD Cameras.
- Quick: In 2017, Shenzhen criminal cases drop 25%, 60% cases solved with surveillance cameras, & almost no case unsolved beyond 48 hours in Longguang, after using Huawei Video Cloud solutions.
Winner of Smart City EXPO World Congress 2017

This project uses a big data platform, one data resource pool and one deep learning system identifying all data to reduce traffic congestions and accidents and improve public safety. This platform allows traffic control, gathers data and enhances data usage by 200%, increasing road capacity by 8% & reducing waiting time by 24%.
Safer Cities, Powered by Heterogeneous Computing, Are Creating a Better Life, Attracting More Talents & Investments, and Promoting More Businesses