Towards Wearable Cognitive Assistance

Mahadev Satyanarayanan

School of Computer Science

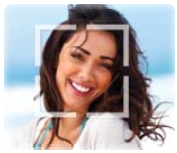
Carnegie Mellon University

Joint work with: Brandon Amos, Zhuo Chen, Benjamin Gilbert, Kiryong Ha, Jan Harkes, Martial Hebert, Wenlu Hu, Roberta Klatzky, Padmanabhan Pillai (Intel), Dan Siewiorek

http://www.istc-cc.cmu.edu/

A Unique Moment in Time

Convergence of Advances in 3 Independent Arenas



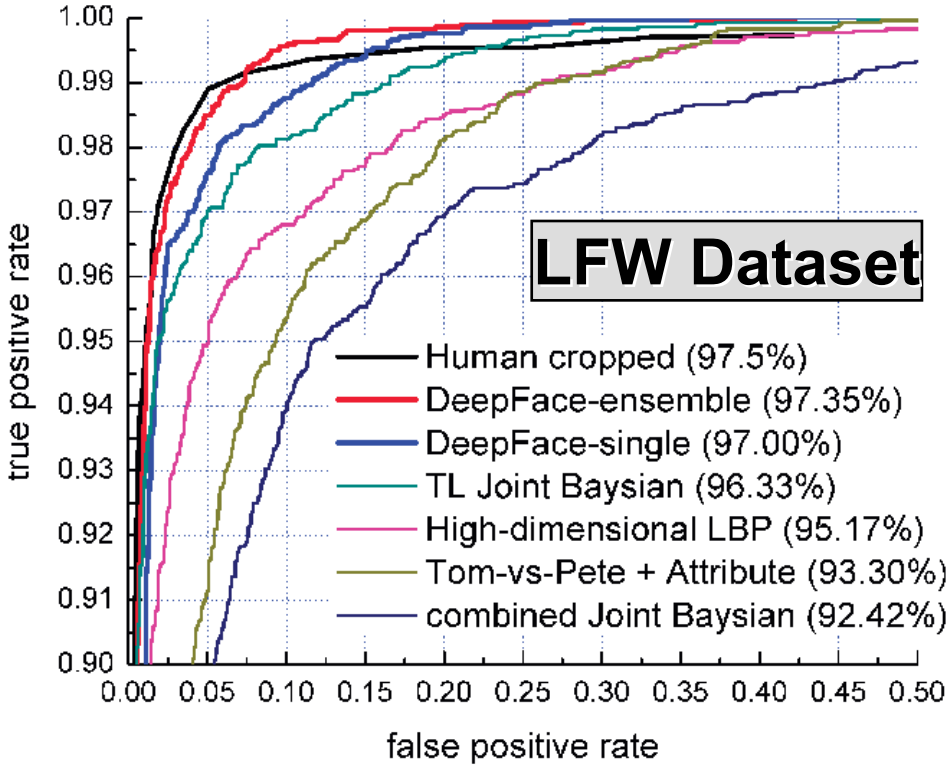
Algorithmic Advances

Cloud-Mobile Convergence

Wearable Hardware

Watson (2011)

