Agile Mobile Cognition
Beyond the Cloud

Augusto Vega, Chung-Ching Lin, Pradip Bose, Sharath Pankanti

This research work was sponsored in part by Defense Advanced Research Projects Agency (DARPA), Microsystems Technology Office (MTO), under contract number HR0011-13-C-0022. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.
This document is: Approved for Public Release, Distribution Unlimited.
A new era of computing…

Cognitive Systems learn and interact naturally with people to amplify what either humans or machines could do on their own. They help us solve problems by penetrating the complexity of Big Data.
The Cognitive Race Has Begun!

- **IBM Watson**: 02/11
- **Apple Siri**: 10/11
- **Google Cat**: 06/12
- **Amazon Firefly**: 06/14
- **Microsoft Adam**: 07/14
Journey to Watson: A deep foundation in computer science

Natural Language Processing
Question Answering Technology
High Performance Computing
Knowledge Representation and Reasoning
Machine Learning
Unstructured Information Management
Watson is creating a new partnership between people and computers that **enhances, scales and accelerates** human expertise.

“Watson on my shoulder makes me smarter”

**Finance**  
*Enhance decision support*

**Healthcare**  
*Surface best protocols to practitioners*

**Legal**  
*Suggest defense/prosecution arguments*

**Telemarketing**  
*Next generation – persuasive – call center*

- Engagement
- Discovery
- Decision-making
New capability: Debating Technologies

**Giving Watson the power to “reason”**

Buildin on Watson’s mastery of Jeopardy-style Q&A IBM Research has begun to train a system to construct natural language arguments

Example: generating *claims* for a given *topic*

- 4M articles in wikipedia
- \( \downarrow \)
- 100B candidate claims
- \( \downarrow \)
- 10 relevant claims
New capability: image analysis and anomaly detection

*Giving Watson the power to “see”*
1 - Healthcare

**FACTS**
- Heart disease costs $312 billion per year\(^1\)
- 25% of heart-related deaths are preventable\(^2\)

**TREND**
- Portable, personal health nurse
- Decentralized health care ("Watson on your shoulder")

**GOAL**
- Well-being improvement
- Health care cost reduction

---

2 - Safety & Security

**Mobile Cognition**
- Mobile Sensors
- Limited Power
- Limited Computing
- Limited Connectivity
- Real-Time Constraints

**FACTS**
- 24-hour, active monitoring leads to the most effective surveillance\(^3\)
- Too expensive when done by humans

**TREND**
- Fixed cameras → Mobile cameras (body, drones, cars)
- Manual analysis → Video analytics

**GOAL**
- 24-hour, ubiquitous and inexpensive public surveillance
- Privacy protection

3 - Automotive

**FACTS**
- Car crashes cost $871 billion per year\(^4\)
- Connected cars can address 80% of the crash scenarios\(^5\)

**TREND**
- Expensive sensors → Inexpensive cams
- 100% of cars will be connected by 2025
- 75% of cars will be autonomous by 2035

**GOAL**
- Safer and efficient driving at lower cost

4 - Financial

**FACTS**
- The Financial sector is one of IBM's biggest sources of revenue.

**TREND**
- Banks/insurance co.'s are always looking for ways to cut costs and deliver better service.

**GOAL**
- Accurate and real-time risk assessments for loans and insurance pay outs for customers
- Lower costs and improve risk exposure of the bank or insurance company

---

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Components of Mobile Cognition

1. Inception of Cognition
   - Lightweight Mobile Agent
   - Basic Knowledge

2. Sensory Perception
   - Raw Data Collection
   - Local Data Registration
   - Extraction of Semantic Information

3. Real-Time Reaction: Decision Making
   - Reasoning for Specific Cases
   - Learning
   - Knowledge Representation

   - Cloud-Connected Scenarios
   - Machine-2-Machine (Swarm) Scenarios
   - Interaction between Connected & Disconnected Scenarios
Use Case – Mobile Cognition for the Military

The Cognitive Drone / UAV

1 – Inception of Cognition
Initially, fundamental cognitive rules are built in the mobile agent.