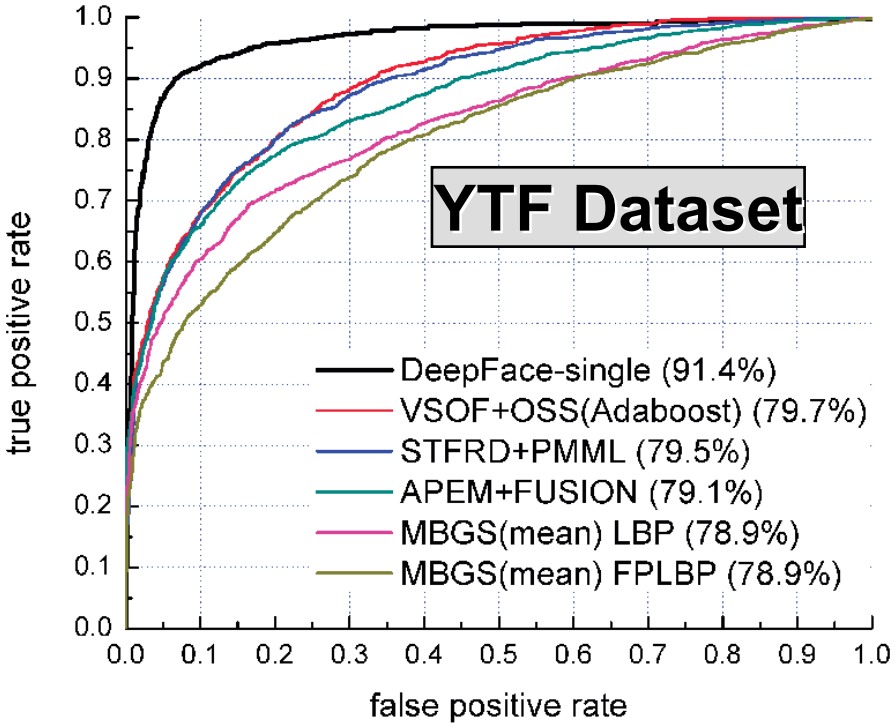
Face Recognition



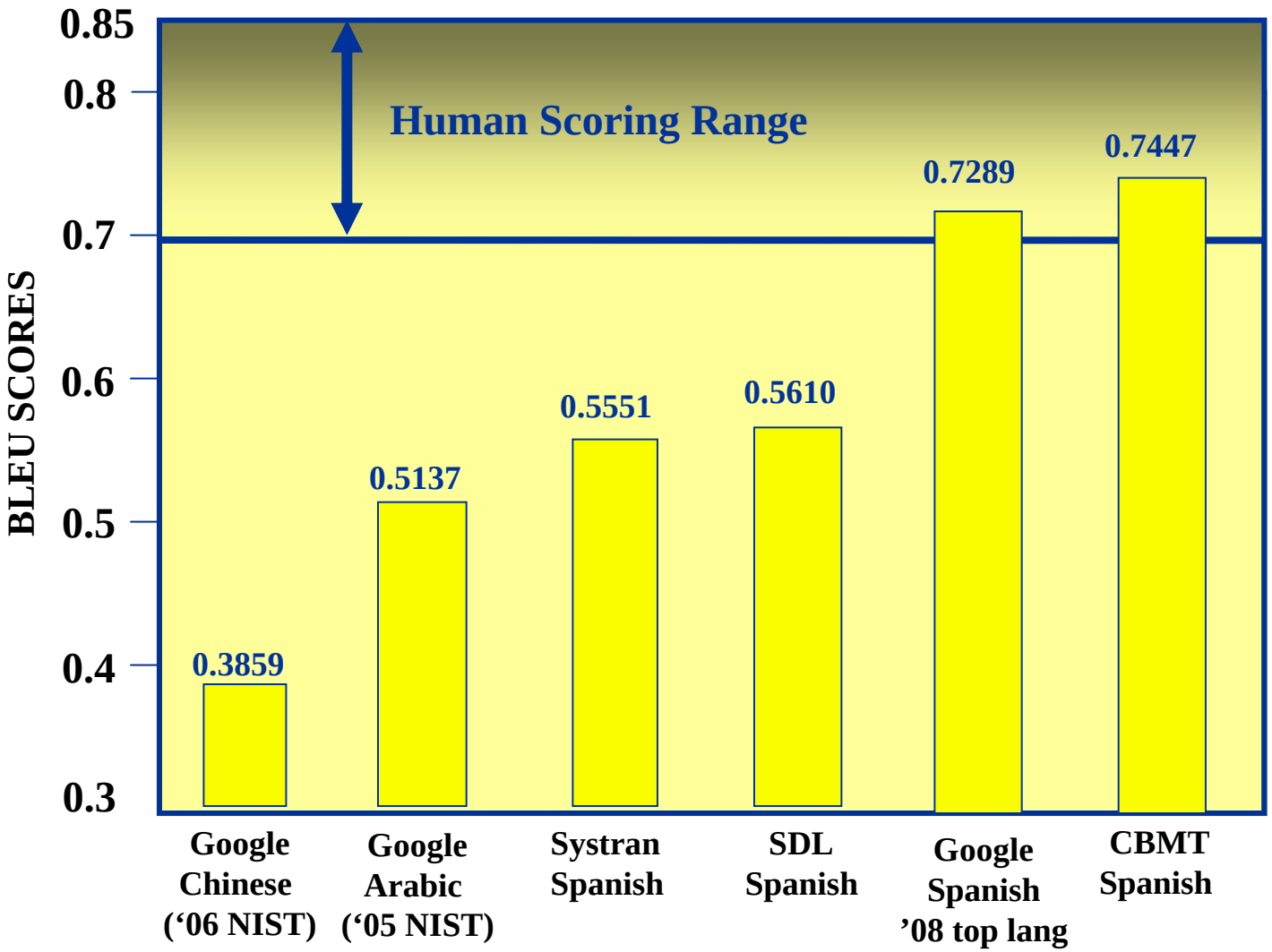
"DeepFace: Closing the Gap to Human-Level Performance in Face Verification"

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf
Proceedings of CVPR 2014 (IEEE Conference on Computer Vision and Pattern Recognition),
Columbus, OH, June 2014

Speed: 330 milliseconds per image on single-core Intel 2.2 GHz CPU



Natural Language Translation



Based on same Spanish test set →
(slides from Carbonell, 2008)

CARBONELL, J., KLEIN, S., MILLER, D., STEINBAUM, M., GRASSIANY, T., AND FREY, J. Context-based Machine Translation. In *Proc. of the 7th Conf. of the Assoc. for Machine Translation in the Americas* (Cambridge, MA, August 2006).

Convergence:
Wearable Cognitive Assistance

Entirely new genre of applications

Combine mobile and cloud with *real-time cognitive engines*

scene analysis, object/person recognition, speech recognition, language translation, planning, navigation, question-answering technology, voice synthesis, …

Seamlessly integrated into inner loop of human perception and cognition

Why?

(more use cases later in talk)

Cognitive decline

- traumatic brain injury
  (accidents, war, sports injuries, …)
- Alzheimer's disease
- survivors of stroke
- mild cognitive impairment
- …

Inability to

- recognize people
- recognize objects
- interpret text and signs
- remember daily routines
- …

Just in the United States

- over 20M Americans affected
- heavy burden on caregivers
- est. savings of $1B+ annually by 1-month delay in nursing home admission

Extrapolate to global scale!

Glass Offers Hope

Glass-based *assistive system*

- **real-time scene interpretation (vision and other sensors)**
- **offer helpful (audio) hints to user when appropriate**

Recognize faces and guides Ron to greet.

Read text from signs.

Translate when necessary.

Reminder for daily routines.

Crisp Interactive Response

Humans are amazingly fast, accurate and robust

- face detection under hostile conditions (low lighting, distorted optics) — < 700 ms
- face recognition — 370 ms – 620 ms
- is this sound from a human? — 4 ms
- VR head tracking — < 16 ms

(2004 NASA study, Ellis et al)

Not enough to just match humans

- we need to be "superhuman"
- allow enough time budget for additional cognitive processing

Safe goal: ***E2E Latency < "few tens of ms"***

Conquering Latency

Choice 1: ***standalone apps***

Choice 2: ***offload to cloud***

- **RTT is too long**
- **optimal Amazon site ~74 ms**

Choice 3: ***offload to cloudlet***

- **"data center in a box"**
- **"bring cloud closer"**
- **1-hop Wi-Fi access**
- **typical RTT < 10 ms**

OCR application on Glass

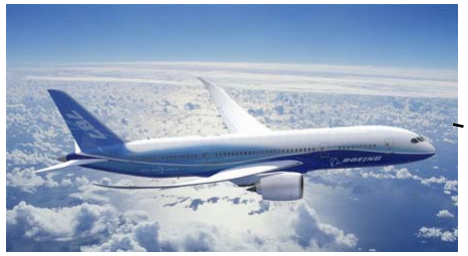
Metric	Standalone	With Offload
Per-image speed (s)	10.49 (0.23)	1.28 (0.12)
Per-image energy (J)	12.84 (0.36)	1.14 (0.11)