2 – Sensory Perception
Sensors on the drone collects video, IR, radar, GPS, etc. data.

3 – Real-Time Reaction
The drone automatically finds real-time targets and/or fires onboard weapons.

4 – Device Interaction
The drone communicates with other drones and soldiers.

- Autonomous location-dependent application of sensors
- Reduction of human interaction with potentially high threat areas
- Based on season and weather
- Based on GPS location
- Based on intelligence

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Video summarization

Event: objects (people/vehicle) are moving

Most videos are non-eventful

To achieve large data reduction without significant loss of events of interest, we summarize UAV videos using:
  - Event: moving object detection and tracking
  - Coverage: create a panorama

Challenges:
  - Low object resolution
  - Scene diversity
  - Camera motion
  - View point change
Event Summarization

Track objects

Group tracks and ranking
  Smoothness
  Length

Decide keyframes

Overlay tracks on the keyframes
Coverage Summarization

Register and relate successive frames
Bring them to common reference frame
Create panoramas
Beyond Cognitive Cloud…Bringing it to the Edge

Cognitive Computing
Today

Mobile Cognition
For Tomorrow

1. Incept Cognition on the Edge
2. Enable Sensory Perception
3. Actuate Real-Time Reaction
4. Build Interactions Between Devices

Mobile Cognition Concept

Cognitive Rules
Knowledge DB
Machine Learning
Data Ingest
Cognitive Cloud
(heavy processing)

Secure, reliable, universal protocol
Unreliable Connection

Cognitive Agent
(light processing)

Mobile Device

- Financial
- Healthcare
- Public Safety
- Industrial
- Autonomous Car
- Military UAV/Drones

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Anatomy of Cloud Latencies

Figure 13: Cumulative distribution function (CDF) of response times in ms


Distribution Statement A: Approved for Public Release, Distribution Unlimited
# State of the Art – Motivation

## Collocated Cameras (e.g. ARGUS-IS)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>1.8 gigapixels</td>
</tr>
<tr>
<td>Raw image data rate</td>
<td>32.4 GB/s</td>
</tr>
<tr>
<td>Channel bandwidth (CDL)</td>
<td>34.25 MB/s</td>
</tr>
<tr>
<td>Compression</td>
<td>Thousand-fold on-board compression using a 32-processor unit</td>
</tr>
<tr>
<td></td>
<td>Low resilience</td>
</tr>
<tr>
<td></td>
<td>Lack of flexibility</td>
</tr>
<tr>
<td></td>
<td>Not easy to scale out</td>
</tr>
<tr>
<td></td>
<td>High power, high cost</td>
</tr>
</tbody>
</table>

### Drawbacks

- Low resilience
- Lack of flexibility
- Not easy to scale out
- High power, high cost

We argue for a decentralized, power-efficient approach, using distributed cameras (*swarm*)

- Lined up with AFRL military strategy
- Transferable to the civilian domain: For example, connected cars will rely on distributed, low-cost cameras

Source: DARPA's new 1.8-gigapixel camera is a super high-resolution eye in the sky. URL: [http://www.gizmag.com/argus-is-darpa-gigapixel-camers/26078/](http://www.gizmag.com/argus-is-darpa-gigapixel-camers/26078/)
Mobile Real-Time Video Summarization

Our Proposal
- Balances processing between the device and the cloud to meet real-time needs
- Provides fault tolerance inherently as a result of its decentralized nature
- Reduces computation and power pressure in both ends: device and cloud

In-the-device processing
- Relies heavily on wireless channel availability to meet both real-time computation and fault-tolerance requirements
- Balances processing between the device and the cloud to meet real-time needs
- Provides fault tolerance inherently as a result of its decentralized nature
- Reduces computation and power pressure in both ends: device and cloud

In-the-cloud processing
- Relies heavily on wireless channel availability to meet both real-time computation and fault-tolerance requirements
- Balances processing between the device and the cloud to meet real-time needs
- Provides fault tolerance inherently as a result of its decentralized nature
- Reduces computation and power pressure in both ends: device and cloud

ARGUS-like approach

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Our Vision

Generate real-time wide-area panoramas from a swarm of camera-equipped mobile devices

Use a highly-effective image stitching algorithm being developed at IBM
Preliminary Demo

In the short-term, we will focus on the scenario of one device (Raspberry Pi board) interacting with the back-end cloud (Linux/x86 laptop).