	Typical Server		Typical Handheld or Wearable	
Year	Processor	Speed	Device	Speed
1997	Pentium® II	266 MHz	Palm Pilot	16 MHz
2002	Itanium®	1 GHz	Blackberry 5810	133 MHz
2007	Intel® Core™ 2	9.6 GHz (4 cores)	Apple iPhone	412 MHz
2011	Intel® Xeon® X5	32 GHz (2x6 cores)	Samsung Galaxy S2	2.4 GHz (2 cores)
2013	Intel® Xeon® E5	64 GHz (2x12 cores)	Samsung Galaxy S4	6.4 GHz (4 cores)
			Google Glass OMAP 4430	2.4 GHz (2 cores)

Source: adapted from [Flinn 2013]

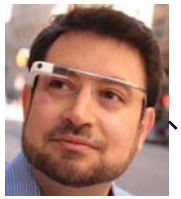
Bring the Cloud Closer

Create a Small Cloudlet Nearby



Aircraft cloudlet

Inflight Internet Service



Automobile Cloudlet

3G/4G/5G WAN

Glass

Landline WAN

Low-latency high-bw 1-hop Wi-Fi

Smartphone

Coffee Shop Cloudlet

other mobile devices

cloudlet ≅ *"data center in a box"*

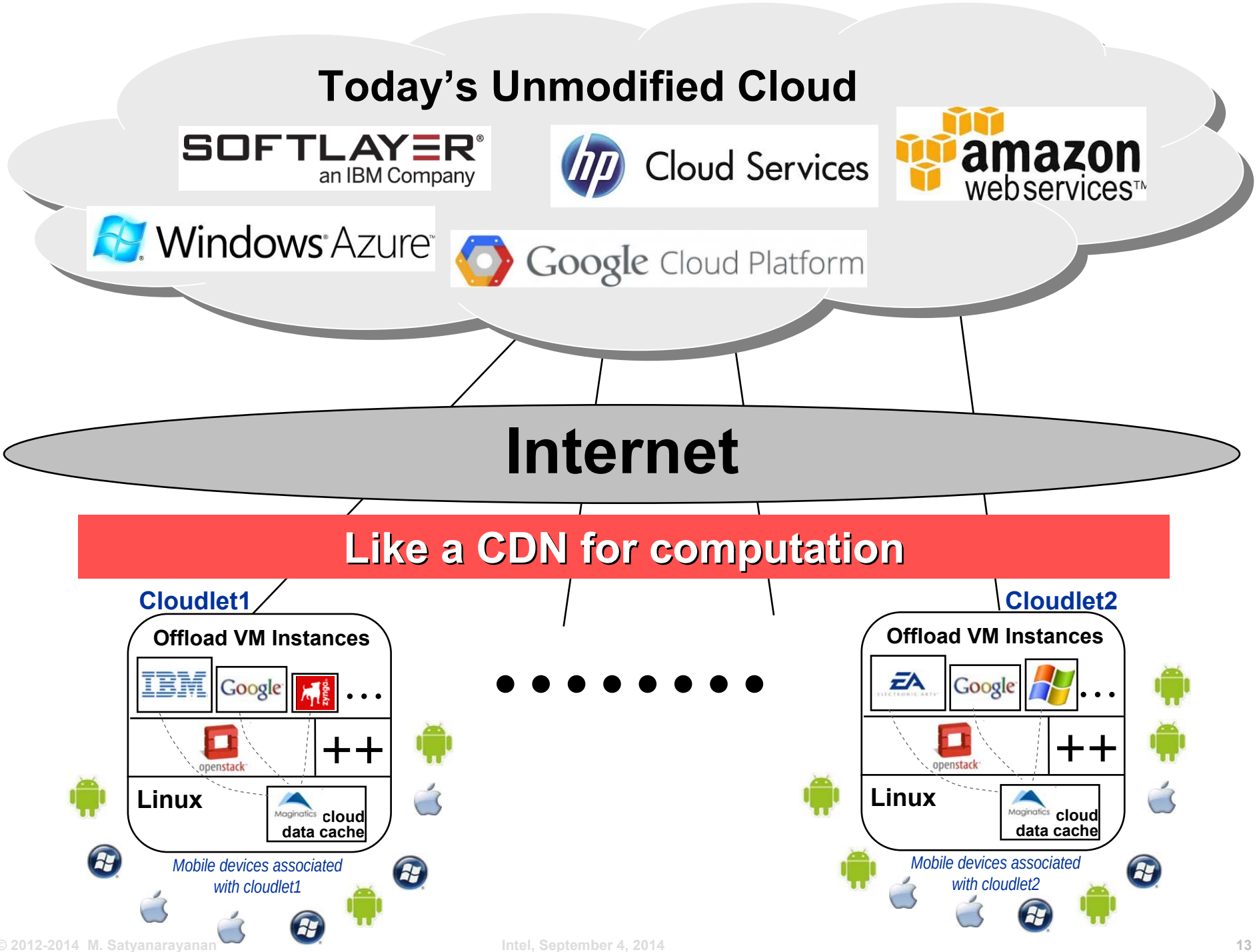
- **compute cluster**
- **"good" Internet access**
- **no battery limitations**
- **VM-based multi-tenancy**

Micro Data Centers Exist Today

Myoonet

Used as private clouds today

Need to be re-purposed as cloudlets i.e., as "2nd-tier infrastructure"


Today's Unmodified Cloud
SOFTLAYER
an IBM Company
hp Cloud Services
amazon web services
Windows Azure
Google Cloud Platform
Internet
Like a CDN for computation
Cloudlet1
Offload VM Instances
IBM
Google
openstack
++
Linux
Maginatics
cloud data cache
Mobile devices associated with cloudlet1
Cloudlet2
Offload VM Instances
Google
openstack
++
Linux
Maginatics
cloud data cache
Mobile devices associated with cloudlet2

Loosely Coupled Architecture

Many human tasks involve distinct cognitive capabilities

- **e.g. human conversation**
- **multiple independent sensor channels**
  analyzed in parallel, combined in real-time
- **strong evidence that brains use *coarse-grain parallelism***

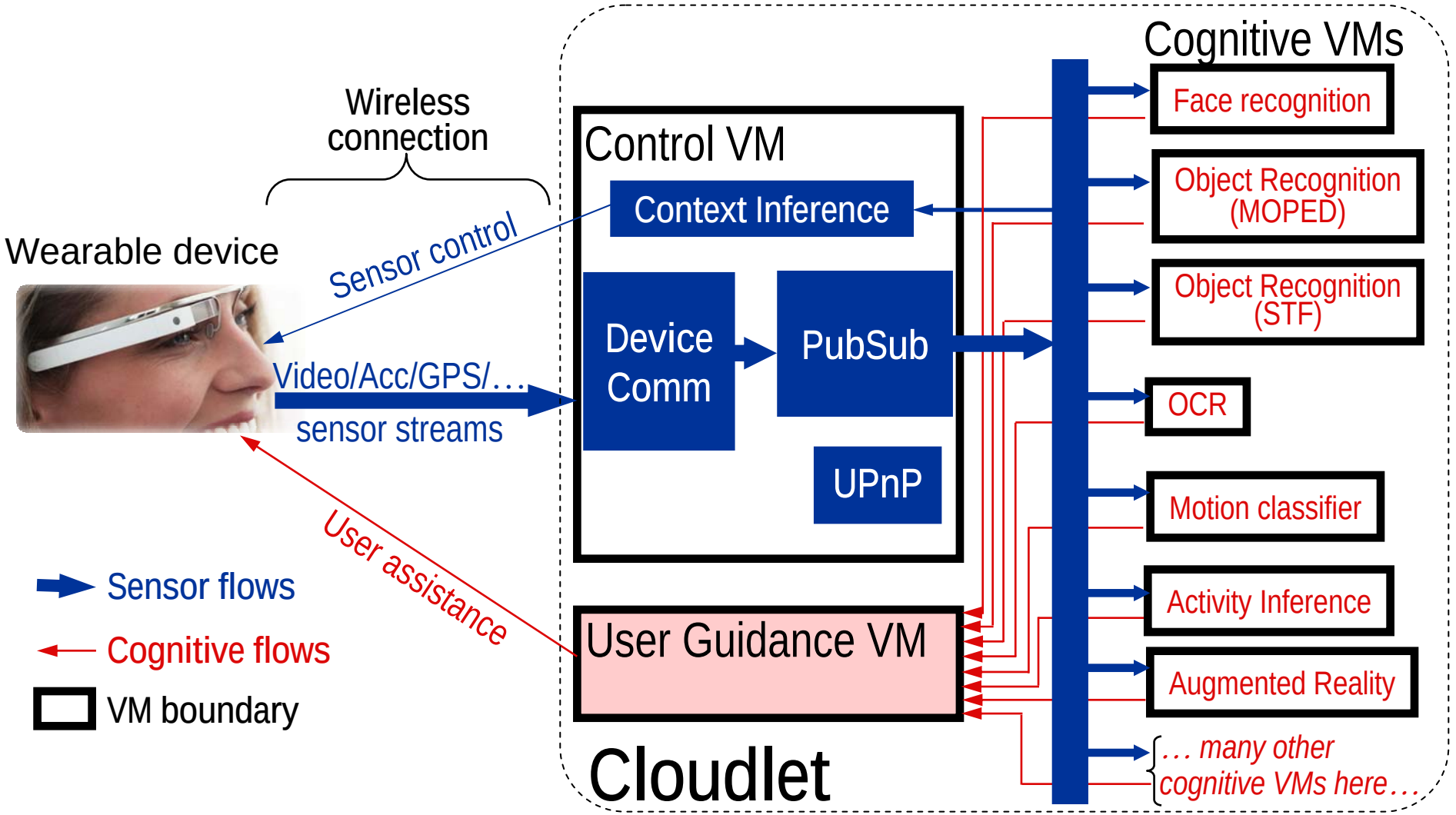
Leverage off-the-shelf building blocks

- **face / object recognition, OCR, speech-to-text, language translation …**
- **use these to catalyze our new class of applications**

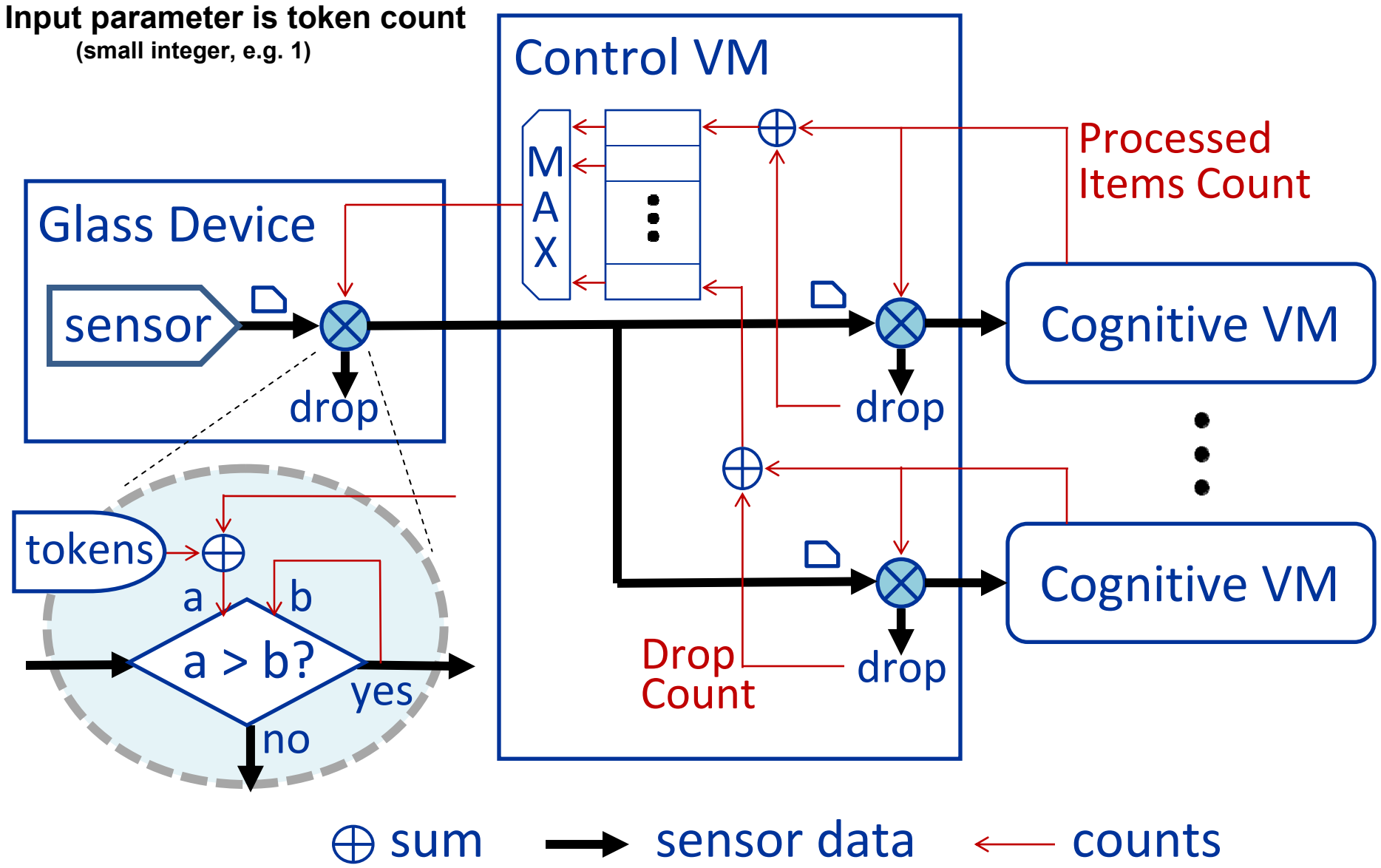
Diverse programming languages, optimizing compilers, OSes