Real-time global panorama + Real-time partial panorama

Bandwidth-variable wireless connection

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Preliminary Demo

We focus on the real-time panorama generation in one mobile device and its different tradeoffs depending on the device-to-cloud connectivity characteristics.

\[ A = \text{camera frame rate} \]
\[ B = f(\text{stage 1 subsampling}) \]
\[ C = f(B, \text{channel bandwidth}) \]
\[(A \geq B, B \geq C)\]
Intermediate Demo (January 2015)

We focus on the real-time panorama generation in one mobile device and its different tradeoffs depending on the device-to-cloud connectivity characteristics.

\[ A = \text{camera frame rate} \]
\[ B = f(\text{stage1 subsampling}) \]
\[ C = f(B, \text{channel bandwidth}) \]

\((A \geq B, B \geq C)\)

✓ **Goal:** to provide real-time or near-real-time user experience to the ground operator

✓ In other words, to refresh the displayed panorama at or close to a rate of \(A\) fps
Preliminary Demo – Three-Tier Testbed

**USER INTERFACE (OPERATOR’S VIEW)**

- Web Browser
- Near real-time update (refresh) of the global panorama
- Different settings can be controlled by the operator
- Several statistics presented

**CLOUD**

- Video Summarization Process
- Step 3: Global panorama
- Step 2: Mini-panoramas
- Wireless Channel
- ✓ High-quality VS process i.e. small down-sampling factor
- ✓ Highly compute intensive

**MOBILE DEVICE**

- ✓ Low-quality VS process i.e. large down-sampling factor

---

**Step 1**
- Raw video

**Step 2**
- Mini-panoramas

**Step 3**
- Global panorama

**Raspberry Pi**
- (ARM11 core)

**MPEG Video File**
- Web Browser
Preliminary Demo – Three-Tier Testbed

- **Raspberry Pi Board**
- **Linux/x86 Laptop**
- **Mobile Device**
- **Web Interface**
- **Cloud**

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Preliminary Demo – Web Interface

Launch/Stop button

Progress bar

Settings:
- Bandwidth cap
- Sub-sampling step
- Total frames
- On-device VS enabled or disabled

Statistics:
- Used bandwidth
- Transferred mini-panoramas
- Frame rate (device and cloud)
- CPU usage (device and cloud)

Connection type (wireless or wired)
### Scenarios Evaluated Using the Demo Testbed

<table>
<thead>
<tr>
<th>Where VS(^1) takes place</th>
<th>Baseline 1: Entirely on the device (Raspberry Pi)</th>
<th>Baseline 2: Entirely on the cloud (Linux/x86 laptop)</th>
<th>Hybrid Approach (our proposal): Distributed across the device and the cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bandwidth</td>
<td>We connect the laptop to the RPi board through Wi-Fi with two configurations: 2 Mbps and 10 Mbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Down-sampling step</td>
<td>3 frames (VS running on Raspberry Pi)</td>
<td>3 frames (VS running on laptop)</td>
<td>3 frames for the cloud part and different steps (10 to 1000 frames) for the mobile part</td>
</tr>
<tr>
<td>Input video file</td>
<td>We use a video file from the Video and Image Retrieval and Analysis Tool (VIRAT) video dataset(^2) Duration: 05:09 min – Frame rate: 30 fps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[1] Video Summarization
Proposed Hybrid Approach: Preliminary Results

- **2 Mbps**: laptop-to-RPi Wi-Fi connection capped to 2 Mbps
- **10 Mbps**: laptop-to-RPi Wi-Fi connection capped to 10 Mbps

- **2 Mbps**: the device requires higher on-board processing (i.e. sub-sampling) when the Wi-Fi channel bandwidth is low
- **10 Mbps**: the device relies more on the cloud (larger sub-sampling) when the Wi-Fi channel bandwidth is larger

#### Frame Rate (fps)

- **2-Mbps case**: 7.7 fps
- **10-Mbps case**: 16.0 fps

Goal: Update the interface (panorama) at 30 fps
Proposed Hybrid Approach: Preliminary Results

- **2 Mbps**: laptop-to-RPi Wi-Fi connection capped to 2 Mbps
- **10 Mbps**: laptop-to-RPi Wi-Fi connection capped to 10 Mbps

Transfer overhead becomes significant at larger steps (200 or more) when BW is higher (10 Mbps).