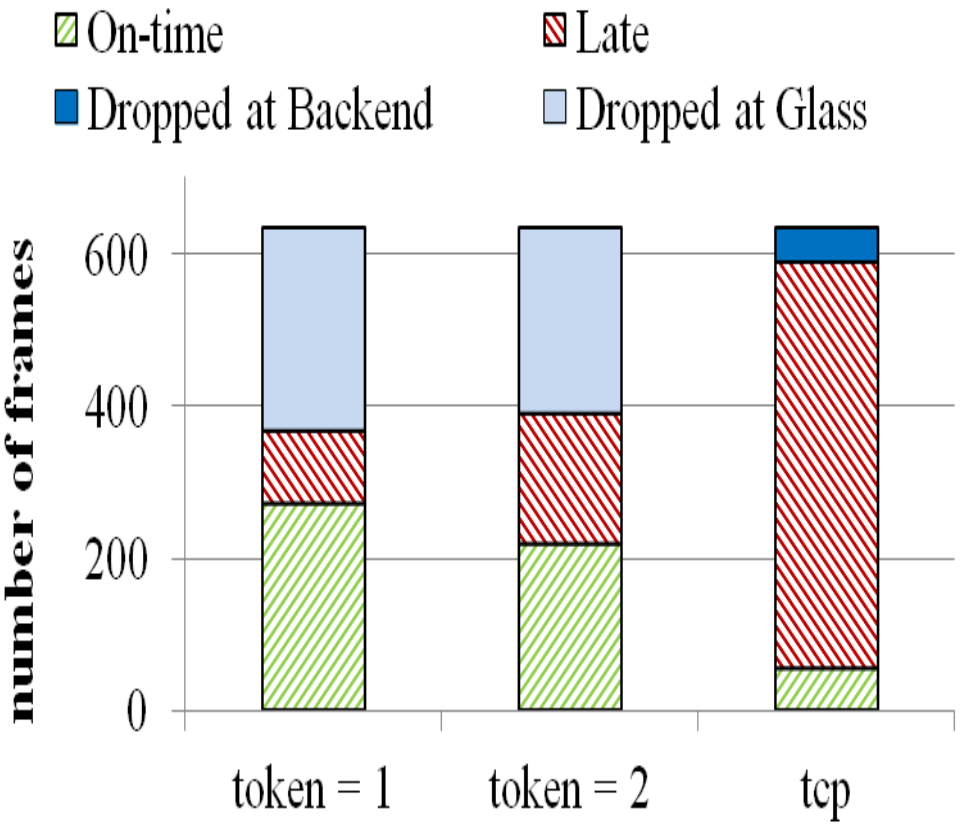
Gabriel Architecture

Ha et al, *"Towards Wearable Cognitive Assistance"*, MobiSys 2014



Don’t Transmit Hopeless Frames



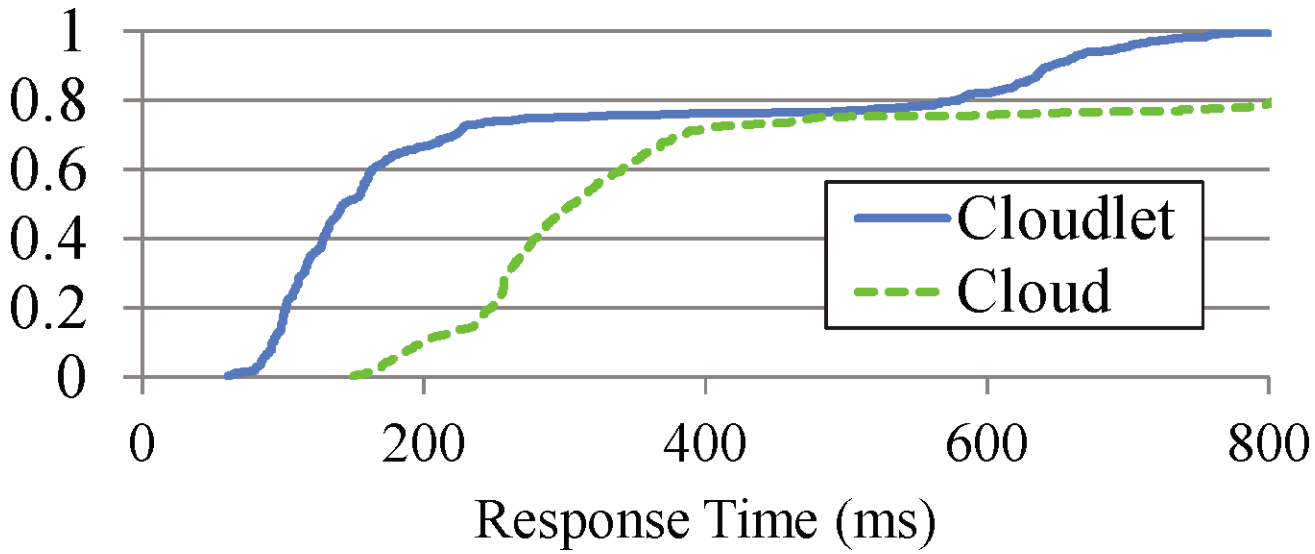

On-time
Late
Dropped at Backend
Dropped at Glass
number of frames
600
400
200
0
token = 1
token = 2
tcp

Many Results in the Paper

(just a sampling here)

Cloudlets are essential

E.g. Face recognition (cloudlet versus Amazon):



Application	Cloudlet (Joule/query)	Cloud (Joule/query)
Face	0.48	0.82
AR	0.19	0.32
OCR(open)	2.01	3.09
OCR(comm)	1.77	2.41

Figure 14: Energy Consumption on Google Glass

Performance Summary

(many more results in paper)

Cognitive Engine	Response time (ms)					Glass Life
	1%	10%	50%	90%	99%	
Face Recognition	196	389	659	929	1175	~1 hour
Object (MOPED)	877	962	1207	1647	2118	
Object (STF)	4202	4371	4609	5055	5684	
OCR (Open)	29	41	87	147	511	
OCR (Comm)	394	435	522	653	1021	
Motion Classifier	126	152	199	260	649	
Augmented Reality	48	72	126	192	498	

1. **Today's cognitive engines are slow** (none of medians are "few tens of ms")
2. **Huge inter-frame variability** (content-specific)
3. **Faster engines/frames not hurt by slower ones** (token-based flow control)

Beyond Cognitive Disabilities

Task-specific Assistance

Example: cooking

passive recipe display

versus active guidance

"Wait, the oil is not hot enough"

Inspiration: GPS Navigation Systems

Turn by turn guidance

- **Ability to detect and recover**
- **Minimally distracting to user**

Uses only one type of sensor: location from GPS

Can we generalize this metaphor?

Many Use Cases …

Assembly instructions

Industrial troubleshooting

Medical training

Strengthening willpower

Prediction

Wearable Cognitive Assistance will be the "killer app" of mobile computing in the next decade

(and we will not get there without cloudlets)

Thank You!