Transfer overhead becomes dominant at small steps when BW is low (2 Mbps).

![Frame Rate (fps) Graph]

- **2-Mbps case**: 7.7 fps
- **10-Mbps case**: 16.0 fps

**Channel saturation**

- **2 Mbps**: laptop-to-RPi Wi-Fi connection capped to 2 Mbps
- **10 Mbps**: laptop-to-RPi Wi-Fi connection capped to 10 Mbps

**Computation vs. Data Transfer Costs**

- **2 Mbps**: Transfer Time Fraction, Computation Time Fraction
- **10 Mbps**: Transfer Time Fraction, Computation Time Fraction

**Distribution Statement A: Approved for Public Release, Distribution Unlimited**
Phase 1 Prototype

Through the demo we will:

- Define a quality metric $Q$ to quantify the effectiveness of the real-time global panorama generation process.

- Study the dependency of $B$ (mini-panorama rate in fps) on the sub-sampling rate (number of discarded input frames).

- Study the dependency of $Q$ on $B$ and $BW_{CHA}$ (variable wireless channel bandwidth).

- Explore control algorithms to dynamically adjust the subsampling rate to maximize $Q$ as much as possible.
Intermediate Demo (February 2015)

Stage 2 video summarization

Stage 1 video summarization

Effectiveness of the real-time global panorama generation process (to be defined)

Quality

Sub-sampling factor (Stage 1)

B > BW_{CHA}
**Hybrid Approach vs. Baselines**

- **Baseline 1:** VS takes place entirely on the RPi device
- **Baseline 2:** VS takes place entirely on the laptop

<table>
<thead>
<tr>
<th>Frame Rates</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
<th>Our Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-Mbps Case</strong></td>
<td>0.3 fps</td>
<td>0.6 fps</td>
<td>7.7 fps</td>
</tr>
<tr>
<td><strong>10-Mbps Case</strong></td>
<td>0.3 fps</td>
<td>2.1 fps</td>
<td>16.0 fps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bottlenecks</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
<th>Our Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-Mbps Case</strong></td>
<td>Limited by RPi CPU capacity</td>
<td>Limited by Wi-Fi bandwidth</td>
<td>Limited by Wi-Fi bandwidth and laptop CPU capacity</td>
</tr>
<tr>
<td><strong>10-Mbps Case</strong></td>
<td>These limitations have lower impact in our “hybrid” proposal (i.e. higher frame rates)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BW Cap: 2 Mbps

CDF

Step Size

Response Time (ms)

Distribution Statement A: Approved for Public Release, Distribution Unlimited
BW Cap: 10 Mbps

CDF

•

Distribution Statement A: Approved for Public Release, Distribution Unlimited
Discussion

• Mobile Cognition is an emerging technology with many critical applications
• To support scale and latencies required by mobile cognition requires a fundamentally different architecture
• Unlike traditional “cloud-backend” approach, we propose an approach amenable to adaptation for existing CPU, bandwidth resources
• We demonstrate efficacy of our approach with preliminary results and prototype demo.
• We plan to further validate our initial results in a real environment.
Backup
Phase 1 Prototype – Testbed

Stage 1 takes place in the device

Back-End Cloud

Stage 2 takes place in the cloud

Step 2
Mini-panoramas

Wireless Channel

Step 3
Global panorama

Summary

Step 1
Raw video

Video
Summarization

Video file

We will consider different options – for example:

- Raspberry Pi board (ARM11 core)
- NVIDIA Jetson TK1 kit
• Analyze the video
• Determine segments automatically
• Create a panorama for each segment